DEEP LEARNING IN COMPUTER-ASSISTED MAXILLOFACIAL SURGERY

Minnema, Jordi

2022

document version
Publisher's PDF, also known as Version of record

citation for published version (APA)
Minnema, J. (2022). *DEEP LEARNING IN COMPUTER-ASSISTED MAXILLOFACIAL SURGERY*.

DEEP LEARNING IN COMPUTER-ASSISTED MAXILLOFACIAL SURGERY

Jordi Minnema

DEEP LEARNING IN COMPUTER-ASSISTED MAXILLOFACIAL SURGERY

Jordi Minnema

VRIJE UNIVERSITEIT

DEEP LEARNING IN COMPUTER-ASSISTED MAXILLOFACIAL SURGERY

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. J.J.G. Geurts,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Geneeskunde
op dinsdag 15 november 2022 om 11.45 uur
in een bijeenkomst van de universiteit,
De Boelelaan 1105

door

Jordi Minnema
geboren te Blaricum

TABLE OF CONTENTS

CHAPTER 1

GENERAL INTRODUCTION

1

Personalized medicine has become an integral part of health care, and refers to the treatments, decisions and interventions that are tailored to the needs of an individual patient. An important pillar of personalized medicine is understanding how a patient's unique characteristics influence their response to treatments. These characteristics include traits such as age and gender, but also less apparent features such as their genetic profile and anatomy. Acquiring such information allows medical practitioners to optimize the impact of existing treatment methods, which can consequently lead to a higher quality of life and reduce the number of complications and adverse side-effects [1].

In the past decades, personalized medicine has taken a significant leap forward due to the availability of novel image-based treatment modalities. One such treatment modality is computer-assisted surgery (CAS) [2, 3]. CAS, also referred to as image-guided surgery, uses spatial information embedded in three-dimensional (3D) medical images to virtually plan surgical interventions, print surgical guides and navigate in the body. Moreover CAS offers the unique possibility to practice surgery in a stress-free virtual environment, and to detect potential obstacles beforehand, which markedly reduces operating times and costs [4].

One of the first studies that focuses on CAS was published in 1990. In this study, Adams et al.[5] developed a novel navigation aid for skull-base surgeries. The ground-breaking software solution and hardware tool developed in the study allowed surgeons for the first time to virtually plan an operation and use the acquired data to safely navigate in specific areas of the skull base. Since then, the number of CAS applications has grown exponentially. For example, 3D virtual or printed models are currently routinely used to visualize anatomical regions of interest prior to performing surgery [6]. In addition, a large number of surgical simulation models (e.g., anthropomorphic phantoms, human cadavers, virtual reality systems) have been developed to facilitate the training of novice surgeons [7]. Furthermore, over the past decades, computer planned robotic surgery has emerged as an increasingly popular technology to facilitate minimally invasive surgeries [8]. As a result, CAS has rapidly become indispensable in many medical disciplines.

One of the disciplines that has particularly benefited from CAS is maxillofacial surgery [9]. Maxillofacial surgeons often require accurate information on the shape and position of the thin bony structures in the skull area in order to correctly plan and execute surgical interventions. Currently the most common CAS applications in maxillofacial surgery include oral and facial trauma surgery [10], ablative oral/head and neck surgery [11], orthognathic surgery [12], implantology [13] and skull surgery[14].

The current maxillofacial CAS workflow can be divided into three main steps: 1) image acquisition, 2) image processing, and 3) surgical planning (Fig. 1). Image acquisition in the maxillofacial CAS workflow is typically performed using computed tomography (CT) and/or cone-beam computed tomography (CBCT), as these technologies offer the best hard-tissue contrast [15]. Alternatively, magnetic resonance imaging (MRI) modalities are occasionally employed to visualize the surrounding soft tissues. In general, the imaging step consists of two parts: acquisition and reconstruction. During the acquisition phase, CT scanners detect x-ray photons that have traveled through an object or human body, and MRI scanners measure the relaxation of hydrogen atoms. The measured data then need to be reconstructed into interpretable images through a series of

1

mathematical operations. To date, many different reconstruction methods have been developed. Choosing the most suitable reconstruction method depends mainly on the imaging modality, the scan settings, and the anatomical region of interest.

After the imaging step, image processing is often required [16]. Currently, the most widely performed image processing task is segmentation [17]. Image segmentation refers to the partitioning of images into multiple anatomical regions of interest. Although a plethora of anatomical structures can be potential targets for this segmentation process, maxillofacial CAS typically focusses on bony structures in the in the head area. Another common image processing task is image registration [18], which is often used to compare two medical images acquired at different moments in time. This is extremely helpful in determining disease progression and in evaluating treatment effectiveness. Artifact reduction is also a frequently required image processing task since artifacts can obscure and deform relevant tissues and organs in medical images.

Once the aforementioned image processing steps have been completed, the processed images can be used for surgical planning. Surgical planning is a very broad term that encompasses a wide variety of different clinical tasks such as creating treatment plans or designing individualized implants [19]. In order to virtually plan surgeries segmented medical images are typically converted into virtual 3D models that represent the anatomical structure of interest. These virtual 3D models allow medical practitioners to evaluate the exact shape and size of the relevant anatomy and to subsequently create a treatment plan that is tailored to the individual characteristics of the patient. Moreover, such virtual 3D models can be used to design personalized implants or surgical guides that fit seamlessly, using dedicated computer-aided design (CAD) software packages. Such personalized constructs can be subsequently manufactured using 3D-printing technologies to translate the treatment plan and transfer it to the operating room.

Even though CAS is an established method for personalized treatments, each step of the current workflow can introduce inaccuracies that can heavily affect the final treatment outcome [20]. For example, an ever recurring problem in the image acquisition step is the high amount of noise that is often present in medical images [21, 22]. High noise levels are often related to short acquisition times, physical limitations of the scanning equipment or the chosen image reconstruction algorithm [21]. The resulting low-quality noisy images can impair the subsequent image processing step. The main reason for this is that current software packages are unable to deal with such image distortions and subsequently require time-consuming manual interventions. Furthermore, due to the variety of shapes and sizes of the bony structures found in the maxillofacial area, it is extremely challenging to develop reliable image processing methods. In clinical settings sub-optimal image processing
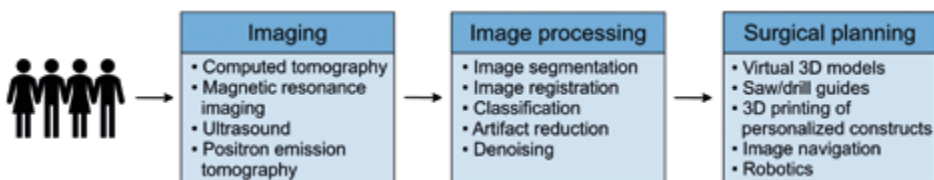

Figure 1. Schematic overview of the maxillofacial computer-assisted surgery workflow.

can lead to erroneous virtual 3D models which, in turn, can result in ill-fitting 3D printed drill/saw guides and implants [23].

Since small inaccuracies in each of the aforementioned CAS steps can amount to severe complications during surgery, experienced medical engineers often need to invest substantial time and effort to guarantee the quality of each step in the CAS workflow. This demanding task comes with challenges such as fatigue, personal preferences, and subjectivity amongst engineers [24]. Moreover, since extensive domain knowledge and manual interactions are often required, the reproducibility of the CAS workflow is often not predictable. Therefore, new methods for automating the different steps in the CAS workflow are sought.

A promising way of circumventing the challenges faced in the CAS workflow is the use of *artificial intelligence*. Artificial intelligence is a branch of computer science that is concerned with automating intelligent behavior [25]. Ever since the term was first coined in 1956, artificial intelligence has grown to become an exceptionally wide field of research. The advantage of artificial intelligence methods over traditional computer algorithms is that they can learn to find patterns in data, without needing explicit rules specified by human experts. Artificial intelligence comprises a wide variety of algorithms, with applications including self-driving vehicles [26], face recognition [27], and quality control [28].

One artificial intelligence algorithm that has revolutionized the field of medicine, is the neural network. Similar to the human brain, neural networks consist of many different neurons that are structured in layers. Each neuron processes incoming data and propagates the processed data to the subsequent layer of neurons through a series of weighted connections. A neural network learns how to process the data by extracting patterns and features from the input data. This means that neural networks essentially learn to find a mathematical function that relates the input data to the desired output. Since neural networks can learn to approximate any mathematical function between input data and desired output data (universal approximation theorem [29]), they offer the opportunity to perform a wide variety of tasks, including those required in the CAS workflow.

In order to extract relevant patterns from the input data, neural networks require training. To date, neural networks are predominantly trained using supervised training strategies. During supervised training, neural networks are provided with a paired dataset of input data and desired output data. The input data are propagated through the network and used to predict an output. The difference between the predicted output and the desired output is subsequently minimized by adjusting the weight of the connections in the neural network. Hence, the weights essentially represent the *memory* of the neural network. Once the weights of the neural network are optimized, the network can be applied to perform the learned task on unseen data.

The first neural network was developed in the 1950s and is commonly known as the multilayer perceptron (MLP) (Fig. 2) [30]. The MLP consists of three types of layers: one input layer, one or more hidden layers, and one output layer (Fig. 2). The input layer, as the name suggest, receives the input data that needs to be processed. The input data are subsequently propagated to the hidden layers of the network and processed so that a desired output can be obtained in the output layer. The number of hidden layers in the MLP is often referred to as its depth. Applying MLPs with many
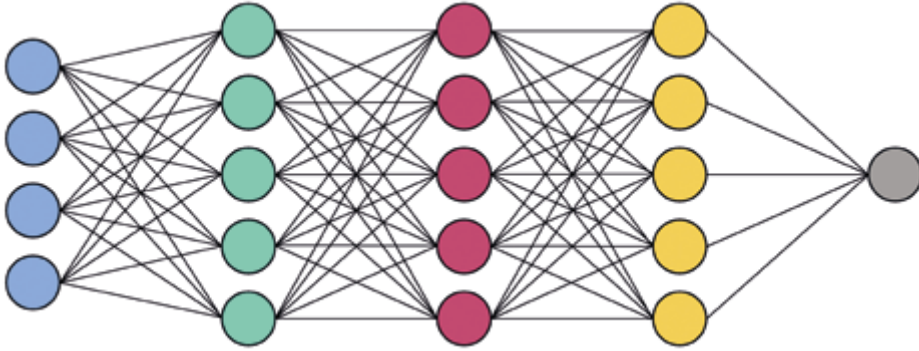
Figure 2. Schematic representation of an ANN with an input layer (left), three hidden layers (middle) and an output layer (right).

hidden layers is therefore often called *deep learning*. Increasing the depth of a network typically increases its ability to learn certain patterns, but also increases the time required to learn them.

Even though MLPs were developed in the 1950s [31], they have only recently gained renewed interest due to advances in computing power and the availability of large training datasets. This has opened new possibilities to train much larger networks (i.e., more hidden layers and neurons) to perform more complex tasks. These developments have inspired researchers to develop a wide variety of new neural network types. One such novel network is the recurrent neural network (RNN), that consists of neurons that not only propagate data to subsequent layers but can also send data back to previous layers of neurons. This allows RNNs to recognize and memorize temporal patterns in input data.

Recently, convolutional neural networks (CNNs) have enjoyed remarkable successes in a wide variety of fields [16]. A CNN is similar to an MLP in that it consists of multiple layers that each extract information from the input data. However, the main difference between the two networks is that CNNs perform convolutional operations instead of regular multiplications, which allows them to capture spatial patterns in the input data. As a result, CNNs are exceptionally good at detecting patterns in pixel-based data such as medical images. Moreover, it has been shown that the convolutional nature of CNNs allows them to process input data much more efficiently than traditional neural networks. As a result, CNNs are generally considered to be the state-of-the-art neural network in the field of medical image analysis.

Since CNNs are capable of capturing spatial patterns in medical images, they are particularly well suited to perform the image processing tasks required in the maxillofacial CAS workflow. Moreover, as CNNs are self-learning algorithms, they can be applied to automatically perform image processing tasks without the need for expert domain knowledge or human supervision. However, relatively few studies have focussed on applying CNNs for maxillofacial CAS. In addition, limited knowledge is currently available in clinical settings on how to set up new deep learning projects and validate new approaches for clinical workflows. As a consequence, the feasibility of applying CNNs for maxillofacial CAS remains largely unknown.

1

GENERAL AIM AND OUTLINE OF THIS THESIS

The aim of thesis is to demonstrate the feasibility of applying CNNs for the image processing tasks of the maxillofacial CAS workflow. More specifically, this thesis focuses on developing, implementing and validating existing and novel CNN approaches for two main applications: medical image segmentation and CT image artifact reduction. This thesis will thus serve as a framework for fellow researchers and an incentive to further incorporate CNN approaches in the maxillofacial CAS workflow. This would markedly reduce the manual work required by medical engineers, and improve the efficiency and reproducibility of treatment planning.

Chapter 2 of this thesis focusses on the application of a relatively simple convolutional neural network to CT image segmentation. An explanation is provided of how CNNs are commonly structured and it demonstrated that a simple CNN can already be extremely useful in automating the segmentation of bony structures. The concept of applying CNNs to CT image segmentation is further developed in **Chapter 3,** where a more advanced CNN was used to simultaneously segment two anatomical regions of interest, namely the jaw and the teeth.

The goal of **Chapter 4** was to compare different approaches of training CNNs for CBCT image segmentation. This chapter describes eight commonly used training strategies and validates the resulting segmentation performance of the CNNs.

Chapters 5 and **6** describe CNNs that were used to process cone-beam CT (CBCT) scans with strong imaging artifacts. More specifically, **Chapter 5** shows that CNNs can accurately segment CBCT scans that were heavily affected by metal artifacts, whereas **Chapter 6** describes a novel symmetry-aware deep learning workflow to reduce cone-angle artifacts.

In **Chapter 7** of this thesis, a broader look is taken at deep learning approaches that can be potentially beneficial for the CAS workflow. This chapter provides a literature review of the neural network approaches that have been applied for three commonly performed tasks of the maxillofacial CAS workflow, namely CT image reconstruction, bone segmentation and surgical planning.

Finally, **Chapter 8** reflects on the developments of deep learning methods, specifically CNNs, for maxillofacial CAS. Moreover, current challenges and promising research avenues are discussed.

1

REFERENCES

1. Mathur S, Sutton J. Personalized medicine could transform healthcare. *Biomedical Reports*. 2017; 7(1): 3–5. doi: 10.3892/br.2017.922.

2. van Baar GJC, Forouzanfar T, Liberton NPTJ, Winters HAH, and Leusink FKJ. Accuracy of computer-assisted surgery in mandibular reconstruction: A systematic review. *Oral Oncology*. 2018; 84: 52–60. doi: 10.1016/j.oraloncology.2018.07.004.

3. Boudissa M, Courvoisier A, Chabanas M, and Tonetti J, Computer assisted surgery in preoperative planning of acetabular fracture surgery: state of the art. *Expert Review of Medical Devices. 2018; 15(1): 81–89. doi: 10.1080/17434440.2017.1413347.*

4. Villanueva-Naquid I, Souberville-Montalvo C, Aguilar-Ponce R, Tovar-Arriaga S, Cuevas-Tello J, et al. Risk assessment methodology for trajectory planning in keyhole neurosurgery using genetic algorithms. *Int J Med Robot*. 2020; 16(2). doi: 10.1002/rcs.2060.

5. Adams L, Krybus W, Meyer-Ebrecht D, Rueger R, Gilsbach J, Moesges R. et al. Computer-assisted surgery. *IEEE Comput. Grap. Appl*. 1990; 10(3): 43–51. doi: 10.1109/38.55152.

6. Swennen GRJ, Mollemans W, Schutyser F. Three-Dimensional Treatment Planning of Orthognathic Surgery in the Era of Virtual Imaging. *Journal of Oral and Maxillofacial Surgery*. 2009; 67(10): 2080-2092. doi: 10.1016/j.joms.2009.06.007.

7. Badash I, Burtt K, Solorzano CA, Carey JN. Innovations in surgery simulation: a review of past, current and future techniques. *Ann. Transl. Med*. 2016; 4(23): 453–453. doi: 10.21037/atm.2016.12.24.

8. Stewart C, Ituarte P, Melstrom K, Warner S, Melstrom L, Lai L, *et al*. Robotic surgery trends in general surgical oncology from the National Inpatient Sample. *Surg Endosc*. 2019; 33(8): 2591–2601. doi: 10.1007/s00464-018-6554-9.

9. Hassfeld S, Mühling J. Computer assisted oral and maxillofacial surgery a review and an assessment of technology. *International Journal of Oral and Maxillofacial Surgery*. 2001; 30(1): 2-13. doi: 10.1054/ijom.2000.0024.

10. Vehmeijer M, van Eijnatten M, Liberton N, Wolff J. A Novel Method of Orbital Floor Reconstruction Using Virtual Planning, 3-Dimensional Printing, and Autologous Bone. *Journal of Oral and Maxillofacial Surgery*. 2016; 74(8): 1608–1612. doi: 10.1016/j.joms.2016.03.044.

11. Bell RB. Computer Planning and Intraoperative Navigation in Cranio-Maxillofacial Surgery. *Oral and Maxillofacial Surgery Clinics of North America*. 2010; 22(1): 135–156. doi: 10.1016/j.coms.2009.10.010.

12. Bengtsson M, Wall G, Miranda-Burgos P, Rasmusson L. Treatment outcome in orthognathic surgery – A prospective comparison of accuracy in computer assisted two and three-dimensional prediction techniques. *Journal of Cranio-Maxillofacial Surgery*. 2018; 46(11): 1867–1874. doi: 10.1016/j.jcms.2017.01.035.

13. Brief J, Edinger D, Hassfeld S, Eggers G, Accuracy of image-guided implantology: Image-guided implantology. *Clinical Oral Implants Research*. 2005; 16(4): 495–501. doi: 10.1111/j.1600-0501.2005.01133.x.

14. van de Vijfeijken SECM, Schreurs R, Dubois L, Becking A. The use of cranial resection templates with 3D virtual planning and PEEK patient-specific implants: A 3 year follow-up. *Journal of Cranio-Maxillofacial Surger*. 2019; 47(4): 542–547. doi: 10.1016/j.jcms.2018.07.012.

15. Huotilainen E, Jaanimets R, Valá ek J, Marcián P, Salmi M, Tuomi J, *et al*. Inaccuracies in additive manufactured medical skull models caused by the DICOM to STL conversion process. *Journal of Cranio-Maxillofacial Surgery*. 2014; 42(5): 259-265. doi: 10.1016/j.jcms.2013.10.001.

16. Litjens G, Kooi T, Bejnordi B, Setio A, Ciompi F, Ghafoorian M, *et al*. A survey *on deep learning in medical image analysis*. *Med. Image. Anal. 2017;42: 60-88.*

17. Maier-Hein L, Eisenmann M, Reinke A, Onogur S, Stankovic M, Scholz P, *et al*. Why rankings of biomedical image analysis competitions should be interpreted with care. *Nat Commun*. 2018; 9(1): 5217. doi: 10.1038/s41467-018-07619-7.

18. Oguro S, Tuncali K, Elhawary H, Morrison PR, Hata N, Silverman SG. Image registration of pre-procedural MRI and intra-procedural CT images to aid CT-guided percutaneous cryoablation of renal tumors. *Int J CARS*. 2011; 6(1): 111–117. doi: 10.1007/s11548-010-0485-9.

19. Su S, Moran K, Robar J. Design and production of 3D printed bolus for electron radiation therapy. *Journal of Applied Clinical Medical Physics*. 2014; 15(4): 4831. doi: 10.1120/jacmp.v15i4.4831.

1

20. van Eijnatten M. Challenges in medical additive manufacturing. 2017.

21. Vaishali S, Rao KK, Rao GVS. A review on noise reduction methods for brain MRI images. *International Conference on Signal Processing and Communication Engineering Systems.* 2015; 363–365. doi: 10.1109/SPACES.2015.7058284.

22. Schulze R, Heil U, Gross D, Bruelmann D, Dranischnikow E, Schwanecke U, *et al.* Artefacts in CBCT: a review. *Dentomaxillofacial Radiology* 2011; 40(5): 265-273. doi: 10.1259/dmfr/30642039.

23. Yong LT, Moy PK. Complications of Computer-Aided-Design/Computer-Aided-Machining-Guided (NobelGuide™) Surgical Implant Placement: An Evaluation of Early Clinical Results. *Clinical Implant Dentistry and Related Research.* 2008; 10(3): 123–127. doi: 10.1111/j.1708-8208.2007.00082.x.

24. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging.* 2016; 35(5): 1153-1159. doi: 10.1109/TMI.2016.2553401.

25. Lugar G, Stubblefield W. *Artificial Intelligence: structures and strategies for complex problem solving.* benjamin/Cummings Pub. Co., 1993.

26. Grigorescu S, Trasnea B, Cocias T, Macesanu G. A survey of deep learning techniques for autonomous driving. *J. Field Robotics.* 2020; 37(3): 362–386. doi: 10.1002/rob.21918.

27. Sun Y, Liang D, Wang X, Tang X. DeepID3: Face Recognition with Very Deep Neural Networks. 2015. *arXiv:1502.00873 [cs].*

28. Iqbal R, Maniak T, Doctor F, Karyotis C. Fault Detection and Isolation in Industrial Processes Using Deep Learning Approaches. *IEEE Trans. Ind. Inf.* 2019;15(5): 3077–3084. doi: 10.1109/TII.2019.2902274.

29. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, *et al.* A guide to deep learning in healthcare. *Nature Medicine.* 2019; 25(1): 24-29. doi: 10.1038/s41591-018-0316-z.

30. Abiodun OI, Jantan A, Omolara AE, Dada KV, Umar AM, Linus OE, *et al.* Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition. *IEEE Access.* 2019; 7: 158820–158846. doi: 10.1109/ACCESS.2019.2945545.

31. McCulloch WS, Pitts W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics.* 1943; 5(4): 115-133. doi: 10.1007/BF02478259.

CHAPTER 2

CT IMAGE SEGMENTATION OF BONE FOR MEDICAL ADDITIVE MANUFACTURING USING A CONVOLUTIONAL NEURAL NETWORK

Jordi Minnema, Maureen van Eijnatten, Wouter Kouw,
Faruk Diblen, Adriënne Mendrik, Jan Wolff

2

ABSTRACT

Background

The most tedious and time-consuming task in medical additive manufacturing (AM) is image segmentation. The aim of the present study was to develop and train a convolutional neural network (CNN) for bone segmentation in computed tomography (CT) scans.

Method

The CNN was trained with CT scans acquired using six different scanners. Standard tessellation language (STL) models of 20 patients who had previously undergone craniotomy and cranioplasty using additively manufactured skull implants served as "gold standard" models during CNN training. The CNN segmented all patient CT scans using a leave-2-out scheme. All segmented CT scans were converted into STL models and geometrically compared with the gold standard STL models.

Results

The CT scans segmented using the CNN demonstrated a large overlap with the gold standard segmentation and resulted in a mean Dice similarity coefficient of 0.92 ± 0.04. The CNN-based STL models demonstrated mean surface deviations ranging between −0.19 mm ± 0.86 mm and 1.22 mm ± 1.75 mm, when compared to the gold standard STL models. No major differences were observed between the mean deviations of the CNN-based STL models acquired using six different CT scanners.

Conclusions

The fully-automated CNN was able to accurately segment the skull. CNNs thus offer the opportunity of removing the current prohibitive barriers of time and effort during CT image segmentation, making patient-specific AM constructs more accessible.

1. INTRODUCTION

Additive manufacturing (AM), also referred to as three-dimensional (3D) printing, is a technique in which successive layers of material are deposited on a build bed, allowing the fabrication of objects with complex geometries [1,2]. In medicine, additive manufactured tangible models are being increasingly used to evaluate complex anatomies [3,4]. Moreover, AM can be used to fabricate patient-specific constructs such as drill guides, saw guides, and medical implants. Such constructs can markedly reduce operating times and enhance the accuracy of surgical procedures [4]. AM constructs have proven to be particularly valuable in the field of oral and maxillofacial surgery due to the plethora of complex bony geometries found in the skull area.

The current medical AM process comprises four different steps: 1) image acquisition; 2) image processing; 3) computer-aided design; and additive manufacturing (Fig. 1). Image acquisition is commonly performed using computed tomography (CT) since it offers the best hard tissue contrast [5]. During step 2 of the medical AM process, the acquired CT scan needs to be converted into a 3D surface model in the standard tessellation language (STL) file format. This STL model can be used to design patient-specific constructs (step 3) that can subsequently be fabricated using a 3D printer (step 4).

The most important step in the CT-to-STL conversion process is image segmentation: the partitioning of images into regions of interest that correspond to a specific anatomical structure (e.g., "bone"). To date, the most commonly used image segmentation method in medical AM is global thresholding [6]. However, global thresholding does not take CT artifacts and noise into account, nor the intensity variations between different CT scanners that often result in inconsistent segmentation results [7]. Therefore, extensive manual post-processing and anatomical modeling is often indispensable. Moreover, due to subjectivity, fatigue, and variance amongst medical engineers, the quality of threshold-based image segmentations can differ markedly.

Many alternative (semi-)automatic image segmentation methods such as edge detection, region growing, statistical shape models, atlas-based methods, morphological snakes, active contouring, and random forests, have been developed over the last decades [6,8]. These automated methods are suitable to some extent for segmenting images with intensity inhomogeneities but often fail when applied to images acquired using different CT scanners and imaging protocols with varying noise levels. The inherent limitations have subsequently dampened the enthusiasm amongst physicians with an interest in adapting AM in clinical settings. Therefore, new methods to automate image segmentation are sought.

Over the past few years, there have been unparalleled advances in the field of artificial intelligence, especially after the ground-breaking performance of the convolutional neural network (CNN) developed by Alex Krizhevsky for the ImageNet challenge in 2012 [9]. A CNN is structured in layers. Each layer comprises multiple computational building blocks called neurons that share weighted connections with neurons in subsequent layers. During training, these layers extract features from training images, after which the CNN can recognize these features in new, unseen images to perform a certain task, such as segmentation.

The aim of the present study was to develop and train a CNN for skull segmentation in CT scans. The CNN was trained using a unique patient dataset that represented variations that are commonly

2

2

found in clinical CT scans. This will hopefully help overcome the aforementioned segmentation issues in medical AM and reduce the time-consuming and costly role of manual processing.

1.1. Related work

Traditionally, (semi-)automatic rule-based methods, such as edge detection [10], region based-methods [11,12], and level sets [13,14], have been used for medical image segmentation. The main strength of such rule-based approaches is their computational efficiency in terms of time and memory. Rule-based methods require the construction of generic priors to ensure correct segmentation. However, defining such generic priors is often a manual task, which can be cumbersome when segmenting images with high noise levels or artifacts. Therefore, data-driven approaches were developed. Data-driven approaches do not depend on a fixed set of manually chosen rules but aim to extract relevant information from large numbers of medical images. Examples of data-driven approaches that have been frequently used for medical image segmentation are random forests [15,16], statistical shape models [17–19], and atlas-based approaches [20,21]. Although many of these approaches offer more accurate segmentation results than rule-based methods, data-driven methods still lack the generalizability to segment medical images of varying shapes, sizes and properties [8]. Moreover, data-driven methods often fail when applied to images acquired using different CT scanners and imaging protocols.

