

3D-3D Deformable Registration and Deep Learning Segmentation based Neck Diseases Analysis in MRI

Author:

Suman, Abdulla Al

Publication Date:

2021

DOI:

<https://doi.org/10.26190/unsworks/2012>

License:

<https://creativecommons.org/licenses/by/4.0/>

Link to license to see what you are allowed to do with this resource.

Downloaded from <http://hdl.handle.net/1959.4/100102> in <https://unsworks.unsw.edu.au> on 2024-03-29

3D-3D Deformable Registration and Deep Learning Segmentation based Neck Diseases Analysis in MRI

Abdulla Al Suman

A thesis submitted in fulfilment of the requirements

for the degree of

Doctor of Philosophy



School of Engineering and Information Technology

The University of New South Wales

Canberra, Australia

October 2020



Australia's
Global
University

Thesis/Dissertation Sheet

Surname/Family Name	: SUMAN
Given Name/s	: ABDULLA AL
Abbreviation for degree as give in the University calendar	: PhD - Doctor of Philosophy
Faculty	: UNSW Canberra at ADFA
School	: Engineering and Information Technology
Thesis Title	: 3D-3D Deformable Registration and Deep Learning Segmentation based Neck Diseases Analysis in MRI

Abstract 350 words maximum: (PLEASE TYPE)

Whiplash, cervical dystonia (CD), neck pain and work-related upper limb disorder (WRULD) are the most common diseases in the cervical region. Headaches, stiffness, sensory disturbance to the legs and arms, optical problems, aching in the back and shoulder, and auditory and visual problems are common symptoms seen in patients with these diseases. CD patients may also suffer tormenting spasticity in some neck muscles, with the symptoms possibly being acute and persisting for a long time, sometimes a lifetime. Whiplash-associated disorders (WADs) may occur due to sudden forward and backward movements of the head and neck occurring during a sporting activity or vehicle or domestic accident. These diseases affect private industries, insurance companies and governments, with the socio-economic costs significantly related to work absences, long-term sick leave, early disability and disability support pensions, health care expenses, reduced productivity and insurance claims. Therefore, diagnosing and treating neck-related diseases are important issues in clinical practice.

The reason for these afflictions resulting from accident is the impairment of the cervical muscles which undergo atrophy or pseudo-hypertrophy due to fat infiltrating into them. These morphological changes have to be determined by identifying and quantifying their bio-markers before applying any medical intervention. Volumetric studies of neck muscles are reliable indicators of the proper treatments to apply. Radiation therapy, chemotherapy, injection of a toxin or surgery could be possible ways of treating these diseases. However, the dosages required should be precise because the neck region contains some sensitive organs, such as nerves, blood vessels and the trachea and spinal cord.

Image registration and deep learning-based segmentation can help to determine appropriate treatments by analyzing the neck muscles. However, this is a challenging task for medical images due to complexities such as many muscles crossing multiple joints and attaching to many bones. Also, their shapes and sizes vary greatly across populations whereas their cross-sectional areas (CSAs) do not change in proportion to the heights and weights of individuals, with their sizes varying more significantly between males and females than ages. Therefore, the neck's anatomical variabilities are much greater than those of other parts of the human body. Some other challenges which make analyzing neck muscles very difficult are their compactness, similar gray-level appearances, intra-muscular fat, sliding due to cardiac and respiratory motions, false boundaries created by intramuscular fat, low resolution and contrast in medical images, noise, inhomogeneity and background clutter with the same composition and intensity. Furthermore, a patient's mode, position and neck movements during the capture of an image create variability. However, very little significant research work has been conducted on analyzing neck muscles.

Although previous image registration efforts form a strong basis for many medical applications, none can satisfy the requirements of all of them because of the challenges associated with their implementation and low accuracy which could be due to anatomical complexities and variabilities or the artefacts of imaging devices. In existing methods, multi-resolution- and heuristic-based methods are popular. However, the above issues cause conventional multi-resolution-based registration methods to be trapped in local minima due to their low degrees of freedom in their geometrical transforms. Although heuristic-based methods are good at handling large mismatches, they require pre-segmentation and are computationally expensive. Also, current deformable methods often face statistical instability problems and many local optima when dealing with small mismatches. On the other hand, deep learning-based methods have achieved significant success over the last few years. Although a deeper network can learn more complex features and yields better performances, its depth cannot be increased as this would cause the gradient to vanish during training and result in training difficulties. Recently, researchers have focused on attention mechanisms for deep learning but current attention models face a challenge in the case of an application with compact and similar small multiple classes, large variability, low contrast and noise. The focus of this dissertation is on the design of 3D-3D image registration approaches as well as deep learning-based semantic segmentation methods for analyzing neck muscles.

In the first part of this thesis, a novel object-constrained hierarchical registration framework for aligning inter-subject neck muscles is proposed. Firstly, to handle large-scale local minima, it uses a coarse registration technique which optimizes a new edge position difference (EPD) similarity measure to align large mismatches. Also, a new transformation based on the discrete periodic spline wavelet (DPSW), affine and free-form-deformation (FFD) transformations are exploited. Secondly, to avoid the monotonous nature of using transformations in multiple stages, affine registration technique, which uses a double-pushing system by changing the edges in the EPD and switching the transformation's resolutions, is designed to align small mismatches. The EPD helps in both the coarse and fine techniques to implement object-constrained registration via controlling edges which is not possible using traditional similarity measures. Experiments are performed on clinical 3D magnetic resonance imaging (MRI) scans of the neck, with the results showing that the EPD is more effective than the mutual information (MI) and the sum of squared difference (SSD) measures in terms of the volumetric dice similarity coefficient (DSC). Also, the proposed method is compared with two state-of-the-art approaches with ablation studies of inter-subject deformable registration and achieves better accuracy, robustness and consistency.

However, as this method is computationally complex and has a problem handling large-scale anatomical variabilities, another 3D-3D registration framework with two novel contributions is proposed in the second part of this thesis. Firstly, a two-stage heuristic search optimization technique for handling large mismatches, which uses a minimal user hypothesis regarding these mismatches and is computationally fast, is introduced. It brings a moving image hierarchically closer to a fixed one using MI and EPD similarity measures in the coarse and fine stages, respectively, while the images do not require pre-segmentation as is necessary in traditional heuristic optimization-based techniques. Secondly, a region of interest (ROI) EPD-based registration framework for handling small mismatches using salient anatomical information (AI), in which a convex objective function is formed through a unique shape created from the desired objects in the ROI, is proposed. It is compared with two state-of-the-art methods on a neck dataset, with the results showing that it is superior in terms of accuracy and is computationally fast.

In the last part of this thesis, an evaluation study of recent U-Net-based convolutional neural networks (CNNs) is performed on a neck dataset. It comprises 6 recent models, the U-Net, U-Net with a conditional random field (CRF-Unet), attention U-Net (A-Unet), nested U-Net or U-Net++, multi-feature pyramid (MFP)-Unet and recurrent residual U-Net (R2Unet) and 4 with more comprehensive modifications, the multi-scale U-Net (MS-Unet), parallel multi-scale U-Net (PMSUnet), recurrent residual attention U-Net (R2A-Unet) and R2A-Unet++ in neck muscles segmentation, with analyses of the numerical results indicating that the R2Unet architecture achieves the best accuracy. Also, two deep learning-based semantic segmentation approaches are proposed. In the first, a new two-stage U-Net++ (TS-UNet++) uses two different types of deep CNNs (DCNNs) rather than one similar to the traditional multi-stage

method, with the U-Net++ in the first stage and the U-Net in the second. More convolutional blocks are added after the input and before the output layers of the multi-stage approach to better extract the low- and high-level features. A new concatenation-based fusion structure, which is incorporated in the architecture to allow deep supervision, helps to increase the depth of the network without accelerating the gradient-vanishing problem. Then, more convolutional layers are added after each concatenation of the fusion structure to extract more representative features. The proposed network is compared with the U-Net, U-Net++ and two-stage U-Net (TS-UNet) on the neck dataset, with the results indicating that it outperforms the others. In the second approach, an explicit attention method, in which the attention is performed through a ROI evolved from ground truth via dilation, is proposed. It does not require any additional CNN, as does a cascaded approach, to localize the ROI. Attention in a CNN is sensitive with respect to the area of the ROI. This dilated ROI is more capable of capturing relevant regions and suppressing irrelevant ones than a bounding box and region-level coarse annotation, and is used during training of any CNN. Coarse annotation, which does not require any detailed pixel wise delineation that can be performed by any novice person, is used during testing. This proposed ROI-based attention method, which can handle compact and similar small multiple classes with objects with large variabilities, is compared with the automatic A-Unet and U-Net, and performs best.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents a non-exclusive licence to archive and to make available (including to members of the public) my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known. I acknowledge that I retain all intellectual property rights which subsist in my thesis or dissertation, such as copyright and patent rights, subject to applicable law. I also retain the right to use all or part of my thesis or dissertation in future works (such as articles or books).

Signature

15.10.2020

Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years can be made when submitting the final copies of your thesis to the UNSW Library. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

Copyright Statement

‘I hereby grant the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all proprietary rights, such as patent rights.

I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation. I have either used no substantial portions of copyright material in my thesis or I have obtained permission to use copyright material; where permission has not been granted I have applied/will apply for a partial restriction of the digital copy of my thesis or dissertation.’

Abdulla Al Suman

UNSW, Australia

Authenticity Statement

‘I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.’

Abdulla Al Suman

UNSW, Australia

Originality Statement

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project’s design and conception or in style, presentation and linguistic expression is acknowledged.’

Abdulla Al Suman
UNSW, Australia

INCLUSION OF PUBLICATIONS STATEMENT

UNSW is supportive of candidates publishing their research results during their candidature as detailed in the UNSW Thesis Examination Procedure.

Publications can be used in their thesis in lieu of a Chapter if:

- The candidate contributed greater than 50% of the content in the publication and is the "primary author", ie. the candidate was responsible primarily for the planning, execution and preparation of the work for publication
- The candidate has approval to include the publication in their thesis in lieu of a Chapter from their supervisor and Postgraduate Coordinator.
- The publication is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in the thesis

Please indicate whether this thesis contains published material or not:

☐

This thesis contains no publications, either published or submitted for publication
(if this box is checked, you may delete all the material on page 2)

☒

Some of the work described in this thesis has been published and it has been documented in the relevant Chapters with acknowledgement
(if this box is checked, you may delete all the material on page 2)

☐

This thesis has publications (either published or submitted for publication) incorporated into it in lieu of a chapter and the details are presented below

CANDIDATE'S DECLARATION

I declare that:

- I have complied with the UNSW Thesis Examination Procedure
- where I have used a publication in lieu of a Chapter, the listed publication(s) below meet(s) the requirements to be included in the thesis.

Candidate's Name	Signature	Date (dd/mm/yy)
Abdulla Al Suman		15/10/2020

*Dedicated to my parents, My wife- Mst. Momota Yeasmin and my son-
Mohammad Shareek Afraz ...*

Acknowledgments

I could not finish this thesis without the Blessings and bountiful Sympathy from Almighty Allah, the most Benevolent and Gracious in the world.

I would like to express my heartfelt appreciation and deep gratitude to my supervisor, Professor Dr. Mark R. Pickering, for his excellent guidance, encouragement, ideas, support, patience and insightful discussions throughout my PhD journey. In spite of his busy routine, he always managed to schedule weekly meetings and helped me a lot when needed. I thank him again for his fantastic proof-reading of all my paper work. I am also extremely grateful to my joint-supervisor, Dr. Md. Asikuzzaman, and co-supervisor, Dr. Murat Tahtali, for their supports and encouragements during my candidature. It would not have been feasible to fulfill this thesis without their uninterrupted assistance.

I would like to express my sincere thank to Denise Russell for her worthy suggestions and proofreading on this dissertation, and Alexandra Louise Webb from the Australian National University, and Diana M. Perriman from Trauma and Orthopaedic Research Unit at the Canberra Hospital for providing me clinical data and annotation with motive clarification.

Also, I wish to extend my warmest appreciation and thanks to my friends and colleagues, Md. Mamunur Rashid, Md. Tanvir Istiak-Ul Haque, Md. Nazrul Islam, Dr. Monir Morshed, Md. Rasel Mahmud, Kartigeso Thanarasi, Dr. Mohammad Omar Khyam, Dr. Md. Forhad Zaman, Jayti Paul, Shabnam Saadat, Asghar Chandio, Yash Khemchandani, and Shubham Sharda for their encouragement and inspiration. In particular, Yash Khemchandani, and Shubham Sharda helped me in chapter-5 to set up deep learning framework.

I also greatly appreciate the financial support and scope to carry out my research work provided by The Australian Government, and all UNSW Canberra staff members for their cordial assistance throughout my research journey.

Last but not least, I am deeply grateful to my parents and parents-in-law for all they have given me and my wife, Mst Momota Yeasmin for her patience and understanding, my son, Mohammad Shareek Afraz for getting inspiration, brother, sister, grandparents and other close relatives for their constant encouragement, love, and prayers.

Abstract

Whiplash, cervical dystonia (CD), neck pain and work-related upper limb disorder (WRULD) are the most common diseases in the cervical region. Headaches, stiffness, sensory disturbance to the legs and arms, optical problems, aching in the back and shoulder, and auditory and visual problems are common symptoms seen in patients with these diseases. CD patients may also suffer tormenting spasticity in some neck muscles, with the symptoms possibly being acute and persisting for a long time, sometimes a lifetime. Whiplash-associated disorders (WADs) may occur due to sudden forward and backward movements of the head and neck occurring during a sporting activity or vehicle or domestic accident. These diseases affect private industries, insurance companies and governments, with the socio-economic costs significantly related to work absences, long-term sick leave, early disability and disability support pensions, health care expenses, reduced productivity and insurance claims. Therefore, diagnosing and treating neck-related diseases are important issues in clinical practice.

The reason for these afflictions resulting from accident is the impairment of the cervical muscles which undergo atrophy or pseudo-hypertrophy due to fat infiltrating into them. These morphological changes have to be determined by identifying and quantifying their bio-markers before applying any medical intervention. Volumetric studies of neck muscles are reliable indicators of the proper treatments to apply. Radiation therapy, chemotherapy, injection of a toxin or surgery could be possible ways of treating these diseases. However, the dosages required should be precise because the neck region contains some sensitive organs, such as nerves, blood vessels and the trachea and spinal cord.

Image registration and deep learning-based segmentation can help to determine appropriate treatments by analyzing the neck muscles. However, this is a challenging task for medical images due to complexities such as many muscles crossing multiple joints and attaching to many bones. Also, their shapes and sizes vary greatly across populations whereas their cross-sectional areas (CSAs) do not change in proportion to the heights and weights of individuals, with their sizes varying more significantly between males and females than ages. Therefore, the neck’s anatomical variabilities are much greater than those of other parts of the human body. Some other challenges which make analyzing neck muscles very difficult are their compactness, similar gray-level appearances, intra-muscular fat, sliding due to cardiac and respiratory motions, false boundaries created by intra-muscular fat, low resolution and contrast in medical images, noise, inhomogeneity and background clutter with the same composition and intensity. Furthermore, a patient’s mode, position and neck movements during the capture of an image create variability. However, very little significant research work has been conducted on analyzing neck muscles.

Although previous image registration efforts form a strong basis for many medical applications, none can satisfy the requirements of all of them because of the challenges associated with their implementation and low accuracy which could be due to anatomical complexities and variabilities or the artefacts of imaging devices. In existing methods, multi-resolution- and heuristic-based methods are popular. However, the above issues cause conventional multi-resolution-based registration methods to be trapped in local minima due to their low degrees of freedom in their geometrical transforms. Although heuristic-based methods are good at handling large mismatches, they require pre-segmentation and are computationally expensive. Also, current deformable methods often face statistical instability problems and many local optima when dealing with small mismatches. On the other hand,

deep learning-based methods have achieved significant success over the last few years. Although a deeper network can learn more complex features and yields better performances, its depth cannot be increased as this would cause the gradient to vanish during training and result in training difficulties. Recently, researchers have focused on attention mechanisms for deep learning but current attention models face a challenge in the case of an application with compact and similar small multiple classes, large variability, low contrast and noise. The focus of this dissertation is on the design of 3D-3D image registration approaches as well as deep learning-based semantic segmentation methods for analyzing neck muscles.

In the first part of this thesis, a novel object-constrained hierarchical registration framework for aligning inter-subject neck muscles is proposed. Firstly, to handle large-scale local minima, it uses a coarse registration technique which optimizes a new edge position difference (EPD) similarity measure to align large mismatches. Also, a new transformation based on the discrete periodic spline wavelet (DPSW), affine and free-form-deformation (FFD) transformations are exploited. Secondly, to avoid the monotonous nature of using transformations in multiple stages, a fine registration technique, which uses a double-pushing system by changing the edges in the EPD and switching the transformation’s resolutions, is designed to align small mismatches. The EPD helps in both the coarse and fine techniques to implement object-constrained registration via controlling edges which is not possible using traditional similarity measures. Experiments are performed on clinical 3D magnetic resonance imaging (MRI) scans of the neck, with the results showing that the EPD is more effective than the mutual information (MI) and the sum of squared difference (SSD) measures in terms of the volumetric dice similarity coefficient (DSC). Also, the proposed method is compared with two state-of-the-art approaches with ablation studies of inter-subject deformable registration and achieves better accuracy, robustness and consistency.

However, as this method is computationally complex and has a problem handling large-scale anatomical variabilities, another 3D-3D registration framework with two novel contributions is proposed in the second part of this thesis. Firstly, a two-stage heuristic search optimization technique for handling large mismatches, which uses a minimal user hypothesis regarding these mismatches and is computationally fast, is introduced. It brings a moving image hierarchically closer to a fixed one using MI and EPD similarity measures in the coarse and fine stages, respectively, while the images do not require pre-segmentation as is necessary in traditional heuristic optimization-based techniques. Secondly, a region of interest (ROI) EPD-based registration framework for handling small mismatches using salient anatomical information (AI), in which a convex objective function is formed through a unique shape created from the desired objects in the ROI, is proposed. It is compared with two state-of-the-art methods on a neck dataset, with the results showing that it is superior in terms of accuracy and is computationally fast.

In the last part of this thesis, an evaluation study of recent U-Net-based convolutional neural networks (CNNs) is performed on a neck dataset. It comprises 6 recent models, the U-Net, U-Net with a conditional random field (CRF-Unet), attention U-Net (A-Unet), nested U-Net or U-Net++, multi-feature pyramid (MFP)-Unet and recurrent residual U-Net (R2Unet) and 4 with more comprehensive modifications, the multi-scale U-Net (MS-Unet), parallel multi-scale U-Net (PMS-Unet), recurrent residual attention U-Net (R2A-Unet) and R2A-Unet++ in neck muscles segmentation, with analyses of the numerical results indicating that the R2Unet architecture achieves the best accuracy. Also, two deep learning-based semantic segmentation approaches are proposed. In the first, a new two-stage U-Net++ (TS-UNet++) uses two different types of deep CNNs (DCNNs) rather than one similar to the traditional multi-stage method, with the U-Net++ in the first stage and the U-Net in the second. More convolutional blocks are added

after the input and before the output layers of the multi-stage approach to better extract the low- and high-level features. A new concatenation-based fusion structure, which is incorporated in the architecture to allow deep supervision, helps to increase the depth of the network without accelerating the gradient-vanishing problem. Then, more convolutional layers are added after each concatenation of the fusion structure to extract more representative features. The proposed network is compared with the U-Net, U-Net++ and two-stage U-Net (TS-UNet) on the neck dataset, with the results indicating that it outperforms the others. In the second approach, an explicit attention method, in which the attention is performed through a ROI evolved from ground truth via dilation, is proposed. It does not require any additional CNN, as does a cascaded approach, to localize the ROI. Attention in a CNN is sensitive with respect to the area of the ROI. This dilated ROI is more capable of capturing relevant regions and suppressing irrelevant ones than a bounding box and region-level coarse annotation, and is used during training of any CNN. Coarse annotation, which does not require any detailed pixel wise delineation that can be performed by any novice person, is used during testing. This proposed ROI-based attention method, which can handle compact and similar small multiple classes with objects with large variabilities, is compared with the automatic A-Unet and U-Net, and performs best.

Contents

Copyright Statement	i
Authenticity Statement	ii
Originality Statement	iii
Dedication	iv
Acknowledgements	v
Abstract	vii
List of Figures	xvii
List of Tables	xxv
List of Abbreviations	xxviii
1 Introduction	1
1.1 Necessity for Neck Muscles Analysis	3
1.2 Roles of Medical Image Registration and Deep Learning-based Segmentation	4
1.3 Challenges of and Motivation for Current Research	6
1.4 Contributions of This Research	10
1.5 Organization of Thesis	13
1.6 List of Publications and Award	14
2 Background and Related Work	16

2.1	Computer Aided Diagnosis (CAD) Techniques	16
2.2	Image Registration	17
2.2.1	Geometrical Transformations	18
2.2.1.1	Parametric Models	19
2.2.1.2	Non-parametric Models	22
2.2.2	Similarity Measures	25
2.2.2.1	Feature-based Methods	25
2.2.2.2	Intensity-based Methods	28
2.2.2.3	Hybrid Methods	32
2.2.3	Optimization	34
2.2.3.1	Continuous Optimization Methods	34
2.2.3.2	Discrete Optimization Methods	36
2.2.3.3	Miscellaneous Optimization Methods	37
2.3	Deep Learning-based Segmentation	37
2.3.1	Convolutional Neural Network (CNN)	37
2.3.2	Fully Convolutional Network (FCN)	39
2.3.3	U-Net	40
2.3.4	Convolutional Residual Networks (CRNs)	42
2.3.5	Recurrent Neural Networks(RNNs)	43
2.3.6	Attention-based CNNs	44
2.4	Summary	47
3	A Deformable 3D-3D Registration Framework using the Discrete Periodic Spline Wavelet and Edge Position Difference	50
3.1	Introduction	51
3.2	Related Work	54

3.3	Deformable 3D-3D Registration Model	57
3.3.1	Overview	57
3.3.1.1	Coarse Section	58
3.3.1.2	Fine Section	59
3.3.2	Geometrical Transformation	60
3.3.2.1	Discrete Periodic Spline Wavelet	61
3.3.2.2	Free Form Deformation	63
3.3.3	EPD Similarity Measure	66
3.3.3.1	Modified Three-Dimensional Chamfer Distance Transform Algorithm	69
3.3.4	Transformation Parameters Optimization	69
3.4	Experimental Procedure and Result Analysis	72
3.4.1	Data and Annotations	72
3.4.2	Evaluation Metric	75
3.4.3	Numerical Results Analysis	76
3.4.3.1	Comparison of the EPD, MI and SSD Similarity Measures	77
3.4.3.2	Performance Comparison	79
3.5	Discussion	84
3.6	Conclusion	87
4	Two-stage Heuristic Search and Region of Interest-based Edge Position Difference Similarity Measure for Neck MRI Registration	90
4.1	Introduction	91
4.1.1	Handling Large Mismatches	91
4.1.2	Handling Small Complex Mismatches	93

4.1.2.1	Spatially Region-weighted Scheme	94
4.1.2.2	Encoding Objects' Local Positions and Shapes . . .	95
4.2	Proposed Registration Framework	98
4.2.1	Overview	98
4.2.2	Two-stage Heuristic Search-based Optimization	99
4.2.3	ROI-based EPD Registration	104
4.3	Experiments and Results	109
4.3.1	Data and Pre-processing	109
4.3.2	Results Analysis	111
4.3.3	Computational Complexity	114
4.4	Discussion and Conclusion	116
5	Deep Learning-based Neck Muscles Segmentation	120
5.1	Introduction	122
5.1.1	Evaluation of U-Net CNN Approaches	122
5.1.2	Multi-stage-based Deep Learning	123
5.1.3	ROI-based Attention	125
5.2	Methods and Materials	129
5.2.1	Image Acquisition and Pre-processing	129
5.2.2	Calculation of Population Average	131
5.2.2.1	Registration Model	132
5.2.3	DCNNs for Evaluation	133
5.2.4	TS-UNet++	135
5.2.5	Dilated-ROI Attention	138
5.2.6	Implementation	140

5.3	Results and Analyses	140
5.3.1	Evaluations	141
5.3.2	TS-UNet++	141
5.3.3	Dilated-ROI Attention	144
5.4	Discussion and Conclusion	145
5.4.1	Evaluations	145
5.4.2	TS-UNet++	148
5.4.3	Dilated-ROI Learning	148
6	Conclusions and Future Directions	149
6.1	Conclusions	149
6.2	Future Works	152
	Bibliography	154

List of Figures

1.1	Challenges involved with neck muscles in MRI: compact and similar objects (left); background clutter (2 nd from left); intra-muscular fat (3 rd from left); and six neck muscles considered in this dissertation denoted as M1, M2, M3, M4, M5 and M6, respectively (right). . . .	7
1.2	Neck MRI (composite form - anatomical variabilities between two individuals overlaid in green and magenta bands; gray regions - both images have same intensity; and green and magenta regions - images have different intensities.	7
2.1	Schematic of image registration process	18
2.2	Diagram of operational flow of image registration algorithm	19
3.1	Proposed hybrid registration framework in which gradient descent (GD) and Gauss-Newton gradient descent (GN GD) optimization techniques are used. The same set of edges is used in the coarse section by the different registration stages. Different sets of edges are used in the fine section by the different stages. The geometrical transformations are separable (not composite), since each stage takes input from the previous stage. The labels Affine-EPD, DPSW-EPD, Coarse-EPD and Fine-EPD are used to describe the experimental results in Section 3.4.	58

3.2	(a) 1D B-spline and (b) spline wavelet. The supports of the B-spline and spline wavelet are 4 and 3, respectively.	64
3.3	(a) 1D cosine and (b) periodic spline wavelet within an image support of $128 \times 128 \times 128$. Basis functions in discrete cosine $\varphi_k(x_i, y_i, z_i) = \cos\left(\frac{(2x_i+1)\pi u}{2M}\right) \cos\left(\frac{(2y_i+1)\pi v}{2N}\right) \cos\left(\frac{(2z_i+1)\pi w}{2O}\right)$. In discrete periodic spline wavelet, $\varphi_k(x_i, y_i, z_i) = \psi_\lambda(xu)\psi_\lambda(yv)\psi_\lambda(zw)$, where M , N and O are the volumetric image's dimensions. The wave and wavelet are at the coarsest scale resolution level for the image support. . . .	64
3.4	Registration results for two magnetic resonance imaging volumes using discrete cosine- and discrete periodic spline wavelet-based basis functions, showing a stretching effect in one slice of the former. The edges of the fixed image are superimposed in red over the unregistered and registered moving images. The registration is performed in 3D-3D, but the results are shown in 2D by taking one image from the 3D volume.	65
3.5	(a) Magnetic resonance imaging slice. (b) Edge image of (a). (c) Chamfer distance image of (b). The binary and Chamfer distance images are calculated for a 3D image but the results are shown in 2D.	68
3.6	(a) Auxiliary schematic diagram for modified 3D Chamfer distance algorithm (voxels of the same color use the same mask modified from the original [1], with O slices in volume and each red number meaning a special condition corresponding to the modified mask). (b) Mask for special condition 3 of (a). The masks are used during traversing for the specified slices in the text boxes.	68

3.7	Annotated contours of an axial magnetic resonance image. Different colors are used for each separate muscle due to annotation convenience; however, the same color is used for symmetric muscles.	74
3.8	Comparison of performances of the edge proposition difference (EPD), mutual information (MI) and sum of squared difference (SSD) similarity measures from the affine registration experiments. (a) Registration accuracies of 110 cases for each muscle (higher volumetric dice similarity coefficient [DSC] values are better). (b) Registration accuracies of 110 cases \times 6 muscles = 660 DSCs for all muscles combined. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients in the neck dataset.	74
3.9	Axial deformation fields and edge maps of an inter-subject case for different sets of edges using different sets of thresholds with different resolution levels of the spline: (a) edge map with lower and higher thresholds pair of 0.08 and 0.7, (b) deformation field of (a) using 4 th resolution level of spline, (c) edge map with thresholds pair of 0.04 and 0.4, (d) deformation field of (c) using 3 rd resolution level of spline.	79

3.10	Axial visual results of one of the 110 inter-subject registration cases to show the effectiveness of every stage of the proposed framework, as shown in Figure 3.1, in terms of the volumetric DSC value. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with superimposed fixed image's edges; superimposed fixed image's edges on the moving image after (d) Affine-EPD stage, (e) DPSW-EPD stage, (f) Coarse-EPD stage, (g) Fine-EPD stage and (h) final stage (proposed method).	81
3.11	Coronal visual results for the same case shown in Figure 3.10. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with superimposed fixed image's edges; superimposed fixed image's edges on the moving image after (d) Affine-EPD stage, (e) DPSW-EPD stage, (f) Coarse-EPD stage, (g) Fine-EPD stage and (h) final stage (proposed method).	82
3.12	Comparison of visual results in terms of muscles contours among proposed, Demons [2] and SyN [3] methods with ground truths: (a) axial view and (b) coronal view.	83

3.13	Comparison of performances of proposed, diffeomorphic Demons [2], SyN [3], Coarse-EPD and Fine-EPD methods. (a) Registration accuracies of 110 cases for each muscle (higher volumetric dice similarity coefficient [DSC] values are better). (b) Registration accuracies of 110 cases \times 6 muscles = 660 DSCs for all muscles combined. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients in our neck dataset.	85
3.14	Comparison of performances of proposed, diffeomorphic Demons [2], SyN [3], Coarse-EPD and Fine-EPD methods on our neck dataset in hausdorff distance (HD) in mm (lower values are better) for all muscles combined for 110 cases \times 6 muscles = 660 HDs. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients. .	88
4.1	Proposed inter-subject framework of MRI-to-MRI registration based on two-stage heuristic search and ROI EPD. These two main components indicated by red boxes, with MRI slices used for only presentation purposes as they are MRI volumes.	98
4.2	Two-stage heuristic search-based optimization. The stages marked by red boxes use same transformation (W_1). The translations T_x^{MI} and T_y^{MI} are varied between main hypothesis values (T_x^{mh} and T_y^{mh} respectively) with step size S_{mh} and translations T_x^{EPD} and T_y^{EPD} are varied between sub-hypothesis values (T_x^{sh} and T_y^{sh} respectively) with step size S_{sh} . Conditions for main and sub-hypotheses $T_x^{mh} > T_x^{sh}$ and $T_y^{mh} > T_y^{sh}$ respectively with step condition $S_{mh} > S_{sh}$. . .	101

4.3	Proposed ROI-based EPD registration process. Although two-dimensional (2D) images are shown in flowchart for representational convenience, all operations are performed in 3D MRI with gradient descent optimization and FFD geometrical transformation.	105
4.4	Comparison of axial visual results of one of the 342 inter-subject cases. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with fixed image's edges superimposed; fixed image's edges superimposed on the moving image after (d) Proposed, (e) Demons, (f) SyN. The Demons and SyN are state-of-the-art deformable registration methods.	113
4.5	Comparison of coronal visual results of one of the 342 inter-subject cases. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with fixed image's edges superimposed; fixed image's edges superimposed on the moving image after (d) Proposed, (e) Demons, (f) SyN. The Demons and SyN are state-of-the-art deformable registration methods.	114
4.6	Comparison of visual results in terms of muscles contours among proposed, Demons [2] and SyN [3] with ground truths: (a) axial view and (b) coronal view	115

4.7	Registration results using the proposed algorithm, without heuristic, without ROI, Demons [2] and SyN [3] on the neck MRI data in volumetric dice similarity coefficient (DSC) (higher values are better). The without heuristic and without ROI versions refer to the proposed framework excluding the two-stage heuristic search-based optimization and ROI-based EPD registration respectively. (a) all muscles separately for the 342 test cases; (b) all muscles combined for 342 cases \times 6 muscles = 2052 DSCs.	115
4.8	Registration results using the proposed, without heuristic, without ROI, Demons [2] and SyN [3] algorithms. For the Hausdorff distance (HD) in mm (lower values are better). The without heuristic and without ROI methods consist of the proposed framework after excluding the two-stage heuristic search-based optimization and ROI-based EPD registration respectively. (a) all muscles separately for the 342 test cases; (b) all muscles combined for 342 cases \times 6 muscles = 2052 HDs.	116
5.1	Anatomical variations in the neck dataset (images of different individuals)	130
5.2	Axial MRIs at the top of the C3 spinal level showing inhomogeneity correction, population average and ground truth	130
5.3	Registration framework for generating population average	131
5.4	The two-stage U-Net++ (TS-UNet++) architecture with a multi-level fusion structure (MLFS) used in stage 2	135

5.5	Multi-level fusion structure (MLFS) integrated with U-Net (feature information encoded from U-Net++ and U-Net in multi-scale fashion)	137
5.6	Types of images in the training dataset (Dilation pixels (DPs) indicate the widths of the structuring elements for the morphological dilation operation)	140
5.7	Visual segmentation results for MRIs and ground truths for A-Unet, R2Unet, U-Net++ and PMS-Unet models (columns represent different patients denoted as P1, P2, P3, P4 and P5), with extracted contours from ground truths indicated by gray lines superimposed on the automatic segmentation results	143
5.8	Visual results for multi-class neck muscles segmentations of five different patients (P1, P2, P3, P4 and P5) (first row original images, second corresponding manual segmentations and following automatic segmentations of two-stage U-Net++ (TS-UNet++), two-stage U-Net (TS-UNet) [4], U-Net++ [5] and U-Net [6], with contours extracted from ground truths denoted by gray lines superimposed on automatic segmentation results)	145
5.9	Visual segmentation results obtained from the dilated-ROI attention experiments for five different patients (P1, P2, P3, P4 and P5) (first row original images, second corresponding manual segmentations and following rows automatic segmentations of U-Net without attention, with dilated-ROI attention by 15 DPs, with dilated-ROI attention by 20 DPs and the A-Unet, respectively, with contours extracted from ground truths denoted by gray lines superimposed on the automatic segmentation results)	147

List of Tables

3.1	Demographic data and magnetic resonance imaging scanner parameters.	73
3.2	Results obtained for the edge proposition difference (EPD), mutual information (MI) and sum of squared difference (SSD) measures from the affine registration experiments in terms of volumetric dice similarity coefficient (DSC) (DSC values for patient-1 [PT-1] fixed and others considered moving; left and right sternocleidomastoid, left and right semispinalis capitis, and left and right splenius capitis muscles denoted as Muscles 1, 2, 3, 4, 5 and 6, respectively, where higher values indicate better performances, with maximum value 1, and best DSC in each column marked in bold).	75
3.3	Results obtained for the edge position difference (EPD), mutual information (MI) and sum of squared difference (SSD) from the affine registration experiments in terms of volumetric dice similarity coefficient (DSC) (mean DSC values for all fixed volumes calculated over all moving ones; values in the first column were obtained from Table 3.2 by taking row-wise mean; bold values indicate better results).	76

3.4	Parameter settings for SyN [3] and Demons [2] registrations in our neck dataset. The SyN and Demons registrations were conducted through ANTs and MATLAB, respectively. Multistage registrations were used in the SyN method, in which the rigid, affine and SyN geometrical transformations used the same smoothing sigmas, shrink factors, and convergence, except that SyN used different convergence. The metric CC means cross-correlation.	78
3.5	Registration results obtained from inter-patient neck magnetic resonance imaging experiments in terms of volumetric dice similarity coefficient (DSC) (proposed method compared with diffeomorphic demons [2], SyN [3], Coarse-EPD, and Fine-EPD; mean DSC values for all fixed volumes calculated over all moving ones; bold values indicate better results).	84
3.6	Computational time to register our dataset using the EPD, SSD, MI, proposed, Demons [2] and SyN [3] algorithms on our computer for a single case. The affine transformation is used for the EPD, SSD and MI registration. The MI and SyN [3] are implemented through advanced normalization tools (ANTs).	88
4.1	Parameter settings for symmetric image normalization (SyN) [3] and D.Demons [2] registrations on the neck dataset conducted through ANTs and MATLAB respectively. Multi-stage registrations were used in the SyN method in which the rigid, affine and SyN geometrical transformations used the same smoothing sigma values, shrink factors and convergence.	110

4.2	Overall registration performance in the neck MRI dataset using the proposed, without heuristic, without ROI, proposed in Chapter 3, Demons [2] and SyN [3] methods over 2052 samples.	117
4.3	Computational time to register two volumes using the proposed, proposed in Chapter 3, Demons [2] and SyN [3] algorithms.	118
5.1	Demographics data and parameters of the MRI scanner.	126
5.2	Average DSC and DHD values for models with best results in the ‘Overall’ column in bold	142
5.3	Comparison of performances of two-stage U-Net++ (TS-UNet++) and baseline CNNs with best results in the ‘Overall’ column in bold	144
5.4	Average DSC and DHD values for U-Net without attention [6], with dilated-ROI attention by 15 DPs, dilated-ROI attention by 20 DPs and automatic attention (A-Unet) [7], with best results in the ‘Overall’ column in bold	146

List of Abbreviations

Abbreviations	Description
CD	Cervical Dystonia
2D	2-dimensional
3D	3-dimensional
CMI	Conditional Mutual Information
CSA	Cross Sectional Area
DSC	Dice Similarity Coefficient
HD	Hausdorff Distance
DHD	Directional Hausdorff Distance
eSSD	Entropy sum-of-Squared Difference
GUI	Graphical User Interface
MRI	Magnetic Resonance Imaging
CT	Computed Tomography
MI	Mutual Information
SSD	Sum of Squared Difference
NMI	Normalized Mutual Information
ROI	Region of Interest
SCV	Sum of Conditional Variance
WADs	Whiplash Associated Disorders
DPSW	Discrete Periodic Spline Wavelet
EPD	Edge Position Difference
FFD	Free-From-Deformation
SM	Similarity Measure
ICP	Iterative Closest Point

DC	Discrete Cosine
GD	Gradient Descent
ANTs	Advanced Normalization Tools
CNN	Convolutional Neural Network

Chapter 1

Introduction

Whiplash, cervical dystonia (CD), work-related upper limb disorder (WRULD) and neck pain are the most common diseases in the cervical region [8–10]. Whiplash covers a range of injuries caused by the sudden acceleration and deceleration associated with a motor vehicle accident. Its symptoms include neck pain, stiffness, headaches, aches in the back and shoulder, sensory disturbances to the legs and arms, and auditory and visual problems [11]. CD is another prevalent neurological condition in the cervical musculature which causes tormenting spasticity in some muscles [10]. The estimated prevalence of significant episodes of neck pain in an affected person’s lifetime is 40 to 70 per cent [12]. Furthermore, neck pain results in significant socio-economic costs related to healthcare expenses, work absences, reduced productivity and insurance claims [13–15]. It also affects governments and other industries which have to pay sufferers long-term sick leave as well as early disability and disability support pensions. More than 430,000 people claimed compensation for whiplash injuries in the United Kingdom in 2007 [14]. Another study by the state insurance regulatory authority in NSW, Australia, revealed that, since 2007, 46% of compulsory third-party health insurance claims in that state were for whiplash-associated disorders (WADs) [15], with almost half these patients never completely recovering and almost a quarter suffering a chronic pain-related disability [16]. Patients may present with either acute pain, particularly as a result of trauma such as that experienced in a motor vehicle accident, or more chronic pain. The cervical muscles of a whiplash patient undergo pseudo-hypertrophy or

atrophy in different inter-vertebral levels [9, 17]. The cost of WRULDs has been estimated at between 0.2 and 0.5 percent of the gross national products (GNPs) of countries in the EU [18]. As neck pain is rated the 4th most serious of 291 disorders in the Global Burden of Disease Study (2010) [19], in order to apply appropriate medical interventions, muscle modeling is required to identify the locations of morphological variations in the cervical region which contains many other sensitive organs, such as blood vessels, nerves, the trachea and spinal cord. Therefore, the diagnosis and treatment of neck-related diseases are important in clinical practice.

Although image registration for the neck region and segmentation of the neck muscles will help a doctor analyze these muscles in order to apply proper medical intervention, none of the existing registration and segmentation methods can handle the related challenges, such as the anatomical variability, compactness, intra-muscular fat and background clutter in medical images and their low resolution and contrast. Therefore, this dissertation focuses on designing image registration algorithms for neck regions and segmentation methods for neck muscles to overcome these issues.

The remainder of this chapter is organized as follows: in Section 1.1, the reasons for analyzing neck muscles are explained; in Section 1.2, the roles of medical image registration and deep learning-based semantic segmentation are described; in Section 1.3, the challenges of and motivation for this research are discussed; in Section 1.4, the contributions of this study are provided; in Section 1.5, an outline of this thesis is presented; and, in Section 1.6, the publications related to this dissertation are listed.

1.1 Necessity for Neck Muscles Analysis

In some types of neck pain, although the cervical muscles appear on magnetic resonance (MR) images to undergo pseudohypertrophy due to fat infiltrating into them or atrophy, these changes are inconsistent between the muscles and vertebral levels [8, 9, 17]. Elliott *et al.* [9] acquired relatively larger cross-sectional areas for the semispinalis capitis, multifidus, splenius capitis, sternocleidomastoid and deep cervical flexors' cervical muscles, and smaller ones for the semispinalis capitis and semispinalis cervicis muscles from whiplash patients than healthy individuals. Bismil and Bismil [20] found that a whiplash event may damage the trapezius muscle through muscle contraction. In clinical practice, physicians often have to identify which neck muscle is affected and where to apply chemotherapy, radiotherapy or laser therapy for chronic neck pain and other neck-related disorders. However, these treatments require high precision in terms of focusing on only the affected regions, otherwise healthy cells and organs may be damaged. The muscles require the injection of a botulinum toxin into them to decrease their amounts of shrinkage and alleviate the pain due to CD [10], with the dose dependent on the sizes of these muscles which are difficult to determine.

A physician needs to know the exact positions as well as the amounts of disturbance of the neck muscles, which their shapes and sizes help to determine, in order to apply the proper medical treatment(s) and plan the correct dosage(s) of medication(s) to reduce their side-effects. Therefore, it is necessary to visualize the muscles in a 3D view using MR imaging (MRI) or computed tomography (CT) which provides volumetric images of a patient in a stack of 2D images, with inconsistencies possibly caused by the measurement method used rather than the

marker [8]. Firstly, quantifying the muscles' sizes and fat infiltrations using cross-sectional measures obtained from single 2D image slices of selected vertebral levels are not representative of 3D muscle measurements [21]. Secondly, when using 2D quantification methods, the data may suffer from errors associated with the effects of a partial volume and the alignment of an axial slice relative to the cervical spine [22]. Therefore, before the use of muscles' sizes and fat infiltrations as reliable markers for neck pain can be verified or refuted, volumetric studies of neck muscles are required.

Generally, in clinical practice, an analysis of neck muscles is conducted manually which is time-consuming and tedious, suffers from inter- and intra-operator variabilities and is unsuitable for large-scale data [23]. Therefore, an automatic analysis is essential to enhance efficiency and reduce the amount of manual labor required.

1.2 Roles of Medical Image Registration and Deep Learning-based Segmentation

Image registration [24–26] is a basic image-processing technique whereby two or more images are aligned by keeping one stationary (called a fixed image) and moving another (called a moving image) towards it [27]. It has many applications in different domains, such as remote sensing, industrial imaging and medical imaging [28]. This dissertation focuses on the medical imaging domain which has a range of applications, as discussed below.

- *Image-guided procedures:* they generally require the three phases of pre-operative planning, intra-operative execution and post-operative assessment [29].

Although pre-operative imaging is performed for surgical planning, diagnosis and intra-operative guidance, these images cannot be used due to a patient's movements during surgery whereas intra-operative imaging can help to monitor up-to-date information. However, to determine spatial relationships, image registration between pre- and intra-operative images is required.

- *Multi-modal fusion:* different modalities have different capabilities to show different information and the images obtained from them could have geometrical mismatches due to a patient's motions, devices' artefacts or calibration errors. To fuse two images to better understand the physiology of a disease, they are registered for alignment.
- *Motion determination:* the physiological motions of organs are determined using image registration for therapy monitoring and diagnosis [30].
- *Change detection:* morphological changes over time are detected through registration and can be used for diagnosis and therapy, for example, hormone therapy, rheumatoid arthritis, multiple sclerosis, Alzheimer's disease and morphological changes due to surgical interventions.
- *Distortion correction:* geometrical distortions in an imaging system can be corrected.
- *Atlas construction:* the population average of an anatomical organ can be created from a group of individuals using image registration.
- *Atlas alignment:* the population average can also be registered using the standard space for analysis, for example, the Talairach space.
- *Segmentation:* the labels from one image can be transferred to another subject image by transformation which provides segmentation, with multi-atlas segmentation currently popular.

A deep learning-based segmentation approach has huge potential in computer-aided diagnosis (CAD) as it can provide better accuracy than other segmentation methods, such as template matching, deformable model fitting, edge detection, clustering-based algorithms and threshold-based techniques. It is considered the primary option for medical image segmentation due to its promising capabilities [31]. However, segmentation is considered as a pre-requisite step for quantifying any anatomical organ and pathology. A quantitative measurement has clinical parameters related to shapes and volumes that are used to determine the planning of treatments and proper dosages.

1.3 Challenges of and Motivation for Current Research

Medical image analysis is generally a challenging task due to the anatomical variability, low resolution and contrast, noise, inhomogeneity and organ diffusion in an image. The registration and segmentation of individual cervical muscles obtained from medical images is more difficult because of the muscles' complex anatomies and changing spatial relationships as they pass through the spine [32, 33]. There are large and complex small mismatches in neck data as, in the small, narrow region of the neck, the 27 muscles that control the movements of the cervical spine are compactly arranged (see Figure 1.1) and their intricate relationships make them difficult to identify accurately. Also, similar muscles have background clutter with similar intensities while the variable quantities of fat present in and between them make their boundaries challenging to demarcate, as shown in Figure 1.1. This may be further complicated by a patient's unconscious movements

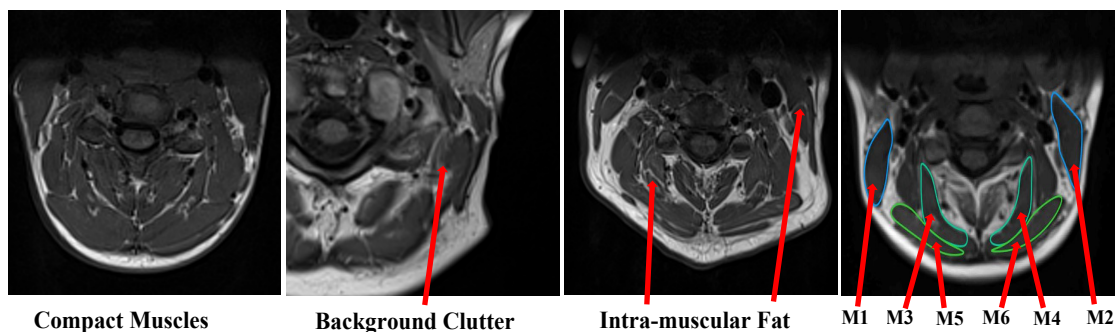


Figure 1.1: Challenges involved with neck muscles in MRI: compact and similar objects (left); background clutter (2nd from left); intra-muscular fat (3rd from left); and six neck muscles considered in this dissertation denoted as M1, M2, M3, M4, M5 and M6, respectively (right).

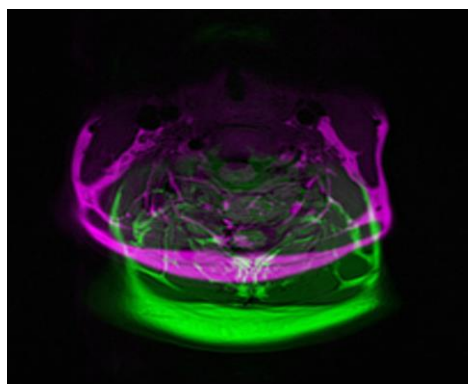


Figure 1.2: Neck MRI (composite form - anatomical variabilities between two individuals overlaid in green and magenta bands; gray regions - both images have same intensity; and green and magenta regions - images have different intensities).

(e.g., swallowing and breathing) during an image-scanning process which can deteriorate the images' quality. Moreover, the anatomical variabilities of individuals' neck muscles (Figure 1.2) are significantly greater than those of other parts of their bodies.

A good automatic medical image registration algorithm should have certain properties: firstly, high accuracy for measuring the amount of alignment between two images; secondly, robustness for handling image noises and large degrees of anatomical variability; thirdly, good computational speed for executing the relevant algorithm. Although the computational speed does not matter for some applications, others require a fast algorithm, for example, an intra-operative image

guidance system; and, finally, topology preservation for continuity of the transformation function with one-to-one mapping and inverse consistency.

