University of Groningen

Automatic Segmentation of the Mandible for Three-Dimensional Virtual Surgical Planning

Qiu, Bingjiang

DOI:
[10.33612/diss.179140520](https://doi.org/10.33612/diss.179140520)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Qiu, B. (2021). *Automatic Segmentation of the Mandible for Three-Dimensional Virtual Surgical Planning*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.179140520>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Automatic Segmentation of the Mandible for Three-Dimensional Virtual Surgical Planning

Bingjiang Qiu

The work described in this thesis was performed in 3D Lab, Machine Learning Lab of Data Science Center in Health (DASH), and Department of Radiation Oncology at University Medical Center Groningen, the Netherlands.

Printing of this thesis was financially supported by:
The University of Groningen (RUG)
The University Medical Center Groningen (UMCG)
Graduate School of Medical Science (GSMS)
The W.J. Kolff Institute

Cover design: Ningli Zhu

Copyright © Bingjiang Qiu 2021

All rights reserved. No part of this thesis may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, without prior permission of the author.

university of
groningen

Automatic Segmentation of the Mandible for Three-Dimensional Virtual Surgical Planning

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Wednesday 6 October 2021 at 9.00 hours

by

Bingjiang Qiu

born on 11 September 1990
in Chaozhou, China

Supervisors

Dr. P.M.A. van Ooijen
Dr. M.J.H. Witjes

Co-supervisors

Dr. R.J.H. Borra
Dr. J. Kraeima

Assessment Committee

Prof. Y. Ren
Prof. J.B.T.M. Roerdink
Prof. C.H. Slump

CONTENTS

I	Introduction	1
1	General introduction	3
1.1	Mandible structure	4
1.2	Three-dimensional virtual surgical planning (3D VSP)	5
1.3	Aims and outline of this thesis	7
	References	7
2	Automatic mandible segmentation methods - A review	11
2.1	Introduction	13
2.2	Method for literature selection	15
2.3	Results	15
2.3.1	Image modality	16
2.3.2	Image database	17
2.3.3	Evaluation metrics	17
2.3.4	Methodology.	18
2.4	Discussion	28
2.5	Conclusion	30
	References	32
II	Mandible segmentation on CT image	43
3	Mandible segmentation using multi-planar volume-to-slice strategy	45
3.1	Introduction	47
3.2	Materials and methods	49
3.2.1	Data preparation.	49
3.2.2	Methods	49
3.3	Experiments	52
3.3.1	Evaluation criterion	52
3.3.2	Experimental results	54
3.4	Discussion	55
3.5	Conclusion	61
	References	62
4	Recurrent Convolutional Neural Networks for Mandible Segmentation	67
4.1	Introduction	69
4.2	Materials and Methods	71
4.2.1	Feedforward, feedback, and recurrent networks	71
4.2.2	Recurrent convolutional neural networks for mandible segmentation (RCNNSeg)	72

4.2.3	Experimental Setup	74
4.3	Results	75
4.3.1	The UMCG head and neck dataset	75
4.3.2	PDDCA dataset	77
4.4	Discussion	79
4.5	Conclusion	81
	References	82
III	Mandible segmentation on CBCT image	87
5	Mandible segmentation on dental CBCT scans using prior shape model	89
5.1	Introduction	91
5.2	Methodology	94
5.2.1	Building a Mean Mandible Shape Model	95
5.2.2	Prior Shape Feature Extractor (PSFE).	96
5.2.3	Recurrent Convolutional Neural Networks for Segmentation	97
5.2.4	Combo Loss Function	98
5.2.5	Dataset.	99
5.2.6	Evaluation Metrics	99
5.2.7	Implementation Details	100
5.3	Results	100
5.3.1	Method Comparison	100
5.3.2	Ablation Experiments	101
5.3.3	Reliability Analysis	104
5.3.4	Experiments on the PDDCA Dataset	105
5.4	Discussion	105
5.5	Conclusions.	107
	References	108
6	Mandible segmentation on CBCT scans using C2F model	113
6.1	Introduction	115
6.2	Methodology	117
6.2.1	Curriculum learning in mandible segmentation	117
6.2.2	Coarse stage: mandible-like organ prediction	118
6.2.3	Fine stage: false positive reduction.	119
6.2.4	Loss	120
6.2.5	Evaluation Metrics	121
6.3	Experiments	122
6.3.1	Datasets	122
6.3.2	Implementation details	123
6.3.3	Results	124
6.4	Discussion	128
6.5	Conclusions.	129
	References	130

IV Discussion and conclusion	135
7 Discussion and conclusion	137
7.1 General discussion	138
7.2 Future perspectives	139
7.2.1 MRI-only 3D VSP	140
7.2.2 Metrics.	141
7.3 Conclusion	141
References	142
Summary	143
Samenvatting	145
Acknowledgements	147
About the author	149
List of Publications	151

I

INTRODUCTION

1

GENERAL INTRODUCTION

1

Oral and maxillofacial surgery (OMFS) procedures are often prepared using 3D medical head and neck images [1]. Although the preoperative medical image data can provide intuitive and clear images for e.g. the mandible structure, the overall 3D mandible information can not be used unedited for rendering a 3D model due to the strong artifacts (because of the teeth (fillings) or metal implants) from the radiography images, which easily lead to the blurred boundaries between anatomical structures. The current standard for surgeons is to preoperatively determine the surgery planning, such as patient specific osteosynthesis plates, fixation method, screw length and screw placement direction. When the 3D planning involves areas of incomplete 3D information of the mandible, this may cause errors during planning and operation, for example, osteosynthesis material failure (poor fixation or material defect), occlusal disorders and temporomandibular joint dysfunction [2, 3].

With the application of image-guided computer-assisted surgery, it is possible to accurately construct the ideal new mandibular shape and restore its original function [2], and overcome the shortcomings of poor repeatability caused by the previous estimation based on the surgeon's clinical experience. Image-guided computer-assisted surgery includes three-dimensional virtual surgical planning (3D VSP), and surgery navigation [4]. Among them, 3D VSP is currently widely used in mandible reconstruction and can result in accurate surgical planning [4, 5].

In this thesis, we focus on segmenting the mandible from computed tomography (CT) or cone beam CT (CBCT), aiming to reduce the inter-observer variability between different software operators when 3D reconstructing the mandible and shorten the time of segmenting the mandible in 3D VSP [6]. Mandible segmentation in CB/CBCT is highly challenging because of complex structure of the mandible and strong artifacts (incomplete rendering of temporomandibular joint, scatter from dental fillings) which is introduced in Chapter 2.

This introduction continues with background information to support the topics of the thesis. In the next section, Section 1.1 we present a short overview about the mandible structure and its function. Subsequently, in Section 1.2, we introduce 3D VSP. Then, in Section 1.3, we present the research objectives and a short overview of this thesis.

1.1. MANDIBLE STRUCTURE

To understand mandible reconstruction, we need to understand its anatomical structure. The mandible is located in the lower 1/3 of the face, is the largest and strongest bone of the face and has complex morphology and function (such as chewing movements). For example, the mandibular body supports the lower teeth in the alveolar process, and it also contains the mandibular canal and nerves [7]. Figure 1.1 shows the anatomical structure of the mandible. The mandible consists of the mandibular body, angles, rami, condyles and coronoid processes. Furthermore, the bilateral mandibular condyles also participate in the formation of the temporomandibular joint, which is one of the most commonly used joints in human body [8]. This joint is used for chewing and speaking, and these movements are performed many times every day. The mandible is the only movable bone in the maxillofacial region. Mandible defects will have a significant impact on the face and function of patients, and can affect physiology and psychology, seriously affecting

quality of life [9]. The main causes of mandibular defects are tumor resection, severe trauma, and congenital developmental malformations [9]. Therefore, the reconstruction of mandibular defects is a challenging problem for maxillofacial head and neck and plastic surgeons.

Figure 1.1 | Schematic illustration of the mandible anatomy.

1.2. THREE-DIMENSIONAL VIRTUAL SURGICAL PLANNING (3D VSP)

Currently, three-dimensional virtual surgical planning (3D VSP) for mandible reconstruction has been widely applied and has achieved satisfactory results in oral and maxillofacial surgery (OMFS), in cases such as during oral tumor removal, orthodontic, orthognathic, and trauma surgery [1, 10]. In Figure 1.2, a schematic illustration is shown of the 3D workflow. 3D VSP is based on radiography, which in most cases already is part of routine preoperative or diagnostic examinations. The strength of 3D VSP imaging in OMFS is that it combines multi-modality imaging (such as CT, CBCT or magnetic resonance imaging (MRI)) into an accurate 3D scenario of the patient and the relevant anatomy and pathology. It allows pre-operative decision-making on the surgical steps, followed by accurate translation towards the actual patient by means of 3D printed guides or surgical navigation. [11–13].

Before 3D VSP, surgeons relied on the preoperative radiography and their clinical experience during operation. However, this does not provide 3D information for mandible resection, bone tissue flap preparation, and bone graft cutting prior to the operation. 3D VSP has the advantage of pre-operative planning and executing the plan by means of

1

Figure 1.2 | A procedure using the 3D VSP workflow.

cutting/drill guides or patient specific implants. Deviations from normal/pre-operative anatomy in mandible reconstruction during operation may affect the functional outcome because of the complex anatomical structure of the mandible. 3D VSP can provide a way to make up for the deficiencies that are easily caused by the fact that surgeons formulate the surgical planning and perform operations using their surgical skills and clinical experience. Reports about the application of 3D VSP for assisting precise mandible reconstruction began in the late twentieth century [14, 15], including the cutting, separation and movement of bone, or the placement and positioning of the osteosynthesis plates and screws. 3D VSP can help engineers formulate a reasonable surgical plan and predict

surgical results, and 3D VSP can guide various operations based on the simulation results to ensure that the surgeon can accurately perform the actual operation [1]. Furthermore, 3D VSP allows the surgeon to evaluate the surgical results in high detail after the operation [16]. 3D VSP has certain teaching significance such that inexperienced surgeons can benefit from virtual surgery by understanding the problem in 3D as well as being assisted intraoperatively by 3D VSP tools. [17].

Moreover, 3D VSP is often combined with surgical navigation which provides surgeons with real-time position information [18]. This can further correct possible errors between preoperative simulation and intra-operative procedures. Mandible reconstruction, the precise adaption of the mandibular angle or the position of the condyles is necessary using 3D VSP. Image segmentation from the radiography images of the head and neck, which is a process to create a 3D volume of the target tissue, is a useful tool to visualize the mandible and quantify geometric parameters [6]. Most software that converts DICOM data for rendering a 3D model of the mandible show inaccuracies of certain anatomical areas of the mandible. During the process of rendering a 3D model, the engineer has to manually adjust the 3D model. Certain areas can be incompletely displayed (condylar head) or hampered by scattering from metals, providing a challenge during 3D VSP. Studies have shown that 3D VSP requires accurate segmentation of the mandible, which is currently performed by medical technicians [19, 20]. Mandible segmentation was usually done manually, which is a time-consuming and poorly reproducible process [21]. The aim here is to overcome these limitations and is to develop an acceptable mandible segmentation method for 3D VSP.

1.3. AIMS AND OUTLINE OF THIS THESIS

This thesis will introduce novel deep learning-based algorithms for mandible segmentation. Our major aim is to develop an accurate and fast mandible segmentation method for 3D VSP. Chapter 2, "Overview of mandible segmentation methods", gives an overview of existing automatic methods for mandible segmentation in the literature. In Chapter 3, we describe a method of utilizing multi-planar mandible segmentation for the 3D mandible segmentation and evaluate the performance in two head and neck datasets. In Chapter 4, a method based on recurrent connections is introduced to consider the anatomical structural continuity of the mandible and is evaluated in two head and neck datasets. In Chapter 5, network architecture using a mean mandible model from a statistical shape model as prior input is introduced, and implementation details are explained and evaluated in two datasets. In Chapter 6, a mandible segmentation method from coarse to fine segmentation is introduced, and this method uses probable mandible-like candidates in the coarse stage and then applies those candidates for fine segmentation in the fine stage. Finally, in Chapter 7, we give our conclusions and outlook.

REFERENCES

- [1] J. Kraeima, *Three dimensional virtual surgical planning for patient specific osteosynthesis and devices in oral and maxillofacial surgery. A new era.*, Ph.D. thesis, University

of Groningen (2019).

- [2] Y. Ghantous, A. Nashef, A. Mohanna, and I. Abu-El-naaj, *Three-dimensional technology applications in maxillofacial reconstructive surgery: Current surgical implications*, *Nanomaterials* **10**, 2523 (2020).
- [3] T. P. Ter Braak, S. G. B. de Koning, M. J. A. van Alphen, F. van der Heijden, W. H. Schreuder, R. L. P. van Veen, and M. B. Karakullukcu, *A surgical navigated cutting guide for mandibular osteotomies: accuracy and reproducibility of an image-guided mandibular osteotomy*, *International journal of computer assisted radiology and surgery* **15**, 1719 (2020).
- [4] D. N. Kirke, R. P. Owen, V. Carrao, B. A. Miles, and J. I. Kass, *Using 3d computer planning for complex reconstruction of mandibular defects*, *Cancers of the head & neck* **1**, 1 (2016).
- [5] J. Kraeima, H. H. Glas, B. B. J. Merema, A. Vissink, F. K. Spijkervet, and M. J. Witjes, *Three-dimensional virtual surgical planning in the oncologic treatment of the mandible*, *Oral diseases* **27**, 14 (2021).
- [6] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. van Ooijen, *Automatic segmentation of the mandible from computed tomography scans for 3d virtual surgical planning using the convolutional neural network*, *Physics in Medicine & Biology* **64**, 175020 (2019).
- [7] *Anatomy and physiology*, <https://courses.lumenlearning.com/hccs-ap1/chapter/the-skull/> (2021).
- [8] B. Bordoni and M. Varacallo, *Anatomy, head and neck, temporomandibular joint*, StatPearls [Internet] (2019).
- [9] B. P. Kumar, V. Venkatesh, K. J. Kumar, B. Y. Yadav, and S. R. Mohan, *Mandibular reconstruction: overview*, *Journal of maxillofacial and oral surgery* **15**, 425 (2016).
- [10] E.-O. Batbayar, *Optimization of mandibular fracture treatment*, Ph.D. thesis, University of Groningen (2019).
- [11] D. X. Ji, K. W. C. Foong, and S. H. Ong, *A two-stage rule-constrained seedless region growing approach for mandibular body segmentation in mri*, *International journal of computer assisted radiology and surgery* **8**, 723 (2013).
- [12] J. K. Udupa, D. Odhner, L. Zhao, Y. Tong, M. M. Matsumoto, K. C. Ciesielski, A. X. Falcao, P. Vaideeswaran, V. Ciesielski, B. Saboury, *et al.*, *Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images*, *Medical image analysis* **18**, 752 (2014).
- [13] J. Dai, X. Wang, Y. Dong, H. Yu, D. Yang, and G. Shen, *Two-and three-dimensional models for the visualization of jaw tumors based on ct-mri image fusion*, *Journal of Craniofacial Surgery* **23**, 502 (2012).

- [14] A. Eckardt and G. R. Swennen, *Virtual planning of composite mandibular reconstruction with free fibula bone graft*, Journal of Craniofacial Surgery **16**, 1137 (2005).
- [15] W. Hallermann, S. Olsen, T. Bardyn, F. Taghizadeh, A. Banic, and T. Iizuka, *A new method for computer-aided operation planning for extensive mandibular reconstruction*, Plastic and reconstructive surgery **117**, 2431 (2006).
- [16] M.-S. Daniel-Amaral-Alves Marlière, A.-R.-M. S. Demétrio, L. A. Caio-Bellini Lovisi, et al., *Accuracy between virtual surgical planning and actual outcomes in orthognathic surgery by iterative closest point algorithm and color maps: A retrospective cohort study*, Medicina oral, patologia oral y cirugia bucal **24**, e243 (2019).
- [17] N. N. Smit, K. Lawonn, A. Kraima, M. DeRuiter, S. Bruckner, E. Eisemann, and A. Vilanova, *Model-based visualization for medical education and training*, (Eurographics, 2019).
- [18] J. Wu, J. Sun, S. G. Shen, B. Xu, J. Li, and S. Zhang, *Computer-assisted navigation: its role in intraoperatively accurate mandibular reconstruction*, Oral surgery, oral medicine, oral pathology and oral radiology **122**, 134 (2016).
- [19] J. Kraeima, B. Dorgelo, H. Gulbitti, R. Steenbakkens, K. Schepman, J. Roodenburg, F. Spijkervet, R. Schepers, and M. Witjes, *Multi-modality 3d mandibular resection planning in head and neck cancer using ct and mri data fusion: A clinical series*, Oral oncology **81**, 22 (2018).
- [20] P. Metzler, E. J. Geiger, A. Alcon, X. Ma, and D. M. Steinbacher, *Three-dimensional virtual surgery accuracy for free fibula mandibular reconstruction: planned versus actual results*, Journal of Oral and Maxillofacial Surgery **72**, 2601 (2014).
- [21] T. J. Huff, P. E. Ludwig, and J. M. Zuniga, *The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning*, Expert review of medical devices **15**, 349 (2018).

2

AUTOMATIC SEGMENTATION OF MANDIBLE FROM CONVENTIONAL METHODS TO DEEP LEARNING - A REVIEW

Bingjiang Qiu, Hylke V. D. Wel, Joep Kraeima, Haye H. Glas, Jiapan Guo, Ronald J. H. Borra, Max J. H. Witjes and Peter M. A. van Ooijen.

Qiu, Bingjiang, et al. "Automatic Segmentation of Mandible from Conventional Methods to Deep Learning—A Review." *Journal of Personalized Medicine* 11.7 (2021): 629.

ABSTRACT

Medical imaging techniques, such as (cone beam) computed tomography and magnetic resonance imaging, have proven to be a valuable component for oral and maxillofacial surgery (OMFS). Accurate segmentation of the mandible from head and neck (H&N) scans is an important step in order to build a personalized 3D digital mandible model for 3D printing and treatment planning of OMFS. Segmented mandible structures are used to effectively visualize the mandible volumes and to evaluate particular mandible properties quantitatively. However, mandible segmentation is always challenging for both clinicians and researchers, due to complex structures and higher attenuation materials, such as teeth (filling) or metal implants that easily lead to high noise and strong artifacts during scanning. Moreover, the size and shape of the mandible vary to a large extent between individuals. Therefore, mandible segmentation is a tedious and time-consuming task and requires adequate training to be performed properly. With the advancement of computer vision approaches, researchers have developed several algorithms to automatically segment the mandible during the last two decades. The objective of this review was to present the available fully (semi)automatic segmentation methods of the mandible published in different scientific articles. This review provides a vivid description of the scientific advancements to clinicians and researchers in this field to help develop novel automatic methods for clinical applications.

Keywords: mandible segmentation; 3D virtual surgical planning; convolutional neural networks; machine learning

2.1. INTRODUCTION

Three-dimensional (3D) medical imaging techniques have a fundamental role in the field of oral and maxillofacial surgery (OMFS) [2, 3]. 3D images are used to guide diagnosis, assess the severity of disease and for pre-operative planning and per-operative guidance using 3D images and virtual surgical planning (VSP) [5]. In the field of oral cancer, where surgical resection requiring the partial removal of the mandible is a common treatment, resection surgery is often based on 3D VSP to accurately design a resection plan around tumor margins [4]. In orthognathic surgery and dental implant surgery, 3D VSP is also extensively used to precisely guide mandibular surgery [5]. Image segmentation from the radiography images of the head and neck (H&N), which is a process to create a 3D volume of the target tissue, is a useful tool to visualize the mandible and quantify geometric parameters [6]. Studies have shown that 3D VSP requires accurate segmentation of the mandible, which is currently performed by medical technicians [4, 7]. Mandible segmentation was usually done manually, which is a time-consuming and poorly reproducible process [8].

The mandible is located in the lower part of the facial skeleton and consists of the body, rami, angle, and condyles. The different anatomical regions of the mandible have varying densities, which is especially true if the teeth and surrounding soft tissue are also considered, which makes accurate manual segmentation of the mandible time-consuming and challenging [9]. Because a 3D model of the mandible is necessary for 3D VSP in OMFS, and time is often a limiting factor in the clinical workflow, fast mandible segmentation has become a frequent topic of research in recent decades [8]. Researchers have attempted to automate the segmentation process of the mandible, as well as all kinds of tissues from medical images, to reduce the processing time and also the inter-observer variability [10]. Recent advances in image segmentation [11–15] have enabled their applications to medical image segmentation, which stimulate progress in automating mandible segmentation. Although significant progress has been made in mandible segmentation, state-of-the-art methods still experience unsatisfactory outcomes, with several challenges to be solved. For oncology segmentation, the segmentation has to be made for an affected bone (with the tumor), which is a different task from segmenting a bone in implantology or for orthognathic purposes. In implantology, many patients have incomplete teeth, which can accordingly lead to resorption of the bone causing shape variation in the mandible among patients. In orthognathic surgery, the population considered varies greatly in shape in addition to asymmetric, which is also a possible challenge for (automatic) segmentation. Figure 2.1 shows research challenges in automatic segmentation algorithms for the mandible. The challenges associated with mandibular bone segmentation can be categorized as follows:

- The presence of anatomically complex bony structures in the scans. As the examples show in Figure 2.1 (a), a normal H&N scan includes other bony structures, with a complex anatomy and a similar density. Determining the correct boundaries and separating the mandibular bone from the other bones may be challenging.
- Artifacts. When X-rays pass through high-density structures or materials, including teeth, postoperative metal implants, etc., the signal on the detectors will change,

which will lead to attenuation calculation errors in the (cone beam) computed tomography (CBCT/CT) reconstruction process and consequently cause high noise and strong artifacts in the visual impression of the scans [16]. The mandible boundaries nearby teeth tend to be blurred and hard to detect. In particular, the boundaries of mandible rami are difficult to be identified when dental braces and metal implants badly affect the image quality [17], as shown in Figure 2.1 (b). Furthermore, the fact that the superior and the inferior teeth are at the same slice and even overlapping that makes segmentation methods challenging, as shown in Figure 2.1 (c).

- **Low contrast.** Due to the tomography process and the thinness of condyles, the slices in condyles commonly have low contrast [18], as shown in Figure 2.1 (d). Especially in cone-beam CT (CBCT) scans, the condyles in the scans are more blurred than in CT because of its intensity inhomogeneity [19, 20].
- **Variation in the appearance.** As the examples show in Figure 2.1 (e), mandibular shapes vary in shape and size between individuals [18]. Additionally, the shape of the mandibular condyles and body is extremely variable among patients and different age groups [21].
- **Annotation Bias.** The manual mandible segmentation often leads to inter-observer variability (Dice score of 94.09% between two clinical experts) [22], which directly influences the quality of treatment planning.

Figure 2.1 | Examples of typical cases that challenge accurate mandible segmentation. (a) Various bone-structured organs in the H&N scans. (b) Metal artifacts. (c) Presence of inferior and superior teeth in the same slice. (d) Lower intensity in the condyles. (e) Large variation in mandibles between patients.

In the present article, a review has been conducted, where the literature available in PubMed and Web of Science databases relating to these studies was considered. In the literature, a variety of segmentation methods, based on statistical shape model, atlas and

machine learning, are proposed for head and neck scanning. The selected publications were divided into several categories depending on the type of method used and discussed extensively.

2.2. METHOD FOR LITERATURE SELECTION

2

In February 2021, a search was conducted on the Web of Science (<https://apps.webofknowledge.com>) and PubMed (<https://www.ncbi.nlm.nih.gov/pubmed/>), with the topic keywords and search builder: (automat* AND segment* AND mandible AND 3D). While searching the literature, no specific time line was set. The literature search was performed by the two observers.

In this article, the lists of literature acquired by the above-mentioned database search were imported into Rayyan QCRI [23]. Using the web application, duplicates were searched and removed accordingly. Subsequently, another search was performed to exclude the articles that did not include “segment” or “mandible” keywords in their title and abstracts. Consequently, the irrelevant studies were removed by manually screening the titles and abstracts of journal articles. The obtained relevant publications were further supplemented with selected publications found in their list of references. A detailed analysis of the resulting publications was then conducted from the perspective of image modality, image database, methodology, and performance metrics.

2.3. RESULTS

In the field of automatic 3D mandible segmentation, the search on the Web of Science and PubMed databases yielded 33 and 45 results, respectively. After detecting the duplicates in Rayyan, 23 duplicates were removed. After reviewing their abstracts by two observers, 12 were considered to be relevant and were further supplemented with selected publications from their list of references. In total, we collected 77 publications focused on mandible segmentation or organs at risk (OARs) (including mandible) segmentation in the H&N region. Here, we emphasize that we collected 41 publications from the reference of [24]. An overview of the results of the literature search is presented in Figure 2.2.

Figure 2.2 | Flowchart of the literature review article selection process.

2.3.1. IMAGE MODALITY

As shown in Table 2.3-2.11, CT imaging is primarily performed for mandibular bone segmentation in their works since CT images provide a promising visibility of the bone structures. There are 70, 7 and 4 publications using CT, CBCT and magnetic resonance imaging (MRI), respectively, to visualize H&N (note: some articles applied the scans from multiple image modalities). CBCT images are applied in H&N because of its lower radiation dose, lower cost and faster scanning time than conventional CT [25]. In particular, CBCT is often used for validating patient settings or adjusting treatment plans to adapt to anatomical changes. While CT images provide a good visibility of the bony anatomy, the contrast differences between various soft tissues are much lower than those in MRI. In the consensus on manual delineation of the mandible based on CT, it is also necessary to use MR images, in addition to CT, to help delineate surrounding tumors and soft tissues. Furthermore, considering the ionizing radiation of CT/CBCT scanning and the development of computer vision techniques, a strategy of MRI-only treatment planning has become increasingly valuable and feasible in the field of 3D VSP [5].

2.3.2. IMAGE DATABASE

Among the reviewed publications, there are four databases of H&N images that are publicly available. The Public Domain Database for Computational Anatomy (PDDCA) (<http://www.imagenglab.com/newsite/pddca/>) [26], an open-access resource of medical images for cancer research, consists of 48 patient CT images from the Radiation Therapy Oncology Group (RTOG) 0522 study of Head-Neck Cetuximab [27] database of The Cancer Imaging Archive (TCIA) [28], together with manual annotation of the mandible, brainstem, etc. [26]. The organizer of PDDCA provides 40 CT scans with mandible annotations. This dataset has been used for the Head and Neck Auto Segmentation MICCAI Challenge (2015) [26]. According to the protocol of the Challenge, 25 of 40 scans (0522c0001-0522c0328) are employed as the training dataset, and the other 15 scans (0522c0555-0522c0878) are employed as the test dataset [26]. 'Organ-at-risk segmentation from head & neck CT scans' is one of four tasks in StructSeg 2019 (<https://structseg2019.grand-challenge.org/Dataset/>), consisting of 50 nasopharynx cancer patient CT scans with several manually delineated organs at risk including mandibles (left and right). The StructSeg 2019 dataset has been used for the Automatic Structure Segmentation for Radiotherapy Planning Challenge 2019. Each of the annotated CT scans is marked by one experienced oncologist and verified by another experienced oncologist.

Furthermore, a dataset has been augmented or combined into new publicly available databases by Tang et al. [29] in 2019, for example, the manual delineations of 28 OARs in 35 CT scans from the Head-Neck Cetuximab and 105 CT scans from Head-Neck-PET-CT [30] databases (<https://github.com/uci-cbcl/UaNet#Data>). A subset of 31 CT scans, which was used for the test and validation set in the study from Nikolov et al. [24], was collected from two datasets of The Cancer Genome Atlas Head-Neck Squamous Cell Carcinoma [31] and Head-Neck Cetuximab [27] from TCIA. The subset with the ground truth added is available at <https://github.com/deepmind/tcia-ct-scan-dataset>.

2.3.3. EVALUATION METRICS

In addition to the differences in the used dataset, the results are evaluated and presented in different ways in the reviewed papers. Moreover, there are no standard metrics for the segmentation evaluation, so different evaluation metrics are used to report the segmentation performance. For segmentation, the evaluation metrics are mainly divided into three categories: overlap-based metrics, distance-based metrics, and volume-based metrics. The metrics are summarized in Table 2.1. The overlap-based metrics indicate the difference in overlap measurement between automatic prediction and manual segmentation, which can be obtained by four indicators: true positive (TP), false positive (FP), false negative (FN) and true negative (TN). The TP value is the number of pixels segmented correctly as foreground. The count of pixels falsely classified as the foreground is given by the FP value. The total count of falsely classified as background pixels is represented by FN. The TN value represents the correctly classified background pixels. The most commonly used overlap-based metrics include the Dice similarity coefficient (Dice), Sensitivity (Sen), false positive volume fraction (FPVF) [32], false negative volume fraction (FNVF) [32], etc. To measure the contour difference between automatic and manual segmentation, the most commonly used metrics are distance-based metrics. In

the context of mandibular segmentation, the following distance-based metrics have been frequently used: average symmetric surface distance (ASD), Hausdorff distance (HD), 95th-percentile Hausdorff distance (95HD), mean square error (MSE) [33], and root mean square error (RMSE) [33]. In some medical image segmentation tasks, the volume of the object is also very important for treatment planning, and the metric based on volume is helpful to evaluate the performance of the segmentation method. Volume overlap error and volume error are the common indices to evaluate the results of mandibular segmentation.

Table 2.1 | Performance metrics applied to performance measurement of automatic segmentation of the mandible and the corresponding references and mathematical definitions. Note: Y_r indicates the pixels in the reference standard (ground truth), and Y_p is the pixels in the automatic segmentation. $|\cdot|$ represents the number of voxels. $\|\cdot\|$ represents the L_2 norm. a and b are corresponding points on the boundary of A and B .

Metric	Abbreviation	Definition
Overlap-based metrics, reported in percents (%)		
Dice similarity index	Dice	$Dice = \frac{2 Y_r \cap Y_p }{ Y_r + Y_p } = \frac{2TP}{2TP + FP + FN}$
Sensitivity	Sen	$Sen = \frac{ Y_r \cap Y_p }{ Y_r } = \frac{TP}{TP + FN}$
Recall	Rec	$Recall = Sen$
Positive Predictive Value	PVV	$PVV = \frac{ Y_r \cap Y_p }{ Y_p } = \frac{TP}{TP + FP}$
Jaccard similarity coefficient	Jac	$Jac = \frac{ Y_r \cap Y_p }{ Y_r \cup Y_p } = \frac{TP}{TP + FP + FN}$
Intersection over union	IoU	$IoU = Jac$
Specificity	Spe	$Specificity = \frac{ (1 - Y_p) \cap (1 - Y_r) }{ 1 - Y_p } = \frac{TN}{TN + FP}$
False positive volume fraction	FPVF	$FPVF = \frac{ Y_p - Y_r }{ Y_r } = \frac{FP}{TP + FN}$
False negative volume fraction	FNVF	$FNVF = \frac{ Y_r - Y_p }{ Y_r } = \frac{FN}{TP + FN}$
Distance-based metrics, reported in millimeters (mm)		
Average symmetric surface distance	ASD	$ASD(A, B) = \frac{d(A, B) + d(B, A)}{2}$, where $d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \ a - b\ $,
Hausdorff distance	HD	$D_{HD}(A, B) = \max(h(A, B), h(B, A))$, where $h(A, B) = \max_{a \in A} \min_{b \in B} \ a - b\ $
95th-percentile Hausdorff distance	95HD	$95HD = \max(h^{95\%}(A, B), h^{95\%}(B, A))$, where $h^{95\%}(A, B) = \max_{a \in A} \min_{b \in B^{95\%}} \ a - b\ $
Mean square error	MSE	$MSE = \frac{1}{n} \sum_{i=1}^n \ A_i - B_i\ ^2$, where A_i is the boundary of the i -th OAR and B_i is the boundary of the i -th prediction.
Root mean square error	RMSE	$RMSE = \sqrt{MSE}$
Volume-based metrics, reported in percents (%)		
Volume overlap error	VOE	$1 - \frac{ Y_r \cap Y_p }{ Y_r \cup Y_p }$
Volume error	VE	$VE = \frac{Y_r - Y_p}{Y_r} \text{ or } \frac{Y_p - Y_r}{Y_r}$

2.3.4. METHODOLOGY

In this paper, the mandible segmentation approaches have been broadly divided into six categories, namely, (1) statistical shape model-based (SSM-based), (2) active shape model-based (ASM-based), (3) active appearance model-based (AAM-based), (4) atlas-based, (5) level set-based, (6) classical machine learning-based and (7) deep learning-based

approaches, as depicted in Table 2.2. There are several hybrid segmentation methods proposed in the collected literature. The hybrid methods are introduced in the listed seven categories.

Table 2.2 | Methodology applied to automatic mandible segmentation in the head and neck region, and the corresponding references.

Methodology categories	Publications	Number of publications
SSM-based	[18, 35–37]	4
ASM-based	[34, 38, 39]	3
AAM-based	[40, 41]	2
Atlas-based	[42–50, 86–94]	18
Level set-based	[9, 51]	2
Classical machine learning-based	[19, 33, 52–64]	15
Deep learning-based	[6, 24, 29, 65–79, 98, 99, 101–106, 110–114, 117, 119]	33

SSM-BASED, ASM-BASED AND AAM-BASED METHODS

Statistical shape model-based (SSM-based) segmentation algorithms take advantage of the prior shape information to extract the structures of the objects. A dataset of object shapes is needed for training in this technique. Procrustes alignment is applied to align the landmark point set of the objects, which is placed at key features and/or along the contour of the object shapes [80]. Principal component analysis (PCA) is employed to build a template shape of the objects [81].

The active shape model (ASM) [82] is the one of the commonly used image segmentation techniques based on the SSM approach. In 2006, Lamecker et al. [39] used a vanilla SSM-based segmentation method to extract the anatomical variability of developed mandible shapes in the training stage and match the statistical mandible shape to a given CT dataset via a deformable model approach in the testing stage. Kainmueller et al. [34] developed a segmentation method based on an SSM and a Dijkstra-based optimization for reconstructing the mandible including the course of the alveolar nerve. Albrecht et al. [38] adopted a multiatlas-based segmentation to obtain the coarse segmentation for the OARs and then employed the active shape model (ASM) to finely segment the OARs based on the results from coarse segmentation.

The active appearance model (AAM) [83] is an extension of SSM and ASM to further statistically model the texture information of the object. Two statistical models of shape and texture are further merged into an appearance model [83]. Babalola et al. [41] applied the AAM approach to automatic mandible and brainstem segmentation. Mannion-Haworth et al. [40] utilized the groupwise image registration method-based minimum description length approach [84] and AAM [83] built from manually segmented objects in CT images.

Moreover, several publications on SSM-based segmentation methods have been used in combination with other strategies to enhance the segmentation outcomes. Hybrid methods have demonstrated good results [18, 35–37]. Kainmueller et al. [37] used SSM adaptation for mandible segmentation and then utilized the graph cut algorithm for fine segmentation in the MICCAI 2019 challenge [85]. Gollmer et al. [35] applied the SSM method with optimized correspondence for mandible segmentation. The authors

established correspondence by optimizing a model-based cost function. Moreover, the authors introduced a relaxed SSM method for mandible segmentation [36]. Abdolali et al. [18] proposed a framework based on SSM and fast marching for automatic segmentation of the mandibular bone and canal. Moreover, the authors utilized low-rank decomposition for preprocessing. Table 2.3 presents the segmentation performance using SSM-based, ASM-based and AAM-based approaches.

Table 2.3 | Summary of SSM-based, ASM-based and AAM-based methods.

Study	Year	Datasets	No. of patients	Performance					Image modalities	Category
				Dice(%)	ASD(mm)	HD(mm)	95HD(mm)	VOE(%)		
Abdolali [18]	2017	In-house	120	91.38 ± 2.06	0.71 ± 0.09	—	—	—	CBCT	SSM-based
Gollmer [35]	2012	In-house	30	—	0.50 ± 0.00	11.30 ± 3.50	—	14.0 ± 1.9	CT	SSM-based
Gollmer [36]	2013	In-house	30+6	—	0.40 ± 0.08	10.26 ± 3.16	—	11.0 ± 2.0	CT(train)/ CBCT(test)	SSM-based
Kainmueller [37]	2009	MICCAI 2009	18	88.40	—	8.40	—	—	CT	SSM-based
Lamecker [39]	2006	In-house	15	—	—	—	—	—	CT	ASM-based
Albrecht [38]	2015	PDDCA	40	88.13 ± 5.55	—	—	2.83 ± 1.18	—	CT	ASM-based
Kainmueller [34]	2009	In-house	106	—	0.50 ± 0.10	6.20 ± 2.30	—	—	CBCT	ASM-based
Mannion-Haworth [40]	2015	PDDCA	48	92.67 ± 1.00	—	—	1.98 ± 0.59	—	CT	AAM-based
Babalola [41]	2009	MICCAI 2009	18	76.10 ± 5.10 (exclude the 13th case)	—	—	—	—	CT	AAM-based

ATLAS-BASED METHODS

Atlas-based segmentation methods first utilize deformable registration approaches to register a known reference segmentation mask (that form an atlas) to a patient [47]. The registration optimization problem is usually solved by searching the deformation space. Then, this deformation is applied to contours made on the atlas to project the contours back to the patient-space.