One way to overcome these limitations is to use deep learning algorithms. Recent advances in Graphical Processing Units (GPU) computing have enabled the development of efficient and intelligent deep learning approaches [22]. More specifically, convolutional neural networks (CNNs) have opened up a wealth of promising opportunities across a number of image-based disciplines. For example, Prasoon et al. (2013) successfully employed a CNN for the segmentation of knee cartilage in magnetic resonance (MR) images [23]. They demonstrated the potential of CNNs and outperformed the then state-of-the-art k-Nearest Neighbor classification method. This has motivated many researchers to use CNNs for various medical segmentation tasks, such as the segmentation of brain tissue [24–26], prostate [27], bone [28,29], and tumors [30–33] in MR images. Furthermore, multiple studies have been conducted on the segmentation of kidneys [34] and the pancreas [35–37] in CT scans. A few studies have investigated the use of CNNs for bone segmentation in CT scans. For example, Vania et al. (2017) employed a CNN for the segmentation of the spine [38]. Moreover, Išgum et al. (2018) proposed an iterative CNN for the segmentation of vertebrae that outperformed alternative segmentation methods [39].

2. NOVELTIES

The novelty of the present study is that it demonstrates the feasibility of training a CNN on a patient dataset for which a unique, high-quality gold standard was available, namely, STL models created by experienced medical engineers. To the best of our knowledge, no studies have been performed in which such "engineered" STL models were used as gold standard. Moreover, CT scans were acquired using different CT scanners and acquisition protocols in order to represent the variability that is commonly found amongst clinical CT datasets.

3. MATERIALS AND METHODS

This study followed the principles of the Helsinki Declaration and was performed in accordance with the guidelines of the Medical Ethics Committee of the VU University Medical Center Amsterdam. The Dutch Medical Research Involving Human Subjects Act (WMO) did not apply to this study (Ref: 2017.145).

3.1. Data acquisition

The CNN was trained using CT scans and STL models of 20 patients who had previously undergone craniotomy and cranioplasty using additively manufactured skull implants. The 20 CT scans were acquired using different CT scanners and imaging protocols in order to represent the variability that is commonly found amongst clinical CT datasets (Table 1). The bony structures in all 20 patient CT scans were initially segmented using global thresholding combined with manual corrections, i.e., removal of noise, artifacts and unrelated structures, such as the head rest in the CT scan, using the available segmentation editing tools in Mimics software (Mimics v20.0, Materialise, Leuven, Belgium). Medical engineers subsequently converted the segmented CT scans into STL models and imported these STL models into medical computer-aided design 3-matic software (3-matic v11.0, Materialise, Leuven, Belgium) for post-processing. The post-processing procedure included the removal of unconnected triangles (noise), the closing of unnatural gaps, and the smoothening of defect edges in the skull (Fig. 1, step 3). Hence, these post-processed STL models contained information that had been directly defined by medical engineers, and therefore served as the gold standard in our study.

3.2. Data processing: generating gold standard labels

All 20 gold standard STL models were subsequently used to create gold standard labels, namely "bone" or "background". To this end, all STL models had to be aligned with their corresponding CT scans (Fig. 1, step 'A'). Each STL model was aligned on a reference model with the same orientation as the CT scan using a local best-fit algorithm in GOM Inspect® software (GOM Inspect 2017, GOM GmbH, Braunschweig, Germany). The aligned STL models were subsequently converted into gold standard labels using the mesh-to-label conversion [40] module in 3D Slicer software (v. 4.6.2) (Fig. 1, step 'B') [41,42].

3.3. Data processing: generating patches

All 20 CT scans were normalized by rescaling the voxel values between 0 and 1 (Fig. 1, step 'C'). Normalization was performed as follows:

$$x_{norm} = \frac{x - CT_{min}}{CT_{max} - CT_{min}} \qquad (1)$$

2
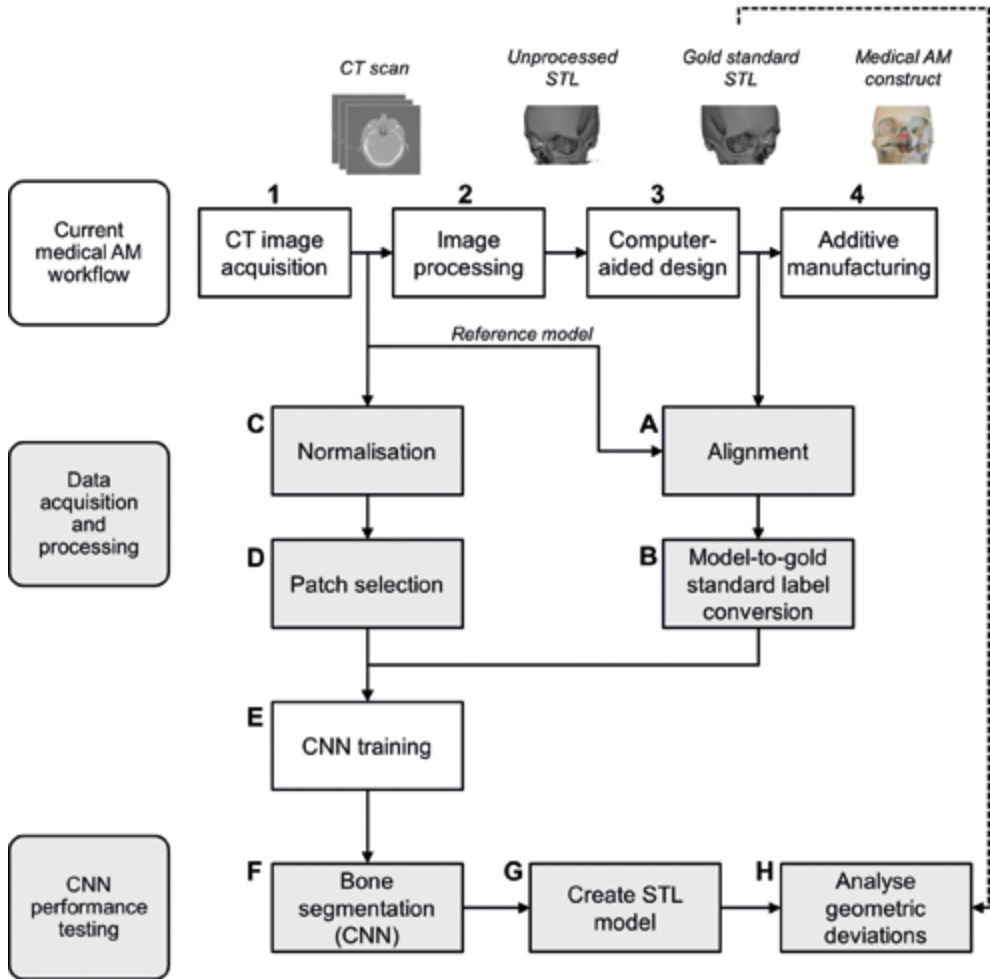
Figure 1. Schematic overview of the study. The current medical additive manufacturing (AM) workflow is presented in the top of the figure. CT scans and STL models acquired in this process were used to train a convolutional neural network (CNN).

where x_{norm} is the normalized voxel value between 0 and 1, x is the voxel intensity (in Hounsfield Units), CT_{min} is the minimum voxel intensity (in Hounsfield units), and CT_{max} is the maximal voxel intensity (in Hounsfield units) in the CT scan.

The normalized CT scans were used to select voxels from a confined rectangular region of interest within each 2D axial CT slice that contained bone (Fig. 2A). These voxels were subsequently used to create 33x33 patches centered on each voxel (Fig. 2B). Thus the created patches contained the intensity values of the surrounding voxels. The patches were then used to train the CNN to classify the center voxel of each patch as either "bone" or "background" (Fig. 1, step 'E').

Selecting patches from CT scans is a highly data imbalanced problem since bone voxels comprise only a small part of the total number of voxels. Training the CNN using the true distribution of

Table 1. Image acquisition parameters of the patients' CT scans used in this study.

Patient ID	Scanner Type	Reconstruction Kernel	Voxel size (mm)	Slice thickness (mm)	Tube voltage (kVp)	Tube Current (mA)
1	GE Discovery CT750 HD	BONEPLUS	0.47	0.625	120	200
2	GE Discovery CT750 HD	BONEPLUS	0.543	0.625	120	200
3	GE Discovery CT750 HD	BONEPLUS	0.49	1.25	140	626
4	GE Discovery CT750 HD	BONEPLUS	0.553	0.625	120	200
5	GE Discovery CT750 HD	BONEPLUS	0.504	0.625	120	200
6	Siemens Sensation 64	H31s	0.482	1	120	213
7	Siemens Sensation 64	H70h	0.568	0.75	120	300
8	Siemens Sensation 64	H70h	0.459	0.75	120	300
9	Siemens Sensation 64	H32s	0.4	0.6	120	380
10	Siemens Sensation 64	H60s	0.443	1	120	323
11	Siemens Somatom Definition AS+	J30s/3	0.404	1.5	120	80
12	Siemens Somatom Definition AS+	J70h/1	0.391	1	120	131
13	Siemens Somatom Definition AS+	J30s/3	0.412	1	120	113
14	Siemens Somatom Definition AS+	J70h/1	0.342	1	120	115
15	Siemens Somatom Force	Hr64h\1	0.432	1	120	184
16	Siemens Somatom Force	Hf38s/3	0.45	1	120	164
17	Siemens Somatom Force	Hr64h\1	0.39	1	120	166
18	Philips iCT 256	UB	0.513	0.9	120	131
19	Philips iCT 256	D	0.52	0.9	120	131
20	Philips Brilliance 64	UC	0.486	1	120	149

bone voxels would cause the CNN to be biased towards classifying background voxels. Therefore, a balanced dataset was used to train the CNN, as proposed by Havaei et al. [30]. This means that an equal number of "bone" patches and "background" patches were randomly selected from the 20 CT scans, which resulted in 1 000 000 patches for each class, hence 2 000 000 patches in total.

3.4. CNN architecture

The CNN architecture used in this study (Fig. 3) was initially developed by N. Aldenborgh for tumor segmentation in MR images [43]. The authors of the present study substantially adapted the aforementioned CNN for bone segmentation in CT images. One of the major differences between the current CNN and the CNN developed by Aldenborgh was the number of labels and input channels used to feed the CNN. Aldenborgh used 5 labels to segment different anatomical structures in MR images of the brain, whereas the modified algorithm implemented in this study used 2 labels to segment CT images into "bone" and "background" (air and soft tissues). In addition, Aldenborgh used 4 input channels to train their CNN on 4 different MRI sequences, whereas we used one input channel. Full details of our CNN architecture and settings can be publicly accessed online [44].
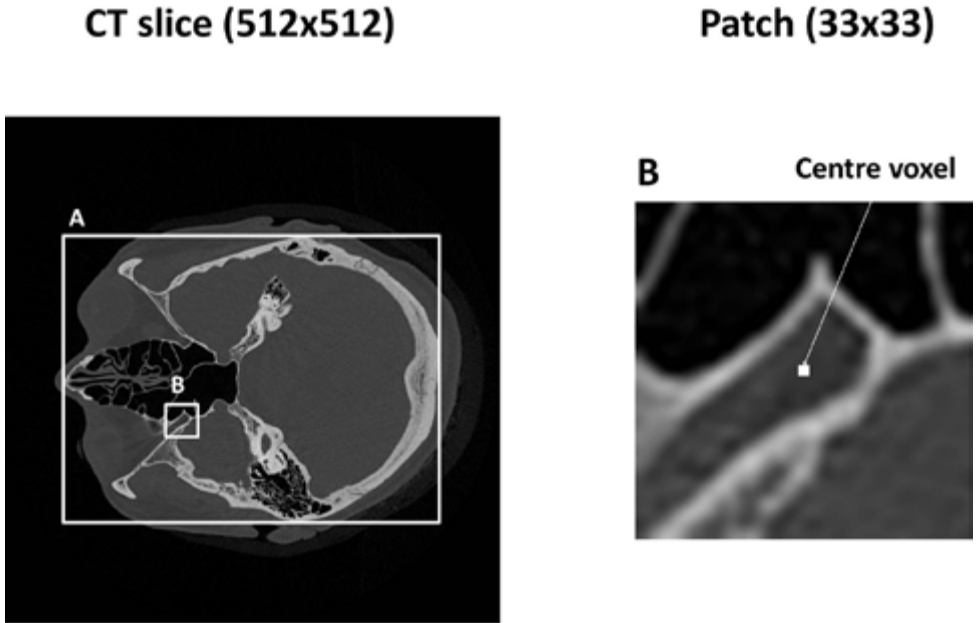
Figure 2. Patch acquisition from CT slices. Box A illustrates the confined rectangular region of interest enclosing all bone voxels in the CT slice, and box B represents the selected patch.

The CNN architecture used in this study consisted of four blocks, each comprising four layers (Fig. 3). The first layer of each block was a convolutional layer (Fig. 3A). Each convolutional layer was composed of a set of different kernels. These kernels are essentially structure detectors that search for particular geometric shapes in the input images by performing a convolution operation. Traditionally, particular kernel shapes are designed by an engineer to perform a certain task. A CNN learns which kernel shapes are the most suited to perform the task at hand.

In order to interpret the output of the convolutional layer, additional layers were included in the CNN architecture. An activation layer (Fig. 3B), i.e., rectified linear units (ReLU), was applied immediately after each convolutional layer to introduce a non-linear property to the CNN. This subsequently increased the flexibility of the CNN to detect different anatomical structures [9]. The output of the activation layer was normalized for numerical stability using a normalization layer (Fig. 3C) [45]. The last layer in each block was the pooling layer (Fig. 3D), which achieved spatial invariance in the detected structures [46].

After the four blocks, a final classification layer was used in the CNN architecture. This layer quantified the difference between the CNN prediction and the gold standard labels in each iteration of the training procedure. After each iteration, the kernels were refined in order to reduce the difference in the next iteration. Generally, a CNN is initiated with random kernel structures that are refined during training until the performance of the CNN no longer improves. In other words, the training of CNNs is characterized by identifying which kernel structures are relevant to solve the task at hand.
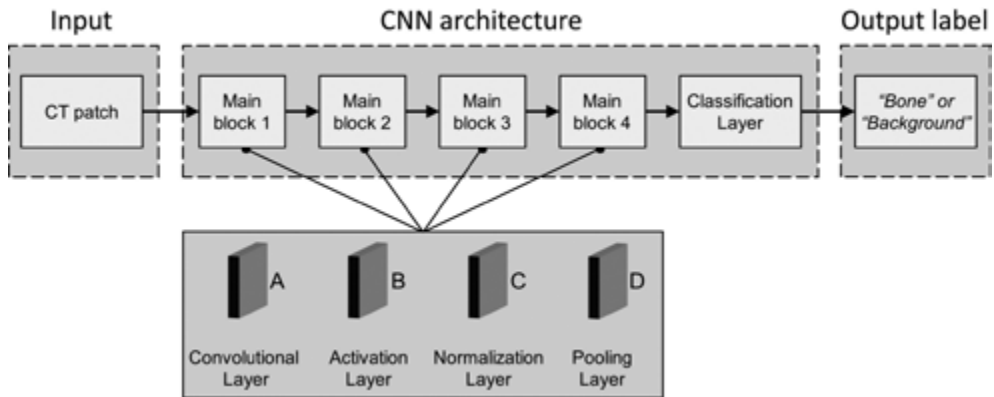
Figure 3. Schematic overview of the CNN architecture.

3.5. Implementation details

The sizes of the convolution kernels were set to 7x7 for the first layer, 5x5 for the second and third layers, and 3x3 for the final layer. Max-pooling operations were performed in the pooling layers using 2x2 kernels. Moreover, dropout [47] was applied after each convolutional layer with a value of 0.1. Training was performed using a batch size of 128 patches. In addition, the RMSprop optimizer [48] was used to update the kernel weights. In this context, the learning rate (α) and the decay (d) were set to α = 0.005 and d = 0.01.

CNN training (Fig. 1, step 'E') was performed using a Linux desktop computer (HP Workstation Z840) with 64 GB RAM, a Xeon E5-2687 v4 3.0GHZ CPU, and a GTX 1080 Ti GPU card. Implementation of the code was performed in Keras [49], a python library that compiles symbolical expressions into C/CUDA code that can then run on GPUs. The training of the CNN took approximately 5 min for each epoch, while the segmentation of one CT slice took approximately 20 s.

3.6. CNN performance testing

The performance of the CNN was evaluated by segmenting CT scans that were not used for training purposes (Fig. 1, step 'F'). To this end, the CNN was trained using a leave-2-out scheme: patches were acquired alternately from 18 of the 20 CT scans, after which the CNN was used to segment the 2 CT scans that were not used for training. Segmentations of the CT scans were performed by classifying each voxel individually. For this purpose, patches (33x33) were generated around each voxel in the CT scan. These patches were subsequently forwarded through the trained CNN, which resulted in label predictions (i.e., "bone" or "background") of all voxels.

The quality of the CNN segmentation was evaluated using the Dice similarity coefficient (DSC). The definition of the DSC is given in Equation (2), where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives.

2

$$DSC = \frac{2TP}{2TP + FP + FN} \qquad (2)$$

The TP, FP, and FN of the CNN segmentations were calculated with respect to the gold standard labels. Since these gold standard labels were derived from STL models that were often cropped to a specific region of interest and thus did not always cover all bony structures in the original CT scan, the TP, FP, and FN values were only calculated within this region of interest.

All 20 CT scans segmented using the CNN were subsequently converted into STL models using 3D Slicer software (Fig. 1, step "G"). The resulting STL models were geometrically compared with the corresponding gold standard STL models using the surface comparison function in GOM Inspect® software. This surface comparison was performed on the largest connected component of the STL models and computes the perpendicular distance between each polygon point on the gold standard STL model and the corresponding CNN-based STL model. Signed deviations between −5.0 mm and +5.0 mm were measured between the CNN-based STL models and the gold standard STL models (Fig. 1, step 'H'). The mean deviations and standard deviations (SDs) were calculated for all CNN-based STL models of the skulls as well as for a manually selected region around the edges of each skull defect.

4. RESULTS

Figure 4 shows axial CT slices of four patients with a skull defect as well as the corresponding gold standard labels created by a medical engineer ("gold standard segmentation"). The labels acquired using the trained CNN ("CNN segmentation") are presented in Figure 4C. Differences between the gold standard segmentation and the CNN segmentation are visualized in Figure 4D: correctly labeled voxels are marked in gray, false negatives in white, and false positives in black. In addition, the DSCs between the gold standard segmentations and the CNN segmentations are presented in Table 2. DSCs ranged between 0.86 (patient 7) and 0.97 (patient 12), with a mean DSC of 0.92 ± 0.04.

Figure 5 shows four typical examples of skull STL models acquired using the trained CNN. The STL models depicted in Figure 5 A, C, E, and G correspond to patient 7, 9, 15, and 18, respectively. The (signed) geometric deviations between these CNN-based STL models and the gold standard STL models are presented using color maps. All CNN-based STL models generally resembled the gold standard STL models created by the medical engineer, with mean deviations ranging between −0.19 mm ± 0.86 mm (patient 1) and 1.22 mm ± 1.75 mm (patient 7) (Fig. 6). The mean of the absolute mean deviations of all CNN-based STL models was 0.44 mm ± 0.36 mm. No differences were observed between the mean deviations of the STL models acquired using the six different CT scanners included in this study, namely, GE Discovery CT750 HD, Siemens Sensation 64, Siemens Somatom Definition AS+, Siemens Somatom Force, Philips iCT 256, and Philips Brilliance 64 (Fig. 6).

Furthermore, the edges of the skull defects in the CNN-based STL models were typically well represented with smooth boundaries. In 17 of the CNN-based STL models, the edges of the skull defects demonstrated smaller mean deviations than the whole skull (Fig. 7). The mean of the absolute mean deviations of all defect edges was 0.27 mm ± 0.29 mm
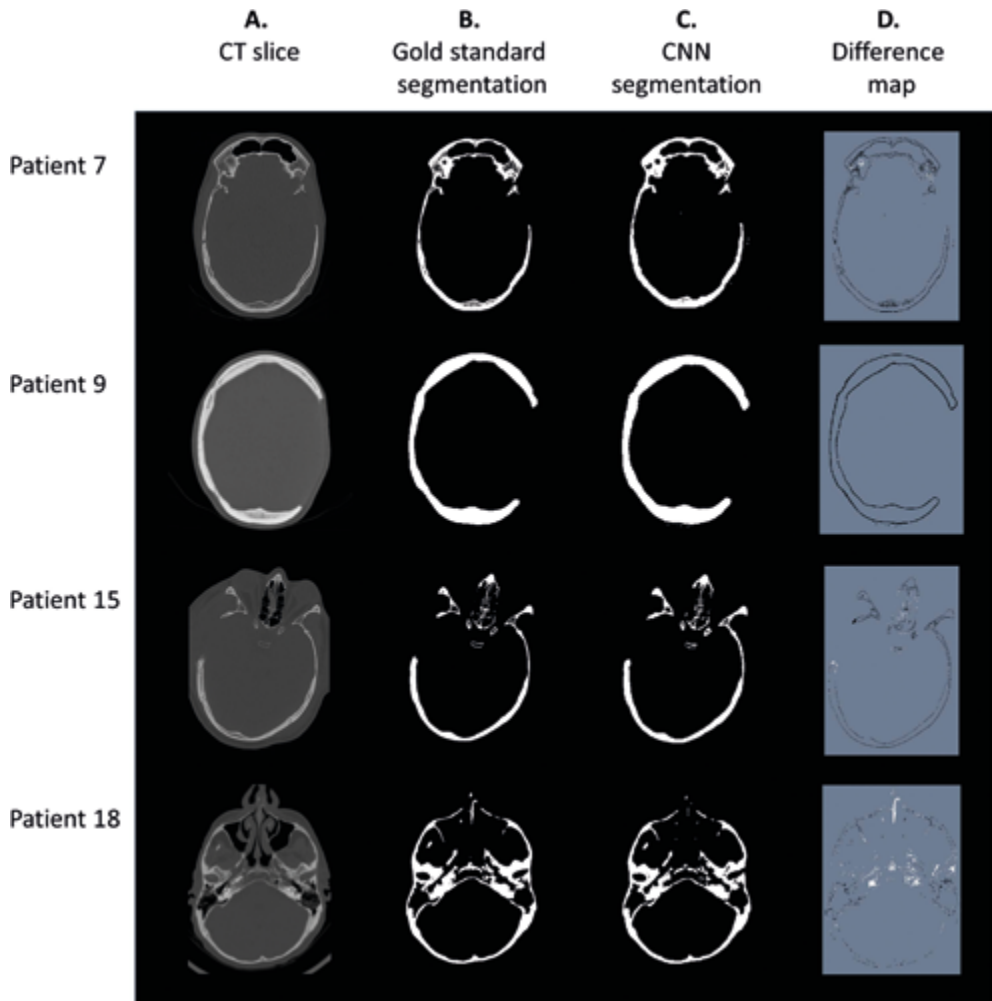
2

Figure 4. Example of four axial CT slices of patients with a skull defect (A), the corresponding gold standard segmentation (B), the CNN segmentation (C), and the differences between the gold standard segmentation and the CNN segmentation (D).

5. DISCUSSION

The CT-to-STL conversion currently required in medical AM remains a challenge. This is impeding the widespread use of additive manufactured constructs in clinical settings. Therefore, the present study aimed to develop and train a CNN for automated CT image segmentation of bone.

The bone segmentation performance of the trained CNN was good (Fig. 4). Differences between the gold standard segmentation and the CNN segmentation were generally in the order of magnitude of one voxel, which is illustrated in Figure 4D. More specifically, the DSCs varied between the different CT scans with a mean DSC of 0.92 ± 0.04 (Table 2). These results are in good agreement with those reported by Powell et al. [34] (2017), who used a fully-automated atlas-based approach for

Table 2. Dice similarity coefficient (DSC) between the gold standard segmentation and the CNN segmentation of all patient CT scans.

Patient ID	DSC
1	0.96
2	0.93
3	0.89
4	0.95
5	0.89
6	0.93
7	0.86
8	0.91
9	0.87
10	0.94
11	0.88
12	0.97
13	0.90
14	0.96
15	0.92
16	0.93
17	0.95
18	0.96
19	0.97
20	0.92
Mean	**0.92 ± 0.04**

the segmentation of temporal bones and obtained DSCs ranging between 0.58 and 0.91. Moreover, the mean DSC found in the present study is comparable to the results reported by Fu et al. [50] (2017) who proposed an atlas-based method and achieved a mean DSC of 0.94 ± 0.01 when segmenting the mandible. Torosdagli et al. [51] (2017) developed a 3D gradient-based fuzzy connectedness method for the segmentation of the mandible and reported a DSC of 0.91. Furthermore, the DSCs found in the present study are higher than those reported by Jafarian et al. [14] (2014) and Ghadimi et al. [52] (2016), who segmented cranial bones of neonates using a level-set method and achieved mean DSCs of 0.87 and 0.81, respectively. It must be noted, however, that the differences between the DSCs across studies must be interpreted with caution due to the variances in the datasets used.