In the past, many registration methods that considered the above properties for tackling the challenges in many medical applications were proposed, with the most widely used multi-resolution free-form deformation (FFD) methods [34–37]. However, they cannot completely avoid local minima due to the low degrees of freedom of geometrical transformations available and the monotonous types of transformations in multiple stages. Information theoretic-based methods are popular in mono- and multi-modal applications [38–40]. However, global information-based approaches are less sensitive to local deformations and often encounter the mis-correspondence problem due to their lack of distinctiveness which is more acute in neck data because of the compactness and similar appearance of neck muscles. To avoid this, researchers have attempted to calculate a local approach and use it with weights to compute the final objective function by summation [39–42]. However, local methods are computationally expensive, face statistical instability problems and ignore anatomical information which is crucial for guiding correspondence detection. While a feature-based registration method establishes correspondences through high-order anatomical information [27, 43, 44], often, some features are partially invariant which causes mis-registration. Although some of its descriptor-type features show distinctive characteristics, its performance faces challenges for scenarios involving large anatomical variations as it embeds undesired information and cannot fully suppress the influence of contrast enhancement. Also, it is incapable of providing distinctiveness in the case of similar patches located near each other. In fact, some features are more suitable for natural than medical image analysis.

Although many techniques for handling large mismatches have been proposed

[45–47], it has been shown that conventional continuous and discrete optimization-based registrations are not successful [48]. A heuristic-based method can provide a wide range of solutions but requires possible intuitive approaches [49–52]. Also, despite having high accuracy, as it is necessary to pre-segment both the fixed and moving images, this compromises automatism which is not feasible in all applications. In addition, it is computationally expensive because it calculates an attribute vector for each driving voxel in both images. In extensions of heuristic-based methods[53–58], local spatial intensity histograms have been used to make their algorithms suitable for general applications. However, these histograms face the problem of statistical power instability.

As, to date, there is no registration algorithm that can tackle the challenges of neck muscles and satisfy the properties of a good registration method, there is room for performance enhancement. This dissertation attempts to overcome these issues.

Like registration algorithms, many deep learning-based segmentation methods for the medical domain have been proposed. In particular, the U-Net [6] architecture is popular in biomedical applications and has attained remarkable success due to its flexibility [59] and, over the last few years, rapid advances have been based on it [60]. However, this approach has some limitations. Firstly, it reduces the resolutions of feature maps due to its consecutive operations of pooling and striding convolution while its detailed spatial information is beneficial for dense predictions. Secondly, although a deeper network can learn more complex features and yield a better performance, the depth of the U-Net cannot be increased as this causes the gradient to vanish increasingly during training and results in training difficulties. Recently, attention-based models have been studied extensively and there are many cascaded ones for different applications, such as abdominal

CT [61], cardiac CT [62], cardiac MRI [63], colorectal tumor [64] and left ventricle [65] segmentation, and lung nodule detection [66]. However, these models use excessive computational resources and repetitively extract similar low-level features. Recently, some automatic attention models [7, 67–71] were proposed. Sometimes, the performance of unsupervised automatic attention is worse than that of a mechanism without attention while an attention map may be affected by noise and cause incorrect attention. However, supervised automatic attention cannot focus on the region of interest (ROI) in an application with compact and similar small multiple classes with large variabilities and low levels of contrast due to region-level coarse supervision. As, in such a complex case, more fine guided attention is required to handle the challenges involved, there are some options for performance improvement using a U-Net approach and attention-based models which are attempted in this study.

1.4 Contributions of This Research

The main objective of this study is to develop an automatic robust 3D-3D image registration algorithm, with high accuracy, for analyzing neck muscles. Another purpose is to design an automatic deep learning-based semantic segmentation method for a neck dataset. The key contributions of this dissertation are summarized below.

- In the first part of this research, a novel object-constrained hierarchical registration framework for aligning inter-subject neck muscles is presented. Firstly, to handle large-scale local minima, it uses a coarse registration technique which optimizes a new edge position difference (EPD) similarity measure to align large mismatches. Also, a new transformation based on the

discrete periodic spline wavelet (DPSW), affine and FFD are exploited. Secondly, to avoid the monotony of using transformations in multiple stages, a fine registration technique for aligning small mismatches, which uses a double-pushing system by changing edges in the EPD and switching transformation resolutions, is designed. The EPD helps both the coarse and fine techniques to implement object-constrained registration via controlling the edges which is not possible using traditional similarity measures. Experiments are performed on clinical 3D MRI scans of the neck, with the results showing that the EPD is more effective than the mutual information (MI) and sum of squared difference (SSD) measures in terms of the volumetric dice similarity coefficient (DSC). Also, the proposed method is compared with the diffeomorphic Demons [2] and SyN [3] state-of-the-art approaches with ablation studies of inter-subject deformable registration. It achieves better accuracy, robustness and consistency than its competitors, with an average volumetric DSC of 0.7029 compared with those of 0.6654 and 0.6606 for the Demons and SyN algorithms, respectively.

- In the second part of this study, two novel contributions for the inter-subject deformable registration problem in a neck MRI application are presented. Firstly, a two-stage heuristic search optimization technique for handling large mismatches using a minimal user hypothesis, which is computationally fast, is proposed. This optimization brings a moving image hierarchically closer to a fixed one using MI and EPD similarity measures in the coarse and fine stages, respectively. Of particular note is that the images do not require pre-segmentation. Secondly, a ROI EPD-based registration framework for handling small mismatches using salient anatomical information, in which a convex objective function is formed through a unique shape created from the desired objects in the ROI, is proposed. This method is compared with

two state-of-the-art algorithms on the MRI neck dataset, with the results showing that it is superior in terms of accuracy.

- In the last part of this thesis, an evaluation study of recent U-Net-based convolutional neural networks (CNNs) is performed on the neck dataset. It comprises 10 recent models, including the U-Net [6], U-Net with a conditional random field (CRF-Unet) [72], attention U-Net (A-Unet) [7], nested U-Net or U-Net++ [73], multi-feature pyramid (MFP)-Unet [74] and recurrent residual U-Net (R2Unet) [75]. Networks with more comprehensive modifications are also evaluated, including the multi-scale U-Net (MS-Unet), parallel multi-scale U-Net (PMS-Unet), recurrent residual attention U-Net (R2A-Unet) and R2A-Unet++, for neck muscles segmentation. Analyses of the numerical results indicate that the R2Unet [75] architecture achieves the best accuracy. Also, two deep learning-based semantic segmentation approaches are proposed. In the first, a new two-stage U-Net++ (TS-Unet++) uses two different types of deep CNNs (DCNNs) rather than one similar to the traditional multi-stage one, that is, the U-Net++ in the first stage and U-Net in the second. Convolutional blocks are added after the input and before the output layers of the multi-stage one to better extract the low- and high-level features. A new concatenation-based fusion structure is incorporated in this architecture to enable deep supervision. Convolutional layers are added after each concatenation of this structure to extract more representative features. In the second approach, an explicit attention method in which the attention is performed through a ROI evolved from the ground truth via dilation is proposed. It does not require any additional CNN to localize the ROI as does a cascaded approach because attention in a CNN is sensitive to the area of a ROI. The dilated ROI is capable of capturing more relevant regions and suppressing irrelevant ones than a bounding box and

region-level coarse annotation. It is used during the training of any CNN and for coarse annotation, which does not require any detailed pixel-wise delineation that can be performed by a novice person, during testing.

1.5 Organization of Thesis

This dissertation is organized as follows.

Chapter 2 begins with a discussion of different CAD techniques used in medical applications. Then, the basics of image registration with a classification of its process obtained from the literature and a review of each class together with their pros and cons discussed. Then, the background to the classification of deep learning-based semantic segmentation is presented. Also, a related literature review of each class of deep learning-based segmentation with its advantages and disadvantages is provided.

In **Chapter 3**, a 3D-3D deformable registration framework using a novel DPSW-based transformation and EPD-based similarity measure for a dataset with neck MRIs is proposed. This method is compared with state-of-the-art ones with ablation studies of inter-subject deformable registrations.

In **Chapter 4**, another 3D-3D deformable registration framework using a two-stage heuristic search and ROI-based EPD is presented. It is fast and handles anatomical variabilities more efficiently than other methods using novel techniques.

In **Chapter 5**, three different types of deep learning-based semantic segmentation analyses are described: the first evaluates existing methods for the application of neck muscles segmentation; the second proposes a new TS-UNet++ semantic

segmentation method; and the third develops a new explicit dilated ROI-based attention approach for a deep learning-based semantic segmentation method.

In **Chapter 6**, the conclusion of this thesis and directions for further research are presented.

1.6 List of Publications and Award

Some findings obtained from this research study have been published, and others are in the pipeline for publication in the international conference papers and journal articles listed below.

Conference Papers

- **Abdulla Al Suman**, Md. Asikuzzaman, Alexandra Louise Webb, Diana M. Perriman and Mark Richard Pickering, “Inter-Subject Image Registration of Clinical Neck MRI Volumes using Discrete Periodic Spline Wavelet and Free form Deformation,” *Digital Image Computing: Techniques and Applications (DICTA)*, 10–13 December. 2018, Canberra, Australia. (Based on Chapter 3).
- Md. Asikuzzaman, **Abdulla Al Suman**, and Mark Richard Pickering, “EPD Similarity Measure and Demons Algorithm for Object-Based Motion Estimation,” *Digital Image Computing: Techniques and Applications (DICTA)*, 10–13 December. 2018, Canberra, Australia.
- **Abdulla Al Suman**, Yash Khemchandani, Md. Asikuzzaman, Alexandra

Louise Webb, Diana M. Perriman, Murat Tahtali and Mark Richard Pickering, “Evaluation of U-Net CNN Approaches for Human Neck MRI Segmentation,” *Digital Image Computing: Techniques and Applications (DICTA)*, 29 November–2 December. 2020, Melbourne, Australia.(Based on Chapter 5 of sections: 5.1.2, 5.2.3, 5.3.1, 5.4.1).

Journal Articles

- **Abdulla Al Suman**, Md. Asikuzzaman, Alexandra Louise Webb, Diana M. Perriman, Murat Tahtali and Mark Richard Pickering, “A Deformable 3D-3D Registration Framework using Discrete Periodic Spline Wavelet and Edge Position Difference,” *IEEE Access.*, vol. 8, pp. 146116-146133, 2020, DOI 10.1109/ACCESS.2020.3015504. (Based on Chapter 3).

Chapter 2

Background and Related Work

This dissertation focuses on image registration and deep learning-based semantic segmentation techniques. This chapter begins with a discussion of computer-aided diagnosis (CAD) techniques used in medical applications. Section 2.2 describes the basics of an image registration algorithm and also presents a classification of existing image registration algorithms with their corresponding merits, demerits and backgrounds as well as a relevant literature review. Section 2.3 provides a review of the literature on different deep learning-based semantic segmentation techniques and, finally, Section 2.4 concludes this chapter.

2.1 Computer Aided Diagnosis (CAD) Techniques

CAD plays an important supporting role for doctors, radiologists and other medical professionals in clinical applications, with many techniques available, such as lesion detection, segmentation, disease classification, object recognition and atlas-based segmentation. It uses the objects' shapes, sizes, textures and inter-object organization in medical images and has potential future applications in pathology using machine learning algorithms.

Medical image registration is considered a vital technique for medical image

analysis [76]. It has many applications in CAD, such as surgical planning, establishing the spatial relationships between a patient's pre- and intra-operative information, analyzing temporal changes, conducting image-guided therapy, assessing brain degeneration in Alzheimer patients, detecting changes in tumors following radiation therapy and studying lung cancer, cross-modality image fusion, atlas construction, distortion correction, population and many other neurosciences. Also, anatomical labels can be transferred automatically from one subject to another using image registration.

Many medical image segmentation methods have been presented in the literature, such as template matching, deformable model fitting, edge detection and learning-based approaches. However, deep learning-based semantic segmentation techniques have recently revolutionized medical image analysis due to their huge success [59]. Segmentation is a first step in CAD [77] and has various applications, such as analyzing pathology, planning treatment and monitoring the progression of disease [78]. It also helps the understanding of the physiologies and characteristics of diseases [78].

2.2 Image Registration

Image registration is a process that establishes the spatial correspondences between two images [48] which is normally considered an ill-posed problem. In it, two or more images are aligned by keeping one stationary (called a fixed image) and moving the other (called a moving image) towards it, as shown in Figure 2.1. A geometric transformation between these images is estimated by optimizing a similarity measure. Therefore, a registration algorithm has three main components:

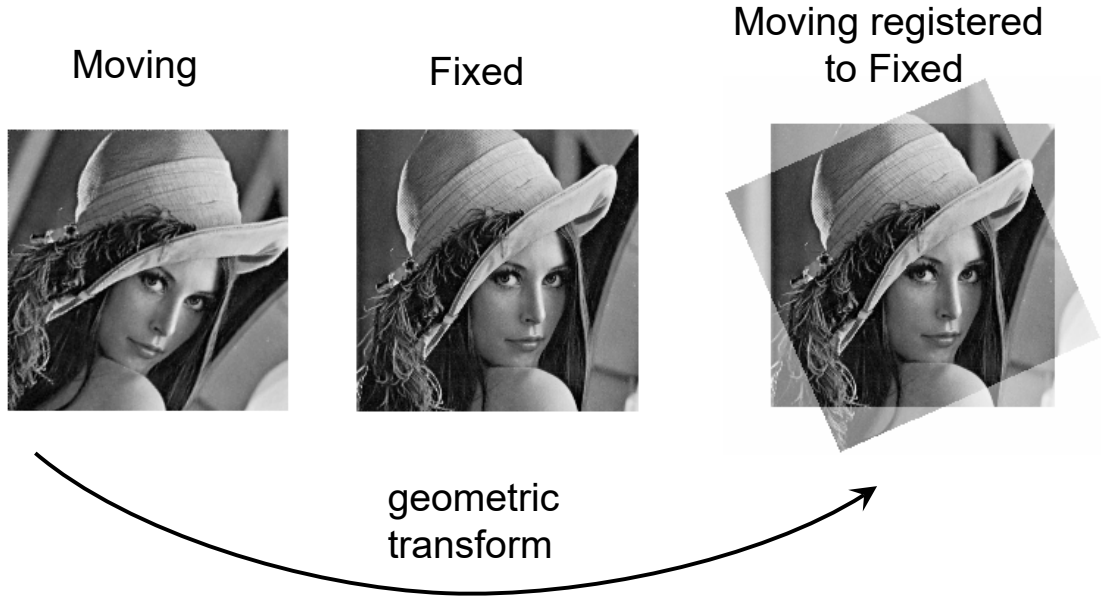


Figure 2.1: Schematic of image registration process

a geometrical transformation; similarity measure; and optimization. Figure 2.2 illustrates the operational flow of a registration algorithm.

Images can be taken from the same or different sensors, with these registrations called mono-modal and multi-modal, respectively. The images can be 2D or 3D and called 2D-2D, 3D-3D and 2D-3D (if one is 2D and another 3D) registrations.

2.2.1 Geometrical Transformations

The choice of transformation has a large effect on the registration process [48], with the most appropriate one not known as a prior [79], and also influences the alignment [48]. A transformation model is selected so that it is close to the underlying deformation required [29], with one that has both translation and rotation, and preserves the distances among points called a rigid body [30]. A transformation which considers the rigid body as sub-set and has shear and scale changes is called affine and matches parallel lines to parallel lines [30].

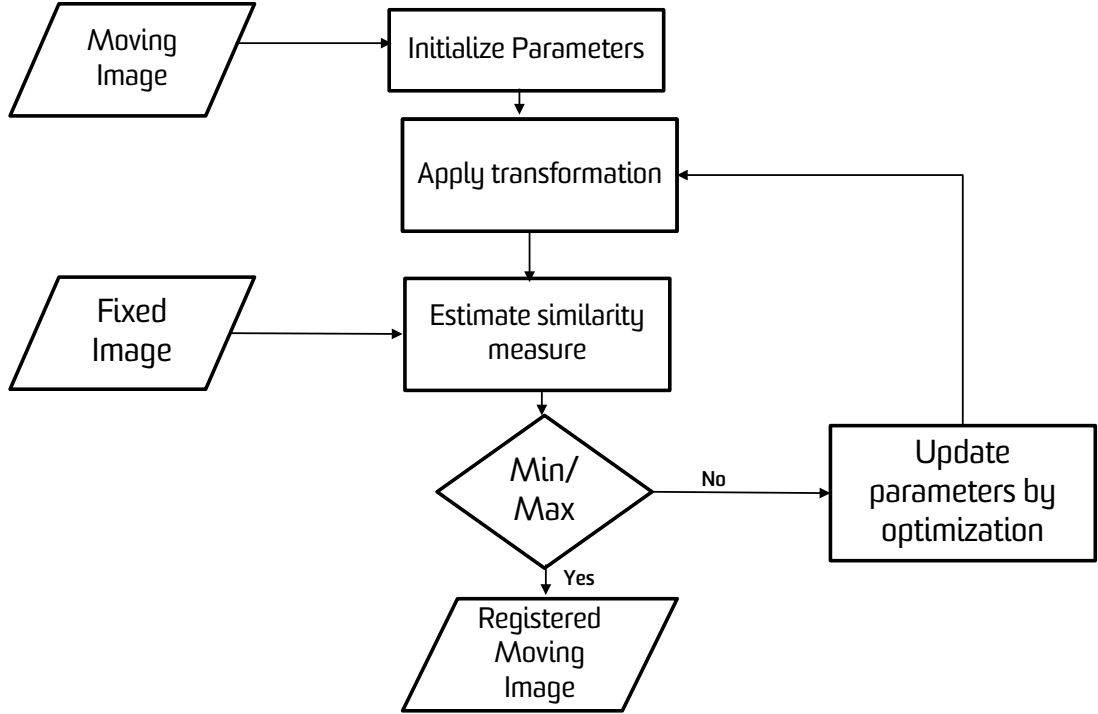


Figure 2.2: Diagram of operational flow of image registration algorithm

On the contrary, models which map straight lines to curves are called non-rigid transformations and can be classified as parametric or function representations and non-parametric or physically-based models. The former have lower degrees of freedom than the latter and their solution spaces can be modeled by parameters [29].

2.2.1.1 Parametric Models

Expanded basis functions are used in this type of transformation to model deformation, with the displacement field a combination of these functions and their corresponding coefficients, which are called the transformation parameters, in the image domain [30]. The number of parameters represents the number of degrees of freedom of the transformation model on which the computational complexity and capability to model spatially varying local deformations depend; in particular, more parameters result in greater computational complexity and more local

spatially varying deformations. The basis functions are derived from either approximation theory or interpolation, with those from the former smoother than those from the latter. There are many types of basis functions in the literature, such as radial basis functions, elastic body splines, free-form deformations, signal representations and locally affine models, with a radial basis function and B-spline more stable than the others [30].

1) *Radial basis functions (RBFs)* are a significant family of interpolation strategies [48] in which the distance from a known sample is used to estimate the value at an interpolation point. An evaluation of RBFs as a transformation in non-rigid image registration can be found in [80] and landmark-based medical image registration used RBFs in [81–84]. There are several families of RBFs, such as thin-plate splines (TPS) [85–90], multi-quadratics [91, 92], Wendland functions [84, 93, 94], Wu [95] and Buhmann [96]. TPSs and multi-quadratics support global deformations [87, 89] whereas the Wendland, Wu [95] and Buhmann [96] functions support local ones.

2) *Elastic body splines (EBSs)*, which use interpolation and approximation theory and may also exploit physical models, were proposed by Davis *et al.* [97]. The splines are obtained from the solution of the Navier-Cauchy equilibrium equation when forces are applied to an isotropic homogeneous elastic body and their support is global. An extension of the EBS using a Gaussian function (Gaussian EBS) of the distance from a landmark in order to impose local deformations was proposed in [98]. Another extension through an approximation strategy incorporates errors in landmark displacements [99]. Also, EBSs were used to handle anisotropic landmark localization errors by exploiting the Tikhonov regularization scheme in [100].

3) *Free-form deformations (FFDs)* were first introduced for computer graphics [101, 102] and later popularized in the medical image analysis community when combined with a cubic-B spline [103–106]. The deformations are designed using a tensor product of a 1D cubic-B spline which has compact support for the capability to design fine and smooth deformations. It supports smooth deformations, simple models, and topology preservation with regularized FFD.

Significant developments of FFDs have been introduced, such as multi-level B-splines in [107] and non-uniform rational B-splines (NURBS) in [108] in an adaptive fashion. Multi-level B-splines were also used with an activating sparse subset of control points in [109]. A symmetric version of FFDs was presented in [110] and an inverse consistent extension of them in [111]. Sotiras and Paragios [112] proposed another symmetric FFD technique by incorporating an invertibility of mappings through discrete labeling. FFDs were applied in group-wise registration through hard constraints in [113–116] and by exploiting splines in the temporal axis in a spatial-temporal domain in [117–119]. An investigation into using lower-order B-spline functions to preserve the smoothness of deformations through a random perturbation technique was proposed in [120]. Different resolution levels of FFDs were used simultaneously for transformation and data in [121] and a FFD was extended by adding extra degrees of freedom near discontinuities to handle sliding motions in [34]

4) *Signal representation* formulates the basis functions as Fourier and wavelet ones which are commonly used in signal processing and a multi-resolution scheme of deformation fields can be obtained from them which is a desirable property for medical image registration.

Fourier-based basis functions were used in non-linear registration in [122, 123] although they are not efficient for image registration as they are localized in only

the frequency domain [124]. On the contrary, as wavelet-based basis functions are localized in both the frequency and spatial domains, they are more suitable for designing local deformations.

A wavelet-based deformation model was used in a multi-resolution manner to deal with coarse-to-fine mismatches in [79, 125]. Topology-preserving registrations using a non-orthogonal Riesz basis of polynomial splines were proposed in [126–128] using a multi-resolution approach. In them, topology-preservation is performed using hard linear constraints in 2D and constrained optimization with user-specified bounds for Jacobian in 3D.

5) *Locally affine models* enable locally linear deformations which are classified as piece-wise affine and poly-affine models. The former divide an image into shaped regions and apply an affine transformation on each. They have good invertibility inside regions but discontinuous deformations on the boundaries. To handle the discontinuous problem, fuzzy regions are used in poly-affine-based models to yield smooth deformations at the interfaces. Piece-wise affine models were used in [129–134] and poly-affine ones in [135, 136].

2.2.1.2 Non-parametric Models

Non-parametric transformation models provide dense deformations for every pixel which are normally derived from physical models deduced from continuum mechanics theory and are controlled by regularizations in the manner like the natural characteristics of permissible deformations. These models are considered to have sub-categories of elasticity, fluid flow, diffusion and diffeomorphisms.

1) *Elasticity model*: in this model, an elastic body is used to design the deformation model, with the deformations described by the Navier-Cauchy Partial

Differential Equation (PDE) in linear models and an image-matching criterion acting as an external deformation force. The displacements were inferred using the finite element method (FEM) in a variational setting in [137]. Symmetric registration using linear elasticity as a regularization constraint was proposed in [122]. Geometric characteristics were incorporated in a linear elastic model with spatially varying elasticity parameters in [138]. Large deformations were estimated in an inverse consistent framework using a linear elastic model in [139]. However, as linear elastic models are not successful for handling large deformations, many non-linear models have been proposed [140–142], with topology preservation guaranteed in them. Non-linearity was incorporated through the St Venant-Kirchoff elasticity energy as well as log-Euclidean metrics in [140] and log-transformation and log-normal distributions in [141].

2) *Fluid flow model*: in this model, a deformation is considered a viscous fluid propelled by the Navier-Stokes equation, with these models capable of recovering large deformations intrinsically due to the deformation rate-dependent strain. Christensen *et al.* [143, 144] applied a viscous fluid flow model after a linear elastic one to align large deformations. It was also applied in an atlas-enhanced registration setting in [145] and in a multi-modal registration with mutual information (MI) in [146]. Chiang *et al.* [147] applied this model for the registration of diffusion tensor images with an inverse-consistent and symmetric Kullback-Leibler divergence (KLD) as a matching criterion.

3) *Diffusion model*: in this model, a diffusion equation is used to model deformations with regularization implemented through the convolution of a Gaussian kernel, which is the Green’s function of the diffusion equation, that smooths the deformation field to preserve the topology and cancel the noise effect.

Thirion [148] proposed a Demons method, inspired by Maxwell's Demons approach, to exploit diffusion regularization, with the smoothness of the displacement field controlled by the width of a Gaussian kernel. In this method, the demon forces are calculated for every demon and used to update the transformation. They can be calculated in different ways, such as considering all of an image's elements as demons using an optical flow constraint, tri-linear interpolation or regularization through Gaussian filtering. Fischer and Modersitzki [149] proposed an extension of Thirion's Demons algorithm to a fast registration one by linearizing a diffusion PDE. Pennec *et al.* [150] employed a Demons algorithm with energy minimization and replaced Thirion's force by a second-order gradient descent (GD) on the sum of squared differences (SSD). Symmetric Demon forces-based image registration with an efficient second-order minimization (ESM) was used in [151, 152] with Gaussian smoothing regularization separated from the matching term.

Vercauteren *et al.* [153] incorporated a diffeomorphic property in the Demons algorithm through a compositional update rule and exponential map of the updated field which ensured diffeomorphism. Later, a symmetric property was incorporated into the algorithm through averaging the forward and backward forces [154]. Stefanescu *et al.* [155] proposed an adaptive smoothing algorithm using the knowledge of tissues and Peyrat *et al.* [156] registered multi-channel time-series images.

4) *Diffeomorphism model*: in this model, non-parametric transformations generally do not guarantee topology preservation as the deformations of physical models can be restrained in the diffeomorphic space. Diffeomorphic deformations can be achieved by modeling velocity over time with respect to the Lagrange transport equation, with the velocity field restrained by a regularization term. Beg *et al.* [157] proposed the large deformation diffeomorphic metric mapping

(LDDMM) method that uses the physics of mechanical motion which is a popular diffeomorphic registration approach and has attracted the attention of many registration researchers for various applications, for example, brain tissue registration and segmentation [3, 158, 159], diffusion tensor imaging [160] and registration of histological and MR volumes [161]. However, as it is not symmetric, several methods for incorporating a symmetric property have been proposed, as in [3, 162]. The LDDMM has high memory and computational requirements due to its calculation of the velocity field over time and slow convergence of the GD optimization.

Although efficient optimization is an option for reducing computational complexity, researchers have tried to simplify the calculation of diffeomorphisms by decreasing the numbers of degrees of freedom using stationary velocity fields [2, 163] which do not require the numerical integration of time-varying fields.

2.2.2 Similarity Measures

Similarity measures quantify the alignments between fixed and moving images, are also known as dissimilarity criteria, matching criteria and distance measures, and are a significant part of any registration approach. They are classified into the two main categories of feature- and intensity-based methods [164]. However, there are also some hybrid methods which are rarely used in medical image registration applications [165]. The following sub-sections describe the three main similarity measure methods.

2.2.2.1 Feature-based Methods

These methods quantify the alignment of corresponding features in fixed and moving images considering the landmark information. These features may include

points, corners, edges, blob-like regions, lines, curvatures, contours and many more, and are also known as points of interest, landmarks and key-points. These methods are computationally efficient because the features can be calculated before registration and this bypasses their re-calculation in every optimization iteration. Furthermore, a calculation may be performed on only the locations of an image's features [166] which do not change due to deformations. These methods have four steps: extracting features; generating feature descriptors; matching the features in two images; and estimating the transformation parameters [167].

Feature detection and extraction are the important first steps in feature-based methods. The quality of the matching of detected features depends on their descriptors which should be capable of distinguishing between close candidate features [48]. While moving images undergo deformations in every optimization iteration, the descriptors should be invariant to these deformations.

There are many techniques in the literature for identifying features. A structure tensor's information was used to identify corners and edges in [168] and points of interest in [169], with many extensions aimed at encoding invariance. An evaluation of corner and point detectors was performed in [170]. An alternative approach using the Laplacian of Gaussian with different sizes of Gaussian kernels for detecting blob-like regions was proposed in [171, 172]. Lowe [173] developed a scale-invariant feature transform (SIFT) algorithm for detecting feature points using a scale-space representation which rejects spurious points using local Hessian information. Many extensions of this method have evolved. The gradient location and orientation histogram (GLOH) was proposed in [174] for reducing the dimensionality of the SIFT using principal component analysis (PCA), which employs a log-polar pattern for spatial sampling. Bay *et al.* [175] proposed the speeded-up robust features (SURF) using a Haar wavelet in the ROI. However, there are very

few applications of feature extraction using the SIFT in medical image analysis.

Cheung and Hamarneh [176] applied their extension of the SIFT on CT and MR images. Han [177] implemented an extension of the SURF in a 3D hybrid registration framework, with a 3D SIFT used in an ultrasound volume [178], CT [179] and optical coherence tomography (OCT) images [180].

The correspondence between fixed and moving images for features used to register the images can be established in two ways. Firstly, the closeness or distance of the descriptor is used to match images and the Euclidean distance to rank potential matches. Different strategies are applied to match images based on their ranks, for example, evaluating the ratio between the distances of the nearest and second-nearest neighbors in the feature space [176, 177, 180]. Secondly, structural or geometric constraints are used as in graph matching. Leordeanu and Hebert [181] proposed a spectral technique that exploits pair-wise constraints to preserve pair-wise geometry. Duchenne *et al.* [182] applied spectral matching to higher-order constraints.

Also, the spatial transformation between estimating two point sets can be used to match images. In this case, the sets are considered rather than every point-to-point explicit assignment which is similar to correspondence matching. A spatial transformation can be estimated in two ways. Firstly, using the known correspondences which are mainly global linear ones and either exact or inexact. In an exact case, a smooth transformation is searched to find the exact correspondences while, alternatively, a compromise between smoothness and matching is desired. Also, non-rigid transformations can be estimated in the case of known correspondences. The Procrustes analysis is a well-known method for matching point sets [183, 184] and exact and inexact matching methods are provided in [185]. Secondly, unknown correspondences are used to estimate spatial transformations. However, as

these methods are more challenging and robust to outliers and missing correspondences, they estimate mainly non-rigid transformations. Point sets are considered as probability distributions for minimizing the distance measured between them. Gaussian mixture models (GMMs) were used in different ways in [186–189]. Using signed distance functions is another way of representing geometric information to accomplish non-rigid registration. This approach was used with the SSD in [190] and MI in [191].

Some methods consider both the correspondences and transformations in an iterative manner. The iterative closest point (ICP) is a popular approach based on the closest distance correspondences, with many variants proposed to improve it [192]. It was used in FFD estimations in [193] and low-order transformations in [194].

2.2.2.2 Intensity-based Methods

These methods are based on the intensities of the images and use the whole image domain. Selecting the appropriate similarity measure is a difficult task which depends on the application and it is desirable for it to have a convexity property that supports the optimization process. However, convexifying a similarity measure may lead to a less realistic solution as the application may be naturally non-convex [48]. A similarity measure should be high for the same tissue classes and low for dissimilar ones. Generally, these methods are classified into two categories: 1) mono-modal; and 2) multi-modal.

1) *Mono-modal methods* can be classified as intensity- and attribute-based, with the images obtained from the same type of imaging device.

Intensity-based methods depend on the intensity relationship between two images for which it is assumed that the same anatomical objects have similar intensities. The SSD, sum of the absolute differences (SAD), correlation coefficient (CCoef), correlation ratio (CR) and cross-correlation (CCor) are the most commonly used mono-modal similarity measures. The choice of intensity difference-based method (SSD or SAD) depends on the noise in the two images, with lower values better for good registration. The CCoef [195], CR [196] and CCor [3] are used for linear relationships between fixed and moving images [164], with higher values preferable for good matching.

Attribute-based methods use standard similarity measures with the geometric structures of objects. They overcome the ambiguous matching problem of intensity-based approaches and reduce the local minima of optimization. Shen and Davatzikos [197] proposed attribute vectors for voxels using a geometric moment to impose a more distinctive property of the voxels so that the number of local minima is reduced and better accuracy achieved, with the number of voxels with attribute vectors increased during the algorithm's progression. However, this method requires pre-segmentation to incorporate spatial local information. Xue *et al.* [198] removed the pre-segmentation requirement by using Daubechies wavelets to populate the attribute vectors. Shen used boundary information and local histograms to make an attribute's vector rotation- and translation-invariant [199]. Also, local frequency representations obtained from Gabor filters [200] and symmetric alpha stable filters [201] can be used to encode local information. The frequency representations can be obtained from different orientations and scales. Liao and Chung [201] showed that symmetric alpha stable filters outperform Gabor filters. Myronenko and Song [202] exploited the residual complexity (RC) to encode complex spatially varying intensity distortions in residual images.

2) *Multi-modal methods*: obtain images from multiple imaging devices and are more challenging than mono-modal methods due to the devices' different characteristics. Applying information theoretic-based methods and converting multi-modal problems to mono-modal ones are the two main approaches for handling this challenge.

MI is a popular information theoretic-based method. In terms of two images, it is defined as the amount of information they contain about each other [203] which increases with registration alignment and decreases with mis-registration. There are two main categories of MI-based registration methods: non-parametric entropy estimation-based MI calculations which are used for multi-modal registration [204, 205]; and histogram entropy estimation-based ones [206]. However, as MI is not overlap-invariant, it may lead to mis-registration. To tackle this problem, Studholme *et al.* [207] proposed normalized MI (NMI) and Cahill *et al.* [208] later developed an appropriate invariant of NMI as NMI was not completely overlap-invariant.

Researchers have tried to use statistical criteria for image registration. Roche *et al.* [209] applied the CR as a similarity measure of the valid functional dependence of the intensities of two images. Different divergence measures have been used in multi-modal registration, such as the KLD [210] and Jensen-Shannon divergence (JSD) [211]. He *et al.* [212] exploited the Jensen-Renyi divergence (JRD) for image registration. The KLD was generalized in [213] based on reformed Bessel functions and performed better than a standard divergence.

As all the above methods use a single-pixel joint probability model, the same criterion may be implemented for the process of image warping in registration [214], for example, shaded artefacts may cause a registration to fail. Also, as global MI is less sensitive to local deformations and often provides mis-correspondences, local

context was introduced to tackle this problem. Hermosillo *et al.* [196] used local probability distribution functions for MI, CCoef and CR. Karaçali [215] applied a deterministic approach for the MI, joint entropy and marginal entropy over local spherical regions. Some other local versions of MI are regional MI (RMI) [216], conditional MI (CMI) [217] and spatially encoded MI (SEMI) [218].

Encoding spatial information is another way of introducing local context and results in a higher-order entropy. Rueckert *et al.* [214] applied second-order MI using a 4D-histogram with local information incorporated through feature extraction. Holden *et al.* [219] incorporated Gaussian-scale space derivatives as an additional information channel for high-dimensional MI. However, a higher-dimensional criterion faces the problem of insufficient samples. Also, prior segmentation is used to perform tissue classification and assists the design of robust registration. Studholme *et al.* [220] exploited threshold-based region segmentation to calculate separate entropies for each region. However, local MI faces the problem of the instability of statistical power and is computationally expensive.

Converting a multi-modal registration to mono-modal can alleviate the problems with the former and can be performed by either simulating one modality based on another or mapping both modalities to a third common domain.

The first process can be achieved by modeling the imaging procedure using machine learning techniques to estimate the intensity relationship. Cao *et al.* [221, 222] proposed bi-directional image synthesis methods using machine learning to simulate the differences between MRI and CT pelvic images in a prostate cancer radiation therapy application. Roche *et al.* [223] predicted ultrasound (US) images from MR using the MR gradient and intensity information. Wein *et al.* [224] simulated US images obtained from CT using the physical theory of US.

The second process extracts the local geometrical information from both images to create new images for which Maintz *et al.* [225] used morphological tools. Haber and Modersitzki [226] exploited the intensity gradient information to extract the borders of anatomical structures while Butz and Thiran [227] applied edge information. Other geometrical information used in multi-modal registration are Gabor filter outputs, the probability of a vessel's presence, local frequency representations, etc. Lee *et al.* [228] proposed a supervised learning-based similarity measure using a support vector machine for multi-modal image registration.

2.2.2.3 Hybrid Methods

These methods have the advantages of both feature- and intensity-based methods and avoid their limitations. Based on their inclusion of features, they are classified as: 1) initialization-; 2) constraint-; and 3) coupled-based approaches.

1) Initialization-based methods: as they use the features and intensity information in an independently sequential way, they can be considered multi-stage methods. Generally, feature-based techniques are initially applied for coarse alignment and then intensity-based ones for fine alignment. Landmark information was used for hybrid registration in [229, 230] for coarse alignment. Surface information is another way of aligning images in hybrid methods in which surface matching is performed either before or after intensity-based registration [231, 232]. The outputs from the registration of segmented structures were further refined through intensity-based registration in [233].

2) Constraint-based methods: they use one type of information independently in the first step and a constraint in the second to estimate correspondences. They

are more robust than initialization-based methods because they preserve correspondences, with the constraint either soft or hard.

In soft constraint methods, constraints are applied point-wise. Papademetris *et al.* [234] used a matching criterion which ensured point correspondences in a deformation field. Rohr *et al.* [235] exploited the local CCoef with point correspondences to register electrophoresis images.

Joshi *et al.* [236] proposed hard constraints through geometric correspondences. They initially established correspondences between cortical gray matter and gray surfaces via sulcal constraints and then used a harmonic map to propagate the correspondences into a whole cortical volume. They applied a hard constraint to ensure no deformation of the previously registered surfaces.

3) *Coupled-based methods*: in them, the feature and intensity information is considered simultaneously in a single objective function. Cachier *et al.* [237] proposed a universal energy function for brain registration in an iterative manner with the three steps of intensity, feature and both of them simultaneously. Joshi *et al.* [238] used a surface- and intensity-based similarity measure and mapped from an interior brain volume to a sphere. Sotiras *et al.* [115] proposed a simultaneous feature-intensity registration method for two images using two sets of landmarks to estimate their correspondences and a FFD to map between them. Honnorat *et al.* [239] developed a coupled similarity measure for a guide-wire tracking problem. This approach was also used with a diffeomorphic Demons algorithm in [240, 241].

2.2.3 Optimization

The optimal transformation which best aligns two images is achieved by maximizing or minimizing their similarity measure, a process called optimization. It is considered an ill-posed multi-dimensional problem which optimizes a similarity measure with respect to the transformation parameters. Generally, optimization methods for medical image registration are classified in three categories based on the nature of their variables, that is continuous, discrete and miscellaneous.

2.2.3.1 Continuous Optimization Methods

In these methods, the variables take real values and the objective functions are differentiable [48]. Some frequently used for medical image registration are the GD, conjugate gradient (CG), Powell's conjugate directions, quasi-Newton (QN), Gauss-Newton (GN), Levenberg-Marquardt (LM) and stochastic GD approaches. All are unconstrained but some can also be applied for a constrained transformation.

A GD algorithm optimizes the objective function by decreasing the energy and requires a gradient calculation in each iteration. It has been used in many medical image registration applications and has two variants [242]. The first uses a decaying step size and the second an inexact line search. The LDDMM exploited a GD method in [157, 243] and a FFD registration in [103].

CG methods have better convergence rates than GD ones. In them, the search direction is the conjugate of that of the previous iteration. Some examples of medical image registration using CG optimization are in [236, 244]. Tustison *et al.* [245]

proposed applying this optimization in FFD registration using a pre-conditioned gradient scheme.

Powell’s conjugate directions method minimizes the objective function in the conjugate direction without using gradient information, with the initial direction set along the basic vectors and each parameter of a transformation optimized along the independent axis. It has been applied for low degrees of freedom [206, 210, 246] and is gradient-free but may fail even for moderately difficult problems.

QN methods exploit the previous iterations’ information to achieve better convergence and use an inverse Hessian matrix for the search direction. The Broyden–Fletcher–Goldfarb–Shanno (BFGS) and Davidon–Fletcher–Powell (DFP) are the two main algorithms used, with the former more efficient than the latter. These approaches for image registration can be found in [119, 188, 217].

A Gauss-Newton method optimizes the objective function in the form of the SSD function values which is common in mono-modal image registration. It calculates the Hessian matrix using the Jacobian and ignores derivatives of more than the first order. Demons registration also uses this optimization method [2, 151, 153] as do some other applications for image registration [247, 248]. Zikic *et al.* [249] recently proposed a pre-conditioning scheme-based GN method for improving convergence.

The Levenberg-Marquardt algorithm is fast and its speed and stability can be controlled by varying a weighting factor. It has been applied in many well-known image registration applications [79, 105, 125]. Thevenaz and Unser [250] proposed an efficient optimization approach based on the LM algorithm for MI registration.

Stochastic GD methods are used to reduce the computational burden of the previous deterministic gradient methods using approximations of the gradient.

There are three types depending on the approximation process [242], that is, the Kiefer-Wolfowitz (KW), simultaneous perturbation (SP) and Robbins and Monro (RM) approaches, with the RM one performing best. These methods are normally applied in transformations with relatively low degrees of freedom, with some examples for image registration provided in [114, 204, 205].

2.2.3.2 Discrete Optimization Methods

These methods are constrained and their variables take discrete values. The three types used for image registration are graph-based, belief propagation (BP) and linear-programming (LP).

Graph-based methods are in accordance with the max-flow min-cut principle. Tang and Chung [251] proposed graph cuts based on non-rigid image registration and So *et al.* [252, 253] used them based on discrete optimization with MI for non-rigid brain image registration.

A BP method is based on local messages passing between nodes in the graph with backtracking which requires a large amount of storage. Yang *et al.* [254] used constant-space BP for stereo matching and Heinrich *et al.* [255] applied BP in non-rigid registration to recover respiratory motion. Shekhovtsovetal *et al.* [256] proposed a non-rigid image registration by decomposing a graph into two layers which reduces the number of operations required to update messages.

LP methods solve LP relaxations which is difficult. FastPD is an approach that exploits primal and dual LP relaxations and is used in image registration to model grid-based displacements in [257, 258]. Sequential tree-reweighted (TRW-S) message passing is another LP relaxation used for various image registration problems [259, 260].

2.2.3.3 Miscellaneous Optimization Methods

In some cases, continuous and discrete methods cannot provide solutions while heuristic and metaheuristic ones can produce large solution spaces but not guarantee optimal solutions. Greedy and evolutionary algorithms are the two main types of algorithms used in image registration. The former require plausible solutions and are intuitive and gradient-free, with some described in [50, 199]. Evolutionary algorithms, which are driven by evolution and the theory of natural selection, are used mainly in linear registration and have shown slow convergence. Practical implementations of them can be found in [261].

2.3 Deep Learning-based Segmentation

These techniques have revolutionized segmentation tasks as they have remarkable segmentation accuracy [59] and are now a well-established robust and primary approach for image segmentation [31], with their success possibly due to advances in hardware. Researchers have focused strongly on deep learning-based segmentation in the past few years, with many different architectures applied [77]. The most common are the convolutional neural network (CNN), fully convolutional network (FCN), U-Net, convolutional residual network (CRN), recurrent NNs (RNNs) and attention-based CNN.

2.3.1 Convolutional Neural Network (CNN)

With its development, a CNN has become a feasible method for segmenting medical images [262]. It consists of a stack of layers, with the first layer called the

input and the last the output. The input is connected to an input image which has neurons equal to number of pixels. Each neuron works on a particular area of the previous layer's output which is known as the receptive field. The intermediate layers between the input and output ones are convolutional and use the previous layer's output as input. Each layer has a fixed number of filters which work as feature extractors through a convolutional operation using the previous layer's output, with each convolutional layer followed by activation and batch normalization ones. The activation layer imposes non-linearity on the network to establish non-linearity between its inputs and outputs. Depending on the design, there may be pooling layers after some convolutional ones which reduce the size of the convolution's output. Finally, the fully connected layers extract high-label feature abstractions and then the weights of the filters' kernels are optimized during back-propagation in the training phase [263].

CNNs have been applied in medical image segmentation by many researchers in different ways, with image patches-based deep learning methods used in the early days [60]. Ciresan *et al.* [264] proposed a sliding window strategy and patches-based deep learning method for segmenting neuronal membranes from microscopic images. Zhang *et al.* [265] developed a CNN for 2D image segmentation using multiple modalities as input channels which are better than single-ones. Bar *et al.* [266] used a transfer learning technique which borrows low-level features from a model pre-trained on a dataset and another dataset was also used to extract high-level features fused with the low-level ones to segment objects in the latter dataset.

A 2.5D approach is another method for segmenting 3D images using 2D labeled data in three orthogonal planes [267, 268]. It has the advantages of richer spatial information and less computational time than 3D, and uses three separate CNNs for each plane. Roth *et al.* [268] considered planes as the channels of an input

image. Moeskops *et al.* [269] proposed a 2.5D approach with a single model for segmenting multiple organs (brain, breast and cardiac) with different modalities. Although it obtains better results than 2D, some researchers [78] believe that it is not the optimal solution due to the possibility of there being many different views of 3D data. Also, there is a conflict between the isotropic kernels and anisotropic behavior of a 3D image [270].

A 3D CNN is more capable of extracting better 3D features than a 2.5D approach since 3D kernels can learn more organized, oriented and precise features. It is constructed by replacing each 2D module in a 2D CNN by a corresponding 3D one. Urban *et al.* [271] proposed the first 3D CNN for brain tumor segmentation which was followed by Kamnitsas *et al.* [272] who developed a dual-pathway multi-scale 3D CNN and achieved an average DSC of 0.66. Dou *et al.* [273] proposed a fast 3D CNN using a set of 3D kernels with spatially shared weights.

2.3.2 Fully Convolutional Network (FCN)

An FCN, which was developed by Long *et al.* [274], replaces the last fully connected layer by a fully convolutional one and supports dense predictions. Its design enables pixel-wise predictions of an entire image in one forward pass and also combines high-resolution activation maps with up-sampled outputs to yield better localization performances. Nie *et al.* [275] showed that an FCN performs better than a CNN using the same dataset.

An FCN was used in a 2.5D approach for segmenting multiple organs [276, 277] and in a 3D method [278]. Roth *et al.* [279] proposed a hierarchical 3D FCN for multi-organ segmentation to improve the accuracy for small organs.

Also, an FCN was applied in a cascaded format to increase the accuracy of a coarse-to-fine approach which enables multi-scale implementation using different kernel sizes. Christ *et al.* [280] applied an cascaded 3D FCN for liver lesion segmentation with 3D conditional random fields (CRFs). This method can be applied in both series and parallel ways but the latter has more computational complexity [78, 281]. A cascaded method is also used to handle the problem of class imbalance, in which segmentation accuracies are low for small objects, by implementing a hierarchical coarse-to-fine approach using two FCNs, where the second network focuses on the boundary.

An FCN can be used in a multi-stream architecture [31, 77] to accept input images in diverse forms, such as multi-resolution and multi-modal, for the same anatomical organ. This enables the variabilities of anatomical structures to be handled through maximizing the contextual information. Zeng and Zheng [282] proposed a multi-stream technique using an 3D FCN for the multi-modal segmentation of a MR image of an isointense infant brain with multi-scale deep supervision. A multi-resolution (multi-scale) technique can handle variable sizes of objects with a fixed receptive field [78, 283] as can using a sliding window strategy over an entire image [283] since the objects in different multi-resolution images necessitate different parameters. Multi-stream methods are also used to detect multiple organs.

2.3.3 U-Net

The U-Net, which was introduced by Ronneberger *et al.* [6], is one of the most well-known architectures, based on an FCN, for medical image analysis. It has two

architectural novelties: firstly, it has equal numbers of down- and up-sampling layers; and, secondly, it has skip connections between these layers which concatenate their features and uses the full context of an image in one forward pass.

Many extensions of the U-Net for medical image segmentation have been proposed [73–75, 284, 285] and applied in many applications [4, 7, 72]. Zhou *et al.* [73] proposed the U-Net++ by decreasing the semantic gap between the feature maps of the encoder and decoder of the U-Net through dense nested skip connections. It provides better segmentation accuracy than the U-Net and wide U-Net for 3D CTs of chest nodules, nuclei microscopic images, liver CTs and polyp colonoscopy videos. Alom *et al.* [75] proposed the R2U-Net which ensures better feature representation than the U-Net with the same number of parameters and yields as good results for segmenting retinal blood vessels, skin cancers and lung lesions. Moradi *et al.* [74] developed the MFP-UNet by exploiting a feature pyramid to extract feature maps from all the blocks of an expanding path in a semantic layer for a segmentation procedure instead of the last block as does the U-Net. It performs excellently for echocardiographic segmentation, better than the U-Net, U-Net++, deeplabv3, an anatomically constrained NN (ACNN) and a stacked hourglass (SHG) network. Song *et al.* [284] proposed the U-NeXt which has attention up-sampling blocks in the expansion path and spatial pyramid pooling in the skip connections. Also, jumping connections are used to connect the convolutional layers to generate multi-scale features. Zhang *et al.* [285] developed the Z-net for prostate segmentation using a multi-level feature technique and modified the U-block by adding an additional convolutional layer. Also, the feature maps are cropped before the pooling layer and concatenated after this layer’s feature maps. Tang *et al.* [4] proposed a multi-stage U-Net for segmenting skin lesions by integrating a contextual information fusion structure (CIFS), which combines

the low-level features in a multi-scale feature space, with a weighted Jaccard distance loss function, that alleviates the gradient-vanishing problem, to improve the network's performance.