In 2007, Zhang et al. [47] used an atlas-based image segmentation method to automatically segment ROIs in H&N CT images. Chuang et al. [42] presented an atlas-based semiautomatic segmentation for the mandible. The authors firstly cropped a minimum enclosing box of the mandible in a raw CT image followed by a user-determined global threshold to remove nonosseous tissue and then used the output from these registrations to transform the 3D mandible template model of the respective template scans into a mandible model that maps to the input test scan. In another paper [86], a semisupervised registration algorithm based on the atlas was developed to accurately segment the OARs with a real ground contour and all other coarse OARs in the atlas. The method concatenates rigid and deformable blocks, takes an atlas image, a group of atlas-space segmentations and a patient image as inputs and outputs the patient-space segmentation of all OARs defined on the atlas. Qazi et al. [48] combined atlas registration and organ-specific model-based segmentation in a common framework to segment various organs at risk in H&N CT images. The authors applied atlas registration and organ-specific model-based segmentation at a global level and then used a probabilistic refinement step to refine at the voxel level. Ayyalusamy et al. [87] applied an atlas-based method for OAR segmentation in H&N and analyzed the dependence of atlas-based automatic segmentation for different levels on anatomy matching between sample patients and atlas patients.

In the case of mandible segmentation, the atlas-based method has become increasingly popular, so it has been frequently implemented in commercial or open-source tools. There are several articles using an atlas-based tool, for instance, Advanced Medical Imaging Registration Engine (ADMIRE) v1.05 (Elekta Software) used by [88], Smart Probabilistic Image Contouring Engine (SPICE) used by [89, 90], and PLASTIMATCH MABS provided by [91]. Moreover, La Macchia et al. [92] quantitatively analyzed three different automatic atlas-based segmentation software offerings for adaptive radiotherapy in a small dataset.

In regard to the fusion strategy, Mencarelli et al. [43] built a knowledge base and thresholding technique for nine high-contrast structures (including the mandible) for an online or offline image-guided radiotherapy (RT) application. Then, they further used a hidden Markov model to identify the structures. Wang et al. [46] developed a mandible and maxilla segmentation approach in an H&N CBCB scan that estimated a patient-specific atlas applying a sparse label fusion strategy from predefined CT atlases and then converted it into a convex optimization problem using maximum a posteriori probability for mandible/maxilla segmentation. Gorthi et al. [49] developed a novel segmentation method that used active contour-joint registration and an atlas selection strategy.

A multiple atlas strategy or its modified version is also commonly used because the propagated label cannot be generalized in conventional atlas-based models. To address this issue, multiatlas approaches have been proposed. Chen et al. [44] presented a multiatlas-based method for multiple structure segmentation in CT scans. The authors registered CT images with the atlases at the global level so that structures of interest are aligned approximately in space. Based on that approach, the multiatlas-based segmentation can be performed at the local level. A method presented by Haq et al. [93] dynamically selected and weighted the proper number of atlases for weighted label fusion and predicted the segmentations and consensus maps, indicating voxel-wise agreement between different atlases that were selected from those exceeding an alignment weight called the dynamic atlas attention index. Alignment weights were computed at the image level (called global weighted voting) or at the structure level (called structure weighted voting). McCarroll et al. [94] studied an automatic segmentation strategy for H&N cancer to better provide a fully automated radiation treatment planning solution for low- and middle-income countries using an atlas-based deformable-image-registration algorithm. Han et al. [45, 50] adopted a novel hierarchical multiatlas registration approach for OAR segmentation. Table 2.4 shows the summary of atlas-based method used.

LEVEL SET-BASED METHODS

The level set-based image segmentation algorithm is modified from the snake algorithm [95]. It requires an initial contour curve, and then, the curve evolution is performed via minimizing the functional energy. To obtain a better prediction, the studies [9, 51] make use of the atlas method to generate an initial contour in their segmentation tasks. Patch-based sparse representation has been applied to estimate the patient-specific atlas and then combined with a level set framework using the rule of maximizing a posteriori probability. With these methods, the mandible segmentation problem can be converted into a convex optimization problem [9]. Table 2.5 shows a summary of the level set-based method used.

Table 2.4 | Summary of atlas-based methods. ⁽¹⁾ Values estimated from figures; actual values not reported.

Study	Year	Datasets	No. of patients	Performance						Image modalities	Category
				Dice(%)	ASD(mm)	HD(mm)	95HD(mm)	VE(%)	Sen(%) / PPV(%)		
Chuang [42]	2019	In-house	54+20	97.60 ± 10.60	—	—	—	—	—	CT	Atlas-based
Mencarelli [43]	2014	In-house	188	—	—	—	—	—	—	CT	Atlas-based
Chen [44]	2015	PDDCA	40	91.70 ± 2.34	—	—	2.49 ± 0.76	—	—	CT	Atlas-based
Han [45]	2008	In-house	10	90.00	—	—	—	—	—	CT	Atlas-based
Wang [46]	2014	In-house	13+30	91.00 ± 2.00	0.61 ± 0.17	0.92 ± 0.47	—	—	—	CBCT+CT	Atlas-based
Zhang [47]	2007	In-house	7	80.00	0.85	—	—	—	—	CT	Atlas-based
Qazi [48]	2011	In-house	25	93.00	—	2.64	—	—	—	CT	Atlas-based
Gorthi [49]	2009	MICCAI 2009	18	77.80 ± 7.40	—	16.87 ± 6.75	—	—	—	CT	Atlas-based
Han [50]	2009	MICCAI 2009	18	90.33 ± 1.49	—	8.07 ± 3.12	—	—	—	CT	Atlas-based
Ayyalusamy [87]	2019	In-house	40	85.00 ⁽¹⁾	—	9 ⁽¹⁾	—	—	—	CT	Atlas-based
Haq [93]	2019	In-house	45	85.00 ⁽¹⁾	—	—	3.00 ⁽¹⁾	—5.00 ⁽¹⁾	—	CT	Atlas-based
		PDDCA	32	83.00 ⁽¹⁾	—	—	3.00 ⁽¹⁾	52.00 ⁽¹⁾	—		
Liu [88]	2016	In-house	6	89.50	—	—	—	—	—	CT	Atlas-based
Zhu [89]	2013	In-house	32	89.00 ± 4.00	—	9.80 ± 4.10	—	—	—	CT	Atlas-based
Walker [90]	2014	In-house	40	98.00 ± 2.00	—	—	—	—	—	CT	Atlas-based
McCarroll [94]	2018	In-house	128	84.00 ± 7.00	1.89 ± 1.55	18.63 ± 14.90	—	—	—	CT	Atlas-based
La Macchia [92]	2012	In-house	5	89.00 ± 2.00	—	—	—	−4.76 ± 7.12	87.00 ± 5.00 / 92.00 ± 2.00	CT	Atlas-based
Zaffino [91]	2016	In-house	25	88.00 ± 7.00	—	—	—	—	—	CT	Atlas-based
Huang [86]	2019	In-house	500	84.50 ± 1.60	—	—	—	—	—	CT	Atlas-based

Table 2.5 | Summary of level set-based methods.

Study	Year	Datasets	No. of patients	Performance			Image modalities	Category
				Dice(%)	ASD(mm)	HD(mm)		
Wang [9]	2014	In-house	15	92.00 ± 2.00	0.65 ± 0.19	0.96 ± 0.53	CBCT	Level set-based
Zhang [51]	2009	MICCAI 2009	18	87.93 ± 2.06	—	8.70 ± 3.28	CT	Level set-based

CLASSICAL MACHINE LEARNING-BASED METHODS

In the past few decades, the role of machine learning in medical applications has greatly increased. Classical machine learning (CML) methods include thresholding techniques, linear regression, support vector machines, random forests, etc. Many studies have used CML methods to segment the mandible, as listed in Table 2.6.

Normally, this technique is implemented in a hybrid way. A multiple threshold method by Otsu [96] was used to calculate proper thresholding values in the segmentation phase in the article from Barandiaran et al. [55]. After thresholding, region growing was applied so that many bone structures could be segmented, such as parts of the jaw bone or spine. The largest connected component of the segmented image volume can be simply selected for mandible segmentation [55]. Later, Wang et al. [53] made use of random forests to segment the mandible and maxilla in their expended CBCT database and achieved better results than their previous method [9]. Linares et al. [19] proposed a semiautomatic two-stage segmentation method for CBCT. The authors first performed bone segmentation using a supervoxel technique and then implemented graph clustering with an interactive user-placed seed process that was used for mandible and skull segmentation. In this way, supervoxel methods can reduce the excessive number of voxels of CBCT 3D volume, leading to shorter processing time. Orbes-Arteaga et al. [56] introduced a new patch-based label fusion approach to weight the label votes using a generative probabilistic approach, in which local classifiers from the dictionary of atlas patches were generated for weighting the probability of the target. Wang et al. [57] integrated shape priors with a hierarchical

learning model, in which they introduced 3 novel strategies: hierarchical critical model vertex identification, joint learning of shape and appearance, and hierarchical vertex regression. Qazi et al. [58] applied point-based registration and model-based segmentation and introduced a novel voxel classification approach to improve the results from model-based segmentation. Torosdagli et al. [59] adopted random forest regression to detect the mandible in 3D and then utilized a 3D gradient-based fuzzy connectedness image segmentation algorithm for final segmentation. Wu et al. [60] built population fuzzy anatomy models to find the hierarchical relationship of the objects. Then, the authors further utilized the anatomical information in the hierarchical relationship to locate the objects. Finally, the segmentation could be generated according to the location results. Tam et al. [33] presented support vector regression to predict all the boundary points of the organs of interest. Gacha et al. [63] performed histogram equalization and grayscale morphology for mandible segmentation. Then, image binarization was performed after grayscale method. Finally, 3D template matching was performed using cross-correlation to find a point inside the mandible followed by a geodesic dilation using the binary image as mask. Tong et al. [61] introduced new features into the automatic anatomy recognition approach from [54] in which they combined texture and intensity information into the recognition procedure. Wu et al. [62] developed a methodology, called AAR-RT, which extended their previous AAR framework [54] to RT planning in H&N. Spampinato et al. [64] applied a fully automatic mandible segmentation method to support orthodontists in assessing facial asymmetry, in which they used morphological operations for extraction of the mandible, and then, the connected components were computed and considered part of the mandible in each 2D slice.

Along with the quick development of computer and MRI imaging technology, MRI images have also been used in mandible segmentation. Ji et al. [52] presented a two-stage rule-constrained seedless region growing approach for the mandible segmentation in MRI images. A 3D/2D seedless region growing approach is used to detect a trabecular bone and cortical bone of mandible after using a thresholding technique. The mandibular body was finally merged with some morphological processes. Udupa et al. [54] developed an automatic anatomy recognition methodology in H&N MRI scans that is based on fuzzy modeling ideas and tightly integrated fuzzy models with the iterative relative fuzzy connectedness delineation algorithm to locate and delineate organs in different body scans.

DEEP LEARNING-BASED METHODS

In recent years, deep learning-based segmentation algorithms have become popular among researchers. With the development of deep learning technology, deep learning methods have shown a tremendous performance in the area of image segmentation. Deep learning techniques provide more flexibility and powerful capabilities than the traditional machine learning methods and require less expert analysis, facilitating extension to the other segmentation tasks [97]. For instance, the popular deep learning architectures, including UNet [13], SegNet [12], etc., are widely used for automatic image segmentation. Consequently, the deep learning approach has shown promising applications in the problem of mandible segmentation. Over the years, many studies have provided better solutions using deep learning architectures than other classical image segmentation methods, as shown in Table 2.7-2.11.

Table 2.6 | Summary of classical machine learning-based methods.

Study	Year	Datasets	No. of patients	Performance					Image modalities	Category	
				Dice(%)	ASD(mm)	HD(mm)	Jac(%)	FPVF/FNVF(%)			MSE/RMSE(mm)
Ji [52]	2013	In-house	12	97.90 ± 1.10	0.20 ± 0.13	—	95.80 ± 2.00	—	—	MRI	CML
Wang [53]	2015	In-house	30	94.00 ± 2.00	0.42 ± 0.15	0.74 ± 0.25	—	—	—	CBCT	CMI
			60	95.00 ± 2.00	0.33 ± 0.11	0.41 ± 0.20	—	—	—	CT	
Udupa [54]	2014	In-house	15	—	—	3.30 ± 0.56	—	1.00 ± 0.00/ 49.00 ± 8.00	—	MRI	CML
Linares [19]	2019	In-house	16	92.88	—	—	86.48	—	—	CBCT	CML
Barandiaran [55]	2009	In-house	12	—	—	—	—	—	—	CT	CML
Orbes-Arteaga [56]	2015	PDDCA	40	93.08 ± 2.36	—	—	—	—	—	CT	CML
Wang [57]	2016	PDDCA	48	94.40 ± 1.30	0.43 ± 0.12	—	—	—	—	CT	CML
Qazi [58]	2010	In-house	25	90.19	—	—	—	—	—	CT	CML
Torosdagli [59]	2017	PDDCA	40	91.00	—	< 1.00	—	—	—	CT	CML
Wu [60]	2018	In-house	216	89.00	—	1.60	—	—	—	CT	CML
Tam [33]	2018	In-house	56	85.20 ± 5.30	—	—	—	—	0.10/ 3.16	CT	CML
Tong [61]	2018	In-house	246	—	—	—	—	—	—	CT	CML
Wu [62]	2019	In-house	216	89.00	—	—	—	—	—	CT	CML
Gacha [63]	2018	PDDCA	30	80.49	—	—	—	—	—	CT	CML
Spampinato [64]	2012	In-house	10	—	—	—	—	—	—	CT	CML

Ibragimov et al. [77] proposed the first attempt of using the CNN-based concept to segment OARs in H&N CT scans. Furthermore, the authors smoothed the obtained segmentation results through a Markov random fields algorithm. To address the challenge of low performance on small low-contrast structures in CT, a new loss function, namely, the batch soft Dice loss function, was developed by Kodym et al. [98]. The authors used the new loss function in UNet and achieved better results than the methods using other loss functions. Yan et al. [79] developed a symmetric convolutional neural network to force convolution and deconvolution computation to be symmetric. Table 2.7 shows the summary of using deep learning-based methods.

Table 2.7 | Summary of deep learning-based methods.

Study	Year	Datasets	No. of patients	Performance				Image modalities	Category
				Dice(%)	ASD(mm)	HD(mm)	IoU(%)		
Ibragimov [77]	2015	In-house	50	89.50 ± 3.60	—	—	—	CT	DL
Kodym [98]	2019	PDDCA	35	94.60 ± 0.70;	0.29 ± 0.03	—	—	CT	DL
Yan [79]	2018	In-house	93	90.76 ± 2.45	—	—	85.44 ± 3.99	CT	DL
Xue [99]	2021	PDDCA	48	94.00 ± 2.00	0.49 ± 0.18	2.36 ± 0.62;	—	CT	DL

Another intuitive method of 3D segmentation is to train a 3D network to process volume data directly. In this way, the 3D segmentation network can extract 3D structure information. AnatomyNet [67] is built in the 3D U-net architecture using residual blocks in encoding layers. Moreover, a new loss function combining the Dice score and focal loss is applied in the training process. Xia et al. [68] presented a methodology to automatically measure the masseter thickness and locate the ideal injection point for botulinum toxin into the masseter from a CT scan, in which a 3D UNet with a Resblock is used for the mandible and masseter segmentation. Willems et al. [72] applied the 3D segmentation network from [100]. The method consists of a 3D CNN and fully connected 3D CRF. Ren et al. [74] introduced multiscale patches as input for representing the center voxel and designed a 3D CNN network for segmentation of OARs. In addition, the authors inter-

leaved the CNNs designated for the individual tissues since the neighboring tissues are often related on the anatomical side. With this strategy, the patch segmentation result of a specific tissue can be refined by the other neighboring tissues. This study can deal with large tissues, while the small tissues are still not that promising in terms of Dice and 95HD. For contending with the imbalance between classes mainly caused by small organs, He et al. [101] proposed a combined UNet model that consists of 2D UNet, 3D UNet and 3D small UNet. The first model is a 2D UNet model, which has advantages in processing thick slice images. The second model is a 3D UNet model, which can cover most organs with the original resolution in the transverse plane via clipping the scans. The third model is a 3D small UNet model, which focuses on the segmentation of small organs, clipping from the boundary box of the 2D UNet model. Rhee et al. [102] proposed a CNN-based autocontouring tool that can be used to detect the errors in autocontours from a clinically validated atlas-based autocontouring tool. They used the DeepMind model [24] to detect the ill-defined contours from an atlas-based autocontouring tool. Nikolov et al. [24] applied a 3D UNet architecture for OAR segmentation and demonstrated that the 3D UNet approach achieved expert-level performance in delineating multiple organ segmentation in the H&N region. Tong et al. [103] proposed a novel deep learning-based approach based on a generative adversarial network (GAN) with a shape constraint (SC-GAN), where a DenseNet was used to predict the segmentation and a CNN-based discriminator network was applied to correct the predicted errors and image-level inconsistency between the predicted image and ground truth. Moreover, a shape representation loss was applied into the segmentation and adversarial loss function for reducing the false positives. Chan et al. [65] provided a lifelong learning protocol to improve the prediction accuracy of OAR segmentation, in which they used a multitask learning scheme coupled with transfer learning to accomplish this. In this way, the network transfers its shared knowledge to single tasks to help improve the generalizability of the network on limited datasets. To cope with the lack of overall shape and smoothness of OARs, Xue et al. [104] developed the feasibility of learning the signed distance map (SDM) directly from medical scans, in which SDM is usually calculated from object boundary contours and the binary segmentation map. To utilize more information in traditional segmentation training, the authors introduced an approximated Heaviside function to train the model by predicting SDMs and segmentation maps simultaneously. Table 2.8 lists the summary of literature that used 3D networks for mandible segmentation.

Table 2.8 | Summary of 3D network strategies of deep learning-based methods.

Study	Year	Datasets	No. of patients	Performance						Image modalities	Category
				Dice(%)	ASD(mm)	HD(mm)	95HD(mm)	Rec/Sen(%)	RMSE(mm)	PPV(%)	
Chan [65]	2019	In-house	200	91.00 ± 9.00	—	—	—	—	0.66 ± 0.31	—	CT
Zhu [67]	2019	PDDCA+TCIA	48+223	92.30	—	—	—	—	—	—	CT
Xia [68]	2019	In-house	53	94.60 ± 1.10	0.24 ± 0.034	—	—	95.60 ± 1.20	—	—	CT
Willem's [72]	2018	In-house	70	95.90	0.60	6.48	—	—	—	—	CT
Ren [74]	2018	PDDCA	48	92.00;	—	—	1.89	—	—	—	CT
He [101]	2020	StructSeg2019	50	90.30(L); 90.80(R)	—	—	—	—	—	—	CT
Rhee [102]	2019	In-house+TCIA	1403+24	86.80 ± 3.30	—	12.80 ± 9.50	—	—	—	—	CT
Nikolov [24]	2018	In-house	459	93.10 ± 1.90	—	—	—	—	—	—	CT
		TCIA	30	92.90 ± 3.50	—	—	—	—	—	—	
		PDDCA	15	93.80 ± 1.90	—	—	—	—	—	—	
Tong [103]	2019	PDDCA	32	93.91 ± 1.30	0.55 ± 0.14	—	2.09 ± 0.63	91.25 ± 2.70	—	96.82 ± 1.70	CT
		In-house	25	81.64 ± 4.44	1.13 ± 0.48	—	2.72 ± 1.31	86.65 ± 6.00	—	78.50 ± 4.30	
Xue [104]	2019	PDDCA	48	95.70 ± 1.80;	—	—	0.60 ± 0.49	—	—	—	CT

However, 3D networks require clipping or resampling the input data into small 3D patches to reduce GPU memory [78]. In addition, integration of 3D patches in post-

preprocessing is a time-consuming workflow. Therefore, to overcome the drawback of 3D networks, using a 2.5D or multiview training strategy instead of 3D is a potential strategy [6, 105]. Qiu et al. [6] adopted 2.5D volume of CTs as input in UNet and then combined the resulting 2D segmentations from three orthogonal planes (axial, sagittal and coronal) into a 3D segmentation. This architecture can take into account the spatial information of adjacent slices in order to preserve the connectivity of anatomical structures. Lei et al. [105] first presented a segmental linear function to enhance the intensity of CT images, which makes the organs more separable than the existing simple window width method. Then, the authors proposed a novel 2.5D network for accurate OAR segmentation. Additionally, they applied a novel hardness-aware loss function to make the network give more attention to hard voxels. Liang et al. [106] developed a novel multiview (i.e., axial, coronal, and sagittal view) spatial aggregation framework for joint localization and segmentation of multiple OARs. Additionally, the authors proposed a region-of-interest-based fine-grained representation CNN for predicting the probability maps of OARs for each 2D view of CT images. In this way, the approach unifies the OAR localization and segmentation tasks and trains them in an end-to-end fashion. Qiu et al. [78] integrated a recurrent unit and a vanilla network, which enable the network to learn the continuity of neighborhood slices for the scans in the 2D architecture. Table 2.9 shows the summary of using 2.5D networks for mandible segmentation.

Table 2.9 | Summary of 2.5D networks deep learning-based methods.

Study	Year	Datasets	No. of patients	Performance				Image modalities	Category
				Dice(%)	ASD(mm)	95FD(mm)	RMSE(mm)		
Qiu [6]	2019	In-house	109	88.10	—	—	0.58	CT	DL
		PDDCA	40	93.28 ± 1.44	—	1.43 ± 0.56	—		
Lei [105]	2021	StructSeg2019	50	91.10 ± 2.90(L); 91.70 ± 1.50(R)	—	2.81 ± 0.45(L); 2.70 ± 0.40(R)	—	CT	DL
		PDDCA+ In-house	50+48+67	90.00 ± 4.20	—	6.54 ± 19.14	—		
Liang [106]	2020	PDDCA	48	94.10 ± 0.70	0.28 ± 0.14	—	—	CT	DL
		In-house	96	91.10 ± 1.00(L); 91.40 ± 2.00(R)	0.76 ± 0.13(L); 0.86 ± 0.14(R)	—	—		
Qiu [78]	2020	In-house	109	97.53 ± 1.65	0.21 ± 0.26	2.40 ± 4.61	—	CT	DL
		PDDCA	40	95.10 ± 1.21	0.14 ± 0.04	1.36 ± 0.45	—		

The attention model has been widely used in various fields of deep learning in recent years [107–109]. The attention mechanism focuses limited attention on key information to save resources and obtain the most effective information quickly [108]. Gou et al. [110] designed a self-channel-and-spatial-attention neural network for H&N OAR segmentation, in which spatial and channelwise attention learning mechanisms can adaptively force the network to emphasize the meaningful features and weaken the irrelevant features at the same time. A novel simple local blockwise self-attention-based segmentation approach was presented by Jiang et al. [111]. This approach can more easily make the network learn the spatial location and interrelation within input images. This study demonstrated that adding the additional attention blocks increases the contextual field and captures focused attention from anatomical structures. Sun et al. [112] presented an end-to-end CNN, called AttentionAnatomy, which was trained with three partially annotated datasets to segment OARs from the whole body. AttentionAnatomy retains the basis structure of UNet and has two branches: a CT region classification path and an OAR segmentation path. The CT region classification path outputs region prediction and attention vectors for OARs, which represent the inference of possible combinations of OARs in the current image. Then, the OAR segmentation path applies these attention

vectors to modulate the final output mask. The authors further proposed a recalibration mechanism to solve the partial annotation problem. Moreover, a hybrid loss function composed of batch Dice loss and spatially balanced focal loss is used to address the extreme category imbalance. Liu et al. [73] developed a cross-layer spatial attention map fusion network to combine different spatial attention maps and establish connections for significant features in the feature maps. Furthermore, the authors adopted a top-k exponential logarithmic Dice loss for the imbalanced dataset in their segmentation tasks. Table 2.10 shows the summary of using attention strategies for mandible segmentation.

Table 2.10 | Summary of attention strategies of deep learning-based methods.

Study	Year	Datasets	No. of patients	Performance					Image modalities	Category
				Dice(%)	ASD(mm)	95HD(mm)	Rec/Sen(%)	PPV(%)		
Gou [110]	2020	PDDCA	48	94.00 ± 1.00	0.47 ± 0.11	1.40 ± 0.02	93.00 ± 2.00	95.00 ± 2.00	CT	DL
Jiang [111]	2019	In-house+PDDCA	48+48	93.00 ± 1.00 (trained in in-house dataset)	—	—	—	—	CT	DL
Sun [112]	2020	In-house	129	94.05 ± 1.08	—	—	—	—	CT	DL
Liu [73]	2020	StructSeg2019	50	90.28(L); 90.81(R)	—	—	—	—	CT	DL

There are several two-stage training strategies, such as cascaded CNNs [66, 69, 70, 113], the combination [75, 114] of Faster R-CNN [115] and CNNs. Zhang et al. [66] developed a cascaded CNN architecture for multitasks segmentation in H&N scans, in which a slice classification network is proposed to classify CT slices into the corresponding target categories. Then, the irrelevant slices without the targets were excluded. Next, the slices in the corresponding categories were pushed to a refined 3D segmentation network for target segmentation. In this case, the prediction performance was further improved with the help of the slice classification network. Tappeiner et al. [69] trained two hierarchical 3D neural networks to segment multiple OARs in H&N CT scans. The authors first implemented a coarse network on size-reduced scans for locating the OARs. Then, a subsequent fine network on the original resolution scans was trained for a final accurate segmentation based on the results from the coarse network. Mu et al. [70] applied two cascade networks for location and fine segmentation of organ segmentation, in which the network is modified based on the squeeze and exception module, the residual module and V-Net [116]. There are still some shortcomings in this research. The network input is accomplished by resampling the input images with the same voxel size. However, the difference of the thickness and the resolution within the slices is often too large. Therefore, the operation of resampling in the first network can easily cause localization failure when organs have small volumes. Wang et al. [71] proposed a framework similar to [70]. Tang et al. [29] presented a novel network U_a -Net which contains detection and segmentation stages. The approximate location of the OARs can be identified in the detection stage, and then, fine segmentation can be further performed utilizing the results from the detection stage as a guide in the fine stage. Lei et al. [75, 76] adopted a 3D Faster R-CNN to locate the H&N OARs in MRI and CT images and then applied an attention UNet to segment the OARs based on the location results. Liang et al. [114] developed an automatic detection-segmentation network for OAR segmentation in H&N CT scans, in which the authors applied a Faster R-CNN [115] to locate the OARs and a fully convolutional network (FCN) [11] to further predict OAR masks based on the organ bounding boxes from the detection stage. Dijk et al. [117] applied a deep learning-based method from Mirada for OAR segmentation in H&N, in which the method was applied in the AAPM Challenge 2017

[118]. The method first used a CNN network to predict all OARs at a coarse resolution and then applied a series of organ-specific 10-layer networks for each organ segmentation using the predicted coarse results with full resolution images as input. Men et al. [113] developed a cascade CNN for the OAR delineation for radiotherapy in head-and-neck squamous cell carcinoma (HNSCC) of TCIA [30]. The cascaded CNN contains two stages (i.e., a coarse detection stage and a fine segmentation stage). The authors first used a shallow network as coarse detection to find the ROIs. In the fine segmentation stage, they used the coarse prediction as input for fine segmentation. Egger et al. [119] utilized a VGG network to classify the slices where the mandible appears and then employed an FCN network to delineate the mandible. Table 2.11 shows the summary of using two-stage strategies for mandible segmentation.

A new automatic segmentation algorithm was presented by Xue et al. [99] that employs a novel hybrid neural-like P system to alleviate the challenges in tasks of segmentation of H&N scans. The new P system has the common advantages of cell-like and neural-like P systems so that it can solve more practical problems in parallel. In the new P system, the effective integrated CNNs are implemented at the same time through different initializations to perform the pixel-level segmentation of OARs in order to obtain more effective features and take advantage of the strength of ensemble learning.

Table 2.11 | Summary of two-stage strategies in deep learning-based methods.

Study	Year	Datasets	No. of patients	Performance			Image modalities	Category
				Dice(%)	HD(mm)	95HD(mm)		
Zhang [66]	2021	In-house	170	89.00 ± 2.00	—	1.66 ± 0.51	CT	DL
Tappeiner [69]	2019	PDDCA	40	91.00 ± 2.00	—	2.4 ± 0.6	CT	DL
Mu [70]	2020	In-house	50	$89.80 \pm 2.70(L)$; $90.40 \pm 2.00(R)$		—	CT	DL
Wang [71]	2018	PDDCA	48	93.00 ± 1.90	—	1.26 ± 0.50	CT	DL
Tang [29]	2019	In-house	175	93.12 ± 1.41	—	2.48 ± 0.83	CT	DL
		HNC+ HNPETCT	35+105	89.31 ± 11.59		3.05 ± 2.60		
		PDDCA	48	95.00 ± 0.80		—		
Lei [75]	2020	In-house	15	85.00 ± 4.00	—	—	MRI	DL
Lei [76]	2020	In-house	15	88.00 ± 3.00	—	—	CT	DL
Liang [114]	2019	In-house	185	$91.40 \pm 0.04(L)$; $91.20 \pm 3.00(R)$		—	CT	DL
Dijk [117]	2020	In-house	693	94.00 ± 1.00	—	1.30 ± 0.50	CT	DL
Men [113]	2019	HNSCC	100	92.00 ± 2.00	2.40 ± 0.40	—	CT	DL
Egger [119]	2018	In-house	20	89.64 ± 1.69	—	—	CT	DL

2.4. DISCUSSION

Numerous studies have been published regarding mandible segmentation. Mandible segmentation methods collected in this paper are classified into seven categories. As shown in Figure 2.3 (a), atlas-based (23.4%), CML-based (19.5%) and DL-based (42.9%) methods are frequently used and developed in mandible segmentation. More than three-quarters of the publications (85.8%) utilized these methods. The atlas-based method is the oldest technique in the field of mandible segmentation and has been frequently implemented in commercial software [88–92]. Low accuracy is reported by using this kind of method due to the image artifacts and low contrast. Atlas construction is a time-consuming task when conducting complex nonrigid registration. An extended approach is to apply level set methods that help the initial curve move towards the boundary of

the object [9, 51]. The level set function implicitly represents the boundary of a curve or a shape through its zero horizontal tangent plane. The main drawback of using the level set-based methods is the requirement to initialize the contour. Thus, using the contours generated from atlas-based methods as initialization is helpful in this task [9, 51]. The SSM-based method is another approach used by various researchers for mandible segmentation, in which the shape information of the object is taken into consideration in ASM. The method performed well compared to the atlas-based method. In addition to considering the shape information, AAM considers the texture information of the image and learns it in an integrated statistical shape model.

The most frequently used image segmentation approach is machine learning, especially deep learning (DL), which produces better results than the other methods. The classical machine learning algorithms demand a considerable human effort in feature extraction. The DL technique has been used frequently since 2018 due to the rapid development of DL techniques, as illustrated in Figure 2.3 (b). Deep learning methods have resolved the issue of automatically extracting the features. The major issue of using deep learning methods is the requirement of a substantial quantity of annotated medical data, which limits the use of supervised learning techniques and avoids overfitting [110]. Furthermore, to increase the size of training data, the deep learning-based methods are trained using data augmentation schemes such as scaling, cropping, affine transformation, elastic deformation, rotations, and noise contamination [65, 67, 69, 110]. Moreover, with the evolution of deep learning, mandible segmentation has achieved remarkable performance compared to classical segmentation approaches.

As shown in Figure 2.3 (c), most of the publications (86.4%) used the CT imaging modality. Very few publications worked with CBCT (8.6%) and MRI (4.9%). Moreover, most of the publications worked with only one single imaging modality. There are 4 publications that worked with two imaging modalities. As shown in Figure 2.3 (c), number of publications using CT modality is much more than that of using CBCT and MRI techniques, which need more effort for manual expert annotation. However, the CBCT and MRI imaging techniques involve very low-dose radiation or no radiation. Mandible segmentation in CBCT and MRI is worthy of further exploration. The ability to derive the exact shape of the mandible from MRI's nonionizing radiation imaging mode provides clinicians with a value-added visualization option to view hard and soft tissues in a visual environment. Especially when high-resolution bone information is not needed, the spatial relationship between soft tissue muscle and the mandibular body can be obtained without additional volume imaging using CBCT. Patients who receive MR imaging but do not require facial CT imaging can benefit from this approach without the need for CBCT or X-ray radiation from CT imaging. This application eliminates the radiation risk to patients [5].

The Head and Neck Auto Segmentation MICCAI Challenge (2015) [26], which used the PDDCA dataset in the challenge, has obtained much attention during the past years. There are 25 publications for mandible segmentation in PDDCA [26]. Among all the metrics summaries, the Dice score is the most popular metric used to evaluate the performance of the segmentation methods. The distribution regarding the publication with the PDDCA test is presented in Figure 2.3 (d). The Dice score increases with the year. Although it is a common metric used in image segmentation, it may not be the most relevant metric for

Figure 2.3 | Analysis of publications pertaining to mandible segmentation. (a) The share of methodology categories used in the publications. (b) Distribution of the publications by year. (c) The share of image modalities used in the publications. (d) Box plot of the distribution of the Dice score of the PDDCA test from the publications. The black lines in the middle of the box indicate the average Dice score.

clinical applications [67]. Identifying a new metric in consultation with the physicians practicing in the field would be an important next step for real clinical applications of the method [67].

2.5. CONCLUSION

Mandible segmentation is a challenging task due to the presence of metal artifacts, the low contrast of condyles and the large variation in mandibles among patients. To date, mandible segmentation methods proposed in the literature have achieved promising results in CT scans. Before the era of machine learning, the atlas-based segmentation methods performed exceptionally well. With the advancements in technology, the use of deep learning-based segmentation methods has increased. Even though the deep learning methods demand considerably more data and higher computational resources than conventional methods in training, they perform well when trained with optimal learning parameters. Moreover, the segmentation performance improves when used in conjunction with an optimal classical segmentation approach. Mandible segmentation

approaches have attained a remarkable performance over the years. There is still room for improvement since the datasets are limited and cannot fully represent the general patient population in clinic. Furthermore, mandible segmentation in CBCT and MRI is valuable of further exploration. Therefore, mandible segmentation is still an open research area to be improved.

REFERENCES

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray, *Global cancer statistics 2020: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries*, CA: A Cancer Journal for Clinicians **n/a**, <https://doi.org/10.3322/caac.21660>, <https://acsjournals.onlinelibrary.wiley.com/doi/pdf/10.3322/caac.21660>.
- [2] J. P. Shah and Z. Gil, *Current concepts in management of oral cancer–surgery*, Oral oncology **45**, 394 (2009).
- [3] L. Ciocca, S. Mazzoni, M. Fantini, F. Persiani, C. Marchetti, and R. Scotti, *Cad/cam guided secondary mandibular reconstruction of a discontinuity defect after ablative cancer surgery*, Journal of Cranio-Maxillofacial Surgery **40**, e511 (2012).
- [4] J. Kraeima, B. Dorgelo, H. Gulbitti, R. Steenbakkers, K. Schepman, J. Roodenburg, F. Spijkervet, R. Schepers, and M. Witjes, *Multi-modality 3d mandibular resection planning in head and neck cancer using ct and mri data fusion: A clinical series*, Oral oncology **81**, 22 (2018).
- [5] J. Kraeima, *Three dimensional virtual surgical planning for patient specific osteosynthesis and devices in oral and maxillofacial surgery. A new era.*, Ph.D. thesis, University of Groningen (2019).
- [6] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. van Ooijen, *Automatic segmentation of the mandible from computed tomography scans for 3d virtual surgical planning using the convolutional neural network*, Physics in Medicine & Biology **64**, 175020 (2019).
- [7] P. Metzler, E. J. Geiger, A. Alcon, X. Ma, and D. M. Steinbacher, *Three-dimensional virtual surgery accuracy for free fibula mandibular reconstruction: planned versus actual results*, Journal of Oral and Maxillofacial Surgery **72**, 2601 (2014).
- [8] T. J. Huff, P. E. Ludwig, and J. M. Zuniga, *The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning*, Expert review of medical devices **15**, 349 (2018).
- [9] L. Wang, K. C. Chen, Y. Gao, F. Shi, S. Liao, G. Li, S. G. Shen, J. Yan, P. K. Lee, B. Chow, et al., *Automated bone segmentation from dental cbct images using patch-based sparse representation and convex optimization*, Medical physics **41**, 043503 (2014).
- [10] T. Heimann and H.-P. Meinzer, *Statistical shape models for 3d medical image segmentation: a review*, Medical image analysis **13**, 543 (2009).
- [11] 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.