All CNN-based STL models generally resembled the gold standard STL models initially created by an experienced medical engineer (Fig. 5). Interestingly, the skull CNN-based STL models resulted in an absolute mean deviation of 0.44 mm ± 0.36 mm (Fig. 6), whereas the selected region around the defect edges of the skull resulted in a smaller absolute mean deviation of 0.27 mm ± 0.29 mm (Fig. 7). These differences could have been caused by the medical engineer manually removing all noise residuals and smoothening the defect edges of the gold standard STL models to ensure the skull implant had a good fit. Since these gold standard STL models were used to generate training data, the CNN learned to reproduce these smooth and accurate defect edges in its segmentation.
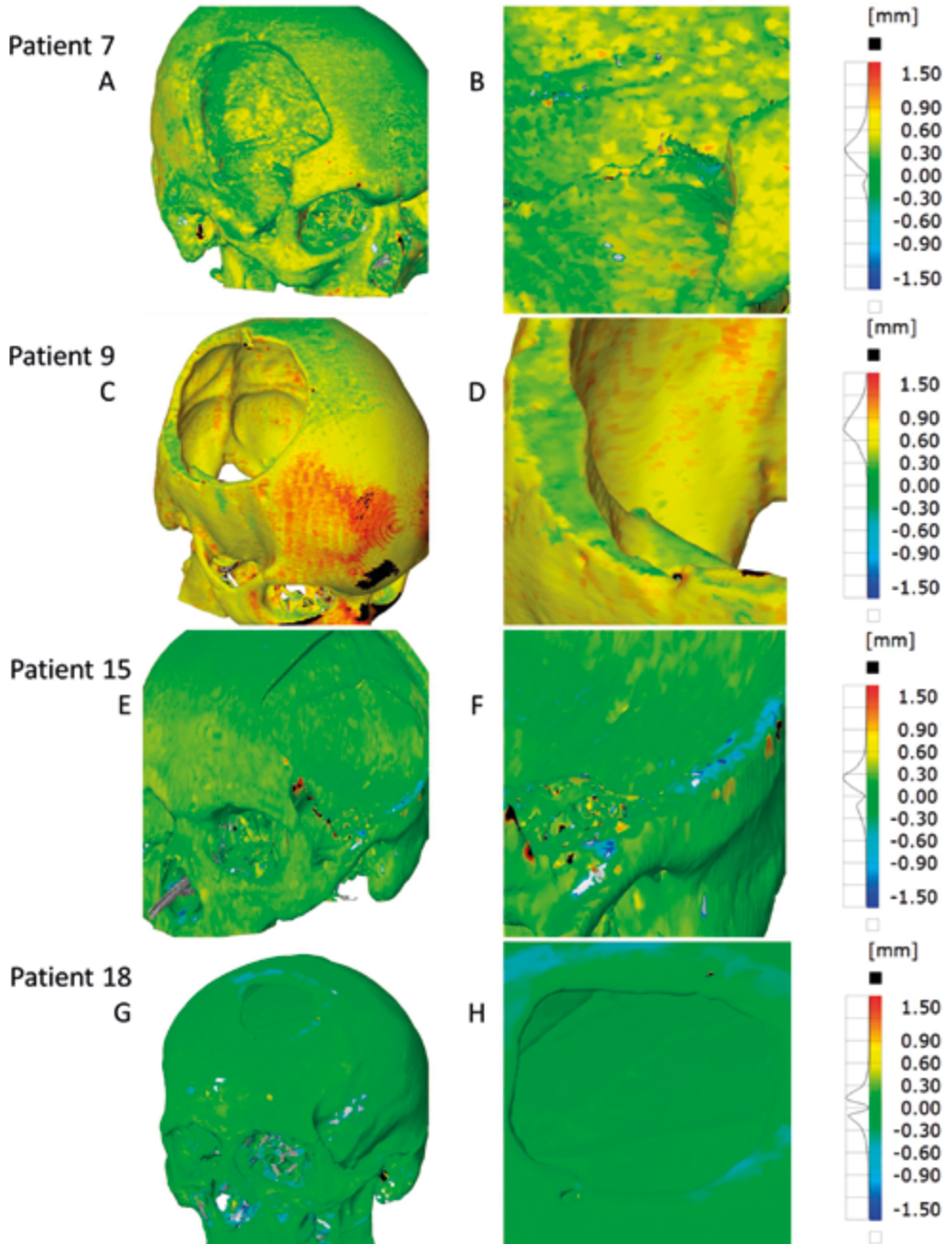
Figure 5. Color maps of the signed geometric deviations between four typical CNN-based STL models and their corresponding gold standard STL models (A, C, E, and G). Images B, D, F, and H present a more detailed visualization of the skull defect edges.

2
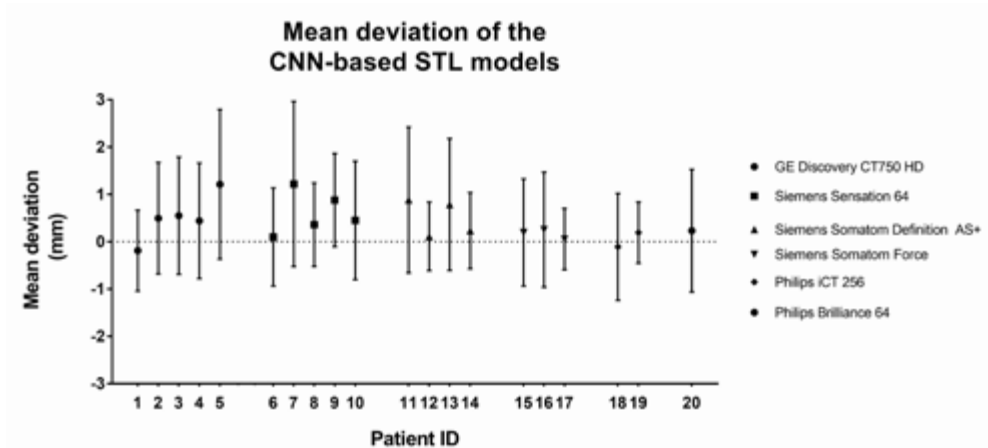
Figure 6. Mean deviation and standard deviation (SD) of all CNN-based STL models with respect to the corresponding gold standard STL models.
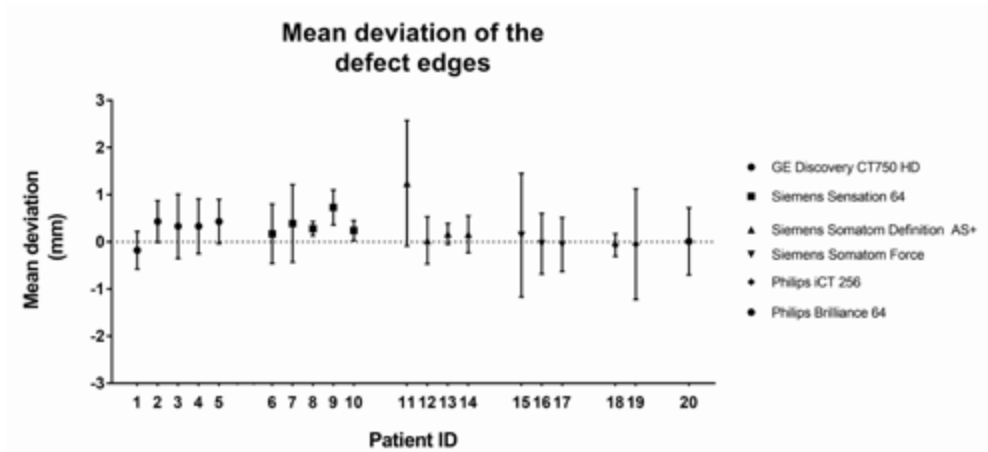
Figure 7. Mean deviation and standard deviation (SD) of the defect edges of all CNN-based STL models with respect to the corresponding gold standard STL models.

Another finding in the present study was that 18 of the 20 CNN-based STL models were larger than the gold standard STL models (Fig. 6). This phenomenon could be the result of the balanced dataset that was used for CNN training containing a higher proportion of bone voxels than the original CT scans. This implies that the CNN learned to label more voxels as bone when compared with the gold standard (Fig. 4). However, there is currently no general consensus in the literature on the effect of the distribution of the training data, namely balanced or unbalanced, on the performance of CNNs [53–55]. The optimal data distribution most likely depends on the specific properties and features of the dataset. Further work is therefore required to establish the viability of considering the data distribution as a tunable hyperparameter that can be optimized for specific datasets. However, this

would require many (cross-validated) training sessions and would thus be a time-consuming and computationally expensive procedure.

The mean deviations of the CNN-based STL models in this study ranged between −0.19 mm ± 0.86 mm and 1.22 mm ± 0.39 mm (Fig. 6). These results differ from those reported by Rueda et al. [56] (2006), who calculated a mean geometrical distance of 1.63 mm ± 1.63 mm using a fully-automated active appearance model for the segmentation of cortical bone in the skull. However, the mean deviations found in the present study are comparable with those acquired using a fully-automatic atlas-based segmentation method developed by Steger et al. [20] (2012) that resulted in a mean deviation of 0.84 mm. The aforementioned findings suggest that CNNs offer comparable bone segmentation performance to the fully-automated segmentation methods currently used in medical AM.

Another interesting finding was that the six different CT scanners used in this study did not seem to have an effect on the mean deviations of the CNN-based STL models (Figs. 6 and 7). This indicates that the CNN was able to generalise intensity variations between different CT scanners and imaging protocols. In comparison, traditional rule-based segmentation methods, such as global thresholding, typically do not generalise well because they are based on a fixed set of features in images, e.g., intensities. A major advantage of CNNs is that they can automatically learn which characteristic features are relevant to segment bone in multiple CT scans, which allows the CNN to segment bone in new, unseen CT scans.

Recent advances in CNN architectures for segmentation have led to the development of fully convolutional neural networks (fCNNs) [27,57,58]. fCNNs take input images of varying sizes and produce a probability map, rather than a classification output for a single voxel. This allows fCNNs to be trained more efficiently compared with CNN architectures for the classification of single voxels [57]. However, the major challenge with fCNNs is that they require large numbers of training images [59,60]. In medical settings, the amount of data that can be acquired is often limited due to privacy regulations and ethical considerations. As a consequence, data augmentation is often necessary [60]. Therefore, the authors of the present study implemented a patch-based CNN for classification of single voxels as initially proposed by Cireşan et al. [61] (2012). This approach has been shown to perform well on multiple image segmentation tasks [62–64]. By extracting a large number of patches from multiple CT slices, sufficient input data can be acquired to train a CNN. Although patch-based CNNs are computationally less efficient than fCNNs, they are easier to train and are more robust to variations within CT slices [65].

A unique feature of the present study is that the gold standard labels that were used to train the CNN were obtained from the STL models of patients who had undergone cranioplasty using AM skull implants created by a medical engineer. Since all gold standard STL models had been successfully used to design patient-specific skull implants, their accuracy can be considered sufficient for medical purposes. One limitation of using STL models as a gold standard is the mesh-to-label conversion algorithm that is required to convert the STL models into gold standard labels. The mesh-to-label conversion algorithm implemented in 3D Slicer resulted in a small number of bone voxels incorrectly labeled as background. However, since this phenomenon was only observed

2

in thin bony structures (≤1 voxel) that were not situated in the region of interest (the skull defect), it can be assumed that it did not affect the segmentation performance of the CNN.

This study presents a framework for fully-automatic CT image seg-mentation of bone using a CNN. The trained CNN was able to accurately segment the skull in a large variety of CT scans of patients with a skull defect. A CNN could thus help to overcome the limitations of the current image segmentation methods commonly used in medical AM, i.e., global thresholding combined with extensive and time-consuming manual post-processing. The current implementation of the CNN takes approximately 20 s to segment a CT slice, which implies that segmenting a full CT scan would take approximately an hour. Nevertheless, since the segmentation process is fully automated, the medical engineer can spend this time on other relevant tasks needed to manufacture patient-specific skull implants. Automating the image segmentation step would therefore not only reduce the subjectivity and the variance amongst medical engineers, it would also improve the cost-effectiveness of patient-specific AM constructs.

Future research should focus on the validation of the CNN using patches from multiple orthogonal planes (i.e., axial, sagittal, and coronal). Furthermore, since low-dose CT and cone-beam CT scans can be affected by higher noise levels than conventional CT scans, future studies should be undertaken to assess the performance of the CNN on low-dose CT and cone-beam CT scans. Moreover, we suggest CNNs are trained and tested for the segmentation of other (bony) structures, such as the mandible [66] and vertebrae [39]. In addition, the segmentation performance could be further enhanced by exploring alternative CNN architectures, such as the TwoPathCNN [30] and the mixed-scale dense CNN [67], which can incorporate global features in CT scans as well as local details. Finally, new platforms and infrastructures (e.g., cloud computing) are required that allow sharing data, reproducing results, and benchmarking algorithms. This will help researchers to adapt to the rapid developments in the field of deep learning.

6. CONCLUSION

This study presents a CNN developed for bone segmentation that was trained using labeled CT scans of patients that had been treated using patient-specific additive manufactured skull implants. The CNN segmentation demonstrated a high overlap with the gold standard segmentation (DSC = 0.92 ± 0.04). The quality of the resulting CNN-based STL models was good (mean deviation = 0.44 mm ± 0.36 mm), particularly around the defect edges (mean deviation = 0.27 mm ± 0.29 mm). CNNs offer the opportunity of removing the prohibitive barriers of time and effort during CT image segmentation, making patient-specific AM constructs more affordable, and thus more accessible to clinicians. Further research should be undertaken to investigate the bone segmentation performance of different CNN architectures.

CONFLICTS OF INTEREST

None declared.

ACKNOWLEDGEMENTS

2

REFERENCES

1. Tapia G, Elwany A. A review on process monitoring and control in metal-based additive manufacturing. *J. Manuf. Sci. Eng*. 2014; 136(6): 060801. doi:10. 1115/1.4028540.

2. Gross BC, Erkal JL, Lockwood SY, Chen C, Spence DM. An evaluation of 3D printing and its potential impact on biotechnology and the chemical sciences. Anal. Chem. 2014; 86(7): 3240–3253. doi:10.1021/ac403397r.

3. Ventola CL. Medical applications for 3D printing: current and projected uses. P T. 2014; 39(10): 704–711. doi:10.1016/j. infsof.2008.09.005.

4. Salmi M, Paloheimo KS, Tuomi J, Wolff J, Mäkitie A. Accuracy of medical models made by additive manufacturing (rapid manufacturing). *J. Cranio-Maxillofacial Surg*. 2013; 41(7): 603–609. doi:10.1016/j.jcms.2012.11. 041.

5. Huotilainen E, Jaanimets R, Valášek J, Marcián P, Salmi M, Tuomi J, *et al*. Inaccuracies in additive manufactured medical skull models caused by the DICOM to STL conversion process. *Journal of Cranio-Maxillofacial Surgery*. 2014; 42(5): 259-265. doi: 10.1016/j.jcms.2013.10.001.

6. M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto and J. Wolff. CT image segmentation methods for bone used in medical additive manufacturing. *Med. Eng. Phys.* 2017; 51: 6–16. doi:10.1016/j.medengphy.2017.10.008.

7. van Eijnatten M, Koivisto J, Karhu K, Forouzanfar T, Wolff J. The impact of manual threshold selection in medical additive manufacturing. *Int. J. Comput. Assist. Radiol. Surg.* 2016; 12(4): 607–615. doi:10.1007/s11548-016-1490-4.

8. Sharma N, Aggarwal LM. Automated medical image segmentation techniques. *J. Med. Phys.* 2010; 35(1): 3-14. doi: 10.4103/ 0971-6203.58777.

9. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst*. 2012: 1097-1105

10. Kaur J, Agrawal S, Vig R. A comparative analysis of thresholding and edge detection segmentation techniques. *Int. J. Comput. Appl.* 2012; 39(15): 29–34. doi: 10.5120/4898-7432.

11. Fripp J, Crozier S, Warfield SK, Ourselin S. Automatic segmentation of articular cartilage in magnetic resonance images of the knee. *Medical Image Computing and Computer-assisted Intervention – MICCAI*. 2007; 10(2): 186–194. doi:10.1007/978-3-540-75759-7_23.

12. Grau V, Mewes AUJ, Alcaniz M, Kikinis R, Warfield SK. Improved watershed transform for medical image segmentation using prior information. *IEEE Trans. Med. Imag.* 2004; 23(4): 447–458. doi:10.1109/TMI.2004.824224.

13. Li C, Huang R, Ding Z, Gatenby JC, Metaxas DN, Gore JC. A level set method for image segmentation in the presence of intensity in-homogeneities with application to MRI. *IEEE Trans. Image Process.* 2011; 20(7): 2007–2016. doi:10.1109/TIP.2011.2146190.

14. Jafarian N, Kazemi K, Moghaddam HA, Grebe R, Fournier M, Helfroush MS, *et al.* Automatic segmentation of newborns' skull and fontanel from CT data using model-based variational level set. *Signal Image Video Process.* 2014; 8: 377–387. doi:10.1007/s11760-012-0300-x.

15. Lindner C, Thiagarajah S, Wilkinson JM, arcOGEN Consortium, Wallis GA, Cootes TF. Fully automatic segmentation of the proximal femur using random forest regression voting. *IEEE Trans. Med. Imag.* 2013; 32(8): 1462–1472. doi:10.1109/ TMI.2013.2258030.

16. Baka N, Leenstra S, van Walsum T. Random forest-based bone segmentation in ultrasound. *Ultrasound Med. Biol*. 2017; 43(10): 2426–2437. doi:10.1016/ j.ultrasmedbio.2017.04.022.

17. Schmid J, Kim J, Magnenat-Thalmann N. Robust statistical shape models for MRI bone segmentation in presence of small field of view. *Med. Image Anal*. 2011; 15(1): 155–168. doi:10.1016/j.media.2010.09.001.

18. Lamecker H, Seebass M, Hege HC, Deuflhard P. A 3D Statistical Shape Model of the Pelvic Bone for Segmentation. *Proc. SPIE 5370*. 2004. doi:10.1117/12.534145.

19. Baldwin MA, Langenderfer JE, Rullkoetter PJ, Laz PJ. Development of subject-specific and statistical shape models of the knee using an efficient segmentation and mesh-morphing approach. *Comput. Methods Progr. Biomed*. 2010; 97(3): 232–240. doi:10.1016/j.cmpb.2009.07.005.

20. Steger S, Kirschner M, Wesarg S. Articulated atlas for segmentation of the skeleton from head & neck CT datasets. *Proceedings - International*

Symposium on Biomedical Imaging. 2012: 1256–1259. doi:10.1109/ISBI.2012. 6235790.

21. Powell KA, Liang T, Hittle B, Stredney D, Kerwin T, Wiet GJ. Atlas-based segmentation of temporal bone anatomy. *Int. J. Comput. Assist. Radiol. Surg.* 2017; 12(11): 1937–1944, doi:10.1007/s11548-017-1658-6.

22. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging.* 2016; 35(5): 1153-1159. doi: 10.1109/TMI.2016.2553401.

23. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Lecture Notes in Computer Science.* 2013: 246–253. doi:10.1007/ 978-3-642-40763-5_31.

24. Zhang W, Li R, Deng H, Wang L, Lin W, Ji S, Shen D. Deep convolutional neural networks for multi-modality isointense infant brain image segmentation. *Neuroimage.* 2015;108: 214–224. doi:10.1016/j.neuroimage.2014.12. 061.

25. de Brébisson A, Montana G. Deep neural networks for anatomical brain segmentation. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops.* 2015: 20–28. doi:10.1109/CVPRW.2015. 7301312.

26. Moeskops P, Viergever MA, Mendrik AM, de Vries LS, Benders MJNL, Išgum I. Automatic segmentation of MR brain images with a convolutional neural network. *IEEE Trans. Med. Imag.* 2016; 35(5): 1252–1261, doi:10.1109/ TMI.2016.2548501.

27. Milletari F, Navab N, Ahmadi SA. V-Net, Fully convolutional neural networks for volumetric medical image segmentation. Fourth International Conference on 3D Vision (3DV). 2016:565-571.

28. Deniz CM, Xiang S, Spencer Hallyburton R, Welbeck A, Babb JS, Honig S, *et al.* Segmentation of the Proximal Femur from MR Images Using Deep Convolutional Neural Networks. *Scientific Reports.* 2018; 8: 164585

29. Liu F, Zhou Z, Jang H, Samsonov A, Gengyan Z, Kijowski R. Deep convolutional neural network and 3D deformable approach for tissue segmentation in musculoskeletal magnetic resonance imaging. *Magn. Reson. Med.* 2017; 79(4): 2379-2391. doi:10.1002/mrm.26841.

30. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, *et al.* Brain tumor segmentation with deep neural networks.

Med. Image Anal. 2017; 35: 18–31. doi:10.1016/j.media.2016.05.004

31. Vivanti R, Ephrat A, Joskowicz L, Lev-Cohain N, Karaaslan OA, Sosna J. Automatic liver tumor segmentation in follow-up CT scans. *Proc. Patch-based Methods in Medical Image Processing Workshop.* 2015: 54-61

32. Elamri C. A New Algorithm for Fully Automatic Brain Tumor Segmentation with 3-DConvolutional Neural Networks. 2016.

33. Charron O, Lallement A, Jarnet D, Noblet V, Clavier JB, Meyer P. Automatic detection and segmentation of brain metastases on multimodal MR images with a deep convolutional neural network. *Comput. Biol. Med.* 2018; 95: 43–54. doi:10.1016/j.compbiomed.2018.02.004.

34. Thong W, Kadoury S, Piché N, Pal CJ. Convolutional networks for kidney segmentation in contrast-enhanced CT scans. *Comput. Methods Biomech. Biomed. Eng.: Imag. Visual.* 2018; 6(3): 277–282. doi:10.1080/21681163.2016. 1148636.

35. Farag A, Lu L, Roth HR, Liu J, Turkbey E, Summers RM. A bottom-up approach for pancreas segmentation using cascaded superpixels and (deep) image patch labeling. *IEEE Trans. Image Process.* 2017; 26(1): 386–399. doi:10.1109/ TIP.2016.2624198.

36. Roth HR, Lu L, Farag A, Shin HC, Liu J, Turkbey EB, et al. DeepOrgan: multi-level deep convolutional networks for automated pancreas seg-mentation. *Medical Image Computing and Computer-assisted Intervention – MICCAI.* 2015: 556–564. doi:10.1007/ 978-3-319-24553-9_68.

37. Cai J, Lu L, Xie Y, Xing F, Yang L. Pancreas segmentation in MRI using graph-based decision fusion on convolutional neural networks. *Med Image Comput Comput Assist Interv.* 2017; 9901: 442-450.

38. Vania M, Mureja D, Lee D. Automatic spine segmentation using convolutional neural network via redundant generation of class labels for 3D spine modeling. *J Comp Design and Eng.* 2019; 6(2): 224-232.

39. Lessmann N, van Ginneken B, Išgum I. Iterative Convolutional Neural Networks for Automatic Vertebra Identification and Segmentation in CT Images. *Proc SPIE 10574.* 2018: 1057408. doi:10.1117/12. 2292731.

2

40. Budin F, Oguz I. MeshToLabelMap. *Github. 2017.* https://github.com/NIRALUser/MeshToLabelMap

41. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, *et al*. 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imag.* 2012; 30(9): 1323–1341. doi:10. 1016/j.mri.2012.05.001.

42. 3D Slicer. 2018. http://www.slicer.org.

43. Aldenborgh N. brain_segmentation. 2016. https://github.com/naldeborgh7575/brain_ segmentation.

44. Minnema J, Kouw W, Diblen F. CNN for bone segmentation. *Github.* 2018 doi:10.5281/zenodo.1164605.

45. Ioffe S, Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *Proc. ICML.* 2015: 448-456.

46. Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. *Lecture Notes in Computer Science.* 2010: 92–101, , doi:10.1007/978-3-642-15825-4_10.

47. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 2014; 15(56): 1929–1958. doi:10.1214/12-AOS1000.

48. Hinton GE, Srivastava N, Swersky K. Lecture 6a- Overview of Mini-batch Gradient Descent, COURSERA: Neural Networks for Machine Learning. 2012: 31.

49. Chollet F. Keras. *GitHub.* 2015. https://github.com/fchollet/keras

50. Fu Y, Liu S, Li HH, Yang D. Automatic and hierarchical segmentation of the human skeleton in CT images. *Phys. Med. Biol.* 2017; 62(7): 2812–2833. doi:10.1088/1361-6560/aa6055.

51. Torosdagli N, Liberton DK, Verma P, Sincan M, Lee J, Pattanaik S, *et al*. Robust and fully automated segmentation of mandible from CT scans. *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). 2017*: 1209–1212. doi:10.1109/ISBI.2017.7950734.

52. Ghadimi S, Abrishami Moghaddam H, Grebe R, Wallois F. Skull segmentation and reconstruction from newborn CT images using coupled level sets. *IEEE J. Biomed. Health Inform.* 2016; 20(2): 563–573. doi:10.1109/JBHI.2015.2391991.

53. Provost F, Weiss GM. Learning when Training Data Are Costly: the Effect of Class Distribution on Tree Induction. *J Art Intel Res.* 2011;19: 315-354

54. He H, Garcia EA. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* 2009; 21(9): 1263–1284. doi:10.1109/TKDE.2008.239.

55. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter - Special Issue on Learning from Imbalanced Datasets.* 2004; 6: 20–29. doi:10.1145/1007730.1007735.

56. Rueda S, Gil JA, Pichery R, Alcañiz M. Automatic segmentation of jaw tissues in CT using active appearance models and semi-automatic landmarking. *Medical Image Computing and Computer-Assisted Intervention: MICCAI.* 2006; 1: 167–174. doi:10.1007/11866565_21.

57. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. *CVPR.* 2015: 3431-3440

58. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention: MICCAI.* 2015: 234–241. doi:10.1007/978-3-319-24574-4_28.

59. Bernal J, Kushibar K, Cabezas M, Valverde S, Oliver A, Lladó X. Quantitative analysis of patch-based fully convolutional neural networks for tissue segmentation on brain magnetic resonance imaging. 2019;7: 89986-90002. doi: 10.1109/ACCESS.2019.2926697.

60. Kayalibay B, Jensen G, van der Smagt P. CNN-based segmentation of medical imaging data. 2017. ArXiv:1701.03056 [Cs]

61. Ciresan D, Giusti A, Gambardella LM, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in Neural Information Processing Systems.* 2012;25: 2843–2851

62. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. *IEEE Trans. Med. Imag.* 2016;35(5): 1240–1251. doi:10.1109/TMI.2016.2538465.

63. Wachinger C, Reuter M, Klein T. DeepNAT: deep convolutional neural network for segmenting neuroanatomy. *Neuroimage. 2018;* 170: 434–445. doi:10. 1016/j.neuroimage.2017.02.035.

64. Trebeschi S, van Griethuysen JJM, Lambregts DMJ, Lahaye MJ, Parmar C, Bakers FCH, *et al*. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric MR. *Sci. Rep*. 2017; 7: 5301. doi:10.1038/s41598-017-05728-9.

65. Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. CVPR. 2016: 2424–2433. doi:10.1109/CVPR.2016.266.

66. Qiu B, Guo J, Kraeima J, Borra R, Witjes M, van Ooijen P. 3D Segmentation of Mandible from Multisectional CT Scans by Convolutional Neural Networks. OpenReview.Net. 2018.

67. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci. Unit. States Am*. 2018; 115: 254–259. doi:10.1073/pnas.1715832114.

2

CHAPTER 3

MULTICLASS CBCT IMAGE SEGMENTATION FOR ORTHODONTICS WITH DEEP LEARNING

Hong Wang, Jordi Minnema, Kees Joost Batenburg,
Tymour Forouzanfar, Feng Jun Hu, Gang Wu

3

ABSTRACT

Accurate segmentation of the jaw (i.e., mandible and maxilla) and the teeth in cone-beam computed tomography (CBCT) scans is essential for orthodontic diagnosis and treatment planning. Although various (semi)automated methods have been proposed to segment the jaw or the teeth, there is still a lack of fully automated segmentation methods that can simultaneously segment both anatomic structures in CBCT scans (i.e., multiclass segmentation). In this study, we aimed to train and validate a mixed-scale dense (MS-D) convolutional neural network for multiclass segmentation of the jaw, the teeth, and the background in CBCT scans. Thirty CBCT scans were obtained from patients who had undergone orthodontic treatment. Gold standard segmentation labels were manually created by 4 dentists. As a benchmark, we also evaluated MS-D networks that segmented the jaw or the teeth (i.e., binary segmentation). All segmented CBCT scans were converted to virtual 3-dimensional (3D) models. The segmentation performance of all trained MS-D networks was assessed by the Dice similarity coefficient and surface deviation. The CBCT scans segmented by the MS-D network demonstrated a large overlap with the gold standard segmentations (Dice similarity coefficient: 0.934 ± 0.019, jaw; 0.945 ± 0.021, teeth). The MS-D network based 3D models of the jaw and the teeth showed minor surface deviations when compared with the corresponding gold standard 3D models (0.390 ± 0.093 mm, jaw; 0.204 ± 0.061 mm, teeth). The MS-D network took approximately 25 s to segment 1 CBCT scan, whereas manual segmentation took about 5 h. This study showed that multiclass segmentation of jaw and teeth was accurate and its performance was comparable to binary segmentation. The MS-D network trained for multiclass segmentation would therefore make patient-specific orthodontic treatment more feasible by strongly reducing the time required to segment multiple anatomic structures in CBCT scans.