The U-Net was extended for 3D image segmentation by Çiçek *et al.* [286] to perform 3D segmentation using 2D annotation. Kleesiek *et al.* [287] used a 3D U-Net for brain extraction by stripping the skull and using a mixing layer in conjunction with two convolutional layers in the up-sampling path to overcome the localization problem. Zeng *et al.* [288] applied multi-level deep supervision with a 3D U-Net by dividing the up-sampling path into low, middle and upper levels, with the de-convolutional blocks in the low and middle levels used to generate the same resolution of the input to enhance the final result. Another popular 3D variant of the U-Net called the V-Net proposed by Milletari *et al.* [289] replaces the requirement for pooling layers in the down-sampling path by re-designing the convolutional layers through appropriate selections of the stride and kernel size. Gibson *et al.* [290] developed an extension of the V-Net that integrates multi-scale and larger receptive fields for multi-organ segmentations of abdominal CTs.

2.3.4 Convolutional Residual Networks (CRNs)

Although deeper networks are good for improving learning capacity, they face problems of the gradient vanishing and accuracy being degraded. He *et al.* [291] proposed a deeper network by introducing a residual one to tackle these problems using skip connections to feed every few layers rather than feed the stacked layers consecutively. Yu *et al.* [283] extended the basic CRN to a fully convolutional residual network (FCRN) to predict pixel-wise for melanoma segmentation and recognition. Kawahara *et al.* [292] proposed a very deep FCRN with 50 layers using global and local contextual features for skin lesion segmentation which performed

better than the VGG-16. Chen *et al.* [293] implemented a 3D CRN for brain MRI using VoxRes modules with small convolutional kernels and stride two for the convolutional layers to reduce the input's resolution and capture a larger receptive field.

2.3.5 Recurrent Neural Networks(RNNs)

An RNN uses recurrent connections to memorize the last input's pattern [31]. This notion was derived from the fact that the anatomical objects in a volumetric image are distributed over multiple slices, that is, there is a correlation between successive slices. Generally, in an RNN, two parts that work on an input slice are: firstly, any type of CNN which extracts intra-slice information; and, secondly, the RNN itself which extracts inter-slice information.

The long short-term memory (LSTM) [294] method is an popular RNN. The standard one takes vectorized inputs which hinders its application for medical image segmentation due to its loss of spatial information. Therefore, researchers have proposed convolutional LSTM (CLSTM) [295, 296] for tackling this problem by exploiting a convolutional operation instead of vector multiplication. Chen *et al.* [270] proposed a bi-directional CLSTM (BDCLSTM) with a modified U-Net for z^+ and z^- directions which performs better than a pyramid LSTM [297] in which six directions (x^+ , x^- , y^+ , y^- , z^+ and z^-) are considered. Another variation of the LSTM is the gated recurrent unit (GRU) in which memory cells are excluded [298]. Xie *et al.* [299] used an clockwork RNN (CW-RNN) to segment a muscle's perimysium. It has long-term dependency using fewer parameters and a 5 percent better accuracy than the U-Net [264]. Xia *et al.* [72] proposed a segmentation method for paraspinal muscles using the U-Net with CRF as an RNN with multi-data training in MRI for the L4-L5 and L5-S1 spinal levels.

2.3.6 Attention-based CNNs

These models have recently been investigated extensively in the deep learning community to more effectively use intermediate feature maps [69]. They focus on the relevant parts of the input without additional supervision and have also been applied in semantic segmentation [69, 300].

Cascaded-based attention models, in which multi-stage CNNs are used to guide attention, have been applied in different medical applications. One stage is used to localize the ROI and another for dense predictions. Some use a bounding box to guide attention in the dense prediction stage during training which is called hard attention. Roth *et al.* [61] used holistically nested convolutional networks (HNNs) in three orthogonal views to generate a 3D bounding box by fusing their probability maps, and then two additional separate realizations of HNNs to segment a pancreas inside the box. Payer *et al.* [62] used a CNN to find the center of a bounding box around a ROI for use later to construct a fixed-size bounding box to crop the ROI and up-sample for the training dataset followed by a 3-stage segmentation CNN focusing on the ROI. Khened *et al.* [63] proposed a ROI extraction process using spatial-temporal variation statistics and a circular Hough transform for deep learning-based cardiac segmentation to reduce the computational and memory requirements. Huang *et al.* [64] used the backbone encoder of the U-Net to select the ROI of a colorectal tumor and combined it with the decoder for segmentation. Ngo *et al.* [65] combined a level set (active contour model) method with two deep belief networks (DBNs) (deep learning) to segment a left ventricle in the MICCAI 2009 challenge dataset (45 MRIs), with one used to locate the ROI and the other to delineate the target objects. Then, a level set optimizes the ROI using the delineation as an evolutionary constraint. However, their method is only semi-automatic as manual selection is used to obtain the initial target volume.

Liao *et al.* [66] used a modified U-net with two modules to obtain the probability of lung cancer. The first module defines a 3D region for all suspicious nodules which is refined by the second one in which the possibilities of cancer in the refined nodules are determined. Then, these possibilities are combined through a leaky noisy OR gate to obtain a subject's final probability of having cancer. Guan *et al.* [301] proposed a three-branch attention-guided CNN for the classification of thoracic diseases. In it, a global image and local region are used to train the global and local CNN branches, respectively, which are finally concatenated through a fusion branch, with a ResNet-50 and DenseNet-121 considered the backbone of the CNN in their architecture. A hard attention model via the global branch is used to generate a heat map which extracts the local context through a mask inference process. Pesce *et al.* [300] proposed a visual attention network using a bounding box for detecting pulmonary lesions by exploiting a few annotated X-rays and a huge number of weakly labeled images. This model consists of two architectures: the first extracts saliency maps from the convolutional layers and then calculates the localization error and back-propagates it along with the softmax classification error; and the second is a recurrent attention model which learns small portions of an image via reinforcement learning. Zhou *et al.* [302] used an iterative process to increase segmentation accuracy by exploiting a bounding box. In this approach, there are two models for training, a coarse-scale one which uses an entire image as input and a fine-scale one which is a ROI cropped by the bounding box with a zoomed-in view, and uses them iteratively during testing. This process is computationally expensive as cascaded models use excessive computational resources and repetitively extract similar low-level features.

Another popular attention-based model is the mask region-based CNN (Mask-RCNN). He *et al.* [303] proposed it by extending the faster R-CNN in which a

branch detects object instances and generates a bounding box like a segmentation mask. It outperforms other methods for many computer vision tasks and researchers have also tried to use it for medical image segmentation. Vuola. *et al.* [304] proposed an ensemble model by combining the U-Net and Mask-RCNN for nuclei segmentation, with the ensemble one outperforming both the others. They found that the Mask-RCNN can find the bounding box accurately but its segmentation performance is worse than that of the U-Net. Johnson. *et al.* [305] used the Mask-RCNN with ResNet-50 and ResNet-101 models as the backbone of a feature pyramid network to segment cell nuclei from microscopic images.

Some automatic attention models have been proposed recently. They avoid the cascaded approach and focus automatically on the ROI during training and implicitly suppress unwanted regions. They can be classified as supervised [71] and unsupervised [7, 67, 69] methods. However, the latter sometimes perform worse than those without an attention mechanism while the former use external supervision through coarse annotation. Schlemper *et al.* [7] proposed an unsupervised attention gate (AG) for focusing structures automatically during the training of CNNs which can implicitly suppress irrelevant regions and concentrate on task-specific salient features using a grid-based gating technique to look at local regions. When evaluated on classification and segmentation applications, it achieves better accuracy than base architectures. However, sometimes the gate cannot focus on the ROI, particularly for applications with compact multiple classes with large variabilities and low levels of contrast.

2.4 Summary

In this chapter, firstly, the different techniques used in CAD for different medical applications are discussed. Image registration and deep learning-based semantic segmentation are the two most important of many techniques, such as detection, classification and recognition, for medical image analysis. Then, the basic elements required for the design of a good image registration algorithm, such as geometrical transformation, a similarity measure and optimization, which are carefully selected, are described. In the literature, many registration algorithms have been developed by designing transformations, similarity measures and optimization for 2D and 3D images. The geometric transformations used in existing methods are discussed and classified in two main categories: parametric; and non-parametric models. The former is further classified into five main types based on RBFs, EBSs, FFDs, signal representations and locally affine models. On the other hand, the latter are further classified into four main types based on elasticity, fluid flow, diffusion and diffeomorphism. The similarity measures in existing registration algorithms are categorized as three main types: feature- and intensity-based and hybrid methods, with the intensity-based ones further classified as mono-modal and multi-modal. Alternatively, the hybrid ones are classified as initialization-, constraint- and coupled-based methods. Finally, continuous, discrete and miscellaneous optimization methods are discussed, with the continuous ones popular in medical image registration. Also, deep learning-based segmentation methods are described, with their most common architectures CNNs, FCNs, the U-Net, CRNs, RNNs and attention-based CNNs. The U-Net-based ones are popular in biomedical image segmentation applications and attention-based models have recently been studied extensively in the deep learning community.

Although previous image registration efforts form a strong base for many medical applications, none can satisfy the requirements of all applications due to some challenges associated with their implementation and low accuracy which could be due to anatomical complexities and variabilities or the artefacts of imaging devices. Therefore, to tackle the complexity and variability of neck muscles, in Chapter 3, a novel object-constrained hierarchical 3D-3D registration framework for aligning inter-subject neck muscles is proposed. Firstly, to handle the large-scale local minima of optimization, the framework uses a coarse registration technique which optimizes a new edge position difference (EPD) similarity measure to align large mismatches. Also, a new transformation based on the discrete periodic spline wavelet (DPSW), and affine and FFD transformations are exploited. Secondly, to avoid the monotony of using transformations in multiple stages, a fine registration technique for aligning small mismatches is designed. It uses a double-pushing system that changes the edges in the EPD and switches the resolutions of the transformations. The EPD helps both coarse and fine techniques to implement object-constrained registration via controlling edges which is not possible using traditional similarity measures. However, this method still has some limitations in terms of computational complexity and the handling of large-scale anatomical variabilities. To overcome these problems, another 3D-3D registration framework with two novel contributions is proposed in Chapter 4. Firstly, a two-stage heuristic search optimization technique for handling large mismatches which uses a minimal user hypothesis regarding large mismatches and is computationally fast is introduced. This optimization brings a moving image hierarchically closer to a fixed one using MI and an EPD similarity measure in the coarse and fine stages, respectively, while the images do not require pre-segmentation as is necessary in traditional heuristic optimization-based techniques. Secondly, a ROI EPD-based

registration framework for handling small mismatches using salient anatomical information (AI), in which a convex objective function is formed through a unique shape created from the desired objects in the ROI, is proposed. Finally, in Chapter 5, three different types of deep learning segmentation-based analysis and methodological propositions in a neck muscles segmentation application are presented. In the first analysis, an evaluation of U-Net-based architectures for neck muscles segmentation, including 10 recent models, is conducted, with the numerical results indicating that the R2Unet architecture achieves the best accuracy. In the second analysis, a new two-stage U-Net++ (TS-UNet++), which uses two different types of deep CNNs (DCNNs), the U-Net++ in the first stage and U-Net in the second, rather than a traditional multi-stage one is proposed. More convolutional blocks are added after the input and before the output layers of this multi-stage method to better extract both low- and high-level features. A new concatenation-based fusion structure is incorporated in the architecture to enable deep supervision, with more convolutional layers added after each concatenation to extract more representative features. In the third analysis, an explicit attention method in which the attention is performed through a ROI evolved from the ground truth via dilation is proposed. It does not require any additional CNN to localize the ROI, as does a cascaded approach, because attention in a CNN is sensitive to the area of a ROI. The dilated ROI is capable of capturing more relevant regions and suppressing irrelevant ones than a bounding box and region-level coarse annotation. It is used during the training of any CNN whereas coarse annotation, which does not require any detailed pixel-wise delineation that can be performed by any novice person, is used during testing.

Chapter 3

A Deformable 3D-3D

Registration Framework using the

Discrete Periodic Spline Wavelet

and Edge Position Difference

Conventional multi resolution-based registration methods trap in local minima due to their low degree of freedom geometrical transforms. This Chapter presents a novel object-constrained hierarchical registration framework for aligning inter-subject neck muscles. First, to handle large scale local minima, the proposed framework uses a coarse registration technique, which optimizes the new edge position difference (EPD) similarity measure, to align large mismatches. Also, a new transformation based on the discrete periodic spline wavelet (DPSW), affine and free-form-deformation (FFD) transformations are exploited. Second, to avoid the

The following research papers have been published based on this chapter:

- **Abdulla Al Suman**, Md. Asikuzzaman, Alexandra Louise Webb, Diana M. Perriman and Mark Richard Pickering, “Inter-Subject Image Registration of Clinical Neck MRI Volumes using Discrete Periodic Spline Wavelet and Free form Deformation,” *Digital Image Computing: Techniques and Applications (DICTA)*, 10–13 December. 2018, Canberra, Australia.
- **Abdulla Al Suman**, Md. Asikuzzaman, Alexandra Louise Webb, Diana M. Perriman, Murat Tahtali and Mark Richard Pickering, “A Deformable 3D-3D Registration Framework using Discrete Periodic Spline Wavelet and Edge Position Difference,” *IEEE Access.*, vol. 8, pp. 146116-146133, 2020, DOI 10.1109/ACCESS.2020.3015504.

monotonous nature of using transformations in multiple stages, a fine registration technique is designed for aligning small mismatches. This technique uses a double-pushing system by changing edges in the EPD and switching transformation resolutions. The EPD helps in both coarse and fine techniques to implement object-constrained registration via controlling edges, which is not possible when using traditional similarity measures. Experiments are performed on clinical 3D magnetic resonance imaging (MRI) scans of the neck, with the results showing that the EPD is more effective than the mutual information (MI) and sum of squared difference (SSD) measures in terms of the volumetric dice similarity coefficient (DSC). Additionally, the proposed method is compared with the state-of-the-art diffeomorphic Demons and SyN approaches with ablation studies in inter-subject deformable registration. The proposed method achieves better accuracy, robustness and consistency than the reference methods, with an average volumetric DSC of 0.7029 compared to 0.6654 and 0.6606 for the Demons and SyN methods, respectively.

The rest of this chapter is organized as follows. An introduction is presented in Section 3.1. Related work is discussed in Section 3.2, the details of the deformable 3D-3D registration method is described in Section 3.3, Section 3.4 presents the experimental procedure and results, and Section 3.5 and Section 3.6 provide a discussion and conclusion, respectively.

3.1 Introduction

Registration methods generally attempt to find a global minimum and avoid local ones, but cannot avoid all local minimums due to the many that are classified as small-scale dips and large-scale basins [306]. Escaping from basins is vital as

they yield large mis-registrations and usually occur in large initial mismatch scenarios. Conversely, dips yield small mis-registrations occurring in small mismatch scenarios and occur more often. Neck medical data contain both scenarios discussed above. Conventional multi resolution-based registration methods [34–37] cannot avoid basins and dips completely due to the lower available degree of freedom of the geometrical transformation used and the monotonous application of the same transformations in multiple stages. Information theoretic-based methods have been widely used for 3D-3D medical image registration [38–40]. However, global information theoretic-based methods are less sensitive to local deformation and often encounter mis-correspondences due to the lack of distinctiveness of the similarity measure (SM). Due to the compactness and similar appearance of neck muscles, the problem of mis-correspondences is more acute in neck data. Further, local information theoretic-based methods are computationally expensive and face statistical instability problems. In addition, information theoretic-based methods ignore anatomical information, which is crucial for guiding correspondence detection and registration. The feature-based registration methods establish correspondence through high-order anatomical information [27]. However, some features are often partially invariant in images with a different appearance, which is common in clinical applications. Some descriptor-type features show distinctive characteristics; nevertheless, they face challenges to performance in scenarios involving large anatomical variation.

A coarse-to-fine 3D-3D registration approach for dealing with the neck’s high level of variability is proposed in this chapter. It uses a hybrid registration framework divided into coarse and fine mismatch correction sections to handle the basins and dips, respectively, and exploits the diffeomorphic Demons algorithm in its last stage to boost alignment. The coarse and fine sections reduce the chances of becoming stuck in basins and dips, respectively, the latter through its double-pushing

system. This method makes the following four key contributions:

- In this multistage framework, multiple 3D transformations at the coarse level are used in multiple stages for large mismatches, which are usually of different types. Therefore, they can push basins from different directions with greater angle differences, as opposed to similar transformations in conventional multi resolution methods, thereby reducing the chance of becoming stuck in basins.
- A novel discrete periodic spline wavelet (DPSW)-based 3D transformation, which requires fewer parameters than the free-form-deformation (FFD) approach but has similar benefits, is developed in the coarse section. It reduces the burden of optimization and provides variations at the global level with no stretching or shrinking effect, unlike a discrete cosine (DC) transform, which is popular in video coding.
- The framework incorporates a new SM called the edge position difference (EPD), which uses a modified 3D Chamfer distance transform algorithm. Since it uses the edges of an object, it provides an opportunity to tune through multiple stages using different sets of edges. It aligns the neck's trunk using the same set of strong edges in the multistage coarse section and, gradually, incorporates the weaker ones for the muscles and other small objects in the fine section. It can do object-wise alignment through multiple transformations, whereas traditional multi resolution methods use SM methods that are incapable of object-wise alignment.
- A double-pushing system is designed for the fine section to reduce the chance of becoming stuck in dips, which occur more often than basins, whereas a

single-pushing system is used in most multi resolution methods. The double-pushing system yields small deformations formed through changing the number of edges of the EPD and the resolution levels of the transformation, rather than changing only the latter, as in traditional multi resolution techniques.

The EPD, mutual information (MI) and sum of squared difference (SSD) SMs are compared using the affine transformation and the EPD achieves good accuracy for a clinical dataset. Additionally, the proposed method is compared with the diffeomorphic Demons [2] and SyN [3] algorithms, which are state-of-the-art registration approaches, and outperforms various other non-rigid registration algorithms [307]. The volumetric dice similarity coefficient (DSC) is computed for a real clinical 3D MRI dataset using the proposed, diffeomorphic Demons and SyN methods, with the proposed method achieving a substantial improvement in accuracy.

3.2 Related Work

Image registration [24, 308, 309] is a basic image-processing technique whereby two or more images are aligned by keeping one stationary (called a fixed image) and moving another (called a moving image) towards it [27]. It comprises a geometrical transformation, similarity measure and optimization, with 3D-2D registration currently being developed commercially and rigid registration [310] practically available [311]. Inter-subject registration (ISR) in 3D, a kind of deformable image registration [56], is a key challenge due to anatomical variability [311] preventing uniformity. Although much work on 3D-3D ISR has been conducted in the last two

decades, mainly on brain images, its accuracy is not clinically acceptable, with specialists considering that further research must yet be undertaken [311, 312]. Moreover, 3D-3D ISR often faces some problems compared to 2D-2D registration. First, the optimization becomes difficult as more parameters are required for geometrical transformation. Second, more dips and basins are encountered due to the higher-dimensionality. Third, the computational cost becomes expensive. Finally, the SM faces discontinuity in intensity problems along out of plane directions, since most medical imaging modalities keep spacing between the slices. Most studies on 3D-3D ISR have focused on optimization [36, 307, 313], local regularization [314, 315], multi resolution FFD [34, 35] and the application of the diffeomorphic log-demons algorithm [45, 316] to 3D-3D ISR. The most recently developed competitive registration algorithms are the multi resolution FFD and diffeomorphic log-demons approaches. Hua *et al.* [34] proposed a 3D-3D deformable registration for handling discontinuities by adding extra degrees of freedom to a multi resolution framework using a parameter up-sampling method that required segmenting a target image to determine discontinuities and allowing more time to optimize additional parameters. Sun *et al.* [120] proposed a random perturbation technique for a multi resolution nonlinear registration framework for 3D-3D and 2D-2D applications using a lower-order B-spline, retaining the same smoothness as a higher-order spline to reduce the execution time. However, in terms of accuracy, it could not perform well in other clinical applications. Sun *et al.* [121] proposed another 3D-3D simultaneous multi resolution strategy in which different resolutions of the spline and data were used together to improve performance, but their method was dependent on a parameter whose value varies for different applications.

Overall, conventional multi resolution methods use only the coarse resolution levels of a transformation to resolve large mismatches. These are not sufficient to achieve proper correction, since, when the optimization process is stuck in a

basin, to escape easily, it must be pushed in a specific direction depending on the particular basin. However, consecutive resolution levels in multiple stages of multi resolution methods have almost the same characteristics with different directions and small variations in angle but are considered different transformations. Therefore, consecutive coarse levels may not thrust in the required direction for some basins because of their small angle variations. Large angle variations among multiple transformations in multiple stages may be required so that a transformation can push basins in the appropriate directions. The optimization process will not face any difficulty to determine the optimal direction because the multiple transformations are not applied simultaneously rather they are applied consecutively as each stage of the algorithm is performed. Therefore, as traditional multi resolution methods cannot push in the required direction for some basins, they are unable to avoid all the basins. Further, these methods try to tackle small mismatches, which cause dips more frequently, by changing only the fine resolution levels of a transformation in multiple stages with the same SM and optimization, and are thus incapable of eliminating all of them.

Feature-based registration methods have been applied in 3D-3D ISR due to their use of anatomical information, which helps to find correspondence detection effectively. There are many types of features in the registration literature, such as histogram of oriented gradient (HOG) [317], gradient location and orientation histograms (GLOH) [318], scale-invariant feature transform (SIFT) [44] and speeded up robust features (SURF) [43]. The gradient information-based features are partially invariant [27], which causes mis-registration. The SIFT and SURF methods require the same features to be detectable in both fixed and moving images, which is not possible in neck MRI data as the muscles are very compact and images are obtained with different acquisition protocols. In fact, SIFT and SURF are

more suitable for natural image analysis than medical image analysis. The Random Sample Consensus (RANSAC) algorithm is often used in conjunction with feature-based registration methods to filter out excellent matches. For example, Kahaki *et al.* [319] proposed a local intensity maxima feature-based registration method for in vivo time lapse microscopy images. They used a two-step feature matching procedure in which features are initially matched coarsely and then the matching features are refined through RANSAC. The iterative closest point (ICP) algorithm is a popular method in shape registration [320–322], which has high accuracy for point set registration. A 3D Canny edge-based objective function is used in medical image registration for pose estimation and shape reconstruction [323], multi-modal geometric matching [324] and respiratory motion correction [325]. A self-similarities-based feature called a modality-independent neighborhood descriptor (MIND) [326] was proposed to provide distinctive correspondences in the objective function by incorporating information from the neighborhood pixels. It showed better results in cases of similar local structural patterns in small regions than other SMs. However, its performance can be limited in cases of large local anatomical variation. Further, it cannot hide the influence of contrast enhancement and embeds unwanted information.

3.3 Deformable 3D-3D Registration Model

3.3.1 Overview

A diagram of the operational flow of the proposed hybrid registration framework is displayed in Figure 3.1. As a pre-processing step, all the original fixed and moving MRI volumes are trimmed and interpolated to volumes of $128 \times 128 \times 128$ voxels,

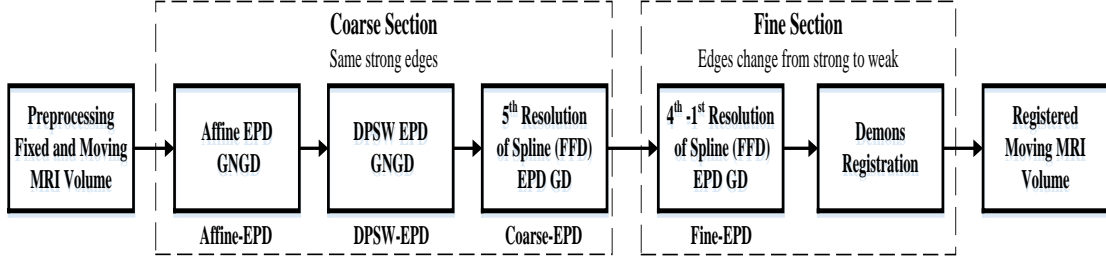


Figure 3.1: Proposed hybrid registration framework in which gradient descent (GD) and Gauss-Newton gradient descent (GNGD) optimization techniques are used. The same set of edges is used in the coarse section by the different registration stages. Different sets of edges are used in the fine section by the different stages. The geometrical transformations are separable (not composite), since each stage takes input from the previous stage. The labels Affine-EPD, DPSW-EPD, Coarse-EPD and Fine-EPD are used to describe the experimental results in Section 3.4.

since they contain some unwanted information. The volume of interest is selected as the volume between the C1 and C7 vertebral levels, which represent the top and bottom vertebrae of the neck and are the landmarks most commonly used to assess muscle morphometry [9]. The processed volumes are then manually delineated to obtain the ground truths, as discussed in detail in Section 3.4. The framework is divided into coarse and fine sections. All the geometrical transformations at the different stages in Figure 3.1 are separable, since each stage uses the registered moving volume of the prior stage as its moving volume and the same fixed volume used by the prior stage as its fixed volume. Thus, the transformations are obtained separately from the framework. To describe experimental results in Section 3.4, the following labels have been assigned for different stages of the algorithm: Affine-EPD, DPSW-EPD, Coarse-EPD and Fine-EPD.

3.3.1.1 Coarse Section

All the stages in this section use the same strong edges of the MRI volumes to align the neck's trunk and the boundaries of other large objects. In each stage,

the geometrical transformation is changed to avoid basins and, importantly, obtain a good alignment. Other elements, such as the SM or optimization method, could also be changed in each stage. A local minimum is considered for a specific combination of the transformation, SM and optimization. The affine, the DPSW and the coarsest level of the FFD are used as transformations to combat basins. Therefore, multiple transformations can attack the optimization from different directions using large angle variations to pull out from basins.

3.3.1.2 Fine Section

In this section, five stages are used to obtain fine deformations. This section helps to reduce the chance of the optimization algorithm converging to a dip. As dips occur more frequently than basins and could cause the optimization to be stuck at any stage, a double pushing system has been designed to combat dips. The double pushing system is implemented by changing the transformation and SM simultaneously at every stage. The transformation change is performed by using different levels of the spline in the FFD. The SM change is accomplished by using the attributes of the EPD which allow different sets of edges to indicate different ranges of values. Different sets of edges for specific volume pairs are used for different stages. These sets of edges are changed gradually from strong to weak. The strong edges are a subset of the set of weaker edges. The first four stages in this section use coarse to fine levels of the spline for the FFD with corresponding different sets of strong to weak edges, respectively. The fine deformations are achieved by using the weak sets of edges and the fine levels of the spline. The strong sets of edges and the coarse levels of the spline are used to correct coarse deformations. The gradual change protects the framework against mis-correspondences. Finally, the Demons algorithm [2, 148] is applied in the fifth

stage of the fine section to correct more fine mismatches. Actually, the stages before Demons bring the moving volume closer to the fixed volume which helps Demons to align more effectively than when only using the Demons method.

3.3.2 Geometrical Transformation

The choice of transformation has a large effect on the registration process [48], with the most appropriate one not known a priori [79]. In the registration process, the transformation parameters are estimated using an optimization technique, with the number of them referred to as the deformation's degrees of freedom. In neck muscles application, the registration needs to be performed between the MRI volumes of two different individuals' necks; this cannot be achieved using only an affine or rigid transformation because both have a limited number of parameters. Therefore, an elastic transformation with a higher degree of freedom is required to tackle the morphological complexity and variability of the population. However, to deal with the neck's variability, a mixture of affine, DPSW and FFD transformations is used to align the neck's trunk first by exploiting the advantages of the EPD. The same strong edges that correspond mainly to the neck's trunk are used to correct a coarse mismatch with a different transformation. This is because a good deformation cannot be achieved through a single transformation.

In this study, considering $F(x, y, z)$ and $I(x', y', z')$ as fixed and moving volumes, respectively, their coordinates are involved in elastic registration as follows:

$$\begin{aligned} x'_i &= x_i + \sum_{k=1}^{P/3} m_k \varphi_k(x_i, y_i, z_i) \\ y'_i &= y_i + \sum_{k=P/3+1}^{2P/3} m_k \varphi_k(x_i, y_i, z_i) \\ z'_i &= z_i + \sum_{k=2P/3+1}^P m_k \varphi_k(x_i, y_i, z_i) \end{aligned} \quad (3.1)$$

where m_k are the motion parameters, k is the parameter index, P is the total number of motion parameters and φ_k are the basis functions for the complex mapping, given as:

$$\varphi_k(x_i, y_i, z_i) = \varphi_{k+P/3}(x_i, y_i, z_i) = \varphi_{k+2P/3}(x_i, y_i, z_i). \quad (3.2)$$

There are many types of basis functions in the literature, including polynomial, Fourier, radial, B-spline, DC and wavelet. Of these, the Fourier, B-spline and wavelet functions support a multi resolution decomposition that provides a coarse-to-fine representation of the displacement field. Hence, these basis functions are usually used in medical image registration. However, wavelet basis functions can achieve a local deformation more effectively than Fourier basis functions due to their localization in both the frequency and spatial domains [327]. The DPSW and FFD transformations will be described in Section 3.3.2.1 and Section 3.3.2.2.

3.3.2.1 Discrete Periodic Spline Wavelet

The Cai-Wang [328] wavelet is used, which is compactly supported, in the proposed DPSW-based basis functions and a fourth-order B-spline as a scaling function. The

wavelet is arranged in a periodic form with the center of the main lobe translated to a coordinate origin.

The fourth-order B-spline is defined as:

$$\phi(x) = \frac{1}{6} \sum_{a=0}^4 \binom{4}{a} (-1)^a (x-a)_+^3 \quad (3.3)$$

where, for any integer (n) :

$$x_+^n = \begin{cases} x^n & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (3.4)$$

Then, the spline wavelet is:

$$\psi(x) = \frac{-3}{7}\phi(2x) + \frac{12}{7}\phi(2x-1) - \frac{3}{7}\phi(2x-2). \quad (3.5)$$

The DPSW-based basis functions are:

$$\begin{aligned} \varphi_k(x_i, y_i, z_i) &= \varphi_{k+P/3}(x_i, y_i, z_i) = \varphi_{k+2P/3}(x_i, y_i, z_i) \\ &= \psi_\lambda(xu)\psi_\lambda(yv)\psi_\lambda(zw) \end{aligned} \quad (3.6)$$

where $k = 2su + sv + w + 1$, $u, v, w = 0, 1, 2, \dots, s-1$, $s = \sqrt[3]{\frac{P}{3}}$ and $\psi_\lambda()$ is the DPSW.

The supports are 4 and 3 for the spline and spline wavelet respectively, as shown in Figure 3.2, with the periodic spline wavelet generated by considering the wavelet as one period. The proposed 3D basis functions are generated from the fifth resolution of the wavelet using a point-to-point multiplication of the 1D functions, whereas the FFD uses a tensor product of the 1D non-periodic third-order

spline to generate basis functions. The DPSW-based transformation requires fewer parameters than the spline or B-spline-based wavelets to represent local deformations. Specifically, one basis function in the DPSW can represent local deformations over an entire image, whereas the B-spline or spline wavelet cannot, since they require more parameters to represent the same level of local deformation for a whole image, which places a burden on the optimization process. In particular, for a $128 \times 128 \times 128$ image, a B-spline-based transformation requires 15, 27 and 51 parameters for the fifth, fourth and third resolutions, respectively, while the DPSW uses only 24 parameters for all resolutions. Although DC-based basis functions can obtain local deformations over an entire image and are widely popular in video and various image-processing applications [329], they cause shrinking and stretching in several parts of the image. There are two reasons for these effects. First, the parameterizations in the available DC-based basis functions do not use the full cycle of a cosine wave within a cubical image support, whereas multiple cycles of the periodical spline wavelet are used in the DPSW-based basis functions, which have greater variations in values, as shown in Figure 3.3. Registrations are also performed using the DPSW and DC on the 3D MRI volumes shown in Figure 3.4 and obtain stretching effects in the latter's results. Second, there is a lower span in the negative lobe of the spline wavelet with regularization when compared to the cosine one.

3.3.2.2 Free Form Deformation

The fourth-order B-spline as defined in (3.3) is used in the FFD-based basis functions to obtain smoother local deformations than those in the traditional FFD

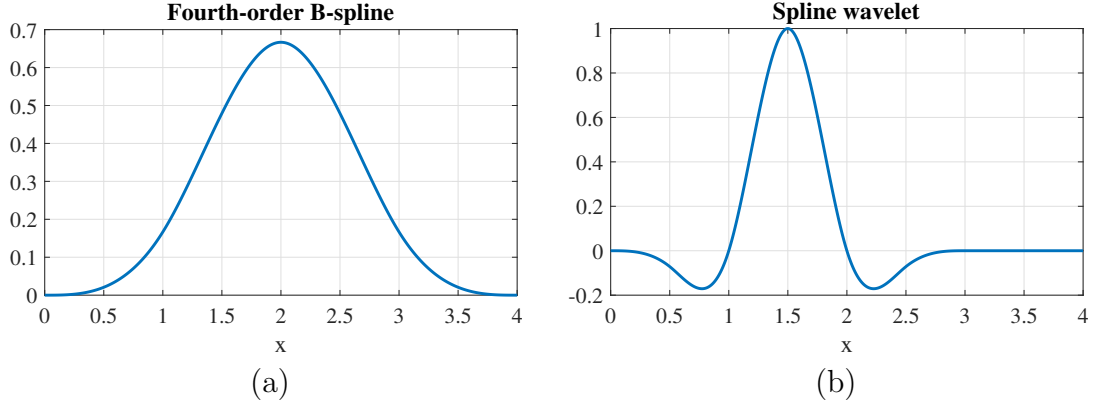


Figure 3.2: (a) 1D B-spline and (b) spline wavelet. The supports of the B-spline and spline wavelet are 4 and 3, respectively.

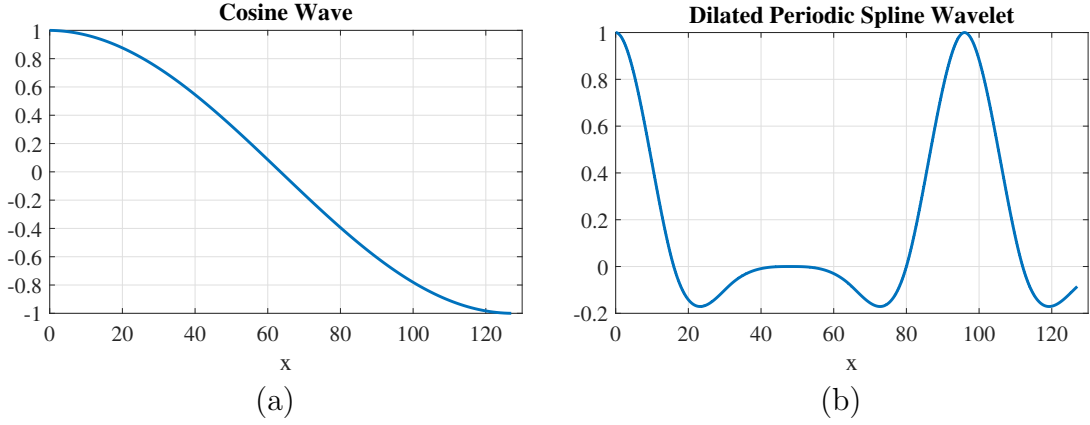


Figure 3.3: (a) 1D cosine and (b) periodic spline wavelet within an image support of $128 \times 128 \times 128$. Basis functions in discrete cosine

$\varphi_k(x_i, y_i, z_i) = \cos\left(\frac{(2x_i+1)\pi u}{2M}\right) \cos\left(\frac{(2y_i+1)\pi v}{2N}\right) \cos\left(\frac{(2z_i+1)\pi w}{2O}\right)$. In discrete periodic spline wavelet, $\varphi_k(x_i, y_i, z_i) = \psi_\lambda(xu)\psi_\lambda(yv)\psi_\lambda(zw)$, where M , N and O are the volumetric image's dimensions. The wave and wavelet are at the coarsest scale resolution level for the image support.

using the third-order B-spline, with the basis functions:

$$\varphi_k(\mathbf{x}) = \sum_{j=R}^L 2^{-j} \sum_{\mathbf{t}=-\mathbf{N}_j/2}^{\mathbf{N}_j/2} \Phi(2^{-j}\mathbf{x} - \mathbf{t}2^j) \quad (3.7)$$

$$\mathbf{N}_j \equiv 2^{-j}\mathbf{N} = 2^{-j}(N_x, N_y, N_z)$$

$$\mathbf{t} = (t_x, t_y, t_z), \mathbf{x} = (x, y, z)$$

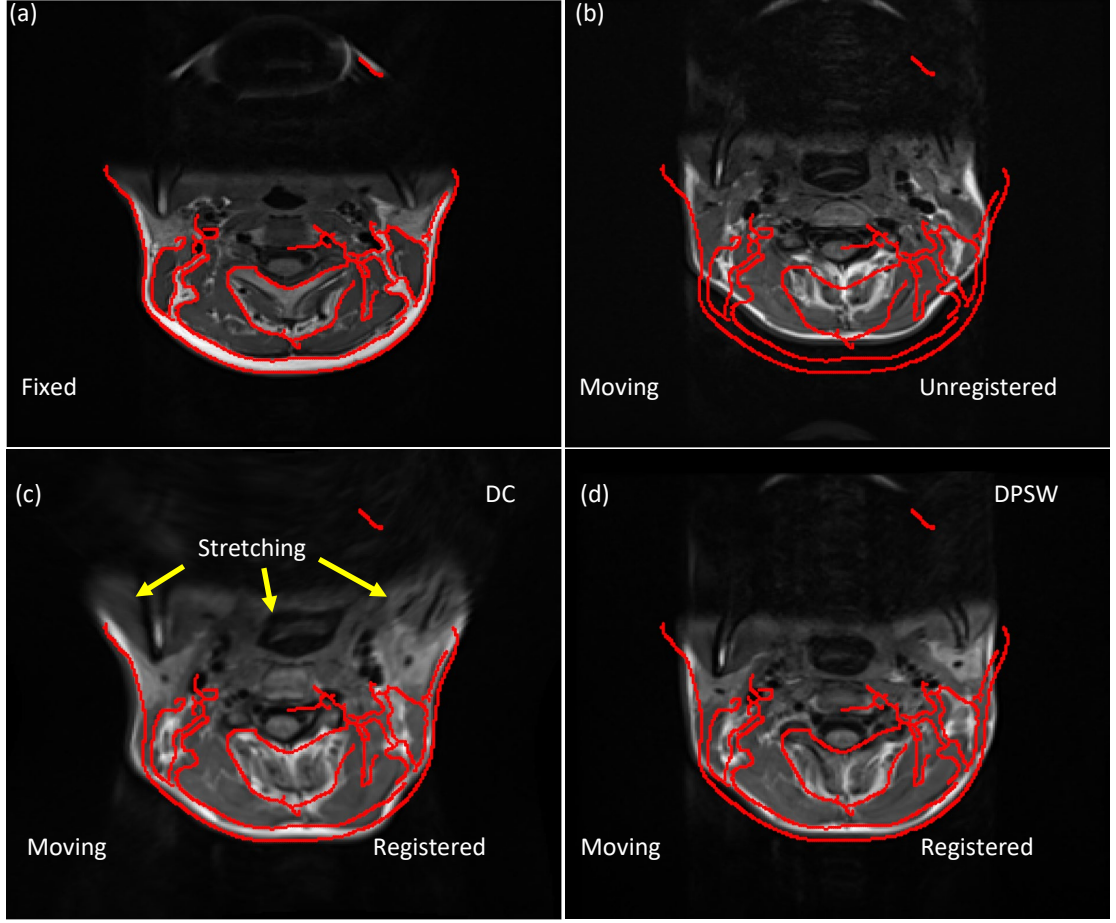


Figure 3.4: Registration results for two magnetic resonance imaging volumes using discrete cosine- and discrete periodic spline wavelet-based basis functions, showing a stretching effect in one slice of the former. The edges of the fixed image are superimposed in red over the unregistered and registered moving images. The registration is performed in 3D-3D, but the results are shown in 2D by taking one image from the 3D volume.

where t is the translation index, j is the resolution level, R is the starting resolution level, L is the maximum resolution level and $\Phi()$ is the basis functions. As L depends on the volume's size, the number of parameters for a resolution level is $(2^{-j}N + 1) \times 3$ if $N = N_x = N_y = N_z$ and the 3D basis functions use the tensor product of the 1D spline, given by:

$$\Phi = \phi(2^{-j}x - t_x 2^j)\phi(2^{-j}y - t_y 2^j)\phi(2^{-j}z - t_z 2^j). \quad (3.8)$$

3.3.3 EPD Similarity Measure

As the SM is another significant part of the registration process, in the proposed method, a new measure called the EPD is leveraged in the proposed coarse-to-fine registration framework to accomplish alignments in both the coarse and fine levels. This measure is based on the hierarchical Chamfer matching algorithm [330] and determines the distance between the corresponding edges in two images. The EPD is not exactly the same as root mean square (RMS), Euclidean distance or closest distance which are commonly used in shape matching. First, the EPD uses the arithmetic mean of Chamfer values of the moving image at the position of edges in the fixed image. Second, the Chamfer distance uses an approximation of the Euclidean distance to calculate distance from a pixel to the nearest edge. Third, the edge points of the fixed image may map to Chamfer values which may correspond to different edges other than the edge in the fixed image. One of the most popular intensity-based SMs in the literature is the MI measure. However, it is not suitable for neck MRI datasets in which multiple muscles are near each other and have similar compositions, with large deformations between subjects. Therefore, using a MI-based SM causes mis-registrations between MRI volumes. Conversely, the EPD is a feature-based SM that uses the edges of the muscles and the neck's trunk. In this registration technique, every edge pixel in a moving image contributes to a registration error with a value proportional to the distance to its closest edge in the reference image. Therefore, the edges of overlapping images are attracted to each other, which leads the EPD SM to have better registration accuracy than the MI measure, as justified in Section 3.4.3. Moreover, the EPD supports coarse-to-fine tuning because it uses edges that are controllable by selecting their detection thresholds. Consequently, since it is more suitable than the MI, this new SM is used in the proposed framework.

To calculate the EPD, first, edge volumes of the moving I and fixed F volumes (E_I and E_F , respectively) are calculated using a canny edge detector, with a value of 1 corresponding to an edge; otherwise, the value is 0. The strong or weak edges are selected by choosing thresholds in the canny edge detector, which assists in designing a double-pushing system for combating dips in the fine section of the registration framework. Then, the locations of the edges (β) in E_F are calculated and the distance transform volume (C) from E_I is determined using the 3D Chamfer distance transform method. This transform is an approximation of the Euclidean distance transform, in which each value of a voxel represents a distance to the closest edge in the edge volume. Finally, the EPD S , which is the arithmetic mean of the values of the voxels in the β positions, is calculated as:

$$S(m_k) = \frac{1}{3T} \sum_{(x,y,z) \in \beta} C(x', y', z') \quad (3.9)$$

where T is the total number of edge voxels in β , 3 is used to compensate local distance in the Chamfer distance.

Figure 3.5 presents an example of a MRI slice with its corresponding edge image and the latter's Chamfer distance image, obtained from a 3D MRI volume using the 3D Chamfer distance algorithm with local distances of $d_1 = 3$, $d_2 = 4$ and $d_3 = 5$ [1]. However, if there are no edges in some consecutive slices of a 3D volume or some local 3D regions of it, infinity remains in the border voxels of the cubical support. Therefore, the original 3D Chamfer distance transform algorithm is modified, as described below.

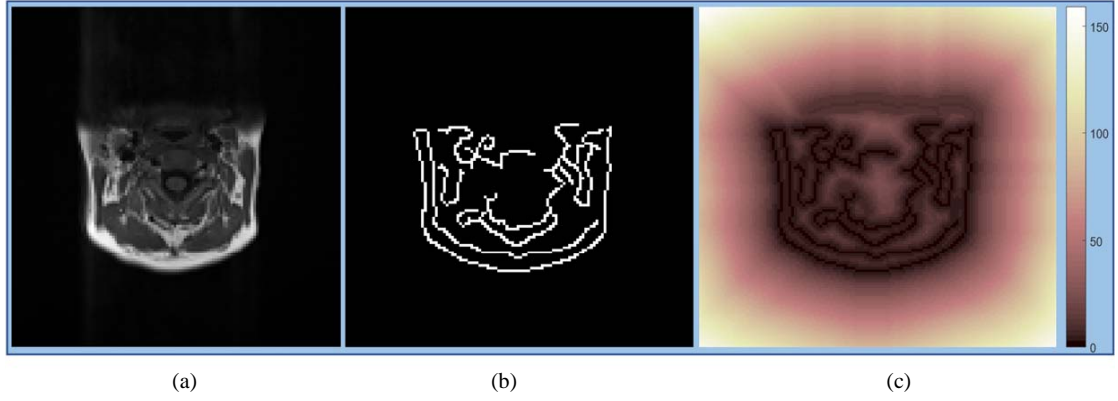


Figure 3.5: (a) Magnetic resonance imaging slice. (b) Edge image of (a). (c) Chamfer distance image of (b). The binary and Chamfer distance images are calculated for a 3D image but the results are shown in 2D.

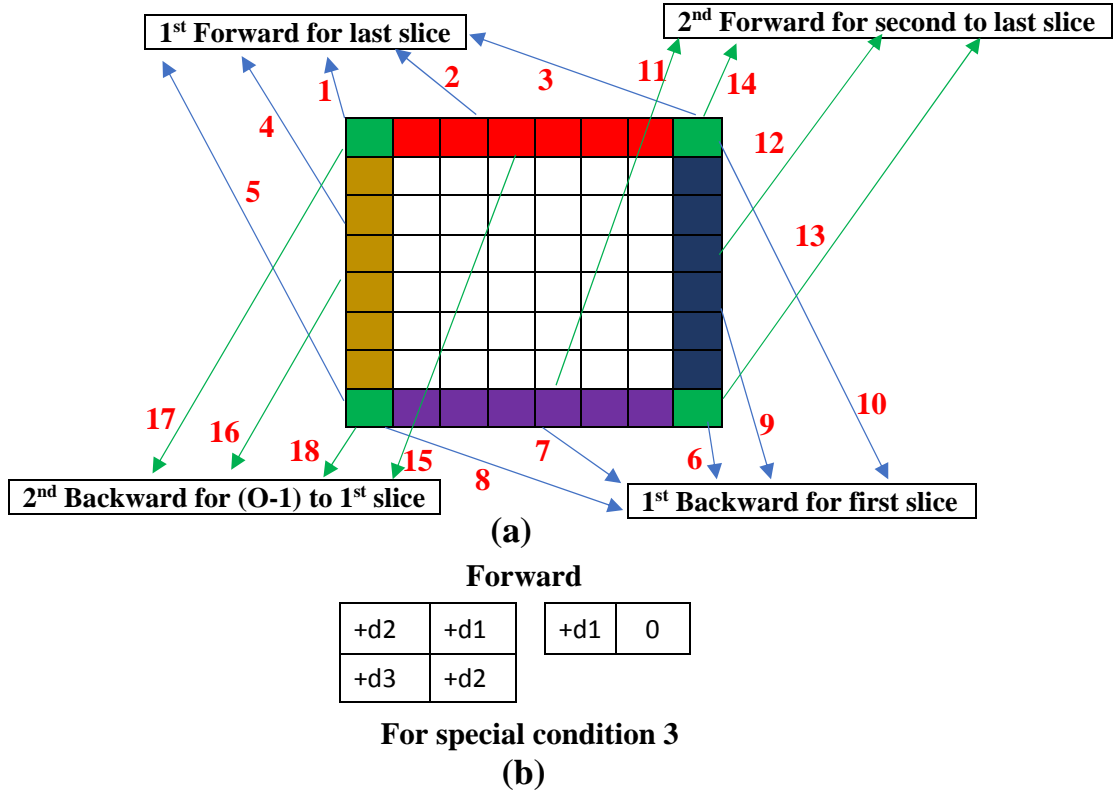


Figure 3.6: (a) Auxiliary schematic diagram for modified 3D Chamfer distance algorithm (voxels of the same color use the same mask modified from the original [1], with O slices in volume and each red number meaning a special condition corresponding to the modified mask). (b) Mask for special condition 3 of (a). The masks are used during traversing for the specified slices in the text boxes.

3.3.3.1 Modified Three-Dimensional Chamfer Distance Transform Algorithm

To tackle the problem mentioned above regarding the boundary voxels, two forward and two backward passes are used rather than only two passes as in the original algorithm. The new algorithm handles the other voxels similarly to those in the original algorithm. The boundary voxels of the last and first slices are processed separately in the first forward and backward passes, respectively. The masks of the original algorithm are changed according to the positions of the boundary voxels and are used in the proposed algorithm, with Figure 3.6 (a) showing their modified positions and Figure 3.6 (b) showing the modified mask for special condition 3. There are five special conditions in the first forward and backward passes, and four in the second ones. The framework is summarized in Algorithm 1.