- [12] V. Badrinarayanan, A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*, IEEE transactions on pattern analysis and machine intelligence **39**, 2481 (2017).
- [13] 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* (Springer, 2015) pp. 234–241.
- [14] F. Yu and V. Koltun, *Multi-scale context aggregation by dilated convolutions*, arXiv preprint arXiv:1511.07122 (2015).
- [15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, *A review on deep learning techniques applied to semantic segmentation*, arXiv preprint arXiv:1704.06857 (2017).
- [16] M. Vaitiekūnas, D. Jegelevičius, A. Sakalauskas, and S. Grybauskas, *Automatic method for bone segmentation in cone beam computed tomography data set*, Applied Sciences **10**, 236 (2020).
- [17] V. Hirschinger, S. Hanke, U. Hirschfelder, and E. Hofmann, *Artifacts in orthodontic bracket systems in cone-beam computed tomography and multislice computed tomography*, Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie **76**, 152 (2015).
- [18] F. Abdolali, R. A. Zoroofi, M. Abdolali, F. Yokota, Y. Otake, and Y. Sato, *Automatic segmentation of mandibular canal in cone beam ct images using conditional statistical shape model and fast marching*, International journal of computer assisted radiology and surgery **12**, 581 (2017).
- [19] O. C. Linares, J. Bianchi, D. Raveli, J. B. Neto, and B. Hamann, *Mandible and skull segmentation in cone beam computed tomography using super-voxels and graph clustering*, The Visual Computer **35**, 1461 (2019).
- [20] U. Vovk, F. Pernus, and B. Likar, *A review of methods for correction of intensity inhomogeneity in mri*, IEEE transactions on medical imaging **26**, 405 (2007).
- [21] A. K. Bag, S. Gaddikeri, A. Singhal, S. Hardin, B. D. Tran, J. A. Medina, and J. K. Curé, *Imaging of the temporomandibular joint: An update*, World journal of radiology **6**, 567 (2014).
- [22] J. Wallner, I. Mischak, and J. Egger, *Computed tomography data collection of the complete human mandible and valid clinical ground truth models*, Scientific data **6**, 1 (2019).
- [23] M. Ouzzani, H. Hammady, Z. Fedorowicz, and A. Elmagarmid, *Rayyan—a web and mobile app for systematic reviews*, Systematic reviews **5**, 1 (2016).
- [24] S. Nikolov, S. Blackwell, R. Mendes, J. De Fauw, C. Meyer, C. Hughes, H. Askham, B. Romera-Paredes, A. Karthikesalingam, C. Chu, et al., *Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy*, arXiv preprint arXiv:1809.04430 (2018).

- [25] Z. Fourie, J. Damstra, R. H. Schepers, P. O. Gerrits, and Y. Ren, *Segmentation process significantly influences the accuracy of 3d surface models derived from cone beam computed tomography*, *European journal of radiology* **81**, e524 (2012).
- [26] P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant, T. Albrecht, T. Gass, C. Langguth, M. Lüthi, *et al.*, *Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015*, *Medical physics* **44**, 2020 (2017).
- [27] W. Bosch, W. Straube, J. Matthews, and J. Purdy, *Head-neck cetuximab. the cancer imaging archive. 2015*, <https://wiki.cancerimagingarchive.net/display/Public/Head-Neck+Cetuximab>.
- [28] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, *et al.*, *The cancer imaging archive (tcia): maintaining and operating a public information repository*, *Journal of digital imaging* **26**, 1045 (2013).
- [29] H. Tang, X. Chen, Y. Liu, Z. Lu, J. You, M. Yang, S. Yao, G. Zhao, Y. Xu, T. Chen, *et al.*, *Clinically applicable deep learning framework for organs at risk delineation in ct images*, *Nature Machine Intelligence* **1**, 480 (2019).
- [30] M. Vallieres, E. Kay-Rivest, L. J. Perrin, X. Liem, C. Furstoss, H. J. Aerts, N. Khaouam, P. F. Nguyen-Tan, C.-S. Wang, K. Sultanem, *et al.*, *Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer*, *Scientific reports* **7**, 1 (2017).
- [31] M. L. Zuley, R. Jarosz, S. Kirk, Y. Lee, R. Colen, K. Garcia, and N. Aredes, *Radiology data from the cancer genome atlas head-neck squamous cell carcinoma [tcga-hnsc] collection*, *Cancer Imaging Archive* (2016).
- [32] X. Fan, J. Yang, and L. Cheng, *A novel segmentation method for mr brain images based on fuzzy connectedness and fcm*, in *International Conference on Fuzzy Systems and Knowledge Discovery* (Springer, 2005) pp. 505–513.
- [33] C. Tam, X. Yang, S. Tian, X. Jiang, J. Beitler, and S. Li, *Automated delineation of organs-at-risk in head and neck ct images using multi-output support vector regression*, in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 10578 (International Society for Optics and Photonics, 2018) p. 1057824.
- [34] D. Kainmueller, H. Lamecker, H. Seim, M. Zinser, and S. Zachow, *Automatic extraction of mandibular nerve and bone from cone-beam ct data*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2009) pp. 76–83.
- [35] S. T. Gollmer and T. M. Buzug, *Fully automatic shape constrained mandible segmentation from cone-beam ct data*, in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* (IEEE, 2012) pp. 1272–1275.

- [36] S. T Gollmer and T. M Buzug, *Relaxed statistical shape models for 3d image segmentation—application to mandible bone in cone-beam ct data*, Current Medical Imaging **9**, 129 (2013).
- [37] D. Kainmueller, H. Lamecker, H. Seim, and S. Zachow, *Multi-object segmentation of head bones*, MIDAS Journal (2009).
- [38] T. Albrecht, T. Gass, C. Langguth, and M. Lüthi, *Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning*, Head and Neck Auto-Segmentation Challenge (MICCAI), Munich (2015).
- [39] H. Lamecker, S. Zachow, A. Wittmers, B. Weber, H. Hege, B. Isholtz, and M. Stiller, *Automatic segmentation of mandibles in low-dose ct-data*, International Journal of Computer Assisted Radiology and Surgery **1**, 393 (2006).
- [40] R. Mannion-Haworth, M. Bowes, A. Ashman, G. Guillard, A. Brett, and G. Vincent, *Fully automatic segmentation of head and neck organs using active appearance models*, MIDAS J. (2015).
- [41] K. Babalola and T. Cootes, *Aam segmentation of the mandible and brainstem*, in *Proc. of the 3rd Workshop on 3D Segmentation in the Clinic: MICCAI* (2009).
- [42] Y. J. Chuang, B. M. Doherty, N. Adluru, M. K. Chung, and H. K. Vorperian, *A novel registration-based semi-automatic mandible segmentation pipeline using computed tomography images to study mandibular development*, Journal of computer assisted tomography **42**, 306 (2018).
- [43] A. Mencarelli, S. Van Beek, L. Zijp, C. Rasch, M. Van Herk, and J. Sonke, *Automatic detection system for multiple region of interest registration to account for posture changes in head and neck radiotherapy*, Physics in Medicine & Biology **59**, 2005 (2014).
- [44] A. Chen and B. Dawant, *A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images*, Head and Neck Auto-Segmentation Challenge (MICCAI), Munich (2015).
- [45] X. Han, M. S. Hoogeman, P. C. Levendag, L. S. Hibbard, D. N. Teguh, P. Voet, A. C. Cowen, and T. K. Wolf, *Atlas-based auto-segmentation of head and neck ct images*, in *International Conference on Medical Image Computing and Computer-assisted Intervention* (Springer, 2008) pp. 434–441.
- [46] L. Wang, K. C. Chen, F. Shi, S. Liao, G. Li, Y. Gao, S. G. Shen, J. Yan, P. K. Lee, B. Chow, *et al.*, *Automated segmentation of cbct image using spiral ct atlases and convex optimization*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2013) pp. 251–258.
- [47] T. Zhang, Y. Chi, E. Meldolesi, and D. Yan, *Automatic delineation of on-line head-and-neck computed tomography images: toward on-line adaptive radiotherapy*, International Journal of Radiation Oncology* Biology* Physics **68**, 522 (2007).

- [48] A. A. Qazi, V. Pekar, J. Kim, J. Xie, S. L. Breen, and D. A. Jaffray, *Auto-segmentation of normal and target structures in head and neck ct images: A feature-driven model-based approach*, Medical physics **38**, 6160 (2011).
- [49] S. Gorthi, V. Duay, M. Bach Cuadra, P.-A. Tercier, A. S. Allal, and J.-P. Thiran, *Active contour-based segmentation of head and neck with adaptive atlas selection*, in *MICCAI workshop on 3D segmentation challenge for clinical applications*, CONF (2009).
- [50] X. Han, L. Hibbard, N. O'Connell, and V. Willcut, *Automatic segmentation of head and neck ct images by gpu-accelerated multi-atlas fusion*, in *Proceedings of Head And Neck Auto-Segmentation Challenge Workshop*, Vol. 219 (2009).
- [51] X. Zhang, J. Tian, Y. Wu, J. Zheng, and K. Deng, *Segmentation of head and neck ct scans using atlas-based level set method*, in *Proceedings of Head And Neck Auto-Segmentation Challenge Workshop* (2009) p. 56.
- [52] D. X. Ji, K. W. C. Foong, and S. H. Ong, *A two-stage rule-constrained seedless region growing approach for mandibular body segmentation in mri*, International journal of computer assisted radiology and surgery **8**, 723 (2013).
- [53] L. Wang, Y. Gao, F. Shi, G. Li, K.-C. Chen, Z. Tang, J. J. Xia, and D. Shen, *Automated segmentation of dental cbct image with prior-guided sequential random forests*, Medical physics **43**, 336 (2016).
- [54] J. K. Udupa, D. Odhner, L. Zhao, Y. Tong, M. M. Matsumoto, K. C. Ciesielski, A. X. Falcao, P. Vaideeswaran, V. Ciesielski, B. Saboury, *et al.*, *Body-wide hierarchical fuzzy modeling, recognition, and delineation of anatomy in medical images*, Medical image analysis **18**, 752 (2014).
- [55] I. Barandiaran, I. Macía, E. Berckmann, D. Wald, M. P. Dupillier, C. Paloc, and M. Grana, *An automatic segmentation and reconstruction of mandibular structures from ct-data*, in *International Conference on Intelligent Data Engineering and Automated Learning* (Springer, 2009) pp. 649–655.
- [56] M. Orbes-Arteaga, D. Pea, and G. Dominguez, *Head and neck auto segmentation challenge based on non-local generative models*, MIDAS J. (2015).
- [57] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, *Hierarchical vertex regression-based segmentation of head and neck ct images for radiotherapy planning*, IEEE Transactions on Image Processing **27**, 923 (2018).
- [58] A. A. Qazi, J. J. Kim, D. A. Jaffray, and V. Pekar, *Probabilistic refinement of model-based segmentation: application to radiation therapy planning of the head and neck*, in *International Workshop on Medical Imaging and Virtual Reality* (Springer, 2010) pp. 403–410.
- [59] N. Torosdagli, D. K. Liberton, P. Verma, M. Sincan, J. Lee, S. Pattanaik, and U. Bagci, *Robust and fully automated segmentation of mandible from ct scans*, in *Biomedical*

- Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on* (IEEE, 2017) pp. 1209–1212.
- [60] X. Wu, J. K. Udupa, Y. Tong, D. Odhner, G. V. Pednekar, C. B. Simone, D. McLaughlin, C. Apinorasethkul, J. Lukens, D. Mihailidis, *et al.*, *Auto-contouring via automatic anatomy recognition of organs at risk in head and neck cancer on ct images*, in *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*, Vol. 10576 (International Society for Optics and Photonics, 2018) p. 1057617.
- [61] Y. Tong, J. K. Udupa, X. Wu, D. Odhner, G. Pednekar, C. B. Simone, D. McLaughlin, C. Apinorasethkul, G. Shammo, P. James, *et al.*, *Hierarchical model-based object localization for auto-contouring in head and neck radiation therapy planning*, in *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 10578 (International Society for Optics and Photonics, 2018) p. 1057822.
- [62] X. Wu, J. K. Udupa, Y. Tong, D. Odhner, G. V. Pednekar, C. B. Simone II, D. McLaughlin, C. Apinorasethkul, O. Apinorasethkul, J. Lukens, *et al.*, *Aar-rt—a system for auto-contouring organs at risk on ct images for radiation therapy planning: Principles, design, and large-scale evaluation on head-and-neck and thoracic cancer cases*, *Medical image analysis* **54**, 45 (2019).
- [63] S. J. Gacha and S. A. G. León, *Segmentation of mandibles in computer tomography volumes of patients with foam cells carcinoma*, in *2018 IX International Seminar of Biomedical Engineering (SIB)* (IEEE, 2018) pp. 1–7.
- [64] C. Spampinato, C. Pino, D. Giordano, and R. Leonardi, *Automatic 3d segmentation of mandible for assessment of facial asymmetry*, in *2012 IEEE International Symposium on Medical Measurements and Applications Proceedings* (IEEE, 2012) pp. 1–4.
- [65] J. W. Chan, V. Kearney, S. Haaf, S. Wu, M. Bogdanov, M. Reddick, N. Dixit, A. Sudhyadhom, J. Chen, S. S. Yom, *et al.*, *A convolutional neural network algorithm for automatic segmentation of head and neck organs at risk using deep lifelong learning*, *Medical physics* **46**, 2204 (2019).
- [66] S. Zhang, H. Wang, S. Tian, X. Zhang, J. Li, R. Lei, M. Gao, C. Liu, L. Yang, X. Bi, *et al.*, *A slice classification model-facilitated 3d encoder–decoder network for segmenting organs at risk in head and neck cancer*, *Journal of Radiation Research* **62**, 94 (2021).
- [67] W. Zhu, Y. Huang, H. Tang, Z. Qian, N. Du, W. Fan, and X. Xie, *Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation*, arXiv preprint arXiv:1808.05238 (2018).
- [68] W.-J. Xia, W.-Q. Han, Y. Zhang, L. Wang, G. Chai, and Y.-J. Huang, *Automatic masseter thickness measurement and ideal point localization for botulinum toxin injection*, in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2019) pp. 2801–2804.

- [69] E. Tappeiner, S. Pröll, M. Hönig, P. F. Raudaschl, P. Zaffino, M. F. Spadea, G. C. Sharp, R. Schubert, and K. Fritscher, *Multi-organ segmentation of the head and neck area: an efficient hierarchical neural networks approach*, International journal of computer assisted radiology and surgery **14**, 745 (2019).
- [70] G. Mu, Y. Yang, Y. Gao, and Q. Feng, *Multi-scale 3d convolutional neural network-based segmentation of head and neck organs at risk*, Nan Fang yi ke da xue xue bao= Journal of Southern Medical University **40**, 491 (2020).
- [71] Y. Wang, L. Zhao, M. Wang, and Z. Song, *Organ at risk segmentation in head and neck ct images using a two-stage segmentation framework based on 3d u-net*, IEEE Access **7**, 144591 (2019).
- [72] S. Willems, W. Crijs, A. L. G. Saint-Estevan, J. Van Der Veen, D. Robben, T. Depuydt, S. Nuyts, K. Haustermans, and F. Maes, *Clinical implementation of deepvoxnet for auto-delineation of organs at risk in head and neck cancer patients in radiotherapy*, in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis* (Springer, 2018) pp. 223–232.
- [73] Z. Liu, H. Wang, W. Lei, and G. Wang, *Csaf-cnn: Cross-layer spatial attention map fusion network for organ-at-risk segmentation in head and neck ct images*, in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2020) pp. 1522–1525.
- [74] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, and Q. Wang, *Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images*, Medical physics **45**, 2063 (2018).
- [75] Y. Lei, J. Zhou, X. Dong, T. Wang, H. Mao, M. McDonald, W. J. Curran, T. Liu, and X. Yang, *Multi-organ segmentation in head and neck mri using u-faster-rcnn*, in *Medical Imaging 2020: Image Processing*, Vol. 11313 (International Society for Optics and Photonics, 2020) p. 113133A.
- [76] Y. Lei, J. Harms, X. Dong, T. Wang, X. Tang, S. Y. David, J. J. Beitler, W. J. Curran, T. Liu, and X. Yang, *Organ-at-risk (oar) segmentation in head and neck ct using u-rcnn*, in *Medical Imaging 2020: Computer-Aided Diagnosis*, Vol. 11314 (International Society for Optics and Photonics, 2020) p. 1131444.
- [77] B. Ibragimov and L. Xing, *Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks*, Medical physics **44**, 547 (2017).
- [78] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. v. Ooijen, *Recurrent convolutional neural networks for mandible segmentation from computed tomography*, arXiv preprint arXiv:2003.06486 (2020).
- [79] M. Yan, J. Guo, W. Tian, and Z. Yi, *Symmetric convolutional neural network for mandible segmentation*, Knowledge-Based Systems (2018).

- [80] W. Krzanowski, *Principles of multivariate analysis*, Vol. 23 (OUP Oxford, 2000).
- [81] S. Wold, K. Esbensen, and P. Geladi, *Principal component analysis*, *Chemometrics and intelligent laboratory systems* **2**, 37 (1987).
- [82] T. Cootes, E. Baldock, and J. Graham, *An introduction to active shape models*, *Image processing and analysis* **328**, 223 (2000).
- [83] T. F. Cootes, G. J. Edwards, and C. J. Taylor, *Active appearance models*, *IEEE Transactions on pattern analysis and machine intelligence* **23**, 681 (2001).
- [84] T. F. Cootes, C. J. Twining, V. S. Petrovic, R. Schestowitz, and C. J. Taylor, *Groupwise construction of appearance models using piece-wise affine deformations*. in *BMVC*, Vol. 5 (2005) pp. 879–888.
- [85] V. Pekar, S. Allaire, J. Kim, and D. A. Jaffray, *Head and neck auto-segmentation challenge*, *MIDAS J* **11**, 2009 (2009).
- [86] C. Huang, M. Badiei, H. Seo, M. Ma, X. Liang, D. Capaldi, M. Gensheimer, and L. Xing, *Atlas based segmentations via semi-supervised diffeomorphic registrations*, *arXiv preprint arXiv:1911.10417* (2019).
- [87] A. Ayyalusamy, S. Vellaiyan, S. Subramanian, A. Ilamurugu, S. Satpathy, M. Nauman, G. Katta, and A. Madineni, *Auto-segmentation of head and neck organs at risk in radiotherapy and its dependence on anatomic similarity*, *Radiation oncology journal* **37**, 134 (2019).
- [88] Q. Liu, A. Qin, J. Liang, and D. Yan, *Evaluation of atlas-based auto-segmentation and deformable propagation of organs-at-risk for head-and-neck adaptive radiotherapy*, *Recent Patents Top Imaging* **5**, 79 (2016).
- [89] M. Zhu, K. Bzdusek, C. Brink, J. G. Eriksen, O. Hansen, H. A. Jensen, H. A. Gay, W. Thorstad, J. Widder, C. L. Brouwer, *et al.*, *Multi-institutional quantitative evaluation and clinical validation of smart probabilistic image contouring engine (spice) autosegmentation of target structures and normal tissues on computer tomography images in the head and neck, thorax, liver, and male pelvis areas*, *International Journal of Radiation Oncology* Biology* Physics* **87**, 809 (2013).
- [90] G. V. Walker, M. Awan, R. Tao, E. J. Koay, N. S. Boehling, J. D. Grant, D. F. Sittig, G. B. Gunn, A. S. Garden, J. Phan, *et al.*, *Prospective randomized double-blind study of atlas-based organ-at-risk autosegmentation-assisted radiation planning in head and neck cancer*, *Radiotherapy and Oncology* **112**, 321 (2014).
- [91] P. Zaffino, P. Raudaschl, K. Fritscher, G. C. Sharp, and M. F. Spadea, *Plastimatch mabs, an open source tool for automatic image segmentation*, *Medical physics* **43**, 5155 (2016).
- [92] M. La Macchia, F. Fellin, M. Amichetti, M. Cianchetti, S. Gianolini, V. Paola, A. J. Lomax, and L. Widesott, *Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer*, *Radiation Oncology* **7**, 1 (2012).

- [93] R. Haq, S. L. Berry, J. O. Deasy, M. Hunt, and H. Veeraraghavan, *Dynamic multiatlas selection-based consensus segmentation of head and neck structures from ct images*, Medical physics **46**, 5612 (2019).
- [94] R. E. McCarroll, B. M. Beadle, P. A. Balter, H. Burger, C. E. Cardenas, S. Dalvie, D. S. Followill, K. D. Kisling, M. Mejia, K. Naidoo, *et al.*, *Retrospective validation and clinical implementation of automated contouring of organs at risk in the head and neck: a step toward automated radiation treatment planning for low-and middle-income countries*, Journal of global oncology **4**, 1 (2018).
- [95] M. Kass, A. Witkin, and D. Terzopoulos, *Snakes: Active contour models*, International journal of computer vision **1**, 321 (1988).
- [96] N. Otsu, *A threshold selection method from gray-level histograms*, IEEE transactions on systems, man, and cybernetics **9**, 62 (1979).
- [97] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, *Deep learning vs. traditional computer vision*, in *Science and Information Conference* (Springer, 2019) pp. 128–144.
- [98] O. Kodym, M. Španěl, and A. Herout, *Segmentation of head and neck organs at risk using cnn with batch dice loss*, arXiv preprint arXiv:1812.02427 (2018).
- [99] J. Xue, Y. Wang, D. Kong, F. Wu, A. Yin, J. Qu, and X. Liu, *Deep hybrid neural-like p systems for multiorgan segmentation in head and neck ct/mr images*, Expert Systems with Applications **168**, 114446.
- [100] 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 **36**, 61 (2017).
- [101] B. HE and F. JIA, *Head and neck ct segmentation based on a combined u-net model*, Journal of Integration Technology **9**, 17.
- [102] D. J. Rhee, C. E. Cardenas, H. Elhalawani, R. McCarroll, L. Zhang, J. Yang, A. S. Garden, C. B. Peterson, B. M. Beadle, and L. E. Court, *Automatic detection of contouring errors using convolutional neural networks*, Medical physics **46**, 5086 (2019).
- [103] N. Tong, S. Gou, S. Yang, M. Cao, and K. Sheng, *Shape constrained fully convolutional densenet with adversarial training for multiorgan segmentation on head and neck ct and low-field mr images*, Medical physics **46**, 2669 (2019).
- [104] Y. Xue, H. Tang, Z. Qiao, G. Gong, Y. Yin, Z. Qian, C. Huang, W. Fan, and X. Huang, *Shape-aware organ segmentation by predicting signed distance maps*, in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34 (2020) pp. 12565–12572.
- [105] W. Lei, H. Mei, Z. Sun, S. Ye, R. Gu, H. Wang, R. Huang, S. Zhang, S. Zhang, and G. Wang, *Automatic segmentation of organs-at-risk from head-and-neck ct using separable convolutional neural network with hard-region-weighted loss*, arXiv preprint arXiv:2102.01897 (2021).

- [106] S. Liang, K.-H. Thung, D. Nie, Y. Zhang, and D. Shen, *Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck ct images*, IEEE transactions on medical imaging **39**, 2794 (2020).
- [107] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, *Attention is all you need*, arXiv preprint arXiv:1706.03762 (2017).
- [108] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, *Attention u-net: learning where to look for the pancreas*, arXiv preprint arXiv:1804.03999 (2018).
- [109] Y. Ji, H. Zhang, Z. Zhang, and M. Liu, *Cnn-based encoder-decoder networks for salient object detection: A comprehensive review and recent advances*, Information Sciences **546**, 835 (2021).
- [110] S. Gou, N. Tong, S. Qi, S. Yang, R. Chin, and K. Sheng, *Self-channel-and-spatial-attention neural network for automated multi-organ segmentation on head and neck ct images*, Physics in Medicine & Biology **65**, 245034 (2020).
- [111] J. Jiang, E. Sharif, H. Um, S. Berry, and H. Veeraraghavan, *Local block-wise self attention for normal organ segmentation*, arXiv preprint arXiv:1909.05054 (2019).
- [112] S. Sun, Y. Liu, N. Bai, H. Tang, X. Chen, Q. Huang, Y. Liu, and X. Xie, *Attentionanatomy: A unified framework for whole-body organs at risk segmentation using multiple partially annotated datasets*, in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)* (IEEE, 2020) pp. 1–5.
- [113] K. Men, H. Geng, C. Cheng, H. Zhong, M. Huang, Y. Fan, J. P. Plastaras, A. Lin, and Y. Xiao, *More accurate and efficient segmentation of organs-at-risk in radiotherapy with convolutional neural networks cascades*, Medical physics **46**, 286 (2019).
- [114] S. Liang, F. Tang, X. Huang, K. Yang, T. Zhong, R. Hu, S. Liu, X. Yuan, and Y. Zhang, *Deep-learning-based detection and segmentation of organs at risk in nasopharyngeal carcinoma computed tomographic images for radiotherapy planning*, European radiology **29**, 1961 (2019).
- [115] S. Ren, K. He, R. Girshick, and J. Sun, *Faster r-cnn: towards real-time object detection with region proposal networks*, IEEE transactions on pattern analysis and machine intelligence **39**, 1137 (2016).
- [116] F. Milletari, N. Navab, and S.-A. Ahmadi, *V-net: Fully convolutional neural networks for volumetric medical image segmentation*, in *3D Vision (3DV), 2016 Fourth International Conference on* (IEEE, 2016) pp. 565–571.
- [117] L. V. van Dijk, L. Van den Bosch, P. Aljabar, D. Peressutti, S. Both, R. J. Steenbakkers, J. A. Langendijk, M. J. Gooding, and C. L. Brouwer, *Improving automatic delineation for head and neck organs at risk by deep learning contouring*, Radiotherapy and Oncology **142**, 115 (2020).

- [118] J. Yang, H. Veeraraghavan, S. G. Armato III, K. Farahani, J. S. Kirby, J. Kalpathy-Kramer, W. van Elmpt, A. Dekker, X. Han, X. Feng, *et al.*, *Autosegmentation for thoracic radiation treatment planning: a grand challenge at aapm 2017*, Medical physics **45**, 4568 (2018).
- [119] J. Egger, B. Pfarrkirchner, C. Gsaxner, L. Lindner, D. Schmalstieg, and J. Wallner, *Fully convolutional mandible segmentation on a valid ground-truth dataset*, in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (IEEE, 2018) pp. 656–660.

II

MANDIBLE SEGMENTATION ON CT IMAGE

3

AUTOMATIC SEGMENTATION OF THE MANDIBLE FROM COMPUTED TOMOGRAPHY SCANS FOR 3D VIRTUAL SURGICAL PLANNING USING THE CONVOLUTIONAL NEURAL NETWORK

Bingjiang Qiu, Jiapan Guo, Joep Kraeima, Haye H. Glas, Ronald J. H. Borra, Max J. H. Witjes and Peter M. A. van Ooijen.

Qiu, Bingjiang, et al. "Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network." *Physics in Medicine & Biology* 64.17 (2019): 175020.

ABSTRACT

3

Segmentation of mandibular bone in CT scans is crucial for 3D virtual surgical planning of craniofacial tumor resection and free flap reconstruction of the resection defect, in order to obtain a detailed surface representation of the bones. A major drawback of most existing mandibular segmentation methods is that they require a large amount of expert knowledge for manual or partially automatic segmentation. In fact, due to the lack of experienced doctors and experts, high quality expert knowledge is hard to achieve in practice. Furthermore, segmentation of mandibles in CT scans is influenced seriously by metal artifacts and large variations in their shape and size among individuals. In order to address these challenges we propose an automatic mandible segmentation approach in CT scans, which considers the continuum of anatomical structures through different planes. The approach adopts the architecture of the U-Net and then combines the resulting 2D segmentations from three orthogonal planes into a 3D segmentation. We implement such a segmentation approach on two head and neck datasets and then evaluate the performance. Experimental results show that our proposed approach for mandible segmentation in CT scans exhibits high accuracy.

Keywords: automatic mandible segmentation; convolutional neural network (CNN); 3D virtual surgical planning; oral and maxillofacial surgery

3.1. INTRODUCTION

Three-dimensional virtual surgical planning (3D VSP) is proposed to be a precise and predictable method for bone related craniofacial tumor resection and free flap reconstruction of the mandible [1][2][3][4][5]. It is performed pre-operatively to determine the resection margins and osteotomy planes. The planning is translated towards the actual surgical procedure by the use of patient specific 3D printed guides. Currently, computed tomography (CT) is the most commonly used modality to implement such a process. The segmentation of anatomical structures plays a critical role in the 3D VSP. Conventional manual segmentation of the mandible in CT scans leads to a tedious procedure in the clinical practice [6]. Moreover, the structural complexity of mandibles and the considerable human rater variability make the segmentation of mandibles in CT scans challenging [7]. Manual segmentation also has limited reproducibility and is very time-consuming. Therefore, semiautomatic or automatic image segmentation would improve efficiency and reliability, as well as reduce the workload of technologists [6].

Automatic and accurate mandible segmentation remains challenging for several reasons. First, head and neck CT scans cover various bone-structured organs of the human body, such as the skull and the spine, Figure 3.1(a). Second, there is a large variation in the appearance of anatomical structures of mandibles, Figure 3.1(b). Third, noise and other artifacts from the teeth (e.g. because of fillings or braces) as well as lower intensity in the condyles very often lead to ambiguous and blurred boundaries in CT scans, Figure 3.1(c-d). Fourth, inconsistent manual labeling can confuse the trained models for mandible segmentation. For instance, the superior teeth are labeled as part of the mandible while inferior teeth are not, and the presence of both the superior and the inferior teeth in the same slice [7], Figure 3.1(e). All these factors make the automatic segmentation of mandibles difficult by using mathematical models that are applicable to various imaging cases.

Conventional approaches for mandible segmentation typically use pixel-based or model-based methods. In the past decade, some traditional semi-automatic and automatic methods have been developed to segment the mandible in CT scans [8] [7] [9] [10] [11] [12] [13]. A statistical shape model for mandible segmentation was presented by Gollmer and Buzug in 2012 [8]. Torosdagli et al. proposed a 3D gradient-based fuzzy connectedness algorithm to segment mandibles [7]. A registration-based semiautomatic mandible segmentation technique was proposed by Chuang et al. in 2017 [9]. The work of Abdi et al. applies an automatic segmentation algorithm via collecting superior, inferior and exterior borders of mandibles in panoramic x-rays [10]. Chen et al. [11] proposed a multi-atlas model that registered CT images with the atlases at the global level, which allowed multi-atlas-based segmentations and correlation-based label fusion to be performed at the local level. The method from [12] applied active appearance models (AAM) built from manually segmented examples. High quality anatomical correspondences for the models are generated using a groupwise registration method. The models are then applied to segmentation of ROIs in CT scans. Albrecht et al. employed a multi-atlas segmentation to obtain an initial segmentation and then apply an Active Shape Model (ASM) segmentation to refine the initial segmentation of the organ [13]. The performances of the conventional methods, however, are often affected by the noise or metal artifacts in

Figure 3.1 | Examples of typical cases that challenge the accurate mandible segmentation in CT scans. (a) Various bone-structured organs in the head and neck CT, such as the skull and the spine. (b) Large variation of mandibles between individuals. (c) Metal artifacts and noises in the teeth. (d) Lower intensity in the condyles. (e) Metal artifacts and the presence of inferior and superior teeth in the same slice.

the CT images. Weak and false edges in the parts of condyles and teeth, often appear in the detected images, which frustrate the accurate segmentation of the mandible. The proposed deformable models, of which the parameters are determined according to the global characteristics of the target contour, are difficult to adapt to some local areas of the contour. [14][15][16]

Since 2013, convolutional neural networks (CNNs) have been successfully applied in computer vision tasks, for example, image classification[17], super-resolution[18], medical imaging[19][20] and semantic segmentation[21]. Medical image segmentation, as one of the research focuses in semantic segmentation, has developed exponentially due to the rapid evolution of CNNs [22][23]. Semantic segmentation aims to assign a semantic label at the pixel level [24]. Recent advances in semantic segmentation [21, 25–34][35] have enabled their applications to medical image segmentation. Most of the above mentioned CNN architectures have been proven to be effective on semantic segmentation of natural scene images. After that, Ibragimo et al. [36] presented the first attempt of using deep learning concept of CNN to segment organs-at-risks in head and neck CT scans. The AnatomyNet [37] is built upon the popular 3D U-net architecture using residual blocks in encoding layers and a new loss function combining Dice score and focal loss in the training process. A fully CNN (FCNN) method was presented by [38]. The multi-planar training strategy was presented by [39]. Such works motivate our implementation of CNNs for automatic mandible segmentation in CT scans.

Here, we propose a CNN-based approach for 3D mandible segmentation in CT scans. Such an approach takes a multi-planar volume-to-slice strategy which takes into ac-

count the spatial information of adjacent slices in order to preserve the connectivity of anatomical structures.

3.2. MATERIALS AND METHODS

3.2.1. DATA PREPARATION

The collection of the patient data sets for medical research purposes was approved by the local Medical Ethical Committee. The data set contains 109 CT scans reconstructed with a kernel of Br64, I70h(s) or B70s. Each scan consists of 221 to 955 slices with a size of 512×512 pixels. The pixel spacing varies from 0.35 to 0.66 mm, and the slice thickness varies from 0.6 to 0.75 mm. The manual mandible segmentation was performed using Mimics software version 20.0 (Materialise, Leuven, Belgium) by a trained researcher and confirmed by a clinician.

We also validate our method on a public dataset of head and neck scans dataset, which was obtained from Public Domain Database for Computational Anatomy (PDDCA) version 1.4.1 [40]. The original CT dataset is provided and maintained by Dr. Gregory C Sharp (Harvard Medical School – MGH, Boston) and his group. PDDCA version 1.4.1 comprises 48 patient CT images from the Radiation Therapy Oncology Group (RTOG) 0522 study, a multi-institutional clinical trial, together with manual segmentation of left and right parotid glands, brainstem, optic chiasm and mandible. Each scan consists of 76 to 360 slices with a size of 512×512 pixels. The pixel spacing varies from 0.76 to 1.27mm, and the slice thickness varies from 1.25 to 3.0 mm. Forty of the 48 patients in PDDCA with manual mandible annotation are used in this study. [40][41]

3.2.2. METHODS

We illustrate our proposed framework for automatic 3D mandible segmentation in Figure 3.2. Such a framework takes as input the 2D slice images of the head and neck CT scans from three orthogonal (axial, coronal and sagittal) planes. The images from each plane are fed into a convolutional neural network that automatically segments the mandible. We then combine the 2D segmentation results obtained from each plane to reconstruct into a 3D segmentation of the mandible.

Our decision to combine the 2D segmentation results from three orthogonal planes for 3D mandible segmentation is motivated by two reasons. First, the anatomical structures of mandibles can be more representative from multi-planar images which are then used to train the computational models for mandible segmentation, comparing to the use of single plane images. Second, it helps to reduce computation complexity in comparison to direct 3D segmentation on volumetric CT scans.

SINGLE-PLANAR CNN MODEL

In this subsection, we elaborate on the design of the single-planar module. Figure 3.3 shows the architecture of the single-planar module, which is inspired by that of the U-Net [26]. In order to consider the similarity and continuity of upper and lower regions of the mandible, we use multi-sectional slices for training in each plane, which can help retain

Figure 3.2 | The proposed framework of the multi-planar volume-to-slice network for 3D segmentation of the mandible.

the structural information of the mandible. We call this strategy multi-planar volume-to-slice. This single-planar network consists of an encoding and a decoding procedure of in total 23 convolutional layers, each of which has convolutional kernels of the same size 3×3 . The number of feature maps is listed on the top of each block that represents the convolutional layer in Figure 3.3. During the encoder procedure, max pooling layers are used to further enlarge the receptive fields [42]. In the configuration of the model, we increase the number of feature maps by a factor of 2 after every max pooling layer. For the technical details on the architecture of the U-Net, we refer the interested readers to [26]. There are two dropout layers [43] with a dropout rate $r = 0.5$ in the encoder, as shown by the grey blocks in Figure 3.3. Dropout is a technique to address the issues of overfitting that usually makes CNN underperform during the test phase. The mechanism of dropout is to randomly drop units along with their connections during the training of CNNs so that it exhibits the bagged ensemble of many neural networks [42]. In such a way it helps to prevent CNN from overfitting in a computationally inexpensive but powerful manner [43]. During the decoder procedure, upsampling layers are used and the feature maps are then concatenated with those of the same resolution from the encoder procedure. Such feature concatenation is indicated by the horizontal arrows in Figure 3.3. In the original U-Net, the resulting feature maps from each level in the encoder are concatenated to those on the same level in the decoder. In order to save computational complexity, we concatenate the feature maps from the first convolutional layer (instead of the second convolutional layer in each level) of the level in the encoder with those in the decoder, which is indicated by the horizontal arrows in Figure 3.3.