1.   INTRODUCTION

Cone-beam computed tomography (CBCT) is being increasingly used in orthodontics because of its low cost and low radiation dose [1]. 3-dimensional (3D) information embedded in CBCT scans allows orthodontists to accurately assess complex dental and skeletal malocclusions, which helps to substantially improve diagnosis and treatment planning [2, 3]. An essential step in the diagnosis of dentofacial abnormalities and deformities is the conversion of CBCT scans into virtual 3D models of relevant anatomic regions of interest, such as the mandible, the maxilla, and the teeth. Orthodontic changes can be visually and quantitatively assessed by superimposing 3D models of a patient at different time points. Moreover, these 3D models can be used to simulate stress distribution in alveolar bone via finite element analysis [4].

Currently, the most challenging step in creating 3D models for orthodontics is CBCT image segmentation—that is, the partitioning of CBCT scans into various anatomic regions of interest. For example, it is laborious to segment the teeth due to the similar intensity of tooth roots and their surrounding alveolar bone. In addition, because of high noise levels, limited image resolution, and cone beam artifacts [5], it is difficult to accurately segment bony structures—for example, the condyle and the ramus. Consequently, bony structures are often erroneously labeled, which leads to cavities and gaps when converted into 3D models, subsequently compromising the quality of the treatment planning and finite element analysis.

In the last decades, several methods have been proposed to (semi)automatically segment various anatomic structures in CBCT scans. Such automatic approaches include edge detection, watershed segmentation, region seed growing, clustering methods, morphologic snakes, random forests, and statistical shape models [6 – 8]. Although these methods are capable of segmenting CBCT scans, accurate segmentation of the mandible, the maxilla, and the teeth remains challenging. Therefore, laborious manual correction remains necessary in clinical settings to achieve accurate segmentation. Hence, new methods for automatic image segmentation are sought.

Deep convolutional neural networks (CNNs) have recently become increasingly utilized in medical image segmentation [9, 10] and have achieved state-of-the-art performances [11 – 13]. The success of CNNs can be mainly attributed to their ability of learning nonlinear spatial features in input images. Several research groups have used CNNs to segment the mandible or the teeth (i.e., binary segmentation) and demonstrated that CNNs were able to perform accurate segmentation tasks [14 – 17]. However, no studies have been published in which CNNs were applied to simultaneously segment the jaw (i.e., mandible and maxilla) and the teeth in CBCT scans, also known as multiclass segmentation. As compared with binary segmentation, multiclass segmentation approaches require training only a single network to segment jaw and teeth, thus reducing the overall training time. Furthermore, multiclass segmentation does not suffer from conflicting segmentation labels. Such conflicting labels are caused when 1 binary segmentation network classifies a pixel as jaw and the other classifies it as teeth.

A novel CNN architecture, namely the mixed-scale dense (MS-D) CNN, has recently shown promising segmentation performances [18, 19]. This MS-D network allows for accurate and automatic segmentation of different bony structures. To reduce the time-consuming and costly

manual labor required to create 3D models for patient-specific orthodontic treatment, our aim was to train the MS-D network to simultaneously segment the jaw and the teeth in CBCT scans.

2. MATERIALS AND METHODS

CBCT scan information, CNN architecture, implementation and training details, and CNN performance evaluation are provided in the supplemental materials.

2.1. Data Acquisition and Preprocessing

Thirty dental CBCT scans were acquired from patients (age, 11 to 24 y; mean ± SD, 14.2 ± 3.4 y; 19 females and 11 males) who previously underwent orthodontic treatment at the Shanghai Xuhui Dental Center. The CBCT scans used in this study were obtained before the orthodontic treatment, and none of the patients had fillings, dental implants, or crowns. Thus, the CBCT scans were free of metal artifacts. Moreover, no patients had missing teeth, and wisdom teeth were not yet erupted in most patients (n = 22). Informed consent was signed by each patient and at least 1 parent. The use of patient data was approved by the medical ethics committee of the Shanghai Xuhui Dental Center (No. 20193).

Since this study focused on segmenting the jaw and the teeth, we cropped all CBCT scans to those anatomic regions, resulting in scans with axial dimensions ranging between 255 to 384. In total, 9507 slices were obtained from the 30 CBCT scans.

To acquire gold standard labels, all 30 CBCT scans were segmented into 3 classes: jaw, teeth, and background. The manual segmentation was carried out by 4 dentists with at least 2 y of working experience in dental clinics. The 4 dentists were well instructed and practiced extensively until they could accurately annotate jaw and teeth in CBCT scans. After that, the 30 CBCT scans were distributed among the 4 annotators, and each CBCT scan was segmented only once by a single annotator. This segmentation was performed with global thresholding, followed by manual correction—that is, removing noise, artifacts, and unrelated parts, as well as adding missing thin bony structures and filling erroneous cavities in the segmented scans through Mimics 21.0 software (Materialise). The resulting segmentation labels were used as the gold standard.

2.2. CNN Architecture

In this study, we employed an MS-D network that was developed by Pelt and Sethian [18]. A schematic overview of an MS-D network with a depth of 3 and a width of 1 is presented in Figure 1A.

2.3. Implementation and Training Details

Three experiments were designed to evaluate the MS-D network's segmentation performance. The first experiment was multiclass segmentation, in which the MS-D network was trained to simultaneously segment 3 labels: jaw, teeth, and background. The second and third experiments were binary segmentation, where the MS-D network segmented jaw or teeth, respectively.

Twenty-eight CBCT scans were divided into 4 subsets (S1, S2, S3, and S4), each containing 7 scans. Each experiment followed a 4-fold cross-validation scheme [20], which means that 3 subsets

Figure 1. MS-D network architecture and 4-fold cross-validation scheme. (**A**) Schematic representation of an MS-D network with 3 convolutional layers and a width of 1; (**B**) 28 CBCT scans were divided into 4 subsets (S1, S2, S3, and S4), each containing 7 CBCT scans. For each iteration, 3 subsets were used for training and 1 for testing. CBCT, cone beam computed tomography; MS-D, mixed-scale dense.

were used for training and 1 for testing. This process was repeated 4 times such that each CBCT scan was used for testing exactly once (Fig. 1B). The 2 CBCT scans that were not included in the 4-fold cross-validation scheme were used to determine the optimal number of epochs for training.

2.4. CNN Performance Evaluation

The segmentation performance of the MS-D network was evaluated with the Dice similarity coefficient (DSC; [21]). DSCs were calculated on the patient level, which means that a single DSC was calculated for each segmented CBCT volume.

Surface deviations between the MS-D network–based 3D models and the gold standards were calculated to evaluate the accuracy of the MS-D segmentation around the edges of bony structures. Additionally, mean absolute deviations (MADs) were calculated between the MS-D network–based 3D models and the gold standards.

After the 4 iterations of the cross-validation scheme, the performance of the MS-D network was averaged over the 28 CBCT scans.

3. RESULTS

The multiclass and binary segmentation approaches achieved similar segmentation accuracies (Fig. 2A, B). The former approach resulted in DSCs of jaw between 0.901 (patient 3) and 0.968 (patient 28), with a mean of 0.934 ± 0.019. The DSCs of teeth ranged between 0.881 (patient 2) and 0.971 (patient 28), with a mean of 0.945 ± 0.021. For the binary segmentation, the DSCs of jaw ranged from 0.892 (patient 3) to 0.966 (patient 28), with a mean of 0.933 ± 0.020. The DSCs of teeth ranged from 0.889 (patient 2) to 0.973 (patient 28), with a mean of 0.948 ± 0.021. The lowest DSC of jaw from patient 3 was due to the larger excluded region of maxilla in its gold standard while this excluded region was segmented by the MS-D (Fig. 2C). The lowest DSC of teeth from patient 2 was attributed to the unerupted teeth not being included in the gold standard while the MS-D segmented these teeth (Fig. 2D).

Examples of the multiclass segmentation of the CBCT scan from patient 9 are presented in Figure 3. Five axial CBCT slices representing different skull anatomies were selected. The difference maps show that the errors mainly occurred at the edges with deviations around 1 pixel (Fig. 3A, B). Some thin bony structures around the maxillary sinus were not segmented by the MS-D network as compared with the gold standard (Fig. 3E).

Figure 4A shows the surface deviations of all 3D jaw models obtained from the multiclass and binary segmentations. Figure 4B illustrates 3 jaw models from the multiclass segmentation. Patient 28 and 12 corresponded to the minimum and maximum MADs, respectively. Patient 25 had an MAD close to the mean MAD. All MAD values of jaw models are presented in Table S1 (supplemental material). When analyzing jaw models, the multiclass segmentation resulted in surface deviations from −0.191 ± 1.095 mm (patient 14) to 0.185 ± 1.011 mm (patient 3) and a mean MAD of 0.390 ± 0.093 mm. The binary segmentation resulted in surface deviations from −0.180 ± 1.069 mm (patient 14) to 0.252 ± 1.058 mm (patient 3) and a mean MAD of 0.410 ± 0.103 mm.

Figure 5A shows the surface deviations of all 3D teeth models obtained from the multiclass and binary segmentations. Figure 5B presents 3 teeth models from the multiclass segmentation. Patients 28 and 23 corresponded to the minimum and maximum MADs, respectively. Patient 14 had an MAD close to the mean MAD. All MAD values of teeth models are presented in Table S1 (supplemental material). When analyzing teeth models, the multiclass segmentation resulted in surface deviations from −0.107 ± 0.546 mm (patient 5) to 0.318 ± 0.347 mm (patient 23) and a mean MAD of 0.204 ± 0.061 mm. The binary segmentation resulted in surface deviations from −0.116 ± 0.534 mm (patient 12) to 0.290 ± 0.272 mm (patient 23) and a mean MAD of 0.163 ± 0.051 mm.

4. DISCUSSION

CBCT is increasingly utilized to create virtual 3D models for quantitative evaluation of orthodontic changes such as tooth resorption, condyle growth, and movement of the chin and teeth. Creating these 3D models requires accurate segmentation of jaw (i.e., mandible and maxilla) and teeth. However, manually segmenting these 2 anatomies is time-consuming, laborious, and expensive. In this study, we trained a novel MS-D network to simultaneously segment jaw and teeth in CBCT scans (i.e., multiclass segmentation). The jaw and teeth segmented by the MS-D network demonstrated
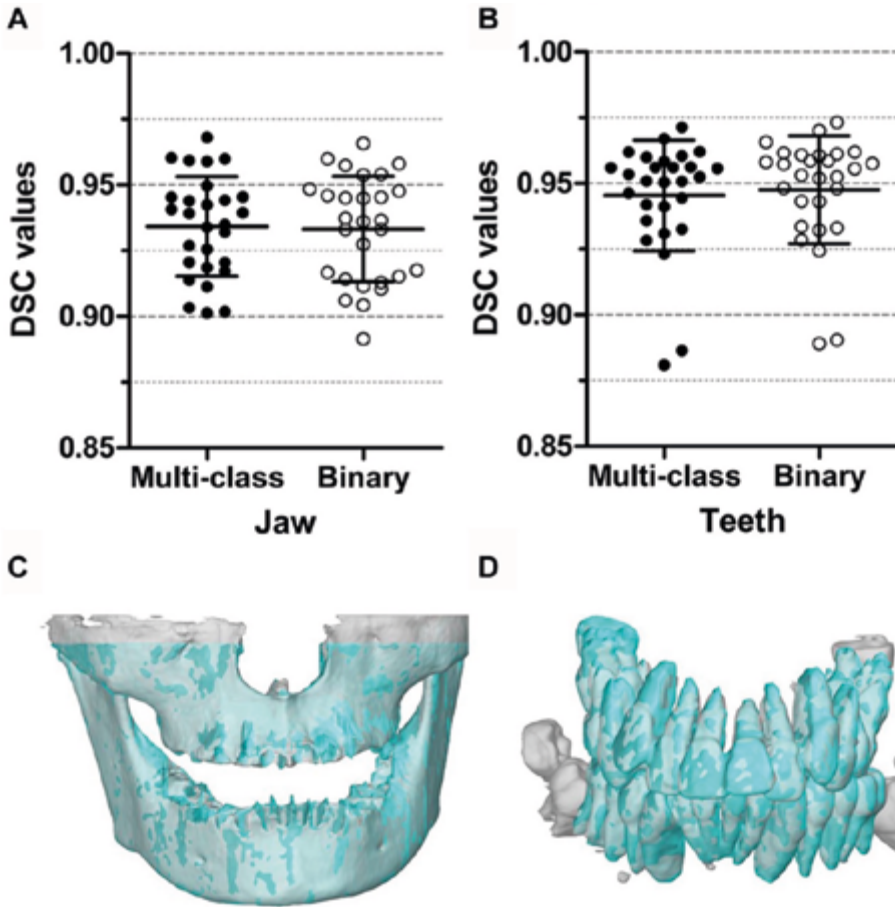
Figure 2. Comparison of DSCs obtained using the multiclass and binary segmentation approaches. For jaw segmentation (**A**) and teeth segmentation (**B**), the multiclass segmentation and the binary segmentation were equivalent. In addition, 3D models of the jaw (**C**) and the teeth (**D**) are presented, which were acquired by using the CBCT scans that resulted in the lowest DSCs. 3D, 3-dimensional; CBCT, cone beam computed tomography; DSC, Dice similarity coefficient.

high DSCs, and their 3D models showed minor surface deviations compared with the gold standard. The MS-D network took approximately 25 s to segment jaw and teeth in 1 CBCT scan, thus markedly reducing the time required for segmentation. Therefore, the MS-D network trained for multiclass segmentation has a promising potential to accurately and automatically segment multiple anatomies of interest for orthodontic diagnosis and treatment.

Multiclass segmentation has been considered challenging since it faces class data imbalance and interclass feature similarity problems [22 - 24]. Compared with multiclass strategies, binary strategies are generally more robust and achieve higher accuracy but come at the cost of increased training time [25, 26]. In this study, the MS-D network trained for multiclass segmentation was able to accurately segment jaw and teeth in CBCT scans, achieving comparable accuracy as binary
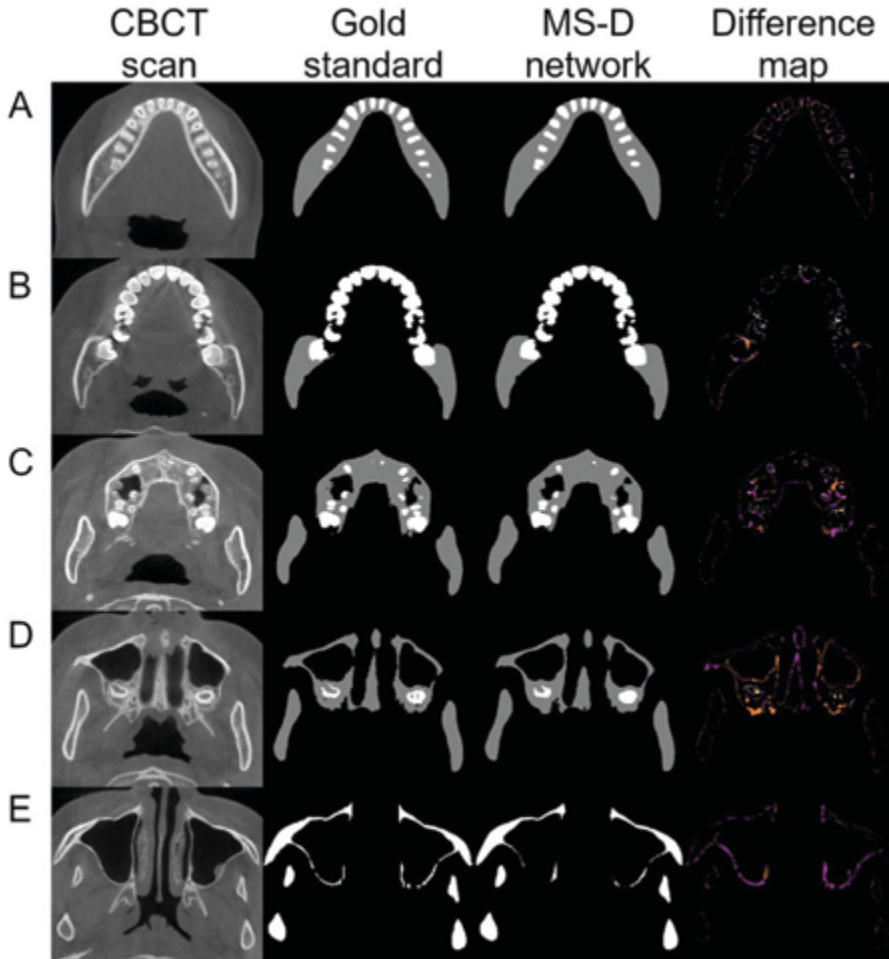
3

Figure 3. Example of jaw and teeth segmentation from 5 axial CBCT slices of patient 9 with the multiclass segmentation approach. The first column contains 5 axial CBCT slices, which represent different skull anatomies. The second and third columns show the gold standard segmentation and the MS-D segmentation, respectively. The last column indicates the difference between the gold standard and the MS-D segmentations. False negatives of the jaw are marked in fuchsia, and false positives are marked in ultra-pink and gamboge. False negatives of the teeth are marked in pink, and false positives are marked in wheat and yellow. CBCT, cone beam computed tomography; MS-D, mixed-scale dense.

segmentation. This indicates that the MS-D network can be trained with 3 classes without losing segmentation accuracy compared to binary segmentation. Moreover, multiclass segmentation has two important advantages over binary segmentation. The first is that multiclass segmentation requires training only a single CNN for segmentation of jaw and teeth, which was twice as fast as training two CNNs needed for binary segmentation. Specifically, training an MS-D network took about 20 h (1 h per epoch), and segmentation of 1 CBCT scan took approximately 25 s. Nevertheless,
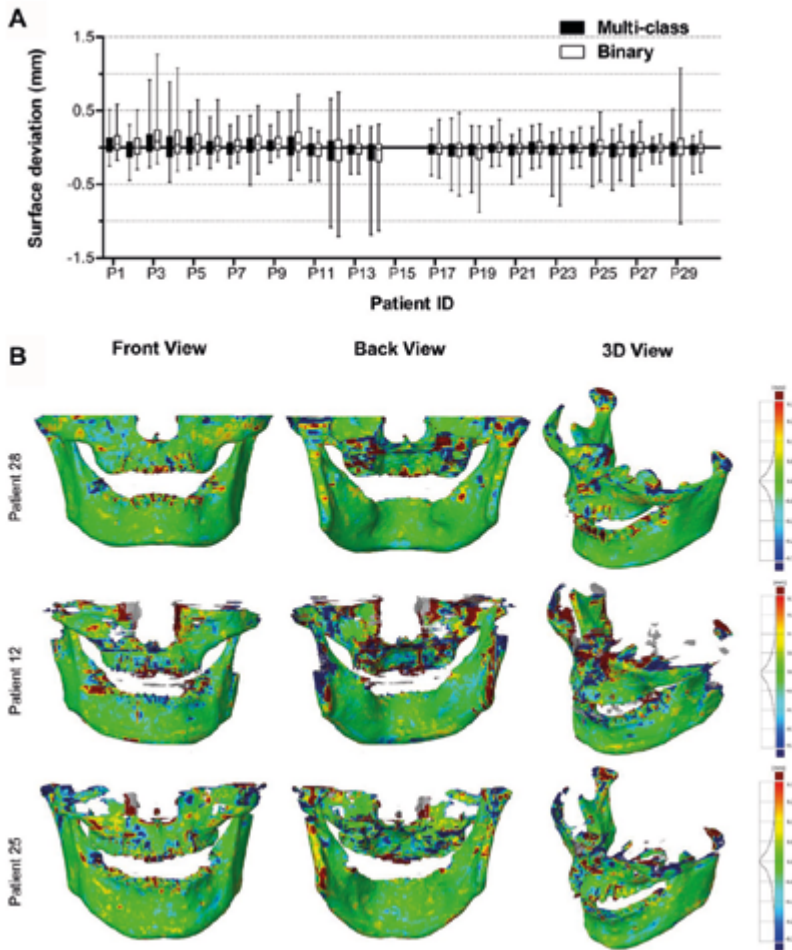
3

Figure 4. Surface deviations between the MS-D network–based 3D models of the jaw and the corresponding gold standard 3D models. The CBCT scans from patients 15 and 16 were used for validation and therefore not included in the analysis. (**A**) Box and whisker plot of the surface deviations. The boxes represent the interquartile range, and the whiskers represent the 10th and 90th percentiles of the surface deviations. (**B**) The front, back, and isotropic views of color maps of surface deviations are presented for 3 jaw models. Patients 28 and 12 corresponded to the minimum and maximum MADs, respectively. Patient 25 had an MAD close to the mean MAD. 3D, 3-dimensional; CBCT, cone beam computed tomography; MAD, mean absolute deviation.

it must be noted that the segmentation time of both the multiclass and the binary segmentation approaches is still significantly less than that of manual segmentation, which took around 5 h per CBCT scan. The second advantage is that multiclass segmentation does not generate conflicting labels, as opposed to binary segmentation. These conflicting labels are caused when 1 binary segmentation network classifies a pixel as jaw and the other classifies it as teeth (Figure S1).

The MS-D networks trained in this study resulted in DSCs that are comparable to those presented in literature. For mandible segmentation, Qiu et al. [15] obtained a mean DSC of 0.896 by training three
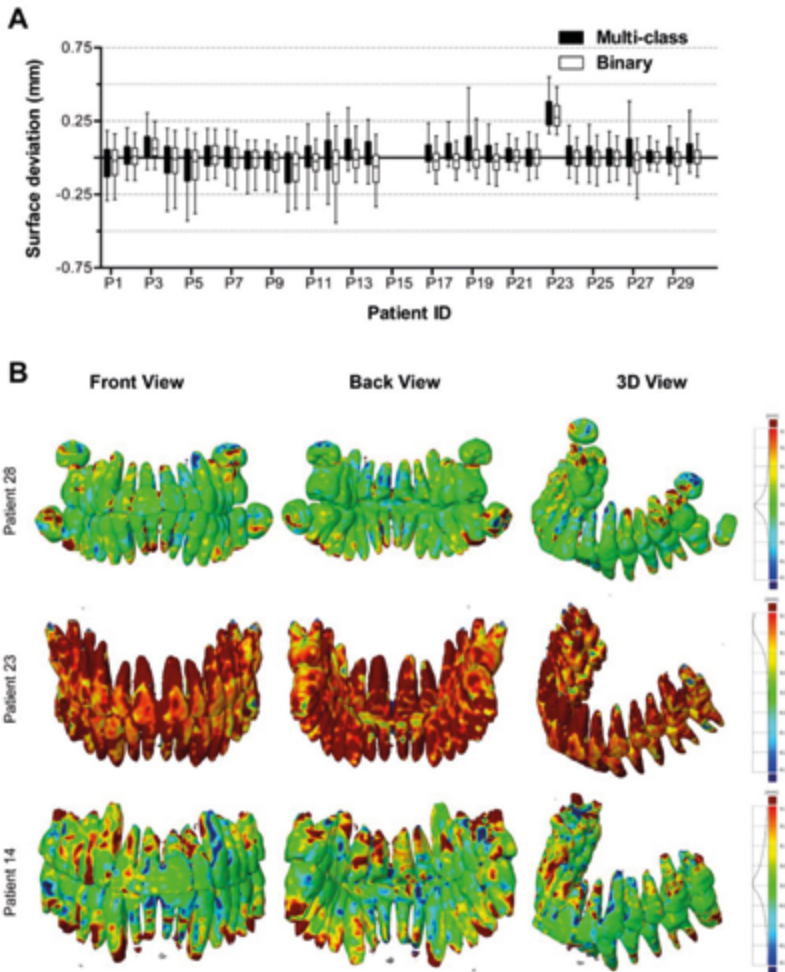
Figure 5. Surface deviations between the MS-D network–based 3D models of the teeth and the corresponding gold standard models. The CBCT scans from patients 15 and 16 were used for validation and therefore not included in the analysis. (A) Box and whisker plot of the surface deviations. The boxes represent the interquartile range, and the whiskers represent the 10th and 90th percentiles of the surface deviations. (B) The front, back, and isotropic views of color maps of surface deviations are presented for 3 teeth models. Patients 28 and 23 corresponded to the minimum and maximum MADs, respectively. Patient 14 had an MAD close to the mean MAD. 3D, 3-dimensional; CBCT, cone beam computed tomography; MAD, mean absolute deviation; MS-D, mixed-scale dense.

CNNs using CBCT slices from axial, sagittal, or coronal planes and then combining the segmentation results from all three CNNs. For maxilla segmentation, a lower mean DSC of 0.800 ± 0.029 was found by S. Chen et al. [27], who used a learning-based multisource integration framework. For teeth segmentation, Lee et al. [17] applied a multiphase strategy to train a U-Net based architecture, which resulted in DSCs ranging from 0.910 to 0.918. Furthermore, Cui et al. [16] employed a 2-stage network consisting of a tooth edge map extraction network and a region proposal network and

reported a mean DSC of 0.926. In comparison with the aforementioned studies, the MS-D network employed in this study achieved comparable DSCs. However, the reviewed studies used different data sets to evaluate their methods, which means that differences among the DSCs should be interpreted with caution.

3

All MS-D network–based 3D models closely resembled the corresponding gold standard 3D models. The surface deviations found in our study were generally lower than those obtained by Wang et al. [28], who developed a random forest method to segment the mandible and the teeth in CBCT scans. When segmenting the mandible, their method resulted in a surface deviation of 0.420 ± 0.150 mm, whereas the segmentation of the upper and lower teeth resulted in surface deviations of 0.312 ± 0.103 mm and 0.346 ± 0.154 mm, respectively. Our MS-D network segmentation also resulted in lower surface deviations of teeth than the multitask fully CNN developed by Y. Chen et al. [29]. Their network was trained for individual tooth segmentation, which resulted in an average surface deviation of 0.363 ± 0.145 mm.

The segmentation errors of the MS-D network trained in this study mainly occurred at the edges of the bony structures (Fig. 3). These segmentation errors are likely due to the partial volume effect. This occurs when tissues with different densities are encompassed within the same voxel, which is typically the case at the border of 2 anatomic regions (e.g., bone and soft tissue). As a consequence, it is extremely hard to exactly define the edge regions of bony structures. This phenomenon might also explain why some thin bony structures in the maxillae were not correctly segmented by the MS-D network (Fig. 3E). The quality of CBCT scans also affects the accuracy of segmentation. For example, the ramus and the condyle were poorly segmented in the CBCT scan of patient 12 (Fig. 4B) because these regions were affected by movement artifacts.