3.3.4 Transformation Parameters Optimization

Optimizing a registration is considered ill posed and is actually a multidimensional problem that maximizes or minimizes the SM with respect to the transformation parameters. Generally, optimization methods for medical image registration are classified into three categories: continuous, discrete and miscellaneous. The first category includes GD, conjugate gradient, Powell's conjugate directions, quasi-Newton, Gauss-Newton (GN), Levenberg-Marquardt and stochastic GD approaches, the second includes graph-based, belief propagation and linear programming techniques and the third includes greedy and evolutionary algorithms. The Powell's conjugate directions and stochastic GD methods have been applied for transformations with low degrees of freedom [48], while evolutionary algorithms, which are used mainly in linear registration, have shown slow convergence. Hence,

Algorithm 1 Modified 3D Chamfer distance transformation.

Input E_I, Q .
Output C .

- 1: Initialize $d_1 = 3, d_2 = 4$ and $d_3 = 5$, obtain dimensions M, N and O from E_I and pre-allocate output matrix (C) with the same dimensions as E_I by zeros.
- 2: **for** $r = 1 : O$ **do**
- 3: **for** $p = 1 : M$ **do**
- 4: **for** $q = 1 : N$ **do**
- 5: Update C by Q where the value of E_I is zero, and by zero otherwise.
- 6: **end for**
- 7: **end for**
- 8: **end for**
- 9: **for** $r = 2 : O$ **do**
- 10: **for** $p = 2 : M$ **do**
- 11: **for** $q = 2 : N$ **do**
- 12: **if** $p < M \wedge q < N$ **then**
- 13: Update C according to forward mask in Figure 17 in [1].
- 14: **else if** $r = O$ **then**
- 15: **for** $p = 1 : M$ **do**
- 16: **for** $q = 1 : N$ **do**
- 17: **if** $p = 1 \wedge q = 1$ **then**
- 18: Update C according to the first special condition in Figure 3.6.
- 19: **else if** $p = 1 \wedge 1 < q < N$ **then**
- 20: Update C according to the second special condition in Figure 3.6.
- 21: **else if** $p = 1 \wedge q = N$ **then**
- 22: Update C according to the third special condition in Figure 3.6.
- 23: **else if** $1 < p < M \wedge q = 1$ **then**
- 24: Update C according to the fourth special condition in Figure 3.6.
- 25: **else if** $p = M \wedge q = 1$ **then**
- 26: Update C according to the fifth special condition in Figure 3.6.
- 27: **end if**
- 28: **end for**
- 29: **end for**
- 30: **end if**
- 31: **end for**
- 32: **end for**
- 33: **end for**
- 34: **for** $r = O - 1 : 1$ **do**
- 35: **for** $p = M - 1 : 1$ **do**
- 36: **for** $q = N - 1 : 1$ **do**
- 37: Update C as first forward pass using backward mask in Figure 17 in [1] and special conditions 6, 7, 8, 9 and 10 in Figure ?? for the first slice.
- 38: **end for**
- 39: **end for**
- 40: **end for**
- 41: **Return** C

GN and GD approaches have been used in many medical image registration approaches. In this application of mono modal registration, the GNGD method is used because the EPD involves summing the function values.

The optimization procedure required to find m_k is described as follows:

$$\frac{\partial S}{\partial m_k} = 0. \quad (3.10)$$

To minimize S , it is necessary to estimate its values in a small neighborhood of m_k , as:

$$\frac{\partial}{\partial m_k} S(m_k + \Delta m) = 0 \quad (3.11)$$

where Δm is a vector added to m_k in each iteration.

The value of Δm is determined using the first-order Taylor series approximation, as:

$$\begin{aligned} \frac{\partial}{\partial m_k} [S(m_k) + S'(m_k)\Delta m] &= 0 \\ \Rightarrow \nabla S(m_k) + \nabla^2 S(m_k)\Delta m &= 0 \\ \Rightarrow \Delta m &= -\nabla S(m_k) \left(\nabla^2 S(m_k) \right)^{-1}. \end{aligned} \quad (3.12)$$

The gradient $\nabla S(m_k)$ is calculated as:

$$\begin{aligned} \nabla S(m_k) &= \frac{\partial S}{\partial x'} \frac{\partial x'}{\partial m_k} + \frac{\partial S}{\partial y'} \frac{\partial y'}{\partial m_k} + \frac{\partial S}{\partial z'} \frac{\partial z'}{\partial m_k} \\ &= \sum_{\beta} \left(\frac{\partial C}{\partial x'} \frac{\partial x'}{\partial m_k} + \frac{\partial C}{\partial y'} \frac{\partial y'}{\partial m_k} + \frac{\partial C}{\partial z'} \frac{\partial z'}{\partial m_k} \right) \end{aligned} \quad (3.13)$$

where $\frac{\partial C}{\partial x'}$, $\frac{\partial C}{\partial y'}$ and $\frac{\partial C}{\partial z'}$ are the spatial gradients of C , while $\frac{\partial x'}{\partial m_k}$, $\frac{\partial y'}{\partial m_k}$ and $\frac{\partial z'}{\partial m_k}$ are obtained from corresponding transformation basis functions (φ_k) in equations (3.6) and (3.7).

The Hessian matrix $\nabla^2 S(m_k)$ in equation (3.12) is defined in the GN optimization technique as:

$$\nabla^2 S(m_k) = J(m_k)J(m_k)^T \quad (3.14)$$

where $J(m_k)$ is the Jacobian of $C(x', y', z')$, calculated as:

$$J_{b,c}(m_k) = \frac{\partial C(x', y', z')}{\partial m_k}, \quad \text{for } (x', y', z') \in \beta \quad (3.15)$$

where b is the parameter indices and c is the position index of β . Finally, the vector Δm is obtained from equation (3.12), which is used to update the transformation parameters as follows:

$$m_k^{l+1} = m_k^l + \Delta m. \quad (3.16)$$

As most of the elements in the fine resolution's basis functions are zero in the FFD transformation, most of those in the Hessian matrix become zero, since this matrix is calculated by multiplying the spatial gradients of C and basis functions in the β positions, making it impossible to invert the matrix. Therefore, this matrix is omitted in the FFD transformation, which changes the optimization to a GD method.

3.4 Experimental Procedure and Result Analysis

3.4.1 Data and Annotations

The experiments were performed on neck MR images as part of a larger study of neck pain undertaken at the Australian National University (ANU) Medical

Table 3.1: Demographic data and magnetic resonance imaging scanner parameters.

Patient index	Age (years)	Weight (kg)	Height (m)	Repetition time (s)	Echo time (s)	Acquisition date (yyyymmdd)
PT-1	29	57	1.61	746	15	20140106
PT-2	19	59	1.65	737	15	20141216
PT-3	25	57	1.67	827	15	20131217
PT-4	20	56	1.55	750	15	20141212
PT-5	32	59	1.64	766	15	20141202
PT-6	23	45	1.60	827	15	20140922
PT-7	23	61	1.72	827	15	20141113
PT-8	27	61	1.62	827	15	20141124
PT-9	24	75	1.75	949	15	20140120
PT-10	27	57	1.65	777	15	20141201
PT-11	32	43	1.50	827	15	20141209

School. The study had ethics approval from the Human Research Ethics Committees of the ANU and Australian Capital Territory Health, with written informed consent from all participants. T1 SE axial spin echo MR images with voxel spacings of $0.8594 \text{ mm} \times 0.8594 \text{ mm} \times 4 \text{ mm}$ and sizes of $256 \times 256 \times 45$ were acquired using a 3 Tesla Skyra scanner (Siemens, Erlangen, Germany) and then cropped and interpolated to $128 \times 128 \times 128$. Data from 11 participants collected at different times using different MRI machine settings were used in this experiments. The participants' demographics and the MR sequence parameters for each scan are shown in Table 3.1.

An ANU graduate-entry medical student, with a degree in anatomy, manually delineated the right and left sternocleidomastoid, semispinalis capitis and splenius capitis muscles between the C1 and C7 vertebral levels. These segmentations were validated and edited by two medical experts from the ANU Medical School and Canberra Hospital. A MATLAB graphical user interface was developed for this research for segmentation which allowed the use of as many vertices as necessary to capture small details of the contours of these muscles. The contours of a slice

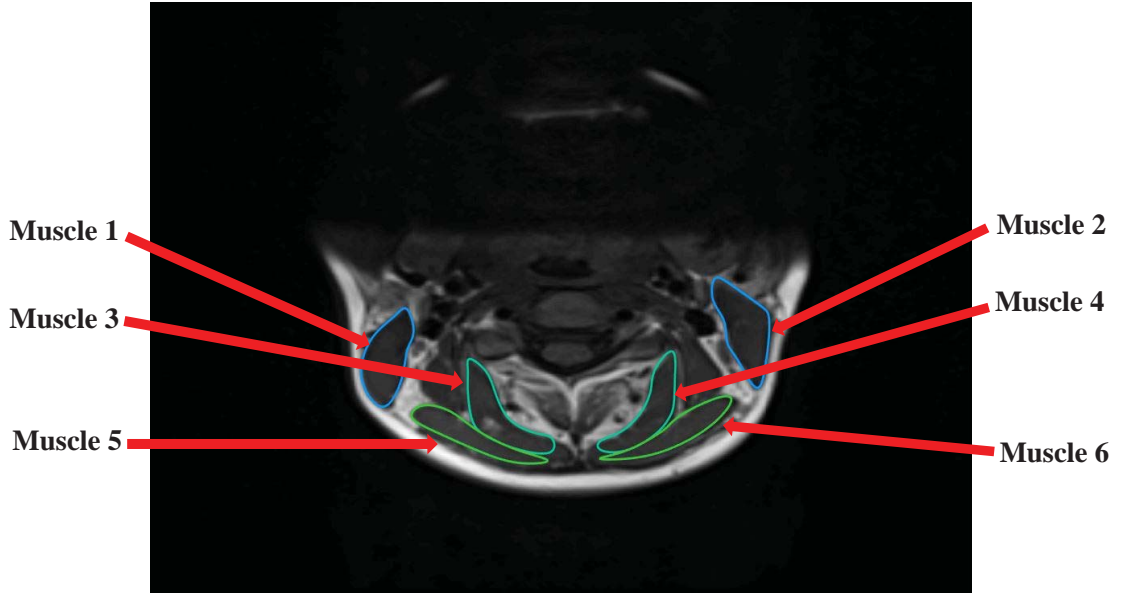


Figure 3.7: Annotated contours of an axial magnetic resonance image. Different colors are used for each separate muscle due to annotation convenience; however, the same color is used for symmetric muscles.

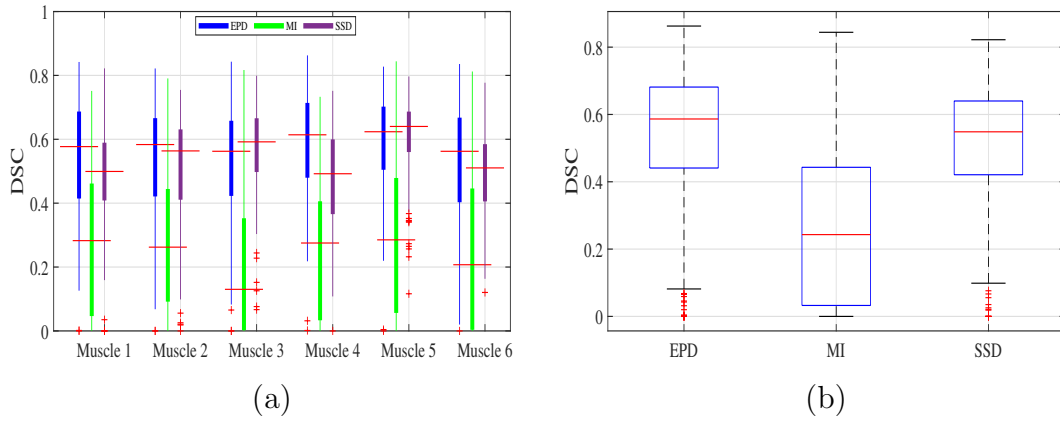


Figure 3.8: Comparison of performances of the edge proposition difference (EPD), mutual information (MI) and sum of squared difference (SSD) similarity measures from the affine registration experiments. (a) Registration accuracies of 110 cases for each muscle (higher volumetric dice similarity coefficient [DSC] values are better). (b) Registration accuracies of 110 cases \times 6 muscles = 660 DSCs for all muscles combined. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients in the neck dataset.

from an MRI volume are shown in Figure 3.7.

Table 3.2: Results obtained for the edge proposition difference (EPD), mutual information (MI) and sum of squared difference (SSD) measures from the affine registration experiments in terms of volumetric dice similarity coefficient (DSC) (DSC values for patient-1 [PT-1] fixed and others considered moving; left and right sternocleidomastoid, left and right semispinalis capitis, and left and right splenius capitis muscles denoted as Muscles 1, 2, 3, 4, 5 and 6, respectively, where higher values indicate better performances, with maximum value 1, and best DSC in each column marked in bold).

Muscles	Methods	Patient Index									
		PT-2	PT-3	PT-4	PT-5	PT-6	PT-7	PT-8	PT-9	PT-10	PT-11
Muscle-1	EPD	0.6065	0.7807	0.6460	0.6302	0.6052	0.7394	0.6736	0.4837	0.6719	0.4375
	MI	0.1164	0.2530	0.1558	0.4991	0.0000	0.1648	0.0374	0.2903	0.6577	0.2877
	SSD	0.4402	0.6611	0.4899	0.6667	0.4147	0.5345	0.5122	0.5044	0.5696	0.5597
Muscle-2	EPD	0.6677	0.6917	0.6869	0.6477	0.5771	0.6315	0.7034	0.4408	0.4446	0.5067
	MI	0.2782	0.3932	0.4868	0.4466	0.0620	0.5772	0.2519	0.1406	0.1889	0.3623
	SSD	0.5315	0.5892	0.6216	0.3935	0.0557	0.1850	0.1438	0.4108	0.2115	0.1518
Muscle-3	EPD	0.6566	0.7023	0.6408	0.6883	0.6687	0.5482	0.5142	0.4342	0.5483	0.7090
	MI	0.2801	0.3503	0.3942	0.4633	0.0246	0.5589	0.3559	0.4906	0.5248	0.3192
	SSD	0.6094	0.6444	0.5412	0.6545	0.5705	0.5786	0.5344	0.3541	0.5076	0.6350
Muscle-4	EPD	0.6530	0.8179	0.8127	0.7368	0.7407	0.5065	0.7731	0.7850	0.7163	0.6897
	MI	0.1734	0.0951	0.5216	0.1663	0.0367	0.0716	0.0227	0.3761	0.6859	0.0316
	SSD	0.6291	0.7201	0.7415	0.6476	0.4828	0.6871	0.3924	0.7843	0.5742	0.7063
Muscle-5	EPD	0.5957	0.6485	0.7734	0.6030	0.6659	0.5480	0.4160	0.5483	0.3643	0.6128
	MI	0.3895	0.1673	0.5672	0.1929	0.0201	0.7845	0.5572	0.2099	0.4460	0.0149
	SSD	0.6735	0.5573	0.5804	0.6017	0.5458	0.5497	0.5629	0.3798	0.4132	0.5171
Muscle-6	EPD	0.5715	0.6392	0.6899	0.7135	0.6785	0.6094	0.6098	0.6559	0.6488	0.5939
	MI	0.2454	0.0000	0.3181	0.0515	0.0028	0.0789	0.0000	0.1309	0.6678	0.0000
	SSD	0.6221	0.3390	0.5980	0.5914	0.5183	0.6797	0.3375	0.5769	0.5769	0.6169

3.4.2 Evaluation Metric

To evaluate registration results, the DSC is used due to its popularity in medical image research. The DSC is expressed as:

$$\text{DSC}(g, s) = \frac{2 |g \cap s|}{|g| + |s|} \quad (3.17)$$

where g is the annotation contour and s is the transformed contour, with the volumetric DSC for each muscle calculated separately.

Table 3.3: Results obtained for the edge position difference (EPD), mutual information (MI) and sum of squared difference (SSD) from the affine registration experiments in terms of volumetric dice similarity coefficient (DSC) (mean DSC values for all fixed volumes calculated over all moving ones; values in the first column were obtained from Table 3.2 by taking row-wise mean; bold values indicate better results).

Muscles	Methods	Patient Index											Overall mean
		PT-1	PT-2	PT-3	PT-4	PT-5	PT-6	PT-7	PT-8	PT-9	PT-10	PT-11	
Muscle 1	EPD	0.6275	0.5917	0.5074	0.5405	0.5049	0.5302	0.6147	0.5612	0.4828	0.5671	0.2790	0.5279 ±0.1784
	MI	0.2462	0.3039	0.3120	0.3962	0.2572	0.0872	0.3915	0.3190	0.2117	0.2101	0.3472	0.2802±0.2242
	SSD	0.5353	0.3908	0.5488	0.4130	0.5259	0.4682	0.5321	0.5834	0.4034	0.5340	0.3778	0.4830±0.1717
Muscle 2	EPD	0.5998	0.4415	0.5638	0.4677	0.4632	0.5953	0.4988	0.6455	0.5607	0.5179	0.2945	0.5135 ±0.2025
	MI	0.3188	0.2760	0.2375	0.3193	0.3174	0.1225	0.2912	0.3964	0.1417	0.1232	0.2607	0.2550±0.2100
	SSD	0.3294	0.6381	0.4362	0.5346	0.5474	0.5959	0.5204	0.5360	0.5208	0.4433	0.4228	0.5023±0.1860
Muscle 3	EPD	0.6111	0.6154	0.6963	0.6096	0.6010	0.6399	0.5831	0.6151	0.4905	0.5754	0.3638	0.5819 ±0.1652
	MI	0.3762	0.2637	0.3370	0.2879	0.2737	0.0422	0.2919	0.3689	0.3842	0.2312	0.3693	0.2933±0.2311
	SSD	0.5630	0.4443	0.6250	0.6457	0.5138	0.6146	0.6097	0.5740	0.3737	0.5873	0.6285	0.5618±0.1493
Muscle 4	EPD	0.7232	0.5359	0.6593	0.6420	0.5560	0.5701	0.5295	0.6434	0.7401	0.5666	0.3636	0.5936±0.1632
	MI	0.2181	0.3325	0.3179	0.2657	0.4075	0.2111	0.3753	0.4436	0.2464	0.0660	0.2820	0.2878±0.2175
	SSD	0.6365	0.5497	0.5745	0.6823	0.5506	0.6335	0.6241	0.5704	0.6172	0.5559	0.6150	0.6009 ±0.1377
Muscle 5	EPD	0.5776	0.5708	0.6270	0.5954	0.5716	0.5504	0.5745	0.4961	0.4823	0.4958	0.3816	0.5385 ±0.1771
	MI	0.3350	0.2070	0.3469	0.2882	0.2379	0.0460	0.2305	0.3654	0.2789	0.1664	0.2218	0.2476±0.2409
	SSD	0.5381	0.3616	0.5057	0.5267	0.5008	0.5777	0.5363	0.4796	0.3715	0.4865	0.5299	0.4922±0.1479
Muscle 6	EPD	0.6411	0.5297	0.5596	0.6137	0.6045	0.4046	0.5887	0.5006	0.5841	0.4673	0.3459	0.5309 ±0.1922
	MI	0.1495	0.2725	0.1831	0.2312	0.3378	0.0857	0.3578	0.2470	0.1900	0.0299	0.1308	0.2014±0.2185
	SSD	0.5457	0.4419	0.3016	0.5025	0.4660	0.4922	0.5285	0.4744	0.4920	0.4183	0.5387	0.4729±0.1548
Overall mean	EPD	0.6300	0.5475	0.6022	0.5781	0.5502	0.5484	0.5649	0.5770	0.5568	0.5317	0.3381	0.5477 ±0.1820
	MI	0.2740	0.2759	0.2891	0.2981	0.3052	0.0991	0.3230	0.3567	0.2421	0.1378	0.2686	0.2609±0.2252
	SSD	0.5247	0.4711	0.4987	0.5508	0.5174	0.5637	0.5585	0.5363	0.4631	0.5042	0.5188	0.5188±0.1648

3.4.3 Numerical Results Analysis

The experiments were performed on a HP z230 tower workstation with a 16 GB RAM and 3.40 GHz Intel(R) Core(TM) i7-4770 processor running the Windows 7 operating system using MATLAB, and C and C++ MEX programming. A canny edge detector with sigma 1.5 was used to calculate the edge image for the EPD similarity measure. The lower and higher threshold pairs of 0.1, 0.9; 0.08, 0.7; 0.04, 0.4; 0.01, 0.2 and 0.001, 0.1 were selected in the 5th, 4th, 3rd, 2nd and 1st resolution levels of the FFD registration respectively. The threshold pair 0.1, 0.9 was also used with affine and DPSW registrations.

3.4.3.1 Comparison of the EPD, MI and SSD Similarity Measures

Affine EPD, affine MI and affine SSD registrations were performed on the dataset to compare the EPD, MI and SSD similarity measures. The affine MI registrations were performed using advanced normalization tools (ANTs). For each fixed patient, the other 10 patients were considered moving, giving a total of $11 \times 10 = 110$ cases (registrations). Table 3.2 shows the experimental results obtained for affine registrations, which are the DSC values for 10 of the 110 cases for the six muscles in which PT-1's volume was considered fixed and the others moving. Similar tables were constructed for the other patients. The bold values, which represent the highest DSC in each column, indicate that the EPD has the highest DSC in eight cases and the MI and SSD each in only one. It is also clear that the PT-2 case has the lowest best value, with best DSCs for the EPD, MI and SSD of 0.8179, 0.7845, and 0.6735 respectively. Overall, the EPD achieved better accuracy than the MI and SSD.

Table 3.3 shows the mean and overall mean DSCs for all the fixed volumes of the affine registration. The last column represents the final means of the 110 cases and the last three rows display the means of 10 cases when considering the other 10 patients as moving. The table indicates that the EPD has better registration accuracy than the MI and SSD for the nine patients, and the SSD has better accuracy than EPD and MI for the two patients, EPD achieving a final DSC of 0.5477, which is significantly better than the MI's 0.2609 and the SSD's 0.5188. This large difference in DSC values proves that the MI similarity measure is not suitable for this dataset and, therefore, for the proposed coarse-to-fine framework. Although the accuracy of the SSD is close to that of EPD, it can not provide object-wise alignment. Therefore, EPD has been selected as the similarity measure for the proposed method. Although the volumetric DSCs for

Table 3.4: Parameter settings for SyN [3] and Demons [2] registrations in our neck dataset. The SyN and Demons registrations were conducted through ANTs and MATLAB, respectively. Multistage registrations were used in the SyN method, in which the rigid, affine and SyN geometrical transformations used the same smoothing sigmas, shrink factors, and convergence, except that SyN used different convergence. The metric CC means cross-correlation.

Parameter	SyN [3]	Demons [2]
Script	antsRegistrationSyN.sh	imregdemons.m
Dimension	3	3
Pyramid levels	-	3
Iterations	-	100x100x100
Transformations	Rigid[0.1], Affine[0.1]	-
Metric	MI[F,I,1,32,Regular,0.25]	-
Smoothing sigmas	3x2x1x0vox	1
Shrink factors	8x4x2x1	-
Convergence	[1000x500x250x100,1e-6,10]	-
Transformation	SyN[0.1,3,0]	-
Metric	CC[F,I,1,4]	-
Convergence	[100x70x50x20,1e-6,10]	-

EPD are better than for the other methods, its values are not in the excellent range because strong edges were used to align the volumes coarsely. The coarse alignment is further refined through the latter multiple stages in the proposed method. These results are obtained using only the single stage "Affine-EPD" from the proposed framework. The EPD yields mediocre results in some circumstances such as PT-11. There could be two possible reasons. First, the patient is demographically and anatomically more different than the other patients. Second, there may be some noise or intensity inhomogeneity problems. These problems can be eliminated by using the proposed complete framework.

To investigate further, box plots are used to statistically analyze the experiments. Figure 3.8 exhibits the registration accuracies obtained by the EPD, MI

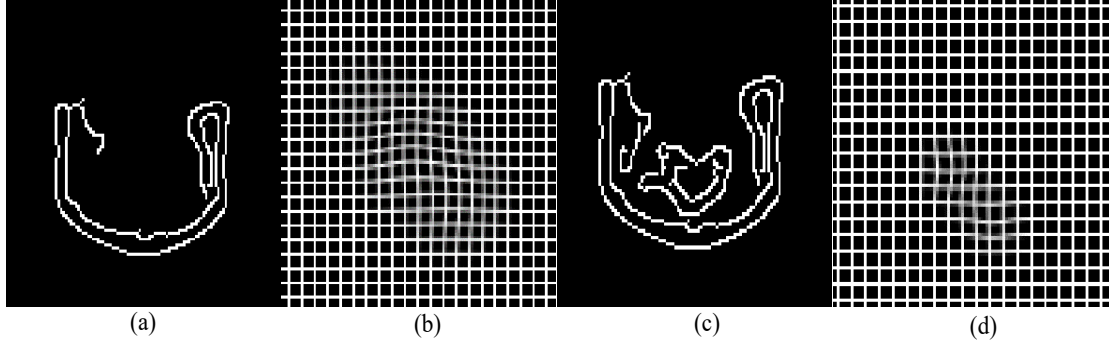


Figure 3.9: Axial deformation fields and edge maps of an inter-subject case for different sets of edges using different sets of thresholds with different resolution levels of the spline: (a) edge map with lower and higher thresholds pair of 0.08 and 0.7, (b) deformation field of (a) using 4th resolution level of spline, (c) edge map with thresholds pair of 0.04 and 0.4, (d) deformation field of (c) using 3rd resolution level of spline.

and SSD for each muscle separately (see Figure 3.8(a)) and combined (see Figure 3.8(b)). Figure 3.8(a) indicates that the EPD has better median DSCs than the MI and SSD for four muscles, and the SSD is better for the other two muscles. In addition, the maximum DSCs for EPD are higher than for the MI and SSD measures for all muscles except for Muscle-5 when using MI.

Similarly, Figure 3.8(b) reveals that the EPD has better overall median and maximum DSCs for all the muscles combined (0.5866 and 0.8627, respectively) than the MI (0.2430 and 0.8441, respectively) and the SSD (0.5484 and 0.8220, respectively).

3.4.3.2 Performance Comparison

Strong and weak sets of edges were used to obtain coarse and fine level deformation. Figure 3.9 shows edge maps and the corresponding deformation fields for an inter-subject registration case using strong and weak sets of edges. These maps illustrate a coarse displacement field for the strong set and a fine displacement field for the weak set. To evaluate the effectiveness of every stage of the proposed method,

the volumetric DSC was calculated for one of the inter-subject registrations at every stage. Figure 3.10 shows the 2D visual results at different stages of the proposed registration method with a corresponding volumetric DSC value. It can be seen from Figure 3.10(c) that the required deformations are very large across subjects. Every stage gradually aligns the two images as indicated by the alignment of the vertebra, muscles and neck trunk with the superimposed edges. The effectiveness of some stages is more clearly visible in some other inter-subject registration cases. The rising trend of volumetric DSC proves the effectiveness of every stage in the proposed framework. Figure 3.11 shows the corresponding coronal views for the patient shown in Figure 3.10. The sagittal views have not been shown due to anterior information loss caused by inhomogeneity problems with the MRI scanner. However, this view exhibits similar alignment improvement with every stage as in the axial views. Figure 3.12 shows the results in terms of muscles contours of different methods compared with the ground truths. It can be seen that the proposed method's contours are more fairly matched than for the SyN and Demons algorithms.

The full proposed method was compared for the 110 inter-subject cases with Coarse-EPD and Fine-EPD using volumetric DSC. Other stages were excluded for presentation convenience. The ISR was also performed using the diffeomorphic Demons [2, 148] and SyN [3] methods on this dataset for the 110 cases. The two methods are powerful and popular registration algorithms considered as a gold standard in the deformable registration field. ANTs and MATLAB were used for implementing the SyN and diffeomorphic Demons, respectively. The parameter information for the SyN and diffeomorphic Demons is shown in Table 3.4. Table 3.5 shows the mean and overall mean volumetric DSCs for all the fixed volumes of the proposed, diffeomorphic Demons [2], SyN [3], Coarse-EPD and Fine-EPD registrations. Tables 3.3 and 3.5 were generated in a similar way. The overall mean

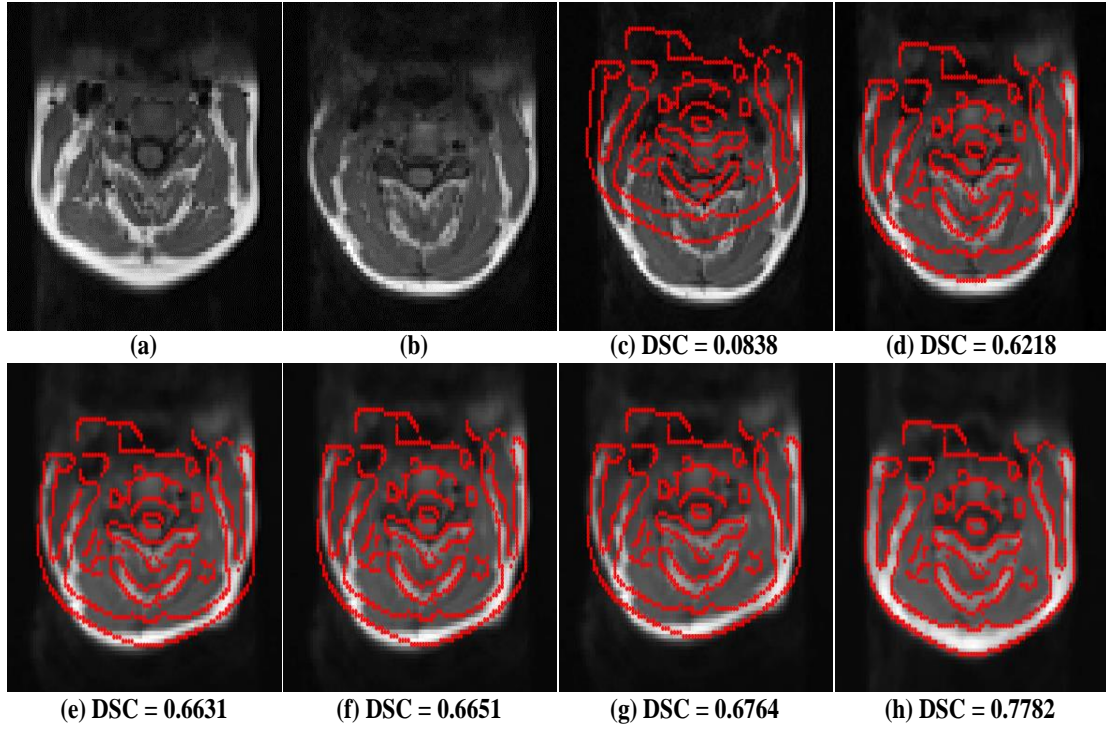


Figure 3.10: Axial visual results of one of the 110 inter-subject registration cases to show the effectiveness of every stage of the proposed framework, as shown in Figure 3.1, in terms of the volumetric DSC value. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with superimposed fixed image's edges; superimposed fixed image's edges on the moving image after (d) Affine-EPD stage, (e) DPSW-EPD stage, (f) Coarse-EPD stage, (g) Fine-EPD stage and (h) final stage (proposed method).

values in the last column represent the final means for the 110 cases and those in the last five rows represent the means for the 10 cases considering the other 10 patients as moving. The proposed method clearly yields better registration accuracy than the others for nine of the 11 participants, whereas Demons [2] and SyN [3] yield good accuracies for only one each of the 11 participants. The proposed method achieves a final DSC of 0.7029, significantly better than those of the Demons (0.6654) and SyN (0.6606), with the overall mean volumetric DSC values obtained for five muscles higher by the proposed method than others, except for Muscle 2 by SyN. However, the rising trend of final DSC values in the Coarse-EPD and Fine-EPD stages indicate that the stages before the Demons stage in the

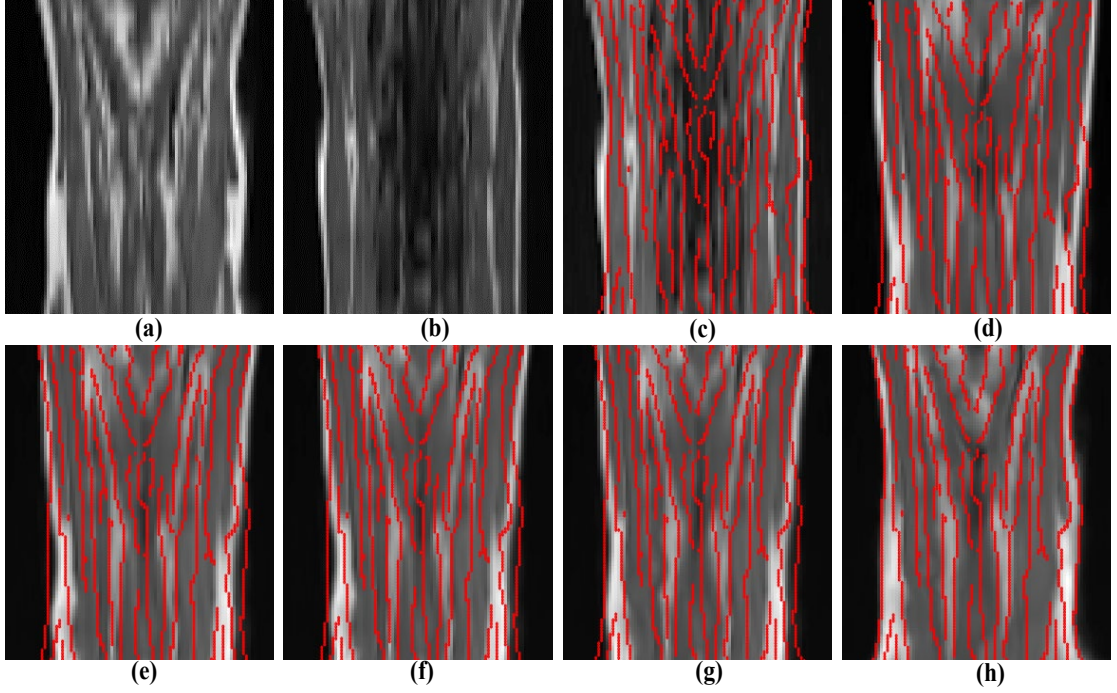


Figure 3.11: Coronal visual results for the same case shown in Figure 3.10. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with superimposed fixed image's edges; superimposed fixed image's edges on the moving image after (d) Affine-EPD stage, (e) DPSW-EPD stage, (f) Coarse-EPD stage, (g) Fine-EPD stage and (h) final stage (proposed method).

proposed method bring the moving volumes closer to the fixed volume which helps Demons to align more effectively than when only the Demons method is used.

Figure 3.13 shows the results obtained from the statistical analysis of the proposed, diffeomorphic Demons, SyN, Coarse-EPD and Fine-EPD experiments for each muscle separately (see Figure 3.13(a)) and combined (see Figure 3.13(b)). Figure 3.13(a) indicates that the proposed method has better median and maximum DSCs than the others for Muscles 1, 4, 5 and 6, whereas the SyN algorithm performs better for Muscles 2 and 3.

Similarly, Figure 3.13(b) reveals that the proposed method has better overall median and maximum volumetric DSCs for all the muscles combined (0.7385

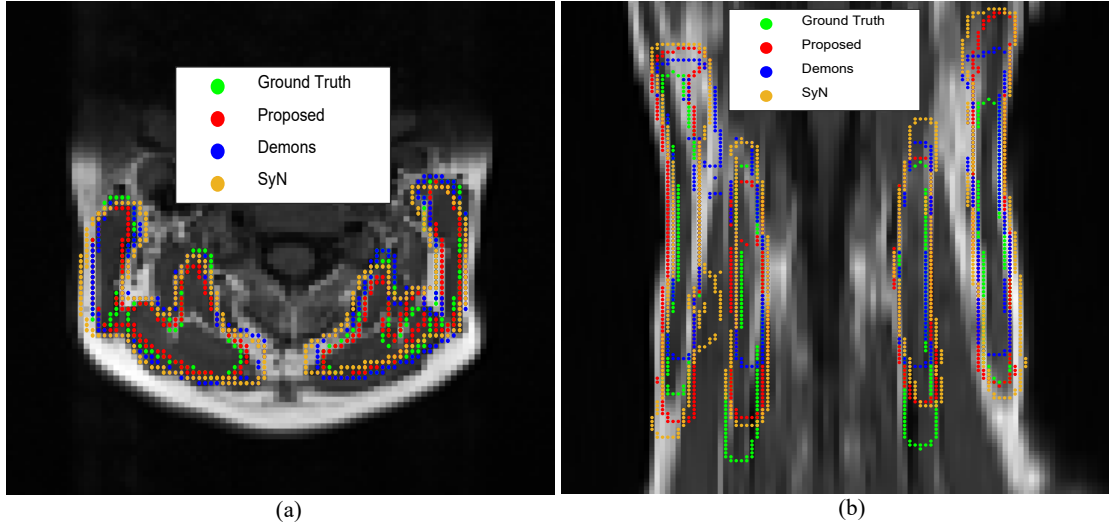


Figure 3.12: Comparison of visual results in terms of muscles contours among proposed, Demons [2] and SyN [3] methods with ground truths: (a) axial view and (b) coronal view.

and 0.9075, respectively) than the diffeomorphic Demons (0.7215 and 0.8680, respectively) and SyN (0.7137 and 0.8940, respectively) methods. Moreover, the diffeomorphic Demons method has more outliers (53) than the proposed method (20) and SyN (32) for all the muscles combined.

To perform a complete analysis, the Hausdorff distance (HD) was also used as a distance error metric. Figure 3.14 shows the HD results for all muscles combined. It reveals that the proposed method has better median HD ($5.7446mm$) than the SyN ($6.0000mm$) and diffeomorphic Demons ($5.9161mm$) methods.

Finally, the experimental results and analyses reveal that the proposed method outperforms the diffeomorphic Demons and SyN algorithms in terms of registration accuracy and consistency.

Table 3.5: Registration results obtained from inter-patient neck magnetic resonance imaging experiments in terms of volumetric dice similarity coefficient (DSC) (proposed method compared with diffeomorphic demons [2], SyN [3], Coarse-EPD, and Fine-EPD; mean DSC values for all fixed volumes calculated over all moving ones; bold values indicate better results).

Muscles	Methods	Patient Index											Overall mean
		PT-1	PT-2	PT-3	PT-4	PT-5	PT-6	PT-7	PT-8	PT-9	PT-10	PT-11	
Muscle 1	Coarse-EPD	0.7021	0.6946	0.6721	0.6339	0.6530	0.6332	0.7226	0.7519	0.3440	0.3745	0.5119	0.6413±0.1841
	Fine-EPD	0.7166	0.7111	0.6940	0.6568	0.6784	0.6646	0.7452	0.7677	0.3603	0.7099	0.5321	0.6614±0.1843
	Proposed	0.7338	0.7213	0.7379	0.7043	0.6942	0.7207	0.7272	0.7613	0.7362	0.7321	0.7668	0.6943 ±0.1814
	SyN [3]	0.7447	0.7661	0.6745	0.6193	0.6424	0.7641	0.7646	0.6981	0.5699	0.6230	0.6501	0.6833±0.1609
	Demons [2]	0.6582	0.7409	0.7140	0.6887	0.4601	0.6704	0.7761	0.6707	0.6900	0.3251	0.7031	0.6452±0.2226
Muscle 2	Coarse-EPD	0.6913	0.5885	0.5868	0.6668	0.6140	0.6500	0.5890	0.7058	0.6639	0.6608	0.3934	0.6227±0.1515
	Fine-EPD	0.7137	0.6054	0.5968	0.6909	0.6539	0.6536	0.6024	0.7201	0.6787	0.6792	0.4094	0.6381±0.1532
	Proposed	0.7585	0.6848	0.6515	0.7429	0.6617	0.6608	0.6781	0.6363	0.6751	0.6850	0.7202	0.6818±0.1409
	SyN [3]	0.7410	0.7842	0.6612	0.7453	0.6627	0.7311	0.7376	0.7263	0.6100	0.6650	0.5546	0.6926 ±0.1650
	Demons [2]	0.7180	0.6932	0.6879	0.6928	0.5462	0.6815	0.7014	0.7064	0.6667	0.4987	0.6100	0.6548±0.1757
Muscle 3	Coarse-EPD	0.7359	0.6286	0.6721	0.7635	0.5946	0.7126	0.7483	0.7574	0.6025	0.6854	0.5534	0.6812±0.1310
	Fine-EPD	0.7560	0.6381	0.6922	0.7753	0.6264	0.7192	0.7647	0.7750	0.6250	0.6985	0.5624	0.6973±0.1311
	Proposed	0.7853	0.6802	0.7618	0.7920	0.6429	0.7347	0.7840	0.8113	0.7895	0.7043	0.7544	0.7422 ±0.1133
	SyN [3]	0.7413	0.6422	0.7389	0.7089	0.6652	0.7942	0.7872	0.7653	0.6932	0.5610	0.5668	0.6967±0.1546
	Demons [2]	0.7539	0.6345	0.7580	0.7113	0.4718	0.7062	0.7827	0.7718	0.7601	0.6839	0.7605	0.7086±0.1396
Muscle 4	Coarse-EPD	0.7345	0.7224	0.6570	0.6846	0.6365	0.7276	0.6519	0.7017	0.6500	0.7644	0.5770	0.6885±0.1053
	Fine-EPD	0.7526	0.7370	0.6795	0.6973	0.7021	0.7438	0.6627	0.7222	0.6765	0.7582	0.5908	0.7058±0.1058
	Proposed	0.7888	0.7461	0.7308	0.7394	0.7197	0.7504	0.7777	0.6832	0.7586	0.7812	0.8002	0.7427 ±0.1087
	SyN [3]	0.7536	0.6927	0.6874	0.7411	0.6790	0.7845	0.7594	0.7354	0.7239	0.5537	0.6435	0.7049±0.1560
	Demons [2]	0.7645	0.7307	0.7467	0.4160	0.7356	0.7820	0.7222	0.7704	0.7418	0.7630	0.6933	0.7100±0.1534
Muscle 5	Coarse-EPD	0.6922	0.4634	0.6602	0.6849	0.5576	0.6530	0.7000	0.6528	0.5327	0.5980	0.5197	0.6122±0.1347
	Fine-EPD	0.7096	0.4815	0.6831	0.7102	0.5771	0.6729	0.7204	0.6731	0.5560	0.6217	0.5348	0.6330±0.1369
	Proposed	0.7175	0.5519	0.7107	0.7593	0.5811	0.6960	0.7460	0.7746	0.7088	0.6220	0.7009	0.6842 ±0.1291
	SyN [3]	0.6476	0.5036	0.7029	0.6234	0.6191	0.7102	0.7233	0.6198	0.5718	0.4791	0.4823	0.6076±0.1737
	Demons [2]	0.6507	0.5561	0.6640	0.6804	0.4492	0.6511	0.7342	0.7221	0.6643	0.5914	0.7251	0.6444±0.1730
Muscle 6	Coarse-EPD	0.6490	0.6509	0.5746	0.6034	0.5617	0.6285	0.5754	0.6036	0.4906	0.5622	0.5059	0.5824±0.1328
	Fine-EPD	0.6620	0.6741	0.5951	0.6278	0.5892	0.6551	0.5952	0.6136	0.5154	0.5904	0.5210	0.6035±0.1310
	Proposed	0.6613	0.7443	0.6576	0.7022	0.6716	0.6640	0.7089	0.6531	0.6392	0.6107	0.6808	0.6721 ±0.1233
	SyN [3]	0.5907	0.5620	0.4523	0.6189	0.6120	0.6454	0.7128	0.5745	0.5277	0.5163	0.5479	0.5782±0.1716
	Demons [2]	0.6072	0.6093	0.6327	0.6457	0.4061	0.6718	0.7202	0.6895	0.6271	0.6080	0.7018	0.6290±0.1740
Overall mean	Coarse-EPD	0.7008	0.6247	0.6371	0.6728	0.6251	0.6719	0.6645	0.6955	0.5473	0.6248	0.5102	0.6380±0.1463
	Fine-EPD	0.7184	0.6412	0.6568	0.6931	0.6300	0.6924	0.6818	0.7120	0.5686	0.6635	0.5251	0.6565±0.1464
	Proposed	0.7409	0.6881	0.7084	0.7400	0.6475	0.6925	0.7370	0.7200	0.7179	0.6848	0.7372	0.7029 ±0.1375
	SyN [3]	0.7031	0.6585	0.6529	0.6761	0.6467	0.7383	0.7475	0.6866	0.6161	0.5663	0.5742	0.6606±0.1704
	Demons [2]	0.6921	0.6453	0.6979	0.6942	0.4582	0.6861	0.7494	0.7138	0.6964	0.5748	0.7106	0.6654±0.1772

3.5 Discussion

A deformable 3D-3D fully automatic registration framework using a novel DPSW transformation and modified 3D Chamfer distance transform of the EPD was developed. The empirical outcomes demonstrated that the proposed method outperformed the well-established diffeomorphic Demons [2, 148] and SyN [3] algorithms in terms of accuracy and consistency. Also, interestingly, the proposed framework

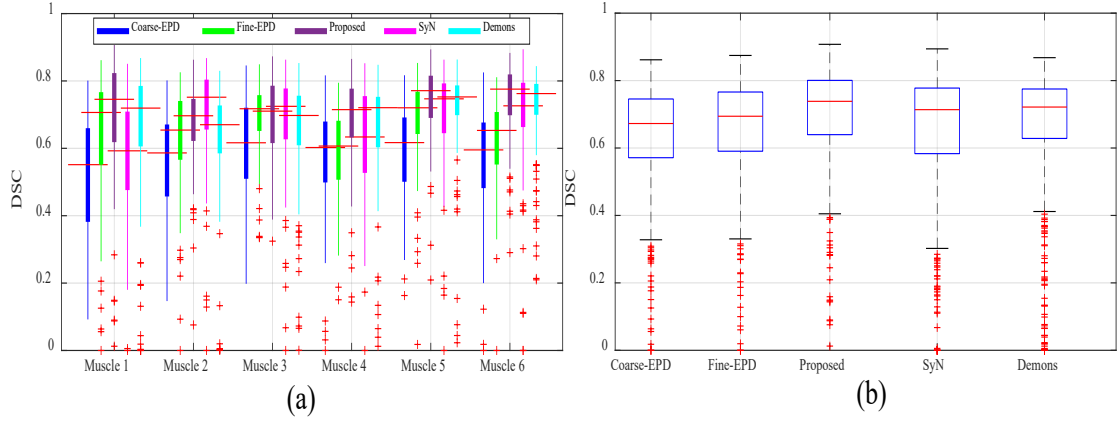


Figure 3.13: Comparison of performances of proposed, diffeomorphic Demons [2], SyN [3], Coarse-EPD and Fine-EPD methods. (a) Registration accuracies of 110 cases for each muscle (higher volumetric dice similarity coefficient [DSC] values are better). (b) Registration accuracies of 110 cases \times 6 muscles = 660 DSCs for all muscles combined. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients in our neck dataset.

was robust in terms of input volumes because it worked on multiple MRI scanner settings. In particular, the same thresholds were used in the EPD SM for the input volumes with different repetition and echo times, which is common when acquiring MR images in clinical practice.

The computational time required to register two neck volumes was calculated for the EPD, SSD, MI, proposed, Demons [2] and SyN [3] methods. These times are shown in Table 3.6. Although, some parts of the proposed method were implemented using C/C++ MEX coding, the computational time of the proposed method could be reduced if the full code was implemented in C/C++. The proposed method is slower than the Demons algorithm but has better accuracy. The low computational time for Demons was due to the professional implementation of this algorithm in MATLAB.

The experiments in this work were conducted on real clinical neck MRI data. The proposed registration framework was thoroughly evaluated against well-recognized

deformable registration methods [2, 3, 148] applied to this dataset, and obtained a better overall mean DSC value, confirming its effectiveness. The reasons for the consistently enhanced accuracy of the proposed method are as follows:

- A conventional multi resolution registration method attempts to avoid large-scale basins in the function to be optimized, using only the coarse resolution levels of the spline. This approach has limited success because of the monotonous SM and optimization. In this work, the framework was constructed based on the novel notion that correct deformations cannot be obtained through a single transformation due to the constraints placed on the parameters of the geometrical transformation. This approach brings the floating volume nearer to the reference one by applying multiple transformations, optimizations and SMs. However, if an exact deformation is near a transformation, it cannot be obtained if the same stage is used repeatedly because it becomes trapped in a local optimum of the SM, to escape from which a different SM and optimization approach is required. In the past, FFD-based registrations attempted to escape large-scale basins by exploiting multi resolution versions of the B-spline with the same SM and optimization. However, they did not succeed due to the similar characteristics of the different B-spline resolutions. In this work, completely different combinations were used in each stage of the coarse section to align the neck's trunk first, which avoids large-scale basins in the optimization function and provides good alignment.
- A registration process can frequently become stuck due to dips in the optimization function, rather than basins [37]. Although all the multi resolution

techniques using the B-spline have tried to avoid these, they could not mitigate all their effects because they considered only variations of the spline resolution and the nature of their cost functions was non-convex [34, 120, 121]. For this framework, a double-pushing system was designed in the fine section to prevent dips yielding small deformations and to reduce the chance of becoming stuck in dips. The first process in this system involved changing the edges of the EPD from the strongest to weakest, while the second involved altering the resolution of the spline. Therefore, this framework obtained better accuracy by avoiding more dips.