Each of the convolutional layer blocks is composed of a linear convolution, a batch normalization [44] and an element-wise nonlinear ReLu function [45], in which batch

Figure 3.3 | The 2D single-planar mandible segmentation based on the U-Net architecture. In the input layer, we denote the height, the width and the number of slices by H , W and s , respectively. All convolutional kernels have a size of 3×3 .

normalization is a technique for improving the performance and stability of CNN. Finally, the resulting responses are turned into probability values using a sigmoid classifier [46]. The output has the same size as that of the input image, as shown in Figure 3.3. In general, the convolutional layers (illustrated by the cubic blocks in Figure 3.3) can be expressed in the following equation:

$$L_i = \Gamma(L_{i-1} \otimes K_{ij} + B_i), \quad (3.1)$$

where L_i is the output of the i -th layer ($i \in [1, 23]$) and the number of feature maps for L_i is shown in the top of each layer in Figure 3.3, Γ represents the nonlinear operator of ReLU or sigmoid function in the output or last layer, $\otimes$ is the convolutional operator, K_{ij} represents the j -th feature kernel in the i -th layer and B_i denotes a bias of the i -th layer.

We use a loss function based on Dice coefficient which is commonly used to evaluate the performance of image segmentation tasks [47]. We provide the detailed information about Dice Coefficient loss in Section 3.2.2.

It is worth noting that the U-Net is able to predict the probability of scores (in our case the probability of the mandible) based on the structural texture information within a given receptive field. This enables the use of all information in a 2D CT plane to directly predict complex structures. The U-Net architecture provides the capability to adapt naturally to input CT images of any resolution.

MULTI-PLANAR MANDIBLE SEGMENTATION FRAMEWORK

The use of consecutive slices from the scans helps improve the 2D segmentation of the mandible. It, however, does not provide satisfactory results since some parts of the mandible shapes are better observed via other planes. Thus, we create a three-planar network which takes into account the structural information of the mandible from different orthogonal planes. Such a multi-planar network consists of three subnetworks, each of which is fed with CT data from one of the three planes. In the network fed with data from the axial plane, the size of the input data has height $H = 512$, width $W = 512$ and number of slices $s = 3$. In the Sagittal (Coronal) planes which often have different numbers of slices in every scan, we use a sliding window to crop the input images into the same size of $H = 400$, $W = 400$ and $s = 7$ ($H = 400$, $W = 400$ and $s = 9$). We use Adam optimization [48] with a learning rate of 10^{-5} . The three subnetworks use the same size of convolution kernels of 3×3 with zero padding.

LOSS METRIC

The Dice similarity index, also called the Dice score, is often used to measure consistency between two samples [47]. Therefore, it is widely applied as a metric to evaluate the performance of image segmentation algorithms. It is defined as:

$$\text{Dice} = \frac{2|Y_r \cap Y_p|}{|Y_r| + |Y_p|}, \quad (3.2)$$

where Y_r is the reference standard, and Y_p is the prediction from the network. Such a score has a value between 0 and 1, in which 0 means the total disagreement between the reference standard and the evaluated segmentation while 1 presents total agreement. A differentiable formulation of the Dice coefficient has been proposed by Milletari [49], for minimizing the loss ($\text{loss} = 1 - \text{Dice}$) between the two binary labels in the training of the proposed multi-planar network.

COMBINATION OF THE MULTI-PLANAR NETWORK FOR 3D MANDIBLE SEGMENTATION

We train each of the subnetworks in the proposed multi-planar framework independently and then apply the trained framework on the test scans. In order to obtain the 3D segmentation of the mandible, we stack the 2D segmentation results from each subnetwork into a 3D volume data, in which each voxel has a probability prediction from the sigmoid function in each subnetwork. For instance, the resulting output (probability prediction) from three subnetworks can be denoted by Y^a , Y^s and Y^c separately. In order to effectively combine the predicted information, we select the maximum probability of each voxel from the three networks, $Y = \max\{Y^a, Y^s, Y^c\}$, to improve accuracy of the final results [39]. Figure 3.4 shows the combination procedures.

3.3. EXPERIMENTS

3.3.1. EVALUATION CRITERION

We use Dice similarity coefficient and the 3D surface error to evaluate the performance of the proposed approach for mandible segmentation. We compute the Dice similarity

Figure 3.4 | Step-by-step illustration on the 3D mandible segmentation.

Figure 3.5 | Examples of the automatic segmentation of mandibles in the axial plane.

coefficient of the automatic segmentation results with respect to the manual segmentation on both 2D slice images and the complete 3D volumetric data. In order to observe the segmentation results in a more straightforward way, we use the Materialise Mimics software to reconstruct the 2D automatic segmentation into a 3D view. We then use the Materialise 3-matic software to automatically post-process the segmentation results in order to remove the disconnected voxels. Afterwards, we compute the root mean square error (RMS) of the mandible surfaces between the manual and the post-processed model.

3.3.2. EXPERIMENTAL RESULTS

We implement the framework proposed in Section 2 by using Keras [50] package with the Tensorflow [51] backend. The CNN models are trained on a workstation equipped with Nvidia Tesla K40m GPU of 12GB memory.

EXPERIMENTS ON 109 CT SCANS

We randomly chose 52 cases as training, 8 cases as validation and 49 cases as test. We repeated the above mentioned process three times and then evaluated the performance of the proposed approach for automatic mandible segmentation. The training of a single-plane model takes less than 60 hours while the duration of the test on one scan is about 10 minutes. Figure 3.5 illustrates several examples of the results achieved in the axial direction by the proposed approach on the test scan. The average Dice score on all test scans from three repeats are 0.893, 0.878 and 0.872, respectively, as listed in Table 3.1.

Figure 3.6 shows several cases of the postprocessed automatic segmentation and their comparison with the manual annotation in the 3D view. The visual comparison of the automatic segmentation with manual segmentation illustrates the effectiveness of our

Table 3.1 | Dice score and RMS surface error on the three random repeats of the experiments.

	Test 1	Test 2	Test 3	Average
Dice score on the axial plane	0.799	0.767	0.789	0.785
Dice score on the sagittal plane	0.559	0.519	0.391	0.490
Dice score on the coronal plane	0.594	0.531	0.545	0.557
Dice score of the combined results	0.893	0.878	0.872	0.881
RMS value (mm) of the combined results	0.4284	0.5349	0.7740	0.5791

method on automatic segmentation of mandible.

Figure 3.7-3.9 illustrates the distribution of the root mean square errors of the surface distances, in which the blue lines indicate the average RMS values that is also listed in Table 3.1. From Figure 3.7-3.9, the majority of the test cases have the surface error under the average errors. We also provide the visual 3D illustration on the cases that have the top three RMS surface errors in each repeat. One of the main reasons for the large surface error is that parts of superior teeth are also automatically segmented while only the inferior teeth are manually annotated.

EXPERIMENTS ON THE PDDCA DATASET

In particular, we compare our proposed method with several state-of-the-art methods on PDDCA dataset. We use the model based on test 1, i.e., we use the weights trained in test 1 of section 3.3.2 as initialization parameters. According to the Challenge description, we follow the same training and testing protocol [40]. A subset of 40 scans was used: 25 scans (0522c0001-0522c0328) are used as training data, and 15 scans (0522c0555-0522c0878) are used for testing [40].

For comparison purpose, Table 3.2 lists Dice score and 95% Hausdorff distance (95HD) used in the Challenge paper [40][38]. According to Table 3.2, the proposed method outperforms most other methods, with the second highest mean DC score and the lowest 95HD. For Dice score results, the segmentation result of our method is only slightly worse than the FCNN+SRM[38], while outperforming the rest of the works.

3.4. DISCUSSION

In this work, we present a CNN-based method for mandible segmentation that is demonstrated to be effective for 3D segmentation of the mandible in a data set of 109 head and neck scans. Our approach uses multi-sectional CT slices from different orthogonal planes as input and then combines the 2D segmentation results from each plane in order to achieve a mandible segmentation in a 3D view. The proposed approach spends around 2.5 minutes to process one scan in the PDDCA dataset. Table 3.3 compares the testing time to segment a new patient using our method versus the state-of-the-art methods in

Figure 3.6 | Visual examples of final segmentations on our data set. From left to right are 3D views of (a) ground truth, (b) the algorithm segmentation, (c) the surface distance maps (mm) from algorithm segmentation to manual segmentation.

Figure 3.7 | Test 1: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS vales in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 0.8494 mm, 1.0261 mm and 1.3885 mm

Figure 3.8 | Test 2: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS vales in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 2.5965 mm, 1.0261 mm and 1.0744 mm

Figure 3.9 | Test 3: Scattered plot of RMS of surface distance from the test data. The green points indicate the RMS of surface distance while the blue line represents the average RMS vales in the 49 testing cases. (a)-(c) the surface distance maps with the three greatest RMS values of 2.9267 mm, 3.4411 mm and 3.6207 mm

Table 3.2 | Comparison of average Dice ($\pm$ standard deviation) between the state-of-the-art methods and our method, bold fonts indicate the best performer for that structure.

Methods	Dice	95HD (mm)
Multi-atlas[11]	0.917($\pm$ 0.0234)	2.4887($\pm$ 0.7610)
AAM[12]	0.9267	1.9767
ASM[13]	0.8813($\pm$ 0.0555)	2.832($\pm$ 1.1772)
CNN[36]	0.895($\pm$ 0.036)	-
AnatomyNet[37]	0.9251($\pm$ 0.02)	(6.28 $\pm$ 2.21)
FCNN[38]	0.9207($\pm$ 0.0115)	2.01($\pm$ 0.83)
FCNN+SRM[38]	0.936 ($\pm$ 0.0121)	1.5($\pm$ 0.32)
The proposed method	0.9328($\pm$ 0.0144)	1.4333 ($\pm$ 0.5564)

the PDDCA dataset. In particular, the testing time costs increase to 10 minutes in the 109 CT scans. We obtain similar accuracies compared to the work of the traditional methods [12][7] but with superiority in efficiency and fully-automatic nature. Furthermore, such a design of the system enables the consideration of similarity and structural continuity of the mandible from different planes. The main contributions of this work can be summarized as follows.

Table 3.3 | Comparison of segmentation time on PDDCA dataset between the state-of-the-art methods and our method.

Methods	Segmentation time	Experimental equipment
Multi-atlas[11]	Over 60 min per patient	CPU
AAM[12]	30 min per image	CPU
ASM[13]	24 min per patient (in the online Challenge)	CPU
CNN[36]	30 min per patient	GPU
AnatomyNet[37]	0.12 sec per patient	GPU
FCNN and FCNN+SRM[38]	9.50 sec per patient	GPU
The proposed method	2.5 min per patient	GPU

First, the proposed approach extracts discriminative features for segmentation from three orthogonal planes (axial, coronal and sagittal) since some parts of the mandible shapes are better observed via other planes. In our study, we make full use of the extracted information from different planes to segment the mandible. Our proposed method indicates that 3D segmentation of medical images can also be achieved by 2D segmentation network which could help alleviate memory issues. Remarkably, this method has a good performance on mandible segmentation. For example, Figure 3.10 shows the intermediate feature maps in the 21-st layer from the Axial plane, which illustrates that the neural network actually learned the structural representation of mandibles. The outputs in Figure 3.10 mainly focus on the mandible; even the parts of condyles and ramus are accurately marked out. The intermediate results demonstrate the feasibility of extracting image features by the proposed CNN-based framework. Moreover, instead of using cropped small patches in some other works [52] [53] for 3D medical image segmentation, our approach uses images with the original size (512×512), which maximally fits to the computation capacity of GPUs and keeps the structural context of mandibles.

Second, we used the multi-planar volume-to-slice training strategy which takes into account the structural continuity of mandibles from different views. Naturally, the mandible anatomical structure is very complicated. The CT imaging quality is severely and easily disrupted by dental prosthesis or fillings in the teeth since they are produced by highly attenuated materials. These materials lead to noisy and ambiguous boundaries in the CT imaging. In order to overcome this shortcoming, we used multi-planar volume-to-slice strategy for training, which can learn the 'upper and lower structure' of the mandible in each CT slice. To increase the robustness of the segmentation method, we only used the original CT data without any pre-processing step.

The achieved experimental results demonstrate the feasibility and effectiveness of

Figure 3.10 | Insight of learned U-Net in Axial plane. (a) The input CT image. (b)-(g) Intermediate output of the 21-st layer. All the intermediate outputs are randomly chosen from the 21-st layer.

the proposed approach in 3D mandible segmentation and could also be applied to other segmentation tasks. The current implementation treats the corresponding 3D CT image segmentation of the mandible as a video object segmentation with continuous structure, in which we consider the special continuous structure of the mandible. Therefore, the technique of taking into account the adjacent slices (frames) information could be transferred to the application of object segmentation in video frames. Moreover, the collection of contextual and shape information from different planes assures the sufficient extraction of useful information from input images, which could provide a future research direction for 3D image segmentation. Finally, the proposed work could be applied to other organ segmentation or in other imaging modalities.

Despite the promising results, this study also has several limitations. First, in this experiment, we use 109 head and neck scans for the training and the validation of the proposed approach. Our data set may be limited and cannot sufficiently represent the general population in clinical practice. The experimental results need to be validated with large variety data sets in future studies. Second, the inconsistency of the manual annotations on the inferior and superior teeth influences the performance of the proposed approach as shown in the examples on top of Figure 3.7-3.9. This could be improved by setting the manual annotation protocol to keep consistency. Third, the accurate evaluation of mandible segmentation remains a challenge and is subjective due to the inconsistency in manual segmentation for the ambiguous boundaries in the parts of teeth and condyles of the mandible. Thus, developing more accurate mandible segmentation and evaluation methods is still one of the research focuses. Further evaluation of our approach is required to assess its performance in clinical practice. This could be done through more massive and intensive experiments in or outside our maxillofacial oncology

center. More experimental data sets should be used to verify this technique in future studies. Besides, it is of great importance to verify the automatic mandible segmentation in the 3D virtual planning of craniofacial tumor resection and free flap reconstruction.

3.5. CONCLUSION

This study proposes an end-to-end approach for automatic segmentation of the mandible from head and neck CT scans. Our approach takes into account the input images from three planes for 2D segmentation of mandibles in multi-planar CT slices and then combines the resulting 2D segmentation into 3D segmentation. We implement the proposed approach on 109 head and neck CT scans and the PDDCA public dataset. We achieve an average Dice coefficient of 0.881 and an average surface error of 0.5791 mm on 109 CT scans and an average Dice coefficient of 0.9328 and 95HD of 1.4333 mm on PDDCA dataset. The experimental results demonstrate the effectiveness of the proposed approach in mandible segmentation and its potential employment in 3D virtual planning of craniofacial tumor resection and free flap reconstruction.

REFERENCES

- [1] G. Bittermann, C. Scheifele, V. Prokic, V. Bhatt, M. Henke, A.-L. Grosu, R. Schmelzeisen, and M. C. Metzger, *Description of a method: computer generated virtual model for accurate localisation of tumour margins, standardised resection, and planning of radiation treatment in head & neck cancer surgery*, Journal of Cranio-Maxillofacial Surgery **41**, 279 (2013).
- [2] H. Essig, M. Rana, H. Kokemueller, C. von See, M. Ruecker, F. Tavassol, and N.-C. Gellrich, *Pre-operative planning for mandibular reconstruction-a full digital planning workflow resulting in a patient specific reconstruction*, Head & neck oncology **3**, 45 (2011).
- [3] R. H. Schepers, G. M. Raghoobar, A. Vissink, M. W. Stenekes, J. Kraeima, J. L. Roodenburg, H. Reintsema, and M. J. Witjes, *Accuracy of fibula reconstruction using patient-specific cad/cam reconstruction plates and dental implants: a new modality for functional reconstruction of mandibular defects*, Journal of Cranio-Maxillofacial Surgery **43**, 649 (2015).
- [4] W. L. Weijs, C. Coppen, R. Schreurs, R. D. Vreeken, A. C. Verhulst, M. A. Merks, S. J. Bergé, and T. J. Maal, *Accuracy of virtually 3d planned resection templates in mandibular reconstruction*, Journal of Cranio-Maxillofacial Surgery **44**, 1828 (2016).
- [5] J. Kraeima, R. H. Schepers, P. M. van Ooijen, R. J. Steenbakkens, J. L. Roodenburg, and M. J. Witjes, *Integration of oncologic margins in three-dimensional virtual planning for head and neck surgery, including a validation of the software pathway*, Journal of Cranio-Maxillofacial Surgery **43**, 1374 (2015).
- [6] T. J. Huff, P. E. Ludwig, and J. M. Zuniga, *The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning*, Expert review of medical devices **15**, 349 (2018).
- [7] N. Torosdagli, D. K. Liberton, P. Verma, M. Sincan, J. Lee, S. Pattanaik, and U. Bagci, *Robust and fully automated segmentation of mandible from ct scans*, in *Biomedical Imaging (ISBI 2017)*, 2017 IEEE 14th International Symposium on (IEEE, 2017) pp. 1209–1212.
- [8] S. T. Gollmer and T. M. Buzug, *Fully automatic shape constrained mandible segmentation from cone-beam ct data*, in *Biomedical Imaging (ISBI)*, 2012 9th IEEE International Symposium on (IEEE, 2012) pp. 1272–1275.
- [9] Y. J. Chuang, B. M. Doherty, N. Adluru, M. K. Chung, and H. K. Vorperian, *A novel registration-based semiautomatic mandible segmentation pipeline using computed tomography images to study mandibular development*, Journal of computer assisted tomography (2017).
- [10] A. H. Abdi, S. Kasaei, and M. Mehdizadeh, *Automatic segmentation of mandible in panoramic x-ray*, Journal of Medical Imaging **2**, 044003 (2015).

- [11] A. Chen and B. Dawant, *A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images*, Head and Neck Auto-Segmentation Challenge (MICCAI), Munich (2015).
- [12] R. Mannion-Haworth, M. Bowes, A. Ashman, G. Guillard, A. Brett, and G. Vincent, *Fully automatic segmentation of head and neck organs using active appearance models*, MIDAS J. (2015).
- [13] T. Albrecht, T. Gass, C. Langguth, and M. Lüthi, *Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning*, Head and Neck Auto-Segmentation Challenge (MICCAI), Munich (2015).
- [14] S. Yuheng and Y. Hao, *Image segmentation algorithms overview*, arXiv preprint arXiv:1707.02051 (2017).
- [15] T. Blaschke, C. Burnett, and A. Pekkarinen, *Image segmentation methods for object-based analysis and classification*, in *Remote sensing image analysis: Including the spatial domain* (Springer, 2004) pp. 211–236.
- [16] I. Bankman, *Handbook of medical image processing and analysis* (Elsevier, 2008).
- [17] 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.
- [18] C. Dong, C. C. Loy, K. He, and X. Tang, *Learning a deep convolutional network for image super-resolution*, in *European conference on computer vision* (Springer, 2014) pp. 184–199.
- [19] J. He, Y. Yang, Y. Wang, D. Zeng, Z. Bian, H. Zhang, J. Sun, Z. Xu, and J. Ma, *Optimizing a parameterized plug-and-play admm for iterative low-dose ct reconstruction*, IEEE transactions on medical imaging **38**, 371 (2019).
- [20] S. Li, D. Zeng, J. Peng, Z. Bian, H. Zhang, Q. Xie, Y. Wang, Y. Liao, S. Zhang, J. Huang, et al., *An efficient iterative cerebral perfusion ct reconstruction via low-rank tensor decomposition with spatial-temporal total variation regularization*, IEEE transactions on medical imaging **38**, 360 (2019).
- [21] 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.
- [22] J. Ker, L. Wang, J. Rao, and T. Lim, *Deep learning applications in medical image analysis*, Ieee Access **6**, 9375 (2018).
- [23] D. Shen, G. Wu, and H.-I. Suk, *Deep learning in medical image analysis*, Annual review of biomedical engineering **19**, 221 (2017).

- [24] H. Fu, *Semantic image understanding: from pixel to word*, Ph.D. thesis, University of Nottingham (2012).
- [25] V. Badrinarayanan, A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*, IEEE transactions on pattern analysis and machine intelligence **39**, 2481 (2017).
- [26] 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* (Springer, 2015) pp. 234–241.
- [27] F. Yu and V. Koltun, *Multi-scale context aggregation by dilated convolutions*, arXiv preprint arXiv:1511.07122 (2015).
- [28] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, *Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs*, IEEE transactions on pattern analysis and machine intelligence **40**, 834 (2018).
- [29] G. Lin, A. Milan, C. Shen, and I. Reid, *Refinenet: Multi-path refinement networks for high-resolution semantic segmentation*, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017).
- [30] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, *Pyramid scene parsing network*, in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)* (2017) pp. 2881–2890.
- [31] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, *Large kernel matters—improve semantic segmentation by global convolutional network*, arXiv preprint arXiv:1703.02719 (2017).
- [32] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, *Rethinking atrous convolution for semantic image segmentation*, arXiv preprint arXiv:1706.05587 (2017).
- [33] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, and J. Garcia-Rodriguez, *A review on deep learning techniques applied to semantic segmentation*, arXiv preprint arXiv:1704.06857 (2017).
- [34] W. Qin, J. Wu, F. Han, Y. Yuan, W. Zhao, B. Ibragimov, J. Gu, and L. Xing, *Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation*, Physics in Medicine & Biology **63**, 095017 (2018).
- [35] J. Wu, K. K. Tha, L. Xing, and R. Li, *Radiomics and radiogenomics for precision radiotherapy*, Journal of radiation research **59**, i25 (2018).
- [36] B. Ibragimov and L. Xing, *Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks*, Medical physics **44**, 547 (2017).
- [37] W. Zhu, Y. Huang, H. Tang, Z. Qian, N. Du, W. Fan, and X. Xie, *Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation*, arXiv preprint arXiv:1808.05238 (2018).

- [38] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, *Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks*, Medical physics **45**, 4558 (2018).
- [39] A. Mortazi, J. Burt, and U. Bagci, *Multi-planar deep segmentation networks for cardiac substructures from mri and ct*, in *International Workshop on Statistical Atlases and Computational Models of the Heart* (Springer, 2017) pp. 199–206.
- [40] P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant, T. Albrecht, T. Gass, C. Langguth, M. Lüthi, *et al.*, *Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015*, Medical physics **44**, 2020 (2017).
- [41] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, and Q. Wang, *Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images*, Medical physics **45**, 2063 (2018).
- [42] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, nature **521**, 436 (2015).
- [43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting*, The Journal of Machine Learning Research **15**, 1929 (2014).
- [44] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, arXiv preprint arXiv:1502.03167 (2015).
- [45] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010) pp. 807–814.
- [46] H.-T. Lin and C.-J. Lin, *A study on sigmoid kernels for svm and the training of non-psd kernels by smo-type methods*, submitted to Neural Computation **3**, 1 (2003).
- [47] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. Uden, C. I. Sanchez, G. Litjens, F.-E. Leeuw, B. Ginneken, E. Marchiori, and B. Platel, *Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities*, Scientific Reports **7**, 5110 (2017).
- [48] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014).
- [49] F. Milletari, N. Navab, and S.-A. Ahmadi, *V-net: Fully convolutional neural networks for volumetric medical image segmentation*, in *3D Vision (3DV), 2016 Fourth International Conference on* (IEEE, 2016) pp. 565–571.
- [50] F. Chollet *et al.*, *Keras*, <https://keras.io> (2015).
- [51] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, *et al.*, *Tensorflow: Large-scale machine learning on heterogeneous distributed systems*, arXiv preprint arXiv:1603.04467 (2016).

- [52] 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* (Springer, 2016) pp. 424–432.
- [53] L. Yu, X. Yang, H. Chen, J. Qin, and P.-A. Heng, *Volumetric convnets with mixed residual connections for automated prostate segmentation from 3d mr images*. in *AAAI* (2017) pp. 66–72.

4

RECURRENT CONVOLUTIONAL NEURAL NETWORKS FOR 3D MANDIBLE SEGMENTATION IN COMPUTED TOMOGRAPHY

Bingjiang Qiu, Jiapan Guo, Joep Kraeima, Haye H. Glas, Weichuan Zhang, Ronald J. H. Borra, Max J. H. Witjes and Peter M. A. van Ooijen.

Qiu, Bingjiang, et al. "Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography." *Journal of Personalized Medicine* 11.6 (2021): 492.

ABSTRACT

4

Purpose: Classic encoder–decoder based convolutional neural network (EDCNN) approaches cannot accurately segment detailed anatomical structures of the mandible in computed tomography (CT), for instance, condyles and coronoids of the mandible, which are often affected by noise and metal artifacts. The main reason is that EDCNN approaches ignore the anatomical connectivity of the organs. In this paper, we propose a novel CNN-based 3D mandible segmentation approach that has the ability to accurately segment detailed anatomical structures.

Methods: Different from the classic EDCNNs that need to slice or crop the whole CT scan into 2D slices or 3D patches during the segmentation process, our proposed approach can perform mandible segmentation on complete 3D CT scans. The proposed method, namely, RCNNSeg, adopts the structure of the recurrent neural networks to form a directed acyclic graph in order to enable recurrent connections between adjacent nodes to retain their connectivity. Each node then functions as a classic EDCNN to segment a single slice in the CT scan. Our proposed approach can perform 3D mandible segmentation on sequential data of any varied lengths and does not require a large computation cost. The proposed RCNNSeg was evaluated on 109 head and neck CT scans from a local dataset and 40 scans from the PDDCA public dataset. The final accuracy of the proposed RCNNSeg was evaluated by calculating the Dice similarity coefficient (DSC), average symmetric surface distance (ASD), and 95% Hausdorff distance (95HD) between the reference standard and the automated segmentation.

Results: The proposed RCNNSeg outperforms the EDCNN-based approaches on both datasets and yields superior quantitative and qualitative performances when compared to the state-of-the-art approaches on the PDDCA dataset. The proposed RCNNSeg generated the most accurate segmentations with an average DSC of 97.48%, ASD of 0.2170 mm, and 95HD of 2.6562 mm on 109 CT scans, and an average DSC of 95.10%, ASD of 0.1367 mm, and 95HD of 1.3560 mm on the PDDCA dataset.

Conclusions: The proposed RCNNSeg method generated more accurate automated segmentations than those of the other classic EDCNN segmentation techniques in terms of quantitative and qualitative evaluation. The proposed RCNNSeg has potential for automatic mandible segmentation by learning spatially structured information.

Keywords: accurate mandible segmentation; oral and maxillofacial surgery; 3D virtual surgical planning (3D VSP); convolutional neural network

4.1. INTRODUCTION

Oral cancer is a type of cancer that originates from the lip, the mouth, or the upper throat [1]. Globally, there are an estimated 354,864 new cases of oral cancer annually, and there were 177,384 deaths in 2018 [2]. Surgical tumor resection is the most common curative treatment for oral cancer [3]. During surgical removal of malignant tumors in the oral cavity, a continuous resection of the jaw bone can be required. Currently, this resection is based on the 3D virtual surgical planning (VSP) [4, 5] that enables accurate planning of the resection margin around the tumor, taking into account the surrounding jaw bone. Researches [5, 6] indicate that 3D VSP requires accurate delineation of mandible organs, which is manually performed by technologists. However, manual mandible delineation in CT scans is very time consuming (about 40 minutes) and has high inter-rater variabilities (Dice score of 94.09% between two clinical experts) [7], and the performances of technologists can also be affected by fatigue [7–9]. In order to help improve the reliability and efficiency of the manual delineation, robust and accurate algorithms for automatic mandible segmentation are highly demanded for the 3D VSP [7, 10].

In general, the existing mandible segmentation approaches can be divided into two categories [11], i.e., traditional and deep learning-based approaches. Traditional approaches [12–17] have been widely investigated for automatic mandible segmentation in CT scans. Gollmer et al. [12] proposed a fully automatic segmentation approach that uses a statistical shape model for mandible segmentation in cone-beam CT. Chen et al. [13] presented an automatic multi-atlas model that registered CT slices with the obtained atlas to enable multi-organ segmentation in head and neck CT scans. Mannion-Haworth et al. [14] proposed automatic active appearance models (AAMs) that rely on a group-wised registration method to generate high-quality anatomical correspondences. Albrecht et al. [15] combined multi-atlas segmentation and the active shape model (ASM) for the segmentation of automatic organs at risk (OARs) in head and neck CT scans. Torosdagli et al. [16] proposed an automatic two-step strategy that first uses random forest regression to localize the mandible region and then performs the 3D gradient-based fuzzy connectedness algorithm for mandible delineation. The aforementioned approaches offer potentially time-saving solutions. However, these methods mostly rely on the averaged shapes or atlases of the structures generated by domain experts [12–15]. Therefore, traditional approaches lead to poor individualization for single cases in mandible segmentation [18–20].

Alternatively, deep learning approaches have demonstrated a strong ability in automatic descriptive and detailed image feature extraction [21, 22]. Ibragimov et al. [23] applied a tri-planar patch-based CNN for mandible segmentation in head and neck CT scans and achieved superior performance for most of the OAR segmentation when compared to the existing approaches. Zhu et al. [24] applied a 3D Unet that uses a loss function combining Dice score and focal loss for the training of the network. This 3D Unet-based method achieved better performances than that of the state-of-the-art approaches in OAR segmentation. Tong et al. [25] proposed a fully CNN (FCNN) with a shape representation model for mandible segmentation in CT scans and achieved better results than those of the conventional approaches. Egger et al. [26] implemented a three-step training strategy of the fully convolutional networks proposed by [27] to segment the mandibles in the CT

scans on a locally acquired dataset. Qiu et al. [11] presented a 2D Unet-based mandible segmentation approach that segments mandibles on orthogonal planes of the CT scans. Liang et al. [28] proposed a multi-view spatial aggregation frame for joint localization and the segmentation of the OARs and achieved competitive segmentation performance.

These aforementioned deep learning approaches for mandible segmentation have the encoder-decoder-based CNN architectures that consist of an encoder and a decoder. The encoder network maps a given input image to a feature space that is then processed by a decoder to produce an output image of the same size as the input [29]. Ye et al. [30] investigated the geometry of the EDCNN and demonstrated that its excellent performances come from the expressiveness of the symmetric encoder and decoder networks and the skip connections that enable feature flows from the encoder to the decoder.

The use of deep learning approaches helped to achieve better performance than the conventional approaches. However, our research indicates that challenges still exist in EDCNN-based approaches for mandible segmentation in CT scans. On the one hand, they are often affected by the noise or metal artifacts that are commonly present in head and neck CT scans [31, 32]. On the other hand, they are not robust in segmenting organs that have weak boundaries, such as the condyles and coronoids of the mandible [11]. The main reason is that the inputs of EDCNN-based segmentation approaches are either truncated into 2D slices or 3D patches due to the computation capacity [33]. These truncated inputs do not represent the complete anatomical structures of the organs and, as a consequence, lead to inaccurate segmentation of the detailed structures and artifact corrupted regions [33].

In this paper, we proposed a novel CNN-based 3D mandible segmentation approach named recurrent convolutional neural networks for mandible segmentation (RCNNSeg). RCNNSeg has the ability to accurately segment the detailed anatomical structures. Different from the classic EDCNN approaches that need to slice or truncate the CT scan into 2D slices or 3D patches during the segmentation process [33], the RCNNSeg approach can perform mandible segmentation on complete 3D CT scans. The proposed method adopts the structure of the recurrent neural networks that form a directed acyclic graph to enable recurrent connections between adjacent nodes to retain their connectivity. Each node then functions as a classic EDCNN to segment a single slice in the CT scan. The proposed segmentation structure enables one to identify the shape of the structures based on their anatomical connectivity. Our approach can perform 3D mandible segmentation on sequential data of any varied lengths and does not require large computation cost. Chen et al. [34] and Bai et al. [35] presented similar ideas that adopt a long short-term memory (LSTM) recurrent neural network [36] and a fully convolutional network (FCN) for image segmentation. However, FCN and convolutional LSTM were trained separately due to the huge demand of computing resources from the convolutional LSTM [34, 35]. Moreover, the method from [35] allows a batch size of 1 in the training of the LSTM. In order to solve the issue of lacking computing resources, we chose the vanilla RNN instead of LSTM, which allows the whole network to be trained in an end-to-end manner, while LSTM needs to be trained in the decoupled manner, as demonstrated in [34, 35]

Our major contributions are threefold. First, we proposed a novel mandible segmentation architecture that uses recurrent neural networks to facilitate connected units between adjacent slices in an anatomical sequence of scanning. Each unit is then implemented as

a classic encoder–decoder segmentation architecture for 2D slice segmentation. Second, the state-of-the-art 3D segmentation algorithms [37, 38] usually down-sample the raw data to overcome memory issues. In contrast, the implementation of the RCNNseg approach enables 3D segmentation in a way that reduces computational complexity of the model without loss of image quality. The 3D mandible segmentation can be performed for any varied lengths of the scan. Third, the proposed approach is able to tackle the problems of truncated inputs, which represent the main reason for the underperformance on detailed structures and regions that are corrupted by metal artifacts in CT scans. Extensive experiments were performed on two head and neck CT scans, and the proposed RCNNseg approach outperformed the existing segmentation approaches for automatic mandible segmentation.

The remainder of the paper is organized as follows: Section 4.2 introduces our proposed RCNNseg approach for automatic mandible segmentation. In Section 4.3, extensive experiments on two head and neck CT datasets are presented. The proposed method is compared both qualitatively and quantitatively with the EDCNN-based approaches as well as state-of-the-art methods for mandible segmentation. In Section 4.4, different aspects of the proposed method are discussed. Conclusions are given in Section 4.5.

4.2. MATERIALS AND METHODS

Due to extensive demands on computational resources, the encoder–decoder based mandible segmentation methods, such as Unet [29], SegNet [39], and AttUnet [40], cannot directly deal with the 3D medical images. Instead, three strategies are widely used for the feeding of 3D volumetric data, which are shown in Figure 4.1. The first two strategies in Figure 4.1a,b use either 2D slices or a few adjacent slices as inputs to feed into the EDCNNs, which are truncated along the z-axis from the volumetric data. The third one in Figure 4.1c uses 3D patches as inputs, which are cropped along all three orthogonal axes. These strategies are easy to implement and can deal with the large memory demands from deep learning algorithms; however, they cannot take into account the complete anatomical structures of the organs in CT scans that often frustrate the accurate segmentation of detailed structures.

Therefore, we propose a novel mandible segmentation approach, which adopts the structure of the recurrent neural network to build successive nodes that enable connectivity between adjacent nodes. Meanwhile, each node then functions as a classic 2D EDCNN to achieve mandible segmentation on a single CT slice.

4.2.1. FEEDFORWARD, FEEDBACK, AND RECURRENT NETWORKS

There are three main neuron connections within the visual cortex, namely feedforward, recurrent synapses, and feedback connections [41]. The feedforward (feedback) connections usually bring inputs (outputs) to a later (earlier) stage of the cortex region along (the opposite direction of) the processing pathway [42], which are widely implemented in most of the deep learning architectures [21, 43], including the classic EDCNNs [29, 39, 40]. Recurrent synapses, however, usually exceed the feedforward and feedback connections and interconnect neurons at the same stage of the pathway in the cortex [41, 42]. Due to

Figure 4.1 | Illustration of three prevalent strategies for the feeding of input data to the classic encoder-decoder-based convolutional neural networks. (a) Use of 2D EDCNN network for 2D slice-based object segmentation. (b) Use of 2D EDCNN for the image segmentation based on several adjacent slices from the volumetric data. (c) Use of 3D EDCNN based on 3D patches cropped from the complete volumetric data.

the interconnection mechanism of the recurrent neural networks, they are more applicable to tasks based on processing of sequential data.

4.2.2. RECURRENT CONVOLUTIONAL NEURAL NETWORKS FOR MANDIBLE SEGMENTATION (RCNNSEG)

We proposed a recurrent convolutional neural networks for mandible segmentation in order to accurately segment mandibles in the head and neck CT scans. Different from the recurrent convolutional neural networks for object recognition [42, 44] that build recurrent connections among the nodes at the same layer, the proposed RCNNSeg enables recurrent connections between adjacent successive units, each of which is then formed as a 2D segmentation network to process a single slice of the scan.

Our proposed architecture for 3D image segmentation can be observed in Figure 4.2a. The RCNNSeg takes a sequence of CT slices as inputs and then outputs a sequence of the corresponding segmented images of the same length. The RCNNSeg architecture can be unfolded to the structure shown in Figure 4.2b. RCNNSeg consists of varied lengths of

Figure 4.2 | The overall graphic scheme of the proposed methods. The architecture of RCNNSeg with two components: (a) The RCNNSeg and its loss drawn with recurrent connections. (b) The same seen as a time-unfolded computational graph, where each node is now associated with one particular time instance.

successive units to be able to process sequential data of different lengths. Each of the unit has a structure of the classic 2D EDCNN to evoke slice-based mandible segmentation.