One challenge for deep learning in medicine and dentistry is to obtain an accurate gold standard [30]. The gold standard is usually created by human annotators, which contains intrinsic inter- and intraobserver variability. However, deep learning can learn from large training data sets and therefore is able to overcome this variability [31]. In this study, the gold standard segmentation labels were annotated by 4 dentists, which introduced subjective variability in the gold standard. For example, the unerupted teeth from 1 patient were not included in the gold standard. However, since the MS-D network was able to learn from all other segmented CBCT images, the unerupted teeth in the CBCT scans were still correctly segmented by the MS-D network. These findings indicate that the MS-D network can reduce the influence of subjective variability. If the inaccurate gold standard labels are included in the test set, this can affect the evaluation of the network, particularly for small data sets. However, since the performance of the MS-D network was evaluated on the 28 CBCT scans, the influence of 1 slightly inaccurate gold standard CT scan in the test set on the performance metrics is minimal.

In the present study, the MS-D network was adopted to evaluate the multiclass segmentation performance. This network was chosen because it has relatively few parameters, making it easier to train and apply than other CNNs [18]. The multiclass segmentation performance of the MS-D network was compared with that of the U-Net, demonstrating that it was able to achieve similar segmentation accuracy as the U-Net [18]. Nevertheless, the MS-D network is not the only CNN architecture capable of performing multiclass segmentation. Several other CNN architectures

3

have been implemented to perform multiclass segmentation of anatomies in brain [22, 24] and lung [23, 32].

In this study, all CBCT scans were obtained from patients without dental fillings, implants, or orthodontic devices to avoid the influence of metal artifacts. In daily clinical practice, there may be such artifacts, so caution should be taken to apply the current findings in clinical practice. Further studies should be performed for CBCT image segmentation with complicated dental status.

To facilitate CNN training, the maxilla and mandible were considered a single class, and the upper and lower teeth were considered another class. However, one may wish to automatically separate the mandible from the maxilla and classify individual teeth during segmentation. The mandible and the maxilla can be easily separated by using region growing methods available with most image processing software packages, but individual teeth segmentation still requires postprocessing. To make the automatic segmentation of individual teeth possible, we aim to include individual labels of different teeth during the training of the MS-D network in future work.

5. CONCLUSION

This study applied a novel MS-D network to segment CBCT scans into jaw, teeth, and background. Multiclass segmentation achieved comparable segmentation accuracy as binary segmentation. In addition, the MS-D network–based 3D models closely resembled the gold standard 3D models. These results demonstrate that deep learning has the potential to accurately and simultaneously segment jaw and teeth in CBCT scans. This will substantially reduce the amount of time and effort spent in clinical settings, thereby facilitating patient-specific orthodontic treatment.

AUTHOR CONTRIBUTIONS

H. Wang, contributed to conception, design, data acquisition, analysis, and interpretation, drafted and critically revised the manuscript; J. Minnema, contributed to design, data analysis, and interpretation, critically revised the manuscript; K.J. Batenburg, F.J. Hu, contributed to data interpretation, critically revised the manuscript; T. Forouzanfar, contributed to conception, critically revised the manuscript; G. Wu, contributed to conception and data acquisition, critically revised the manuscript. All authors gave final approval and agree to be accountable for all aspects of the work.

ACKNOWLEDGEMENTS

DECLARATION OF CONFLICTING INTERESTS

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

FUNDING

3

REFERENCES

1. Carter JB, Stone JD, Clark RS, Mercer JE. Applications of cone-beam computed tomography in oral and maxillofacial surgery: an overview of published indications and clinical usage in United States academic centers and oral and maxillofacial surgery practices. *J Oral Maxillofac Surg*. 2016; 74(4): 668–679.

2. Kapila SD, Nervina JM. CBCT in orthodontics: assessment of treatment outcomes and indications for its use. *Dentomaxillofac Radiol*. 2015; 44(1): 20140282.

3. Abdelkarim A. Cone-beam computed tomography in orthodontics. *Dent J*. 2019; 7(3): 89.

4. Likitmongkolsakul U, Smithmaitrie P, Samruajbenjakun B, Aksornmuang J. Development and validation of 3D finite element models for prediction of orthodontic tooth movement. *Int J Dent*. 2018: 4927503.

5. Schulze R, Heil U, Gross D, Bruelmann D, Dranischnikow E, Schwanecke U, *et al*. Artefacts in CBCT: a review. *Dentomaxillofacial Radiology* 2011; 40(5): 265-273. doi: 10.1259/dmfr/30642039.

6. Khan MW. A survey: image segmentation techniques. *IJFCC*. 2014; 3(2): 89–93.

7. Mustafa ID, Hassan MA, Mawia A. A comparison between different segmentation techniques used in medical imaging. *Am J Biomed Eng*. 2018; 6(2):59–69.

8. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phys*. 2018; 51: 6–16

9. Litjens G, Kooi T, Bejnordi B, Setio A, Ciompi F, Ghafoorian M, *et al*. A survey *on deep learning in medical image analysis*. *Med. Image. Anal. 2017;42: 60-88.*

10. Altaf F, Islam SM, Akhtar N, Janjua NK. Going deep in medical image analysis: concepts, methods, challenges, and future directions. *IEEE Access*. 2019; 7: 99540–99572.

11. Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput Biol Med*. 2018; 103: 130–139.

12. Casalegno F, Newton T, Daher R, Adbelaziz M, Lodi-Rizzini A, Schürmann F, *et al*. Caries detection with near-infrared transillumination using deep learning. *J Dent Res*. 2019; 98(11): 1227–1233.

13. Nguyen KCT, Duong DQ, Almeida FT, Major PW, Kaipatur NR, Phan TT, *et al*. Alveolar bone segmentation in intraoral ultrasonographs with machine learning. *J Dent Res*. 2020; 99(9): 1054–1061.

14. Egger J, Pfarrkirchner B, Gsaxner C, Lindner L, Schmalstieg D, Wallner J. Fully convolutional mandible segmentation on a valid ground-truth dataset. *Annu Int Conf IEEE Eng Med Biol Soc*. 2018: 656–660.

15. Qiu B, Guo J, Kraeima J, Borra R, Witjes M, van Ooijen P. 3D segmentation of mandible from multisectional CT scans by convolutional neural networks. *arXiv*. 2018. https://arxiv.org/abs/1809.06752

16. Cui Z, Li C, Wang W. ToothNet: automatic tooth instance segmentation and identification from cone beam CT images. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019: 6361–6370

17. Lee S, Woo S, Yu J, Seo J, Lee J, Lee C. Automated CNN-based tooth segmentation in cone-beam CT for dental implant planning. *IEEE Access*. 2020; 8: 50507–50518.

18. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. Proc Natl Acad Sci U S A. 2018; 115(2): 254–259.

19. Minnema J, van Eijnatten M, Hendriksen AA, Liberton N, Pelt DM, Batenburg KJ, *et al*. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med Phys*. 2019; 46(11): 5027–5035.

20. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The "k" in k-fold cross validation. *Paper presented at: ESANN 2012 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; April 25–27, 2012; Bruges, BE,* 2012: 441–446.

21. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, haker SJ, *et al*. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004; 11(2): 178–189

22. Chen X, Liew JH, Xiong W, Chui CK, Ong SH b. Focus, segment and erase: an efficient network for multi-label brain tumor segmentation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer vision— ECCV 2018. Lecture notes in computer science. Vol. 11217. Cham (Switzerland): Springer. 2018: 674–689

3

23. Novikov AA, Lenis D, Major D, Hladůvka J, Wimmer M, Bühler K. Fully convolutional architectures for multi-class segmentation in chest radiographs. *IEEE Trans Med Imaging*. 2018; 37(8): 1865–1876.

24. Jafari M, Li R, Xing Y, Auer D, Francis S, Garibaldi J, Chen X. FU-net: multi-class image segmentation using feedback weighted U-net. *ICIGP*. 2019: 529–537.

25. Berstad TJD, Riegler M, Espeland H, de Lange T, Smedsrud PH, Pogorelov K, et al. Tradeoffs using binary and multiclass neural network classification for medical multidisease detection. *In: 2018 IEEE International Symposium on Multimedia. Washington (DC): IEEE Computer Society*. 2018; 1–8.

26. Gómez O, Mesejo P, Ibáñez O, Valsecchi A, Cordón O. Deep architectures for high-resolution multi-organ chest X-ray image segmentation. *Neural Comput Appl*. 2020; 32: 15949–15963.

27. Chen S, Wang L, Li G, Wu TH, Diachina S, Tejera B, *et al*. Machine learning in orthodontics: introducing a 3D autosegmentation and auto-landmark finder of CBCT images to assess maxillary constriction in unilateral impacted canine patients. *Angle Orthod*. 2020; 90(1): 77–84.

28. Wang L, Gao Y, Shi F, Li G, Chen KC, Tang Z, *et al*. Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Med Phys*. 2016; 43(1): 336.

29. Chen Y, Du H, Yun Z, Yang S, Dai Z, Zhong L, *et al*. Automatic segmentation of individual tooth in dental CBCT images from tooth surface map by a multi-task FCN. *IEEE Access*. 2020; 8: 97296–97309.

30. Schwendicke F, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. *J Dent Res*. 2020; 99(7): 769–774.

31. Naylor CD. On the prospects for a (deep) learning health care system. JAMA. 2018; 320(11): 1099–1100.

32. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: UNET vs. SegNET. *BMC Medical Imaging*. 2021; 21:19

3

SUPPLEMENTAL MATERIAL

CBCT scan information

All CBCT scans were acquired using a NewTom VGi scanner, with a tube voltage of 75 kV - 110 kV and tube current of 1 mA - 32 mA. All CBCT scans contained axial slices of 512 x 512 voxels with a size of 0.3 mm.

CNN architecture

In this study we employed a MS-D network that was originally developed by Pelt and Sethian [1]. This MS-D network uses dilated convolutional filters to capture relevant patterns at different image scales. In addition, all layers of the MS-Dnetwork are densely connected, which means that relevant patterns can be directly passed to deeper layers in the network. As a result, the MS-D network consists of far fewer trainable parameters than alternative CNN architectures such as U-Net or ResNet. This reduces the risk of overfitting on the training data [1], without suffering from lower segmentation performances [2]. Moreover, the MS-D network has demonstrated strong performance in improving the quality of tomographic data [3]. A schematic overview of an MS-D network with a depth of 3 and a width of 1 is presented in Figure 1A. A detailed description of the MS-Dnetwork can be found in [1].

Implementation and training details

The depth and the dilation factors of the network were adopted from the study by Minnema et al. [2]. Specifically, the depth was 100, and the dilation factor was 1 for the first convolutional layer and increased by 1 for each subsequent layer. After 10 layers, the dilation factor was reset to 1, and the same scheme was applied. The network width was chosen as 1.

Three different experiments were designed to evaluate the MS-D network's segmentation performance. The first experiment was multi-class segmentation, in which the MS-D network was trained to simultaneously segment 3 labels: (1) jaw, (2) teeth, and (3) background. The second and third experiments were binary segmentation, where the MS-D network segmented either jaw, or teeth, respectively.

In this study, training the MS-D network was performed following a modified version of the k-fold cross-validation scheme. The k-fold cross-validation is typically used to tune the hyper-parameters of the network [4]. In the standard procedure, the data sets are split into k folds. The data in k-1 folds are used for training and 1 remaining fold is used for validation. This process is repeated until all folds are used exactly once as validation set. The hyper-parameters of the network can then be chosen based on the highest possible performance on the validation set. The model is subsequently tested on an independent hold-out test set.

However, this typical k-fold cross-validation scheme can lead to unreliable results when the hold-out test set consists of few CBCT scans, as it heavily depends on the properties of the randomly chosen CBCT scans in the hold-out test set. In order to overcome the limitation of available data for testing, we applied a 4-fold cross-validation scheme on the test set (Fig. 1B), while using a hold-out validation set. More specifically, 28 CBCT scans were divided into 4 subsets (S1, S2,

S3, and S4), each containing 7 scans. The number of slices was 2226, 2214, 2216, and 2187 in S1, S2, S3, and S4 respectively. Each experiment followed a 4-fold cross-validation scheme, which means that 3 subsets were used for training and 1 subset was used for testing. This process was repeated 4 times such that each CBCT scan was used for testing exactly once. Performing a 4-fold cross-validation scheme on the test set allowed us to evaluate the segmentation performance of the MS-D network on all CBCT scans (28 in total), thus making the evaluation robust to differences between the CBCT scans and insensitive to the random choice of test set.

An independent hold-out validation set was used to determine the optimal number of epochs for training. This validation set consisted of 2 CBCT scans which were not included in the 4-fold cross-validation scheme. The number of epochs was chosen as 20 for all training iterations, as the segmentation performance on the validation set did not improve when trained longer. It should be noted that the validation set consisted of relatively few CBCT scans. However, because the MS-D network has a low risk of overfitting the training data [1], and only a single hyper-parameter was tuned (i.e., number of epochs), 2 CBCT scans were sufficient to reliably determine the number of epochs in our study.

The MS-D network was implemented by Hendriksen [5] and the python code for training the MS-D network is publicly available at https://github.com/ahendriksen/msd_pytorch. Implementation of the MS-D network was performed using the deep learning platform PyTorch (version 0.3.1) in Python (version 3.6.1). Training and testing were performed on 2D axial CBCT slices using a batch size of 1 and the default Adam optimizer [6] on a Linux desktop computer (HP Workstation Z840) with 64 GB RAM, a Xeon E5-2687 v4 3.0 GHZ CPU and a GTX 1080 Ti GPU card. Each training epoch took approximately 1 hour.

CNN performance evaluation

The segmentation performance of the MS-D network was evaluated using the Dice similarity coefficient (DSC) which is a well-known metric in the medical image segmentation domain [7]. DSCs were calculated on the patient level, which means that a single DSC was calculated for each segmented CBCT volume.

All segmented CBCT scans (i.e., MS-D network segmentations and gold standard segmentations) were also converted into 3D models using 3D Slicer software [8]. Surface deviations between the MS-D network-based 3D models and the gold standard 3D models were calculated to evaluate the accuracy of the MS-D segmentation around the edges of bony structures. These surface deviations were analyzed within the range of -5.0 mm and +5.0 mm using GOM Inspect software (GOM Inspect 2018, GOM GmbH, Braunschweig, Germany). Additionally, mean absolute deviations (MADs) were calculated between all the MS-D network-based 3D models and the corresponding gold standard 3D models.

After the 4 iterations of the cross-validation scheme, the performance of the MS-D network was averaged over the 28 segmented CBCT scans. All results are presented as means ± standard deviation (SD). The data analysis was performed using GraphPad Prism 8 (GraphPad). Equivalence tests were performed with a threshold difference of 0.005. If the 90% CIs were within (-0.005, 0.005), the two groups were considered to be equivalent with a confidence of 95%.

Table S1. Mean absolute surface deviation of jaw and teeth 3D models

3

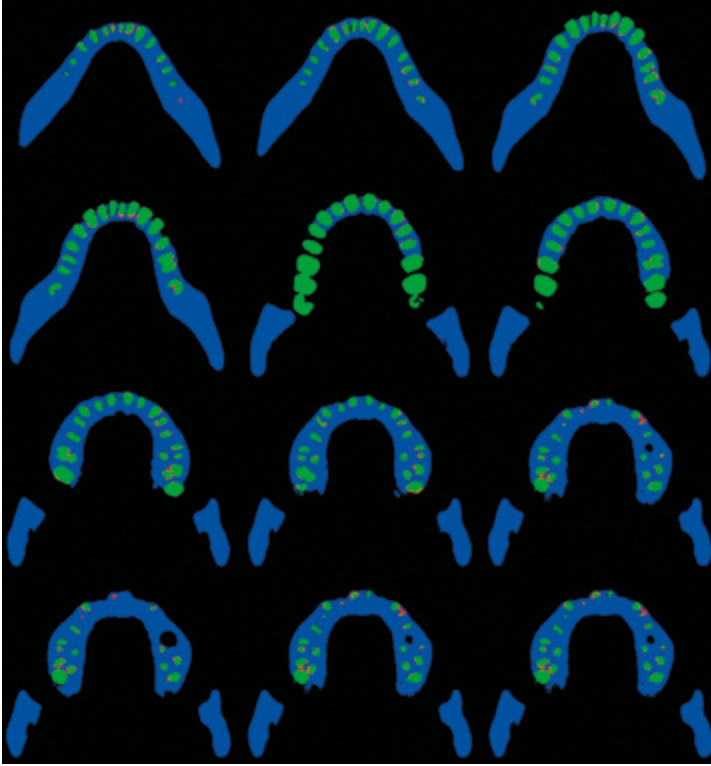| Patient ID	Jaw segmentation		Teeth segmentation	
	Multi-class (mm)	**Binary (mm)**	**Multi-class (mm)**	**Binary (mm)**
P1	0.407	0.403	0.184	0.170
P2	0.443	0.450	0.238	0.211
P3	0.484	0.538	0.230	0.203
P4	0.540	0.527	0.227	0.197
P5	0.364	0.408	0.251	0.208
P6	0.371	0.372	0.132	0.126
P7	0.345	0.340	0.133	0.141
P8	0.438	0.434	0.168	0.132
P9	0.297	0.327	0.152	0.124
P10	0.401	0.432	0.189	0.178
P11	0.369	0.324	0.297	0.149
P12	0.615	0.630	0.258	0.267
P13	0.330	0.345	0.244	0.194
P14	0.524	0.516	0.199	0.197
P15	Validation	Validation	Validation	Validation
P16	Validation	Validation	Validation	Validation
P17	0.375	0.416	0.178	0.136
P18	0.446	0.483	0.159	0.118
P19	0.415	0.454	0.280	0.156
P20	0.308	0.333	0.255	0.179
P21	0.371	0.368	0.110	0.090
P22	0.268	0.265	0.146	0.132
P23	0.469	0.488	0.352	0.321
P24	0.273	0.296	0.168	0.118
P25	0.387	0.434	0.202	0.157
P26	0.370	0.336	0.186	0.160
P27	0.378	0.349	0.291	0.168
P28	0.198	0.244	0.100	0.077
P29	0.490	0.669	0.166	0.134
P30	0.257	0.305	0.206	0.128
min	0.198	0.244	0.100	0.077
max	0.615	0.669	0.352	0.321
Mean ± SD	0.390±0.093	0.410±0.103	0.204±0.061	0.163±0.051

3

Figure S1. The conflicting labels induced by the binary segmentation (12 CBCT slices from patient 1). The conflicting labels are marked in red. The blue color represents the segmented jaw and the green color represents the segmented teeth.

REFERENCES

1. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. Proc. Natl. Acad. Sci. Unit. States Am. 2018; 115: 254–259.

2. Minnema J, van Eijnatten M, Hendriksen AA, Liberton N, Pelt DM, Batenburg KJ, *et al*. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med Phys*. 2019; 46(11): 5027–5035.

3. Pelt DM, Batenburg KJ, Sethian JA. Improving tomographic reconstruction from limited data using mixed-scale dense convolutional neural networks. J. Imaging. 2018; 4(11): 128.

4. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The "k" in k-fold cross validation. *Paper presented at: ESANN 2012 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; April 25–27, 2012; Bruges, BE,* 2012: 441–446.

5. Hendriksen AA. ahendriksenh/msd_pytorch: v0.7.2. *Zenodo*. 2019

6. Kingma DP, Ba J, Adam: a method for stochastic optimization. ArXiv:1412.6980 [Cs]. 2015. http://arxiv.org/abs/1412.6980.

7. Zou KH, Warfield SK, Bharatha A, Tempany CMC, Kaus MR, haker SJ, *et al*. Statistical validation of image segmentation quality based on a spatial overlap index. *Acad Radiol*. 2004; 11(2): 178–189

8. 3D Slicer. 2018. http://www.slicer.org.

CHAPTER

4

COMPARISON OF CONVOLUTIONAL NEURAL NETWORK TRAINING STRATEGIES FOR CONE-BEAM CT IMAGE SEGMENTATION

Jordi Minnema, Jan Wolff, Juha Koivisto,
Felix Lucka, Kees Joost Batenburg,
Tymour Forouzanfar, Maureen van Eijnatten

ABSTRACT

Background and objective

Over the past decade, convolutional neural networks (CNNs) have revolutionized the field of medical image segmentation. Prompted by the developments in computational resources and the availability of large datasets, a wide variety of different two-dimensional (2D) and three-dimensional (3D) CNN training strategies have been proposed. However, a systematic comparison of the impact of these strategies on the image segmentation performance is still lacking. Therefore, this study aimed to compare eight different CNN training strategies, namely 2D (axial, sagittal and coronal slices), 2.5D (3 and 5 adjacent slices), majority voting, randomly oriented 2D cross-sections and 3D patches.

Methods

These eight strategies were used to train a U-Net and an MS-D network for the segmentation of simulated cone-beam computed tomography (CBCT) images comprising randomly-placed non-overlapping cylinders and experimental CBCT images of anthropomorphic phantom heads. The resulting segmentation performances were quantitatively compared by calculating Dice similarity coefficients. In addition, all segmented and gold standard experimental CBCT images were converted into virtual 3D models and compared using orientation-based surface comparisons.

Results

The CNN training strategy that generally resulted in the best performances on both simulated and experimental CBCT images was majority voting. When employing 2D training strategies, the segmentation performance can be optimized by training on image slices that are perpendicular to the predominant orientation of the anatomical structure of interest. Such spatial features should be taken into account when choosing or developing novel CNN training strategies for medical image segmentation.

Conclusions

The results of this study will help clinicians and engineers to choose the most-suited CNN training strategy for CBCT image segmentation.

1.   INTRODUCTION

Convolutional neural networks (CNNs) are becoming increasingly popular for a wide range of medical image segmentation tasks [1]. The first CNNs developed to segment three-dimensional (3D) images acquired using e.g. magnetic resonance imaging (MRI) or computed tomography (CT) were trained on two-dimensional (2D) image slices [2]. The majority of these CNNs used axial slices as input [3-5] due to the high in-plane resolution with respect to the slice thickness [1]. Recent advances in Graphics Processing Unit (GPU) computing and efficient CNN architectures such as U-Net [6] led to an increasing number of studies on 3D CNNs [7-9]. However, due to the memory constraints of current GPUs, fully 3D CNN approaches remain limited in terms of volume size (e.g., 256 × 256 × 256). Therefore, most 3D CNNs are trained using cropped images (e.g., 128 × 128 × 128) [9] or small patches (e.g., 23 × 23 × 23) [10] that are extracted from the original images. As a consequence, global anatomical information is not adequately modeled within the segmentation approach. Another disadvantage of 3D CNNs is their large number of trainable parameters [11], which makes these networks more prone to overfitting and leads to considerably longer training times when compared to 2D CNNs [12].

The aforementioned limitations have sparked a wide range of different CNN training strategies to exploit the 3D nature of anatomical features in image segmentation tasks without the need to train 3D CNNs. In the present study, we refer to these approaches as *augmented 2D* training strategies. An example of such a training strategy is the "2.5D" approach, in which a small number of 2D slices are combined for CNN training purposes. This can be achieved either by combining three orthogonal slices (i.e., axial, sagittal and coronal) to classify the voxel at the intersection of the three slices [13,14], or by combining three adjacent slices in the same plane [15]. Another augmented 2D training strategy is to train separate CNNs for each of the three orthogonal slices and combine the predictions of these CNNs using majority voting [16]. This strategy was developed to mimic the thought process of radiologists, who typically first analyze multiple 2D image slices from different orthogonal planes and then combine them to interpret the shape and size of the anatomical structure of interest [17].

To date, there has been little agreement in the literature [7,13,18-23] on which CNN training strategy is the best for 3D medical image segmentation. In addition, it remains unclear how the segmentation performance of different strategies is influenced by the spatial characteristics of the anatomical structure of interest. Therefore, the aim of this study was to compare different 2D, augmented 2D and 3D training strategies for the segmentation of CBCT images containing structures with diverse spatial orientations. Segmentation of bony structures in CBCT images is often required for diagnostic purposes [24], virtual treatment planning [25], personalized implant design [26] and post-operative analysis [27]. However, this task is notoriously difficult since CBCT images are typically affected by high noise levels, directionally dependent imaging artifacts and partial volume effects [28]. Although previous studies have experimented with using CNNs to automate CBCT image segmentation [29,30], it remains unclear which training strategy is best suited for this task.

The contributions of this study are as follows:

1. This study is the first to provide a comprehensive and quantitative comparison between different 2D, augmented 2D and 3D CNN training strategies for CBCT image segmentation.

2. All CNN training strategies were evaluated using simulated CBCT images with a known ground truth, as well as real CBCT images of physical anthropomorphic phantom heads for which high-quality gold standard segmentation labels were acquired using an industrial micro-CT scanner.

3. In addition to calculating Dice similarity coefficients commonly used in the field, we propose a novel metric to quantify segmentation performance by converting the segmented images into virtual 3D surface models and performing orientation-based surface comparisons.

4. This study demonstrates that majority voting can improve a CNN's segmentation performance compared to conventional 2D and 3D CNN training strategies.

2. MATERIALS AND METHODS

The performance of 2D, augmented 2D and 3D CNN training strategies was quantitatively compared using both simulated CBCT images and experimental CBCT images. The simulation of CBCT images offered the unique possibility to create images with a known ground truth in which the spatial orientation of the inner structures could be precisely controlled. The experimental CBCT images were obtained by imaging five anthropomorphic phantom heads that were also scanned using an industrial micro-CT scanner. This allowed us to create highly accurate gold standard segmentation labels, which are indispensable for a fair comparison between training strategies.

2.1. CBCT simulations

A simulation phantom was generated by removing 10,500 randomly-placed non-overlapping cylinders from a large homogeneous cylinder with a value of one. The large cylinder had a height of 300 mm and a radius of 100 mm. The height of the smaller cylinders ranged from 7 to 12 mm and their radius ranged from 1.4 to 2.4 mm. Three different types of this simulation phantom were generated (Fig. 1) by varying the orientation of the smaller cylinders in the XZ-plane as follows: (1) fixed orientation of the smaller cylinders parallel to the Z-axis; (2) randomly rotating the smaller cylinders independently between -20 and 20°; and (3) randomly rotating the smaller cylinders independently between -90 and 90°. For each of these three phantom types, ten different simulation phantoms were created. The code to construct the phantoms was adapted from a recent study by Hendriksen et al. [31,32].

All 30 simulation phantoms were subsequently used to simulate CBCT projections using the Astra Toolbox [33,34] (v. 1.8) that provides high-performance GPU implementations for tomographic operations with flexible geometries. In order to remain as close as possible to clinical practice, a limited-data CBCT geometry was simulated by calculating 180 projections with an angular increment of 2° for each simulation phantom. These simulated projections were used to reconstruct CBCT images using the Feldkamp Davis and Kress (FDK) algorithm [35]. The dimensions of the reconstructed CBCT images were set to 512 × 512 × 512. Finally, Gaussian noise ($\mu = 0$ and $= 0.2$) was added to all reconstructed CBCT images.

2.2. Experimental CBCT data

CBCT images of five anthropomorphic phantom heads were acquired in this study. Four of these heads contained real human bone (The Phantom Laboratory, Salem, NY, USA; Erler-Zimmer GmbH & Co.KG, Lauf, Germany), and one contained acrylic bony structures (CIRS Inc., Norfolk, VA, USA). Such anthropomorphic phantom heads are commonly used by CBCT device manufacturers to evaluate their scanners for clinical use, and are specifically designed to mimic the tissue densities and morphologies of real human heads. As a result, these phantom heads cause realistic imaging artifacts without the need of exposing patients to harmful X-ray radiation.