- The negative lobe of the spline wavelet provides an implicit regularization in the transformation and, therefore, a smooth deformation. The DPSW has fewer parameters than the FFD, which facilitates the optimization process to avoid large-scale basins and improves matching.
- The EPD provides an opportunity to achieve an alignment through multiple stages for the same objects or edges. Therefore, tuning an alignment can be performed object-wise, a capability not available using any other SM.

Although a limitation of this work is the lack of explicit regularization, which normally introduces a parameter that imposes a burden on the user, in this framework, the DPSW was preferred, which provides an implicit regularization and obtains smooth transformations through the fourth-order spline.

3.6 Conclusion

In this research, a registration framework was proposed with different combinations of transformations, optimizations and SMs with a new transformation DPSW

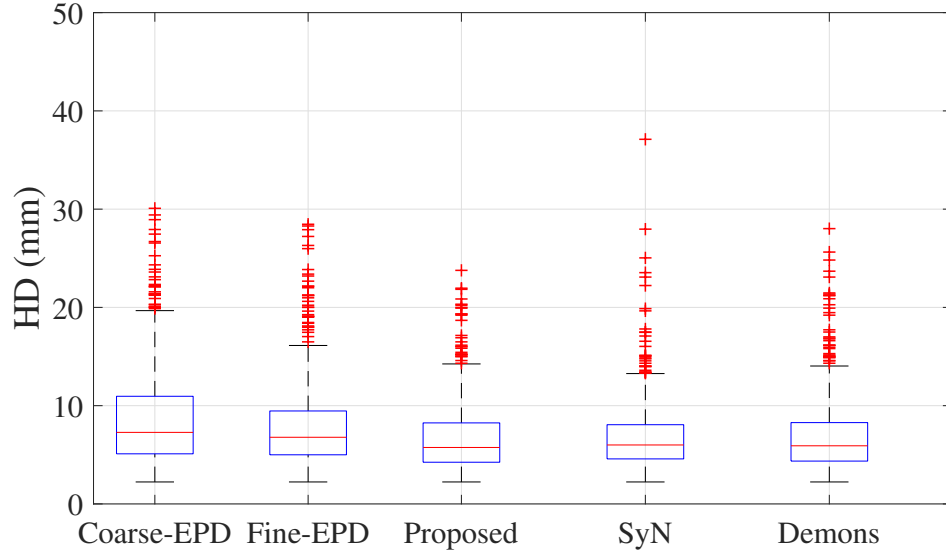


Figure 3.14: Comparison of performances of proposed, diffeomorphic Demons [2], SyN [3], Coarse-EPD and Fine-EPD methods on our neck dataset in hausdorff distance (HD) in mm (lower values are better) for all muscles combined for 110 cases \times 6 muscles = 660 HDs. The 110 cases are assessed by considering all other patients' images as moving and a specific patient's image as fixed for the total 11 patients.

Table 3.6: Computational time to register our dataset using the EPD, SSD, MI, proposed, Demons [2] and SyN [3] algorithms on our computer for a single case. The affine transformation is used for the EPD, SSD and MI registration. The MI and SyN [3] are implemented through advanced normalization tools (ANTs).

Methods	Time (minutes)
EPD	3.21
SSD	24
MI	2.21
Proposed	33
Demons [2]	0.81
SyN [3]	69.95

and a modified 3D Chamfer distance transform algorithm. The DPSW requires fewer parameters than the FFD to represent similar local deformations, which decreases the burden on the optimization process. The framework enhances the probability of finding the global minimum by reducing the effects of basins and

dips, with outcomes obtained from experiments on real clinical datasets confirming its effectiveness.

Chapter 4

Two-stage Heuristic Search and Region of Interest-based Edge Position Difference Similarity Measure for Neck MRI Registration

Although traditional heuristic optimization-based techniques achieve good results when handling large mismatches, they require pre-segmentation on both the fixed and moving images and are computationally expensive. On the other hand, current deformable methods often face statistical instability problems and many local optima when dealing with small mismatches. Also, they ignore anatomical information (AI) which is important for detecting correspondences. This chapter investigates the inter-subject deformable registration problem in a neck MRI application and provides two novel contributions, the practical merits of which are demonstrated in the experimental results. Firstly, it propose a two-stage heuristic search optimization technique for handling large mismatches, using a minimal user hypothesis about the large mismatches, which is computationally fast. The optimization brings a moving image hierarchically closer to a fixed one using mutual information (MI) and edge position difference (EPD) similarity measures in coarse

and fine stages respectively. Of particular note is that the images do not require pre-segmentation. Secondly, a region of interest (ROI) EPD-based registration framework for handling small mismatches using salient AI in which a convex objective function is formed through a unique shape created from the desired objects in the ROI is proposed. The proposed method is compared to two state-of-the-art methods on a neck MRI dataset, with the results showing that the proposed method is superior in terms of accuracy.

The rest of this chapter is organized as follows: an introduction is presented in Section 4.1; details of our registration method are presented in Section 4.2; in Section 4.3, the experimental procedure and analyses of the results are provided; and, finally, a discussion and conclusion are presented in Section 4.4.

4.1 Introduction

In order to tackle the two major challenges of handling large and small mismatches, many deformable registration methods have been proposed in the literature. A recent review of them can be found in [48] but, as few published works consider neck medical images, this chapter deals with these mismatches using different novel techniques. In the following sections, the challenges, current solutions and their limitations, and proposed approaches for overcoming them are discussed.

4.1.1 Handling Large Mismatches

In the literature, although many techniques for handling large mismatches between subjects have been proposed [45–47], it has been shown that conventional continuous and discrete optimization-based registrations cannot achieve this [48].

Although heuristic-based methods can provide a wide range of solutions, they require possible solutions to be intuitive. They were used in [49–52] to register brain medical images, whereby the authors approximated a sequence of the energy function by selecting the driving voxels hierarchically. Initially, they used a few driving voxels, typically located on landmarks, through prior segmentation of both the fixed and moving images. These voxels have distinctive attribute vectors and more are added as the algorithm progresses. One heuristic search algorithm is known as the hierarchical attribute-matching mechanism for elastic registration (HAMMER). Although its accuracy is high for registering inter-subject brain MRIs, as its main limitation is that it requires pre-segmentation of both fixed and moving images, it compromises automatic registration which is not feasible in all applications. Also, it is computationally expensive because it calculates an attribute vector for each driving voxel in both images which is gradually increased to the total number of voxels in a volume. Each attribute vector comprises edge types, a geometric moment and image intensity. In extensions of the HAMMER algorithm [53–58], local spatial intensity histograms have been used to calculate a different type of attribute vector for each point in the fixed and moving images, which makes the algorithm suitable for general applications. However, local histograms face the problem of statistical power instability.

In order to deal with the above limitations in the neck MRI dataset, a new, totally different, two-stage heuristic optimization technique is proposed that uses the MI and EPD similarity measures as well as a translation transformation. It automatically brings a moving image closer to a fixed one in a rapid and efficient way. Initially, MI is applied on the coarse scale as it is not good for fine-scale matching, especially of compact or similar compositional objects in medical images, due to its low sensitivity to small translational changes between two images. Then, the EPD is used for fine-scale alignment as it considers the salient edges of

objects which are very sensitive to small changes. The two-stage heuristic search does not require the pre-segmentation of any image in the registration process and only minimal user interaction in terms of approximating large mismatches in a specific application, such as of the neck, brain, cardiac region, etc., which are very easy to determine. The proposed heuristic optimization is fast for two reasons: firstly, it does not require calculations of the attribute vectors which is very time-consuming; and, secondly, it requires fewer iterations due to using a coarse-to-fine step size. Only the in-plane translations are optimized for large mismatches of these parameters, unlike the HAMMER algorithm which optimizes all the parameters of an affine transformation.

4.1.2 Handling Small Complex Mismatches

To confront small complex mismatches, many registration methods have been proposed. The objective function is a constituent element in an image registration process, with a suitable one being a crucial part of image registration for achieving accurate deformations [48, 331]. Usually, the objective function is defined in terms of the point-to-point or region-to-region correspondences between two images, an assumption which is reliable in the case of a linear intensity relationship between the images. However, this relationship is non-linear for magnetic resonance (MR) and multi-modal images as the former experience space-variant intensity distortions. Several methods proposed to tackle the non-linearity problem have incorporated local orientation features instead of intensities [39, 331, 332], most of which can be classified in the following two major groups.

4.1.2.1 Spatially Region-weighted Scheme

The registration methods in this group place more importance on some regions than others by using a weighting scheme, with the objective function being a weighted sum of the local similarity measures. Although MI and its variants are widely used in image registration, global MI is less sensitive than its local counterpart to local deformations and often provides mis-correspondences. This problem is more acute in the case of neck muscles in MRIs because similar muscles are compactly grouped together. To avoid this, researchers have attempted to calculate local MI and use it with weights to compute the final objective function by summation [39–42]. Studholme *et al.* [41] used regional labels with an intensity joint histogram as an extra channel to calculate regional MI (RMI). This method divided the image into overlapping squares and is robust to variations in local intensity. Another similarity measure called localized MI (LMI) [42] used random partitioning around the stochastic points. Both the RMI and LMI used the same weight inside a sub-region. In contrast, the conditional MI (CMI) [333] used a spatial distribution function to assign weights to the voxels with respect to their distances from the center of the relevant sub-region. Its experimental results showed that it performed better than the RMI and LMI similarity measures but all three are computationally expensive as they used sums of the local measures. To tackle computational complexity, a spatially encoded MI (SEMI) approach [39] used a hierarchical weighting scheme by exploiting a Gaussian function and dividing an image into spheres as well as local gradient ascent optimization to reduce the computational time. However, global convergence was deteriorated in this approach. Rivaz *et al.* [40] proposed a self-similarity α -MI (SeSaMI) metric which exploited local structural information via weighted gradient information and was implemented through a weighted graph. However, although it is invariant to local

affine intensity distortions and rotations, it is computationally expensive due to its multi-dimensional features.

All the local MI-based similarity measures, such as RMI, LMI, CMI, SEMI and SeSaMI, face the problem of statistical instability and are computationally expensive. Also, they often ignore anatomical information (AI) which is crucial for guiding the detection and registration of correspondences. Albeit their registration results are better than those for global MI, finding accurate correspondences remains difficult due to the many local optima in deformable registrations. Moreover, some other metrics also include local structural information such as local entropy [332] and spatially region-weighted correlation ratio (SRWCR) [38]. The method in [332] used entropy images to avoid the effect of intensity variations at corresponding locations in different images. It exploits the sum of squared differences (SSD) of the entropy images as the objective function and is sensitive to noise due to the histograms being derived from small patches. Gong *et al.* [38] proposed a SRWCR metric using a three-dimensional (3D) joint probability density function which incorporated a spatial information channel using a cubic B-spline spatial distribution function with intensity channels.

4.1.2.2 Encoding Objects' Local Positions and Shapes

The registration methods in this group embed local AI in the objective function. A similarity measure called the modality-independent neighborhood descriptor (MIND) [326], which includes local structural information and information of neighborhood pixels in the vector representation of each pixel, obtains good registration results. However, although it provides distinctiveness, it embeds undesired information and cannot fully suppress the influence of contrast enhancement. Also, it is incapable of providing distinctiveness in the case of similar patches located

near each other, such as neck muscles and cardiac ventricles. Another descriptor called the self-similarity context (SSC) [334] was proposed to eliminate the problems of the MIND measure. It is robust to noise as it considers six-neighborhood patches for self-similarities instead of only the center one, as the MIND does, but also responds weakly to variations in contrast. Li *et al.* in [331] embedded local phase features in the MIND measure which rely on the auto-correlation of local structures. Cao *et al.* in [27] extracted AI from two modalities to guide the detection of correspondences through normalized cross-correlation (NCC) in a multi-modal registration problem. They used computed tomography (CT) and MRI for correspondences of bone and soft tissue regions respectively.

All the methods, described in the two groups above, treat local regions with special consideration and incorporate local structural information in the objective function to attempt to make it convex by introducing distinctiveness.

In the proposed method, edge features are also used as local structural information but local regions are treated in a totally different way from partitioning a whole image into regions and assigning a different weight to each region which is similar to the spatially region-weighted scheme. A unique shape is formed from only the targeted objects to introduce distinctiveness into the objective function, to make it convex, as the aim is to segment specific muscles in an atlas-based application from MRIs with intensity distortions. This technique is called a region of interest (ROI)-based registration that focuses on only the targeted objects because, in most medical image segmentation problems, the application is related to the segmented objects. Also, this ROI-based image registration approach can estimate transformations more efficiently and accurately than a whole-image approach [27] because the ambiguity of a similarity measure is less for a ROI than for a global image. The EPD is used as a similarity measure in this process because it can use

salient AI which is crucial for image matching and is sensitive to small geometric transformations. Although a global EPD may face a mis-correspondence problem, the ROI-based approach does not because of the small amount of geometrical transformation of the unique shape. The main contributions of this chapter are summarized below.

- A new heuristic search-based optimization technique in which an approximation of large mismatches is required for a new application. The search is fully automatic in the same application as it is not required to be conducted again for any new subjects. In this search, an image is warped within a range with a specific step size to find the best correspondence. It avoids many local minima and significantly increases accuracy. It is computationally fast and does not require any pre-segmentation as do traditional heuristic methods.
- This heuristic search concept is extended to a novel coarse-to-fine approach using two stages to align images more effectively, whereby the first stage uses MI for coarse alignment and the second, the EPD, to correct global fine mismatches.
- A novel ROI-based registration is proposed for aligning local fine mismatches. As it helps to avoid many local minima by concentrating on specific regions, it improves registration accuracy. It imposes distinctiveness in the objective function by forming a unique shape to help the optimization process. Its processing is computationally fast and it estimates transformations more efficiently and accurately than global registration.
- A coarse-to-fine approach-based framework using the heuristic search and ROI-based approaches, as well as other registration stages. As these stages are also computationally fast, the overall computational time of the framework is low with improved accuracy.

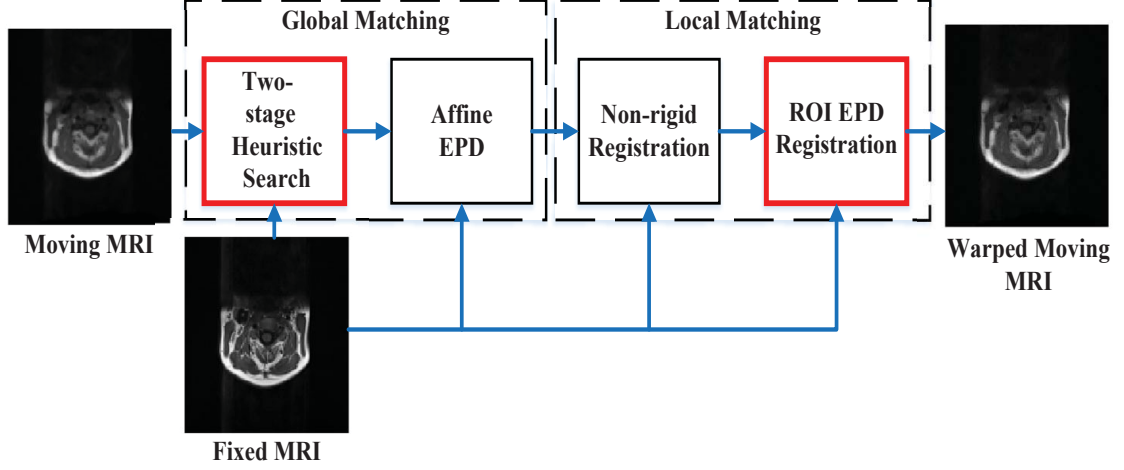


Figure 4.1: Proposed inter-subject framework of MRI-to-MRI registration based on two-stage heuristic search and ROI EPD. These two main components indicated by red boxes, with MRI slices used for only presentation purposes as they are MRI volumes.

- This proposed framework and two state-of-the-art methods are applied for the challenging neck muscle segmentation task and their performances are compared. The evaluation results show that the proposed approach achieves a substantially better performance than its state-of-the-art competitors.

4.2 Proposed Registration Framework

4.2.1 Overview

The proposed registration framework consists of two main components, a two-stage heuristic search and ROI EPD registration, as shown in Fig. 4.1. Let $F(\mathbf{x}) : \Omega_F \subset \mathbb{R}^D \mapsto \mathbb{R}$ be a fixed image and $M(\mathbf{x}') : \Omega_M \subset \mathbb{R}^D \mapsto \mathbb{R}$ a moving one, where $D = 3$ represents their dimensions, $\mathbf{x} \in \mathbb{R}^D$ are the images' coordinates and Ω a closed domain. The geometrical transformation is considered as $\mathbf{W}(\mathbf{x}, \mu) : \Omega_F \times \mathbb{R}^P \mapsto$

Ω_M , where $\mu \in \mathbb{R}^P$ represents the vector of the transformation parameters and P the total number of parameters.

The two-stage heuristic search resolves the large or global mismatches using the MI and EPD similarity measures in its coarse and fine levels respectively. Details of this optimization step are described in Section 4.2.2. The affine stage removes the remainder of the coarse mismatches. Details of the affine EPD registration are provided in [308]. As it is difficult to achieve the required deformations by a single transformation in a non-linear application, the same set of edges is used in both the affine and second stages of the heuristic search so that the same objects are targeted for alignment through multiple transformations. In this case, the strong edges are used so that the neck's trunk and large muscles are aligned first. In the local matching part, firstly, a non-rigid registration aligns the fine mismatches using the diffeomorphic demons (D.Demons) [2] algorithm as this registration is computationally inexpensive and supports invertibility of the transformation. Secondly, the ROI-based EPD registration removes the rest of the small mismatches in the targeted regions by exploiting local AI using a spline-based geometrical transformation. This estimates the local transformation which is more effective than a global estimation and also avoids many local minima by neglecting major parts of the images. The ROI-based EPD registration algorithm is described in Section 4.2.3.

4.2.2 Two-stage Heuristic Search-based Optimization

When there is a large misalignment between two images, traditional continuous and discrete optimization methods have limited success, a problem that becomes more acute in the case of 3D medical images. Heuristic search-based optimization methods can manage this situation as they can be used before continuous

or discrete optimization to correct large mismatches. In the neck dataset, there are large in-plane translational mismatches among individuals. To resolve this problem, a new two-stage hierarchical heuristic search-based optimization that uses MI, EPD similarity measures and translational transformations, as shown in Fig. 4.2, is adopted. It requires an approximate intuitive hypothesis to mitigate any possible large misalignment for a specific application that is easy to obtain for a specific dataset. This optimization has a great impact on the registration framework as it reduces the local minima by decreasing variations between the fixed and moving images. A traditional translation-based registration method optimizes all the transformation parameters which results in a poor match because it treats all the parameters equally irrespective of the amounts they contribute to the overall number of mismatches between two images. On the other hand, the proposed heuristic search-based approach optimizes only the in-plane translation parameters.

The hypothesis regarding large mismatches is that, as they are estimated once for a specific application and do not need to be estimated again for any new images in the same application, proposed heuristic search algorithm can be considered a fully automatic one. A specific application means one that can be used to register neck muscles, brain organs, thigh muscles, and pelvic and heart organs. This hypothesis could depend on the dataset's demographics and image acquisition settings. A smaller sub-hypothesis is estimated from the main hypothesis. In the proposed two-stage heuristic search, the MI is used in the first stage with a large translation step-size and the main hypothesis for aligning the coarse mismatches. To align the large-scale fine mismatches, the EPD is used in the second stage with a small translation step size and the sub-hypothesis. The first stage brings the moving volume closer to the fixed one and then the second aligns the rest of the large mismatches. The MI is used before the EPD in the heuristic search because

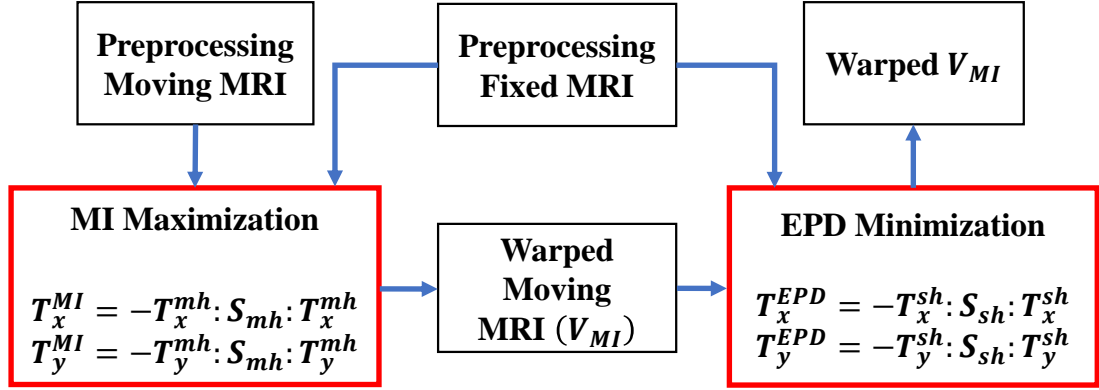


Figure 4.2: Two-stage heuristic search-based optimization. The stages marked by red boxes use same transformation (W_1). The translations T_x^{MI} and T_y^{MI} are varied between main hypothesis values (T_x^{mh} and T_y^{mh} respectively) with step size S_{mh} and translations T_x^{EPD} and T_y^{EPD} are varied between sub-hypothesis values (T_x^{sh} and T_y^{sh} respectively) with step size S_{sh} . Conditions for main and sub-hypotheses $T_x^{mh} > T_x^{sh}$ and $T_y^{mh} > T_y^{sh}$ respectively with step condition $S_{mh} > S_{sh}$.

it can perform well for alignment in the coarse level as it is an intensity-based similarity measure whereas the EPD can do this in the fine level as it is a feature-based, specifically an edge-based, measure. The MI is incapable of matching well in the fine level because neck muscles have almost a similar intensity which results in a local statistical instability problem. Therefore, the MI faces ambiguity when aligning two MRIs during a small translational change. However, during a large change, it can align them, because, as the neck's trunk is almost the same for all individuals and extends over nearly all the image, it does not encounter any statistical instability problem. On the contrary, the EPD is not used during a large translational change in order to avoid false correspondence as, in some cases, the outer strong edge of the moving MRI may correspond to the inner strong edge of the fixed image. However, when the moving image is closer, the chance of false correspondence is greatly reduced.

Since the large mismatches are due to in-plane translations, the translational

transformation W_1 , which is used in our heuristic search in both stages, is computationally fast and expressed as

$$\begin{aligned}x' &= x + T_x \\y' &= y + T_y \\z' &= z + T_z\end{aligned}\tag{4.1}$$

where (x, y, z) and (x', y', z') represent the coordinates of the fixed and moving images respectively and (T_x, T_y, T_z) the translations to the corresponding coordinates.

The parameter vector of W_1 is $\mu_1 = \{T_x, T_y, T_z\}$ and becomes $\mu_{1h} = \{T_x, T_y\}$ for the heuristic search optimization due to the large mismatches of in-plane translations. The numbers of parameters for optimization and transformation using heuristic search-based optimization depend on the nature of the dataset, with the objective function for its first stage

$$\hat{\mu}_{1hMI} = \arg \max_{\mu_{1hMI}} \mathbf{MI}(F, M \circ W_1)\tag{4.2}$$

where μ_{1hMI} represents the parameters vector for the MI stage which is subject to the main hypothesis.

Let the main hypothesis be T_x^{mh} and T_y^{mh} with a step size of S_{mh} and the translations varied in the first stage as

$$\begin{aligned}T_x^{MI} &= -T_x^{mh} : S_{mh} : T_x^{mh} \\T_y^{MI} &= -T_y^{mh} : S_{mh} : T_y^{mh}\end{aligned}\tag{4.3}$$

The translations are varied between the negative and positive values of the

main hypothesis to make the search symmetric and the objective function for the second stage of the heuristic search is

$$\hat{\mu}_{1hEPD} = \arg \min_{\mu_{1hEPD}} \mathbf{EPD}(F, M \circ W_1) \quad (4.4)$$

where μ_{1hEPD} represents the parameters vector for the EPD stage which is subject to the sub-hypothesis.

Let the sub-hypothesis be T_x^{sh} and T_y^{sh} with a step size of S_{sh} , where $T_x^{mh} > T_x^{sh}$, $T_y^{mh} > T_y^{sh}$ and $S_{mh} > S_{sh}$ make the heuristic search a hierarchical optimization. The translations are varied in the second stage using equation (4.3).

The parameters vector for the heuristic search optimization is the concatenation of $\hat{\mu}_{1hMI}$ and $\hat{\mu}_{1hEPD}$ since both stages are considered cascaded and given as

$$\mu_{1h} = \hat{\mu}_{1hMI}(\hat{\mu}_{1hEPD}) \quad (4.5)$$

The EPD similarity measure, which is feature-based and computationally efficient, is used in the second stage of the heuristic search as well as in the affine and ROI-based registration stages of the framework. It uses the edges of images as features and is intended to reduce the distances of the corresponding edges between two images. The numerical measure of the EPD is calculated by averaging the pixel values of the chamfer distance transform of one image at the positions of the superimposed edge points of the other. The chamfer distance transform is an approximation of the Euclidean distance transform, where each pixel value represents the approximate distance to its nearest edge in an image. The chamfer distance image is generated using the 3D HCMA algorithm [330]. The chamfer distance image and the corresponding edge image are shown in Fig. 4.3. Some additional merits and aspects of the EPD are discussed in Section 4.2.3 and it is

Algorithm 2 Two-stage heuristic search algorithm

Input $F, M, T_x^{mh}, T_y^{mh}, T_x^{sh}, T_y^{sh}, S_{mh}, S_{sh}$.
Output $\hat{\mu}_{1hMI}, \hat{\mu}_{1hEPD}$.

- 1: Initialize S_{max} and $\hat{\mu}_{1hMI}$.
- 2: **for** $T_x^{MI} = -T_x^{mh} : S_{mh} : T_x^{mh}$ **do**
- 3: **for** $T_y^{MI} = -T_y^{mh} : S_{mh} : T_y^{mh}$ **do**
- 4: Update T_x^{MI} and T_y^{MI} in $\hat{\mu}_{1hMI}$.
- 5: Warp M .
- 6: Calculate $MI(S)$ using warped M and F .
- 7: **if** $S > S_{max}$ **then**
- 8: Update S_{max} by S .
- 9: Save corresponding T_x^{MI} in tx and T_y^{MI} in ty .
- 10: **end if**
- 11: **end for**
- 12: **end for**
- 13: Update $\hat{\mu}_{1hMI}$ by tx and ty .
- 14: Warp M by $\hat{\mu}_{1hMI}$ to get V_{MI} .
- 15: Repeat statements 1-13 to minimize EPD using $V_{MI}, F, T_x^{sh}, T_y^{sh}$ and S_{sh} .
- 16: **Return** $\hat{\mu}_{1hMI}, \hat{\mu}_{1hEPD}$.

mathematically expressed as

$$S(\mu_k) = \frac{1}{3T} \sum_{(x,y,z) \in \beta} C(x', y', z') \quad (4.6)$$

where β denotes the list of edge points in the fixed image (F), C is the chamfer distance image of the moving image (M) and T is the total number of edge voxels in β . The value of 3 in equation 4.6 compensates the unit distance in the chamfer distance transform, with the proposed heuristic search-based optimization algorithm described in Algorithm 2.

4.2.3 ROI-based EPD Registration

After the two-stage heuristic approach and non-rigid registrations are performed, most of the large mismatches are removed while the small ones remain. To eliminate small local mismatches, a novel ROI-based EPD registration algorithm is

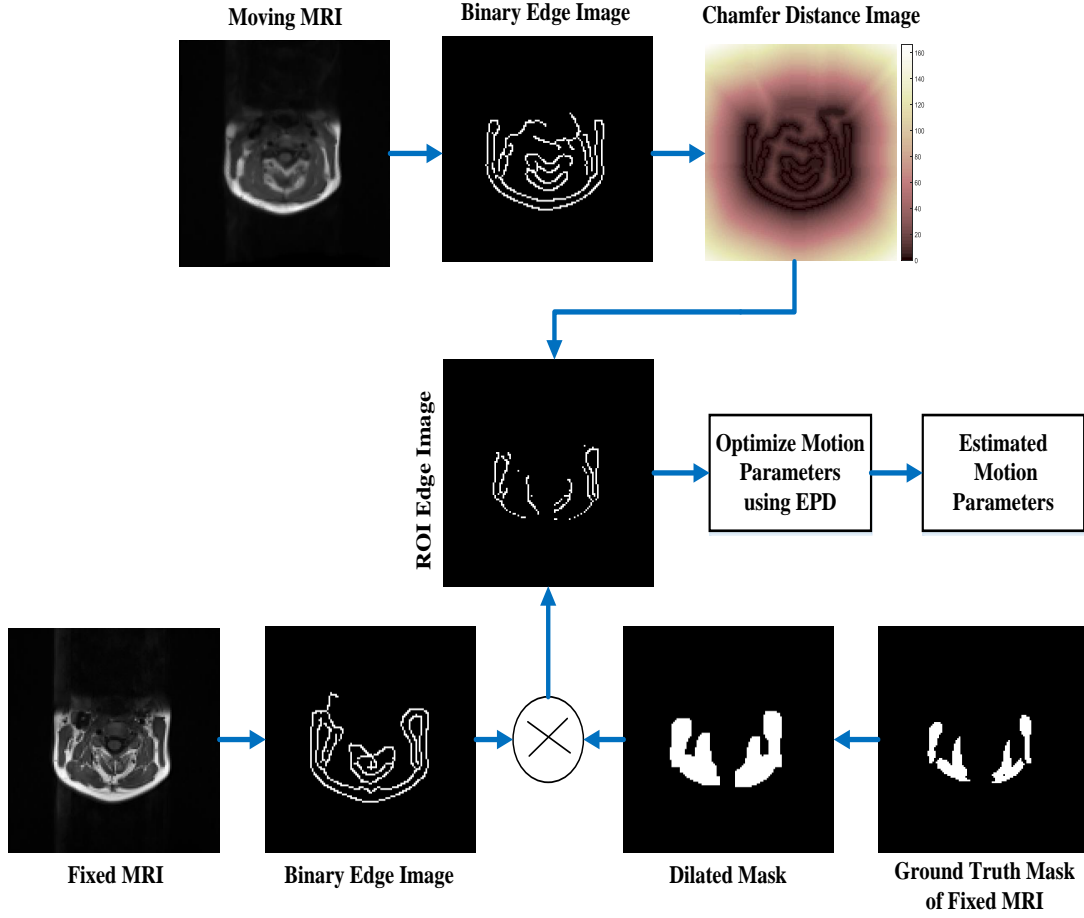


Figure 4.3: Proposed ROI-based EPD registration process. Although two-dimensional (2D) images are shown in flowchart for representational convenience, all operations are performed in 3D MRI with gradient descent optimization and FFD geometrical transformation.

proposed, with the ROI selected using the ground truth available in atlas-based segmentation methods. These methods use manual boundary segmentation information of the anatomical organs of individuals in the database to segment them in a new individual's medical image [335, 336].

The EPD, which is a geometric type-matching criteria, is used in the proposed ROI-based registration process for the following reasons. Firstly, the EPD can exploit AI from medical images whereas intensity-based similarity measures, such as the MI, NCC and SSD, cannot use AI which is crucial for robust correspondence matching. Secondly, it can detect AI in even smaller regions and does not face

the problem of local statistical instability whereas MI and NCC do [39]. In the proposed approach, AI is extracted in the form of morphological shapes which, although sometimes not complete for neck muscles in medical images, help to guide local correspondences. Thirdly, the MI, NCC and SSD are incapable of providing a measure for distinguishing muscles in a neck dataset in the ROI where those with similar intensities remain close to each other. In particular, the MI or local MI, NCC and SSD may provide mis-correspondences for neck muscles in MRIs due to their compactness and similar intensity whereas the ROI EPD can provide distinguishing values because the shapes inside the ROI form the unique shape shown in Fig. 4.3. Finally, the distinguishing nature of the ROI EPD creates a convex objective function for the registration process which provides leverage to the optimization.

Some other popular geometric types of matching criteria, such as the scale-invariant feature transform (SIFT) and speeded-up robust features (SURF) could be considered alternatives to EPD, require detection of key points in the images. However, medical images typically do not have sufficient details for accurate key-point detection which make the SIFT and SURF methods unsuitable for our application.

Moreover, the ROI-based EPD has advantages over a global EPD-based approach. It can avoid most of the local minima faced by a global EPD approach as it only uses the edges located inside the ROI. As a result, the optimization process can easily find the optimum value of the EPD for the edges in the ROI and hence provide better alignment.

A flowchart of the proposed ROI-based EPD registration is shown in Fig. 4.3. A spline-based free form deformation (FFD) model is used to obtain an estimated deformation field and the ground truth of the fixed MRI volume to generate a

binary mask of the muscles inside the ROI. The fixed volume can be considered an atlas (an individual with known muscle boundaries) and the moving image represents a new individual whose neck muscles are required to be segmented. After registration is completed, the muscle boundaries are transferred from the fixed volume to the moving volume since the exploited transformations in the proposed framework in Fig. 4.1 are inversely consistent. The binary mask is dilated by a square structuring element so that it can properly extract all the necessary anatomical shapes from the binary edge volume of the fixed volume generated using a canny edge detector. If no dilation is performed, some necessary shapes may be missed. Then, the volumes of the dilated mask and binary edge are multiplied to obtain an extracted shape from the edge volume to guide the registration to find local deformations.

The muscles inside the ROI are considered to be a unique shape rather than separately because parts of a muscle's contour can easily mis-correspond with another muscle's contour owing to the almost similarly shaped muscles that are close to each other in the neck. Although parts of this shape may also mis-coincide, the percentage will be less than that of the individual muscle's shape due to the former's larger size. Therefore, the real coincident parts of the unique shape will dominate in the numerical value of the EPD. As a result, this shape can provide better alignment than separate muscles and, moreover, considering all the muscles together, rather than separately, in the ROI process makes the algorithm computationally less expensive.

The binary edge volume is also calculated from the moving one and then the chamfer distance transform volume ($C(x', y', z')$) is generated. The locations of the edge voxels (β_{ROI}) are calculated from the extracted binary shape volume,

Algorithm 3 ROI-based EPD registration process

Input F, M, labels .

Output $\hat{\mu}_{4ROI}$.

- 1: Initialize $\hat{\mu}_{4ROI}^0$, maximum iteration number $i_{max} = 100$ $LT = 0.0002$, $\sigma = 1.5$ and $UT = 0.25$.
 - 2: Calculate edge volume from F .
 - 3: Generate binary mask from the **labels**.
 - 4: Dilate the mask by 7 pixels.
 - 5: Extract edges for ROI using the mask and edge volume.
 - 6: Calculate β_{ROI} from the extracted edges.
 - 7: Warp M^0 using $\hat{\mu}_{4ROI}^0$.
 - 8: **for** $i = 1 : i_{max}$ **do**
 - 9: Calculate edge volume (E_M) from M .
 - 10: Calculate Chamfer distance volume C_W^M from E_M .
 - 11: Calculate EPD using C_W^M and β_{ROI} .
 - 12: Update $\hat{\mu}_{4ROI}$ using Gradient-descent optimization and FFD with C_W^M and β_{ROI} .
 - 13: Warp moving volume: $M^i = D(M^{i-1}, \hat{\mu}_{4ROI}^i)$.
 - 14: **end for**
 - 15: **Return** $\hat{\mu}_{4ROI}$.
-

with the ROI EPD value expressed as

$$S_{ROI}(\mu_{4ROI}) = \frac{1}{3N} \sum_{(x,y,z) \in \beta_{ROI}} C(x', y', z') \quad (4.7)$$

where N is the total number of edge voxels in β_{ROI} and μ_{4ROI} is the parameters vector of the ROI-based EPD registration.

The S_{ROI} measure is optimized using the gradient descent optimization algorithm [243] and then the objective function is computed as

$$\hat{\mu}_{4ROI} = \arg \min_{\mu_{4ROI}} \mathbf{S}_{ROI}(F, M \circ W_4) \quad (4.8)$$

where W_4 is the geometrical transformation of the ROI-based EPD registration.

The overall ROI-based EPD registration process is described in Algorithm 3 in which it should be noted that the chamfer distance volume is updated in every

iteration.

4.3 Experiments and Results

The experiments for this work were performed on a HP z230 tower workstation with a 3.40 GHz Intel(R) Core(TM) i7-4770 processor and 16 GB RAM running the Windows 10 operating system. The major parts of the framework were implemented using MATLAB and minor ones through interfacing C and C++ with MATLAB via a MATLAB executable (MEX).

4.3.1 Data and Pre-processing

A sample of patients was selected by the Canberra Imaging Group at the John James Calvary Hospital, Canberra, Australia to study whiplash injury of the neck. Ethics approval to conduct the study was obtained from the Office of Research Integrity of the Australian National University (ANU), Australia, and Health Directorate of the Australian Capital Territory (ACT) Government, Australia. Informed consent was obtained from the patients whose ages, weights and heights ranged from 19 to 35 years, 45 to 121 kg and 1.5 to 1.9 m respectively. A 3 Tesla Skyra (Siemens, Erlangen, Germany) MRI scanner was used to capture axial T1-weighted spin echo MRIs with sizes of $256 \times 256 \times 45$ and voxel spacings of $0.8594 \text{ mm} \times 0.8594 \text{ mm} \times 4 \text{ mm}$ for each patient's cervical region. The images were collected using the following scanner settings: echo time of 15 to 16 ms, repetition time of 750 to 1120 ms and a 100 cm^2 phase field of view. Nineteen patients for these experiments were selected from the ongoing study, with images cropped

Table 4.1: Parameter settings for symmetric image normalization (SyN) [3] and D.Demons [2] registrations on the neck dataset conducted through ANTs and MATLAB respectively. Multi-stage registrations were used in the SyN method in which the rigid, affine and SyN geometrical transformations used the same smoothing sigma values, shrink factors and convergence.

Parameter Name	SyN [3]	D.Demons [2]
Script	antsRegistrationSyN.sh	imregdemons.m
Dimension	3	3
Pyramid level	-	3
Iterations	-	100x100x100
Transformations	Rigid[0.1], Affine[0.1]	-
Metric	MI[F,I,1,32,Regular,0.25]	-
Smoothing sigmas	3x2x1x0vox	1
Shrink factors	8x4x2x1	-
Convergence	[1000x500x250x100,1e-6,10]	-
Transformation	SyN[0.1,3,0]	-
Metric	CC[F,I,1,4]	-
Convergence	[100x70x50x20,1e-6,10]	-

to avoid unnecessary regions and interpolated to $128 \times 128 \times 128$ pixel volumes. Intensity inhomogeneity correction was not performed for the images.

The interpolated volumes of the left sternocleidomastoid (Muscle 1), right sternocleidomastoid (Muscle 2), left semispinalis capitis (Muscle 3), right semispinalis capitis (Muscle 4), left splenius capitis (Muscle 5) and right splenius capitis (Muscle 6) cervical muscles were then segmented manually by an ANU medical student and used as the ground truth. Delineations were performed by a MATLAB graphical user interface (GUI) designed specifically for this study and later validated by two other senior medical experts from the ANU. The GUI enabled the MRIs to be segmented in 2D with as many vertices as necessary to achieve smooth contours.

4.3.2 Results Analysis

The main and sub-hypotheses were empirically determined for the dataset as $T_x^{mh} = 8$, $T_y^{mh} = 14$ and $S_{mh} = 2$, and $T_x^{sh} = 3$, $T_y^{sh} = 4$ and $S_{mh} = 1$ respectively. The dilation in the ROI EPD registration was performed by a square structuring element with a width of 7 pixels in the 3D binary volume. The registration results were evaluated numerically using the volumetric dice similarity coefficient (DSC) and hausdorff distance (HD). Higher DSC and lower HD values are desirable for good registration results.

The inter-subject registration was performed by keeping one patient fixed and the others moving, and this approach was repeated for all the patients in the dataset. Therefore, the possibilities for one fixed image was 18 and, for the 19 patients, was 342. To assess the effect of the two-stage heuristic search-based optimization and ROI-based EPD registration on the proposed framework, experiments were performed without these techniques and denoted as 'Without Heuristic' and 'Without ROI' respectively. The proposed method was evaluated against two state-of-the-art deformable registration methods, Symmetric Image Normalization (SyN) [3] and diffeomorphic Demons (Demons) [2]. The SyN algorithm was implemented using advanced normalization tools (ANTs) [337] on a Linux platform via a virtual box using Ubuntu version 18.04 and the Demons algorithm used was the one available in MATLAB version R2017b. The ANTs is a well-recognized library for medical image registration and segmentation while the SyN algorithm uses an initial moving transform, with rigid-MI and affine-MI linear registrations for initialization. Details of the parameter settings for the SyN and Demons registrations are shown in Table 4.1, with the script from the ANTs used for good-quality deformable registration.

Fig. 4.4 shows visual result of the proposed method compared to the Demons and SyN methods for one of the 342 cases. The initial mis-alignment is very high as shown in Fig. 4.4(c). The proposed method matches the boundaries of the muscles better than other methods. Specifically, the sternocleidomastoid muscles' boundaries are aligned more accurately than others. Fig. 4.5 shows a visual results comparison in the coronal plane for the proposed, Demons and SyN methods for one of the 342 cases. It exhibits similar alignment improvement for the proposed method over other methods as for the axial visual results. Fig. 4.6 exhibits the visual results in terms of muscles contours for the different methods, when compared with the ground truths, in both the axial and coronal views. The proposed method's contours are more fairly matched when compared to the other algorithms.

Fig. 4.7 displays the registration accuracy for the DSC on the neck MRI data by the proposed, without heuristic, without ROI, Demons [2] and SyN [3] methods for all muscles separately (Fig. 4.7(a)) and combined (Fig. 4.7(b)). Fig. 4.7(a) shows that the proposed method is significantly better than the without heuristic, without ROI and state-of-the-art methods for all the muscles. Fig. 4.7(b) shows that the proposed method substantially outperforms the other compared methods. It is clear that the two-stage heuristic search-based optimization and ROI-based EPD has great effect on the registration performance.

Fig. 4.8 exhibits the registration accuracy in HD (mm) on the neck MRI data by the proposed, without heuristic, without ROI, Demons [2] and SyN [3] algorithms for all muscles separately (Fig. 4.8(a)) and combined (Fig. 4.8(b)). Fig. 4.8(a) shows that the proposed method performs better than the others for all the muscles except for Muscle 4 and Muscle 6. The proposed, without heuristic and without ROI exhibit almost similar performance except for some more outliers

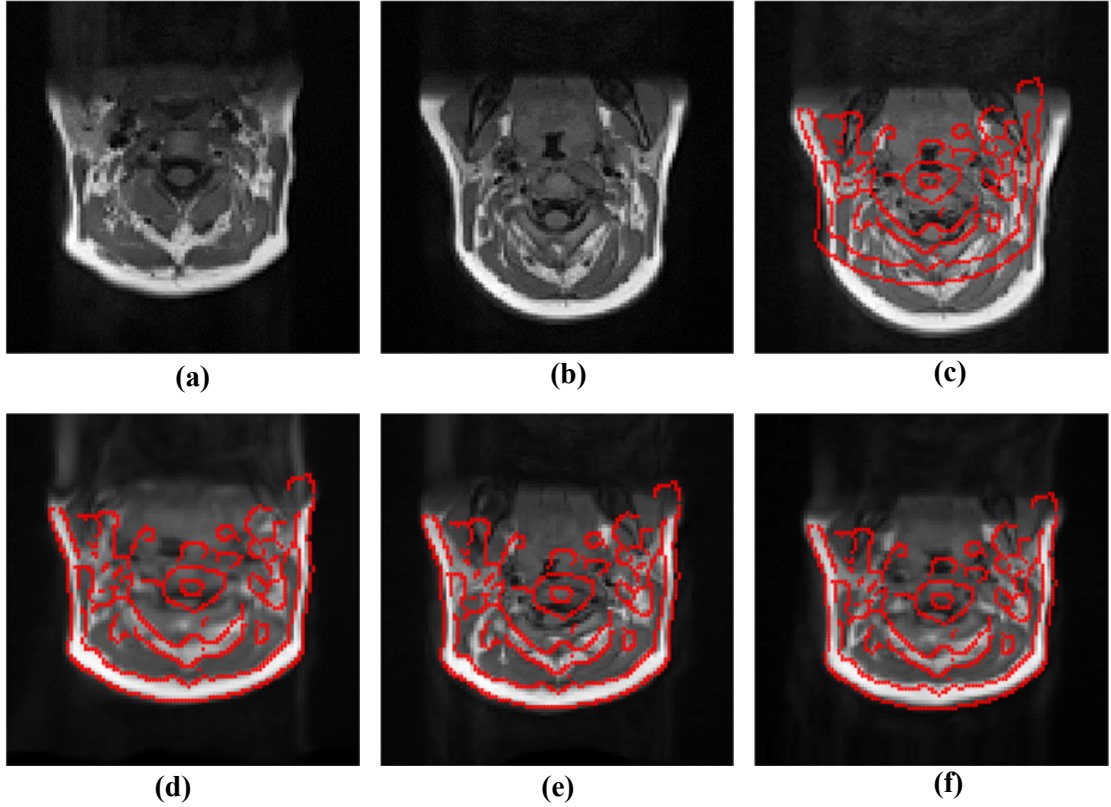


Figure 4.4: Comparison of axial visual results of one of the 342 inter-subject cases. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with fixed image's edges superimposed; fixed image's edges superimposed on the moving image after (d) Proposed, (e) Demons, (f) SyN. The Demons and SyN are state-of-the-art deformable registration methods.

in the without heuristic method. However, the proposed method performs much better in terms of the DSC than the without heuristic and without ROI methods. Fig. 4.8(b) shows that the proposed method also performs better than the other methods in terms of the HD. The samples in the without heuristic method are more dispersed than the proposed method which shows that the heuristic search not only increases the accuracy but also the consistency of the model.

Table 4.2 provides the overall mean and median DSC and HD values for the proposed, without heuristic, without ROI, proposed in Chapter-3, Demons [2] and SyN [3] methods. The median values are obtained from Fig. 4.7(b) and Fig. 4.8(b).

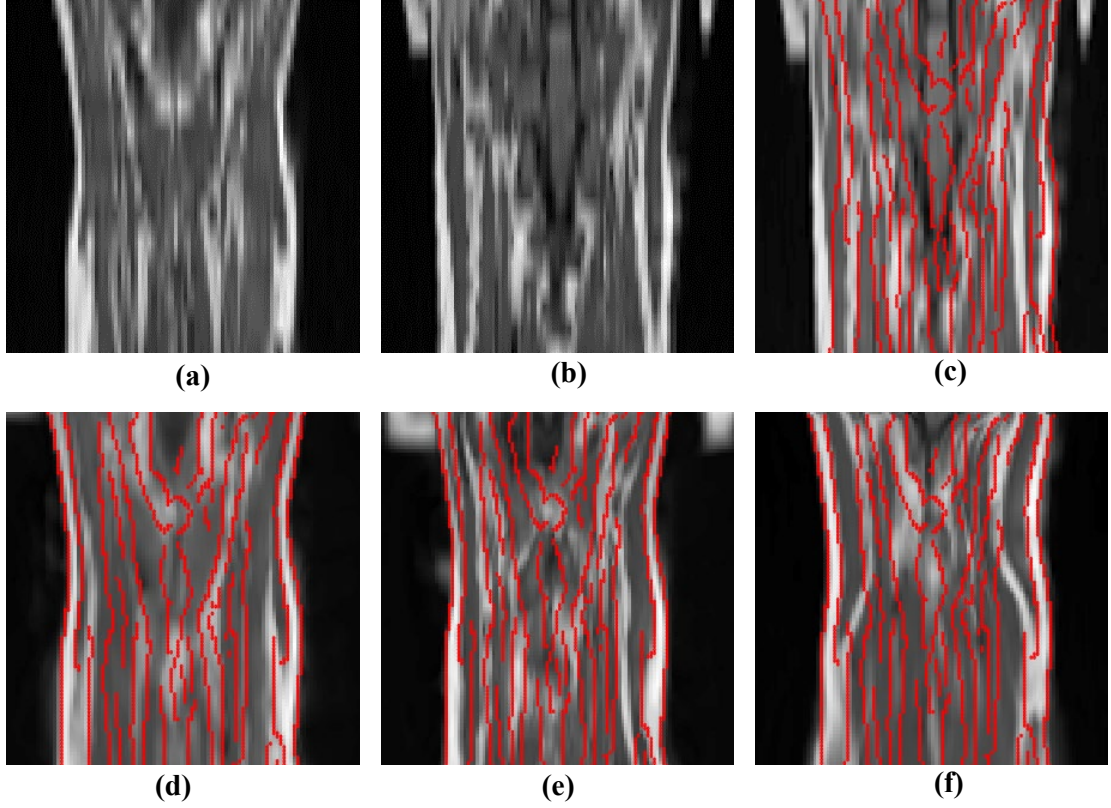


Figure 4.5: Comparison of coronal visual results of one of the 342 inter-subject cases. It should be noted that, although the registration is performed between the two 3D MRI volumes, the results are illustrated in 2D for presentation convenience. (a) fixed image, (b) moving image before registration, (c) moving image before registration with fixed image's edges superimposed; fixed image's edges superimposed on the moving image after (d) Proposed, (e) Demons, (f) SyN. The Demons and SyN are state-of-the-art deformable registration methods.