At the t -th unit, the 2D EDCNN segmentation unit takes as input the t -th CT slice I_t concatenated with the output O_{t-1} from the previous time step $t - 1$. Then, the output of this unit is given as

$$O_t = w_f^T * \text{concat}(I_t, w_r^T * O_{t-1}), \quad (4.1)$$

where T is the transpose. w_f and w_r denote the feedforward weights in the 2D segmentation and the recurrent weights between the adjacent units, respectively.

The 2D segmentation unit adopts the classic EDCNN-based mandible segmentation approaches, such as U-Net [29], SegU-Net [39], and Att U-Net [40]. During the training, the feedforward weights w_f from the EDCNN and the recurrent weights w_r are shared with all the units at different time steps.

Loss function Given an input scan, we optimize the feedforward weights with the gradient-based optimization technique. The cost function at unit t is a combination of the Dice loss [45] and the binary cross entropy (BCE) loss [46], which is given as

$$\mathcal{L}^t = \alpha \times \mathcal{L}_{\text{BCE}}^t + \beta \times \mathcal{L}_{\text{Dice}}^t, \quad (4.2)$$

where α and β control the amount of contributions the BCE term and Dice term give in the loss function $\mathcal{L}$, respectively. For a detailed explanation and implementation, we refer the interested readers to [45, 46]. It is worth noting that these two loss functions have the potential to deal with imbalanced data, which fits well with our case.

Training of the model For the training of the RCNNSeg, we use the backpropagation through time (BPTT) [47] technique that begins by unfolding a recurrent neural network into successive units. As shown in Figure 4.2, the unfolded network contains N units, each

of which takes a slice as the input in the ordered sequence and outputs the corresponding segmentation result. According to Equation (6.4), the loss at the t -th step is denoted as $\mathcal{L}^t$. The total loss for a given ordered sequence of slices is the sum of the loss over all of the steps. We then optimize all of the parameters of the model based on the chain rule [48] to minimize the total loss.

4.2.3. EXPERIMENTAL SETUP

IMPLEMENTATION DETAILS

Our proposed approach was implemented in Pytorch [50]. All of our experiments were performed on a workstation equipped with Nvidia K40 GPUs of 12GB memory. We set the weights of the loss function to 0.5 for both the BCE loss term α and the Dice loss term β . We used Adam optimization [51] with a learning rate of $r = 10^{-4}$. The total number of epochs was 40 and 80 for the UMCG dataset and the PDDCA dataset, respectively. Moreover, an early stopping strategy was utilized if there was no improvement in the loss of the validation set for 10 epochs in order to avoid over-fitting.

EVALUATION METRICS

For quantitative analysis of the experimental results, several performance metrics are considered, including the Dice similarity coefficient (DSC), average symmetric surface distance (ASD), and 95% Hausdorff distance (95HD).

The Dice similarity coefficient (DSC) is often used to measure the consistency between two objects [52]. Therefore, it is widely applied as a metric to evaluate the performance of image segmentation algorithms. We do not elaborate further about the DSC and refer the readers to [52].

The average symmetric surface distance (ASD) [25] computes the average distance between the boundaries of two object regions. It is used to measure the error between the surfaces of the ground truth and the segmented regions. It is defined as

$$D_{ASD}(A, B) = \frac{d(A, B) + d(B, A)}{2}, \quad (4.3)$$

$$d(A, B) = \frac{1}{N} \sum_{\mathbf{a} \in A} \min_{\mathbf{b} \in B} \|\mathbf{a} - \mathbf{b}\|, \quad (4.4)$$

where $\|\cdot\|$ is the L_2 norm. $\mathbf{a}$ and $\mathbf{b}$ are the coordinates of the points on the boundary of objects A and B , respectively.

Hausdorff distance (HD) measures the maximum distance of a point in a set A to the nearest point in the other set B . It is defined as

$$D_{HD}(A, B) = \max(h(A, B), h(B, A)), \quad (4.5)$$

$$h(A, B) = \max_{\mathbf{a} \in A} \min_{\mathbf{b} \in B} \|\mathbf{a} - \mathbf{b}\|, \quad (4.6)$$

where $h(A, B)$ is often called the directed HD. The maximum HD is sensitive to contours. When the image is contaminated by noise or occlusion, the original Hausdorff distance is prone to mismatch [49, 53]. Huttenlocher et al. [54] proposed the concept of partial

Hausdorff distance. The 95HD is similar to maximum HD and selects 95% of the closest points in set B to the point in set A in Equation 6.13 to calculate $h(A, B)$

$$D_{95HD} = \max(h^{95\%}(A, B), h^{95\%}(B, A)), \quad (4.7)$$

$$h^{95\%}(A, B) = \max_{a \in A} \min_{b \in B^{95\%}} \|a - b\|. \quad (4.8)$$

The purpose of using this metric is to eliminate the impact of a very small subset of inaccurate segmentation on the evaluation of the overall segmentation performance.

4.3. RESULTS

In this section, we demonstrate the effectiveness of the proposed approach for mandible segmentation in CT scans. We verify this by both quantitative and qualitative experimental results on two datasets, namely, the local UMCG dataset and the public PDDCA dataset [49], and facilitate comparisons with the state-of-the-art methods.

4

4.3.1. THE UMCG HEAD AND NECK DATASET

The UMCG head and neck dataset was collected in the department of oral and maxillofacial surgery in the University Medical Center Groningen¹. This dataset contains 109 head and neck CT scans reconstructed with a kernel of Siemens Br64, I70h(s) or B70s. Each scan consists of 221 to 955 slices with a size of 512×512 pixels. The pixel spacing varies from 0.35 to 0.66 mm, and the slice thickness varies from 0.6 to 0.75 mm. The corresponding manual mandible segmentations were obtained by an experienced researcher using Mimics software version 20.0 (Materialise, Leuven, Belgium) and then confirmed by a clinician.

We compare our proposed RCNNSeg approach with the classic EDCNN approaches for mandible segmentation. For brevity, we refer to our methods as RUnetSeg, RAttUnetSeg, and RSegUnetSeg, which use U-Net [29], Att U-Net [40], and SegU-Net [39] as the base units of the proposed RCNNSeg, respectively. We also investigated the EDCNN approaches with three different strategies for the feeding of the 3D volumetric data. The three strategies are in the 2D, 2.5D, and 3D forms that take 2D slices, three consecutive 2D slices, and 3D patches as inputs, respectively. We randomly chose 90 cases as training, 2 cases as validation, and 17 cases as test models. The models were trained from scratch, and the training took approximately 40 hours, while the test on one scan took about 1.5 minutes.

Quantitative Results We evaluated the proposed RCNNSeg for mandible segmentation based on the metrics mentioned in 4.2.3. In Table 4.1 we list the results on DSC , D_{ASD} and D_{95HD} as well as the corresponding standard deviations. In general, our proposed RCNNSeg based approaches outperformed the EDCNN-based methods on mandible segmentation. For the DSC , all RCNNSeg-based approaches performed better than the corresponding EDCNNs with different strategies, with an improvement of at least 0.79%. Our proposed RCNNSeg also achieved the minimum error on D_{ASD} . On D_{95HD} , except for

¹<https://3dlabgroningen.nl/>

RUNetSeg, which underperformed the 2.5D U-Net, all other RCNNSeg-based approaches outperformed the corresponding EDCNNs. It is also surprising that the 3D patch EDCNNs dramatically underperformed when compared with all the other approaches.

Table 4.1 | Quantitative comparison of segmentation performance in the UMCG dataset between the proposed RCNNSeg and the classic EDCNNs. The values in the square brackets indicate the standard deviation of the corresponding measurements. We mark in bold the best performance in each metric.

	$DSC(\%)$	$D_{ASD}(mm)$	$D_{95HD}(mm)$
2D U-Net	95.95 [± 2.24]	0.3615 [± 0.3366]	4.0145 [± 4.6487]
2.5D U-Net	96.34 [± 1.99]	0.4053 [± 0.7565]	2.0154 [± 1.5949]
3D U-Net	77.73 [± 6.74]	17.2808 [± 6.0045]	133.7464 [± 22.8779]
RUNetSeg	97.53 [± 1.65]	0.2070 [± 0.2623]	2.3975 [± 4.6051]
2D SegU-Net	96.30 [± 2.06]	0.2794 [± 0.2447]	3.7958 [± 4.3662]
2.5D SegU-Net	96.69 [± 2.12]	0.4210 [± 0.6111]	4.9574 [± 7.5637]
3D SegU-Net	81.88 [± 7.14]	19.0109 [± 6.7765]	137.2283 [± 17.1781]
RSegUNetSeg	97.48 [± 1.70]	0.2170 [± 0.3491]	2.6562 [± 5.7014]
2D Att U-Net	94.21 [± 3.34]	0.6929 [± 0.8370]	5.1368 [± 3.2194]
2.5D Att U-Net	93.87 [± 2.89]	0.5188 [± 0.3327]	4.9223 [± 4.6204]
3D Att U-Net	83.92 [± 5.43]	16.2428 [± 3.9300]	124.1773 [± 14.2461]
RAttUNetSeg	96.57 [± 1.69]	0.2978 [± 0.2340]	2.4068 [± 1.5479]

Qualitative results We illustrate the 3D view of an example taken from the UMCG head and neck dataset in Figure 4.3. Compared to the ground truth in Figure 4.3a, the 2D and 2.5D EDCNN-based segmentation approaches shown in Figure 4.3b,c,f,g,j,k failed to segment the mandible detailed structures, such as the coronoids (indicated by red circles) and parts of the mandible body (indicated by the yellow circles). The 3D patch EDCNN approaches shown in Figure 4.3d,h,l managed to acquire most of the mandible segmented, but they also segmented part of the skull as mandible by mistake. The results of our proposed RCNNSeg approaches in Figure 4.3e,f,m show more accurate segmentation of those detailed structures with much less segmentation on other bone structures.

Figure 4.4 illustrates examples of three slices in the aforementioned encircled regions in a 2D slice view. The three images in Figure 4.4a are the corresponding ground truth (GT) segmentations. The images shown in Figure 4.4b,f,j (the middle column) are results from the 2D EDCNNs. The third and fourth columns in Figure 4.4 show the results from 2.5D EDCNNs and 3D patch EDCNNs. The last column gives the results from the proposed RCNNSeg approaches. Pink indicates the regions that are missed by the corresponding segmentation approaches. In general, the 2D and 2.5D EDCNN approaches missed more regions than the RCNNSeg approaches in segmenting the detailed structures. The 3D patch-based EDCNNs seem to segment the mandible quite well, but they also segment other bone structures, such as the skull, as mandibles.

Figure 4.3 | 3D view of a case from the UMCG Head and neck dataset. (a-m) Ground truth, 2D U-Net, 2.5D U-Net, 3D U-Net, RUnetSeg, 2D SegU-Net, 2.5D SegU-Net, 3D SegU-Net, RSegUnetSeg, 2D Att U-Net, 2.5D Att U-Net, 3D Att U-Net, and RAttUnetSeg. We use cyan color to indicate the correctly segmented mandible comparing to the ground truth. The pink ones are the regions that were missed by the algorithms while the yellow indicates the segmented regions that are not man. The red circles indicate the coronoids of the mandibles that are often missed by the traditional EDCNNs, yellow circles indicate parts of the mandible body.

4.3.2. PDDCA DATASET

We also evaluated the proposed pipeline on the public dataset PDDCA [49]. This dataset contains 48 patient CT scans from the Radiation Therapy Oncology Group (RTOG) 0522 study with manual segmentation of the left and right parotid glands, brainstem, optic chiasm, and mandible. Each scan consists of 76 to 360 slices with a size of 512×512 pixels. The pixel spacing varies from 0.76 to 1.27mm, and the slice thickness varies from 1.25 to 3.0 mm. We followed the same training and testing split as described in [49]. Forty out of the 48 patients in PDDCA with manual mandible annotations were used in previous studies [49, 55], in which the dataset was split into training and test subsets with 25 (0522c0001-0522c0328) and 15 (0522c0555-0522c0878) cases, respectively [49]. We used the pre-trained models obtained from the UMCG dataset and fine-tuned them on the training subset of the PDDCA dataset. We evaluated the performances of the models

Figure 4.4 | Examples of the automatic segmentation of mandibles in UMCG dataset. (a) Ground truth segmentation on the three examples (b-m) Segmentation results obtained on the example slices from 2D U-Net, 2.5D U-Net, 3D U-Net, RUnetSeg, 2D SegU-Net, 2.5D SegU-Net, 3D SegU-Net, RSegUnetSeg, 2D Att U-Net, 2.5D Att U-Net, 3D Att U-Net, and RAttUnetSeg. We use cyan color to indicate the correctly segmented mandible comparing to the ground truth. The pink ones are the regions that were missed by the algorithms while the yellow is the segmented non-mandible regions.

on the test subset.

Quantitative results Table 4.2 shows the quantitative evaluation based on the average DSC, ASD, and 95HD used in the challenge [25, 49]. The achieved performance indicates that our proposed RCNNSeg-based methods outperformed the corresponding EDCNN-based approaches with different input feeding strategies regarding all evaluation metrics.

In Table 4.3 we also compare our proposed approach with the state-of-the-art methods on the PDDCA dataset. We mark in bold the best three performances for each evaluation metric. In general, our RCNNSeg-based approaches outperformed the other existing approaches. Only the RAttUnetSeg model underperformed when compared with some of

Table 4.2 | Quantitative comparison of the segmentation performance between the proposed RCNNSeg-based approaches and the EDCNN-based methods on the PDDCA dataset. The values in the square brackets indicate the standard deviation of the corresponding measurements. We mark in bold the best performance in each metric.

	$DSC(\%)$	$D_{ASD}(\text{mm})$	$D_{95HD}(\text{mm})$
2D U-Net	94.15[± 1.31]	0.1827[± 0.0915]	2.0547[± 1.4431]
2.5D U-Net	94.19[± 1.25]	0.1915[± 0.0669]	1.7512[± 0.6539]
3D U-Net	91.85[± 5.32]	3.7577[± 6.1869]	36.9138[± 66.9059]
RUNetSeg	94.71[± 1.35]	0.1353[± 0.0614]	1.4098[± 0.8573]
2D SegU-Net	94.69[± 1.33]	0.1765[± 0.0671]	1.5067[± 0.6938]
2.5D SegU-Net	94.76[± 1.20]	0.1532[± 0.0622]	1.6856[± 0.6426]
3D SegU-Net	93.08[± 2.80]	2.4289[± 5.8637]	24.1133[± 62.1808]
RSegUnetSeg	95.10[± 1.21]	0.1367[± 0.0382]	1.356[± 0.4487]
2D Att U-Net	92.99[± 1.25]	0.2924[± 0.2523]	3.1848[± 4.0571]
2.5D Att U-Net	92.75[± 1.34]	0.2502[± 0.0887]	2.1815[± 1.0656]
3D Att U-Net	90.14[± 8.50]	6.3894[± 11.7528]	54.2182[± 72.3141]
RAttUnetSeg	93.87[± 1.29]	0.1773[± 0.0515]	1.6397[± 0.6219]

the existing approaches [56–58] on DSC and 95HD.

4.4. DISCUSSION

Over the last few decades, many traditional computer vision algorithms [13–15] and CNN-based methods [23–25, 28, 56–59] have been proposed for head and neck CT segmentation. However, there are still significant challenges to completely automate the segmentation of the mandible from CT scans, whereas manual delineation is time consuming and has high inter-rater variabilities [7]. The quantitative comparison in Table 4.3 show that the CNN algorithms outperform the traditional SSM- or atlas-based methods in general. Therefore, the development of CNN algorithms can bring significant benefits to automatically segment the mandible. However, these methods are proposed for OAR segmentation in radiotherapy rather than 3D VSP.

In this paper, we present a robust end-to-end deep learning approach for accurate mandible segmentation in 3D VSP. The proposed RCNNSeg approach adopts the structure of the recurrent convolutional neural networks to enable connections between adjacent slices in the CT volume. Unlike the classic encoder–decoder-based approaches that need to truncate the 3D volume, our approach can perform 3D mandible segmentation on sequential data of any varied lengths and does not require large computational resources.

Quantitative and qualitative evaluation on two datasets demonstrates that our proposed approach is robust in mandible segmentation, as illustrated in Figure 4.3–4.4 and Table 4.1–4.3. Comparisons with the classic EDCNN approaches and state-of-the-art approaches illustrate that the RCNNseg approach significantly improves segmentation of the mandible. It is worth noting that RCNNSeg is very robust for the segmentation of the weak and blurry boundaries (for instance, the coronoids of mandibles).

The experimental results show that the proposed RCNNSeg is feasible and effective

Table 4.3 | Comparison of segmentation performance between the state-of-the-art methods and our proposed RCNNSeg approach; bold font indicates the best three performers for each measurement.

	DSC (%)	D_{ASD} (mm)	D_{95HD} (mm)
Multi-atlas [13]	91.70[±2.34]	-	2.49[±0.76]
AAM [14]	92.67[±1.00]	-	1.98[±0.59]
ASM [15]	88.13[±5.55]	-	2.83[±1.18]
CNN [23]	89.50[±3.60]	-	-
NLGM [59]	93.08[±2.36]	-	-
AnatomyNet [24]	92.51[±2.00]	-	6.28[±2.21]
FCNN [25]	92.07[±1.15]	0.51[±0.12]	2.01[±0.83]
FCNN+SRM [25]	93.60[±1.21]	0.37[±0.11]	1.50[±0.32]
CNN+BD [57]	94.60[±0.70]	0.29[±0.03]	-
HVR [56]	94.40[± 1.30]	0.43[± 0.12]	-
Cascade 3D U-Net [58]	93.00[±1.90]	-	1.26[±0.50]
Multi-view[28]	94.10[±0.70]	0.28[±0.14]	-
RUnetSeg	94.71[±1.35]	0.135[±0.061]	1.410[±0.857]
RSegUnetSeg	95.10[±1.21]	0.137[±0.038]	1.356[±0.449]
RAttUnetSeg	93.87[±1.29]	0.177[±0.052]	1.640[±0.622]

for 3D mandible segmentation in CT scans. The RCNNSeg approach can also be applied to other segmentation tasks that need to deal with 3D sequential data. The architecture of the RCNNSeg adopts the structure of the recurrent neural network that facilitates recurrent connections between adjacent slices. The proposed approach implemented the strategy that takes advantage of the anatomical prior information from the previous slice and then segments the current CT slice.

This strategy takes advantage of shape prior information, which considers the segmentation result of the previous slice as the shape prior to the segmentation of the current slice. Therefore, the proposed approach can help to identify the continuous structure of the mandible in 2D segmentation networks. In addition, the proposed approach utilizes the RNN module that helps the extraction of spatial information of the object based on the collection of context and shape information. This strategy can support further research on 3D image segmentation. It can also help alleviate memory issues for 3D medical image segmentation as well as segmentation tasks on sequential data, such as video images.

In this study, we used the 109 H&N CT dataset and a small public dataset to validate our method, which cannot satisfactorily represent the average patient population that requires tumor resection surgery. In practical automatic segmentation simulation, larger amounts of data across regions should be further explored. Moreover, orthognathic surgery and complex trauma have been widely reported in using 3D VSP [60]. Most of CT scans in the datasets exclude metal implants or dental braces that often lead to noisy and blurry structures and also make the segmentation tasks difficult. Validation on even more scans with dental braces or metal implants can be performed. In addition, automatic mandible segmentation from MRI images may be required in the future, since MRI-based 3D surgery planning workflow has been developed [61]. In our future research, further validation and evaluation will be performed to determine whether our proposed

approach is effective for real clinical practice.

4.5. CONCLUSION

We propose an end-to-end approach for the accurate segmentation of the mandible from H&N CT scans. Our approach incorporates the encoder–decoder-based segmentation algorithms into recurrent connections and uses a combination of Dice and BCE as the loss function. We implemented the proposed approach on 109 H&N CT scans from our dataset and 40 scans from the PDDCA public dataset. The experimental results show that the RSegCNN strategy can yield more significant performance than that of the conventional algorithms in terms of quantitative and qualitative evaluation. The proposed RSegCNN has potential for automatic mandible segmentation by learning spatial structured information.

REFERENCES

- [1] S. B. Edge, D. R. Byrd, M. A. Carducci, C. C. Compton, A. Fritz, F. Greene, *et al.*, *AJCC cancer staging manual*, Vol. 649 (Springer New York, 2010).
- [2] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, *Global cancer statistics 2018: Globocan estimates of incidence and mortality worldwide for 36 cancers in 185 countries*, a cancer journal for clinicians **68**, 394 (2018).
- [3] J. P. Shah and Z. Gil, *Current concepts in management of oral cancer–surgery*, Oral oncology **45**, 394 (2009).
- [4] L. Ciocca, S. Mazzoni, M. Fantini, F. Persiani, C. Marchetti, and R. Scotti, *CAD/CAM guided secondary mandibular reconstruction of a discontinuity defect after ablative cancer surgery*, Journal of cranio-maxillo-facial surgery **40**, 511 (2012).
- [5] J. Kraeima *et al.*, *Multi-modality 3D mandibular resection planning in head and neck cancer using CT and MRI data fusion: A clinical series*, Oral Oncology **81**, 22 (2018).
- [6] G. Succo, M. Berrone, B. Battiston, P. Tos, F. Goia, P. Appendino, and E. Crosetti, *Step-by-step surgical technique for mandibular reconstruction with fibular free flap: Application of digital technology in virtual surgical planning*, European archives of oto-rhino-laryngology **272**, 1491 (2014).
- [7] J. Wallner, I. Mischak, and J. Egger, *Computed tomography data collection of the complete human mandible and valid clinical ground truth models*, Scientific data **6**, 190003 (2019).
- [8] W. Jeanneret-Sozzi, R. Moeckli, J.-F. Valley, A. Zouhair, E. M. Ozsahin, R.-O. Mirimanoff, *et al.*, *The reasons for discrepancies in target volume delineation*, Strahlentherapie und Onkologie **182**, 450 (2006).
- [9] A. K. Hoang Duc, G. Eminowicz, R. Mendes, S.-L. Wong, J. McClelland, M. Modat, M. J. Cardoso, A. F. Mendelson, C. Veiga, T. Kadir, *et al.*, *Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer*, Medical physics **42**, 5027 (2015).
- [10] T. J. Huff, P. E. Ludwig, and J. M. Zuniga, *The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning*, Expert review of medical devices **15**, 349 (2018).
- [11] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. van Ooijen, *Automatic segmentation of the mandible from computed tomography scans for 3d virtual surgical planning using the convolutional neural network*, Physics in Medicine & Biology **64**, 175020 (2019).
- [12] S. T. Gollmer and T. M. Buzug, *Fully automatic shape constrained mandible segmentation from cone-beam CT data*, in *IEEE 9th International Symposium on Biomedical Imaging* (2012) pp. 1272–1275.

- [13] A. Chen and B. Dawant, *A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images*, in *The MIDAS Journal - Head and Neck Auto Segmentation Challenge* (2015).
- [14] R. Mannion-Haworth, M. Bowes, A. Ashman, G. Guillard, A. Brett, and G. Vincent, *Fully automatic segmentation of head and neck organs using active appearance models*, in *The MIDAS Journal - Head and Neck Auto Segmentation Challenge* (2015).
- [15] T. Albrecht, T. Gass, C. Langguth, and M. Lüthi, *Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning*, in *The MIDAS Journal - Head and Neck Auto Segmentation Challenge* (2015).
- [16] N. Torosdagli et al., *Robust and fully automated segmentation of mandible from CT scans*, in *IEEE 14th International Symposium on Biomedical Imaging* (2017) pp. 1209–1212.
- [17] Y. J. Chuang, B. M. Doherty, N. Adluru, M. K. Chung, and H. K. Vorperian, *A novel registration-based semiautomatic mandible segmentation pipeline using computed tomography images to study mandibular development*, *Journal of computer assisted tomography* **42**, 306 (2017).
- [18] T. Blaschke, C. Burnett, and A. Pekkarinen, *Image segmentation methods for object-based analysis and classification*, in *Remote sensing image analysis: Including the spatial domain* (Springer, 2004) pp. 211–236.
- [19] I. Bankman, *Handbook of medical image processing and analysis* (Elsevier, 2008).
- [20] S. Yuheng and Y. Hao, *Image segmentation algorithms overview*, arXiv preprint arXiv:1707.02051 (2017).
- [21] K. He, X. Zhang, S. Ren, and J. Sun, *Deep residual learning for image recognition*, in *2016 IEEE Conference on Computer Vision and Pattern Recognition* (2016) pp. 770–778.
- [22] K. Simonyan and A. Zisserman, *Very deep convolutional networks for large-scale image recognition*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (2015).
- [23] B. Ibragimov and L. Xing, *Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks*, *Medical Physics* **44**, 547 (2017).
- [24] W. Zhu et al., *Anatomynet: Deep 3D squeeze-and-excitation U-Nets for fast and fully automated whole-volume anatomical segmentation*, arXiv preprint arXiv:1808.05238 (2018).
- [25] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, *Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks*, *Medical Physics* **45**, 4558 (2018).

- [26] J. Egger, B. Pfarrkirchner, C. Gsaxner, L. Lindner, D. Schmalstieg, and J. Wallner, *Fully convolutional mandible segmentation on a valid ground truth dataset*, in *Annual International Conference IEEE Engineer Medical Biology Society* (2018) pp. 656–660.
- [27] J. Long, E. Shelhamer, and T. Darrell, *Fully convolutional networks for semantic segmentation*, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2015) pp. 3431–3440.
- [28] S. Liang, K. Thung, D. Nie, Y. Zhang, and D. Shen, *Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images*, *IEEE Transactions on Medical Imaging* **39**, 2794 (2020).
- [29] 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* (Springer, 2015) pp. 234–241.
- [30] J. C. Ye and W. K. Sung, *Understanding geometry of encoder-decoder CNNs*, in *Proceedings of the 36 th International Conference on Machine Learning, Long Beach, California*, PMLR 97 (2019).
- [31] R. B. Ger, D. F. Craft, D. S. Mackin, S. Zhou, R. R. Layman, A. K. Jones, H. Elhalawani, C. D. Fuller, R. M. Howell, H. Li, *et al.*, *Practical guidelines for handling head and neck computed tomography artifacts for quantitative image analysis*, *Computerized Medical Imaging and Graphics* **69**, 134 (2018).
- [32] J. Minnema, M. van Eijnatten, A. A. Hendriksen, N. Liberton, D. M. Pelt, K. J. Batenburg, T. Forouzanfar, and J. Wolff, *Segmentation of dental cone-beam ct scans affected by metal artifacts using a mixed-scale dense convolutional neural network*, *Medical Physics* **46**, 5027 (2019).
- [33] Q. Yu, Y. Xia, L. Xie, E. K. Fishman, and A. L. Yuille, *Thickened 2d networks for efficient 3d medical image segmentation*, arXiv preprint arXiv:1904.01150 (2019).
- [34] J. Chen, L. Yang, Y. Zhang, M. Alber, and D. Z. Chen, *Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation*, *Advances in neural information processing systems* **29**, 3036 (2016).
- [35] W. Bai, H. Suzuki, C. Qin, G. Tarroni, O. Oktay, P. M. Matthews, and D. Rueckert, *Recurrent neural networks for aortic image sequence segmentation with sparse annotations*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2018) pp. 586–594.
- [36] S. Hochreiter and J. Schmidhuber, *Long short-term memory*, *Neural computation* **9**, 1735 (1997).
- [37] 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.

- [38] 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* (2016) pp. 565–571.
- [39] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, *Automatic traffic sign detection and recognition using SegU-Net and a modified Tversky loss function with L1-constraint*, *IEEE Transactions on Intelligent Transportation Systems* **21**, 1467 (2020).
- [40] O. Oktay et al., *Attention U-Net: learning where to look for the pancreas*, arXiv preprint arXiv:1804.03999 (2018).
- [41] L. F. Abbott and P. Dayan, *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems* (The MIT Press, 2001).
- [42] M. Liang and X. Hu, *Recurrent convolutional neural network for object recognition*, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2015) pp. 3367–3375.
- [43] C. Szegedy et al., *Going deeper with convolutions*, in *IEEE Proceedings Computer Vision and Pattern Recognition* (2015).
- [44] S. Lai, L. Xu, K. Liu, and J. Zhao, *Recurrent convolutional neural networks for text classification*, in *AAAI conference on artificial intelligence* (2015) p. 2267–2273.
- [45] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, *Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations*, in *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Springer, 2017) pp. 240–248.
- [46] S. A. Taghanaki et al., *Combo loss: Handling input and output imbalance in multi-organ segmentation*, *Computerized Medical Imaging and Graphics* **75**, 24 (2019).
- [47] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *Nature* **521**, 436 (2015).
- [48] R. L. Magin, *Fractional calculus in bioengineering* (Begell House Redding, 2006).
- [49] P. F. Raudaschl et al., *Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015*, *Medical Physics* **44**, 2020 (2017).
- [50] A. Paszke et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems* (2019) pp. 8024–8035.
- [51] D. P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, arXiv preprint arXiv:1412.6980 (2014).
- [52] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. Uden, C. I. Sanchez, G. Litjens, F.-E. Leeuw, B. Ginneken, E. Marchiori, and B. Platel, *Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities*, *Scientific Reports* **7**, 1 (2017).

- [53] A. A. Taha and A. Hanbury, *Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool*, BMC medical imaging **15**, 1 (2015).
- [54] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, *Comparing images using the Hausdorff distance under translation*, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1992) pp. 654–656.
- [55] X. Ren et al., *Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images*, Medical Physics **45**, 2063 (2018).
- [56] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, *Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning*, IEEE Transactions on Image Processing **27**, 923 (2017).
- [57] O. Kodym, M. Španěl, and A. Herout, *Segmentation of head and neck organs at risk using CNN with batch Dice loss*, in *German Conference on Pattern Recognition* (2018) pp. 105–114.
- [58] Y. Wang, L. Zhao, Z. Song, and M. Wang, *Organ at risk segmentation in head and neck ct images by using a two-stage segmentation framework based on 3D U-Net*, arXiv preprint arXiv:1809.00960 (2018).
- [59] M. Orbes-Arteaga, D. Pea, and G. Dominguez, *Head and neck auto segmentation challenge based on non-local generative models*, in *The MIDAS Journal - Head and Neck Auto Segmentation Challenge* (2015).
- [60] J. Castro-Núñez, J. M. Shelton, S. Snyder, and J. Sickels, *Virtual surgical planning for the management of severe atrophic mandible fractures*, Craniomaxillofacial trauma & reconstruction **11**, 150 (2018).
- [61] A. M. Hoving, J. Kraeima, R. H. Schepers, H. Dijkstra, J. H. Potze, B. Dorgelo, and M. J. Witjes, *Optimisation of three-dimensional lower jaw resection margin planning using a novel black bone magnetic resonance imaging protocol*, PloS one **13**, e0196059 (2018).

III

MANDIBLE SEGMENTATION ON CBCT IMAGE

5

ROBUST AND ACCURATE MANDIBLE SEGMENTATION ON DENTAL CBCT SCANS AFFECTED BY METAL ARTIFACTS USING A PRIOR SHAPE MODEL

Bingjiang Qiu, Hylke V. D. Wel, Joep Kraeima, Haye H. Glas, Jiapan Guo, Ronald J. H. Borra, Max J. H. Witjes and Peter M. A. van Ooijen.

Qiu, Bingjiang, et al. "Robust and Accurate Mandible Segmentation on Dental CBCT Scans Affected by Metal Artifacts Using a Prior Shape Model." *Journal of Personalized Medicine* 11.5 (2021): 364.

ABSTRACT

Accurate mandible segmentation is significant in the field of maxillofacial surgery to guide clinical diagnosis and treatment and develop appropriate surgical plans. In particular, cone-beam computed tomography (CBCT) images with metal parts, such as those used in oral and maxillofacial surgery (OMFS), often have susceptibilities when metal artifacts are present such as weak and blurred boundaries caused by a high-attenuation material and a low radiation dose in image acquisition. To overcome this problem, this paper proposes a novel deep learning-based approach (SASeg) for automated mandible segmentation that perceives overall mandible anatomical knowledge. SASeg utilizes a prior shape feature extractor (PSFE) module based on a mean mandible shape, and recurrent connections maintain the continuity structure of the mandible. The effectiveness of the proposed network is substantiated on a dental CBCT dataset from orthodontic treatment containing 59 patients. The experiments show that the proposed SASeg can be easily used to improve the prediction accuracy in a dental CBCT dataset corrupted by metal artifacts. In addition, the experimental results on the PDDCA dataset demonstrate that, compared with the state-of-the-art mandible segmentation models, our proposed SASeg can achieve better segmentation performance.

Keywords: accurate mandible segmentation; oral and maxillofacial surgery; 3D virtual surgical planning (3D VSP); convolutional neural network

5.1. INTRODUCTION

Currently, the three-dimensional (3D) virtual surgical planning (VSP) technique is commonly used for oral and maxillofacial surgery (OMFS), and planning since it allows for pre- or postoperative simulation of surgical options [1]. 3D surface models of the mandible in 3D VSP are created and superimposed to visually and quantitatively demonstrate the orthodontic/orthognathic changes and provide postoperative follow-up of patients with cranio-maxillofacial deformities [1]. Cone-beam computed tomography (CBCT) is widely applied in 3D VSP because of its lower radiation dose and faster scanning time than conventional CT [2]. In orthodontic or orthognathic treatment, the dentist or maxillofacial surgeon needs visual information about the location and movement of their patient's teeth and mandible. A requirement for this process is to accurately segment the mandible from the dental CBCT scans and then to generate 3D surface mandible model. Therefore, accurate mandible segmentation plays an important role in 3D VSP for OMFS. Dental CBCT scans are noisier and have more metal artifacts than conventional CTs because dental CBCTs use a low-radiation technique and teeth, dental braces in orthodontic treatment and metal implants in orthognathic treatment are higher attenuation materials, easily leading to high noise and strong metal artifacts in the visual impression of the scans [3]. The boundaries of mandibles are difficult to be identified since dental braces and metal implants badly affect the image quality in CBCT [4], as shown in Figure 5.1. Additionally, low contrast in the condylar process very often leads to ambiguous and blurred boundaries in CBCT scans because of its low radiation dose, as illustrated in Figure 5.1. Consequently, the main difficulty in orthodontics or orthognathic visualization is precise mandible segmentation in CBCT scans. Currently, manual segmentation for 3D modeling of the mandible is widely adopted in clinical practice, but this is a time-consuming and labor-intensive approach so that it is impractical to perform on a large number of subjects. Moreover, manual segmentation often suffers from large interoperator variability (Dice score of 94.09% between two clinical experts) [5], which directly influences the quality of treatment planning. To date, there are still no reliable automatic segmentation approaches that can adapt to badly affected CBCT scans. Accordingly, it is meaningful to develop an accurate and automatic technique to segment the mandible for orthodontic or orthognathic treatment from CBCT images.

To reduce the workload of mandible segmentation, a number of (semi)automatic segmentation methods on CBCT have been developed. Wand et al. [6] proposed an automated segmentation method from dental CBCT images using patch-based sparse representation and convex optimization. They used the B-spline registration algorithm provided by the Elastix toolbox for the deformable registration. Furthermore, the average computational time was approximately 5 h for segmentation of a scan of size $400 \times 400 \times 400$ [6]. Fan et al. [7] used a marker-based watershed transform method for fully automatic mandible segmentation from CBCT images. This approach used a Gaussian filter and manually selected an adequate threshold for preprocessing initialization. Gollmer et al. [8] employed a statistical shape model (SSM) with optimized correspondence, which can help to improve the segmentation accuracy of the mandible. This method needs to manually and intentionally choose the proper mandible prior as initialization and a suitable window width of the segmentation object. Oscar et al. [9] proposed an interactive segmentation method

Figure 5.1 | Example illustrations that challenge mandible segmentation in CBCT images. (a) Original CBCT image. The mandible and teeth appear with almost invisible boundaries. (b) Example of manual annotation. Low contrast often appears in condyles. The purple region indicates the manual annotation of the mandible.

to aid specialists in segmenting CBCT via applying supervoxels and graph clustering techniques. Wang et al. [10] presented a majority voting method and combined it with a random forest for mandible segmentation in dental CBCT. These approaches have been proven to be useful in utilizing prior knowledge in the segmentation tasks [11]. However, these methods suffer from the problem of hyperparameter selection and the manual positioning of landmarks in the initialization steps [12]. The performance of the methods is often affected by high noise or metal artifacts caused by dental braces or implants [13, 14]. In other words, the application of these traditional techniques requires expert analysis and adjustment in every specific patient, which makes it difficult to handle the massive amount of medical data encountered in practice.