All phantom heads were scanned using a Planmeca ProMax Mid CBCT scanner (Planmeca Oy., Helsinki, Finland) that is widely used in dentistry and maxillofacial surgery. All scans were performed using a tube voltage of 90 kVp, a tube current of 10 mA and an isotropic voxel size of 0.2 mm using two non-overlapping imaging protocols. The first imaging protocol covered the top part of the phantom heads and the second imaging protocol covered the bottom part of the phantom heads (Fig. 2), resulting in a total of ten CBCT images with dimensions between $1001 \times 1001 \times 153$ and $1001 \times 1001 \times 380$. Since the acquisition of annotated medical imaging data is a common challenge in training deep learning algorithms, we believe that the relative small size of the datasets used in the present study (10 CBCT scans per dataset) is representative for the datasets that can be typically obtained in clinical settings.
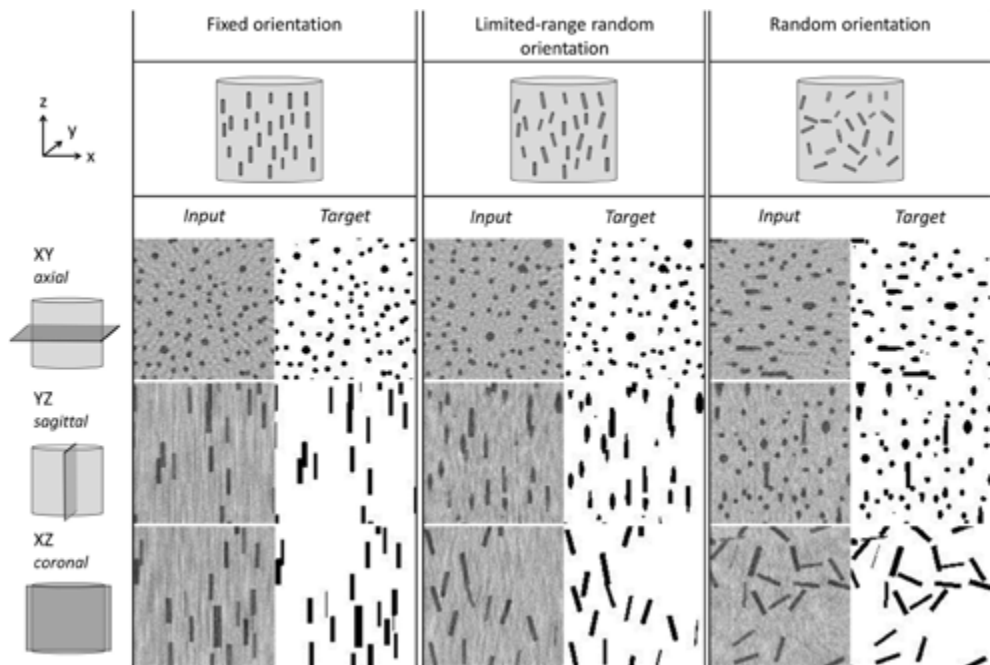

Figure 1. Schematic representation of the three different simulation phantoms used in this study (top) and magnified examples (4x) of the central slices of the resulting CBCT images (input and gold standard segmentation labels).
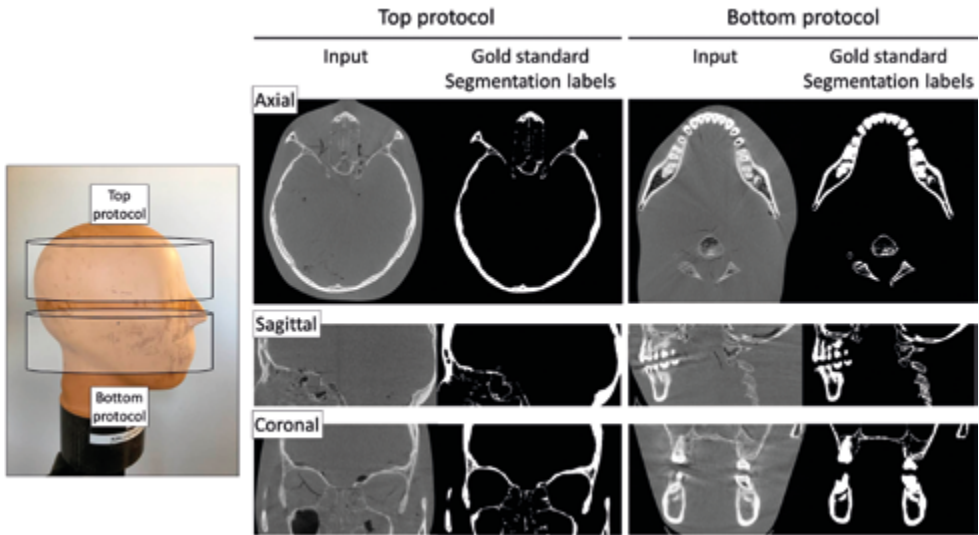
4

Figure 2. Example of one anthropomorphic phantom head with the corresponding input CBCT slices and gold standard segmentation labels.

In order to create gold standard segmentation labels, all five phantom heads were also scanned using an industrial GE phoenix v|tome|x m cone-beam micro-CT scanner (GE Sensing & Inspection Technologies GmbH, Wunstorff, Germany) using a tube voltage of 100 kVp, a tube current of 1.2 mA and 250 ms exposure time. All micro-CT images were reconstructed with an isotropic voxel size of 0.12 mm. Since the voxel sizes of the micro-CT images were smaller than those of the CBCT images, the voxels of the micro-CT images were rescaled to 0.2 mm. The high radiation dose and long scanning times (roughly 2 h per phantom head) resulted in micro-CT images with a superior signal-to-noise ratio (SNR) when compared to the CBCT images, which enabled accurate segmentation of the bony structures. Segmentation of all five micro-CT images was performed using global thresholding, followed by manual post-processing using the open-source 3D slicer software package [36,37]. The segmented micro-CT images were subsequently aligned on the CBCT images and cropped to the same dimensions. The aligned micro-CT segmentation labels served as gold standard segmentation labels (Fig. 2).

2.3. Training strategies

Eight different CNN training strategies were evaluated (Fig. 3). The first three training strategies were traditional 2D approaches in which axial, sagittal and coronal CBCT slices were used to train a 2D CNN (Fig. 3a–c).

In addition, three augmented 2D training strategies were evaluated. First, we implemented a 2.5D approach proposed by Ben-Cohen et al. [15] in which a 2D CNN was trained using input images consisting of 3 or 5 channels: one axial slice of interest and either two or four adjacent slices above and below this axial slice (Fig. 3d and e). Second, we evaluated a majority voting (MV)
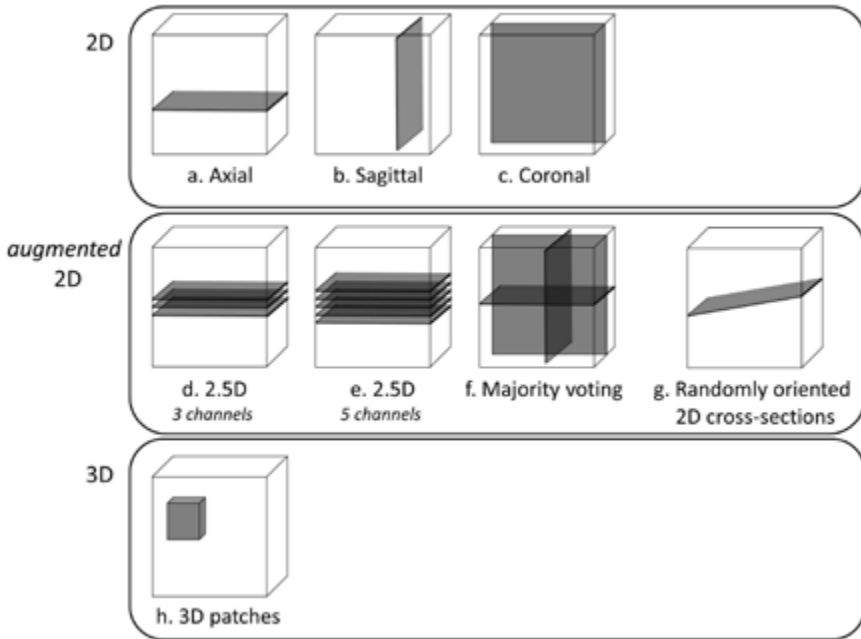
Figure 3. Schematic overview of the eight CNN training strategies evaluated in the present study.

scheme proposed by Zhou et al. [16,17] (Fig. 3f). In this scheme, separate 2D CNNs were trained for each of the three orthogonal slices, after which a voxel was labeled as foreground if at least two of the three trained CNNs also labeled that voxel as foreground. Third, we developed a novel augmented 2D training strategy in which a 2D CNN was trained using randomly oriented 2D cross-sections of CBCT images (Fig. 3g). In order to compare this cross-section strategy with the 2D axial strategy, we acquired the same number of randomly oriented cross-sections from each CBCT image as the original number of axial slices. The cross-sections were equally spread over the Z-axis of the CBCT image and the orientation of each cross-section was randomly and independently sampled with an angle of at most 10° with respect to the axial plane. We refer to this novel approach as the *randomly oriented 2D cross-sections* strategy.

Finally, we evaluated a fully 3D training strategy (Fig. 3h). Existing 3D CNN approaches typically use cropped [38] or downsampled [7] 3D images due to GPU memory constraints. However, cropping leads to a loss of global information, whereas downsampling results in lower spatial resolution and thus a loss of fine details in the images. Therefore, the most popular 3D CNN training strategies use multiple 3D patches extracted from the original images [11,39]. In the present study, we implemented a 3D patch-based training strategy that allowed us to use all voxels of the CBCT image when training the CNN without reducing the image resolution. All CBCT images were padded with reflective boundaries to ensure that image dimensions were a multiple of the patch size. Partially overlapping 3D patches of 128 × 128 × 128 voxels were extracted from the padded CBCT images with a stride of 64 voxels, resulting in 343 3D patches acquired from each simulated CBCT

image and between 675 and 1125 3D patches acquired from each experimental CBCT image. Thus, a large number of training instances (i.e., 2D slices or 3D patches) were extracted for all training strategies, ensuring convergence of the CNN during the training process.

2.4. CNN architecture

In order to compare the effectiveness of the aforementioned training strategies, CNN architectures were required that (1) can be broadly applied for many different medical image segmentation tasks and (2) are robust and easy to train on a small number of CBCT images. In order to comply with (1), we employed the commonly used U-net architecture that has been used for a wide variety of medical image segmentation tasks [40 – 42]. However, since U-net consists of a relatively large number of trainable parameters, it tends to overfit when few training images are available. Therefore, to comply with criterion (2), we also used the mixed-scale dense convolutional neural network (MS-D network) initially developed by Pelt and Sethian [43]. The MS-D network uses dense connections to directly pass relevant feature maps to deeper layers of the network. As a result, fewer trainable parameters are necessary compared to U-Net [29], thereby reducing the risk of overfitting and making the MS-D network particularly suited to process small imaging datasets.

The U-Net and the MS-D network were implemented in Python (v. 3.7.4) using the deep learning framework PyTorch (v. 1.1.0) and are both publicly accessible online [44,45]. The U-Net used in the present study was comparable to the one described by Ronneberger et al., [6] except that we used batch normalization [46] after each Rectified Linear Units (ReLU) activation function and applied reflection padding on input images of which the dimensions were not divisible by 16. The MS-D network was the same as the one proposed by Pelt and Sethian [43], with a depth of 50 convolutional layers and a width of 1.

In order to ensure a fair comparison between the different CNN training strategies, both CNN architectures (i.e., U-Net and MS-D network) were trained using the default Adam optimizer [47] (i.e., learning rate = 0.001, $\varepsilon = 1 \times 10^{-8}$) with a batch size of 1 on a server with 192 GB RAM and one NVidia GeForce GTX 1080 Ti GPU. Both CNNs were trained for 10 epochs on the experimental CBCT images and for 50 epochs on the simulated CBCT images. In order to avoid overfitting, a relatively small number of epochs (i.e., 10) was used to train the CNNs on the experimental data. The chosen number of epochs was based on the results of a previous study [29], in which we found that training the networks for 10 epochs already was sufficient to achieve satisfactory performances in image segmentation tasks. Training of the CNNs on the simulated datasets was performed for 50 epochs. The reason for this was that the risk of overfitting on the simulated dataset was substantially smaller, since this dataset consisted of manually created CBCT scans that shared many similar image properties and features compared to CBCT scans of the experimental dataset (e.g., same intensity values, same range of cylinder sizes).

2.5. Evaluation

All eight training strategies were evaluated on each of the four different datasets, i.e., the three types of simulated CBCT images and the experimental CBCT images. A leave-2-out scheme was

used in which eight of the ten CBCT images were alternately used for training and two for testing. In the leave-2-out scheme performed using the experimental CBCT images, the test set always consisted of the two CBCT images acquired from the same phantom head (i.e., the top and the bottom imaging protocol). This ensured that training and testing of the CNNs was fully independent. The 3D patch-based training strategy was only evaluated for U-Net since the MS-D network implementation used in this study does currently not support 3D convolutions.

All segmentation performances of the trained CNNs were assessed using the Dice similarity coefficient (DSC). The DSC is the most commonly used measure for the overlap between a segmented image and the corresponding gold standard segmentation labels, and is given by

$$DSC = \frac{2TP}{2TP + FP + FN}$$ (1)

, where TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives. To enable mutual comparisons between the segmentation performances achieved by the different training strategies, all DSCs were normalized by dividing them by the DSC achieved by the U-net trained on axial slices. Statistical differences were calculated using a paired nonparametric Wilcoxon signed-rank test at a predefined significance level of $P < 0.05$.

As an additional evaluation step, all segmented experimental CBCT images and gold standard CBCT images were converted into virtual 3D models in the standard tessellation language (STL) file format using 3D slicer software [36,37]. The resulting STL models were geometrically compared to the corresponding gold standard STL model using the surface comparison module in GOM Inspect software (GOM Inspect 2018, GOM GmbH, Braunschweig, Germany). Mean absolute deviations (MADs) between the gold standard STL models and the CNN-based STL models were calculated for bony structures that were oriented between -20 and 20° with respect to each of the three orthogonal planes (Fig. 4). The MADs were then separately analyzed for each orthogonal plane. This novel metric allowed us to investigate the influence of the spatial orientation of bony structures on the segmentation performance of the different CNN training strategies.

3. RESULTS

Generally, the highest mean DSCs were achieved using majority voting (Table 1 and 2; Figs. 5 and 6). The 3D patch-based strategy resulted in mean DSCs comparable to those achieved using majority voting when segmenting simulated CBCT images (Table 1; Fig. 5), but resulted in the lowest mean DSCs when segmenting experimental CBCT images (Table 2; Fig. 6). Both 2.5D strategies resulted in significantly lower mean DSCs compared to training on axial slices ($P < 0.001$) when segmenting simulated CBCT images. Moreover, the 5-channel strategy resulted in significantly lower mean DSCs compared to the 3-channel strategy ($P < 0.001$). In the experimental CBCT images, no significant differences were observed between the 2.5D strategies and the axial strategy. The randomly oriented 2D cross-sections strategy always resulted in similar DSCs compared to training on axial slices. Training on coronal slices resulted in significantly higher mean DSCs($P < 0.001$) compared
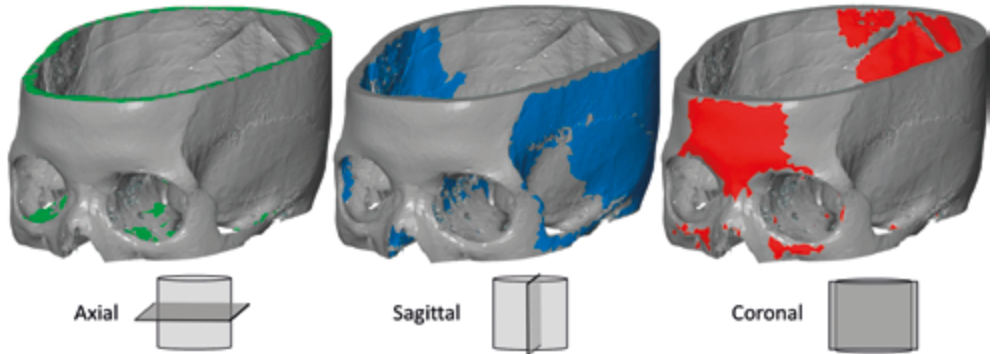
Figure 4. Visual representation of the bony structures that were separately analyzed for each orthogonal plane.

to training on axial slices in all simulated experiments, whereas training on sagittal slices only outperformed the axial strategy when segmenting fixed orientation cylinders ($P < 0.001$). No significant differences were observed between the three 2D training strategies when segmenting experimental CBCT images. In addition, no significant differences were observed between the two CNN architectures, i.e., U-Net and MS-D network.

The segmentation performances of the eight CNN training strategies on experimental CBCT images were also compared using orientation-based surface comparisons. Generally, the majority voting strategy resulted in the lowest MADs, i.e., the least deviations from the gold standard STL models (Table 3; Fig. 7). Interestingly, the 2D strategies generally resulted in higher MADs when segmenting bony structures oriented in the same plane as the training slices. Furthermore, the randomly oriented 2D cross-sections strategy resulted in higher MADs when segmenting bony structures oriented in the axial plane compared to structures oriented in the sagittal and coronal planes.

The segmentation performances of the eight CNN training strategies on experimental CBCT images were also compared using orientation-based surface comparisons. Generally, the majority voting strategy resulted in the lowest MADs, i.e., the least deviations from the gold standard STL models (Table 3; Fig. 7). Interestingly, the 2D strategies generally resulted in higher MADs when segmenting bony structures oriented in the same plane as the training slices. Furthermore, the randomly oriented 2D cross-sections strategy resulted in higher MADs when segmenting bony structures oriented in the axial plane compared to structures oriented in the sagittal and coronal planes.

Table 4 shows the number of trainable parameters of the two network architectures used in this study, as well as the training time per epoch and the segmentation time required to segment one simulated CBCT image. The 2D U-Net comprised roughly 1000 time more trainable parameters and required longer training and segmentation times compared to the MS-D network. The 3D U-Net comprised approximately 3 times more trainable parameters compared to the 2D U-Net and the training and segmentation times were 30 and 10 time longer, respectively.

Table 1. Mean Dice similarity coefficients (± standard deviation) of the segmented simulated CBCT images comprising cylinders with a fixed, limited-range random and random orientation.

	Fixed orientation		Limited-range random orientation		Random orientation	
	U-Net	MS-D network	U-Net	MS-D network	U-Net	MS-D network
2D						
Axial	0.9930 ± 0.0006	0.9913 ± 0.0004	0.9899 ± 0.0002	0.9869 ± 0.0002	0.9879 ± 0.0008	0.9849 ± 0.0003
Sagittal	0.9992 ± 0.0001	0.9993 ± 0.0001	0.9908 ± 0.0002	0.9888 ± 0.0002	0.9869 ± 0.0001	0.9842 ± 0.0003
Coronal	0.9991 ± 0.0001	0.9993 ± 0.0001	0.9930 ± 0.0001	0.9909 ± 0.0002	0.9910 ± 0.0002	0.9868 ± 0.0003
Augmented 2D						
2.5D						
3 channels	0.9878 ± 0.0001	0.9857 ± 0.002	0.9754 ± 0.0001	0.9727 ± 0.0003	0.9733 ± 0.0002	0.9668 ± 0.0003
5 channels	0.9782 ± 0.0001	0.9752 ± 0.0005	0.9608 ± 0.0003	0.9560 ± 0.0003	0.9577 ± 0.0002	0.9458 ± 0.0006
MV	**0.9996 ± 0.0001**	**0.9995 ± 0.0001**	0.9946 ± 0.0001	**0.9924 ± 0.0001**	0.9927 ± 0.0001	**0.9890 ± 0.0002**
Randomly oriented cross-sections	0.9929 ± 0.0002	0.9918 ± 0.0003	0.9891 ± 0.0006	0.9867 ± 0.0003	0.9882 ± 0.0006	0.9845 ± 0.0003
3D						
3D patches	0.9991 ± 0.0003	n.a.	**0.9958 ± 0.0001**	n.a.	**0.9940 ± 0.0003**	n.a.

4

Table 2. Mean Dice similarity coefficients (± standard deviation) of the segmented experimental CBCT images.

	U-Net	MS-D network
2D		
Axial	0.805 ± 0.10	0.809 ± 0.10
Sagittal	0.802 ± 0.11	0.806 ± 0.10
Coronal	0.799 ± 0.11	0.813 ± 0.10
Augmented 2D		
2.5D		
3 channels	0.813 ± 0.10	0.807 ± 0.09
5 channels	0.803 ± 0.10	0.801 ± 0.09
MV	**0.821 ± 0.11**	**0.821 ± 0.10**
randomly oriented cross-sections	0.811 ± 0.09	0.808 ±0.09
3D		
3D patches	0.782 ± 0.13	n.a.

4. DISCUSSION

Although CNNs are being employed for an increasing number of 2D and 3D medical image segmentation tasks, it remains unclear which CNN training strategy is best in terms of segmentation performance and computational cost. In this study, we therefore compared eight different 2D, augmented 2D and 3D CNN training strategies for the segmentation of CBCT images comprising structures with diverse spatial orientations. The strategy that generally resulted in the highest DSCs and lowest MADs was majority voting (Table 1–3; Figs. 5–7). This finding is in agreement with a recent study by Zhou et al., who improved the segmentation of multiple organs in abdominal CT images by training three different CNNs and combining the resulting segmentations [16]. These findings are also supported by the study of Mlynarski et al., who reported that a U-Net trained using the majority voting strategy more accurately segmented various anatomical regions in brain MRI scans than with a 2D axial strategy [48]. Majority voting strategies thus seem to perform well on various imaging datasets and anatomical regions of interest. The high segmentation performance achieved by majority voting strategies can be explained by the fact that they combine distinct anatomical features from different orthogonal slices, thereby correcting erroneously labeled voxels in one slice if these voxels are correctly labeled in the other two slices.

Both 2.5D CNN training strategies evaluated in this study resulted in comparable or worse segmentation performances compared to training on axial slices. These results are in line with those presented by Desai et al, who reported that their 2.5D strategy did not lead to more accurate segmentation of femur cartilage in MRI scans compared to training with axial slices [49]. However, contradicting findings were presented by Ben-Cohen et al. [15] and Vu et al. [50] who reported that their 2.5D strategies generally outperformed training on axial slices when segmenting different CT and MR images. Similarly, Zhang et al., showed that their U-Net trained with a 2.5D approach resulted in higher DSCs than training with axial slices when segmenting the heart and the spleen [51]. These contradicting findings may be explained by the fact that Ben-Cohen et al., Vu et al., and Zhang et al. segmented relatively large and connected soft tissue structures such as pelvic region organs or
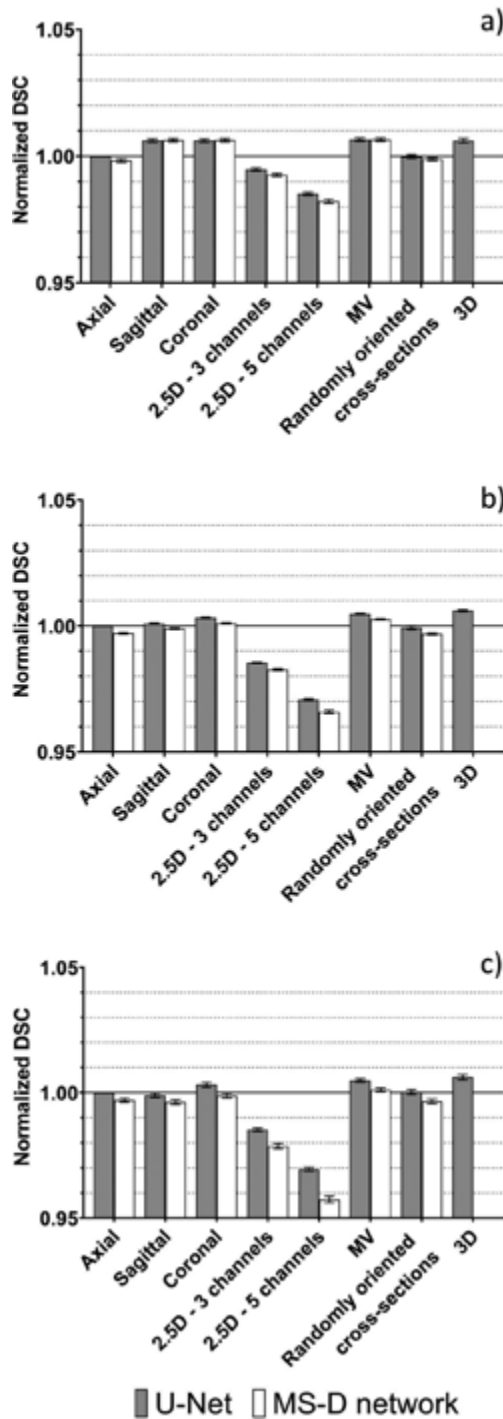
4

Figure 5. Mean normalized Dice similarity coefficients (DSCs) of the segmented simulated CBCT images comprising cylinders with (a) fixed orientation; (b) limited-range random orientation; and (c) random orientation. DSCs were normalized by dividing them by the DSC obtained using U-Net trained on axial slices.
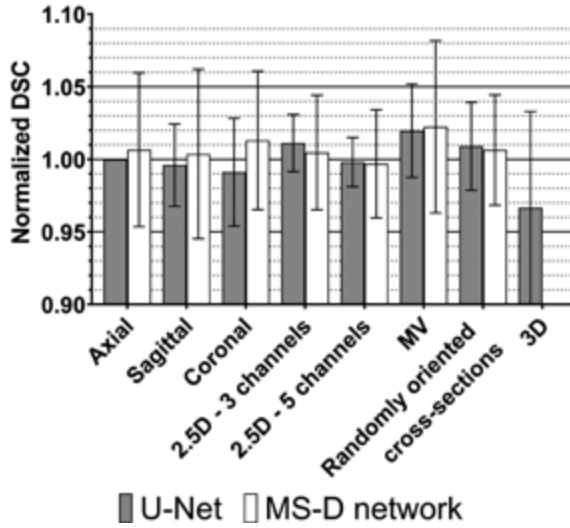
4

Figure 6. Mean normalized Dice similarity coefficients (DSC) of the segmented experimental CBCT images. DSCs were normalized by dividing them by the DSC obtained using U-Net trained on axial slices.

Table 3. Mean absolute surface deviations (MADs) (± standard deviation) between the gold standard STL models and the STL models acquired using the eight different CNN training strategies. MADs were calculated separately for bony structures oriented in the axial, sagittal and coronal plane.