This table shows that the proposed method significantly outperforms the other methods in both metrics.

4.3.3 Computational Complexity

The computational times for the compared methods to register two volumes were also calculated, as shown in Table 4.3. These times show that the proposed method is sufficiently fast and, although slower than the Demons algorithm, has better

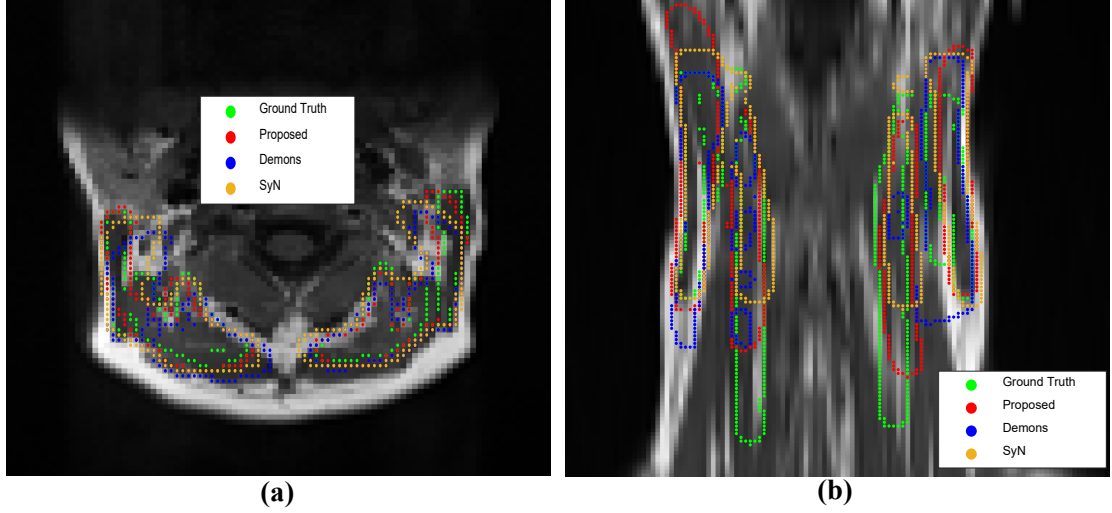


Figure 4.6: Comparison of visual results in terms of muscles contours among proposed, Demons [2] and SyN [3] with ground truths: (a) axial view and (b) coronal view

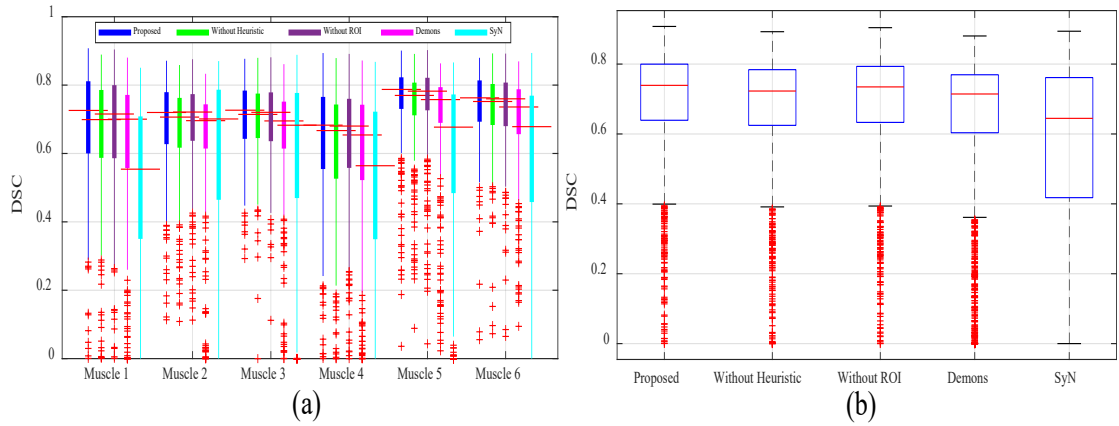


Figure 4.7: Registration results using the proposed algorithm, without heuristic, without ROI, Demons [2] and SyN [3] on the neck MRI data in volumetric dice similarity coefficient (DSC) (higher values are better). The without heuristic and without ROI versions refer to the proposed framework excluding the two-stage heuristic search-based optimization and ROI-based EPD registration respectively. (a) all muscles separately for the 342 test cases; (b) all muscles combined for $342 \text{ cases} \times 6 \text{ muscles} = 2052 \text{ DSCs}$.

accuracy. The low computational time for the Demons algorithm can also be attributed to the professional implementation in MATLAB.

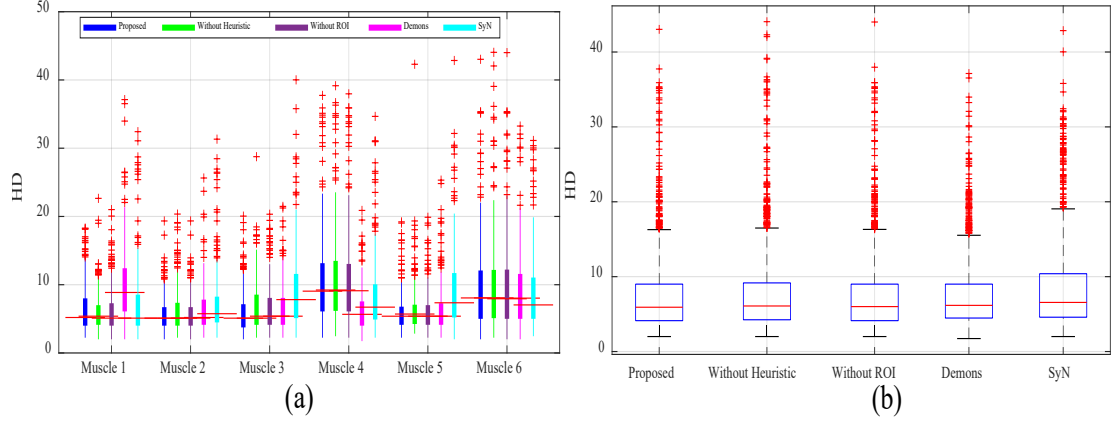


Figure 4.8: Registration results using the proposed, without heuristic, without ROI, Demons [2] and SyN [3] algorithms. For the Hausdorff distance (HD) in mm (lower values are better). The without heuristic and without ROI methods consist of the proposed framework after excluding the two-stage heuristic search-based optimization and ROI-based EPD registration respectively. (a) all muscles separately for the 342 test cases; (b) all muscles combined for 342 cases \times 6 muscles = 2052 HDs.

4.4 Discussion and Conclusion

An automatic 3D-3D deformable registration framework using novel two-stage heuristic search and ROI-EPD techniques was developed. It outperformed the two state-of-the-art methods, Demons [2] and SyN [3], on a real clinical neck MRI dataset in terms of the DSC and HD. Also, the proposed method was reasonably fast compared to these approaches, (Table 4.3). Its execution time could be further reduced if it was implemented entirely in the C++ language using optimized coding. It was proven to be robust as it performed well on clinical data with inhomogeneity and different MRI scanner settings.

As the heuristic search brought a moving volume closer to the fixed volume from which it was far away, the overall registration performance increased more than without it. The two-stage heuristic search helped to align large mismatches whereas traditional optimization methods failed to achieve convergence in these

Table 4.2: Overall registration performance in the neck MRI dataset using the proposed, without heuristic, without ROI, proposed in Chapter 3, Demons [2] and SyN [3] methods over 2052 samples.

Methods	Mean DSC	Median DSC	Mean HD	Median HD
Proposed	0.6982 ± 0.1462	0.7391	7.3022 ± 4.7583	5.9161
Without Heuristic	0.6812 ± 0.1503	0.7228	7.6233 ± 5.3345	6.0828
Without ROI	0.6892 ± 0.1537	0.7346	7.4100 ± 4.8445	6.0000
Proposed in Chapter-3	0.6649 ± 0.1677	0.7145	8.1349 ± 5.5490	6.4031
Demons [2]	0.6487 ± 0.1885	0.7144	7.4312 ± 4.3719	6.1644
SyN [3]	0.5575 ± 0.2566	0.6446	7.8884 ± 5.1996	6.5574

cases. The proposed search was fast as it was gradient-free and did not require searching over a large range. Also, it used an intensity-based similarity measure in its coarse level and a feature-based one in its fine level, with the latter greatly reducing the computational complexity, along with the translational transformation which was also fast to compute. Furthermore, as the two-stage heuristic search found an optimum transformation within a predefined set of plausible solutions, it reduced the time complexity. Moreover, it required negligible user interaction compared with conventional heuristic search methods as it was not necessary for a user to interact with a specific application for a new image. This resulted in a single application of the algorithm being fully automatic. Although interaction would be required for a new application, it would be very easy as the user would be required to know only approximate prior knowledge of the large mismatches in the dataset, not the pre-segmentation of objects. Alternatively, conventional heuristic search methods require comprehensive user interaction in terms of the pre-segmentation of objects on both registration images. If a new image in the same application has to be registered, it is required to have pre-segmentation. Furthermore, the heuristic search greatly reduced local minima by minimizing the number of targeted large mismatches which helped the framework more effectively align the images in its subsequent stages using continuous or discrete optimization methods.

Table 4.3: Computational time to register two volumes using the proposed, proposed in Chapter 3, Demons [2] and SyN [3] algorithms.

Methods	Time (minutes)
Proposed	2.26
Proposed in Chapter-3	33
Demons [2]	0.81
SyN [3]	69.95

An ROI-based registration was used in the proposed framework because it can estimate transformations more effectively and accurately than a whole-image registration [27]. Also, it detects correspondences more accurately and robustly than a global one because it only needs to search small regions. Furthermore, it is compatible with our practical domain as, for example, controlling a large population is much more difficult than controlling a small one and whole-image registration is more difficult. Also, as specific organs are required to segment in clinical applications to analyze diseases, ROI-based registration is consistent with relevant clinical motives. Due to the ROI-based EPD’s capabilities to exploit AI in small regions, not face local statistical instability problems and provide distinguishing values owing to its unique shape inside the ROI, it increases good alignment even using coarse levels of a spline-based transformation which reduces the computational time. On the other hand, traditional multi-resolution methods use finer levels of the spline to deal with small mismatches and are computationally expensive. Also, some other conventional registration methods try to deal with small mismatches using the sum of the local measures with local optimization and are computationally expensive [39]. Furthermore, the proposed ROI-based EPD registration considers all the objects simultaneously which enables fast computation.

The proposed method was applied on 19 clinical neck MRIs for experimental validation. Generating the 3D ground truth was a time-consuming task. Further

studies of a larger neck dataset and other clinical scenarios to validate our proposed method will be ongoing using diverse datasets. More research is required to investigate applications of this method to neck pain and, thereby, medical interventions such as radiation therapy, chemotherapy and injection dosages. A fine resolution of the FFD transformation could also be examined with local gradient descent optimization using the ROI-based EPD algorithm.

Chapter 5

Deep Learning-based Neck Muscles Segmentation

In this chapter, deep learning-based segmentation methods using three different types of methodological propositions are presented and analyzed for the application of neck muscles segmentation.

In the first analysis, evaluations of U-Net architecture-based approaches for neck muscles segmentation are presented. Numerous versions of the U-Net architecture have emerged for the task of medical image segmentation due to their significant success over the last few years for semantic medical image segmentation. 10 versions of the U-Net convolutional neural network (CNN) are analysed, 6 direct (U-Net, CRF-Net, A-Net, MFP-Net, R2U-Net and U-Net++) and 4 modified (R2A-Net, R2A-Net++, PMS-Net and MS-Net) methods. The modifications are inspired by recent multi-scale and multi-stream techniques for deep learning algorithms. T1-weighted MR images of the distal ends of the C3 vertebrae of 45 subjects are used in the evaluations. An analysis of the numerical results indicates that the R2U-Net architecture achieves the best accuracy.

The following research paper has been accepted based on this chapter of sections: 5.1.2, 5.2.3, 5.3.1, 5.4.1

- **Abdulla Al Suman**, Yash Khemchandani, Md. Asikuzzaman, Alexandra Louise Webb, Diana M. Perriman, Murat Tahtali and Mark Richard Pickering, “Evaluation of U-Net CNN Approaches for Human Neck MRI Segmentation,” *Digital Image Computing: Techniques and Applications (DICTA)*, 29 November–2 December. 2020, Melbourne, Australia.

In the second analysis, a new two-stage U-Net++ (TS-UNet++) architecture that uses two different types of deep CNNs (DCNNs) rather than a traditional multi-stage network is proposed. It uses the U-Net++ in the first stage and the U-Net in the second. Extra convolutional block is added before the output layer of the multi-stage network to better extract the high-level features. A new concatenation-based fusion structure is incorporated in this architecture to enable deep supervision. More convolutional layers are added after each concatenation of the fusion structure to extract more representative features. The performance of this proposed method on a neck dataset is compared with those of the U-Net, U-Net++ and two-stage TS-UNet ones, with the results indicating that it provides the best performance.

In the third analysis, an explicit attention method in which the attention is performed through a region of interest (ROI) evolved from the ground truth via dilation is proposed. It does not require any additional CNN, such as a cascaded one, to localize the ROI. Attention in a CNN is sensitive. This dilated ROI is capable of capturing more relevant regions and suppressing irrelevant ones than a bounding box and region-level coarse annotation. The dilated ROI is used during the training of the CNN and coarse annotation, which does not require any detailed pixel-wise delineation that can be undertaken by a novice person, during testing. When compared with the automatic attention U-Net (A-Unet) and U-Net methods, this proposed ROI-based attention network provides better performance.

The remainder of this chapter is organized as follows: Section 5.1 presents a relevant literature review for each analysis; the image acquisition and pre-processing stages as well as the corresponding methodology for each analysis is presented

in Section 5.2; Section 5.3 presents the experimental numerical results; and Section 5.4 contains the discussion and conclusions.

5.1 Introduction

5.1.1 Evaluation of U-Net CNN Approaches

There are many medical image segmentation methods in the literature, such as: template matching, deformable model fitting, edge detection and learning-based approaches. However, deep learning-based techniques have revolutionized segmentation tasks [59]; in particular, the U-Net [6] architecture is popular for biomedical image segmentation and has achieved remarkable success due to its flexibility [59]. Over the last few years, there have been rapid advances based on it for medical and natural image segmentations [60]. To choose the best algorithms for further research, a comparative study of the many U-Net-based approaches is required.

In this chapter, an evaluation of U-Net-based approaches is presented for the application of neck muscles segmentation in MRIs. These approaches include direct and modified models, with the former based on slight variations in the U-Net architecture (no major architectural change) and the latter on the integration of some recent techniques for deep learning-based segmentation.

Deep learning-based segmentation for cell segmentation was recently evaluated by Caicedo *et al.* [338] who considered 5 strategies: deep learning (U-Net, DeepCell), classical machine learning (Random Forest) and classical image processing (advanced and basic CellProfiler) for segmenting the cell nuclei in fluorescent images. In their experiments, they used 200 images with 23,165 manually annotated

nuclei. The performances of the Random Forest and advanced and basic CellProfiler methods were found to be worse than those of the deep learning strategies. Hansch *et al.* [339] compared the 2D U-Net, 2D ensemble U-Net and 3D U-Net techniques for parotid gland segmentation from a head and neck computed tomography (CT) in the 2015 Medical Image Computing and Computer-assisted Intervention (MICCAI) challenge. The 2D ensemble U-Net was actually a combination of three 2D U-Nets arranged in parallel to work on sagittal, coronal and axial patches, respectively, from which predictions of the individual 2D models were combined through a majority voting rule.

In this chapter, 10 popular deep learning-based segmentation models, 6 direct (U-Net, CRF-Unet, A-Unet, MFP-Unet, R2Unet and U-Net++) and 4 modified (R2A-Unet, R2A-Unet++, PMS-Unet and MS-Unet) are evaluated, for neck muscles segmentation, using the T1-weighted MRIs selected from the distal ends of the C3 cervical levels of 45 subjects. The dice similarity coefficient (DSC) and directional Hausdorff distance (DHD) are used to analyze the results. Brief descriptions of the models and an explanation of the modifications incorporated are provided in Section 5.2.3. The codes for the models have been made available to the public by the corresponding authors. This study will help further research by the medical image segmentation community.

5.1.2 Multi-stage-based Deep Learning

The U-Net architecture has some limitations. Firstly, it reduces the resolution of a feature map due to its consecutive operations of pooling and striding convolution whereas detailed spatial information is beneficial for dense predictions. If high resolution is maintained, it does not accelerate and makes optimization difficult. Secondly, although a deeper network can learn more complex features and yield

a better performance, the depth of the U-Net cannot be increased due to its gradient disappearing with an increasing depth during training which results in training problems.

Bi *et al.* [340] proposed a multi-stage fully convolutional network (FCN) method for skin lesion segmentation. They also introduced a parallel integration technique for combining the outputs of the stages. Tang *et al.* [4] presented a multi-stage U-Net (MS-Unet) with a contextual information fusion structure (CIFS) to combine low- and high-level features in a multi-scale feature space for skin lesion segmentation. This architecture has some advantages. Firstly, it can provide a coarse-to-fine approach for segmentation in which the early stages can produce coarse and localization information, which is refined in the later stages. Secondly, distinctive additional features can be learned, which may not be possible when using a single stage. Thirdly, the depth can be increased without the gradient vanishing through fusing features from multiple stages to provide an additional information flow. Finally, it has the benefits of a large training dataset even when used on a small one.

Inspired by Bi *et al.* [340] and Tang *et al.* [4], a new two-stage U-Net++ (TS-Unet++) using two different types of DCNNs is proposed. This network is applied on a multi-class neck muscles segmentation task and makes use of the following three key contributions.

- The U-Net++ is used in the first stage and U-Net in the second.
- Extra convolutional block is added before the output layer of the multi-stage to better extract its high-level features.
- A new concatenation-based fusion structure is incorporated in its architecture to enable deep supervision.

- More convolutional layers are added after each concatenation of the fusion structure to extract more representative features.

The proposed method is compared with the U-Net, U-Net++ and TS-UNet approaches on the neck dataset using the DSC and DHD metrics, with the results indicating that it outperforms these competing approaches.

5.1.3 ROI-based Attention

Attention-based models have recently been studied extensively in the deep learning community to more effectively use intermediate feature maps. There are many cascaded attention models for different applications, such as abdominal CT [61], cardiac CT [62], cardiac MRI [63], colorectal tumor [64] and left ventricle [65] segmentations. For lung nodule detection in [66], one stage is used to localize the ROI and another for dense predictions. Some cascaded models use a bounding box to guide attention in the dense prediction stage during training. Example applications of this approach include: medical report generation [341, 342], text classification [343], disease classification [301], lesion detection [300] and pancreas segmentation [302]. This is called hard attention, with some models using the first stage to find the parameters of the bounding box for the prediction stage [61, 62]. However, cascaded models require excessive computational resources and repetitively extract similar low-level features.

Recently, some automatic attention models [7, 67–71] without cascaded schemes have been proposed. They automatically focus on structures during the training of the CNNs and implicitly suppress irrelevant regions. These methods can be further classified as supervised [71] and unsupervised [7, 67, 69] attention mechanisms. The performance of unsupervised attention may be worse than that of an

approach without an attention mechanism and its attention map may be affected by noise which causes incorrect attention. A supervised attention mechanism uses external supervision through an explicit region-level coarse annotation as well as self-attention which performs better than only self-attention [344]. However, as supervised attention cannot focus on the ROI in a complex case of compact and similar small multi classes with large variabilities and low contrast, finer guided attention is required to handle the challenges involved.

An explicit attention method in which the attention is performed through a ROI evolved from the ground truth via dilation is proposed. It does not require any additional CNN like a cascaded method to localize the ROI as attention in a CNN is sensitive to the ROI's area. This dilated ROI can capture more relevant regions and suppress irrelevant ones when compared to bounding box or region-level coarse annotation. The dilated ROI is used during the training of the CNN whereas coarse annotation, which does not require any detailed pixel wise delineation and can be undertaken by a novice person, is used during testing. When the proposed ROI-based attention method is compared with the A-Unet and U-Net ones using the DSC and DHD, it provides the best performance.

Table 5.1: Demographics data and parameters of the MRI scanner.

Patient index	Age (years)	Weight (kg)	Height (m)	Repetition time (s)	Echo time (s)	Acquisition date (yyyymmdd)
PT-1	29	57	1.61	746	15	20140106
PT-2	19	59	1.65	737	15	20141216
PT-3	25	57	1.67	827	15	20131217
PT-4	20	56	1.55	750	15	20141212

Continued on next page

Table 5.1 – continued from previous page

Patient index	Age (years)	Weight (kg)	Height (m)	Repetition time (s)	Echo time (s)	Acquisition date (yyyymmdd)
PT-5	32	59	1.64	766	15	20141202
PT-6	23	45	1.60	827	15	20140922
PT-7	23	61	1.72	827	15	20141113
PT-8	27	61	1.62	827	15	20141124
PT-9	24	75	1.75	949	15	20140120
PT-10	27	57	1.65	777	15	20141201
PT-11	32	43	1.50	827	15	20141209
PT-12	30	121	1.78	1120	16	20140315
PT-13	30	105	1.78	1140	15	20140526
PT-14	22	70	1.75	827	15	20150305
PT-15	28	69	1.70	843	15	20140311
PT-16	27	53	1.72	827	15	20140709
PT-17	31	90	1.80	1100	16	20141103
PT-18	27	71	1.67	1110	15	20141113
PT-19	27	52	1.65	827	15	20140510
PT-20	25	80	1.58	827	15	20140417
PT-21	29	68	1.78	827	15	20140711
PT-22	27	50	1.55	827	15	20140701
PT-23	23	77	1.92	1070	15	20140331
PT-24	29	69	1.79	827	15	20131223
PT-25	22	74	1.58	941	15	20140912
PT-26	27	72	1.60	1080	15	20140215

Continued on next page

Table 5.1 – continued from previous page

Patient index	Age (years)	Weight (kg)	Height (m)	Repetition time (s)	Echo time (s)	Acquisition date (yyyymmdd)
PT-27	19	60	1.70	827	15	20140828
PT-28	25	75	1.80	827	16	20141121
PT-29	35	90	1.63	755	15	20150302
PT-30	24	64	1.67	864	15	20140602
PT-31	26	102	2.18	1110	15	20140508
PT-32	25	84	1.80	827	15	20141201
PT-33	34	56	1.58	827	15	20141117
PT-34	22	97	1.71	1110	15	20141119
PT-35	34	75	1.63	901	15	20150220
PT-36	28	69	1.70	843	15	20140311
PT-37	28	75	1.75	848	15	20140827
PT-38	30	80	1.80	1120	16	20140516
PT-39	36	52	1.64	827	15	20140701
PT-40	34	70	1.75	898	15	20141020
PT-41	18	68	1.87	827	15	20140721
PT-42	34	110	1.94	857	15	20140611
PT-43	29	84	1.87	1070	15	20140311
PT-44	28	59	1.65	827	15	20131204
PT-45	29	67	1.72	1020	15	20140603

5.2 Methods and Materials

5.2.1 Image Acquisition and Pre-processing

The dataset was adopted from an ongoing larger study entitled “Characterizing whiplash injury using magnetic resonance imaging” conducted by the Australian National University (ANU) Medical School with ethics approval granted. It consists of 45 patients (28 females and 17 males) aged from 18 to 36 years, weighing 50 to 100 kilograms and with heights of 1.5 to 1.8 meters. All the subjects, who were suffering neck pain, provided written informed consent to participate in the experiments, with their personal identifying information removed prior to the research being conducted. The data was obtained from multiple medical centers over a two-year period using 3 Tesla MR scanners (Siemens, Skyra, Erlangen, Germany) with different MRI protocols. Table 5.1 shows the dataset’s demographics information and MRI scanner’s parameters. Examples of the very large anatomical variabilities are shown in Figure 5.1, with many more similar and different scenarios found in the neck dataset. Although the images were initially different in their spatial resolutions, intensity ranges and levels of contrast, they were later normalized. Axial T1-weighted spin echo images with sizes of 256×256 were obtained from between each participant’s occiput and T1 vertebral levels (45 slices, each with a repetition time of 746-1140 ms, echo time of 15-16 ms, slice thickness of 4 mm, inter-slice gap of 4 mm, flip angle of 70° , pixel bandwidth of 300 Hz/pixel and percent phase field of view of 100). As the images were collected using different MRI protocols, the inhomogeneity of their MRI intensities was corrected individually by varying the parameters of the multiplicative intrinsic component optimization (MICO) method [345]. Five blinded and independent students from the ANU’s medical school, with extensive training on cervical spine anatomy, manually segmented

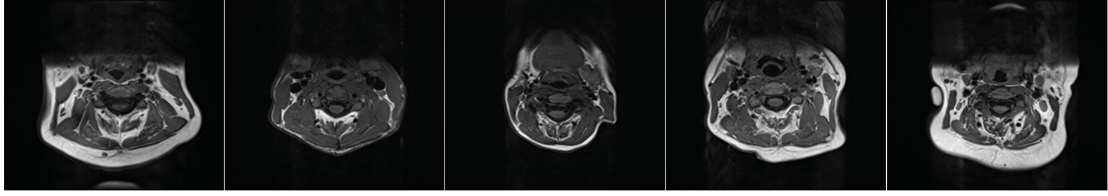


Figure 5.1: Anatomical variations in the neck dataset (images of different individuals)

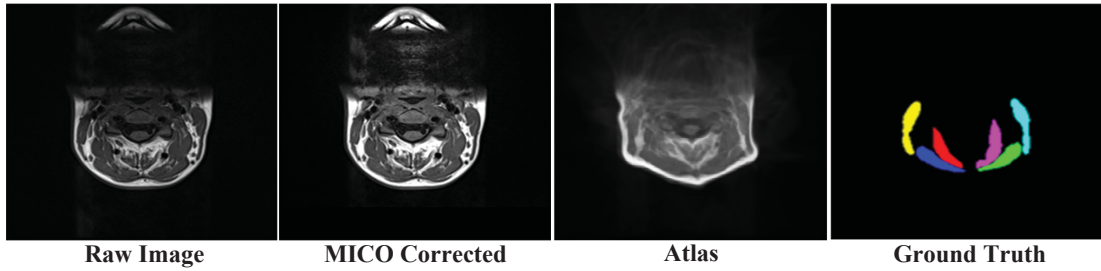


Figure 5.2: Axial MRIs at the top of the C3 spinal level showing inhomogeneity correction, population average and ground truth

each image's left and right muscles (sternocleidomastoid, semispinalis capitis and splenius capitis) at the top of the C3 vertebral level using a Matlab graphical user interface (GUI) specially constructed for this study. Later, their annotations were validated by two other anatomical experts from the medical school. The segmentation masks using six colors for the six muscles (yellow (M1), cyan (M2), red (M3), magenta (M4), blue (M5) and green (M6)) of each patient are shown in the last column in Figure 5.2. The inhomogeneity-corrected and ground-truth images were aligned with a population average using an affine transformation and edge position difference (EPD) similarity measure-based registration technique [308]. The image of the population average was generated using a registration algorithm, as described in Section 5.2.2.1. Figure 5.2 shows a raw MR image, its MICO-corrected version, the population average for this study and the ground-truth segmentation for one subject.

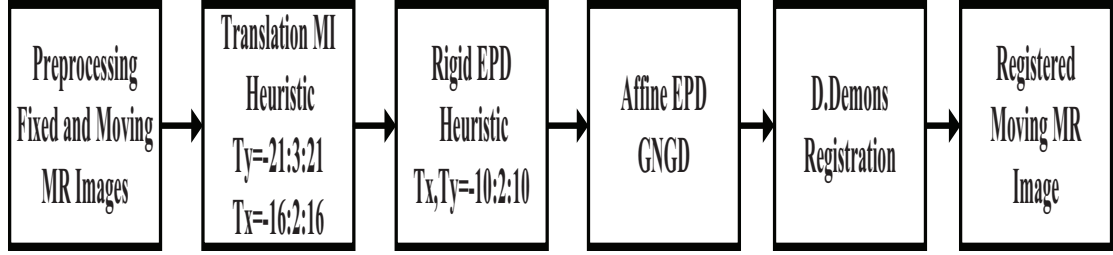


Figure 5.3: Registration framework for generating population average

5.2.2 Calculation of Population Average

In order to train a CNN efficiently, there should be few variations among the training images. Therefore, the training images are required to align with a reference one (also called a population average or atlas). It is the average of all the aligned training ones with respect to an initial reference one selected from them, based on criteria which should be common to the maximum number of images in the training dataset, and is calculated using a registration framework.

Some candidate images are selected as the initial reference image based on a rough visual inspection of their common criteria. The shapes of a neck's trunk, upper black shade and muscles as well as information on their positions relative to the horizontal and vertical axes and depths of their subcutaneous fat are considered as the criteria for observations. These criteria are: round, elliptical and angular shapes for a neck's trunk and high, low and medium for the depths of the subcutaneous fat. Then, a score is recorded for each criterion by visually inspecting all the training images. Finally, the scores of all the candidate images are compared and the initial reference one selected from those with the highest scores.

5.2.2.1 Registration Model

Once the initial reference image is selected, a registration model is required to align all the training images with the initial reference one. Designing this model for a neck dataset is very difficult because it contains very large anatomical variabilities, as can be seen in Figure 5.1, and the deformations necessary for the images to align are very complex and large; in particular, variations in the depth of the subcutaneous fat, irregular shapes of the neck's trunk and the top black region are the main hindrances.

To tackle the challenge of large variabilities, a multi-stage registration framework was developed, as shown in Figure 5.3. This framework is similar to the registration method proposed in chapter 4 except that the last stage is absent. This last stage requires the ground which is not possible in this scenario. The images registered from it are then co-registered using a different non-linear registration method called diffeomorphic demons (D.Demons) [2] in which heuristic, linear and non-linear registration stages are used. The heuristic stages, which tackle very large deformations as traditional continuous or discrete optimization-based registration algorithms, are limited to these deformations [48]. Heuristic or greedy optimization approaches can be found in [54, 58]. However, in this framework, which is intuitive and gradient-free, a heuristic approach is implemented using different transformations and SMs. The MI-based stage corrects coarse mismatches but cannot correct fine ones due to its statistical instability in local regions whereas the EPD-based stage does so as it is feature-based and performs well in local regions. The MI heuristic registration is used before the EPD one, with the former good for matching globally and the latter locally, with translational and rigid transformations used. Different types of transformations are helpful for avoiding local minima during the registration process while the affine registration,

which is used to combat global differences among subjects, is adopted from the paper in [308]. Finally, the D.Demons non-linear registration handles complex deformations resulting from the algorithm [2].

5.2.3 DCNNs for Evaluation

Deep learning-based segmentation methods have recently been shown to achieve more promising results than traditional techniques in medical image segmentation applications; in particular, the U-Net [6] architecture is a popular approach due to its good performance. Many researchers have tried to improve its performance by integrating additional techniques into its original architecture. Some recent prominent U-Net-based networks, including the U-Net [6], U-Net with a conditional random field (CRF-Unet) [72], attention U-Net (A-Unet) [7], nested U-Net or U-Net++ [73], multi-feature pyramid (MFP)-Unet [74] and recurrent residual U-Net (R2Unet) [75] are evaluated through experiments. Also, networks with more comprehensive modifications, including the multi-scale U-Net (MS-Unet), parallel multi-scale U-Net (PMS-Unet), recurrent residual attention U-Net (R2A-Unet) and R2A-Unet++, are evaluated for neck muscles segmentation.

The A-Unet is included in this evaluation study because its attention gate can automatically increase the prediction accuracy and sensitivity of any model by concentrating on target structures and neglecting irrelevant regions. It demonstrates better accuracy than the U-Net for two 3D CT abdominal datasets [7] while the CRF-Unet uses CRFs as recurrent neural networks [346]. A CRF establishes spatial constraints among labels to reduce false labeling which occurs due to local minima in training and image noise, and increases segmentation accuracy in natural images and medical images [59, 72]. The MFP-Unet uses a feature pyramid to extract feature maps from all the blocks of an expanding path in a semantic

layer for a segmentation procedure instead of the last block as does the U-Net. The MFP-Unet performs excellently for echocardiographic segmentation, better than the U-Net, U-Net++, deeplabv3, an anatomically constrained neural network (ACNN) and a stacked hourglass (SHG) network [74]. The R2Unet ensures better feature representation than the U-Net with the same number of parameters and yields good results compared with those of the SegNet and U-Net for segmenting retinal blood vessels, skin cancers and lung lesions [75]. The U-Net++ decreases the semantic gap between the feature maps of the encoder and decoder of the U-Net through dense nested skip connections. It provides better segmentation accuracy than the U-Net and wide U-Net for 3D CTs of chest nodules, nuclei microscopic images, liver CTs and polyp colonoscopy videos [73].

A multi-scale strategy is integrated with the U-Net in a serial and parallel way because it has shown good performances for anatomical variabilities and the dataset contains neck MRIs with muscles of various sizes and shapes [78]. It encodes both local and global contexts, textures and shapes. In the MS-Unet, it is implemented in a U-Net architecture using kernels of different sizes in different blocks of contracting and expanding paths. Another popular way of implementing a multi-scale technique is in the multi-stream architecture [78, 347, 348] used in the PMS-Unet to enable the processing of a large context through different resolutions of input images without increasing the receptive fields, thereby reducing the memory and computational requirements of DCNNs. Three pathways with different kernel sizes are decoupled and merged in the contraction path for the first 3 convolutional blocks of the U-Net and replicated in the expanding path for the 7th and 8th ones of the PMS-Unet. Also, this multi-stream concept is extended for the two different networks in the R2A-Unet++ in which the R2A-Unet and U-Net++ are considered different pathways and merged in the output

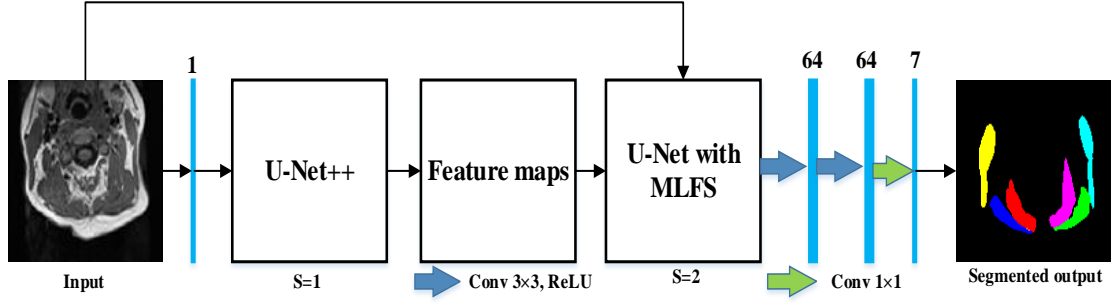


Figure 5.4: The two-stage U-Net++ (TS-UNet++) architecture with a multi-level fusion structure (MLFS) used in stage 2

layer. Furthermore, an attention gate is integrated with the R2Unet in the R2A-UNet to boost the prediction performance.

The networks are trained using a combination of the loss function of cross-entropy and the dice coefficient in a multi-data training technique [72] in which both a gradient magnitude and images are used, that had shown improved performances. Also, batch normalization and ReLU activation are applied in each layer after convolution.

5.2.4 TS-UNet++

Figure 5.4 shows the proposed TS-UNet++ architecture. It has two stages rather than multiple ones like the MS-UNet [4], which makes it memory-efficient. It uses the U-Net++ instead of the U-Net in its first stage and, at each layer, batch normalization and ReLU activation are applied after convolution. The architecture is trained by the loss function of the combination of a dice coefficient and cross-entropy in an end-to-end learning manner. Gradient-magnitude images of corresponding images with data augmentation are used during training to improve the learning efficiency [72] as they help the architecture to extract salient features quickly.

An additional convolutional block is inserted in the architecture between the U-Net with a multi-level fusion structure (MLFS) and output layer. This block can be considered as better for extracting information about the high-level semantics of an output feature map. This also makes the network deeper and wider without accelerating the problem of the gradient vanishing as in the traditional U-Net in which it is proportional to the number of convolutional blocks. This is possible due to the two stages in the architecture, whereby the network learns more representative features and performs better. There are two reasons for the vanishing of the gradient not increasing. Firstly, both stages of the network use the input images directly and, secondly, the network re-uses the feature maps from the first stage through fusing them with those from the second via concatenation. Although this fusion does not incur information loss, if the feature maps are repeatedly convolved in the same encoding path with deeper layers, the vanishing-gradient problem becomes acute.

A multi-stage-based CNN method has some advantages. Firstly, it can be considered to have many learners, each of which learns different additional features whereas their fused output may not be obtained by a single learner. Secondly, it can provide a coarse-to-fine approach for segmentation. Thirdly, it has the capability to boost training data by learning from images as well as the results of previous stages. Finally, it can control high-level semantic information obtained from multiple stages for effective segmentation. In the proposed TS-UNet++, two different types of learners with different capabilities, rather than the same as those in a traditional multi-stage method [4, 340], are used. This approach is adopted because, although the same types of learners may extract different additional features in multiple stages, these features may be almost similar or only subtly different. In contrast, differences may be large if different types of learners are used in a multi-stage technique and result in a better performance.

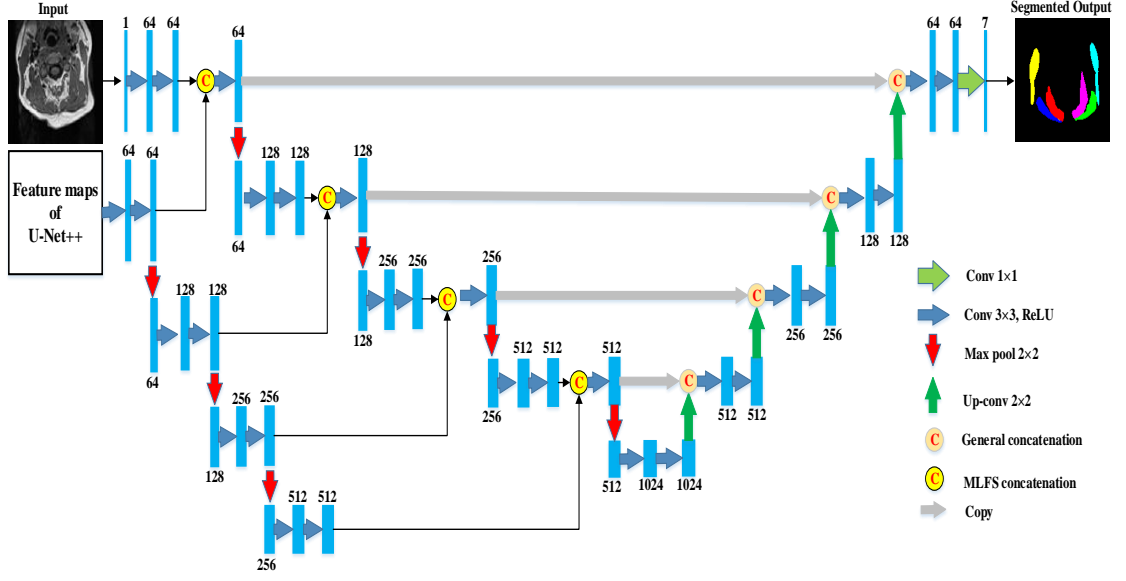


Figure 5.5: Multi-level fusion structure (MLFS) integrated with U-Net (feature information encoded from U-Net++ and U-Net in multi-scale fashion)

The proposed method uses the U-Net++ in the first stage and U-net in the second. It is not possible to use the opposite order of learners or have both of them as U-Net++ due to the architectural complexity of the U-Net++ which can yield a significantly better performance than the U-Net and provide good distinctive additional features and high-level semantic information. The reason for this is the fusing of the semantically similar feature maps of the encoder and decoder. Actually, the U-Net++ network reduces the semantic gap between these feature maps through using nested and dense skip connections which gradually enrich the encoder's high-resolution feature maps to rich ones of the decoder. As a result, the network can learn fine-grained details of foreground objects more effectively and simplify the learning task.

The MLFS is implemented in the proposed method through concatenation, rather than addition as occurs in a traditional multi-stage method [4], as it does not incur any information loss which enables deep supervision [59]. It also helps

to propagate the gradient easily which simplifies the training and combines multi-level features hierarchically, as shown in Fig. 5.5. To deepen the proposed model, another convolutional layer is added after each MLFS concatenation which does not worsen the gradient-vanishing problem because it obtains information from two sources via its concatenation while the TS-UNet++ can learn more representative features. However, the number of convolutional layers added should be limited according to the number of information sources in the concatenation.

5.2.5 Dilated-ROI Attention

The proposed dilated-ROI attention method is simple and effective for increasing the accuracy of semantic segmentation. In this method, any DCNN for semantic segmentation can be used as a backbone network while the training images are cropped for the ROI using dilated ground truths generated by a morphological dilation operation that exploits a square structuring element. Selecting the dilation pixels (DPs) (the width of the square) is an important aspect of this method since any DCNN for semantic segmentation is sensitive to the input features. If a DCNN is trained using a full input image, its unnecessary features generate some undesired values during testing which act as noise during the prediction stage and, consequently, the softmax layer produces false semantic labels. Therefore, the ROI selected encompasses all the boundaries of the desired objects and avoids extra regions by keeping an appropriate margin of distance.

Fig. 5.6 shows the types of images which can be used to train a semantic DCNN, with the middle and right-hand ones generated by 15 and 20 DPs, respectively. To obtain better semantic accuracy, the middle image is preferred because it includes all the desired muscles with decent margins of distance. On the other hand, the right-hand one covers some additional regions with extra features which

produce some scattered values that result in a false probability distribution in the output image.

Previous attention-based models used a bounding box or region-level coarse annotation to select the ROI, neither of which can select the regions required for the dilated-ROI process because: firstly, the muscles are not fully aligned either horizontally or vertically; secondly, the shape of the bounding box can be either square or rectangular which cannot optimally encompass the muscles due to their blobbed shapes; and thirdly, if a region-level coarse annotation is not performed carefully, excessive regions will be enclosed. Therefore, these methods may not produce better semantic segmentation than the dilated-ROI attention method.

Also, the testing images are cropped to have similar margins to those of the training ones in order to maintain symmetry. As this matches the network's training pattern, the performance of the network will improve. In this case, as any novice person can undertake this task with a small period of training by an anatomical expert, coarse annotation will be achieved with a minimal amount of labor and in less time because detailed pixel-wise annotation is not required, thereby reducing the annotation's cost and time. This part could also be conducted using a registration technique in which the contours of the desired objects can be easily transferred from an atlas of a visually similar person or through exploiting the output from another semantic DCNN.

The loss function, batch normalization, activation and multi-data technique for the models' training are performed in a similar manner to those in sections 5.2.3 and 5.2.4 except that the dataset's images are not aligned with the average of the population.

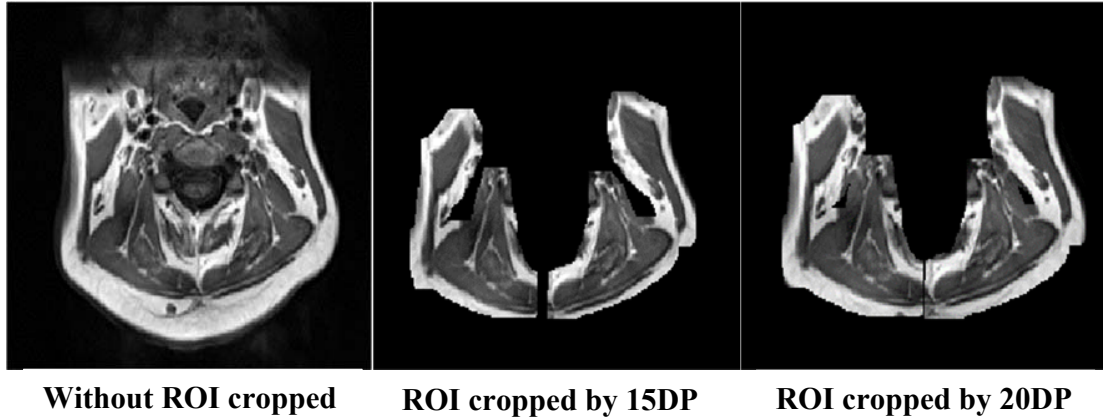


Figure 5.6: Types of images in the training dataset (Dilation pixels (DPs) indicate the widths of the structuring elements for the morphological dilation operation)

5.2.6 Implementation

The CNNs were trained on a NVIDIA GeForce GTX Titan Linux GPU (12 GB) using TensorFlow and Keras, and an Adam optimizer with a 0.001 learning rate. The images were divided into 30 training, 5 validation and 10 testing ones with almost gender equality. The number of epochs and batch size were set to 1500 and 5, respectively, during training. Online data augmentation was performed through an affine transformation (rotation, translation, shearing and scaling) and the nearest fill mode for the image and mask. The models were initialized by drawing weights from a 0-centered truncated normal distribution.

5.3 Results and Analyses

The deep learning based semantic segmentation methods in this Chapter were developed for 2D images. On the other hand, the proposed registration algorithms in Chapter 3 and 4 were developed for 3D images. Therefore, the registration methods cannot be compared directly with the segmentation methods.

5.3.1 Evaluations

The DSC and DHD metrics [349] were used to quantitatively analyze the segmentation results. The DHD metric was used because the DSC ignores the positions of the pixels which are important in segmentation. Table 5.2 shows the average DSC and DHD values achieved by the models and it can be seen that the U-Net++ yielded the best results in terms of the DSC with an overall mean of 0.89. Although the A-Unet obtained better results than the U-Net for two 3D CT abdominal datasets, it achieved a lower overall mean DSC than most of the models. On the other hand, the A-Unet and R2Unet models produced better overall mean DHDs, each with a value of 1.89. It is also noticeable that the PMS-Unet provided relatively good performances in terms of both metrics with overall DSC and DHD mean values of 0.88 and 1.90, respectively. Considering both evaluation measures when comparing performances, it can be said that the R2Unet outperformed the others. Of the muscles, M1 obtained the best results from both metrics for almost all the models, possibly because of its larger size.

To further analyze the A-Unet, R2Unet, U-Net++ and PMS-Unet models, the visual segmentation results for some test patients are shown in Fig. 5.7. It can be seen that the R2Unet was better than the other models and, although there were some scattered labels and holes for the M1 muscle (cyan), all the other muscles' boundaries were quite smooth except for a few holes and additional false labeling.