With the development of deep learning technology, deep learning methods have been proven to show powerful capabilities in detailed image feature extraction in automatic segmentation tasks [15–17]. The deep learning approach provides much more flexibility than the traditional methods [6–10], which require less expert analysis and fine-tuning and can easily exploit the other objects [18]. However, these studies [15–17] are still not robust in segmenting organs because the medical image usually has a 3D volume form, but 2D slices are usually fed as the input. For instance, Minnema et al. [13] employed a mixed-scale dense convolutional neural network to segment the mandible in CBCT. However, the other bone-structured organs are also segmented in this network, as shown

in the figures of [13], due to the fact that the spatial information was not considered in their 2D network. To use 3D spatial information from the volume data, researchers first proposed to use a 3D network instead of the original 2D network. Çiçek et al. [19], Milletari et al. [20], Zhu et al. [21], and Wang et al. [22] explored a 3D convolution kernel in their network instead of the original 2D kernel in the medical image segmentation task. However, fully implementing the 3D network for image segmentation requires cropping the input volume into a small patch in training, which limits the maximum receptive field of the network and leads to the loss of global information. It is difficult for the 3D network to learn the overall structure information of the target. Thus, researchers started to investigate learning 3D spatial information and voxel connectivity of the upper and lower context of the object via a 2D network. Mortazi et al. [23] proposed the multiplanar training strategy, which utilized the images' three perpendicular planes (axial, coronal, and sagittal) as input. Novikov et al. [24] used a sequence of slices as input in the network. Qiu et al. [12] adopted a 2.5D volume as input in their network and then combined the resulting 2D segmentations from three orthogonal planes into a 3D segmentation. Ghavami et al. [25] incorporated different numbers of neighboring slices as input for prostate segmentation from ultrasound images. Qiu et al. [26] developed a novel technique that combined a regular segmentation network with a recurrent module in their network for mandible segmentation in conventional CT scans. In general, these works [12, 23–26] have shown that using adjacent slices can help the network obtain more accurate and reliable results in terms of anatomy. Nevertheless, these works [12, 23–26] have also shown that obtaining 3D spatial information from a 2D network still leaves room for improvement. As illustrated in Figure 5.2, an example shows the comparison between SegUnet [27], recurrent SegUnet (RSegUnet) [26] and the proposed method in mandible segmentation. These methods are not suitable for CBCT images that are strongly corrupted by the metal braces and low contrast due to the inherent characteristic of CBCT.

Figure 5.2 | An example illustration that shows the comparison with SegUnet and RSegUnet. The existing state-of-the-art convolutional neural network (CNN) approaches for segmentation tasks perform poorly when the input data are strongly degenerated by noise. (a) the ground truth segmentation; (b)–(d) the automatic segmentation results obtained from SegUnet [27], RSegUnet [26] and the proposed SASeg. The purple region indicates the manual annotation, while the green regions indicate automatic segmentations.

With the development of conventional approaches and the deep segmentation neural network, introducing prior information to the network has become a popular topic in research. Chen et al. [28] tried to use a shape prior in their segmentation model. They first used the deep Boltzmann machine to extract the prior shape hierarchy, which can capture global and local structures of the prior shapes. Then, the learned structure is introduced in an energetic form to regularize the target shape for image segmentation [28]. Duan et al. [29] proposed a context-aware 3D fully convolutional network (FCN) for vessel enhancement and segmentation in coronary computed tomography angiography volume, which used 3D vessel-like structures as spatial prior knowledge to feed the 3D FCN. Tong et al. [30] developed a novel automated segmentation approach incorporating shape priors as a constraint term, where they combined an FCN with a shape representation model. However, how to define an appropriate prior shape model to guide object segmentation is still an open problem.

Motivated by the above analysis, we propose a novel convolutional neural network (CNN) framework based on shape-aware segmentation for mandible segmentation (SASeg), which follows the classical encoder–decoder structure. We first adopt a prior shape feature extractor (PSFE) to capture the shape information from the mean mandible shape model, in which the mean model is implemented in the average mandible shapes in the training set. In this way, the network can be aware of the general mandible shape information. To avoid the large computation and memory demand in the 3D network and utilize spatial information from the 3D volume, we feed the mean mandible shape feature from PSFE into a recurrent FCN for mandible segmentation at the pixel level.

In particular, our main contributions are as follows: (1) We present a novel end-to-end method for dental CBCT mandible segmentation based on prior shape information. (2) We propose a PSFE module to extract the spatial mandible information from the mean mandible shape model. (3) The proposed method provides the potential capability of removing the need for post-processing steps, even in cases where the images are corrupted by metal artifacts or noise due to the limitations in their image acquisition.

5.2. METHODOLOGY

In this section, we elaborate on the construction of the proposed SASeg and its core modules (i.e., PSFE, recurrent SegUnet). Figure 5.3 demonstrates the proposed SASeg, which builds recurrent SegUnet (RSegUnet) connections between adjacent units to retain the connectivity of mandible anatomy and utilizes the common mandible shape feature by extracting the mean mandible model from PSFE. Each unit in the RSegUnet is implemented as a classic end-to-end segmentation architecture for 2D slice segmentation. Moreover, every unit consists of an encoder and a decoder, as indicated by the blue and yellow ladder blocks in Figure 5.3. The PSFE module is inserted at the last layer of the encoder to capture mandible prior information, as indicated by the green ladder block in Figure 5.3.

Specifically, we first use the preprocessing technique of the statistical shape model (SSM) [31, 32] to generate a mean mandible shape structure in the training set based on manual expert segmentations. In this way, every mandible shape is represented in the average mandible shape. The average mandible shape is then introduced into the

proposed framework as a shape prior input in the network. To obtain more abundant structure knowledge, we employ a prior shape feature extractor (PSFE) as an auxiliary path to encode the average mandible shape so that the common features of the mandibles can be obtained. PSFE makes use of the mandible shape information from the mean model to help supervise the modeling of the target area, which is helpful to refine the segmentation performance. Furthermore, we employ the recurrent SegUnet in the framework, which has been proven helpful to segment 3D objects in 2D networks [26]. The proposed network that extracts prior knowledge from the mean shape model is able to constrain the mandible shape consistency in the segmentation task. Details of PSFE and the recurrent SegUnet in the proposed method are introduced as follows.

Figure 5.3 | Overview of the proposed SASeg and its corresponding unfolded computational graph. The PSFE module is leveraged to extract general shape features from a mean mandible shape, and recurrent SegUnet connections are used to conduct the slice-by-slice segmentation.

5.2.1. BUILDING A MEAN MANDIBLE SHAPE MODEL

We present a brief overview of the mean mandible shape model generation process, as illustrated in Figure 5.4. For generating a mean mandible shape model, we use a similar method for the preprocessing of SSM [31]. A dataset of mandible shapes is needed for generating a mean mandible shape model. The required mandible shapes are obtained via manual segmentation by an experienced technical physician and surface processing methods in the training set. The mandible shapes vary widely in terms of rotation, scale, and position of the object [33]. Therefore, all n mandibles need to be aligned into a common coordinate frame. This can be achieved by applying a generalized Procrustes alignment to remove all information that is unrelated to shape (i.e., rotation, position and scaling information) [34]. The mean mandible shape model is built from the set of n rigidly-aligned input shapes, which have been remeshed so the mandible shape is described by l landmark points which are forming a triangulation mesh over the contour of the mandible with edge lengths of ± 1 mm. Using a surface registration method based on elastic deformation [35], correspondence between the landmark points on the input shapes was established. Due to the registration steps, the variation in position of the

corresponding landmarks on the input mandibles now represents the shape variation for each landmark. Based on this, the mean mandible shape was calculated. To adapt the mean mandible model to the network, we voxelize the mean 3D mandible geometry into the same image coordinate system. The data processing pipeline for generation of the mean mandible shape model fed to the network is presented in Figure 5.4. The main stages of the method are illustrated as follows: preparation of training data, manual segmentation, surface processing (i.e., Procrustes alignment), building a mean mandible model, and voxelization back to the image coordinate system. For convenience, we still use the mean mandible shape model to represent the voxelized mean mandible shape model.

Figure 5.4 | The procedure of generation of the mean mandible shape model. The generation consists of preparing the mandible model, generating the mesh surfaces, creating the mean mandible shape model, and calculating the corresponding mask stack.

5.2.2. PRIOR SHAPE FEATURE EXTRACTOR (PSFE)

Using a mean shape model as the global contextual prior has proven to be a promising approach and is commonly used in conventional image segmentation tasks [36, 37]. These works motivated our development of the prior shape feature extractor (PSFE) using the mean mandible shape model as a shape prior in the segmentation network. The PSFE architecture consists of one convolution block (a 3×3 convolution with a stride of 2, batch normalization (BN), ReLU) and two residual blocks (two depthwise separable convolutions (DSConv) [38], BN, ReLU, Maxpooling), as shown in Figure 5.5. To capture sufficiently large amount of contextual information, the feature maps are gradually downsampled in PSFE architectures. The mean mandible shape model is fed into the PSFE to learn mandible prior information to improve the performance of mandible segmentation with regard to complex situations such as blurred dental CBCTs, as illustrated in PSFE of Figure 5.5. The PSFE module is connected to the last bottom layer of the encoder, which is introduced in Section 5.2.3. The PSFE module learns θ_{PSFE} from

the prior mandible information obtained from the mean mandible shape model. It can reduce oversegmentation (i.e., false positive prediction) in the segmentation task.

DSConv [38] in PSFE is improved from the Inception v3 structure [39]. DSConv makes the network processing simpler and more effective [38], especially in regard to increasing the computational efficiency. DSConv consists of depthwise convolutions and pointwise convolutions (i.e., 1×1 convolutions). Depthwise convolution first performs a 3×3 convolution operation for each input channel. Then, pointwise convolution performs 1×1 convolution to fuse the feature maps after the depthwise convolution. Assume that the number of channels in the i -th layer is C_i . The channels are independent of each other, so the number of parameters of the convolution kernel is $3 \times 3 \times C_i$, which is much less than $3 \times 3 \times C_i \times C_{i-1}$ of standard convolution. After channel-by-channel convolution, a 1×1 convolution kernel is used for feature fusion between channels. Therefore, the number of parameters of the second half of the convolution kernel is $1 \times 1 \times C_{i-1} \times C_i$. Furthermore, the calculation amount of the DSConv is $3 \times 3 \times C_{i-1} \times H \times W + 1 \times 1 \times C_{i-1} \times C_i \times H \times W$, where H and W present the weight and height of the feature maps, respectively. The calculation amount of using DSConv is $\frac{1}{C_i} + \frac{1}{9}$ of the standard convolution, which is $3 \times 3 \times C_{i-1} \times C_i \times H \times W$. Therefore, the amount of calculation is greatly reduced. Moreover, PSFE does not require the training of multiple models and a large number of extra model parameters. Here, the number of trainable parameters of PSFE in SASeg is 149,440. Furthermore, using PSFE as an auxiliary path to extract the mean mandible shape information can be easily applied in a segmentation network with a small increase in the amount of memory and computational resources.

5.2.3. RECURRENT CONVOLUTIONAL NEURAL NETWORKS FOR SEGMENTATION

We start from SegUnet [27] as the node in the recurrent architecture. Similar to U-Net [15], the basis network we use in SASeg consists of an encoder and a decoder path, each with four resolution steps. In the encoder path, each layer has two 3×3 convolutions, each followed by a batch normalization (BN) [40], a rectified linear unit (ReLU) [41] and a 2×2 max pooling with strides of two in each dimension. In the decoder path, each layer has an upsampling layer of 2×2 followed by two convolutions with a kernel size of 2×2 , each followed by a BN and an ReLU as well. The resulting feature maps from each resolution in the encoder are concatenated to those of the same resolution in the decoder. In the last layer, a 1×1 convolution is used to reduce the number of output channels to the number of labels, which is 1 in a binary classification problem. The SegUnet network architecture is shown in the encoder and decoder in Figure 5.5. A recurrent convolutional neural network (CNN) algorithm is introduced to segment the mandible iteratively. At each iteration, the original input slice and previous probability map are fed into the network. This kind of network was robust for segmentation of the mandible from CT images due to learning the 3D spatial information and voxel connectivity of the upper and lower context of the object [26]. Different from the original encoding and decoding structure, we concatenate the feature maps including prior information in the bottom layer, which is obtained from a PSFE.

Figure 5.5 | Details of each unit of SASeg. Each unit of SASeg consists of the PSFE, an encoder and a decoder. The number of channels is indicated in the left of each convolution. The PSFE module is inserted at the last layer of the encoder to capture the mandible prior information.

5.2.4. COMBO LOSS FUNCTION

Motivated by a combo loss used in [42], a combination of Dice and binary cross-entropy (BCE) loss have been applied to train the proposed SASeg. These loss functions are selected due to their potential to contend with imbalanced data [42].

$$\mathcal{L} = \omega_1 \times \mathcal{L}_{BCE} + \omega_2 \times \mathcal{L}_{Dice}, \quad (5.1)$$

where ω_1 and ω_2 are the trade-offs between the BCE term and Dice term contribution in the loss $\mathcal{L}$, which are set as $\omega_1 = 1$ and $\omega_2 = 1$ to obtain the gradient update. $\mathcal{L}_{BCE}$ and

$\mathcal{L}_{Dice}$ are defined as follows:

$$\mathcal{L}_{BCE}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i), \quad (5.2)$$

$$\mathcal{L}_{Dice}(\hat{y}, y) = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \hat{y}_i}. \quad (5.3)$$

Here, y_i and $\hat{y}_i$ represent the ground truth and the predicted probability of pixel i , respectively, and N is the number of pixels.

5.2.5. DATASET

A total of 59 dental CBCT scans that had been heavily affected by metal artifacts were used in this study. All the CBCT scans are obtained on a Vatech PaXZenith3D or Planmeca promax. Each scan consists of 431 to 944 slices with a size of 992×992 to 495×495 pixels. The pixel spacing varies from 0.2 to 0.4 mm, and the slice thickness varies from 0.2 to 0.4 mm. Of these CBCT scans, 38 are used for training, 1 is used for validation, and 20 are used for testing. To train a CNN for bone segmentation in these CBCT scans, gold standard segmentation labels were required. These gold standard labels were created by manually segmenting all CBCT scans using Mimics software (Mimics Medical 23.0, Materialise, Leuven, Belgium) by one experienced technical physician. Gold standard segmentations are actually segmentations augmented with a dentition model. This task took approximately 30–60 min per scan.

To investigate the generalization ability of our model, we also compare our proposed method with several state-of-the-art methods on a public CT dataset, which is the Public Domain Database for Computational Anatomy (PDDCA) version 1.4.1 used for the “Head and Neck Auto Segmentation MICCAI Challenge (2015)” [43]. There are 48 patient CT scans from the Radiation Therapy Oncology Group (RTOG) 0522 study in the PDDCA dataset, in which 40 out of the 48 patients in PDDCA with manual mandible annotations were used in previous studies [43, 44]. According to the protocol setting in the PDDCA challenge description, 40 cases are separated into 25 cases (0522c0001-0522c0328) as a training subset and 15 cases (0522c0555-0522c0878) as a testing subset [43]. Each scan consists of 76 to 360 slices with a size of 512×512 pixels. The pixel spacing varies from 0.76 to 1.27 mm, and the slice thickness varies from 1.25 to 3.0 mm [43].

5.2.6. EVALUATION METRICS

For quantitative analysis of the experimental results, four performance metrics are used, including the Dice coefficient (Dice), the average symmetric surface distance (ASD), and the 95% Hausdorff distance (95HD).

The Dice coefficient is widely used to assess the performance of image segmentation algorithms [45]. It is defined as:

$$\text{Dice} = \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \hat{y}_i}, \quad (5.4)$$

where $y_i, \hat{y}_i$ represents the ground truth and the predicted probability of pixel i , respectively, and N is the number of pixels.

The average symmetric surface distance (ASD) is a measure of computing the average distance between the boundaries of two object regions [30]. It is defined as:

$$\text{ASD}(A, B) = \frac{d(A, B) + d(B, A)}{2}, \quad (5.5)$$

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\|, \quad (5.6)$$

where $\|\cdot\|$ represents the L_2 norm. a and b are corresponding points on the boundary of A and B .

The Hausdorff distance (HD) measures the maximum distance of a point in a set A to the nearest point in the other set B [46]. It is defined as:

$$\text{HD}(A, B) = \max(h(A, B), h(B, A)) \quad (5.7)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (5.8)$$

where $h(A, B)$ means the directed HD. The maximum HD is sensitive to contours. When the image is contaminated by noise or occluded, the original Hausdorff distance is prone to mismatch [43, 47]. Thus, Huttenlocher proposed the concept of partial Hausdorff distance in 1933 [46]. The 95HD metric is similar to the maximum HD. In brief, 95HD selects 95% of the closest points in set B to the point in set A in Equation (6.13) to calculate $h(A, B)$:

$$95\text{HD} = \max(h^{95\%}(A, B), h^{95\%}(B, A)) \quad (5.9)$$

$$h^{95\%}(A, B) = \max_{a \in A} \min_{b \in B^{95\%}} \|a - b\| \quad (5.10)$$

The purpose of using 95HD is to reduce the impact of a small subset of inaccurate prediction outliers on the overall assessment of segmentation quality.

5.2.7. IMPLEMENTATION DETAILS

The proposed CNN model is implemented on an NVIDIA GeForce Tesla V100 or Quadro P6000 by using PyTorch 1.4.0 [48]. In this study, the batch size, epochs, and learning rate are set to 3, 50, and 0.0001, respectively. Furthermore, the Adam optimizer is used to minimize the combo loss. Finally, an early stopping strategy is applied if no improvement in the loss of the validation set for five epochs occurs to avoid overfitting. We apply $n = 39/25$ mandible shapes in the training set to build mean mandible shape model for the CBCT/PDDCA dataset.

5.3. RESULTS

5.3.1. METHOD COMPARISON

We compared the proposed SASeg with recent fashionable methods, such as U-Net [15], SegNet [16], SegUnet [27], and AttUnet [17], which are widely used in medical image

segmentation. In order for fair comparison, the same parameter settings are applied in those methods. In Table 5.1, we list the quantitative results as well as the number of parameters of the corresponding approaches. As shown in Table 5.1, the proposed SASeg gets the best performance and achieves approximately 0.55% to 4%, 0.76 mm to 3 mm and 13.4 mm to 37 mm gains in terms of $Dice$, D_{ASD} and D_{95HD} , respectively. Furthermore, the results in Table 5.1 show that our proposed SASeg is more effective in segmenting the mandible, though there is a small increase of approximately 0.45 million parameters compared with SegUnet.

Table 5.1 | Quantitative comparison of segmentation performance in the CBCT dataset between the proposed SASeg and the state-of-the-art methods.

Methods	$Dice$ (%)	D_{ASD} (mm)	D_{95HD} (mm)	#Params (M)
U-Net	94.79 (± 1.77)	2.0698 (± 0.6137)	32.6401 (± 22.0779)	3.35
SegNet	94.93 (± 1.74)	1.7762 (± 1.5937)	15.9851 (± 26.5286)	2.96
SegUnet	91.27 (± 5.13)	3.1436 (± 3.6049)	26.3569 (± 34.9539)	3.35
AttUnet	93.34 (± 3.79)	3.9705 (± 4.6460)	35.1859 (± 42.3474)	8.73
SASeg	95.35 (± 1.54)	0.9908 (± 0.4128)	2.5723 (± 4.1192)	3.80

We also illustrate the 3D view of three examples taken from the dataset in Figure 5.6. Comparing the ground truth in Figure 5.6(a), the 3D-based segmentation approaches shown in Figure 5.6 failed to segment the mandible detailed structures, such as the coronoids and parts of the mandible body indicated by the red rectangle. More concretely, the U-Net, SegNet, and SegUnet methods have a weaker ability to handle the mandible body that is close to the teeth and are affected by strong metal artifacts, and the segmentation results appear undersegmented, as shown in Figure 5.6(b)-(d). Additionally, the SegUnet misclassifies the maxilla bone as the mandible, while the proposed SASeg can accurately segment the mandible. The AttUnet approach can address the mandible body but cannot correctly recognize the mandible angle, as illustrated in Figure 5.6(e). As shown in Figure 5.6, these above-mentioned fashionable methods have problems with oversegmentation or undersegmentation while addressing the dental CBCT scans, while our method can precisely segment the entire mandible structure even if metal artifacts appear. Moreover, only a low number of voxels is misclassified by SASeg as the mandible. This can be explained by the fact that the proposed SASeg method aggregates pixelwise contextual information, resulting in better segmentation predictions and being able to learn spatial discrepancies between real mandible areas and other structures with high intensity in dental CBCT scans.

5.3.2. ABLATION EXPERIMENTS

Ablation tests are performed to analyze the influence of the components, i.e., the recurrent module and the PSFE module, and the loss function of the proposed SASeg approach. The final comparison of the experiments is also evaluated by calculating $Dice$, D_{ASD} and D_{95HD} between the ground truth and the automated segmentation.

Figure 5.6 | Results with the fashionable segmentation approaches. The ground truth is shown in the column (a). Columns (b)–(f) show the mandible predictions generated by the U-Net, SegNet, SegUnet, AttUnet, and the proposed SASeg, respectively. The red rectangle indicates the zoom-in views of bad predictions.

ABLATION ANALYSIS OF THE PSFE MODULE

To validate the effectiveness of our prior shape extractor module, we conduct a set of ablation experiments on the CBCT dataset. The experimental results are summarized in Table 5.2. From Table 5.2, we can see the use of PSFE and RNN modules brings the most gains in $Dice$, D_{ASD} , and D_{95HD} . When the network utilizes RNN modules, the continuous relationship between current slices is utilized and mined. In addition, the PSFE module is employed for further integrating refined contextual information after RNN modules from cross slices. From Table 5.2, by fully extracting the features of the mean mandible shape prior and integrating the information of different adjacent slices, our model achieves promising results.

We also show the effectiveness of the RNN and PSFE modules based on a mean shape model as a prior input by showing visualized results, as shown in Figure 5.7. We can obviously see that the method without the RNN and PSFE cannot address the mandible body affected by metal artifacts and condyles, as illustrated in Figure 5.7b. The method without PSFE easily causes oversegmentation in the angle of the mandible and is still not sensitive enough in the thick part such as condyles, as shown in Figure 5.7c. The method without the RNN in Figure 5.7d has a slightly stronger ability to handle the mandible body compared with the method without the RNN and PSFE. Incorporation of prior

Table 5.2 | Ablation analysis on different modules of the RNN and PSFE.

RNN	PSFE	<i>Dice</i> (%)	<i>D_{ASD}</i> (mm)	<i>D_{95HD}</i> (mm)
✓	✓	93.78 (± 2.50)	1.5851 (± 1.0680)	17.2517 (± 24.2681)
		92.26 (± 5.66)	1.3133 (± 0.7276)	7.2442 (± 8.9275)
✓	✓	95.09 (± 1.47)	1.2083 (± 0.3354)	4.7629 (± 8.1762)
		95.35 (± 1.54)	0.9908 (± 0.4128)	2.5723 (± 4.1192)

knowledge and spatial information into the mandible segmentation task could provide more accurate and reliable results in terms of the mandible anatomy.

Figure 5.7 | Results with different ablations of our method. The ground truth is shown in column (a). Columns (b)–(e) illustrate the segmentation results derived from the model without RNN and PSFE modules, without the PSFE module, and without the RNN module and the proposed SASeg, respectively. The red rectangle indicates the zoom-in views of bad predictions. (w/o RNN and PSFE: without RNN and PSFE modules, w/o PSFE: without the PSFE module, and w/o RNN: without the RNN module.)

ABLATION ANALYSIS OF THE LOSS FUNCTIONS

In a similar way, we also perform a set of experiments on different loss settings based on the proposed method. The experimental results are listed in Table 5.3. As illustrated in Table 5.3, only BCE loss has a slightly higher performance than the combined loss of BCE and Dice in the *Dice* score, while the combined loss of BCE and Dice has higher performance in *D_{ASD}* and *D_{95HD}*. It is worth noting that the number of predicted background pixels using BCE and Dice is more than the number of only using BCE, and the model using BCE and Dice is able to predict the fine appearance features of the mandible, as shown in Figure 5.8. In brief, the network with BCE and Dice as loss functions easily

generates fewer false positives in mandible segmentation and is more focused on the structure of the object, as illustrated in Table 5.3 and Figure 5.8.

Table 5.3 | Ablation analysis on different loss function settings.

BCE	Dice	<i>Dice</i> (%)	<i>D_{ASD}</i> (mm)	<i>D_{95HD}</i> (mm)
✓		95.45 (±1.39)	1.3934 (±0.6228)	10.6093 (±20.8654)
	✓	83.75 (±15.59)	3.2537 (±3.4075)	16.7939 (±23.4980)
✓	✓	95.35 (±1.54)	0.9908 (±0.4128)	2.5723 (±4.1192)

5

Figure 5.8 | Results with different ablations of loss functions. The ground truth is shown in column (a). Columns (b)–(d) illustrate the segmentation results derived from the model without Dice loss, BCE loss and the proposed SASeg, respectively. The red rectangle indicates the zoom-in views of bad predictions. (w/o Dice: without Dice loss, w/o BCE: without BCE loss.)

5.3.3. RELIABILITY ANALYSIS

To investigate the reliability of mandible segmentation, five randomly selected scans are used to evaluate the intraobserver variability and interobserver variability. For the intraobserver variability study, the second annotations are done six months after the first annotation, and for evaluating interobserver variability, we also employ another technical physician to annotate the mandible.

Intraobserver variability is the variability between the first and the second annotations of

the first observer. The interobserver variability is the average variability between the second observer's annotation and the first observer's two annotations. Intra- and interobserver reliability tests for the mandible were computed using Dice, ASD, and 95HD. The intra- and interobserver agreements for manual segmentation are given in Table 5.4. For intraobserver variability, 98.76%, 0.0690 mm, and 0.6347 mm are found for mean *Dice*, D_{ASD} and D_{95HD} , respectively. For interobserver variability, 91.56%, 0.3555 mm, and 2.0780 mm are obtained for mean *Dice*, D_{ASD} and D_{95HD} , respectively, as shown in Table 5.4. It is worth noting that the *Dice* score for SASeg is higher than the interobserver reliability (95.35% > 91.56%), and the D_{95HD} value is slightly larger than interobserver reliability (2.5723 > 2.0780), demonstrating the reliability and robustness of the automatic segmentation with SASeg.

Table 5.4 | Intra- and interobserver reliabilities and agreement of manual segmentation by the first expert.

	<i>Dice</i> (%)	D_{ASD} (mm)	D_{95HD} (mm)
Intraobserver	98.76 (± 0.96)	0.0690 (± 0.534)	0.6347 (± 0.6176)
Interobserver	91.56 (± 4.45)	0.3555 (± 0.1701)	2.0780 (± 1.1699)
SASeg	95.35 (± 1.54)	0.9908 (± 0.4128)	2.5723 (± 4.1192)

5.3.4. EXPERIMENTS ON THE PDDCA DATASET

We evaluate the performance of the models on the test subset of PDDCA. As shown in Table 5.5, the proposed SASeg method achieves the best performance compared with the state-of-the-art methods, with the highest *Dice* score of 95.29%, the lowest D_{ASD} of 0.1353mm and the second lowest D_{95HD} of 1.3054 mm.

5.4. DISCUSSION

In this work, we developed and validated a novel deep learning-based approach (SASeg) for automated mandible segmentation that utilizes the PSFE module based on a mean mandible shape as a prior and a recurrent network to train the neural network model. In this way, the network makes segmentation predictions that are in agreement with the learned shape model of the underlying anatomy, which is referred to as a shape prior. Most importantly, the proposed approach allows us to perform fully 3D mandible segmentation via slice-by-slice 2D segmentation even in the presence of strong metal artifacts.

We first demonstrate the applicability of the proposed approach SASeg for a dental CBCT dataset of orthodontic treatment that is composed of 59 patient scans. Automated segmentation correlates well with manual segmentations, and the promising segmentation results are shown in Tables 5.1–5.3 and Figures 5.6–5.8. As can be seen in the results, the existing state-of-the-art convolutional neural network (CNN) approaches for segmentation tasks perform poorly when the input data are strongly noisy and blurred by metal artifacts. The experiments described in Figure 5.6 and Table 5.1 show that the proposed segmentation models become more robust against CBCT imaging metal artifacts that

Table 5.5 | Quantitative comparison of the segmentation performance on the PDDCA dataset between the proposed method and the state-of-the-art methods.

Methods	<i>Dice</i> (%)	<i>D_{ASD}</i> (mm)	<i>D_{95HD}</i> (mm)
Multiatlas [49]	91.70 (± 2.34)	-	2.49 (± 0.76)
AAM [50]	92.67 (± 1.00)	-	1.98 (± 0.59)
ASM [51]	88.13 (± 5.55)	-	2.83 (± 1.18)
CNN [52]	89.50 (± 3.60)	-	-
NLGM [53]	93.08 (± 2.36)	-	-
AnatomyNet [21]	92.51 (± 2.00)	-	6.28 (± 2.21)
FCNN [30]	92.07 (± 1.15)	0.51 (± 0.12)	2.01 (± 0.83)
FCNN+SRM [30]	93.60 (± 1.21)	0.37 (± 0.11)	1.5 (± 0.32)
CNN+BD [54]	94.60 (± 0.70)	0.29 (± 0.03)	-
HVR [55]	94.40 (± 1.30)	0.43 (± 0.12)	-
Cascade 3D U-Net [56]	93.00 (± 1.90)	-	1.26 (± 0.50)
Multipplanar [12]	93.28 (± 1.44)	-	1.43 (± 0.56)
Multiview [57]	94.10 (± 0.70)	0.28 (± 0.14)	-
RSegUnet [26]	95.10 (± 1.21)	0.137 (± 0.038)	1.356 (± 0.448)
SASeg	95.29 (± 1.16)	0.135 (± 0.048)	1.305 (± 0.320)

are shown in Figure 5.1. The experimental results show that the classical network can benefit from the learned prior in cases where the images are corrupted by metal artifacts. Compared with the state-of-the-art methods that directly segment a single slice without considering the prior information and the continuity of neighborhood slices, the SASeg network provided better D_{ASD} and D_{95HD} scores. Figure 5.7 and Table 5.2 demonstrate that the mandible shape prior learned by the proposed shape extractor module (PSFE) is useful for mandible segmentation on dental CBCTs in orthodontics. As discussed in the above analysis, the added PSFE module that extracts the prior information from the mean mandible model plays an important role in the segmentation operation in that it can provide more focus on a certain region containing the mandible and avoid the influence of metal artifacts and other bone-structure organs. Figure 5.8 and Table 5.3 show that the combo loss function of Dice and BCE facilitate the network to focus more on the mandible and accordingly generate fewer false positives. Furthermore, we also investigate the intraobserver variability and interobserver variability for manual segmentation. Table 5.4 shows that a higher accuracy is achieved on mandible segmentation using the proposed SASeg when compared to interobserver variability. Furthermore, the PDDCA dataset, consisting of 40 patients scans, is employed for testing the proposed method. The quantitative results shown in Table 5.5 demonstrate that the proposed SASeg method has a good generalization ability in a conventional CT dataset. Overall, SASeg enables the algorithm to not only solve the challenge of mandible segmentation in a dental CBCT dataset with strong metal artifacts but also to provide a good approach to segment the mandible in a CT dataset.

There are a few limitations to the study. First, the CBCT dataset belonged to only a select patient group that required orthodontic treatment. The metal implants in orthog-

nathic surgery should be included in the future. Second, the CBCT and PDDCA datasets are limited and cannot fully represent the general patient population in the clinical setting. Third, a mean mandible model is required before training the model. Fourth, we only use CBCT data from local institutes to train the SASeg model, and the PDDCA dataset in the external validation is a CT dataset and not a CBCT dataset.

To summarize, the proposed SASeg can anatomically and sequentially learn the 3D underlying anatomy through the auxiliary path and the recurrent module, respectively, which is able to enforce that network predictions follow the learned mean shape of the mandible structure and consider the continuity of neighborhood slices for the input scans. More importantly, it is easy to combine with any of the state-of-the-art medical image segmentation networks and can potentially improve their prediction accuracy and robustness with a slight increase in computational resources and memory. Last but not the least, training with the proposed mandible shape prior almost removes the need for postprocessing steps, which provides the capability of simplifying the postprocessing in segmentation tasks, especially in cases where the images are corrupted by metal artifacts or are noisy due to the limitations in their image acquisition. The accurate automated mandible segmentation offers an improved and faster procedure than clinical assessment in 3D VSP, thus providing more practical and faster therapy planning for surgeons or medical technician. The proposed SASeg model can be regarded as an application-specific training target. The presented SASeg framework is not limited to only the mandible segmentation task but can also be extended to other segmentation tasks in which prior knowledge can provide model guidance and robustness. In that regard, future research will focus on the application of SASeg to problems such as other anatomical organ segmentation even on low-quality scans.

5.5. CONCLUSIONS

In this paper, we propose an end-to-end approach (SASeg) for accurate mandible segmentation from CBCT scans that are badly influenced by metal artifacts. First, we adopt a PSFE module that encodes the shape's prior information from a mean mandible model, and then we add it as an auxiliary path to assist the recurrent segmentation network to further learn the structure of the mandible. In this way, the proposed SASeg can automatically aggregate contextual information of the mandible at pixel level and capture the blurry mandible area without any interaction. The quantitative and qualitative results demonstrate that the proposed SASeg can yield better segmentation results compared to the other state-of-the-art algorithms.

REFERENCES

- [1] J. Kraeima, *Three dimensional virtual surgical planning for patient specific osteosynthesis and devices in oral and maxillofacial surgery. A new era.*, Ph.D. thesis, University of Groningen (2019).
- [2] Z. Fourie, J. Damstra, R. H. Schepers, P. O. Gerrits, and Y. Ren, *Segmentation process significantly influences the accuracy of 3d surface models derived from cone beam computed tomography*, *European journal of radiology* **81**, e524 (2012).
- [3] M. Vaitiekūnas, D. Jegelevičius, A. Sakalauskas, and S. Grybauskas, *Automatic method for bone segmentation in cone beam computed tomography data set*, *Applied Sciences* **10**, 236 (2020).
- [4] V. Hirschinger, S. Hanke, U. Hirschfelder, and E. Hofmann, *Artifacts in orthodontic bracket systems in cone-beam computed tomography and multislice computed tomography*, *Journal of Orofacial Orthopedics/Fortschritte der Kieferorthopädie* **76**, 152 (2015).
- [5] J. Wallner, I. Mischak, and J. Egger, *Computed tomography data collection of the complete human mandible and valid clinical ground truth models*, *Scientific data* **6**, 1 (2019).
- [6] L. Wang, K. C. Chen, Y. Gao, F. Shi, S. Liao, G. Li, S. G. Shen, J. Yan, P. K. Lee, B. Chow, *et al.*, *Automated bone segmentation from dental cbct images using patch-based sparse representation and convex optimization*, *Medical physics* **41**, 043503 (2014).
- [7] Y. Fan, R. Beare, H. Matthews, P. Schneider, N. Kilpatrick, J. Clement, P. Claes, A. Penington, and C. Adamson, *Marker-based watershed transform method for fully automatic mandibular segmentation from cbct images*, *Dentomaxillofacial Radiology* **48**, 20180261 (2019).
- [8] S. T. Gollmer and T. M. Buzug, *Fully automatic shape constrained mandible segmentation from cone-beam ct data*, in *IEEE 9th International Symposium on Biomedical Imaging* (IEEE, 2012) pp. 1272–1275.
- [9] O. C. Linares, J. Bianchi, D. Raveli, J. B. Neto, and B. Hamann, *Mandible and skull segmentation in cone beam computed tomography using super-voxels and graph clustering*, *The Visual Computer* **35**, 1461 (2019).
- [10] L. Wang, Y. Gao, F. Shi, G. Li, K.-C. Chen, Z. Tang, J. J. Xia, and D. Shen, *Automated segmentation of dental cbct image with prior-guided sequential random forests*, *Medical physics* **43**, 336 (2016).
- [11] M. S. Nosrati and G. Hamarneh, *Incorporating prior knowledge in medical image segmentation: a survey*, arXiv preprint arXiv:1607.01092 (2016).
- [12] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. H. Borra, M. J. H. Witjes, and P. M. A. van Ooijen, *Automatic segmentation of the mandible from computed tomography scans*

- for 3d virtual surgical planning using the convolutional neural network, *Physics in Medicine & Biology* **64**, 175020 (2019).
- [13] J. Minnema, M. van Eijnatten, A. A. Hendriksen, N. Liberton, D. M. Pelt, K. J. Batenburg, T. Forouzanfar, and J. Wolff, *Segmentation of dental cone-beam ct scans affected by metal artifacts using a mixed-scale dense convolutional neural network*, *Medical physics* **46**, 5027 (2019).
 - [14] R. Pauwels, K. Araki, J. Siewerdsen, and S. S. Thongvigitmanee, *Technical aspects of dental cbct: state of the art*, *Dentomaxillofacial Radiology* **44**, 20140224 (2015).
 - [15] 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* (Springer, 2015) pp. 234–241.
 - [16] V. Badrinarayanan, A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*, *IEEE transactions on pattern analysis and machine intelligence* **39**, 2481 (2017).
 - [17] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, *Attention u-net: learning where to look for the pancreas*, arXiv preprint arXiv:1804.03999 (2018).
 - [18] N. O'Mahony, S. Campbell, A. Carvalho, S. Harapanahalli, G. V. Hernandez, L. Krpalkova, D. Riordan, and J. Walsh, *Deep learning vs. traditional computer vision*, in *Science and Information Conference* (Springer, 2019) pp. 128–144.
 - [19] 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* (Springer, 2016) pp. 424–432.
 - [20] 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)* (IEEE, 2016) pp. 565–571.
 - [21] W. Zhu, Y. Huang, H. Tang, Z. Qian, N. Du, W. Fan, and X. Xie, *Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation*, arXiv preprint arXiv:1808.05238 (2018).
 - [22] C. Wang, T. MacGillivray, G. Macnaught, G. Yang, and D. Newby, *A two-stage 3d unet framework for multi-class segmentation on full resolution image*, arXiv preprint arXiv:1804.04341 (2018).
 - [23] A. Mortazi, J. Burt, and U. Bagci, *Multi-planar deep segmentation networks for cardiac substructures from mri and ct*, in *International Workshop on Statistical Atlases and Computational Models of the Heart* (Springer, 2017) pp. 199–206.