	MAD of bony structures in axial plane (mm)		MAD of bony structures in sagittal plane (mm)		MAD of bony structures in coronal plane (mm)	
	U-Net	MS-D network	U-Net	MS-D network	U-Net	MS-D network
2D						
Axial	0.48 ± 0.17	0.54 ± 0.17	0.36 ± 0.18	0.41 ± 0.20	0.42 ± 0.20	0.56 ± 0.33
Sagittal	0.43 ± 0.15	0.48 ± 0.14	0.51 ± 0.23	0.53 ± 0.16	0.42 ± 0.18	0.49 ± 0.13
Coronal	0.40 ± 0.12	0.41 ± 0.13	0.37 ± 0.16	0.41 ± 0.16	0.52 ± 0.25	0.52 ± 0.20
Augmented 2D						
2.5D						
3 channels	0.43 ± 0.14	0.43 ± 0.15	0.36 ± 0.18	0.40 ± 0.18	0.42 ± 0.22	0.43 ± 0.19
5 channels	0.44 ± 0.15	0.44 ± 013	**0.34 ± 0.16**	0.41 ± 0.16	**0.41 ± 0.20**	0.43 ± 0.17
MV	**0.38 ± 0.14**	**0.36 ± 0.09**	0.37 ± 0.18	**0.38 ± 0.17**	**0.41 ± 0.23**	**0.39 ± 0.15**
randomly oriented cross-sections	0.50 ± 0.17	0.49 ± 0.16	0.35 ± 0.18	0.42 ± 0.16	0.42 ± 0.20	0.46 ± 0.16
3D						
3D patches	0.48 ± 0.20	n.a.	0.44 ± 0.24	n.a.	0.50 ± 0.36	n.a.

4
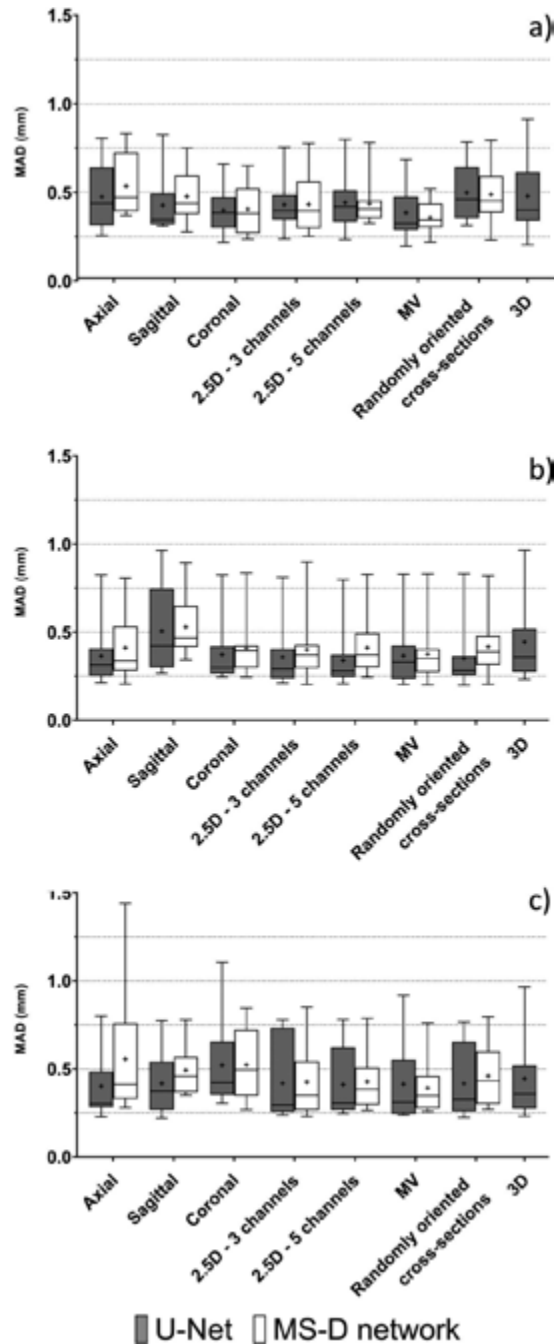
Figure 7. Box and whisker plot of the mean absolute surface deviations (MADs) between the gold standard STL models and the STL models acquired using the eight different CNN training strategies. MADs were calculated separately for bony structures oriented in the axial (a), sagittal (b) and coronal (c) plane. The boxes represent the interquartile range and the whiskers represent the lowest and highest MAD. The median and mean are represented by the lines and dots in the boxes, respectively.

Table 4. Overview of the number of trainable parameters, training times per epoch, and the times required to segment one simulated CBCT image.

	Number of trainable parameters		Training time per epoch (s)		Segmentation time per CBCT image (s)	
	U-Net	MS-D network	U-Net	MS-D network	U-Net	MS-D network
2D						
Axial/sagittal/coronal	14,787,844	11,631	377.8 ± 0.4	260.2 ±1.3	25.6 ± 0.8	12.4 ± 0.8
Augmented 2D						
2.5D						
3 channels	14,789,006	12,545	377.4 ± 1.4	274.6 ± 1.0	29.3 ± 1.7	14.8 ± 0.7
5 channels	14,790,176	13,467	379.8 ± 2.6	288.6 ± 0.5	27.3 ± 0.3	15.6 ± 0.7
MV	*3 x*	*3 x*	*3 x*	*3 x*	*3 x*	*3 x*
	14,787,844	11,631	377.8 ± 0.4	260.2 ±1.3	25.6 ± 0.8	12.4 ± 0.8
randomly oriented cross-sections	14,787,844	11,631	374,6 ± 0.4	260.3 ± 0.7	22,9 ± 0.1	10,8 ± 1.1
3D						
3D Patches	42,944,900	n.a	9960.7 ± 58	n.a	259.3 ± 13.5	n.a

brain tumors, whereas the present study, and the study by Desai et al., focused on relatively small and thin structures of which the appearance can differ considerably between consecutive slices. The large variations in the shape of thin anatomical structures may have impeded the CNNs' ability to learn spatial relations between multiple adjacent image slices.

Another interesting finding was that, in all simulated experiments, training on coronal slices outperformed training on axial slices (Table 1). This phenomenon may be due to the fact that coronal slices contained more voxels of each inner cylinder compared to axial slices (Fig. 1), which facilitated segmentation of the cylinders. When segmenting the experimental CBCT images, the best segmentation results were generally achieved by training on slices perpendicular to the orientation of the bony structures (Table 3 and Fig. 7). A possible explanation could be that structures in perpendicular slices are less affected by the partial volume effect, resulting in better contrast between the bony structures and the surrounding soft structures of the anthropomorphic phantom heads.

The 3D patch-based strategy evaluated in this study resulted in the highest mean DSC when segmenting simulated CBCT images, whereas it resulted in the lowest mean DSC when segmenting experimental CBCT images (Table 1, 2 and Figs. 5, 6). This difference is likely due to the fact that the simulated CBCT images only comprised inner cylinders with little morphological variation, whereas the experimental CBCT images comprised bony structures with large variations in shape, size and intensity. Consequently, optimization of the large number of trainable parameters in the 3D U-net (Table 4) was challenging with only 8 different CBCT images available for training. Another recent study that showed that 3D CNNs are not always better than 2D CNNs was conducted by Mlynarski et al., [52] who found that a conventional 3D U-Net did not outperform a 2D U-Net when segmenting brain tumors in MR images.

The segmentation performances of the 2D and augmented 2D strategies evaluated in this study were comparable between the two CNN architectures, i.e. U-Net and the MS-D network. The observed performances are therefore likely to be generalizable to different CNN architectures. This finding is consistent with a recent study by Isensee et al., who showed that non-architectural modifications such as training strategies can be more powerful than using different CNN architectures [53].

An important advantage of 2D and augmented 2D CNN training strategies over 3D strategies are the short computational times needed for training and segmentation (Table 4). In the present study, training of the 2D CNNs took approximately 7 min per epoch and segmentation of one simulated CBCT image (512 × 512 × 512) took less than 30 s. In comparison, training the 3D U-Net took almost 3 h per epoch and segmentation of a single CBCT image took more than 4 min. These longer computational times were caused by the 3D U-Net's larger number of trainable parameters and the fact that more voxels needed to be processed during training and segmentation because of the overlapping 3D patches.

The insights gained from this study will hopefully help clinicians and engineers to choose a suitable CNN training strategy for the segmentation task at hand. Nevertheless, further research is needed to investigate to which extent the results of the present study can be generalized to other imaging modalities (e.g., MRI), different anatomical structures, or anisotropic images. Another possible direction for future research is to assess the impact of different training strategies when simultaneously segmenting multiple anatomical regions, i.e. multi-class segmentation. Finally, additional studies are necessary to determine whether the findings of this study also hold when applying different CNN architectures.

5. CONCLUSION

The present study provides a comprehensive comparison between eight different 2D, augmented 2D and 3D CNN training strategies for the segmentation of CBCT scans of the head and neck area. The empirical findings suggest that majority voting is a robust CNN training strategy that generally results in the best segmentation performances. However, if training three separate CNNs is infeasible, it is recommended to train on image slices perpendicular to the predominant orientation of the anatomical structure of interest.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

1. Litjens G, Kooi T, Bejnordi B, Setio A, Ciompi F, Ghafoorian M, *et al*. A survey *on deep learning in medical image analysis*. *Med. Image. Anal. 2017;42: 60-88.*

2. Ciresan D, Giusti A, Gambardella LM, Schmidhuber J. Deep neural networks segment neuronal membranes in electron microscopy images, in: F. Pereira, C.J.C. Burges, L. Bottou, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems 25, Curran Associates, Inc, 2012, pp. 2843–2851 .

3. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, *et al*. Brain tumor segmentation with deep neural networks. *Med. Image Anal.* 2017; 35: 18–31 https://doi.org/10.1016/j.media.2016.05.004 .

4. Vivanti R, Ephrat A, Joskowicz L, Lev-Cohain N, Karaaslan OA, Sosna J. Au- tomatic liver tumor segmentation in follow-up CT scans, in: Proceeding of the Patch-Based Methods in Medical Image Processing Workshop, 2015, pp. 53–61 .

5. Shin HC, Roth HR, Gao M, Lu L, Xu Z, Nogues I, *et al*. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging.* 2016; 35: 1285–1298. doi: 10.1109/TMI.2016.2528162 .

6. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Med. Image Comput. Comput. Assist. Interv. MICCAI.* 2015: 234–241. doi:10.1007/978-3-319-24574-4_28 .

7. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation, in: S. Ourselin, L. Joskowicz, M.R. Sabuncu, G. Unal, W. Wells (Eds.), *Med. Image Comput. Comput. Assist*. *Interv. MICCAI.* 2016: 424–432. doi: 10.1007/978-3-319-46723-8_49 .

8. Dou Q, Chen H, Yu L, Zhao L, Qin J, Wang D, *et al*. Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks. *IEEE Trans. Med. Imaging.* 2016; 35: 1182–1195. doi:10.1109/TMI.2016.2528129 .

9. Milletari F, Navab N, Ahmadi SA. V-NET: fully convolutional neural networks for volumetric medical image segmentation. in: *Proceeding of the 2016 Fourth International Conference on 3D Vision (3DV), Stanford, CA , USA , IEEE.* 2016: 565–571. doi: 10.1109/3DV.2016.79 .

10. Wachinger C, Reuter M, Klein T. DeepNAT: deep convolutional neural network for segmenting neuroanatomy. *NeuroImage.* 2018; 170: 434–445, doi:10.1016/j. neuroimage.2017.02.035 .

11. Kamnitsas K, Ledig C, Newcombe VFJ, Simpson JP, Kane AD, Menon DK, *et al*. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* 2017; 36: 61–78, doi: 10.1016/j.media.2016.10.004 .

12. Zhou X, Yamada K, Takayama R, Zhou X, Hara T, Fujita H, *et al*. Performance evaluation of 2D and 3D deep learning approaches for automatic segmentation of multiple organs on CT images. *in: K. Mori, N. Petrick (Eds.), Medical Imaging 2018, Computer-Aided Diagnosis, SPIE, Houston, United States.* 2018: 83. doi:10.1117/12.2295178 .

13. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *in: C. Salinesi, M.C. Norrie, Ó. Pastor (Eds.), Advanced Information Systems Engineering.* 2013: 246–253. doi:10.1007/978-3-642-40763- 5 _ 31 .

14. Roth HR, Lu L, Seff A, Cherry KM, Hoffman J, Wang S, *et al*. A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations. *Med. Image Comput. Comput. Assist. Interv. MICCAI.* 2014; 17: 520–527.

15. Ben-Cohen A, Diamant I, Klang E, Amitai M, Greenspan H. Fully convolutional network for liver segmentation and lesions detection. Deep Learning and Data Labeling for Medical Applications. 2016: 77–85. doi:10.1007/978-3-319-46976-8_9.

16. Zhou X, Takayama R, Wang S, Hara T, Fujita H. Deep learning of the sectional appearances of 3D CT images for anatomical structure segmentation based on an FCN voting method. *Med. Phys.* 2017; 44(10): 5221–5233. doi: 10.1002/mp.12480 .

17. Zhou X, Ito T, Takayama R, Wang S, Hara T, Fujita H, Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting. *Deep Learning and Data Labeling for Medical Applications.* 2016: 111–120. doi: 10.1007/978-3-319-46976-8_12 .

18. Sobhaninia Z, Rezaei S, Noroozi A, Ahmadi M, Zarrabi H, Karimi N, et al. Brain tumor segmentation using deep learning by type specific sorting of images. 2018. ArXiv:1809.07786 [Cs, Eess].

19. Cheng R, Lay N, Mertan F, Turkbey B, Roth HR, Lu L, et al. Deep learning with orthogonal volumetric HED segmentation and 3D surface reconstruction model of prostate MRI. in: Proceeding of the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE. 2017: 749–753. doi: 10.1109/ISBI.2017.7950627 .

20. Banerjee S, Mitra S, Shankar BU. Multi-planar spatial-ConvNet for segmentation and survival prediction in brain cancer. Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. 2019: 94–104. doi: 10.1007/978-3-030-11726-9_9 .

21. Hänsch A, Schwier M, Morgas T, Klein J, Hahn HK, Gass T, et al. Comparison of different deep learning approaches for parotid gland segmentation from CT images. Medical Imaging 2018: Computer- Aided Diagnosis, SPIE. 2018: 44. doi: 10.1117/12.2292962 .

22. Chen J, Yang L, Zhang Y, Alber M, Chen D. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. Conference on Neural Information Processing Systems (NIPS 2016). 2016.

23. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. 2018: 120–129. doi:10.1007/978-3-319-75541-0_13 .

24. Vallaeys K, Kacem A, Legoux H, Tenier ML, Hamitouche C, Arbab-Chirani R, 3D dentomaxillary osteolytic lesion and active contour segmentation pilot study in CBCT: semi-automatic vs manual methods. Dentomaxillofac. Radiol. 2015; 44: 20150079. doi:10.1259/dmfr.20150079.

25. dos Santos RMG, De Martino JM, Passeri LA, Attux RRdF, Neto FH. Automatic repositioning of jaw segments for three-dimensional virtual treatment planning of orthognathic surgery. J. Cranio Maxillofac. Surg. 2017; 45(9): 1399–1407. doi: 10.1016/j.jcms.2017.06.017 .

26. Todorov G, Nikolov N, Sofronov Y, Gabrovski N, Laleva M, Gavrilov T. Computer aided design of customized implants based on CT-Scan data and virtual prototypes. Future Access Enablers for Ubiquitous and Intelligent Infrastructures. 2019: 339–346. doi:10.1007/978-3-030-23976-3_30.

27. Verhelst PJ, Shaheen E, Vasconcelos KdF, Cruyssen FVd, Shujaat S, Coudyzer W, et al. Validation of a 3D CBCT-based protocol for the follow-up of mandibular condyle remodeling. Dentomaxillofac. Radiol. 2019: 20190364, doi:10.1259/dmfr.20190364 .

28. Schulze R, Heil U, Gross D, Bruelmann D, Dranischnikow E, Schwanecke U, et al. Artefacts in CBCT: a review. Dentomaxillofacial Radiology 2011; 40(5): 265-273. doi: 10.1259/dmfr/30642039.

29. Minnema J, van Eijnatten M, Hendriksen AA, Liberton NPTJ, Pelt DM, Batenburg KJ, et al. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. Med. Phys. 2019; 46(11): 5027–5035, doi: 10.1002/mp.13793 .

30. Tian S, Dai N, Zhang B, Yuan F, Yu Q, Cheng X. Automatic classification and segmentation of teeth on 3D dental model using hierarchical deep learning networks. IEEE Access. 2019; 7: 84817–84828. doi: 10.1109/ACCESS.2019.2924262 .

31. Pelt DM. foam_ct_phantom. Github. https://github.com/conda-forge/foam _ ct _ phantom-feedstock.

32. Hendriksen AA, Pelt DM, Palenstijn WJ, Coban SB, Batenburg KJ. On-the- fly machine learning for improving image resolution in tomography. Appl. Sci. 2019; 9: 2445. doi: 10.3390/app9122445 .

33. van Aarle W, Palenstijn WJ, Cant J, Janssens E, Bleichrodt F, Dabravolski A, et al. Fast and flexible X-ray tomography using the ASTRA toolbox. Opt. Express. 2016; 24(22): 25129. doi: 10.1364/OE. 24.025129 .

34. van Aarle W, Palenstijn WJ, Beenhouwer JD, Altantzis T, Bals S, Batenburg KJ, et al. The Astra toolbox: a platform for advanced algorithm development in electron tomography. Ultramicroscopy. 2015;157:35–47. doi: 10.1016/ j.ultramic.2015.05.002 .

35. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. J. Opt. Soc. Am. 1984; 1(6): 612-619. doi:10.1364/JOSAA.1.000612.

36. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, et al. 3D slicer as an image computing platform for the quantitative imaging network. Magn. Reson. Imag. 2012; 30(9): 1323–1341. doi:10. 1016/j.mri.2012.05.001.

4

4

37. 3D Slicer. 2018. http://www.slicer.org.

38. Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. *Med. Phys.* 2019;46(5): 2169–2180. doi:10.1002/mp.13466 .

39. Chen L, Wu Y, Dsouza AM, Abidin AZ, Wismüller A, Xu C. MRI tumor segmentation with densely connected 3D CNN. *Medical Imaging 2018: Image Processing, SPIE.* 2018: 50. doi: 10.1117/12.2293394.

40. Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Med. Image Anal.* 2019; 53: 142–155. doi:10.1016/j.media.2019.02.005 .

41. Qiu B, Guo J, Kraeima J, Borra RJH, Witjes MJH, Ooijen PMAV. 3D segmentation of mandible from multisectional CT scans by convolutional neural networks. 2018. http://arxiv.org/abs/1809.06752 .

42. Klein A, Warszawski J, Hillengaß J, Maier-Hein KH. Automatic bone segmentation in whole-body CT images. *Int. J. CARS.* 2019;14(1): 21–29. doi:10.1007/s11548-018-1883-7 .

43. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci.* 2018; 115: 254–259. doi:10.1073/pnas.1715832114 .

44. Hendriksen AA. ahendriksen/msd_pytorch: v0.7.2, Zenodo, 2019. https://doi.org/ 10.5281/ ZENODO.3560114 .

45. Hendriksen AA. On the fly. *Github.* https:// github.com/ahendriksen/on _ the _ fly .

46. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. ArXiv:1502.03167 [Cs]. 2015. http://arxiv. org/abs/1502.03167 (accessed June 21, 2019).

47. Kingma DP, Ba J, Adam: a method for stochastic optimization. ArXiv:1412.6980 [Cs]. 2015. http://arxiv.org/abs/1412.6980.

48. Mlynarski P, Delingette H, Alghamdi H, Bondiau PY, Ayache N. Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy. *J. Med. Imaging.* 2020; 7(1): 014502. doi:10.1117/1.JMI.7.1.014502 .

49. Desai AD, Gold GE, Hargreaves BA, Chaudhari AS. Technical considerations for semantic segmentation in MRI using convolutional neural networks. ArXiv:1902.01977 [Cs, Eess]. 2019. http://arxiv.org/abs/1902.01977 .

50. Vu MH, Grimbergen G, Nyholm T, Löfstedt T. Evaluation of multi-slice inputs to convolutional neural networks for medical image segmentation. ArXiv:1912.09287 [Cs, Eess, Stat]. 2019. http://arxiv.org/abs/1912.09287 .

51. Zhang Y, Liao Q, Zhang J. Exploring efficient volumetric medical image segmentation using 2.5D method: an empirical study. ArXiv:2010.06163 [Cs, Eess]. 2020. http://arxiv.org/abs/2010.06163 .

52. Mlynarski P, Delingette H, Criminisi A, Ayache N. 3D convolutional neural networks for tumor segmentation using long-range 2D context. *Comput. Med. Imaging Gr.* 2019; 73: 60–72. doi:10.1016/j.compmedimag.2019.02.001 .

53. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, *et al.* nnU-Net: self-adapting framework for U-Net-based medical image segmentation. ArXiv:1809.10486 [Cs]. 2018. http://arxiv.org/abs/1809.10486.

CHAPTER 5

SEGMENTATION OF DENTAL CONE-BEAM CT SCANS AFFECTED BY METAL ARTIFACTS USING A MIXED-SCALE DENSE CONVOLUTIONAL NEURAL NETWORK

Jordi Minnema, Maureen van Eijnatten, Allard A. Hendriksen,
Niels Liberton, Daniël M. Pelt, K. Joost Batenburg,
Tymour Forouzanfar, Jan Wolff

ABSTRACT

Purpose

In order to attain anatomical models, surgical guides and implants for computer-assisted surgery, accurate segmentation of bony structures in cone-beam computed tomography (CBCT) scans is required. However, this image segmentation step is often impeded by metal artifacts. Therefore, the current study aimed to develop a mixed-scale dense convolutional neural network (MS-D network) for bone segmentation in CBCT scans affected by metal artifacts.

5

Method

Training data were acquired from 20 dental CBCT scans affected by metal artifacts. An experienced medical engineer segmented the bony structures in all CBCT scans using global thresholding and manually removed all remaining noise and metal artifacts. The resulting gold standard segmentations were used to train an MS-D network comprising 100 convolutional layers using far fewer trainable parameters than alternative convolutional neural network (CNN) architectures. The bone segmentation performance of the MS-D network was evaluated using a leave-2-out scheme and compared with a clinical snake evolution algorithm and two state-of-the-art CNN architectures (U-Net and ResNet). All segmented CBCT scans were subsequently converted into standard tessellation language (STL) models and geometrically compared with the gold standard.

Results

CBCT scans segmented using the MS-D network, U-Net, ResNet and the snake evolution algorithm demonstrated mean Dice similarity coefficients of 0.87 ± 0.06, 0.87 ± 0.07, 0.86 ± 0.05 and 0.78 ± 0.07, respectively. The STL models acquired using the MS-D network, U-Net, ResNet and the snake evolution algorithm demonstrated mean absolute deviations of 0.44 mm ± 0.13 mm, 0.43 mm ± 0.16 mm, 0.40 mm ± 0.12 mm and 0.57 mm ± 0.22 mm, respectively. In contrast to the MS-D network, the ResNet introduced wave-like artifacts in the STL models, whereas the U-Net incorrectly labelled background voxels as bone around the vertebrae in 4 of the 9 CBCT scans containing vertebrae.

Conclusion

The MS-D network was able to accurately segment bony structures in CBCT scans affected by metal artifacts.

1. INTRODUCTION

The spatial information embedded in medical three-dimensional (3D) images is being increasingly used to personalize treatment by means of computer-assisted surgery (CAS) [1]. This new field of medicine encompasses virtual surgical planning [2], 3D printing of personalized constructs [3], such as anatomical models, surgical saw guides, or implants [4], virtual and augmented reality [5,6] and robot-guided surgery [7]. The use of such emerging technologies in medicine has resulted in better treatment outcomes and a reduction in both operating times and costs[3,8,9]. In recent years, CAS has reached a state of high technology readiness in maxillofacial surgery, where cone-beam computed tomography (CBCT) is rapidly becoming the imaging modality of choice due to the low costs and radiation dose [10].

An essential step in the maxillofacial CAS workflow is the conversion of these CBCT images into a virtual 3D model of the anatomical region of interest [11]. This conversion process requires accurate segmentation of bony structures in dental CBCT images [12]. Image segmentation is, however, often impeded by metal artifacts [13]. Such artifacts are caused by high-density metal objects, such as amalgam fillings, crowns, dental implants, and retainers [14]. The presence of high-density metal objects in the radiation beam path induces photon starvation and scattering that lead to characteristic bright and dark streak artifacts in the resulting CBCT images (see Fig. 1) [15]. These streak artifacts can obscure anatomical structures and reduce the contrast between adjacent regions [16], and thereby impede the segmentation process of the teeth and bony structures in the mandible and maxilla.

To overcome these challenges, various metal artifact reduction (MAR) methods have been proposed. Such methods commonly aim to reduce metal artifacts during the reconstruction phase of CBCT scans [17]. More specifically, an initial CBCT reconstruction is performed, followed by the segmentation of the metal structures and the removal of the segmented metal structures from the sinogram. Thereafter, a new reconstruction is performed based on the corrected sinogram, which results in a reduced incidence of metal artifacts in the reconstructed CBCT scan [18]. However, the performance of such MAR methods depends strongly on the quality of the initial metal artifact segmentation [19], and is often limited by the introduction of secondary artifacts [18,20] and incomplete metal artifact correction [21]. As a consequence, metal artifacts remain a challenge in CAS.

In recent years, deep learning has been increasingly used for MAR. The majority of these approaches are based on convolutional neural networks (CNNs). CNNs can learn to extract information from a large number of training images to perform certain tasks in the MAR workflow, such as CBCT sinogram correction [22–24]. In a recent study by Zhang and Yu (2018), a CNN-based MAR framework was developed that fused the information from original and corrected MDCT images to suppress metal artifacts [25]. These corrected MDCT images were obtained by combining multiple conventional MAR methods. A major drawback of such MAR frameworks is that they first need to be trained using two sets of images of the same patient — one set of artifact free images and one set of images affected by artifacts. Since such paired datasets are often unavailable in clinical settings, most deep learning-based MAR methods rely on mathematical simulations of

metal artifacts that typically do not fully represent the photon and detector physics of individual MDCT or CBCT scanners.

Instead of relying solely on such mathematical simulations, it is also possible to use deep learning for MAR during the CBCT image segmentation step. A major advantage of CNNs is that they can be efficiently trained using high quality "gold standard" CBCT segmentations of teeth and bony structures created by human experts during CAS. To date, a variety of CNN architectures have been proposed for medical image segmentation [26]. However, since it is relatively difficult to acquire a sufficient number of gold standard segmentations in clinical settings, it is important to choose a CNN architecture with few trainable parameters that can be trained using few datasets. Therefore, in this study, the authors for the first time employed a novel mixed-scale dense CNN (MS-D network) architecture [27] to segment dental CBCT scans affected by metal artifacts. Furthermore, the performance of this MS-D network was compared with two state-of-the-art CNN architectures, namely U-Net [28] and ResNet [29] In addition, a clinical snake evolution algorithm [30] that is commonly used for medical image segmentation was evaluated. Specifically, the main contributions of this study are as follows:

1. CNNs were used to deal with metal artifacts in dental CBCT scans during image segmentation, rather than image reconstruction.
2. A novel mixed-scale dense CNN was trained on a relatively small dataset of dental CBCT images.
3. The mixed-scale dense CNN resulted in comparable segmentation performances as U-Net and ResNet CNN architectures, while using far fewer trainable parameters.
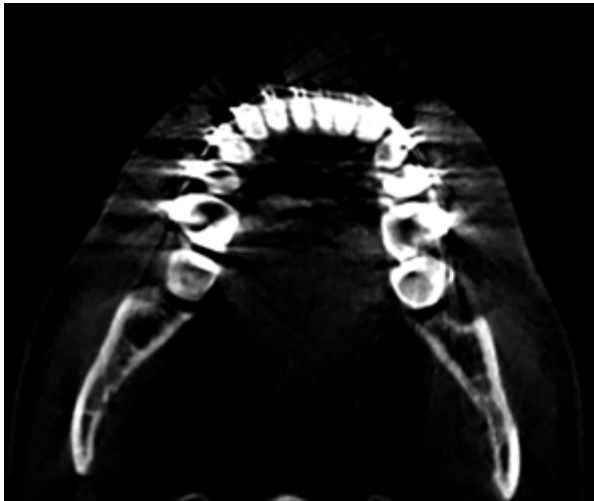4. All CNNs outperformed a widely-used clinical snake evolution method.