5.3.2 TS-UNet++

The TS-UNet++ model was compared with state-of-the-art approaches such as the U-Net [6], U-Net++ [5] and two-stage U-Net (TS-UNet) [4]. The TS-UNet

Table 5.2: Average DSC and DHD values for models with best results in the ‘Overall’ column in bold

	Networks	M1	M2	M3	M4	M5	M6	Overall
DSC	U-Net	0.94±0.03	0.86±0.03	0.91±0.05	0.85±0.07	0.86±0.04	0.86±0.06	0.88±0.06
	CRF-Unet	0.94±0.04	0.86±0.04	0.91±0.04	0.85±0.05	0.84±0.04	0.85±0.06	0.88±0.06
	A-Unet	0.93±0.03	0.86±0.04	0.91±0.05	0.85±0.06	0.84±0.06	0.85±0.06	0.87±0.06
	MFP-Unet	0.93±0.03	0.86±0.05	0.90±0.05	0.85±0.06	0.84±0.05	0.87±0.04	0.88±0.06
	R2Unet	0.92±0.03	0.86±0.04	0.93±0.03	0.85±0.08	0.86±0.04	0.85±0.07	0.88±0.06
	U-Net++	0.94±0.03	0.86±0.04	0.93±0.03	0.86±0.06	0.86±0.05	0.86±0.06	0.89±0.06
	R2A-Unet	0.94±0.03	0.85±0.06	0.92±0.04	0.84±0.07	0.86±0.06	0.85±0.06	0.88±0.07
	R2A-Unet++	0.92±0.04	0.86±0.05	0.90±0.04	0.85±0.06	0.87±0.04	0.86±0.06	0.88±0.05
	PMS-Unet	0.94±0.03	0.86±0.03	0.92±0.03	0.85±0.06	0.86±0.05	0.85±0.08	0.88±0.06
	MS-Unet	0.94±0.03	0.85±0.04	0.90±0.04	0.84±0.07	0.84±0.06	0.85±0.06	0.87±0.07
DHD (mm)	U-Net	1.49±0.26	1.93±0.27	1.60±0.59	2.26±0.52	2.21±0.37	2.23±0.37	1.95±0.52
	CRF-Unet	1.49±0.26	1.96±0.26	1.72±0.36	2.36±0.46	2.26±0.28	2.33±0.26	2.02±0.46
	A-Unet	1.56±0.41	1.83±0.26	1.56±0.60	2.13±0.43	2.04±0.36	2.20±0.42	1.89±0.50
	MFP-Unet	1.51±0.15	1.83±0.26	1.64±0.55	2.36±0.50	2.38±0.39	2.18±0.38	1.98±0.52
	R2Unet	1.67±0.32	1.76±0.33	1.51±0.15	2.10±0.52	2.09±0.38	2.20±0.38	1.89±0.45
	U-Net++	1.46±0.27	1.89±0.38	1.54±0.47	2.25±0.57	2.08±0.27	2.37±0.40	1.93±0.53
	R2A-Unet	1.63±0.19	2.03±0.41	1.50±0.50	2.27±0.51	2.17±0.31	2.24±0.32	1.97±0.49
	R2A-Unet++	1.70±0.63	2.11±0.35	1.56±0.24	2.19±0.35	2.28±0.34	2.38±0.34	2.04±0.50
	PMS-Unet	1.39±0.28	1.91±0.26	1.59±0.42	2.20±0.39	2.19±0.32	2.13±0.28	1.90±0.45
	MS-Unet	1.42±0.30	2.07±0.35	1.62±0.52	2.28±0.54	2.32±0.37	2.19±0.35	1.98±0.54

is a network in which the U-Net is used in both the first and second stages and no convolutional block is added before the output layer. Table 5.3 shows the average DSC and DHD values for the test images obtained by the comparative methods. The TS-UNet++ performed better than the others in terms of both metrics and would improve if more stages were incorporated. It can be conjectured that the convolutional layers added in a multi-stage manner with concatenation of the feature maps in a multi-level feature space can capture more high-level and abstract features for multi-class dense prediction problems.

In Fig. 5.8, the visual results for the five images randomly selected from the test dataset are displayed. The predictions more finely match the contours of the ground truths for the TS-UNet++ than for the other methods. Although there are some small under- and over-segmentations for the TS-UNet++, there are no holes as there are for the TS-UNet [4] for P4, U-Net++ [5] for P5 and U-Net [6] for

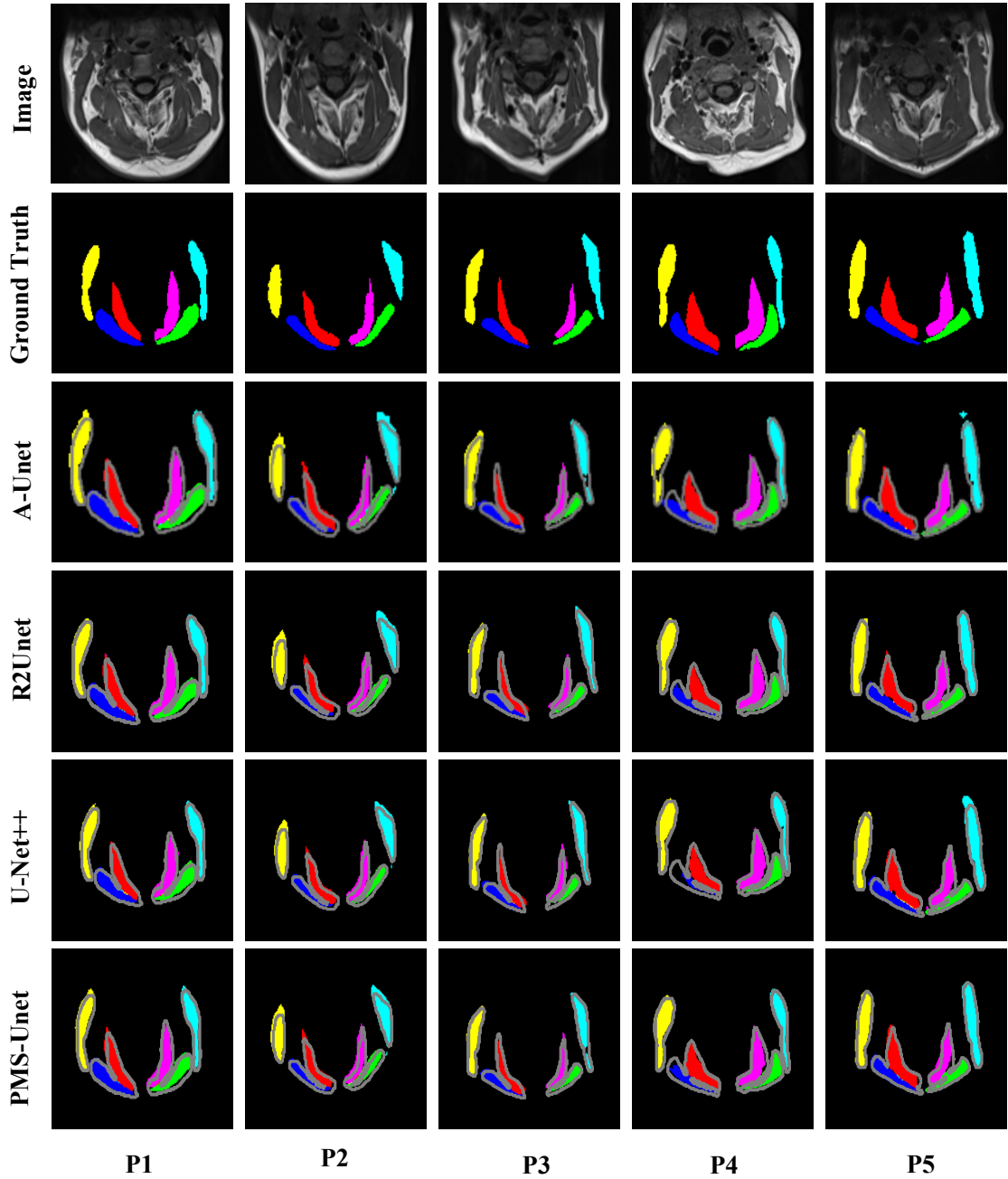


Figure 5.7: Visual segmentation results for MRIs and ground truths for A-Unet, R2Unet, U-Net++ and PMS-Unet models (columns represent different patients denoted as P1, P2, P3, P4 and P5), with extracted contours from ground truths indicated by gray lines superimposed on the automatic segmentation results

P4 and P5. Moreover, there are no large over-segmentations for the TS-UNet++ as for the TS-UNet [4] for P3 and P4, U-Net++ [5] for P1 and U-Net [6] for P2 and P3.

Table 5.3: Comparison of performances of two-stage U-Net++ (TS-U-Net++) and baseline CNNs with best results in the ‘Overall’ column in bold

	Models	M1	M2	M3	M4	M5	M6	Overall
DSC	TS-U-Net++	0.94±0.03	0.87±0.03	0.94±0.03	0.86±0.08	0.87±0.06	0.87±0.05	0.8911±0.06
	TS-U-Net [4]	0.94±0.03	0.85±0.04	0.92±0.03	0.85±0.06	0.86±0.05	0.86±0.06	0.8806±0.06
	U-Net++ [5]	0.94±0.03	0.86±0.04	0.93±0.03	0.86±0.06	0.86±0.05	0.86±0.06	0.8863±0.06
	U-Net [6]	0.94±0.03	0.86±0.03	0.91±0.05	0.85±0.07	0.86±0.04	0.86±0.06	0.8797±0.06
DHD (mm)	TS-U-Net++	1.55±0.31	1.92±0.35	1.27±0.29	2.15±0.50	2.17±0.33	2.08±0.39	1.8574±0.50
	TS-U-Net [4]	1.56±0.40	1.98±0.31	1.60±0.57	2.34±0.55	2.28±0.32	2.31±0.38	2.0134±0.54
	U-Net++ [5]	1.46±0.27	1.89±0.39	1.54±0.47	2.25±0.57	2.08±0.27	2.37±0.40	1.9321±0.53
	U-Net [6]	1.49±0.26	1.93±0.27	1.60±0.59	2.26±0.52	2.21±0.37	2.23±0.37	1.9534±0.52

5.3.3 Dilated-ROI Attention

To understand the effect of the dilated-ROI attention mechanism, the U-net [6] DCNN model was evaluated on the neck MRI dataset using four different cases: without attention; with dilated-ROI attention by 15 DPs; with dilated-ROI attention by 20 DPs; and with automatic attention (A-U-Net). Table 5.4 shows the average DSC and DHD values for the test images obtained from the comparative methods. The U-Net with dilated-ROI attention by 15 DPs performed the best of all the others in terms of both metrics. Although the A-U-Net was superior on the abdominal CT dataset for pancreas segmentation, it was worse than the without attention mechanism which may have been due to the complexity involved in the neck dataset or the attention model causing incorrect attention. Also, the area of the ROI affected the accuracy as the performance of the U-Net with dilated-ROI attention by 20 DPs was worse than that of the U-Net with dilated-ROI attention by 15 DPs.

In Fig. 5.9, the visual results for the five images randomly selected from the test dataset are presented. It can be seen that the predicted labels for the U-Net with dilated-ROI attention by 15 DPs are more closely matched than the others.

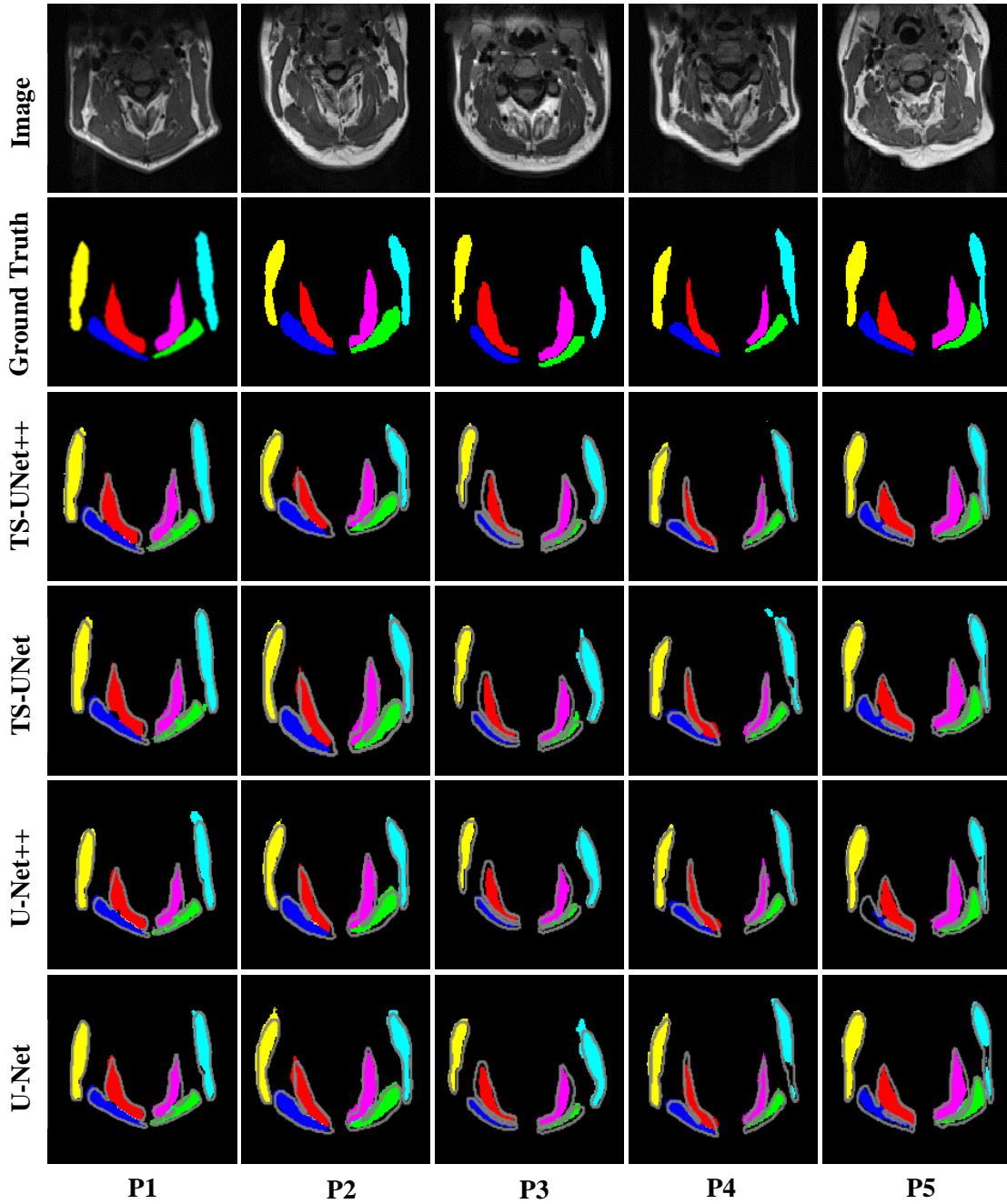


Figure 5.8: Visual results for multi-class neck muscles segmentations of five different patients (P1, P2, P3, P4 and P5) (first row original images, second corresponding manual segmentations and following automatic segmentations of two-stage U-Net++ (TS-UNet++), two-stage U-Net (TS-UNet) [4], U-Net++ [5] and U-Net [6], with contours extracted from ground truths denoted by gray lines superimposed on automatic segmentation results)

5.4 Discussion and Conclusion

5.4.1 Evaluations

In this study, 10 different versions of the U-Net deep learning segmentation method

Table 5.4: Average DSC and DHD values for U-Net without attention [6], with dilated-ROI attention by 15 DPs, dilated-ROI attention by 20 DPs and automatic attention (A-Unet) [7], with best results in the ‘Overall’ column in bold

	Models	DP	M1	M2	M3	M4	M5	M6	Overall
DSC	U-Net [6]	-	0.94±0.04	0.88±0.04	0.93±0.04	0.86±0.06	0.88±0.03	0.87±0.05	0.8954±0.05
	U-Net	15	0.95±0.04	0.90±0.04	0.94±0.03	0.90±0.04	0.85±0.05	0.89±0.05	0.9056±0.06
	U-Net	20	0.95±0.04	0.90±0.03	0.94±0.05	0.89±0.05	0.87±0.05	0.88±0.06	0.9028±0.05
	A-Unet [7]	-	0.95±0.04	0.87±0.04	0.92±0.05	0.86±0.06	0.88±0.04	0.86±0.06	0.8900±0.06
DHD (mm)	U-Net [6]	-	1.46±0.27	1.85±0.26	1.32±0.23	2.18±0.50	2.25±0.23	2.30±0.31	1.8936±0.50
	U-Net	15	1.38±0.30	1.89±0.33	1.42±0.28	2.10±0.42	2.27±0.37	2.04±0.37	1.8518±0.49
	U-Net	20	1.42±0.28	1.89±0.36	1.41±0.45	2.22±0.43	2.26±0.32	2.19±0.35	1.8974±0.52
	A-Unet [7]	-	1.44±0.34	1.98±0.42	1.47±0.21	2.22±0.41	2.21±0.32	2.22±0.38	1.9254±0.49

manually delineated neck muscles. The modified models were based on recent popular techniques for machine learning, such as multi-stream and multi-scale approaches. This type of study is unprecedented in the deep learning-based segmentation literature and, although its scale is small, it will be beneficial for the deep learning-based medical image segmentation community.

The main finding from this study was that the performances of the models were quite similar except for the R2Unet which achieved slightly better accuracy in terms of both metrics. Although it has been reported in the literature that the performances of some of the models were better on other medical datasets, that was not the case for the neck dataset; in particular, the A-Unet’s performance was worse than that of the U-Net in terms of the DSC but was superior when using CT data for pancreatic segmentation. This might have been for two reasons; firstly, the neck MRI data were more complex due to the compactness, similar compositions and small sizes of the muscles, specifically, the M3-M6 muscles which had adjoining boundaries whereas the pancreas was comparatively large and not as compact; and, secondly, the same hyper-parameters were used for all the models which may not have been optimal for the A-Unet. However, all the other models may also have performed better if the hyper-parameters were optimized. Therefore, there is no recommendation to use any particular model at this stage as further

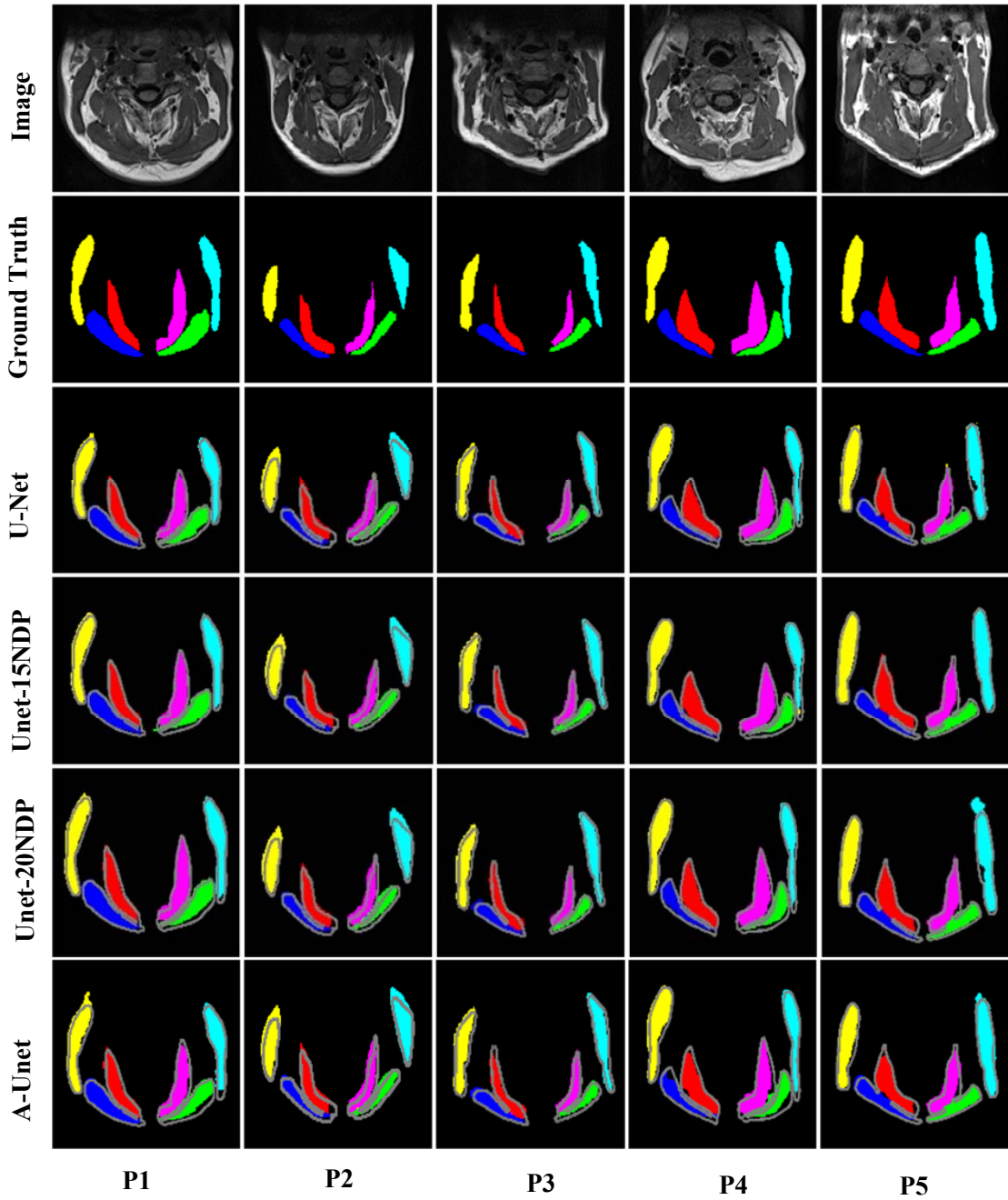


Figure 5.9: Visual segmentation results obtained from the dilated-ROI attention experiments for five different patients (P1, P2, P3, P4 and P5) (first row original images, second corresponding manual segmentations and following rows automatic segmentations of U-Net without attention, with dilated-ROI attention by 15 DPs, with dilated-ROI attention by 20 DPs and the A-Unet, respectively, with contours extracted from ground truths denoted by gray lines superimposed on the automatic segmentation results)

research regarding their hyper-parameters and other issues needs to be conducted. In future, some additional networks with more data for other spinal levels will be analyzed using more evaluation measures and a diverse dataset for 3D learning.

5.4.2 TS-UNet++

A two-stage DCNN using the U-Net and U-Net++ in a cascaded manner with deep supervision and end-to-end learning was proposed. In this method, convolutional layers were added to extract more high-level features which avoided the gradient-vanishing problem of the traditional DCNN due to the fusing of feature maps that combined multiple sources of information through concatenation in a multi-level space. The experimental evaluation showed that the TS-UNet++ yielded state-of-the-art results for neck muscles segmentation. In future, this model will be evaluated using public and large-scale datasets with many other medical segmentation applications.

5.4.3 Dilated-ROI Learning

A simple and effective dilated-ROI-based attention mechanism is proposed. As DCNNs for semantic segmentation are sensitive to learning features in terms of feeding image regions, any automatic attention-based approach is not suitable for all applications. In future, this method will be made fully automatic through using either a registration technique or another DCNN output to select the ROI for testing images.

Chapter 6

Conclusions and Future Directions

This chapter presents the major findings and contributions of this research in Section 6.1 as well as an outline of potential extensions of it in Section 6.2.

6.1 Conclusions

In this thesis, new 3D-3D non-rigid image registration and deep learning-based semantic segmentation methods for supporting analyses of neck diseases in order to apply proper medical interventions were presented. This study will assist medical doctors in diagnosing these diseases by helping them understand the sizes and shapes of neck muscles. These methods can be applied for image-guided surgery and therapeutic processes as well as determinations of the proper doses of toxin injections. The important findings and developments are summarized below.

- In Chapter 3, a novel object-constrained hierarchical registration framework for aligning inter-subject neck muscles is proposed. Firstly, to handle large-scale local minima, a coarse registration technique, which optimizes the new EPD similarity measure, is used to align large mismatches. Also, a new transformation based on the DPSW, affine and FFD transformations are

exploited. Secondly, to avoid the monotony of using the same transformations in multiple stages, a fine registration technique for aligning small mismatches, which uses a double-pushing system by changing edges in the EPD and switching the transformation's resolutions, is designed. The EPD helps both the coarse and fine techniques to implement object-constrained registration via controlling edges which is not possible using traditional similarity measures. Also, a modified 3D chamfer distance transformation algorithm is developed for the EPD. Experiments are performed on clinical 3D MRI scans of the neck, with the results showing that the EPD is more effective than the MI and SSD measures in terms of the volumetric DSC. The proposed method is compared with two state-of-the-art approaches with ablation studies of inter-subject deformable registration and achieves better accuracy, robustness and consistency.

- However, the above framework is computationally complex and has difficulty handling large-scale anatomical variabilities. To overcome these problems, another 3D-3D registration framework with two novel contributions is proposed in Chapter 4. Firstly, a two-stage heuristic search optimization technique for handling large mismatches, which uses a minimal user hypothesis regarding these mismatches and is computationally fast, is introduced. It brings a moving image hierarchically closer to a fixed one using MI and EPD similarity measures in the coarse and fine stages, respectively, while the images do not require pre-segmentation as is necessary in traditional heuristic optimization-based techniques. Secondly, a ROI EPD-based registration framework for handling small mismatches using salient anatomical information, in which a convex objective function is formed through a unique shape created from the desired objects in the ROI, is proposed. It is compared with

two state-of-the-art methods on the neck dataset, with the results showing that it is superior in terms of accuracy and is computationally fast.

- In Chapter 5, an evaluation study of recent U-Net-based CNNs is performed on the neck dataset. It comprises 6 recent models, the U-Net, CRF-Unet, A-Unet, U-Net++, MFP-Unet and R2Unet, and 4 networks with more comprehensive modifications, the MS-Unet, PMS-Unet, R2A-Unet and R2A-Unet++, for neck muscles segmentation, with analyses of the numerical results indicating that the R2Unet architecture achieves the best accuracy. Also, two deep learning-based semantic segmentation approaches are proposed. In the first, a new TS-Unet++ algorithm uses two different types of deep CNNs (DCNNs) rather than one, as is the case for the traditional multi-stage approaches, with the U-Net++ in the first stage and the U-Net in the second. More convolutional blocks are added after the input and before the output layers of this multi-stage method to better extract the low- and high-level features. A new concatenation-based fusion structure, which is incorporated in the architecture to allow deep supervision, helps to increase the depth of the network without accelerating the vanishing gradient problem of a traditional CNN. More convolutional layers are added after each concatenation of the fusion structure to extract more representative features. This network is compared with the U-Net, U-Net++ and TS-Unet on the neck dataset, with the results indicating that it outperforms the others. In the second approach, an explicit attention method, in which the attention is performed through a ROI evolved from the ground truth via dilation, is proposed. It does not require any additional CNN, as does a cascaded approach, to localize the ROI. Attention in a CNN is sensitive with respect to the area of the ROI. This dilated ROI is more capable of capturing relevant regions and suppressing irrelevant ones than a bounding box and region-level coarse

annotation and is used during the training of any CNN. Coarse annotation, which does not require any detailed pixel wise delineation and can be performed by a novice person, is used during testing. This ROI-based attention method can handle compact and similar small multiple classes with objects with large variabilities. It is compared with the A-Unet and U-Net, and performs the best.

6.2 Future Works

Although the proposed registration and segmentation methods obtain good accuracy, it is still possible to improve their levels of accuracy, robustness and computational complexity for which the following aspects could be further investigated.

- The framework proposed in Chapter 3 could be validated using some public brain and lung datasets using a more theoretical insight into DPSW. Also, some other optimization methods, such as stochastic ones, could be explored and other evaluation metrics applied. Furthermore, the size of the neck dataset could be increased for experimental analyses.
- A fine resolution of the FFD transformation in Chapter 4 could be examined using local gradient descent optimization for the ROI-based EPD algorithm. Also, regularization could be imposed inside the objective function to confirm the smoothness of the transformation.
- Other spinal levels than those in Chapter 5 could be analyzed to validate the effectiveness of the proposed methods. The 2D models could be transformed to corresponding 3D ones for which it would be necessary to generate 3D ground truths. The explicit ROI-based attention model could be made fully

automatic using either a registration technique or another DCNN output to select the ROI for testing images.

- In general, the registration methods could be implemented in C++ with a GPU to speed up their execution times. Also, other neck muscles could be investigated and the inhomogeneity of the MRI corrected to obtain better performances of the proposed methods.
- An investigation could be conducted to translate the proposed methods to the clinical domain, with the developed models integrated in commercialized software. The registration methods could be used in radiation therapy applications in which a patient could be immobilized in an almost identical position during the pre-operative and treatment imaging stages.

Bibliography

- [1] G. Borgefors, “Distance transformations in arbitrary dimensions,” *Computer vision, graphics, and image processing*, vol. 27, no. 3, pp. 321–345, 1984.
- [2] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, “Diffeomorphic demons: Efficient non-parametric image registration,” *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.
- [3] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, “Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain,” *Medical image analysis*, vol. 12, no. 1, pp. 26–41, 2008.
- [4] Y. Tang, F. Yang, S. Yuan *et al.*, “A multi-stage framework with context information fusion structure for skin lesion segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE, 2019, pp. 1407–1410.
- [5] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: Redesigning skip connections to exploit multiscale features in image segmentation,” *IEEE Transactions on Medical Imaging*, 2019.
- [6] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, 2015, pp. 234–241.
- [7] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, “Attention gated networks: Learning to leverage salient regions in medical images,” *Medical image analysis*, vol. 53, pp. 197–207, 2019.

- [8] D. S. Owers, D. M. Perriman, P. N. Smith, T. Neeman, and A. L. Webb, “Evidence for cervical muscle morphometric changes on magnetic resonance images after whiplash: A systematic review and meta-analysis,” *Injury*, vol. 49, no. 2, pp. 165–176, 2018.
- [9] J. M. Elliott, A. R. Pedler, G. A. Jull, L. Van Wyk, G. G. Galloway, and S. P. O’Leary, “Differential changes in muscle composition exist in traumatic and nontraumatic neck pain,” *Spine*, vol. 39, no. 1, pp. 39–47, 2014.
- [10] R. J. Cunningham, P. J. Harding, and I. D. Loram, “Real-time ultrasound segmentation, analysis and visualisation of deep cervical muscle structure,” *IEEE transactions on medical imaging*, vol. 36, no. 2, pp. 653–665, 2017.
- [11] Y. Hagström and J. Carlsson, “Prolonged functional impairments after whiplash injury.” *Scandinavian journal of rehabilitation medicine*, vol. 28, no. 3, pp. 139–146, 1996.
- [12] M. W. Devereaux, “Neck pain,” *Primary Care: Clinics in Office Practice*, vol. 31, pp. 19–31, 2004.
- [13] W. F. Stewart, J. A. Ricci, E. Chee, D. Morganstein, and R. Lipton, “Lost productive time and cost due to common pain conditions in the us workforce,” *Jama*, vol. 290, no. 18, pp. 2443–2454, 2003.
- [14] “Warning over whiplash ‘epidemic’,” [Online]. Available: <http://news.bbc.co.uk/2/hi/health/7729336.stm>.
- [15] “Acute whiplash,” [Online]. Available: <http://www.maa.nsw.gov.au/for-professionals/for-health-professionals/acute-whiplash>.
- [16] M. Sterling, J. Hendrikz, and J. Kenardy, “Developmental trajectories of pain/disability and ptsd symptoms after whiplash injury,” *Pain*, vol. 150, pp. 22–28, 2010.

- [17] J. Elliott, G. Jull, J. T. Noteboom, and G. Galloway, "MRI study of the cross-sectional area for the cervical extensor musculature in patients with persistent whiplash associated disorders (WAD)," *Manual therapy*, vol. 13, no. 3, pp. 258–265, 2008.
- [18] V. C. Hoe, D. M. Urquhart, H. L. Kelsall, and M. R. Sim, "Ergonomic design and training for preventing work-related musculoskeletal disorders of the upper limb and neck in adults," *Cochrane database of systematic reviews*, no. 8, 2012.
- [19] E. Smith, D. G. Hoy, M. Cross, T. Vos, M. Naghavi, R. Buchbinder, A. D. Woolf, and L. March, "The global burden of low back pain: estimates from the global burden of disease 2010 study," *ANNALS OF THE RHEUMATIC DISEASES*.
- [20] Q. Bismil and M. Bismil, "Myofascial-entheseal dysfunction in chronic whiplash injury: an observational study," *JRSM short reports*, vol. 3, no. 8, p. 57, 2012.
- [21] M. E. Vidt, A. C. Santago II, C. J. Tuohy, G. G. Poehling, M. T. Freehill, R. A. Kraft, A. P. Marsh, E. J. Hegedus, M. E. Miller, and K. R. Saul, "Assessments of fatty infiltration and muscle atrophy from a single magnetic resonance image slice are not predictive of 3-dimensional measurements," *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, vol. 32, no. 1, pp. 128–139, 2016.
- [22] D. W. McRobbie, E. A. Moore, M. J. Graves, and M. R. Prince, "Mri from protons to pictures," 2006.

- [23] S. Andrews and G. Hamarneh, “The generalized log-ratio transformation: Learning shape and adjacency priors for simultaneous thigh muscle segmentation,” *IEEE Transactions on Medical Imaging*, vol. 34, no. 9, pp. 1773–1787, Sept 2015.
- [24] A. Eresen, S. M. Birch, L. Alic, J. F. Griffin IV, J. N. Kornegay, and J. X. Ji, “New similarity metric for registration of mri to histology: Golden retriever muscular dystrophy imaging,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 5, pp. 1222–1230, 2018.
- [25] Z. Tang, P.-T. Yap, and D. Shen, “A new image similarity metric for improving deformation consistency in graph based groupwise image registration,” *IEEE Transactions on Biomedical Engineering*, 2018.
- [26] J. Lotz, J. Olesch, B. Müller, T. Polzin, P. Galuschka, J. Lotz, S. Heldmann, H. Laue, M. González-Vallinas, A. Warth *et al.*, “Patch-based nonlinear image registration for gigapixel whole slide images,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1812–1819, 2016.
- [27] X. Cao, J. Yang, J. Zhang, Q. Wang, P.-T. Yap, and D. Shen, “Deformable image registration using a cue-aware deep regression network,” *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1900–1911, 2018.
- [28] A. A. Goshtasby, *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.
- [29] S. Reaungamornrat *et al.*, “Deformable image registration for surgical guidance using intraoperative cone-beam ct,” Ph.D. dissertation, Johns Hopkins University, 2017.

- [30] M. Holden, “A review of geometric transformations for nonrigid body registration,” *IEEE transactions on medical imaging*, vol. 27, no. 1, pp. 111–128, 2008.
- [31] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, “Deep learning techniques for medical image segmentation: Achievements and challenges,” *Journal of digital imaging*, vol. 32, no. 4, pp. 582–596, 2019.
- [32] L. K. Kamibayashi and F. J. Richmond, “Morphometry of human neck muscles,” *Spine*, vol. 23, no. 12, pp. 1314–1323, 1998.
- [33] J. Au, D. M. Perriman, M. R. Pickering, G. Buirski, P. N. Smith, and A. L. Webb, “Magnetic resonance imaging atlas of the cervical spine musculature,” *Clinical Anatomy*, vol. 29, no. 5, pp. 643–659, 2016.
- [34] R. Hua, J. M. Pozo, Z. A. Taylor, and A. F. Frangi, “Multiresolution extended free-form deformations (XFFD) for non-rigid registration with discontinuous transforms,” *Medical image analysis*, vol. 36, pp. 113–122, 2017.
- [35] A. Pai, S. Sommer, L. Sørensen, S. Darkner, J. Sporring, and M. Nielsen, “Kernel bundle diffeomorphic image registration using stationary velocity fields and wendland basis functions,” *IEEE transactions on medical imaging*, vol. 35, no. 6, pp. 1369–1380, 2016.
- [36] Y. Sun, J. Yuan, W. Qiu, M. Rajchl, C. Romagnoli, and A. Fenster, “Three-dimensional nonrigid MR-TRUS registration using dual optimization,” *IEEE transactions on medical imaging*, vol. 34, no. 5, pp. 1085–1095, 2015.
- [37] W. Sun, D. H. Poot, I. Smal, X. Yang, W. J. Niessen, and S. Klein, “Stochastic optimization with randomized smoothing for image registration,” *Medical image analysis*, vol. 35, pp. 146–158, 2017.

- [38] L. Gong, C. Zhang, L. Duan, X. Du, H. Liu, X. Chen, and J. Zheng, “Nonrigid image registration using spatially region-weighted correlation ratio and gpu-acceleration,” *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 766–778, 2018.
- [39] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, “A nonrigid registration framework using spatially encoded mutual information and free-form deformations,” *IEEE transactions on medical imaging*, vol. 30, no. 10, pp. 1819–1828, 2011.
- [40] H. Rivaz, Z. Karimaghloo, and D. L. Collins, “Self-similarity weighted mutual information: a new nonrigid image registration metric,” *Medical image analysis*, vol. 18, no. 2, pp. 343–358, 2014.
- [41] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, “Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change,” *IEEE transactions on Medical Imaging*, vol. 25, no. 5, pp. 626–639, 2006.
- [42] S. Klein, U. A. Van Der Heide, I. M. Lips, M. Van Vulpen, M. Staring, and J. P. Pluim, “Automatic segmentation of the prostate in 3d mr images by atlas matching using localized mutual information,” *Medical physics*, vol. 35, no. 4, pp. 1407–1417, 2008.
- [43] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [44] D. G. Lowe, “Object recognition from local scale-invariant features,” in *Proceedings of the seventh IEEE international conference on computer vision*, vol. 2, 1999, pp. 1150–1157.

- [45] H. Lombaert, L. Grady, X. Pennec, N. Ayache, and F. Chriet, “Spectral log-demons: diffeomorphic image registration with very large deformations,” *International journal of computer vision*, vol. 107, no. 3, pp. 254–271, 2014.
- [46] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, “Computing large deformation metric mappings via geodesic flows of diffeomorphisms,” *International journal of computer vision*, vol. 61, no. 2, pp. 139–157, 2005.
- [47] T. Brox and J. Malik, “Large displacement optical flow: descriptor matching in variational motion estimation,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 3, pp. 500–513, 2010.
- [48] A. Sotiras, C. Davatzikos, and N. Paragios, “Deformable medical image registration: A survey,” *IEEE transactions on medical imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.
- [49] D. Shen and C. Davatzikos, “Hammer: hierarchical attribute matching mechanism for elastic registration,” *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.
- [50] G. Wu, F. Qi, and D. Shen, “Learning-based deformable registration of mr brain images,” *IEEE transactions on medical imaging*, vol. 25, no. 9, pp. 1145–1157, 2006.
- [51] T. Liu, D. Shen, and C. Davatzikos, “Deformable registration of cortical structures via hybrid volumetric and surface warping,” *NeuroImage*, vol. 22, no. 4, pp. 1790–1801, 2004.
- [52] Z. Xue, D. Shen, and C. Davatzikos, “Determining correspondence in 3-d mr brain images using attribute vectors as morphological signatures of voxels,” *IEEE transactions on medical imaging*, vol. 23, no. 10, pp. 1276–1291, 2004.

- [53] D. Shen, “Image registration by local histogram matching,” *Pattern Recognition*, vol. 40, no. 4, pp. 1161–1172, 2007.
- [54] J. Zhang, M. Liu, L. An, Y. Gao, and D. Shen, “Alzheimer’s disease diagnosis using landmark-based features from longitudinal structural mr images,” *IEEE journal of biomedical and health informatics*, vol. 21, no. 6, pp. 1607–1616, 2017.
- [55] G. Wu, M. Kim, Q. Wang, and D. Shen, “S-hammer: Hierarchical attribute-guided, symmetric diffeomorphic registration for mr brain images,” *Human brain mapping*, vol. 35, no. 3, pp. 1044–1060, 2014.
- [56] G. Wu, M. Kim, Q. Wang, B. C. Munsell, and D. Shen, “Scalable high-performance image registration framework by unsupervised deep feature representations learning,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 7, pp. 1505–1516, 2016.
- [57] Q. Wang, G. Wu, P.-T. Yap, and D. Shen, “Attribute vector guided group-wise registration,” *NeuroImage*, vol. 50, no. 4, pp. 1485–1496, 2010.
- [58] Y. Wu, G. Wu, L. Wang, B. C. Munsell, Q. Wang, W. Lin, Q. Feng, W. Chen, and D. Shen, “Hierarchical and symmetric infant image registration by robust longitudinal-example-guided correspondence detection,” *Medical physics*, vol. 42, no. 7, pp. 4174–4189, 2015.
- [59] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, “DRINet for medical image segmentation,” *IEEE transactions on medical imaging*, vol. 37, no. 11, pp. 2453–2462, 2018.
- [60] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, “Ce-net: Context encoder network for 2d medical image segmentation,” *IEEE transactions on medical imaging*, 2019.

- [61] H. R. Roth, L. Lu, N. Lay, A. P. Harrison, A. Farag, A. Sohn, and R. M. Summers, “Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation,” *Medical image analysis*, vol. 45, pp. 94–107, 2018.
- [62] C. Payer, D. Štern, H. Bischof, and M. Urschler, “Multi-label whole heart segmentation using cnns and anatomical label configurations,” in *International Workshop on Statistical Atlases and Computational Models of the Heart*, 2017, pp. 190–198.
- [63] M. Khened, V. A. Kollerathu, and G. Krishnamurthi, “Fully convolutional multi-scale residual densenets for cardiac segmentation and automated cardiac diagnosis using ensemble of classifiers,” *Medical image analysis*, vol. 51, pp. 21–45, 2019.
- [64] Y.-J. Huang, Q. Dou, Z.-X. Wang, L.-Z. Liu, Y. Jin, C.-F. Li, L. Wang, H. Chen, and R.-H. Xu, “3D RoI-aware U-net for accurate and efficient colorectal tumor segmentation,” *arXiv preprint arXiv:1806.10342*, 2018.
- [65] T. A. Ngo, Z. Lu, and G. Carneiro, “Combining deep learning and level set for the automated segmentation of the left ventricle of the heart from cardiac cine magnetic resonance,” *Medical image analysis*, vol. 35, pp. 159–171, 2017.
- [66] F. Liao, M. Liang, Z. Li, X. Hu, and S. Song, “Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network,” *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3484–3495, 2019.
- [67] B. Xu, J. Liu, X. Hou, B. Liu, J. Garibaldi, I. O. Ellis, A. Green, L. Shen, and G. Qiu, “Attention by selection: A deep selective attention approach to

- breast cancer classification,” *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1930–1941, 2019.
- [68] W. Wang, C. Ye, S. Zhang, Y. Xu, and K. Wang, “Improving whole-heart ct image segmentation by attention mechanism,” *IEEE Access*, vol. 8, pp. 14 579–14 587, 2019.
- [69] A. Sinha and J. Dolz, “Multi-scale self-guided attention for medical image segmentation,” *IEEE Journal of Biomedical and Health Informatics*, 2020.
- [70] Y. Lei, Y. Tian, H. Shan, J. Zhang, G. Wang, and M. K. Kalra, “Shape and margin-aware lung nodule classification in low-dose ct images via soft activation mapping,” *Medical Image Analysis*, vol. 60, p. 101628, 2020.
- [71] H. Yang, J.-Y. Kim, H. Kim, and S. P. Adhikari, “Guided soft attention network for classification of breast cancer histopathology images,” *IEEE transactions on medical imaging*, vol. 39, no. 5, pp. 1306–1315, 2019.
- [72] W. Xia, M. Fortin, J. Ahn, H. Rivaz, M. C. Battié, T. M. Peters, and Y. Xiao, “Automatic paraspinal muscle segmentation in patients with lumbar pathology using deep convolutional neural network,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2019, pp. 318–325.
- [73] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, “Unet++: A nested u-net architecture for medical image segmentation,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Springer, 2018, pp. 3–11.

- [74] S. Moradi, M. G. Oghli, A. Alizadehasl, I. Shiri, N. Oveisi, M. Oveisi, M. Maleki, and J. Dhooze, “MFP-Unet: A novel deep learning based approach for left ventricle segmentation in echocardiography,” *Physica Medica*, vol. 67, pp. 58–69, 2019.
- [75] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, “Recurrent residual u-net for medical image segmentation,” *Journal of Medical Imaging*, vol. 6, no. 1, p. 014006, 2019.
- [76] S. K. Zhou, D. Rueckert, and G. Fichtinger, *Handbook of medical image computing and computer assisted intervention*. Academic Press, 2019.
- [77] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [78] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, “Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation,” *Medical image analysis*, vol. 36, pp. 61–78, 2017.
- [79] Y.-T. Wu, T. Kanade, C.-C. Li, and J. Cohn, “Image registration using wavelet-based motion model,” *International Journal of Computer Vision*, vol. 38, no. 2, pp. 129–152, 2000.
- [80] L. Zagorchev and A. Goshtasby, “A comparative study of transformation functions for nonrigid image registration,” *IEEE transactions on image processing*, vol. 15, no. 3, pp. 529–538, 2006.

- [81] K. Rohr, M. Fornefett, and H. S. Stiehl, "Spline-based elastic image registration: integration of landmark errors and orientation attributes," *Computer Vision and Image Understanding*, vol. 90, no. 2, pp. 153–168, 2003.
- [82] K. Rohr, H. S. Stiehl, R. Sprengel, T. M. Buzug, J. Weese, and M. Kuhn, "Landmark-based elastic registration using approximating thin-plate splines," *IEEE Transactions on medical imaging*, vol. 20, no. 6, pp. 526–534, 2001.
- [83] K. Rohr, H. S. Stiehl, R. Sprengel, W. Beil, T. M. Buzug, J. Weese, and M. Kuhn, "Point-based elastic registration of medical image data using approximating thin-plate splines," in *International Conference on Visualization in Biomedical Computing*, 1996, pp. 297–306.
- [84] M. Fornefett, K. Rohr, and H. S. Stiehl, "Radial basis functions with compact support for elastic registration of medical images," *Image and vision computing*, vol. 19, no. 1-2, pp. 87–96, 2001.
- [85] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 11, no. 6, pp. 567–585, 1989.
- [86] —, "Thin-plate splines and the atlas problem for biomedical images," in *Biennial International Conference on Information Processing in Medical Imaging*, 1991, pp. 326–342.
- [87] J. Li, X. Yang, and J. Yu, "Compact support thin plate spline algorithm," *Journal of Electronics (China)*, vol. 24, no. 4, pp. 515–522, 2007.
- [88] H. J. Johnson and G. E. Christensen, "Landmark and intensity-based, consistent thin-plate spline image registration," in *Biennial International Conference on Information Processing in Medical Imaging*, 2001, pp. 329–343.

- [89] K. Rohr and S. Wörz, “An extension of thin-plate splines for image registration with radial basis functions,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*. IEEE, 2012, pp. 442–445.
- [90] G. Donato and S. Belongie, “Approximate thin plate spline mappings,” in *European conference on computer vision*, 2002, pp. 21–31.
- [91] D. Ruprecht, R. Nagel, and H. Müller, “Spatial free-form deformation with scattered data interpolation methods,” *Computers & Graphics*, vol. 19, no. 1, pp. 63–71, 1995.
- [92] J. A. Little, D. L. Hill, and D. J. Hawkes, “Deformations incorporating rigid structures,” *Computer Vision and Image Understanding*, vol. 66, no. 2, pp. 223–232, 1997.
- [93] H. Wendland, “Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree,” *Advances in computational Mathematics*, vol. 4, no. 1, pp. 389–396, 1995.
- [94] —, “Error estimates for interpolation by compactly supported radial basis functions of minimal degree,” *Journal of approximation theory*, vol. 93, no. 2, pp. 258–272, 1998.
- [95] Z. Wu, “Compactly supported positive definite radial functions,” *Advances in Computational Mathematics*, vol. 4, no. 1, p. 283, 1995.
- [96] M. D. Buhmann, “Radial basis functions,” *Acta numerica*, vol. 9, pp. 1–38, 2000.
- [97] M. H. Davis, A. Khotanzad, D. P. Flamig, and S. E. Harms, “A physics-based coordinate transformation for 3-d image matching,” *IEEE transactions on medical imaging*, vol. 16, no. 3, pp. 317–328, 1997.

- [98] J. Kohlrausch, K. Rohr, and H. S. Stiehl, “A new class of elastic body splines for nonrigid registration of medical images,” *Journal of Mathematical Imaging and Vision*, vol. 23, no. 3, pp. 253–280, 2005.
- [99] S. Wörz and K. Rohr, “Physics-based elastic registration using non-radial basis functions and including landmark localization uncertainties,” *Computer Vision and Image Understanding*, vol. 111, no. 3, pp. 263–274, 2008.
- [100] —, “Physics-based elastic image registration using splines and including landmark localization uncertainties,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 678–685.
- [101] T. W. Sederberg and S. R. Parry, “Free-form deformation of solid geometric models,” in *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 1986, pp. 151–160.
- [102] W. M. Hsu, J. F. Hughes, and H. Kaufman, “Direct manipulation of free-form deformations,” *ACM Siggraph Computer Graphics*, vol. 26, no. 2, pp. 177–184, 1992.
- [103] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, “Nonrigid registration using free-form deformations: application to breast MR images,” *IEEE transactions on medical imaging*, vol. 18, no. 8, pp. 712–721, 1999.
- [104] J. Declerck, J. Feldmar, M. L. Goris, and F. Betting, “Automatic registration and alignment on a template of cardiac stress and rest reoriented spect images,” *IEEE transactions on medical imaging*, vol. 16, no. 6, pp. 727–737, 1997.
- [105] J. Kybic and M. Unser, “Fast parametric elastic image registration,” *IEEE transactions on image processing*, vol. 12, no. 11, pp. 1427–1442, 2003.