- [24] A. A. Novikov, D. Major, M. Wimmer, D. Lenis, and K. Bühler, *Deep sequential segmentation of organs in volumetric medical scans*, IEEE transactions on medical imaging **38**, 1207 (2018).
- [25] N. Ghavami, Y. Hu, E. Bonmati, R. Rodell, E. Gibson, C. Moore, and D. Barratt, *Integration of spatial information in convolutional neural networks for automatic segmentation of intraoperative transrectal ultrasound images*, Journal of Medical Imaging **6**, 011003 (2018).
- [26] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. v. Ooijen, *Recurrent convolutional neural networks for mandible segmentation from computed tomography*, arXiv preprint arXiv:2003.06486 (2020).
- [27] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, *Automatic traffic sign detection and recognition using segunet and a modified tversky loss function with l1-constraint*, IEEE Transactions on Intelligent Transportation Systems **21**, 1467 (2020).
- [28] F. Chen, H. Yu, R. Hu, and X. Zeng, *Deep learning shape priors for object segmentation*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013) pp. 1870–1877.
- [29] Y. Duan, J. Feng, J. Lu, and J. Zhou, *Context aware 3d fully convolutional networks for coronary artery segmentation*, in *International Workshop on Statistical Atlases and Computational Models of the Heart* (Springer, 2018) pp. 85–93.
- [30] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, *Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks*, Medical Physics **45**, 4558 (2018).
- [31] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, *Active shape models-their training and application*, Computer vision and image understanding **61**, 38 (1995).
- [32] F. Ambellan, H. Lamecker, C. von Tycowicz, and S. Zachow, *Statistical shape models: understanding and mastering variation in anatomy*, in *Biomedical Visualisation* (Springer, 2019) pp. 67–84.
- [33] G. Zheng, S. Li, and G. Szekely, *Statistical shape and deformation analysis: methods, implementation and applications* (Academic Press, 2017).
- [34] W. Krzanowski, *Principles of multivariate analysis*, Vol. 23 (OUP Oxford, 2000).
- [35] Manu, *Nonrigidicp*, <https://www.mathworks.com/matlabcentral/fileexchange/41396-nonrigidicp> MATLAB Central File Exchange, Retrieved February 23, 2021.
- [36] A. Saito, S. Nawano, and A. Shimizu, *Joint optimization of segmentation and shape prior from level-set-based statistical shape model, and its application to the automated segmentation of abdominal organs*, Medical image analysis **28**, 46 (2016).

- [37] D. Grosgeorge, C. Petitjean, J.-N. Dacher, and S. Ruan, *Graph cut segmentation with a statistical shape model in cardiac mri*, Computer Vision and Image Understanding **117**, 1027 (2013).
- [38] F. Chollet, *Xception: Deep learning with depthwise separable convolutions*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) pp. 1251–1258.
- [39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, *Rethinking the inception architecture for computer vision*, in *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016) pp. 2818–2826.
- [40] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, arXiv preprint arXiv:1502.03167 (2015).
- [41] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010) pp. 807–814.
- [42] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, *Combo loss: Handling input and output imbalance in multi-organ segmentation*, Computerized Medical Imaging and Graphics **75**, 24 (2019).
- [43] P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant, T. Albrecht, T. Gass, C. Langguth, M. Lüthi, *et al.*, *Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015*, Medical Physics **44**, 2020 (2017).
- [44] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, and Q. Wang, *Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images*, Medical Physics **45**, 2063 (2018).
- [45] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. Uden, C. I. Sanchez, G. Litjens, F.-E. Leeuw, B. Ginneken, E. Marchiori, and B. Platel, *Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities*, Scientific Reports **7**, 5110 (2017).
- [46] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, *Comparing images using the hausdorff distance under translation*, in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 1992) pp. 654–656.
- [47] A. A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool*, BMC medical imaging **15**, 29 (2015).
- [48] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems* (2019) pp. 8024–8035.

- [49] A. Chen and B. Dawant, *A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images*, in *Presented in Head and Neck Auto-Segmentation Challenge (MICCAI), Munich* (2015).
- [50] R. Mannion-Haworth, M. Bowes, A. Ashman, G. Guillard, A. Brett, and G. Vincent, *Fully automatic segmentation of head and neck organs using active appearance models*, in *Presented in Head and Neck Auto-Segmentation Challenge (MICCAI), Munich* (2015).
- [51] T. Albrecht, T. Gass, C. Langguth, and M. Lüthi, *Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning*, in *Presented in Head and Neck Auto-Segmentation Challenge (MICCAI), Munich* (2015).
- [52] B. Ibragimov and L. Xing, *Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks*, *Medical Physics* **44**, 547 (2017).
- [53] M. Orbes-Arteaga, D. Pea, and G. Dominguez, *Head and neck auto segmentation challenge based on non-local generative models*, in *Presented in Head and Neck Auto-Segmentation Challenge (MICCAI), Munich* (2015).
- [54] O. Kodym, M. Španěl, and A. Herout, *Segmentation of head and neck organs at risk using cnn with batch dice loss*, in *German Conference on Pattern Recognition* (Springer, 2018) pp. 105–114.
- [55] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, *Hierarchical vertex regression-based segmentation of head and neck ct images for radiotherapy planning*, *IEEE Transactions on Image Processing* **27**, 923 (2017).
- [56] Y. Wang, L. Zhao, Z. Song, and M. Wang, *Organ at risk segmentation in head and neck ct images by using a two-stage segmentation framework based on 3d u-net*, *arXiv preprint arXiv:1809.00960* (2018).
- [57] S. Liang, K. Thung, D. Nie, Y. Zhang, and D. Shen, *Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images*, *IEEE Transactions on Medical Imaging* **39**, 2794 (2020).

6

MANDIBLE SEGMENTATION OF DENTAL CBCT SCANS AFFECTED BY METAL ARTIFACTS USING COARSE-TO-FINE LEARNING MODEL

Bingjiang Qiu, Hylke V. D. Wel, Joep Kraeima, Haye H. Glas, Jiapan Guo, Ronald J. H. Borra, Max J. H. Witjes and Peter M. A. van Ooijen.

Qiu, Bingjiang, et al. "Mandible Segmentation of Dental CBCT Scans Affected by Metal Artifacts Using Coarse-to-Fine Learning Model." *Journal of Personalized Medicine* 11.6 (2021): 560.

ABSTRACT

Accurate segmentation of the mandible from cone-beam computed tomography (CBCT) scans is an important step for building a personalized 3D digital mandible model for maxillofacial surgery and orthodontic treatment planning because of the low radiation dose and short scanning duration. CBCT images, however, exhibit lower contrast and higher levels of noise and artifacts due to extremely low radiation in comparison with the conventional computed tomography (CT), which makes automatic mandible segmentation from CBCT data challenging. In this work, we propose a novel coarse-to-fine segmentation framework based on 3D convolutional neural network and recurrent SegUnet for mandible segmentation in CBCT scans. Specifically, the mandible segmentation is decomposed into two stages: localization of the mandible-like region by rough segmentation and further accurate segmentation of the mandible details. The method was evaluated using a dental CBCT dataset. In addition, we evaluated the proposed method and compared it with state-of-the-art methods in two CT datasets. The experiments indicate that the proposed algorithm can provide more accurate and robust segmentation results for different imaging techniques in comparison with the state-of-the-art models with respect to these three datasets.

6

Keywords: Accurate Mandible Segmentation; Oral and Maxillofacial Surgery; 3D virtual surgical planning (3D VSP); Convolutional Neural Network

6.1. INTRODUCTION

Three-dimensional (3D) virtual surgical planning (VSP) technique is commonly used for orthodontic diagnosis, orthognathic diagnosis and surgery planning because it allows for pre- or post-operative simulation of surgical options [1]. Accurate mandible segmentation plays a critical role in the 3D VSP. 3D mandible surface models in 3D VSP are created and superimposed to demonstrate the orthodontic changes both visually and quantitatively (including pre- and post- operation). Cone-beam computed tomography (CBCT) is widely applied in 3D VSP because of its low radiation doses and short scanning duration. However, teeth, tooth fillings, and dental braces in orthodontic treatment and metal implants in orthognathic treatment are high attenuation materials which cause high noise and low contrast in visual impressions of CBCT images. Specifically, weak and false edges in parts of condyles and teeth often appear in the CBCT images. Furthermore, it is difficult to identify the boundaries of mandibles since the dental braces and metal implants negatively affect the image quality in CBCT, as shown in Figure 6.1. Therefore, it is challenging for orthodontic or orthognathic VSP to accurately perform mandibular segmentation in CBCT. Consequently, a large amount of manual work is required to reconstruct 3D mandible models. The patient-specific orthodontic or orthognathic treatment planning is restricted and delayed by this time-consuming procedure.

6

Figure 6.1 | Example of manual annotation of the mandible in a CBCT image with strong metal artifacts.

To reduce the workload of mandible segmentation, a number of (semi-)automatic segmentation methods have been developed in the past decade. Sebastian et al. [2] presented a statistical shape model (SSM) based mandible segmentation approach. They introduced an optimized correspondence to their SSM model. Wang et al. [3] employed a majority voting method and combined it with random forest for mandible segmentation.

Rarasmaya et al. [4] proposed a method based on histogram thresholding and polynomial fitting to segment mandibular cortical bone in CBCT scans. Oscar et al. [5] used super-voxels and graph clustering for mandible segmentation in CBCT images. Fan et al. [6] proposed an automatic approach for segmenting mandibles from CBCT using a marker-based watershed transform. However, some of these traditional techniques require mandible shape prior to initialization, and the performances of these methods are often affected by the noise or metal artifacts. Furthermore, it is difficult to adjust the model parameters according to the overall characteristics of the target contour [7].

With the development of convolutional neural networks (CNN), many approaches have introduced the CNN for mandible segmentation. Ibragimo et al. [8] presented the first attempt of using the deep learning concept of CNN to segment organs at risk (OARs) in head and neck CT scans. The AnatomyNet [9] is built upon the popular 3D Unet architecture using residual blocks in encoding layers and a new loss function combining Dice score and focal loss in the training process. A fully CNN (FCNN) method with a shape representation model for segmentation of organs at risk in CT scans was presented in [10]. Qiu et al. [11] developed a novel technique, RSegUnet, for mandible segmentation in conventional CT scans. This kind of network architecture combines the recurrent unit and the normal segmentation network. RSegUnet has been proven able to accurately segment the mandible parts with weak boundaries, such as condyles and ramus, since the network considers the continuity of neighborhood slices for the scans [11]. The recurrent segmentation network relies on the spatial connections between pixels of the mandible. However, this approach is vulnerable to spatial discontinuities such as metal artifacts in the tooth region and can suffer from oversegmentation once the upper and lower teeth are connected in the scans. Although these methods have led to some performance improvements, they still exhibit some disadvantages such as missing parts of the ramus when metal artifacts are present.

As mentioned in the literature, automatic mandible segmentation is still far from a solved problem, especially for CBCT scans, and post-processing for the mandible often requires significant manual interaction to achieve useful results for clinical practice. In this work, we aim to develop an accurate mandible segmentation algorithm to overcome the inaccurate prediction for 3D VSP in CBCT.

Motivated by the concept of curriculum learning [12], we propose a novel mandible segmentation approach based on a coarse-to-fine learning framework to solve the aforementioned challenges in mandible segmentation of CBCT. Curriculum learning draws upon a learning idea that follows a learning order from easy to difficult tasks [12]. Specifically, a complex task can be solved by dividing it into simple sub-tasks, then starting from the simplest sub-task and progressing to the more difficult sub-tasks. In this study, we propose a hybrid method which consists of a coarse stage and fine stage, in which the coarse stage makes use of 3D CNN for predicting the mandible-like organ and the fine stage utilizes the recurrent segmentation CNN for fine mandible segmentation in CBCT images which are mostly affected by metal artifacts. The proposed approach aims at overcoming the oversegmentation in some parts of the tooth regions and undersegmentation in the weak edges of the ramus and condyles. The coarse segmentation from the coarse stage guides the segmentation of the mandible in the fine stage, and therefore decreases the difficulty of segmenting the mandible from CBCT. Along with this coarse-to-fine

segmentation (named as C2FSeg) task, we design a network by stacking a 3D SegUnet and a recurrent SegUnet. In addition, we extend the proposed segmentation network with a hybrid loss proposed by Taghanaki et al. [13], which has been demonstrated to offer superior performance in many visual applications.

This paper proposes a novel mandible segmentation approach for artifact-affected CBCT/CT with two main contributions:

- First, we apply the concept of curriculum learning to split the mandible segmentation into two sub-tasks. We extract the mandible-like organ using a 3D Unet in the coarse stage and then apply the mandible-like organ into the recurrent segmentation network in the fine stage. In comparison with other CNN approaches, the proposed segmentation approach is robust against metal artifacts.
- Second, the proposed model achieves promising performance on the dataset of CBCT scans of dental braces. Furthermore, the proposed model achieves a promising performance on the conventional CT dataset and Public Domain Database of the Computational Anatomy (PDDCA) dataset.

6.2. METHODOLOGY

From the perspective of the framework, a coarse-to-fine mandible segmentation approach (C2FSeg) is proposed according to curriculum learning [12]. The C2FSeg consists of two main components: coarse stage and fine stage, in which the coarse stage obtains the mandible-like organs, while the fine stage reduces the false positive rate by embedding the overall information of the mandible-like organs and the neighboring information. In the coarse stage, potential mandible candidates are first identified, and then reduction of the false positives (FPs) within the candidates is performed in the fine stage. The coarse model identifies potential mandible candidates, and the fine model reduces the false positives (FPs) within the candidates. The overview of the mandible segmentation framework is given in Fig. 6.2.

6.2.1. CURRICULUM LEARNING IN MANDIBLE SEGMENTATION

Curriculum learning describes a type of learning in which tasks can start with simple tasks before the number of difficult tasks is gradually increased. This learning method is proposed by [12]. It assumes that the curriculum learning can improve the convergence speed of the training process and find a better local minimum[12]. To elaborate upon the proposed C2FSeg approach, we first formulate a segmentation model that can be generalized to both coarse stage and fine stage; we will customize the segmentation model to these two stages in Sections 6.2.2 and 6.2.3, respectively.

Let $X = \{x^1, \dots, x^t, \dots, x^n\}$ be the head and neck scan volume, where X belongs to the CBCT image domain, $X \in \Omega = R^{n \times w \times h}$, where n , w and h represent slice number, width and height, respectively. The corresponding ground truth is $Y = \{y^1, \dots, y^t, \dots, y^n\} \in \{0, 1\}^{n \times w \times h}$, where t denotes the t -th slice of the CT scan. Let $\hat{Y} = \{\hat{y}^1, \dots, \hat{y}^t, \dots, \hat{y}^n\} \in S \in \{0, 1\}^{n \times w \times h}$ denote the predicted segmentation ($\Omega \rightarrow S$). We denote a segmentation task by an operator F , i.e., $\hat{Y} = F(X, \theta)$, where θ indicates model parameters. Specif-

Figure 6.2 | Overview of the proposed method consisting of (a) a 3D CNN for rough mandible segmentation and (b) a 2D recurrent segmentation network for further accurate mandible segmentation. The implementation of stage (a) is as follows: input scan $\rightarrow$ 3D patch extraction $\rightarrow$ 3D CNN (3D SegUnet based) $\rightarrow$ 3D patch segmentation $\rightarrow$ reconstruct the 3D patches into 3D segmentation of the scan. (b) The implementation of stage (b) is as follows: fusion with the coarse mask, the output probability maps and input data $\rightarrow$ RSegCNN (recurrent SegUnet) $\rightarrow$ accurate mandible prediction.

6

ically, in a CNN model with L layers and parameters $\theta = \{w^1, w^2, \dots, w^L; b^1, b^2, \dots, b^L\}$, $\{w^1, w^2, \dots, w^L\}$ is a set of weights and $\{b^1, b^2, \dots, b^L\}$ is a set of biases. According to the concept of curriculum learning, in which a task can be divided into several simple sub-tasks, the task is defined as $F = F_1, \dots, F_s$, where s represents the number of sub-tasks. Therefore, the predicted segmentation $\hat{Y} = F(X, \theta)$ can be rewritten as $\hat{Y} = F_s(F_{s-1}, \dots, F_2(F_1(X, \theta_1), X, \theta_2), \dots, X, \theta_{s-1}), X, \theta_s)$. Although this structure can improve the performance of the task and reduce the difficulty of the task, this model exponentially increases required computing resources, resulting in low efficiency. A small s is sufficient to handle the problem of mandible segmentation. We use $s = 2$ in this study, i.e.,

$$\hat{Y} = F_2(F_1(X, \theta_1), X, \theta_2), \quad (6.1)$$

where F_1 and F_2 denote the models from coarse stage and fine stage, respectively.

6.2.2. COARSE STAGE: MANDIBLE-LIKE ORGAN PREDICTION

One of the obstacles to training 3D deep networks is the problem of "insufficient memory". A common solution is to train a 3D CNN from smaller sub-volumes (3D patches) and test it by sliding window [14]: that is, to perform 3D segmentation on densely and uniformly sampled sub-volumes. In the coarse stage, the input of the coarse stage is cropped from the whole CBCT scan volume X denoted by X^c , $X^c = \text{Crop}(X) \in \text{sub}(\Omega) = R^{n^c \times w^c \times h^c}$, ($n^c \leq n, w^c \leq w, h^c \leq h$), where n^c , w^c and h^c represent the depth, width and height of the cropped volume, respectively. The coarse segmentation model can be formulated as:

$$\hat{Y}_1^c = F_1(X^c, \theta_1). \quad (6.2)$$

The goal of this stage is to efficiently produce the rough mandible segmentations $\hat{Y}_1^c$ from the complex background, which can remove regions that are segmented as non-mandible with high confidence in order to obtain an approximate mandible volume. To be used in the following fine stage, we need to map the sub-volume predictions $\hat{Y}_1^c$ back to exactly the same location given by $X^c = Crop(X)$ after all the positions are traversed. The mathematical definition is $\hat{Y}_1 = UnCrop(\hat{Y}_1^c)$.

As illustrated in Figure 6.3 (a), the coarse stage of our proposed C2Fseg network for capturing the mandible-like candidates is based on the 3D SegUnet, which is the original SegUnet [15] expanded from 2D to 3D. The 3D SegUnet consists of an encoder and a decoder, each of which has four convolutional blocks followed by 3D maxpooling or 3D uppooling layers. The 3D convolutional block includes two convolution operations with a kernel of $3 \times 3 \times 3$, each of which is followed by a batch normalization [16] and a rectified linear unit (ReLU) [17] activation function. The number of filters in the encoder starts at 32 and increases by a factor of 2 after every 3D convolutional block, while it declines by a factor of 2 in the decoder path. The number of feature maps is listed on the left of each convolutional block, and the convolutional layers are represented in Figure 6.3. In addition to the encoder and the decoder, we also use a cross connection to bridge the short-cut connection between the low-level and high-level layers, and we use a cross connection to transfer coordinates from maxpooling to uppooling. In the forward phase, the low-level feature maps extracted from the encoder are directly concatenated to the high-level feature maps, which can improve fine-scaled segmentation [18]. As for the backward phase, the high-level feature maps can be propagated backward through the connections. This approach can prevent the network from enduring gradient vanishing, which will hinder the convergence of the network in the training process [19]. The output of the 3D SegUnet is obtained by applying a convolutional layer with a kernel of $1 \times 1 \times 1$ followed by the sigmoid function.

6.2.3. FINE STAGE: FALSE POSITIVE REDUCTION

In the fine stage as shown in Figure 6.2 (b), a recurrent SegUnet (RSegUnet) is utilized to predict the segmentation map. SegUnet [15] is used as a basic element in the recurrent network. The network setting is the same as 3D SegUnet on the coarse stage and is performed by using a 2D kernel instead of using a 3D kernel, as illustrated in Figure 6.3 (b). RSegUnet adopts the structure of the recurrent neural network, which forms a directed acyclic graph, so that the recurrent connection between adjacent nodes can maintain its connectivity. Furthermore, RSegUnet can further learn the shape of the mandible based on its anatomical connectivity by using the spatial information from neighboring predictions. The coarse network spotted the potential mandible candidates, and we further refine the segmentation results by reducing the FPs of the coarse predictions.

To further utilize the information obtained from the prediction of the previous neighborhood slice, we use RSegUnet [11] to accurately segment the mandible in the fine stage. The sequential design of RSegUnet allows the network to learn anatomical structure continuity in 3D form. In the fine stage, RSegUnet, F_2 , processes each slice sequentially. The input of RSegUnet is sampled from the scan volume X and the rough predictions from $\hat{Y}_1$ of the coarse stage. Here, $\hat{Y}_2 = \{\hat{y}_2^1, \dots, \hat{y}_2^t, \dots, \hat{y}_2^n\} \in S_2 \in \{0, 1\}^{n \times w \times h}$ is the binary seg-

Figure 6.3 | The detailed architecture setting of the proposed method consists of a 3D CNN (a) and a 2D recurrent segmentation network (b).

mentation map generated from the fine stage ($\{\Omega, S_1\} \rightarrow S_2$). In general, $\hat{Y}_2 = F_2(\hat{Y}_1, X, \theta_2)$. In this task, RSegUnet maps a sequence input slice ($x^t, \hat{y}_1^t, \hat{y}_2^{t-1}$) to a sequence output $\hat{y}_2^t$ of the same length, i.e., the output of the unfolded RSegUnet after t steps is represented as:

$$\hat{y}_2^t = F_2(\hat{y}_2^{(t-1)}, x^t, \hat{y}_1^t). \quad (6.3)$$

All in all, the fine stage of the proposed C2Fseg framework is illustrated in Figure 6.3 (b).

6.2.4. LOSS

These two stages are trained separately using the same unified loss function. The loss function of each stage is a combination of Dice and binary cross entropy (BCE) loss. These loss functions are selected due to their potential to deal with imbalanced data.

$$\mathcal{L} = \omega_1 \times \mathcal{L}_{BCE} + \omega_2 \times \mathcal{L}_{Dice}, \quad (6.4)$$

where ω_1 and ω_2 are the hyperparameters which adjust the amounts of BCE and Dice contribution in the loss function $\mathcal{L}$. $\mathcal{L}_{BCE}$ and $\mathcal{L}_{Dice}$ are defined as follows:

$$\mathcal{L}_{BCE}(\hat{y}, y) = -\frac{1}{N} \sum_{i=1}^N y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i), \quad (6.5)$$

$$\mathcal{L}_{Dice}(\hat{y}, y) = 1 - \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \hat{y}_i}. \quad (6.6)$$

Here, y_i and $\hat{y}_i$ represent the ground truth and the predicted probability of pixel i , respectively, and N is the number of pixels.

According to equation (6.5), the term $(1 - y_i) \log(1 - \hat{y}_i)$ penalizes false positives (FPs), as it is 0 when the prediction probability is correct, and $y_i \log(\hat{y}_i)$ penalizes false negatives (FNs) [13]. Therefore, the BCE term is able to control the trade-off between FPs and FNs in the pixelwise segmentation task. In spite of that, the networks with only BCE as loss function are often prone to generate more false positives in the segmentation [20]. [21] have proven that Dice loss yields better performance for one-target segmentation and is able to predict the fine appearance features of the object. Dice loss is based on the Dice coefficient metric, which measures the proportion of overlap between the resulting segmentation and the ground truth. Thus, the combination of loss functions can control the penalization of both FPs and FNs by the BCE term and simultaneously boost the model parameters out of local minima via Dice term.

The training procedure of RSegCNN is the same as that of the traditional CNN, where the trainable weights are updated with the backpropagation through time (BPTT) algorithm [22]. According to equation (6.4), the loss for the t -th step with prediction $\hat{y}$ with respect to ground truth y is:

$$\mathcal{L}^t(\hat{y}^t, y^t) = \omega_1 \times \mathcal{L}_{BCE}(\hat{y}^t, y^t) + \omega_2 \times \mathcal{L}_{Dice}(\hat{y}^t, y^t), \quad (6.7)$$

in which each $\mathcal{L}^t$ is used only at step t .

The gradient $\frac{\partial \mathcal{L}^t}{\partial \hat{y}_j^t}$ of the outputs at step t , for all j, t , is as follows:

$$\begin{aligned} \frac{\partial \mathcal{L}^t}{\partial \hat{y}_j^t} &= \omega_1 \frac{\partial \mathcal{L}_{BCE}(\hat{y}^t, y^t)}{\partial \hat{y}_j^t} + \omega_2 \frac{\partial \mathcal{L}_{Dice}(\hat{y}^t, y^t)}{\partial \hat{y}_j^t} \\ &= -\frac{\omega_1}{N} \left(\frac{y_j^t}{\hat{y}_j^t} - \frac{1 - y_j^t}{1 - \hat{y}_j^t} \right) - \omega_2 \frac{2y_j^t}{\sum_{i=1}^N y_i^t + \hat{y}_i^t} + \\ &\quad \omega_2 \frac{2 \sum_{i=1}^N y_i^t \hat{y}_i^t}{(\sum_{i=1}^N y_i^t + \hat{y}_i^t)^2} \end{aligned} \quad (6.8)$$

6.2.5. EVALUATION METRICS

For quantitative analysis of the experimental results, four performance metrics are used, including Dice coefficient (Dice), average symmetric surface distance (ASD) and 95% Hausdorff distance (95HD).

Dice coefficient is widely used to assess the performance of image segmentation algorithms [23]. It is defined as:

$$\text{Dice} = \frac{2 \sum_{i=1}^N y_i \hat{y}_i}{\sum_{i=1}^N y_i + \hat{y}_i}. \quad (6.9)$$

The average symmetric surface distance (ASD) is a measure for computing the average distance between the boundaries of two object regions [10]. It is defined as:

$$\text{ASD}(A, B) = \frac{d(A, B) + d(B, A)}{2}, \quad (6.10)$$

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\|, \quad (6.11)$$

where $\|\cdot\|$ represents the L_2 norm. a and b are corresponding points on the boundary of A and B .

Hausdorff distance (HD) measures the maximum distance of a point in a set A to the nearest point in the other set B [24]. It is defined as:

$$\text{HD}(A, B) = \max(h(A, B), h(B, A)) \quad (6.12)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\| \quad (6.13)$$

where $h(A, B)$ denotes the directed HD. The maximum HD is sensitive to contours. When the image is contaminated by noise or occluded, the original Hausdorff distance is prone to mismatch [25, 26]. Thus, Huttenlocher proposed the concept of partial Hausdorff distance in 1933 [24]. The 95HD metric is similar to maximum HD. In brief, 95HD selects 95% of the closest points in set B to the point in set A in equation (6.13) to calculate $h(A, B)$:

$$95\text{HD} = \max(h^{95\%}(A, B), h^{95\%}(B, A)) \quad (6.14)$$

$$h^{95\%}(A, B) = \max_{a \in A} \min_{b \in B^{95\%}} \|a - b\| \quad (6.15)$$

The purpose of using 95HD is to reduce the impact of a small subset of inaccurate prediction outliers on the overall assessment of segmentation quality.

6.3. EXPERIMENTS

We evaluate our method on three datasets and compare our performance with state-of-the-art methods.

6.3.1. DATASETS

CBCT DATASET

A total of 59 orthodontic CBCT scans that had been heavily affected by metal artifacts were used in this study. All the CBCT scans were obtained on a Vatech PaXZenith3D (or Planmeca promax). Each scan consists of 431 to 944 slices with size of 992×992 to 495×495 pixels. The pixel spacing varies from 0.2 to 0.4 mm and the slice thickness varies from 0.2 to 0.4 mm. Of these CBCT scans, 38 are used for training, 1 is used for validation and 20 are used for testing. To train a CNN for bone segmentation in these CBCT scans, gold standard segmentation labels were required. These gold standard labels were created by the manual segmentation of all CBCT scans by three experienced medical engineers using Mimics software 20.0 (Materialise, Leuven, Belgium).

CT DATASET

The collection of the patient datasets for medical research purposes was approved by the local medical ethical committee. The dataset contains 109 CT scans reconstructed with a kernel of Br64, I70h(s) or B70s. Each scan consists of 221 to 955 slices with size of 512×512 pixels. We randomly choose 52 cases as training, 8 cases as validation and 49 cases as test. The images have axial dimensions of 512 by 512 with slice numbers varying from 221 to 955. The pixel spacing varies from 0.35 to 0.66 mm, and the slice thickness varies from 0.6 to 0.75 mm. The manual mandible segmentation was performed using Mimics software version 20.0 (Materialise, Leuven, Belgium) by a trained researcher and confirmed by a clinician.

PDDCA DATASET

We also test the proposed strategy on the public dataset PDDCA [25]. This dataset contains 48 patient CT scans from the Radiation Therapy Oncology Group (RTOG) 0522 study, a multi-institutional clinical trial, together with manual segmentation of left and right parotid glands, brainstem, optic chiasm and mandible. Each scan consists of 76 to 360 slices with size of 512×512 pixels. The pixel spacing varies from 0.76 to 1.27 mm, and the slice thickness varies from 1.25 to 3.0 mm. According to the Challenge description, we follow the same training and testing protocol [25]. Forty of the 48 patients in PDDCA with manual mandible annotations are used in this study [25, 27], in which the dataset is split into the training and test subsets, each with 25 (0522c0001-0522c0328) and 15 (0522c0555-0522c0878) cases, respectively [25].

6.3.2. IMPLEMENTATION DETAILS

We implement all the experiments based on the PyTorch [28] platform developed by Facebook. The experiments are trained on a workstation equipped with an Nvidia P6000 or Tesla V100 GPU. For the data pre-processing, we simply truncated the raw intensity values to be within $[-1000, 2000]$, and then normalized each raw CT case to $[0, 1]$ to decrease the data variance caused by physical considerations of the medical device. Note that different CBCT/CT cases have different physical resolutions. As described in Section 6.3.1, we maintain their resolutions in a unified resolution of 512×512 . The weights of the BCE loss term ω_1 and the Dice loss term ω_2 in the loss function are both set to 0.5. We use Adam optimization with a learning rate of $r = 10^{-4}$.

For the coarse stage, we randomly sampled $n^c \times w^c \times h^c = 64 \times 128 \times 128$ sub-volumes from the whole CT scan in the training phase. In this case, a sub-volume can either cover a portion of mandible voxels or be cropped from regions with non-mandible voxels, which acts as a hard negative mining to reduce the false positives. In the testing phase, a sliding window is carried out for the entire CT volume with a coarse step size that has small overlaps within each neighboring sub-volume. Specifically, for a testing volume with a size of $(64, 128, 128)$, we have a total number of sub-volumes to be fed into the network and then combined to obtain the final prediction. For the fine stage, we sequentially sample the slices from the medical scan, the coarse predictions from the coarse stage and apply the mask from the previous unit.

6.3.3. RESULTS

EXPERIMENTS ON THE CBCT DATASET

We compare our methods with numerous standard segmentation architectures such as Unet [18], Segnet [29], SegUnet [15], AttUnet [30] and RSegUnet [11]. Table 6.1 shows the performance comparison as well as the corresponding standard deviation for the mandible segmentation. The average *Dice*, D_{ASD} and D_{95HD} values of the proposed method are 95.31%, 1.2827 mm and 3.1258 mm, respectively. From the Table 6.1, it can be observed that the proposed method outperforms the existing approaches with respect to *Dice*, D_{ASD} and D_{95HD} . These experimental results indicate that our proposed model with the C2Fseg learning method performs significantly better and achieves the highest overall Dice scores compared to other segmentation methods. According to Table 6.1, the proposed method also outperforms most other methods, with the lowest ASD and 95HD scores.

Table 6.1 | Quantitative comparison of segmentation performance for the CBCT dataset between the proposed method and the state-of-the-art methods.

Methods	<i>Dice</i> (%)	D_{ASD} (mm)	D_{95HD} (mm)
Unet[18]	94.79(± 1.77)	2.0698(± 0.6137)	32.6401(± 22.0779)
SegNet[29]	94.93(± 1.74)	1.7762(± 1.5937)	15.9851(± 26.5286)
SegUnet[15]	91.27(± 5.13)	3.1436(± 3.6049)	26.3569(± 34.9539)
AttUnet[30]	93.34(± 3.79)	3.9705(± 4.6460)	35.1859(± 42.3474)
RSegUnet[11]	92.26(± 5.66)	1.3133(± 0.7276)	7.2442(± 8.9275)
Ours	95.31(± 1.11)	1.2827(± 0.2780)	3.1258(± 3.2311)

To better demonstrate the performance of the presented approach, several 2D and 3D view examples of the different algorithms are depicted in Figure 6.4 and Figure 6.5. Figure 6.4 shows some examples of ground truth (GT), Unet [18], Segnet [29], SegUnet [15], AttUnet [30], RSegUnet [11] and the proposed method. As shown in the first two rows of Figure 6.4, the other methods fail to obtain satisfactory results for the main mandible body, while the results from our proposed approach are much better. The third row in Figure 6.4 show that the proposed algorithm achieves better performances when the upper jaw teeth and lower jaw teeth appear within the same slice. The final row in Figure 6.4 illustrates that the proposed method can deal with the ambiguity and blurred boundaries common to CBCT scans of the condyles area.

Figure 6.5 illustrates two 3D view examples (the 1st, and 3rd rows) with the corresponding post-processed examples (the 2nd, and 4th rows) obtained from Unet [18], Segnet [29], SegUnet [15], AttUnet [30], RSegUnet [11] and the proposed method, respectively. The first case shown in Figure 6.5 demonstrate that the proposed method can effectively segment the angle area of the mandible. The second examples shown in Figure 6.5 show that the ramus, the coronoid process area and the teeth are missed by the other methods while the proposed method can tackle the thin parts of the mandible, which are almost always challenging mandible segmentation tasks.

Table 6.1, Figure 6.4 and Figure 6.5 indicate that the proposed approach is quite accurate in segmenting mandibles affected by metal artifacts in CBCT.

Figure 6.4 | 2D examples from CBCT dataset. From left to right: Original CT slice, Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

Figure 6.5 | Visual 3D examples of final segmentations from the CBCT dataset. From left to right: Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

EXPERIMENTS ON THE CT DATASET

We also test the proposed method on a CT dataset. To quantitatively compare the proposed approach with other methods, we compute the Dice scores, ASD and 95HD values of the five methods. Table 6.2 lists the average Dice, ASD and 95HD, as well as the corresponding standard deviation. In general, the average values of these metrics obtained from our proposed method are better than those of the other methods. As shown in Table 6.2, it can be observed that our method yields the highest mean Dice score and the smallest mean errors in 95HD.

Table 6.2 | Quantitative comparison of segmentation performance for the CT dataset between the proposed method and the state-of-the-art methods.

Methods	$Dice(\%)$	$D_{ASD}(mm)$	$D_{95HD}(mm)$
Unet[18]	87.61(± 5.13)	1.8779(± 0.7407)	9.2152(± 17.0825)
SegNet[29]	86.11(± 7.69)	1.6028(± 0.7194)	7.6235(± 15.1696)
SegUnet[15]	83.14(± 12.65)	2.4753(± 1.9507)	15.4372(± 25.1890)
AttUnet[30]	86.11(± 11.63)	1.6033(± 1.4386)	16.7041(± 24.2038)
RSegUnet[11]	86.48(± 7.98)	1.3907(± 0.7566)	7.6591 (± 16.7968)
Ours	88.62(± 4.98)	1.2582(± 0.4102)	4.9668(± 5.0592)

6

Figure 6.6 shows some examples from the original CT slice, the ground truth (GT) and the results obtained from Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method. As shown in the first two rows of Figure 6.6, the other methods fail to obtain satisfactory results for the main mandible body and some parts of teeth, while the results from our proposed approach are much better. The three row in Figure 6.6 shows that the proposed algorithm achieves better performances when the upper jaw teeth and lower jaw teeth appear within the same slice. The final row in Figure 6.6 illustrates that the proposed method can process the condyles area, which is often ambiguous due to blurred boundaries in CT scans.

Moreover, two cases of the automatic segmentation (the 1st, and 3rd rows) and the corresponding post-processed examples (the 2nd, and 4th rows) in the 3D views of Unet, Segnet, SegUnet, AttUnet, RSegUnet and the proposed method are displayed in Figure 6.7. The first case shown in Figure 6.7 indicate that the proposed method can effectively segment the ramus area and body of the mandible. The second examples shown in Figure 6.7 show that the angle area and the teeth in the case are missed by the other methods, while the proposed method can tackle the thin parts of the mandible, which are almost always challenging mandible segmentation tasks. The conventional methods usually lead to erroneous segmentation within the whole mandible, as shown in Figure 6.7. The visual comparison of the automatic segmentation with manual segmentation demonstrates the effectiveness of our method with respect to automatic mandible segmentation. To summarize, Figure 6.6, Figure 6.7 and Table 6.2 indicate that the proposed approach also works well with respect to the other datasets.

EXPERIMENTS ON THE PDDCA DATASET

To further investigate the proposed method, it is preferable to test it with respect to a public dataset and measure the performance of the proposed method on the dataset. Here,

Figure 6.6 | 2D examples from CT dataset. From left to right: Original CT slice, Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

Figure 6.7 | Visual 3D examples of final segmentations from the CT dataset. From left to right: Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

we compare our proposed method with several state-of-the-art methods with respect to the PDDCA dataset. Table 6.3 also lists Dice, ASD and 95HD used in the Challenge paper [25] [10]. According to Table 6.3, the performance of the proposed model surpasses the

majority of the other methods. The proposed method outperforms other methods, with the third-highest mean Dice score, the lowest ASD and the lowest 95HD. For Dice score results, the segmentation result of our method is only slightly worse than RSegUnet [10], while it is better than RSegUnet in terms of ASD and 95HD.

Table 6.3 | Quantitative comparison of segmentation performance for the PDDCA dataset between the proposed method and the state-of-the-art methods.

Methods	$DSC(\%)$	$D_{ASD}(\text{mm})$	$D_{95HD}(\text{mm})$
Multi-atlas [31]	91.7(± 2.34)	-	2.4887(± 0.7610)
AAM [32]	92.67(± 1)	-	1.9767(± 0.5945)
ASM [33]	88.13(± 5.55)	-	2.832(± 1.1772)
CNN [8]	89.5(± 3.6)	-	-
NLGM [34]	93.08(± 2.36)	-	-
AnatomyNet [9]	92.51(± 2)	-	6.28(± 2.21)
FCNN [10]	92.07(± 1.15)	0.51(± 0.12)	2.01(± 0.83)
FCNN+SRM [10]	93.6(± 1.21)	0.371(± 0.11)	1.5(± 0.32)
CNN+BD [35]	94.6(± 0.7)	0.29(± 0.03)	-
HVR [36]	94.4(± 1.3)	0.43(± 0.12)	-
Cascade 3D Unet [37]	93(± 1.9)	-	1.26(± 0.5)
Multi-planar[7]	93.28(± 1.44)	-	1.4333(± 0.5564)
Multi-view[38]	94.1(± 0.7)	0.28(± 0.14)	-
RSegUnet[11]	95.10(± 1.21)	0.1367(± 0.0382)	1.3560(± 0.4487)
SASeg [39]	95.29 (± 1.16)	0.1353 (± 0.0481)	1.3054 (± 0.3195)
Our	94.57(± 1.21)	0.1252(± 0.0275)	1.1813(± 0.4028)

6.4. DISCUSSION

In this paper, we present a novel C2Fseg method for mandible segmentation that utilizes the curriculum learning strategy [12] of separating the task into two simpler sub-tasks (coarse-to-fine). We apply 3D SegUnet to look for mandible-like organs in the coarse stage. In the fine stage, recurrent SegUnet is then employed to finely segment the mandible based on the results from the coarse stage. Quantitative evaluation results shown in Table 6.1-6.3 demonstrate that our proposed approach outperforms the state-of-the-art methods for mandible segmentation. In addition, qualitative visual inspection in Figure 6.4-6.7 illustrates that our automatic segmentation approach performs quite well in comparison with the ground truth. The direct comparison in PDDCA, as shown in Table 6.3, illustrates that the proposed method significantly improves mandible segmentation. Remarkably, we found that this segmentation architecture is very robust for weak and blurry edge segmentation. For instance, the networks can segment both condyles and ramus of the mandible quite well, even under the influence of strong metal artifacts. In addition, the results based on the two datasets (CT and PDDCA) indicate that our proposed approach offers excellent generalization ability, since the images in the two datasets were produced by different imaging technologies.

This method takes advantage of the concept of curriculum learning to simplify the

difficult task to several easy tasks. Therefore, the proposed approach can help to learn the coarse mandible structure in the coarse stage. Furthermore, the proposed approach utilizes a recurrent network in the fine stage that helps with the extraction of spatial information of objects based on mandible-like candidates from the first stage. The experimental results show that the proposed approach is feasible and effective in 3D mandible segmentation and that it can also be applied to other segmentation tasks. This method can support further research on the 3D image segmentation. It can also help overcome the disadvantage of cropping volume for the 3D network because of the memory problem, as well as accomplish 3D segmentation tasks.

In this study, 3D SegUnet is used for searching mandible-like organs, and many other networks which have similar ability for seeking mandible candidates can be used to replace the 3D SegUnet in the coarse stage. Therefore, experimenting with more network architectures in the coarse stage is a potential future research direction. In addition, the number of subjects in the PDDCA dataset [25] is small, which indicates that it may not serve as a good representative dataset in the segmentation tasks. In the future, verifying whether the segmentation accuracy of the proposed method can be further improved by testing on more data is necessary. Moreover, testing whether the proposed method is available for clinical use in 3D VSP and whether it can help improve surgical treatment planning workflows are important aims.

6.5. CONCLUSIONS

6

In this paper, we attempt to address the problem of mandible segmentation using the coarse-to-fine approach. The coarse stage of our algorithm attempts to obtain probable mandible-like candidates in 3D volume. The fine stage of our approach attempts to reduce the false positives detected from the estimated mandible-like organs. First, we employ a patch-based mandible detector, wherein scans are divided into overlapping patches which are classified as mandible or non-mandible. Second, we utilize the recurrent CNN to finely segment the mandible following the coarse stage. The proposed algorithm is evaluated on three datasets: an orthodontic CBCT dataset polluted by metal artifacts, a CT dataset and a PDDCA dataset. Experimental results show that our method can achieve high accuracy for mandible segmentation in comparison with ground truth. The method overcomes the problem of weak mandible boundaries caused by low radiation and strong metal artifacts.

REFERENCES

- [1] J. Kraeima, *Three dimensional virtual surgical planning for patient specific osteosynthesis and devices in oral and maxillofacial surgery. A new era.*, Ph.D. thesis, University of Groningen (2019).
- [2] S. T. Gollmer and T. M. Buzug, *Fully automatic shape constrained mandible segmentation from cone-beam ct data*, in *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on* (IEEE, 2012) pp. 1272–1275.
- [3] L. Wang, Y. Gao, F. Shi, G. Li, K.-C. Chen, Z. Tang, J. J. Xia, and D. Shen, *Automated segmentation of dental cbct image with prior-guided sequential random forests*, *Medical physics* **43**, 336 (2016).
- [4] R. Indraswari, A. Z. Arifin, N. Suciati, E. R. Astuti, and T. Kurita, *Automatic segmentation of mandibular cortical bone on cone-beam ct images based on histogram thresholding and polynomial fitting*, *Int. J. Intell. Eng. Syst* **12**, 130 (2019).
- [5] O. C. Linares, J. Bianchi, D. Raveli, J. B. Neto, and B. Hamann, *Mandible and skull segmentation in cone beam computed tomography using super-voxels and graph clustering*, *The Visual Computer* **35**, 1461 (2019).
- [6] Y. Fan, R. Beare, H. Matthews, P. Schneider, N. Kilpatrick, J. Clement, P. Claes, A. Penington, and C. Adamson, *Marker-based watershed transform method for fully automatic mandibular segmentation from cbct images*, *Dentomaxillofacial Radiology* **48**, 20180261 (2019).
- [7] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. van Ooijen, *Automatic segmentation of the mandible from computed tomography scans for 3d virtual surgical planning using the convolutional neural network*, *Physics in Medicine & Biology* **64**, 175020 (2019).
- [8] B. Ibragimov and L. Xing, *Segmentation of organs-at-risks in head and neck ct images using convolutional neural networks*, *Medical physics* **44**, 547 (2017).
- [9] W. Zhu, Y. Huang, H. Tang, Z. Qian, N. Du, W. Fan, and X. Xie, *Anatomynet: Deep 3d squeeze-and-excitation u-nets for fast and fully automated whole-volume anatomical segmentation*, *arXiv preprint arXiv:1808.05238* (2018).
- [10] N. Tong, S. Gou, S. Yang, D. Ruan, and K. Sheng, *Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks*, *Medical physics* **45**, 4558 (2018).
- [11] B. Qiu, J. Guo, J. Kraeima, H. H. Glas, R. J. Borra, M. J. Witjes, and P. M. v. Ooijen, *Recurrent convolutional neural networks for mandible segmentation from computed tomography*, *arXiv preprint arXiv:2003.06486* (2020).

- [12] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, *Curriculum learning*, in *Proceedings of the 26th annual international conference on machine learning* (2009) pp. 41–48.
- [13] S. A. Taghanaki, Y. Zheng, S. K. Zhou, B. Georgescu, P. Sharma, D. Xu, D. Comaniciu, and G. Hamarneh, *Combo loss: Handling input and output imbalance in multi-organ segmentation*, *Computerized Medical Imaging and Graphics* **75**, 24 (2019).
- [14] F. Milletari, N. Navab, and S.-A. Ahmadi, *V-net: Fully convolutional neural networks for volumetric medical image segmentation*, in *3D Vision (3DV), 2016 Fourth International Conference on* (IEEE, 2016) pp. 565–571.
- [15] U. Kamal, T. I. Tonmoy, S. Das, and M. K. Hasan, *Automatic traffic sign detection and recognition using segunet and a modified tversky loss function with l1-constraint*, *IEEE Transactions on Intelligent Transportation Systems* **21**, 1467 (2019).
- [16] S. Ioffe and C. Szegedy, *Batch normalization: Accelerating deep network training by reducing internal covariate shift*, arXiv preprint arXiv:1502.03167 (2015).
- [17] V. Nair and G. E. Hinton, *Rectified linear units improve restricted boltzmann machines*, in *Proceedings of the 27th international conference on machine learning (ICML-10)* (2010) pp. 807–814.
- [18] 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* (Springer, 2015) pp. 234–241.
- [19] X. Glorot and Y. Bengio, *Understanding the difficulty of training deep feedforward neural networks*, in *Proceedings of the thirteenth international conference on artificial intelligence and statistics (JMLR Workshop and Conference Proceedings, 2010)* pp. 249–256.
- [20] S. M. Abulnaga and J. Rubin, *Ischemic stroke lesion segmentation in ct perfusion scans using pyramid pooling and focal loss*, in *International MICCAI Brainlesion Workshop* (Springer, 2018) pp. 352–363.
- [21] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, *Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations*, in *Deep learning in medical image analysis and multimodal learning for clinical decision support* (Springer, 2017) pp. 240–248.
- [22] Y. LeCun, Y. Bengio, and G. Hinton, *Deep learning*, *nature* **521**, 436 (2015).
- [23] M. Ghafoorian, N. Karssemeijer, T. Heskes, I. W. Uden, C. I. Sanchez, G. Litjens, F.-E. Leeuw, B. Ginneken, E. Marchiori, and B. Platel, *Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities*, *Scientific Reports* **7**, 5110 (2017).

- [24] D. P. Huttenlocher, W. J. Rucklidge, and G. A. Klanderman, *Comparing images using the hausdorff distance under translation*, in *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE, 1992) pp. 654–656.
- [25] P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant, T. Albrecht, T. Gass, C. Langguth, M. Lüthi, *et al.*, *Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015*, *Medical physics* **44**, 2020 (2017).
- [26] A. A. Taha and A. Hanbury, *Metrics for evaluating 3d medical image segmentation: analysis, selection, and tool*, *BMC medical imaging* **15**, 29 (2015).
- [27] X. Ren, L. Xiang, D. Nie, Y. Shao, H. Zhang, D. Shen, and Q. Wang, *Interleaved 3d-cnn s for joint segmentation of small-volume structures in head and neck ct images*, *Medical physics* **45**, 2063 (2018).
- [28] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimeshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems* (2019) pp. 8024–8035.
- [29] V. Badrinarayanan, A. Kendall, and R. Cipolla, *Segnet: A deep convolutional encoder-decoder architecture for image segmentation*, *IEEE transactions on pattern analysis and machine intelligence* **39**, 2481 (2017).
- [30] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, *et al.*, *Attention u-net: learning where to look for the pancreas*, arXiv preprint arXiv:1804.03999 (2018).
- [31] A. Chen and B. Dawant, *A multi-atlas approach for the automatic segmentation of multiple structures in head and neck ct images*, *Head and Neck Auto-Segmentation Challenge (MICCAI)*, Munich (2015).
- [32] R. Mannion-Haworth, M. Bowes, A. Ashman, G. Guillard, A. Brett, and G. Vincent, *Fully automatic segmentation of head and neck organs using active appearance models*, *MIDAS J.* (2015).
- [33] T. Albrecht, T. Gass, C. Langguth, and M. Lüthi, *Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning*, *Head and Neck Auto-Segmentation Challenge (MICCAI)*, Munich (2015).
- [34] M. Orbes-Arteaga, D. Pea, and G. Dominguez, *Head and neck auto segmentation challenge based on non-local generative models*, *MIDAS J.* (2015).
- [35] O. Kodym, M. Španěl, and A. Herout, *Segmentation of head and neck organs at risk using cnn with batch dice loss*, arXiv preprint arXiv:1812.02427 (2018).

- [36] Z. Wang, L. Wei, L. Wang, Y. Gao, W. Chen, and D. Shen, *Hierarchical vertex regression-based segmentation of head and neck ct images for radiotherapy planning*, IEEE Transactions on Image Processing **27**, 923 (2017).
- [37] Y. Wang, L. Zhao, Z. Song, and M. Wang, *Organ at risk segmentation in head and neck ct images by using a two-stage segmentation framework based on 3d u-net*, arXiv preprint arXiv:1809.00960 (2018).
- [38] S. Liang, K. Thung, D. Nie, Y. Zhang, and D. Shen, *Multi-view spatial aggregation framework for joint localization and segmentation of organs at risk in head and neck CT images*, IEEE Transactions on Medical Imaging **39**, 2794 (2020).
- [39] B. Qiu, H. van der Wel, J. Kraeima, H. H. Glas, J. Guo, R. J. Borra, M. J. H. Witjes, and P. van Ooijen, *Robust and accurate mandible segmentation on dental cbct scans affected by metal artifacts using a prior shape model*, Journal of Personalized Medicine **11**, 364 (2021).

IV

DISCUSSION AND CONCLUSION

7

DISCUSSION AND CONCLUSION

7.1. GENERAL DISCUSSION

Three-dimensional virtual surgical planning (3D VSP) has become a critical component of oral and maxillofacial surgery (OMFS) [1]. Over the years, 3D VSP has been shown to improve the accuracy and speed of OMFS procedures [1]. This thesis aims to improve the workflow of 3D VSP and provides four deep learning-based methods of mandible segmentation in CT and CBCT to reduce the time and inter-observer variability of mandible segmentation during OMFS.

In the first part, we introduced 3D VSP during OMFS and reviewed and summarized the available automatic segmentation methods from the collected papers. In the second part, we proposed two automatic mandible segmentation methods using 2D networks to determine the spatial volume information from head and neck CT scan. In the third part, we developed two automatic mandible segmentation approaches against strong artifacts in dental CBCT.

Chapter 2 summarized the available automatic mandible segmentation techniques and provided a vivid description of the scientific advancements to clinicians and researchers in this field to help develop novel automatic methods for clinical applications. The collected publications were divided into several categories depending on the type of method used. An analysis of the resulting publications was then conducted from the perspective of image modality, image database, methodology, and performance metrics. Chapter 2 showed that mandible segmentation is still a challenging task due to strong artifacts and low contrast in the condyles, although some work has achieved promising results. With the development of machine learning and deep learning, mandible segmentation performance has rapidly increased. However, these methods require more data and greater computational resources. Additionally, there is still room of improvement in cases of MRI and (CB)CT corrupted by artifacts.

Chapter 3 introduced an end-to-end approach which extracts the spatial information from the mandible using three orthogonal planes of CT scans using 2D networks. Furthermore, we used a 2.5D training strategy in the network architecture. The use of these methods can overcome the memory problem shortcoming when using a 3D network. We achieved an average Dice score of 0.881 and an average surface error of 0.5791 mm on 109 CT scans and an average Dice coefficient of 0.9328 and 95HD of 1.4333 mm on the PDDCA dataset. The quantitative results demonstrated the effectiveness of the proposed approach in mandible segmentation and its potential use in 3D VSP.

Although the proposed method in the Chapter 3 achieved promising results, the method combined the results from three orthogonal directions in three sub-networks, which required training the three sub-networks separately and is still not the most effective method. To deal with the separate training strategy, Chapter 4 provided an end-to-end approach that combined normal 2D networks with the recurrent connections, namely RCNNSeg. The proposed RCNNSeg allows the networks to learn information between adjacent slices in the CT volume. Unlike the classic 3D networks that need to crop the 3D volume into smaller 3D patches, our approach can perform 3D segmentation on sequential data of any varied lengths and does not require large computational resources. We achieved best results in 109 CT scans comparing with normal 2D, 2.5D and 3D networks in terms of the Dice, ASD and 95HD metrics. Furthermore, we also obtained best

results in the PDDCA dataset in terms of the Dice, ASD and 95HD metrics. The qualitative results showed the proposed RCNNSeg is effective for the segmentation of weak and blurry boundaries (for instance, the mandible coronoids).

However, the RCNNSeg approach in Chapter 4 failed to delineate the condyle of the mandible from CBCT images when the images were severely distorted by metal artifacts. Chapter 5 focused on solving the mandible segmentation with respect to dental CBCT damaged by metal artifacts. We developed SASeg for automatic mandible segmentation that applies prior shape feature extractor (PSFE) module based on a mean mandible shape as a prior and recurrent network, in which the mean mandible is generated by a statistical shape model (SSM). The proposed SASeg method makes the segmentation predictions under the mean mandible shape. The proposed SASeg method performs fully 3D mandible segmentation using a slice-by-slice manner in a 2D network, even in the presence of strong metal artifacts. We performed the SASeg method on 59 dental CBCT scans and achieved best performance compared with the classical networks. Furthermore, the ablation experiments demonstrate the PAFE module and recurrent connection strategy are effective in mandible segmentation. The results demonstrated that the proposed SASeg method is more robust against CBCT imaging metal artifacts. We also performed the SASeg method in PDDCA, and the quantitative experiments showed that the proposed method has a good generalization ability in conventional CT. Additionally, SASeg provides potential for simplifying the post-processing during the segmentation task even in dental CBCT that is corrupted by metal artifacts.

Although SASeg from Chapter 5 had its first successful trial in the artifact-corrupted CBCT, the mean mandible model from SSM is required as a shape prior in SASeg. An approach without prior information is worthy to study. Chapter 6 proposed a novel coarse-to-fine segmentation approach (C2FSeg) for mandible. This approach adopted the idea from curriculum learning. In a way, the mandible segmentation task was divided into two simple sub tasks (a coarse stage and a fine stage). In the coarse stage, a 3D SegUnet was utilized to find mandible-like organs. Based on that, we applied recurrent SegUnet for fine segmentation of the mandible. We performed the C2FSeg in three datasets which contained an orthodontic patient CBCT dataset, a CT dataset and a PDDCA dataset. Quantitative and qualitative results showed that C2FSeg achieved promising performance compared with the state-of-the-art methods, and is effective in 3D mandible segmentation.

7.2. FUTURE PERSPECTIVES

The proposed work can be extended in the following directions for future research. First, the CT, CBCT and PDDCA datasets are limited and cannot fully represent the general patient population in the clinical setting. In future work, we aim to investigate the mandible segmentation algorithms in larger datasets, or even extend the segmentation algorithms to MRI images. Second, we only studied mandible segmentation in medical images. However, 3D VSP involves multi-organ segmentation. Therefore, exploring multi-organ segmentation is another researchable direction. Third, we did not clinically analyze the segmentation methods. More clinical references and feasibility analyses for the proposed methods should be performed in the future.

In addition to the above methods that need to be further expanded, there are two research directions worthy of discussion.

7.2.1. MRI-ONLY 3D VSP

Typically, 3D VSP of OMFS commonly requires combination CT/CBCT and MRI image modalities. Since MRI imaging is not sensitive for bone tissue, CT images are always acquired and segmented for bone-structured organs in 3D VSP of OMFS treatment. In the consensus guidelines for CT-based manual segmentation of mandibles, 3D VSP strongly recommends to the use of MRI images for segmenting tumors and their surrounding tissues [1]. Furthermore, 3D VSP needs to employ MRI-to-CT/CBCT registration technique to merge the segmentation of tumor, bone, and their surrounding tissues. In this way, 3D VSP relies on an accurate intra-patient multi-modal image registration. Geometric transformations of the corresponding MR-to-CT/CBCT image registration is a tedious procedure. Moreover, this multimodality image fusion has an error of 1–2 mm [5]. MRI-only 3D VSP does not require the registration procedure. Moreover, considering the ionizing radiation of CT/CBCT scanning, MRI-only and synthetic CT based on MRI are increasingly prioritized in 3D VSP. In this way, elimination of the radiation risk to patients in 3D VSP has become possible and explorable. There are two potential solutions: synthetic CT based on MRI [2–4] and fully MRI segmentation [5–7]. However, it is hard to distinguish air and bone tissue from MRI [6, 7]. Another good research direction is to study MRI sequences that are sensitive to bone or hard tissues. For example, the study from Hoving et al. [5], which tried to apply black bone sequence to 3D VSP, makes MRI-only 3D VSP feasible. The first MRI-to-CT synthesis trial was performed using a deep learning approach. Figure 7.1 shows an example of mandibles segmented from synthetic CT and real CT, along with a comparison. In order to apply this idea to the daily practice of 3D VSP, further research is needed, such as study more accessible MRI sequences to hard tissues, develop accurate and reproducible synthetic CT generation methods, evaluate accuracy error in clinics.

7

Figure 7.1 | A visual example of mandible segmentation based on synthetic CT from MRI (a), real CT (b) and its comparison (c).

7.2.2. METRICS

With the promotion of the MICCAI challenge protocol [8], Dice similarity coefficient (Dice), average symmetric surface distance (ASD), Hausdorff distance (HD) and 95-percentile Hausdorff distance (95HD) metrics are used to evaluate the segmentation tasks. These metrics are only used to evaluate the whole target. However, when the ramus is really thin, the evaluation metrics like Dice or ASD are not sensitive for these thin structures. As shown the example in Figure 7.2, the evaluation of model (a) is better than model (b). However, the mandibular angle region is important in most cases; consequently, missing parts of the angle during the segmentation task is unacceptable in the clinic. Model (b) is more useful for 3D VSP in clinical practices, as shown in Figure 7.2. Therefore, to obtain a more accurate understanding of the predicted results, developing an appropriate metric for evaluating mandibles should be further explored to address this problem.

Figure 7.2 | A visual case of mandible with its performance metrics.

7.3. CONCLUSION

This thesis presents four algorithms for mandible segmentation from CT and CBCT and contributes to some novel ideas for the development of automatic mandible segmentation for 3D VSP, which makes it possible to simplify automated mandible segmentation. Furthermore, the automatic methods can markedly shorten the time in preparing the 3D mandible model in 3D VSP of OMFS. For example, manual mandible delineation takes about 40 minutes for one CT scan, while the RCNNSeg method takes about 1.5 minutes. Surgeons or medical technicians can utilize these deep learning-based segmentation methods to decrease the preparing time and the inter-observer variety when designing 3D VSP of OMFS, thus making the 3D VSP procedure more practical and faster.

REFERENCES

- [1] J. Kraeima, *Three dimensional virtual surgical planning for patient specific osteosynthesis and devices in oral and maxillofacial surgery. A new era.*, Ph.D. thesis, University of Groningen (2019).
- [2] N. Burgos, M. J. Cardoso, F. Guerreiro, C. Veiga, M. Modat, J. McClelland, A.-C. Knopf, S. Punwani, D. Atkinson, S. R. Arridge, *et al.*, *Robust ct synthesis for radiotherapy planning: application to the head and neck region*, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015) pp. 476–484.
- [3] A. Bahrami, A. Karimian, E. Fatemizadeh, H. Arabi, and H. Zaidi, *A new deep convolutional neural network design with efficient learning capability: Application to ct image synthesis from mri*, *Medical Physics* **47**, 5158 (2020).
- [4] J. Lee, A. Carass, A. Jog, C. Zhao, and J. L. Prince, *Multi-atlas-based ct synthesis from conventional mri with patch-based refinement for mri-based radiotherapy planning*, in *Medical Imaging 2017: Image Processing*, Vol. 10133 (International Society for Optics and Photonics, 2017) p. 101331I.
- [5] A. M. Hoving, J. Kraeima, R. H. Schepers, H. Dijkstra, J. H. Potze, B. Dorgelo, and M. J. Witjes, *Optimisation of three-dimensional lower jaw resection margin planning using a novel black bone magnetic resonance imaging protocol*, *PloS one* **13**, e0196059 (2018).
- [6] Y. Lei, J. Zhou, X. Dong, T. Wang, H. Mao, M. McDonald, W. J. Curran, T. Liu, and X. Yang, *Multi-organ segmentation in head and neck mri using u-faster-rcnn*, in *Medical Imaging 2020: Image Processing*, Vol. 11313 (International Society for Optics and Photonics, 2020) p. 113133A.
- [7] Y. Lei, J. Harms, X. Dong, T. Wang, X. Tang, S. Y. David, J. J. Beitler, W. J. Curran, T. Liu, and X. Yang, *Organ-at-risk (oar) segmentation in head and neck ct using u-rcnn*, in *Medical Imaging 2020: Computer-Aided Diagnosis*, Vol. 11314 (International Society for Optics and Photonics, 2020) p. 113144A.
- [8] P. F. Raudaschl, P. Zaffino, G. C. Sharp, M. F. Spadea, A. Chen, B. M. Dawant, T. Albrecht, T. Gass, C. Langguth, M. Lüthi, *et al.*, *Evaluation of segmentation methods on head and neck ct: Auto-segmentation challenge 2015*, *Medical physics* **44**, 2020 (2017).

SUMMARY

Three-dimensional virtual surgical planning (3D VSP) has been shown to improve the accuracy and speed of oral and maxillofacial surgery (OMFS) procedures. The aim of this thesis is to improve the workflow of 3D VSP and provides four deep learning-based methods of mandible segmentation in CT and CBCT to reduce the time and inter-observer variability of mandible segmentation during OMFS. In **Part I**, we introduced 3D VSP during OMFS in Chapter 1. Chapter 2 summarized the available automatic mandible segmentation techniques and provided a vivid description of the scientific advancements to clinicians and researchers in this field to help develop novel automatic methods for clinical applications. In **Part II**, we proposed two automatic mandible segmentation methods using 2D networks to extract the spatial volume information from head and neck CT scan. Chapter 3 introduced a 2.5D training strategy in three orthogonal planes of CT scans for mandible segmentation. In order to modify the separate training strategy in the proposed method in Chapter 3, Chapter 4 provided an end-to-end approach that combined normal 2D networks with the recurrent connections, namely RCNNSeg. In **Part III**, we developed two automatic mandible segmentation approaches against strong artifacts in dental CBCT. In Chapter 5, we developed SASeg for automatic mandible segmentation that applies prior shape feature extractor (PSFE) module based on a mean mandible shape as a prior and recurrent network, in which the mean mandible is generated by a statistical shape model (SSM). Chapter 6 proposed a novel coarse-to-fine segmentation approach for mandible using curriculum learning. In **Part IV**, the main findings of studies in this thesis and future perspectives are summarized and discussed.

SAMENVATTING

Van driedimensionale virtuele chirurgische planning (3D VCP) is aangetoond dat het de nauwkeurigheid en snelheid van orale en mondziekten, kaak- en aangezichtschirurgie (MKA) verbetert. Het doel van dit proefschrift is om de workflow van 3D VCP te verbeteren. We bieden vier op deep learning gebaseerde methoden van onderkaaksegmentatie in CT en CBCT aan om de benodigde tijd en interobserver variabiliteit van onderkaaksegmentatie tijdens MKA te verminderen. In **Deel I** introduceerden we 3D VCP tijdens MKA in Hoofdstuk 1. Hoofdstuk 2 vatte de beschikbare automatische onderkaaksegmentatietechnieken samen en gaf een levendige beschrijving van de wetenschappelijke vooruitgang op dit gebied aan klinici en onderzoekers om nieuwe automatische methoden voor klinische toepassingen te helpen ontwikkelen. In **Deel II** hebben we twee automatische onderkaaksegmentatiemethoden voorgesteld die gebruik maken van 2D-netwerken om de ruimtelijke volume-informatie te extraheren uit de CT-scan van het hoofd en de nek. Hoofdstuk 3 introduceerde een 2.5D trainingsstrategie in drie orthogonale vlakken van CT-scans voor onderkaaksegmentatie. Om de afzonderlijke trainingsstrategie in de voorgestelde methode in Hoofdstuk 3 aan te passen, verschaftte Hoofdstuk 4 een end-to-end benadering die normale 2D-netwerken combineerde met terugkerende verbindingen, zogenaamde RCNNSeg. In **Deel III** hebben we twee automatische onderkaaksegmentatiebenaderingen ontwikkeld tegen sterke artefacten in tandheelkundige CBCT. In Hoofdstuk 5 hebben we SASeg ontwikkeld voor automatische segmentatie van de onderkaak, waarbij een voorafgaande vorm feature extractor module wordt toegepast op basis van een gemiddelde onderkaakvorm als een voorafgaand en terugkerend netwerk, waarbij de gemiddelde onderkaak wordt gegenereerd door een statistisch vormmodel. Hoofdstuk 6 stelde een nieuwe grof-naar-fijn segmentatie benadering voor onderkaak voor met behulp van curriculum learning. In **Deel IV** worden de belangrijkste bevindingen van studies in dit proefschrift en toekomstperspectieven samengevat en besproken.

ACKNOWLEDGEMENTS

Completing a PhD is quite an unforgettable journey. It has been quite a different experience living and working here in Groningen, Netherlands. At the end of my PhD journey, I would like to express my deepest gratitude and acknowledge those who provide me make this journey happen and come to this nice end.

I would like to express my gratitude to Peter. I felt lucky to do this unique opportunity in the MII team. Your vision for the future of this research direction gave me a lot of inspiration during my PhD. Your critical feedback has inspired me to think deeper into my research and often has resulted in creative solutions. Thank you very much for everything. I greatly value the trust you placed in me to do my work in your group, the exciting research directions you offered me, and the numerous opportunities you have encouraged me.

I would also like to thank Max for providing the unique opportunity in 3D Lab, which allows me to research machine learning in 3D virtual surgical planning. Thank you for the inspiration from your work related to 3D technology in head and neck oncology. You are a kind and patient person, and I appreciate your explanations about oral and maxillofacial surgery. Your passion for science and your hard work is a great inspiration for me. Thank you for all the constructive discussions. Thank you very much for your guidance during my PhD.

I would like to thank Ronald for your critical comments. I enjoyed the discussion in our regular project meetings.

I would like to express my thanks to Joep. I always enjoyed discussing with you; thanks for sharing your views on so many things over the years! Thank you for all the helps you provided during these years.

I would also like to express my appreciation to Jiapan. I very much appreciate your valuable advice on my research. Our discussions about research and life were always very valuable to me. Thank you for all the kind help and guidance in my daily PhD life.

Furthermore, I would like to thank the many lovely colleagues in MII group: Jiapan, Ricardo, Sunyi, Daan, Yeshu, Xueping, Ludo, Alessia, Qiao, Erfan, Jingxuan, Nikos. They made the MII group into a pleasant research environment. I have learned a lot from our bi-weekly meeting. Special thanks to Daan, Ricardo, Sunyi, Yeshu, Xueping; I enjoyed the parties and game nights that we had.

Many thanks to all my lovely colleagues at 3D Lab: Joep, Ricardo, Bram, Haye, Hylke, Anne, Nick, Peter, Lisanne, Jelbrich. Big thanks to Joep, Haye and Hylke for their expertise and assistance in collecting data for my research.

I would also like to express my appreciation to my colleagues at G2/Bloemsingel: Stella, Runlei, Daiwei, Congying, Marleen, Rik, Ivan, Marly, Mirjam, Gert Jan, Grigory, Moniek, Daan Ties, Randy, Jiali, Yihui, Xiaonan, Qiong, Yaping, Lingyun, Jiehui, Herbert. We had a wonderful time together, and I am delighted to have worked with you. Special thanks to Stella. Thank you for helping me to solve many problems all the time. And

thank you for creating an excellent atmosphere for researching in G2/Bloemsingel. I also really enjoyed our ECR trip in Vienna.

Another big thank you to our collaborators, Antje, Adrian and Bas. Thank you for their kind assistance in collecting data for my research trial.

I would like to thank Prof. Oudkerk and Jiehui. Thank you for giving me a chance to work in i-DNA at the end of my PhD. I hope that i-DNA will continue to thrive in the future.

I would like to thank my friends from the dual summit academic forum: An, Lianmin, Mengfei, Shixian, Yuntao. Thanks for exchanging ideas in our parties. It is a shame that we may not be able to drink as usual in the future.

I would like to express my thanks to Bohuan, Cancan, Ji, Jing, Mian, Qianliang, Sevrin, Shandong, Xinyu, Xiaodong, Yi, Yanni, Yichen, Zhenhua and all my friends. I felt happy to meet you and thank you for making my daily life more colorful.

I also want to thank the members of the reading committee, Prof. Y. Ren, Prof. J.B.T.M. Roerdink and Prof. C.H. Slump for their critical comments.

Lastly, my family has been a great support to me all these years. I would like to thank their support and love.

ABOUT THE AUTHOR

Bingjiang Qiu was born on 11 September 1990 in Chaozhou, China. In 2013, he completed his Bachelor education in applied mathematics at Shaoguan University in China. In 2014, he started studying applied mathematics at Shenzhen University in China. After completing his master on medical image processing, he started his PhD programme in 2017, focused on automatic mandible segmentation for 3D virtual surgical planning, supervised by Dr. Peter M.A. van Ooijen and Dr. Max J.H. Witjes. The results of the PhD research are presented in this thesis.

LIST OF PUBLICATIONS

1. Qiu, Bingjiang, Jiapan Guo, Joep Kraeima, Haye H. Glas, Ronald JH Borra, Max JH Witjes, and Peter MA van Ooijen. "Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network." *Physics in Medicine & Biology* 64, no. 17 (2019): 175020.
2. Qiu, Bingjiang, Hylke van der Wel, Joep Kraeima, Haye Hendrik Glas, Jiapan Guo, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter van Ooijen. "Robust and Accurate Mandible Segmentation on Dental CBCT Scans Affected by Metal Artifacts Using a Prior Shape Model." *Journal of personalized medicine* 11, no. 5 (2021): 364.
3. Qiu, Bingjiang, Jiapan Guo, Joep Kraeima, Haye Hendrik Glas, Weichuan Zhang, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter van Ooijen. "Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography." *Journal of Personalized Medicine* 11, no. 6 (2021): 492.
4. Qiu, Bingjiang, Hylke van der Wel, Joep Kraeima, Haye Hendrik Glas, Jiapan Guo, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter van Ooijen. "Mandible Segmentation of Dental CBCT Scans Affected by Metal Artifacts Using Coarse-to-Fine Learning Model." *Journal of Personalized Medicine* 11, no. 6 (2021): 560.
5. Qiu, Bingjiang, Hylke van der Wel, Joep Kraeima, Haye Hendrik Glas, Jiapan Guo, Ronald JH Borra, Max Johannes Hendrikus Witjes, and Peter van Ooijen. "Automatic Segmentation of Mandible from Conventional Methods to Deep Learning—A Review." *Journal of Personalized Medicine* 11, no. 7 (2021): 629.