- [106] M. Sdika, “A fast nonrigid image registration with constraints on the jacobian using large scale constrained optimization,” *IEEE transactions on medical imaging*, vol. 27, no. 2, pp. 271–281, 2008.
- [107] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit *et al.*, “A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 573–581.
- [108] J. Wang and T. Jiang, “Nonrigid registration of brain mri using nurbs,” *Pattern Recognition Letters*, vol. 28, no. 2, pp. 214–223, 2007.
- [109] W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K.-P. Tung, P. Edwards, and D. Rueckert, “Registration using sparse free-form deformations,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2012, pp. 659–666.
- [110] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach, “Symmetric nonrigid image registration: application to average brain templates construction,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2008, pp. 897–904.
- [111] W. Feng, S. J. Reeves, T. S. Denney, S. Lloyd, L. Dell’Italia, and H. Gupta, “A new consistent image registration formulation with a b-spline deformation model,” in *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*. IEEE, 2009, pp. 979–982.
- [112] A. Sotiras and N. Paragios, “Discrete symmetric image registration,” in *2012 9th IEEE International Symposium on Biomedical Imaging (ISBI)*, 2012, pp. 342–345.

- [113] K. K. Bhatia, J. V. Hajnal, B. K. Puri, A. D. Edwards, and D. Rueckert, “Consistent groupwise non-rigid registration for atlas construction,” in *2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821)*, 2004, pp. 908–911.
- [114] S. K. Balci, P. Golland, M. E. Shenton, and W. M. Wells, “Free-form b-spline deformation model for groupwise registration,” *Med Image Comput Comput Assist Interv*, 2007.
- [115] A. Sotiras, Y. Ou, B. Glocker, C. Davatzikos, and N. Paragios, “Simultaneous geometric-iconic registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 676–683.
- [116] C. T. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen, “Non-rigid registration of dynamic medical imaging data using nd+ t b-splines and a groupwise optimization approach,” *Medical image analysis*, vol. 15, no. 2, pp. 238–249, 2011.
- [117] D. Perperidis, R. H. Mohiaddin, and D. Rueckert, “Spatio-temporal free-form registration of cardiac mr image sequences,” *Medical image analysis*, vol. 9, no. 5, pp. 441–456, 2005.
- [118] M. J. Ledesma-Carbayo, J. Kybic, M. Desco, A. Santos, M. Suhling, P. Hunziker, and M. Unser, “Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation,” *IEEE transactions on medical imaging*, vol. 24, no. 9, pp. 1113–1126, 2005.
- [119] J. Vandemeulebroucke, S. Rit, J. Kybic, P. Clarysse, and D. Sarrut, “Spatiotemporal motion estimation for respiratory-correlated imaging of the lungs,” *Medical physics*, vol. 38, no. 1, pp. 166–178, 2011.

- [120] W. Sun, W. J. Niessen, and S. Klein, “Randomly perturbed b-splines for nonrigid image registration,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 7, pp. 1401–1413, 2017.
- [121] W. Sun, W. J. Niessen, M. van Stralen, and S. Klein, “Simultaneous multiresolution strategies for nonrigid image registration,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4905–4917, 2013.
- [122] G. E. Christensen and H. J. Johnson, “Consistent image registration,” *IEEE transactions on medical imaging*, vol. 20, no. 7, pp. 568–582, 2001.
- [123] J. Ashburner and K. J. Friston, “Nonlinear spatial normalization using basis functions,” *Human brain mapping*, vol. 7, no. 4, pp. 254–266, 1999.
- [124] Y. Amit, “A nonlinear variational problem for image matching,” *SIAM Journal on Scientific Computing*, vol. 15, no. 1, pp. 207–224, 1994.
- [125] S. Gefen, O. Tretiak, and J. Nissanov, “Elastic 3-D alignment of rat brain histological images,” *IEEE transactions on medical imaging*, vol. 22, no. 11, pp. 1480–1489, 2003.
- [126] O. Musse, F. Heitz, and J.-P. Armspach, “Topology preserving deformable image matching using constrained hierarchical parametric models,” *IEEE transactions on image processing*, vol. 10, no. 7, pp. 1081–1093, 2001.
- [127] V. Noblet, C. Heinrich, F. Heitz, and J.-P. Armspach, “3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization,” *IEEE Transactions on image processing*, vol. 14, no. 5, pp. 553–566, 2005.
- [128] —, “Retrospective evaluation of a topology preserving non-rigid registration method,” *Medical image analysis*, vol. 10, no. 3, pp. 366–384, 2006.

- [129] P. Hellier, C. Barillot, E. Mémin, and P. Pérez, “Hierarchical estimation of a dense deformation field for 3-d robust registration,” *IEEE transactions on medical imaging*, vol. 20, no. 5, pp. 388–402, 2001.
- [130] H. Zhang, P. A. Yushkevich, D. C. Alexander, and J. C. Gee, “Deformable registration of diffusion tensor mr images with explicit orientation optimization,” *Medical image analysis*, vol. 10, no. 5, pp. 764–785, 2006.
- [131] A. Pitiot, E. Bardinet, P. M. Thompson, and G. Malandain, “Piecewise affine registration of biological images for volume reconstruction,” *Medical image analysis*, vol. 10, no. 3, pp. 465–483, 2006.
- [132] O. Commowick, V. Arsigny, A. Isambert, J. Costa, F. Dhermain, F. Bidault, P.-Y. Bondiau, N. Ayache, and G. Malandain, “An efficient locally affine framework for the smooth registration of anatomical structures,” *Medical Image Analysis*, vol. 12, no. 4, pp. 427–441, 2008.
- [133] T. F. Cootes, C. J. Twining, V. S. Petrovic, K. O. Babalola, and C. J. Taylor, “Computing accurate correspondences across groups of images,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 11, pp. 1994–2005, 2009.
- [134] C. Buerger, T. Schaeffter, and A. P. King, “Hierarchical adaptive local affine registration for fast and robust respiratory motion estimation,” *Medical image analysis*, vol. 15, no. 4, pp. 551–564, 2011.
- [135] V. Arsigny, X. Pennec, and N. Ayache, “Polyrigid and polyaffine transformations: a novel geometrical tool to deal with non-rigid deformations—application to the registration of histological slices,” *Medical image analysis*, vol. 9, no. 6, pp. 507–523, 2005.

- [136] V. Arsigny, O. Commowick, N. Ayache, and X. Pennec, “A fast and log-euclidean polyaffine framework for locally linear registration,” *Journal of Mathematical Imaging and Vision*, vol. 33, no. 2, pp. 222–238, 2009.
- [137] J. C. Gee and R. K. Bajcsy, “Elastic matching: Continuum mechanical and probabilistic analysis,” *Brain warping*, vol. 2, pp. 183–197, 1998.
- [138] C. Davatzikos, “Spatial transformation and registration of brain images using elastically deformable models,” *Computer Vision and Image Understanding*, vol. 66, no. 2, pp. 207–222, 1997.
- [139] J. He and G. E. Christensen, “Large deformation inverse consistent elastic image registration,” in *Biennial International Conference on Information Processing in Medical Imaging*, 2003, pp. 438–449.
- [140] X. Pennec, R. Stefanescu, V. Arsigny, P. Fillard, and N. Ayache, “Riemannian elasticity: A statistical regularization framework for non-linear registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2005, pp. 943–950.
- [141] A. D. Leow, I. Yanovsky, M.-C. Chiang, A. D. Lee, A. D. Klunder, A. Lu, J. T. Becker, S. W. Davis, A. W. Toga, and P. M. Thompson, “Statistical properties of jacobian maps and the realization of unbiased large-deformation nonlinear image registration,” *IEEE transactions on medical imaging*, vol. 26, no. 6, pp. 822–832, 2007.
- [142] C. Le Guyader and L. A. Vese, “A combined segmentation and registration framework with a nonlinear elasticity smoother,” *Computer Vision and Image Understanding*, vol. 115, no. 12, pp. 1689–1709, 2011.

- [143] G. E. Christensen, R. D. Rabbitt, and M. I. Miller, “Deformable templates using large deformation kinematics,” *IEEE transactions on image processing*, vol. 5, no. 10, pp. 1435–1447, 1996.
- [144] G. E. Christensen, S. C. Joshi, and M. I. Miller, “Volumetric transformation of brain anatomy,” *IEEE transactions on medical imaging*, vol. 16, no. 6, pp. 864–877, 1997.
- [145] Y. Wang and L. H. Staib, “Physical model-based non-rigid registration incorporating statistical shape information,” *Medical image analysis*, vol. 4, no. 1, pp. 7–20, 2000.
- [146] E. D’agostino, F. Maes, D. Vandermeulen, and P. Suetens, “A viscous fluid model for multimodal non-rigid image registration using mutual information,” *Medical image analysis*, vol. 7, no. 4, pp. 565–575, 2003.
- [147] M.-C. Chiang, A. D. Leow, A. D. Klunder, R. A. Dutton, M. Barysheva, S. E. Rose, K. L. McMahon, G. I. De Zubicaray, A. W. Toga, and P. M. Thompson, “Fluid registration of diffusion tensor images using information theory,” *IEEE transactions on medical imaging*, vol. 27, no. 4, pp. 442–456, 2008.
- [148] J.-P. Thirion, “Image matching as a diffusion process: an analogy with maxwell’s demons,” *Medical image analysis*, vol. 2, no. 3, pp. 243–260, 1998.
- [149] B. Fischer and J. Modersitzki, “Fast diffusion registration,” *Contemporary Mathematics*, vol. 313, pp. 117–128, 2002.
- [150] X. Pennec, P. Cachier, and N. Ayache, “Understanding the “demon’s algorithm”: 3d non-rigid registration by gradient descent,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1999, pp. 597–605.

- [151] T. Vercauteren, X. Pennec, E. Malis, A. Perchant, and N. Ayache, “Insight into efficient image registration techniques and the demons algorithm,” in *Biennial International Conference on Information Processing in Medical Imaging*, 2007, pp. 495–506.
- [152] P. Cachier, E. Bardinet, D. Dormont, X. Pennec, and N. Ayache, “Iconic feature based nonrigid registration: the pasha algorithm,” *Computer vision and image understanding*, vol. 89, no. 2-3, pp. 272–298, 2003.
- [153] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, “Non-parametric diffeomorphic image registration with the demons algorithm,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2007, pp. 319–326.
- [154] —, “Symmetric log-domain diffeomorphic registration: A demons-based approach,” in *International conference on medical image computing and computer-assisted intervention*, 2008, pp. 754–761.
- [155] R. Stefanescu, X. Pennec, and N. Ayache, “Grid powered nonlinear image registration with locally adaptive regularization,” *Medical image analysis*, vol. 8, no. 3, pp. 325–342, 2004.
- [156] J.-M. Peyrat, H. Delingette, M. Sermesant, X. Pennec, C. Xu, and N. Ayache, “Registration of 4d time-series of cardiac images with multichannel diffeomorphic demons,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2008, pp. 972–979.
- [157] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, “Computing large deformation metric mappings via geodesic flows of diffeomorphisms,” *International journal of computer vision*, vol. 61, no. 2, pp. 139–157, 2005.

- [158] A. R. Khan, L. Wang, and M. F. Beg, “Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping,” *Neuroimage*, vol. 41, no. 3, pp. 735–746, 2008.
- [159] K. Oishi, A. Faria, H. Jiang, X. Li, K. Akhter, J. Zhang, J. T. Hsu, M. I. Miller, P. C. van Zijl, M. Albert *et al.*, “Atlas-based whole brain white matter analysis using large deformation diffeomorphic metric mapping: application to normal elderly and alzheimer’s disease participants,” *Neuroimage*, vol. 46, no. 2, pp. 486–499, 2009.
- [160] C. Ceritoglu, K. Oishi, X. Li, M.-C. Chou, L. Younes, M. Albert, C. Lyketsos, P. C. van Zijl, M. I. Miller, and S. Mori, “Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging,” *Neuroimage*, vol. 47, no. 2, pp. 618–627, 2009.
- [161] C. Ceritoglu, L. Wang, L. D. Selemon, J. G. Csernansky, M. I. Miller, and J. T. Ratnanather, “Large deformation diffeomorphic metric mapping registration of reconstructed 3d histological section images and in vivo mr images,” *Frontiers in human neuroscience*, vol. 4, p. 43, 2010.
- [162] M. F. Beg and A. Khan, “Symmetric data attachment terms for large deformation image registration,” *IEEE transactions on medical imaging*, vol. 26, no. 9, pp. 1179–1189, 2007.
- [163] J. Ashburner, “A fast diffeomorphic image registration algorithm,” *Neuroimage*, vol. 38, no. 1, pp. 95–113, 2007.
- [164] F. P. Oliveira and J. M. R. Tavares, “Medical image registration: a review,” *Computer methods in biomechanics and biomedical engineering*, vol. 17, no. 2, pp. 73–93, 2014.

- [165] M. N. Aktar, “Registration and fusion of multi-modal medical images,” Ph.D. dissertation, The University of New South Wales Australia, 2016.
- [166] R. A. McLaughlin, J. Hipwell, D. J. Hawkes, J. A. Noble, J. V. Byrne, and T. Cox, “A comparison of 2d-3d intensity-based registration and feature-based registration for neurointerventions,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 517–524.
- [167] B. Zitova and J. Flusser, “Image registration methods: a survey,” *Image and vision computing*, vol. 21, no. 11, pp. 977–1000, 2003.
- [168] C. G. Harris, M. Stephens *et al.*, “A combined corner and edge detector.” in *Alvey vision conference*, vol. 15, no. 50, 1988, pp. 10–5244.
- [169] J. Shi *et al.*, “Good features to track,” in *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, 1994, pp. 593–600.
- [170] C. Schmid, R. Mohr, and C. Bauckhage, “Evaluation of interest point detectors,” *International Journal of computer vision*, vol. 37, no. 2, pp. 151–172, 2000.
- [171] T. Lindeberg, “Feature detection with automatic scale selection,” *International journal of computer vision*, vol. 30, no. 2, pp. 79–116, 1998.
- [172] T. Kadir and M. Brady, “Saliency, scale and image description,” *International Journal of Computer Vision*, vol. 45, no. 2, pp. 83–105, 2001.
- [173] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

- [174] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [175] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (surf),” *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.
- [176] W. Cheung and G. Hamarneh, “ n -sift: n -dimensional scale invariant feature transform,” *IEEE Transactions on Image Processing*, vol. 18, no. 9, pp. 2012–2021, 2009.
- [177] X. Han, “Feature-constrained nonlinear registration of lung ct images,” *Medical image analysis for the clinic: a grand challenge*, pp. 63–72, 2010.
- [178] D. Ni, Y. Qu, X. Yang, Y. P. Chui, T.-T. Wong, S. S. Ho, and P. A. Heng, “Volumetric ultrasound panorama based on 3d sift,” in *International conference on medical image computing and computer-assisted intervention*, 2008, pp. 52–60.
- [179] S. Allaire, J. J. Kim, S. L. Breen, D. A. Jaffray, and V. Pekar, “Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis,” in *2008 IEEE computer society conference on computer vision and pattern recognition workshops*, 2008, pp. 1–8.
- [180] M. Niemeijer, M. K. Garvin, K. Lee, B. van Ginneken, M. D. Abràmoff, and M. Sonka, “Registration of 3d spectral oct volumes using 3d sift feature point matching,” in *Medical Imaging 2009: Image Processing*, vol. 7259, 2009, p. 72591I.

- [181] M. Leordeanu and M. Hebert, “A spectral technique for correspondence problems using pairwise constraints,” in *Tenth IEEE International Conference on Computer Vision (ICCV’05) Volume 1*, vol. 2, 2005, pp. 1482–1489.
- [182] O. Duchenne, F. Bach, I.-S. Kweon, and J. Ponce, “A tensor-based algorithm for high-order graph matching,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 12, pp. 2383–2395, 2011.
- [183] N. Duta, A. K. Jain, and M.-P. Dubuisson-Jolly, “Automatic construction of 2d shape models,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 5, pp. 433–446, 2001.
- [184] T. F. Cootes and C. J. Taylor, “A mixture model for representing shape variation,” *Image and Vision Computing*, vol. 17, no. 8, pp. 567–573, 1999.
- [185] H. Guo, A. Rangarajan, and S. Joshi, “Diffeomorphic point matching,” in *Handbook of mathematical models in computer vision*, 2006, pp. 205–219.
- [186] B. Jian and B. C. Vemuri, “Robust point set registration using gaussian mixture models,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 8, pp. 1633–1645, 2010.
- [187] A. Myronenko and X. Song, “Point set registration: Coherent point drift,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 12, pp. 2262–2275, 2010.
- [188] F. Wang, B. C. Vemuri, A. Rangarajan, and S. J. Eisenschenk, “Simultaneous nonrigid registration of multiple point sets and atlas construction,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, no. 11, pp. 2011–2022, 2008.

- [189] N. J. Tustison, S. P. Awate, G. Song, T. S. Cook, and J. C. Gee, "Point set registration using havrda-charvat-tsallis entropy measures," *IEEE transactions on medical imaging*, vol. 30, no. 2, pp. 451–460, 2010.
- [190] N. Paragios, M. Rousson, and V. Ramesh, "Non-rigid registration using distance functions," *Computer Vision and Image Understanding*, vol. 89, no. 2-3, pp. 142–165, 2003.
- [191] X. Huang, N. Paragios, and D. N. Metaxas, "Shape registration in implicit spaces using information theory and free form deformations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 8, pp. 1303–1318, 2006.
- [192] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. International Society for Optics and Photonics, 1992, pp. 586–606.
- [193] Y. Liu, "Improving icp with easy implementation for free-form surface matching," *Pattern Recognition*, vol. 37, no. 2, pp. 211–226, 2004.
- [194] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE transactions on medical imaging*, vol. 22, no. 11, pp. 1379–1394, 2003.
- [195] J. Kim and J. A. Fessler, "Intensity-based image registration using robust correlation coefficients," *IEEE transactions on medical imaging*, vol. 23, no. 11, pp. 1430–1444, 2004.
- [196] G. Hermosillo, C. Chefd'Hotel, and O. Faugeras, "Variational methods for multimodal image matching," *International Journal of Computer Vision*, vol. 50, no. 3, pp. 329–343, 2002.

- [197] D. Shen and C. Davatzikos, “Hammer: hierarchical attribute matching mechanism for elastic registration,” *IEEE transactions on medical imaging*, vol. 21, no. 11, pp. 1421–1439, 2002.
- [198] Z. Xue, D. Shen, and C. Davatzikos, “Determining correspondence in 3-d mr brain images using attribute vectors as morphological signatures of voxels,” *IEEE transactions on medical imaging*, vol. 23, no. 10, pp. 1276–1291, 2004.
- [199] D. Shen, “Image registration by local histogram matching,” *Pattern Recognition*, vol. 40, no. 4, pp. 1161–1172, 2007.
- [200] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, “Dramms: Deformable registration via attribute matching and mutual-saliency weighting,” *Medical image analysis*, vol. 15, no. 4, pp. 622–639, 2011.
- [201] S. Liao and A. C. Chung, “Feature based nonrigid brain mr image registration with symmetric alpha stable filters,” *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 106–119, 2009.
- [202] A. Myronenko and X. Song, “Intensity-based image registration by minimizing residual complexity,” *IEEE transactions on medical imaging*, vol. 29, no. 11, pp. 1882–1891, 2010.
- [203] J. P. Pluim, J. A. Maintz, and M. A. Viergever, “Mutual-information-based registration of medical images: a survey,” *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 986–1004, 2003.
- [204] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis, “Multi-modal volume registration by maximization of mutual information,” *Medical image analysis*, vol. 1, no. 1, pp. 35–51, 1996.

- [205] P. Viola and W. M. Wells III, "Alignment by maximization of mutual information," *International journal of computer vision*, vol. 24, no. 2, pp. 137–154, 1997.
- [206] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE transactions on Medical Imaging*, vol. 16, no. 2, pp. 187–198, 1997.
- [207] C. Studholme, D. L. Hill, and D. J. Hawkes, "An overlap invariant entropy measure of 3d medical image alignment," *Pattern recognition*, vol. 32, no. 1, pp. 71–86, 1999.
- [208] N. D. Cahill, J. A. Schnabel, J. A. Noble, and D. J. Hawkes, "Revisiting overlap invariance in medical image alignment," in *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1–8.
- [209] A. Roche, G. Malandain, X. Pennec, and N. Ayache, "The correlation ratio as a new similarity measure for multimodal image registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 1998, pp. 1115–1124.
- [210] A. C. Chung, W. M. Wells, A. Norbash, and W. E. L. Grimson, "Multimodal image registration by minimising kullback-leibler distance," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 525–532.
- [211] R. Liao, C. Guetter, C. Xu, Y. Sun, A. Khamene, and F. Sauer, "Learning-based 2d/3d rigid registration using jensen-shannon divergence for image-guided surgery," in *International Workshop on Medical Imaging and Virtual Reality*, 2006, pp. 228–235.

- [212] Y. He, A. B. Hamza, and H. Krim, "A generalized divergence measure for robust image registration," *IEEE Transactions on Signal Processing*, vol. 51, no. 5, pp. 1211–1220, 2003.
- [213] S. Martin and T. S. Durrani, "A new divergence measure for medical image registration," *IEEE transactions on image processing*, vol. 16, no. 4, pp. 957–966, 2007.
- [214] D. Rueckert, M. J. Clarkson, D. L. Hill, and D. J. Hawkes, "Non-rigid registration using higher-order mutual information," in *Medical Imaging 2000: Image Processing*, vol. 3979, 2000, pp. 438–447.
- [215] B. Karaçali, "Information theoretic deformable registration using local image information," *International journal of computer vision*, vol. 72, no. 3, pp. 219–237, 2007.
- [216] C. Studholme, C. Drapaca, B. Iordanova, and V. Cardenas, "Deformation-based mapping of volume change from serial brain mri in the presence of local tissue contrast change," *IEEE transactions on Medical Imaging*, vol. 25, no. 5, pp. 626–639, 2006.
- [217] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid image registration using conditional mutual information," *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 19–29, 2009.
- [218] X. Zhuang, S. Arridge, D. J. Hawkes, and S. Ourselin, "A nonrigid registration framework using spatially encoded mutual information and free-form deformations," *IEEE transactions on medical imaging*, vol. 30, no. 10, pp. 1819–1828, 2011.

- [219] M. Holden, L. D. Griffin, N. Saeed, and D. L. Hill, “Multi-channel mutual information using scale space,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2004, pp. 797–804.
- [220] C. Studholme, D. Hill, and D. Hawkes, “Incorporating connected region labelling into automated image registration using mutual information,” in *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*, 1996, pp. 23–31.
- [221] X. Cao, J. Yang, Y. Gao, Y. Guo, G. Wu, and D. Shen, “Dual-core steered non-rigid registration for multi-modal images via bi-directional image synthesis,” *Medical image analysis*, vol. 41, pp. 18–31, 2017.
- [222] X. Cao, J. Yang, Y. Gao, Q. Wang, and D. Shen, “Region-adaptive deformable registration of ct/mri pelvic images via learning-based image synthesis,” *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3500–3512, 2018.
- [223] A. Roche, X. Pennec, G. Malandain, and N. Ayache, “Rigid registration of 3-d ultrasound with mr images: a new approach combining intensity and gradient information,” *IEEE transactions on medical imaging*, vol. 20, no. 10, pp. 1038–1049, 2001.
- [224] W. Wein, S. Brunke, A. Khamene, M. R. Callstrom, and N. Navab, “Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention,” *Medical image analysis*, vol. 12, no. 5, pp. 577–585, 2008.
- [225] J. A. Maintz, P. A. van den Elsen, and M. A. Viergever, “3d multimodality medical image registration using morphological tools,” *Image and vision computing*, vol. 19, no. 1-2, pp. 53–62, 2001.

- [226] E. Haber and J. Modersitzki, “Intensity gradient based registration and fusion of multi-modal images,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2006, pp. 726–733.
- [227] T. Butz and J.-P. Thiran, “Affine registration with feature space mutual information,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 549–556.
- [228] D. Lee, M. Hofmann, F. Steinke, Y. Altun, N. D. Cahill, and B. Scholkopf, “Learning similarity measure for multi-modal 3d image registration,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 186–193.
- [229] H. J. Johnson and G. E. Christensen, “Consistent landmark and intensity-based image registration,” *IEEE transactions on medical imaging*, vol. 21, no. 5, pp. 450–461, 2002.
- [230] Y. Yin, E. A. Hoffman, K. Ding, J. M. Reinhardt, and C.-L. Lin, “A cubic b-spline-based hybrid registration of lung ct images for a dynamic airway geometric model with large deformation,” *Physics in Medicine & Biology*, vol. 56, no. 1, p. 203, 2010.
- [231] E. Gibson, A. R. Khan, and M. F. Beg, “A combined surface and volumetric registration (savor) framework to study cortical biomarkers and volumetric imaging data,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2009, pp. 713–720.
- [232] T. Liu, D. Shen, and C. Davatzikos, “Deformable registration of cortical structures via hybrid volumetric and surface warping,” *NeuroImage*, vol. 22, no. 4, pp. 1790–1801, 2004.

- [233] O. Camara, G. Delso, O. Colliot, A. Moreno-Ingelmo, and I. Bloch, “Explicit incorporation of prior anatomical information into a nonrigid registration of thoracic and abdominal ct and 18-fdg whole-body emission pet images,” *IEEE transactions on medical imaging*, vol. 26, no. 2, pp. 164–178, 2007.
- [234] X. Papademetris, A. P. Jackowski, R. T. Schultz, L. H. Staib, and J. S. Duncan, “Integrated intensity and point-feature nonrigid registration,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2004, pp. 763–770.
- [235] K. Rohr, P. Cathier, and S. Wörz, “Elastic registration of electrophoresis images using intensity information and point landmarks,” *Pattern recognition*, vol. 37, no. 5, pp. 1035–1048, 2004.
- [236] A. A. Joshi, D. W. Shattuck, P. M. Thompson, and R. M. Leahy, “Surface-constrained volumetric brain registration using harmonic mappings,” *IEEE transactions on medical imaging*, vol. 26, no. 12, pp. 1657–1669, 2007.
- [237] P. Cachier, J.-F. Mangin, X. Pennec, D. Rivière, D. Papadopoulos-Orfanos, J. Régis, and N. Ayache, “Multisubject non-rigid registration of brain mri using intensity and geometric features,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 734–742.
- [238] A. Joshi, R. Leahy, A. W. Toga, and D. Shattuck, “A framework for brain registration via simultaneous surface and volume flow,” in *International Conference on Information Processing in Medical Imaging*, 2009, pp. 576–588.
- [239] N. Honnorat, R. Vaillant, and N. Paragios, “Graph-based geometric-iconic guide-wire tracking,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2011, pp. 9–16.

- [240] V. Siless, P. Guevara, X. Pennec, and P. Fillard, “Joint t1 and brain fiber diffeomorphic registration using the demons,” in *International Workshop on Multimodal Brain Image Analysis*, 2011, pp. 10–18.
- [241] S. Durrleman, P. Fillard, X. Pennec, A. Trouvé, and N. Ayache, “Registration, atlas estimation and variability analysis of white matter fiber bundles modeled as currents,” *NeuroImage*, vol. 55, no. 3, pp. 1073–1090, 2011.
- [242] S. Klein, M. Staring, and J. P. Pluim, “Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines,” *IEEE transactions on image processing*, vol. 16, no. 12, pp. 2879–2890, 2007.
- [243] Y. Cao, M. I. Miller, R. L. Winslow, and L. Younes, “Large deformation diffeomorphic metric mapping of vector fields,” *IEEE transactions on medical imaging*, vol. 24, no. 9, pp. 1216–1230, 2005.
- [244] G. Postelnicu, L. Zollei, and B. Fischl, “Combined volumetric and surface registration,” *IEEE transactions on medical imaging*, vol. 28, no. 4, pp. 508–522, 2008.
- [245] N. J. Tustison, B. B. Avants, and J. C. Gee, “Directly manipulated free-form deformation image registration,” *IEEE transactions on image processing*, vol. 18, no. 3, pp. 624–635, 2009.
- [246] R. Gan, A. C. Chung, and S. Liao, “Maximum distance-gradient for robust image registration,” *Medical image analysis*, vol. 12, no. 4, pp. 452–468, 2008.
- [247] J. Modersitzki, “Flirt with rigidity—image registration with a local non-rigidity penalty,” *International Journal of Computer Vision*, vol. 76, no. 2, pp. 153–163, 2008.

- [248] E. Haber and J. Modersitzki, “Image registration with guaranteed displacement regularity,” *International journal of computer vision*, vol. 71, no. 3, pp. 361–372, 2007.
- [249] D. Zikic, M. Baust, A. Kamen, and N. Navab, “A general preconditioning scheme for difference measures in deformable registration,” in *2011 International Conference on Computer Vision*, 2011, pp. 49–56.
- [250] P. Thévenaz and M. Unser, “Optimization of mutual information for multiresolution image registration,” *IEEE transactions on image processing*, vol. 9, no. 12, pp. 2083–2099, 2000.
- [251] T. W. Tang and A. C. Chung, “Non-rigid image registration using graph-cuts,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2007, pp. 916–924.
- [252] R. W. So and A. C. Chung, “Non-rigid image registration by using graph-cuts with mutual information,” in *2010 IEEE International Conference on Image Processing*, 2010, pp. 4429–4432.
- [253] R. W. So, T. W. Tang, and A. C. Chung, “Non-rigid image registration of brain magnetic resonance images using graph-cuts,” *Pattern Recognition*, vol. 44, no. 10-11, pp. 2450–2467, 2011.
- [254] Q. Yang, L. Wang, and N. Ahuja, “A constant-space belief propagation algorithm for stereo matching,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1458–1465.
- [255] M. P. Heinrich, M. Jenkinson, J. M. Brady, and J. A. Schnabel, “Non-rigid image registration through efficient discrete optimization.” in *MIUA*, 2011, pp. 187–192.

- [256] A. Shekhovtsov, I. Kovtun, and V. Hlaváč, “Efficient mrf deformation model for non-rigid image matching,” *Computer Vision and Image Understanding*, vol. 112, no. 1, pp. 91–99, 2008.
- [257] B. Glocker, N. Komodakis, G. Tziritas, N. Navab, and N. Paragios, “Dense image registration through mrfs and efficient linear programming,” *Medical image analysis*, vol. 12, no. 6, pp. 731–741, 2008.
- [258] B. Glocker, A. Sotiras, N. Komodakis, and N. Paragios, “Deformable medical image registration: setting the state of the art with discrete methods,” *Annual review of biomedical engineering*, vol. 13, 2011.
- [259] D. Kwon, K. J. Lee, I. D. Yun, and S. U. Lee, “Nonrigid image registration using dynamic higher-order mrf model,” in *European Conference on Computer Vision*, 2008, pp. 373–386.
- [260] K. J. Lee, D. Kwon, I. D. Yun, and S. U. Lee, “Optical flow estimation with adaptive convolution kernel prior on discrete framework,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2504–2511.
- [261] J. Santamaría, O. Cordon, and S. Damas, “A comparative study of state-of-the-art evolutionary image registration methods for 3d modeling,” *Computer Vision and Image Understanding*, vol. 115, no. 9, pp. 1340–1354, 2011.
- [262] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [263] F. Commandeur, M. Goeller, J. Betancur, S. Cadet, M. Doris, X. Chen, D. S. Berman, P. J. Slomka, B. K. Tamarappoo, and D. Dey, “Deep learning for quantification of epicardial and thoracic adipose tissue from non-contrast

- ct,” *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1835–1846, 2018.
- [264] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, “Deep neural networks segment neuronal membranes in electron microscopy images,” in *Advances in neural information processing systems*, 2012, pp. 2843–2851.
- [265] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, “Deep convolutional neural networks for multi-modality isointense infant brain image segmentation,” *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [266] Y. Bar, I. Diamant, L. Wolf, and H. Greenspan, “Deep learning with non-medical training used for chest pathology identification,” in *Medical Imaging 2015: Computer-Aided Diagnosis*, vol. 9414, 2015, p. 94140V.
- [267] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, “Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network,” in *International conference on medical image computing and computer-assisted intervention*, 2013, pp. 246–253.
- [268] H. R. Roth, L. Lu, A. Farag, A. Sohn, and R. M. Summers, “Spatial aggregation of holistically-nested networks for automated pancreas segmentation,” in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 451–459.
- [269] P. Moeskops, J. M. Wolterink, B. H. van der Velden, K. G. Gilhuijs, T. Leiner, M. A. Viergever, and I. Išgum, “Deep learning for multi-task medical image segmentation in multiple modalities,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 478–486.

- [270] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, “Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation,” in *Advances in neural information processing systems*, 2016, pp. 3036–3044.
- [271] J. Kleesiek, A. Biller, G. Urban, U. Kothe, M. Bendszus, and F. Hamprecht, “Ilastik for multi-modal brain tumor segmentation,” *Proceedings MICCAI BraTS (brain tumor segmentation challenge)*, pp. 12–17, 2014.
- [272] K. Kamnitsas, L. Chen, C. Ledig, D. Rueckert, and B. Glocker, “Multi-scale 3d convolutional neural networks for lesion segmentation in brain mri,” *Ischemic stroke lesion segmentation*, vol. 13, p. 46, 2015.
- [273] Q. Dou, L. Yu, H. Chen, Y. Jin, X. Yang, J. Qin, and P.-A. Heng, “3d deeply supervised network for automated segmentation of volumetric medical images,” *Medical image analysis*, vol. 41, pp. 40–54, 2017.
- [274] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.
- [275] D. Nie, L. Wang, Y. Gao, and D. Shen, “Fully convolutional networks for multi-modality isointense infant brain image segmentation,” in *2016 IEEE 13Th international symposium on biomedical imaging (ISBI)*, 2016, pp. 1342–1345.
- [276] X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, and H. Fujita, “Three-dimensional ct image segmentation by combining 2d fully convolutional network with 3d majority voting,” in *Deep Learning and Data Labeling for Medical Applications*, 2016, pp. 111–120.

- [277] X. Zhou, R. Takayama, S. Wang, T. Hara, and H. Fujita, “Deep learning of the sectional appearances of 3d ct images for anatomical structure segmentation based on an fcn voting method,” *Medical physics*, vol. 44, no. 10, pp. 5221–5233, 2017.
- [278] P. Hu, F. Wu, J. Peng, Y. Bao, F. Chen, and D. Kong, “Automatic abdominal multi-organ segmentation using deep convolutional neural network and time-implicit level sets,” *International journal of computer assisted radiology and surgery*, vol. 12, no. 3, pp. 399–411, 2017.
- [279] H. R. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, K. Misawa, and K. Mori, “Hierarchical 3d fully convolutional networks for multi-organ segmentation,” *arXiv preprint arXiv:1704.06382*, 2017.
- [280] P. F. Christ, M. E. A. Elshaer, F. Ettlinger, S. Tatavarty, M. Bickel, P. Bilic, M. Rempfler, M. Armbruster, F. Hofmann, M. D’Anastasi *et al.*, “Automatic liver and lesion segmentation in ct using cascaded fully convolutional neural networks and 3d conditional random fields,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 415–423.
- [281] H. Zhang, Z. Kyaw, J. Yu, and S.-F. Chang, “PPR-FCN: Weakly supervised visual relation detection via parallel pairwise R-FCN,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4233–4241.
- [282] G. Zeng and G. Zheng, “Multi-stream 3d fcn with multi-scale deep supervision for multi-modality isointense infant brain mr image segmentation,” in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 2018, pp. 136–140.

- [283] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, “Automated melanoma recognition in dermoscopy images via very deep residual networks,” *IEEE transactions on medical imaging*, vol. 36, no. 4, pp. 994–1004, 2016.
- [284] T. Song, F. Meng, A. Rodríguez-Patón, P. Li, P. Zheng, and X. Wang, “U-next: A novel convolution neural network with an aggregation u-net architecture for gallstone segmentation in ct images,” *IEEE Access*, vol. 7, pp. 166 823–166 832, 2019.
- [285] Y. Zhang, J. Wu, W. Chen, Y. Chen, and X. Tang, “Prostate segmentation using z-net,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 11–14.
- [286] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, “3d u-net: learning dense volumetric segmentation from sparse annotation,” in *International conference on medical image computing and computer-assisted intervention*, 2016, pp. 424–432.
- [287] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein, M. Bendszus, and A. Biller, “Deep mri brain extraction: A 3d convolutional neural network for skull stripping,” *NeuroImage*, vol. 129, pp. 460–469, 2016.
- [288] G. Zeng, X. Yang, J. Li, L. Yu, P.-A. Heng, and G. Zheng, “3d u-net with multi-level deep supervision: fully automatic segmentation of proximal femur in 3d mr images,” in *International workshop on machine learning in medical imaging*, 2017, pp. 274–282.
- [289] F. Milletari, N. Navab, and S.-A. Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*, 2016, pp. 565–571.

- [290] E. Gibson, F. Giganti, Y. Hu, E. Bonmati, S. Bandula, K. Gurusamy, B. Davidson, S. P. Pereira, M. J. Clarkson, and D. C. Barratt, “Automatic multi-organ segmentation on abdominal ct with dense v-networks,” *IEEE transactions on medical imaging*, vol. 37, no. 8, pp. 1822–1834, 2018.
- [291] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [292] J. Kawahara, A. BenTaieb, and G. Hamarneh, “Deep features to classify skin lesions,” in *2016 IEEE 13th international symposium on biomedical imaging (ISBI)*, 2016, pp. 1397–1400.
- [293] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, “Voxresnet: Deep voxelwise residual networks for brain segmentation from 3d mr images,” *NeuroImage*, vol. 170, pp. 446–455, 2018.
- [294] J. Schmidhuber and S. Hochreiter, “Long short-term memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [295] N. Srivastava, E. Mansimov, and R. Salakhudinov, “Unsupervised learning of video representations using lstms,” in *International conference on machine learning*, 2015, pp. 843–852.
- [296] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, “Convolutional lstm network: A machine learning approach for precipitation nowcasting,” in *Advances in neural information processing systems*, 2015, pp. 802–810.

- [297] M. F. Stollenga, W. Byeon, M. Liwicki, and J. Schmidhuber, “Parallel multi-dimensional lstm, with application to fast biomedical volumetric image segmentation,” in *Advances in neural information processing systems*, 2015, pp. 2998–3006.
- [298] D. Cheng and M. Liu, “Combining convolutional and recurrent neural networks for alzheimer’s disease diagnosis using pet images,” in *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, 2017, pp. 1–5.
- [299] Y. Xie, Z. Zhang, M. Sapkota, and L. Yang, “Spatial clockwork recurrent neural network for muscle perimysium segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 185–193.
- [300] E. Pesce, S. J. Withey, P.-P. Ypsilantis, R. Bakewell, V. Goh, and G. Montana, “Learning to detect chest radiographs containing pulmonary lesions using visual attention networks,” *Medical image analysis*, vol. 53, pp. 26–38, 2019.
- [301] Q. Guan, Y. Huang, Z. Zhong, Z. Zheng, L. Zheng, and Y. Yang, “Diagnose like a radiologist: Attention guided convolutional neural network for thorax disease classification,” *arXiv preprint arXiv:1801.09927*, 2018.
- [302] Y. Zhou, L. Xie, W. Shen, Y. Wang, E. K. Fishman, and A. L. Yuille, “A fixed-point model for pancreas segmentation in abdominal ct scans,” in *International conference on medical image computing and computer-assisted intervention*. Springer, 2017, pp. 693–701.

- [303] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2961–2969.
- [304] A. O. Vuola, S. U. Akram, and J. Kannala, “Mask-rcnn and u-net ensembled for nuclei segmentation,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 208–212.
- [305] J. W. Johnson, “Adapting mask-rcnn for automatic nucleus segmentation,” *arXiv preprint arXiv:1805.00500*, 2018.
- [306] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, “Improved optimization for the robust and accurate linear registration and motion correction of brain images,” *Neuroimage*, vol. 17, no. 2, pp. 825–841, 2002.
- [307] S. Klein, J. P. Pluim, M. Staring, and M. A. Viergever, “Adaptive stochastic gradient descent optimisation for image registration,” *International journal of computer vision*, vol. 81, no. 3, p. 227, 2009.
- [308] A. Al Suman, M. Asikuzzaman, A. L. Webb, D. M. Perriman, and M. R. Pickering, “Inter-subject image registration of clinical neck MRI volumes using discrete periodic spline wavelet and free form deformation,” in *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 2018, pp. 1–5.
- [309] A. A. Suman, M. N. Aktar, M. Asikuzzaman, A. L. Webb, D. M. Perriman, and M. R. Pickering, “Segmentation and reconstruction of cervical muscles using knowledge-based grouping adaptation and new step-wise registration with discrete cosines,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, pp. 1–13, 2017.

- [310] D. Yu, F. Yang, C. Yang, C. Leng, J. Cao, Y. Wang, and J. Tian, “Fast rotation-free feature-based image registration using improved n-sift and gmm-based parallel optimization,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 8, pp. 1653–1664, 2016.
- [311] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim, “A survey of medical image registration—under review,” *Medical image analysis*, vol. 33, pp. 140–144, 2016.
- [312] J. A. Schnabel, M. P. Heinrich, B. W. Papież, and J. M. Brady, “Advances and challenges in deformable image registration: From image fusion to complex motion modelling,” *Medical image analysis*, vol. 33, pp. 145–148, 2016.
- [313] S. Klein, M. Staring, and J. P. Pluim, “Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines,” *IEEE transactions on image processing*, vol. 16, no. 12, pp. 2879–2890, 2007.
- [314] B. W. Papież, M. P. Heinrich, J. Fehrenbach, L. Risser, and J. A. Schnabel, “An implicit sliding-motion preserving regularisation via bilateral filtering for deformable image registration,” *Medical image analysis*, vol. 18, no. 8, pp. 1299–1311, 2014.
- [315] I. J. Simpson, M. J. Cardoso, M. Modat, D. M. Cash, M. W. Woolrich, J. L. Andersson, J. A. Schnabel, and S. Ourselin, “Probabilistic non-linear registration with spatially adaptive regularisation,” *Medical image analysis*, vol. 26, no. 1, pp. 203–216, 2015.
- [316] S. Reaungamornrat, T. De Silva, A. Uneri, S. Vogt, G. Kleinszig, A. J. Khanna, J.-P. Wolinsky, J. L. Prince, and J. H. Siewerdsen, “MIND demons:

- symmetric diffeomorphic deformable registration of MR and CT for image-guided spine surgery,” *IEEE transactions on medical imaging*, vol. 35, no. 11, pp. 2413–2424, 2016.
- [317] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, vol. 1, pp. 886–893.
- [318] K. Mikolajczyk and C. Schmid, “A performance evaluation of local descriptors,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.
- [319] S. M. Kahaki, S.-L. Wang, and A. Stepanyants, “Accurate registration of in vivo time-lapse images,” in *Medical Imaging 2019: Image Processing*, vol. 10949, 2019, p. 109491D.
- [320] Y. Peng, W. Lin, S. Ying, and J. Peng, “Soft shape registration under lie group frame,” *IET Computer Vision*, vol. 7, no. 6, pp. 437–447, 2013.
- [321] S. Ying, Y. Peng, and Z. Wen, “Iwasawa decomposition: a new approach to 2D affine registration problem,” *Pattern Analysis and Applications*, vol. 14, no. 2, pp. 127–137, 2011.
- [322] Z. Wu, H. Chen, S. Du, M. Fu, N. Zhou, and N. Zheng, “Correntropy based scale ICP algorithm for robust point set registration,” *Pattern Recognition*, vol. 93, pp. 14–24, 2019.
- [323] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. Giphart, W. J. Niessen, and B. P. Lelieveldt, “2d–3d shape reconstruction of the distal femur from stereo x-ray imaging using statistical shape models,” *Medical image analysis*, vol. 15, no. 6, pp. 840–850, 2011.

- [324] L.-Y. Hsu and M. H. Loew, “Fully automatic 3d feature-based registration of multi-modality medical images,” *Image and Vision Computing*, vol. 19, no. 1-2, pp. 75–85, 2001.
- [325] C. C. Conlin, J. L. Zhang, F. Rousset, C. Vachet, Y. Zhao, K. A. Morton, K. Carlston, G. Gerig, and V. S. Lee, “Performance of an efficient image-registration algorithm in processing mr renography data,” *Journal of Magnetic Resonance Imaging*, vol. 43, no. 2, pp. 391–397, 2016.
- [326] M. P. Heinrich, M. Jenkinson, M. Bhushan, T. Matin, F. V. Gleeson, M. Brady, and J. A. Schnabel, “Mind: Modality independent neighbourhood descriptor for multi-modal deformable registration,” *Medical image analysis*, vol. 16, no. 7, pp. 1423–1435, 2012.
- [327] Y. Amit, “A nonlinear variational problem for image matching,” *SIAM Journal on Scientific Computing*, vol. 15, no. 1, pp. 207–224, 1994.
- [328] W. Cai and J. Wang, “Adaptive multiresolution collocation methods for initial-boundary value problems of nonlinear pdes,” *SIAM Journal on Numerical Analysis*, vol. 33, no. 3, pp. 937–970, 1996.
- [329] A. A. Muhit, M. R. Pickering, M. R. Frater, and J. F. Arnold, “Video coding using elastic motion model and larger blocks,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 5, pp. 661–672, 2010.
- [330] G. Borgefors, “Hierarchical chamfer matching: A parametric edge matching algorithm,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 10, no. 6, pp. 849–865, 1988.

- [331] Z. Li, D. Mahapatra, J. A. Tielbeek, J. Stoker, L. J. van Vliet, and F. M. Vos, “Image registration based on autocorrelation of local structure,” *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 63–75, 2015.
- [332] C. Wachinger and N. Navab, “Entropy and laplacian images: Structural representations for multi-modal registration,” *Medical image analysis*, vol. 16, no. 1, pp. 1–17, 2012.
- [333] D. Loeckx, P. Slagmolen, F. Maes, D. Vandermeulen, and P. Suetens, “Non-rigid image registration using conditional mutual information,” *IEEE transactions on medical imaging*, vol. 29, no. 1, pp. 19–29, 2009.
- [334] M. P. Heinrich, M. Jenkinson, B. W. Papież, M. Brady, and J. A. Schnabel, “Towards realtime multimodal fusion for image-guided interventions using self-similarities,” in *International conference on medical image computing and computer-assisted intervention*, 2013, pp. 187–194.
- [335] J. E. Iglesias and M. R. Sabuncu, “Multi-atlas segmentation of biomedical images: A survey,” *Medical image analysis*, vol. 24, no. 1, pp. 205–219, 2015.
- [336] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, “Multi-atlas segmentation with joint label fusion,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 611–623, 2012.
- [337] B. B. Avants, N. J. Tustison, G. Song, P. A. Cook, A. Klein, and J. C. Gee, “A reproducible evaluation of ants similarity metric performance in brain image registration,” *Neuroimage*, vol. 54, no. 3, pp. 2033–2044, 2011.
- [338] J. C. Caicedo, J. Roth, A. Goodman, T. Becker, K. W. Karhohs, M. Broisin, C. Molnar, C. McQuin, S. Singh, F. J. Theis *et al.*, “Evaluation of deep learning strategies for nucleus segmentation in fluorescence images,” *Cytometry Part A*, vol. 95, no. 9, pp. 952–965, 2019.

- [339] A. Hänsch, M. Schwier, T. Gass, T. Morgas, B. Haas, V. Dicken, H. Meine, J. Klein, and H. K. Hahn, “Evaluation of deep learning methods for parotid gland segmentation from CT images,” *Journal of Medical Imaging*, vol. 6, no. 1, p. 011005, 2018.
- [340] L. Bi, J. Kim, E. Ahn, A. Kumar, M. Fulham, and D. Feng, “Dermoscopic image segmentation via multistage fully convolutional networks,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2065–2074, 2017.
- [341] Z. Zhang, P. Chen, M. Sapkota, and L. Yang, “Tandemnet: Distilling knowledge from medical images using diagnostic reports as optional semantic references,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2017, pp. 320–328.
- [342] Z. Zhang, Y. Xie, F. Xing, M. McGough, and L. Yang, “Mdnet: A semantically and visually interpretable medical image diagnosis network,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6428–6436.
- [343] X. Wang, Y. Peng, L. Lu, Z. Lu, and R. M. Summers, “Tienet: Text-image embedding network for common thorax disease classification and reporting in chest x-rays,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9049–9058.
- [344] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu, “Tell me where to look: Guided attention inference network,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9215–9223.

- [345] C. Li, J. C. Gore, and C. Davatzikos, “Multiplicative intrinsic component optimization (MICO) for MRI bias field estimation and tissue segmentation,” *Magnetic resonance imaging*, vol. 32, no. 7, pp. 913–923, 2014.
- [346] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, “Conditional random fields as recurrent neural networks,” in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1529–1537.
- [347] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [348] Y. Song, L. Zhang, S. Chen, D. Ni, B. Lei, and T. Wang, “Accurate segmentation of cervical cytoplasm and nuclei based on multiscale convolutional network and graph partitioning,” *IEEE Transactions on Biomedical Engineering*, vol. 62, no. 10, pp. 2421–2433, 2015.
- [349] A. A. Taha and A. Hanbury, “An efficient algorithm for calculating the exact hausdorff distance,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2153–2163, 2015.