Figure 1. Example of metal artifacts in a CBCT image of the mandible.

2. MATERIALS AND METHODS

2.1. Data acquisition

A total of 20 dental CBCT scans that had been heavily affected by metal artifacts caused by dental restorations and appliances were used in this study. Of these CBCT scans, 2 were used for validation (see section 2.3. Implementation and training details) and 18 were used for training (see section 2.4. Evaluation). All scans were obtained on a Vatech PaX-Zenith3D (Vatech, Gyeonggi-do, South-Korea) CBCT scanner using a tube voltage of 105 kVp, a tube current of 6 mA and an isotropic voxel size of 0.2 mm. Each CBCT scan was cropped to a confined region of interest that included the lower part of the maxilla, the mandible and both condyles, resulting in variable scan dimensions ranging from 800 × 412 × 190 (patient 9) to 1000 × 724 × 383 (patient 10). All CBCT scans were normalized by subtracting the mean voxel value of the training CBCT scans and dividing the resulting values by the standard deviation.

In order to train a CNN for bone segmentation in these CBCT scans, gold standard segmentation labels were required. These gold standard labels were created by segmenting all CBCT scans using global thresholding, followed by extensive manual postprocessing by an experienced medical engineer using Mimics software (Mimics v20.0, Materialise, Leuven, Belgium). This postprocessing step was necessary to remove the noise and metal artifacts caused by dental fillings and appliances. This task took approximately 2 h per scan to complete.

2.2. CNN architecture

In this study, we used a mixed-scale dense CNN (MS-D network) architecture originally proposed by Pelt and Sethian [27]. This network architecture combines small- and large-scale features with far fewer trainable parameters compared with state-of-the-art U-Net architectures [28]. These properties enable an MS-D network to be trained more efficiently and reduce the risk of overfitting [27].
A schematic overview of an MS-D network architecture with three convolutional layers is presented in Figure 2. Note that the MS-D network architecture used in this study comprised 100 convolutional layers. Each convolutional layer performs a convolutional operation on its input to produce an intermediate image, also known as a feature map. All feature maps are used to compute the final output segmentation.

A feature map z_i in convolutional layer i is calculated as

$$z_i = \sigma(g_i(\{z_0,...,z_{i-1}\}) + b_i) \tag{1}$$

where σ is a rectified linear unit (ReLU) activation function [31] and b_i is a constant bias term. The function g_i performs a 2D 3x3 dilated convolution $D_{h,s}$ on all previously computed feature maps $\{z_0,...,z_{i-1}\}$ and sums the resulting feature maps in a pixel-wise manner, giving

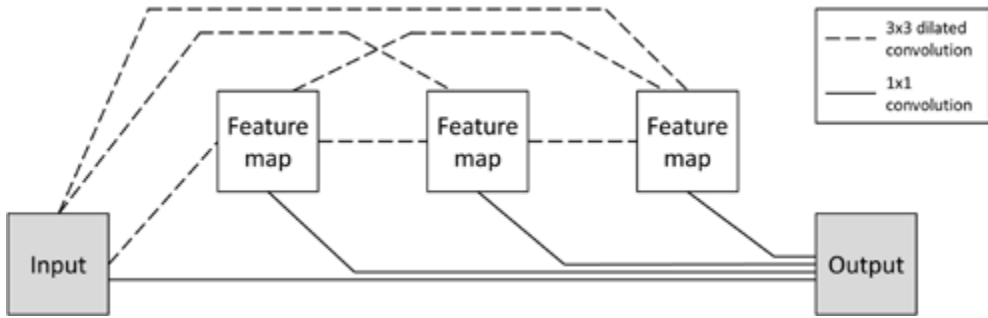$$g_i(\{z_0,...,z_{i-1}\}) = \sum_{j=0}^{i-1} D_{h_{ij},s_i} z_j \tag{2}$$

Figure 2. Schematic representation of an MS-D network architecture with 3 convolutional layers.

where h is a convolutional kernel and s is the dilation factor. In a dilated convolution, a kernel h is expanded by a dilation factor s and filled with zeros at distances that are not a multiple of s voxels from the kernel center. Thus, by increasing the dilation factor, the MS-D network is able to detect large features [27] without increasing the number of kernel weights [32]. In the present study, the dilation factor was initialized as 1 in the first convolutional layer, and then increased by 1 in each subsequent convolutional layer. After 10 convolutional layers, the dilation factor was reset to 1 and the process was repeated. This enabled the MS-D network to extract mixed-scale features from the input CBCT slices. In addition, all dilated convolutions were performed using reflective boundaries. As a result, the size and shape of all feature maps remained equal to those of the initial input. The major advantage of equally sized feature maps is that the convolutional layers are not restricted to using only the feature map of the previous layer to compute a new feature map. Instead, all previously computed feature maps, including the initial input, are used to compute a new feature map, resulting in a densely connected network (Fig. 2).

The output y of an MS-D network is computed by applying 1x1 convolutional kernels w_i to all previously computed feature maps $\{z_0,...,z_{i-1}\}$, adding a constant bias term b and applying a softmax activation function σ'. This can be written as follows:

$$y = \sigma'\left(\sum_i w_i z_i + b\right) \tag{3}$$

Since y is a continuous variable with values between 0 and 1, a cut-off value was required to obtain a binary segmentation (i.e., "bone" or "background"). In the present study, we treated this cut-off value as an additional hyper-parameter of the MS-D network.

2.3. Implementation and training details

The hyper-parameters of the MS-D network, that is, the number of layers and the cut-off value, were determined by validating the network on two CBCT scans. During these validation experiments, the number of layers was varied between 30 and 150 (30, 50, 80, 100 and 150), and the cut-off value was varied between 0.1 and 0.9 with a step size of 0.2. Optimal performance of the MS-D network

was achieved using 100 convolutional layers and a cut-off value of 0.7. The validation dataset was also used to find the optimal number of epochs (10) for training.

The MS-D network was implemented in Python (version 3.6.1) using Pytorch (version 0.3.1). Training of the MS-D network was performed on 2D axial CBCT slices using a batch size of 1 and the default Adam optimizer [33] on a Linux desktop computer (HP Workstation Z840) with 64 GB RAM, a Xeon E5-2687 v4 3.0 GHZ CPU and a GTX 1080 Ti GPU card. Training took approximately 5 h for each epoch.

2.4. Evaluation

The segmentation performance of the MS-D network was evaluated using the 18 CBCT scans available for training (see Section 2.1. Data acquisition). A leave-2-out scheme [34] was used so that 16 of the 18 training CBCT scans were alternately used for training and 2 for testing. As a clinical benchmark, these 18 CBCT scans were also segmented using a snake evolution algorithm that is commonly used for various clinical segmentation purposes [35-38]. This algorithm is available in the open-source ITK-SNAP software package [30] and requires an initial segmentation using global thresholding, followed by selection of seed points in the region of interest (i.e., bone).

In addition, the performance of the MS-D network was compared to two state-of-the-art CNN architectures available on Github, namely U-Net [39] and ResNet [40]. The U-Net used in this study is comparable to the one described by Ronneberger et al. [40], except that our implementation performed batch normalization [[32]] after each ReLU and used reflection padding on images of which the dimensions were not divisible by 16. The ResNet used in this study was a residual network comprising 50 layers as described by He et al. [40]. Both CNNs were trained using 4 epochs and a cut-off value of 0.3.

The segmentation performance of all three CNNs and the clinical snake evolution algorithm was evaluated using the Dice similarity coefficient (DSC). The DSC indicates the overlap between a segmented CBCT scan and the corresponding gold standard segmentation. This can be written as follows:

$$DSC \ = \ \frac{2TP}{2TP \ + \ FP \ + \ FN} \tag{4}$$

where TP is the number of true positives, FP is the number of false positives and FN is the number of false negatives.

All segmented CBCT scans and corresponding gold standard segmentations were subsequently converted into virtual 3D models in the standard tessellation language (STL) file format using 3D Slicer software [42,43] The resulting STL models were geometrically compared with the corresponding gold standard STL models using the surface comparison function in GOM Inspect software (GOM Inspect 2018, GOM GmbH, Braunschweig, Germany). Signed deviations between 5.0 and +5.0 mm were measured between the acquired STL models and the gold standard STL models. The mean absolute deviations (MADs) were calculated for all STL models.

3.   RESULTS

In all CBCT scans affected by metal artifacts, the MS-D network resulted in fewer erroneously labeled voxels in the dental region than the snake evolution algorithm (Fig. 3). Moreover, the MS-D network resulted in a more complete segmentation of the condyles and the rami than the snake evolution algorithm in 13 of the 18 CBCT scans (Fig. 3). Furthermore, in 8 out of 9 CBCT scans that contained parts of the vertebrae, the MS-D network segmented these vertebrae, whereas the snake evolution algorithm incorrectly labeled the vertebrae as the background in all 9 CBCT scans. Quantitatively, the snake evolution algorithm and the MS-D network resulted in a mean DSC of 0.78 ± 0.07 and 0.87 ± 0.06, respectively (Table 1). U-Net and ResNet resulted in mean DSCs of 0.87 ± 0.07 and 0.86 ± 0.05, respectively.

Generally, all STL models acquired using the CNNs, i.e., the MS-D network, U-Net and ResNet, contained fewer outliers in the dental region than the STL models acquired using the snake evolution algorithm (Fig. 4). However, in contrast to the MS-D network, the ResNet introduced wave-like artifacts in all 18 STL models (Fig. 4), whereas the U-Net incorrectly labeled
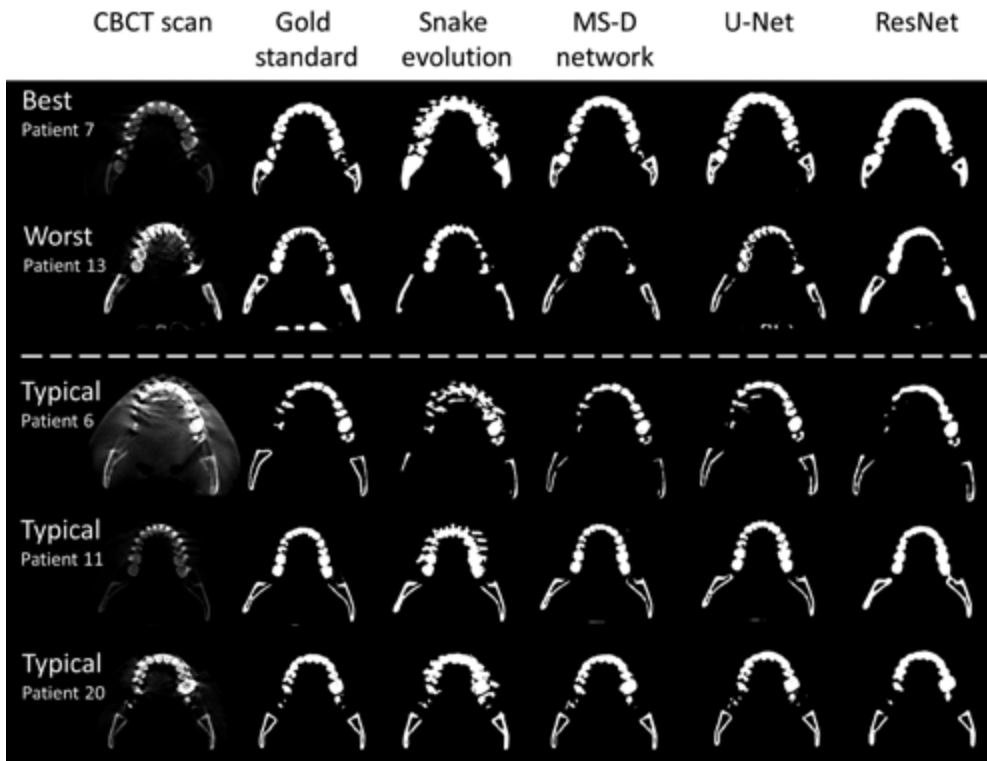

Figure 3. Examples of the best (patient 7) and worst (patient 13) MS-D network performances and three typical examples (patients 6, 11, and 20). Each example comprises an axial CBCT slice, the gold standard (manual) segmentation, and the segmentations acquired using the snake evolution algorithm, MS-D network, U-Net and ResNet.

background voxels as bone around the vertebrae in 4 of the 9 CBCT scans containing vertebrae (Fig. 4; patients 7, 13 and 20).

Figure 5 visualizes the surface deviations between all STL models and their corresponding gold standard STL models. In 11 of the 18 patients, the 10-90 percentile range acquired using the snake evolution algorithm was larger than those acquired using the CNNs. When compared with the gold standard STL models, the STL models acquired using the MS-D network resulted in a mean MAD of 0.44 ± 0.13 mm; whereas the STL models acquired using the snake evolution algorithm resulted in a mean MAD of 0.57 ± 0.22 mm. The STL models acquired using U-Net and ResNet resulted in mean MADs of 0.43 ± 0.16 mm and 0.40 ± 0.12 mm, respectively.

4. DISCUSSION

High-density metal fillings and appliances are very common in the oral cavity. For example, more than half of the American population has at least one dental filling and approximately 25% are estimated to have more than 7 fillings [44]. Consequently, metal artifacts caused by such objects
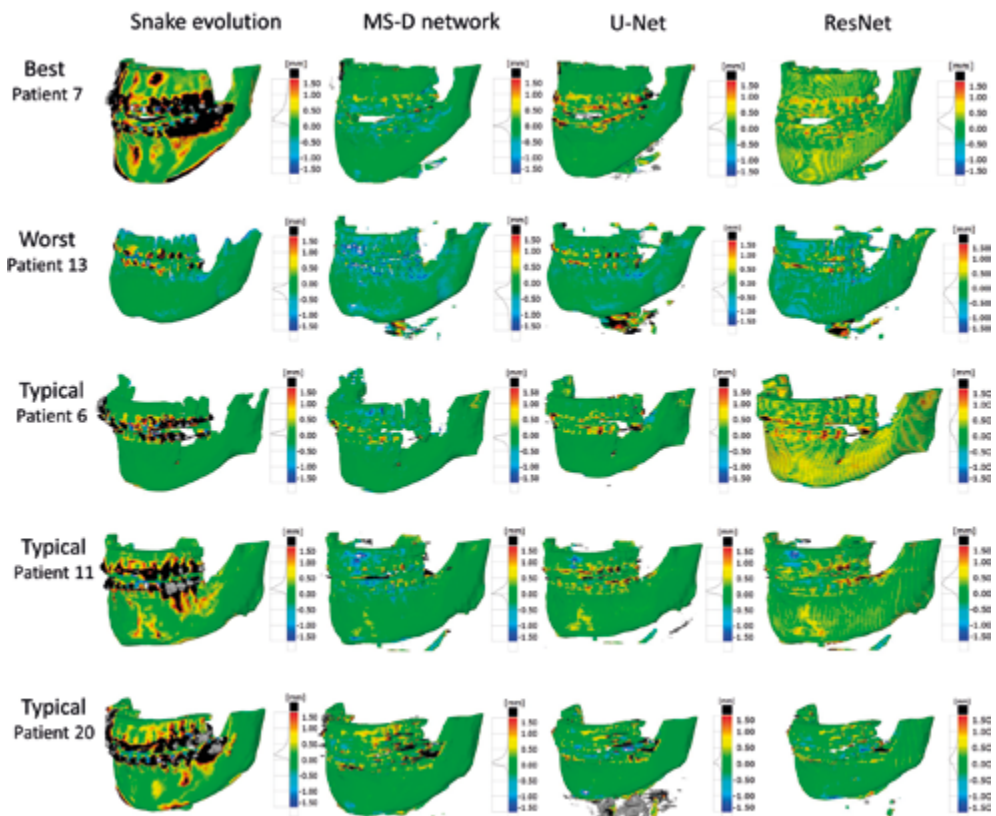

Figure 4. Color maps of the surface deviations of five STL models acquired using the snake evolution algorithm, MS-D network, U-Net and ResNet. All depicted surface deviations were calculated with respect to the corresponding gold standard STL model.
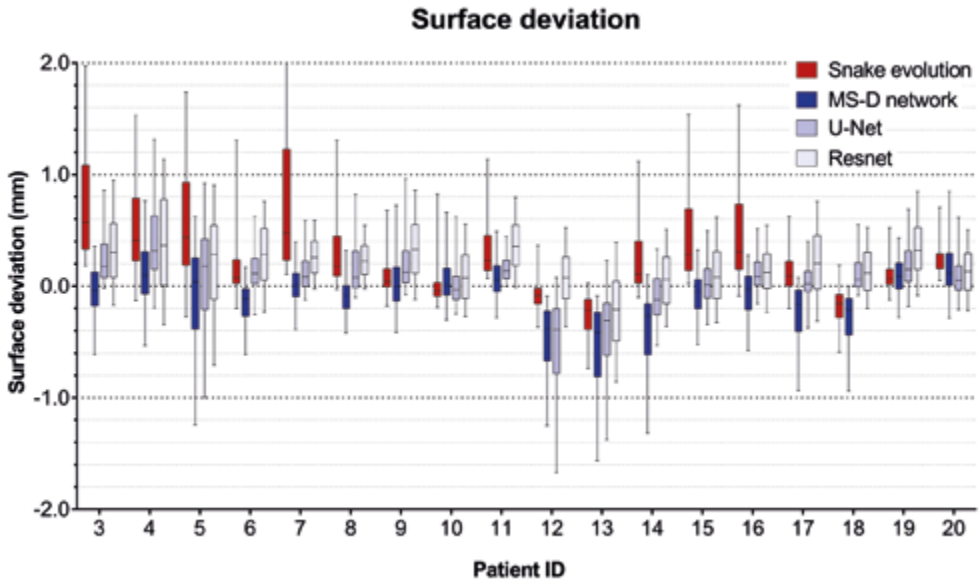
Figure 5. Box and whisker plot of the surface deviations between the STL models acquired using the snake evolution algorithm, MS-D network, U-Net, ResNet, and the corresponding gold standard STL models. The boxes represent the interquartile range and the whiskers represent the 10th and 90th percentiles of the surface deviations.

remain a challenge in CBCT imaging. Such artifacts can obscure bony regions in the mandible and maxilla and can lead to inaccuracies and time constraints during the image segmentation process required for computer-assisted maxillofacial surgery.

All CNNs trained in this study (MS-D network, U-Net and ResNet) were able to segment bony structures in CBCT scans and classify metal artifacts as background more accurately than the current clinical benchmark, i.e., the snake evolution algorithm (Fig. 3 and Table 1). This finding is likely due to the CNNs' ability to learn characteristic features that distinguish bone from metal artifacts. The snake evolution algorithm, on the other hand, is a model-driven segmentation method that is solely based on identifying intensity gradients in images. Although such intensity-based image segmentation methods generally perform well in identifying the edges of bony structures in CBCT images [35], they tend to fail in the presence of metal artifacts due to the introduction of strong intensity gradients in the reconstructed CBCT images.

The DSCs found in our study (Table 1) are comparable to those reported by Wang et al. (2015), who used a prior-guided random forest to segment the maxilla and mandible in 30 CBCT scans and reported a mean DSC of 0.91 ± 0.03 for the maxilla and 0.94 ± 0.02 for the mandible [45]. However, their dataset only included 4 CBCT scans that were affected by metal artifacts. Evain et al. (2017) recently developed a graph-cut approach for the segmentation of individual teeth [46]. Although their algorithm achieved a high mean DSC of 0.958 ± 0.023, they also reported that false edges were induced in images affected by metal artifacts [46].

Table 1. Dice similarity coefficients (DSCs) of all CBCT scans segmented using the snake evolution algorithm, MS-D network, U-Net and ResNet.

Patient ID	Snake evolution	MS-D network	U-Net	ResNet
1	Validation set	Validation set	Validation set	Validation set
2	Validation set	Validation set	Validation set	Validation set
3	0.67	0.89	0.86	0.82
4	0.72	0.85	0.79	0.75
5	0.60	0.75	0.78	0.78
6	0.80	0.85	0.88	0.82
7	0.76	0.94	0.93	0,90
8	0.83	0.92	0.91	0.90
9	0.80	0.89	0.87	0.78
10	0.90	0.92	0.92	0.90
11	0.84	0.92	0.92	0.85
12	0.84	0.78	0.69	0.90
13	0.73	0.73	0.78	0.82
14	0.86	0.83	0.88	0.90
15	0.75	0.91	0.90	0.88
16	0.77	0.91	0.92	0.90
17	0.76	0.86	0.91	0.86
18	0.81	0.88	0.94	0.92
19	0.86	0.90	0.87	0.82
20	0.79	0.91	0.90	0.90
Mean	**0.78 ± 0.07**	**0.87 ± 0.06**	**0.87 ± 0.07**	**0.86 ± 0.05**

As an additional evaluation step in our study, all segmented CBCT scans were converted into STL models and geometrically compared with the corresponding gold standard STL models. Interestingly, fewer outliers were observed in the STL models acquired using the CNNs than in the STL models acquired using the snake evolution algorithm (Figs. 4 and 5). The MADs acquired in the present study are smaller than those obtained by Lamecker et al. (2006), who developed a statistical shape model for the segmentation of the mandible in CBCT scans and found mean surface deviations larger than 1 mm, even though they excluded all teeth from statistical analysis due to severe metal artifacts [47] The MADs obtained in this study are, however, higher than those reported by Gan et al. (2014), who segmented individual teeth in CBCT scans using a level-set method and achieved a MAD of 0.3 ± 0.08 mm [48]. Nevertheless, it must be noted that Gan et al. did not include any CBCT scans affected by metal artifacts because their level-set algorithm failed to identify teeth contours in these scans.

The novel MS-D network resulted in accurate segmentations that were comparable to those achieved by U-Net and ResNet, using fewer trainable parameters (Table 2). Reducing the number of parameters is crucial in clinical settings since it minimizes the risk of overfitting [27] and prevents common deep learning issues such as vanishing gradients and local minima [49]. Another major advantage of MS-D networks over U-Net and ResNet is the use of dilated convolutional kernels

instead of standard convolutional kernels. This allows MS-D networks to learn which combinations of dilations are most suited to solve the task at hand and offers the unique possibility to use the same MS-D network architecture for a broad range of different applications such as segmenting organelles in microscopic cell images [27], image denoising [27] and improving the resolution of tomographic reconstructions [50]. Finally, all layers of an MS-D network are interconnected using the same set of standard operations [see Section 2.2, Eqs. (1) and (2)], which greatly simplifies implementation and training of an MS-D network in clinical settings [27].

Another interesting finding in this study was that the MS-D network was able to accurately segment bony regions that were not affected by metal artifacts, such as the medial parts of the rami, the condyles and the vertebrae (Fig. 3; patients 6 and 13). On the other hand, the segmentations obtained using ResNet demonstrated less anatomical details (Figs. 3 and 4), which can result in ill-fitting personalized constructs during CAS [4]. Furthermore, the segmentations obtained using U-Net were less accurate in the vicinity of the vertebrae when compared to those obtained using the MS-D network. A possible explanation for this phenomenon is that the MS-D network was better capable to learn features of relatively rare structures in the training dataset such as the vertebrae. These results demonstrate that the MS-D network is well suited for "real-world" clinical segmentation purposes.

An important advantage of all three CNNs evaluated in this study over alternative clinical segmentation methods is the short computational time required for image segmentation. More specifically, all three CNNs automatically segmented each CBCT scan in <5 min. In comparison, the semi-automatic clinical snake evolution algorithm segmented a single CBCT scan in 20 min to 1 h. All CNN segmentation times found in this study are markedly quicker than the atlas-based method described by Wang et al. (2015) that segmented a single CBCT scan in 5 h [45]. Moreover, a previously published patch-based CNN for MDCT image segmentation of the skull took approximately 1 h to segment a single MDCT scan [51]. The short segmentation times of the CNNs in this study were primarily due to their fully-convolutional nature that allows the CNNs to segment CBCT images using far fewer convolutional operations than patch-based CNNs [26].

Taking the aforementioned advantages in terms of performance and speed into account, deep learning is now coming of age for medical image segmentation, especially with advanced architectures such as the MS-D network. The next step toward making deep learning-based solutions available for challenging image segmentation tasks in CAS would be to develop, test and certify interactive plugins for medical image processing software packages.

Table 2. The number of trainable parameters used by the MS-D network, U-Net and ResNet.

CNN model	Number of trainable parameters
MS-D network	45,756
U-Net	14,787,842
ResNet	32,940,996

5. LIMITATIONS

One challenge that all supervised deep learning algorithms have in common is the overall accuracy of the gold standard segmentations. Especially the presence of metal artifacts can negatively influence the judgements of experienced medical engineers and subsequently affect the quality of their gold standard segmentations. Furthermore, the process of creating sufficient gold standard segmentations can be very time-consuming. One solution could be to adopt an iterative training strategy in which a pretrained CNN is used to perform an initial segmentation of a CBCT scan, after which a medical engineer only has to correct the errors and retrain the CNN. Another interesting direction for future research is the potential use of 3D CNNs due to the 3D characteristics of metal artifacts in dental CBCT scans.

6. CONCLUSION

This study presents a mixed-scale dense CNN (MS-D net-work) to segment teeth and bony structures in CBCT images heavily affected by metal artifacts. Experimental results demonstrated that the segmentation performance of the MS-D network was comparable to those of state-of-the-art U-Net and ResNet CNN architectures, while preserving more anatomical details in the resulting STL models and using fewer trainable parameters. Moreover, all CNNs outperformed a commonly used clinical snake evolution algorithm. These promising results show that deep learning offers unique possibilities to eliminate the inaccuracies caused by metal artifacts in the CAS workflow.

ACKNOWLEDGEMENTS

ETHICAL CONSIDERATIONS

This study followed the principles of the Helsinki Declaration and was performed in accordance with the guidelines of the Medical Ethics Committee of the Amsterdam UMC. The Dutch Medical Research Involving Human Subjects Act (WMO) did not apply to this study (Ref: 2017.145).

REFERENCES

1. Swennen GRJ, Mollemans W, Schutyser F. Three-dimensional treatment planning of orthognathic surgery in the era of virtual imaging. *J Oral Maxillofac Surg*. 2009;67(10):2080–2092.

2. Zhao L, Patel PK, Cohen M. Application of virtual surgical planning with computer assisted design and manufacturing technology to cranio-maxillofacial surgery. *Arch Plast Surg*. 2012;39(4):309.

3. Ventola CL. Medical applications for 3D printing: current and projected uses. *P & T*. 2014;39(10):704–711.

4. Stoor P, Suomalainen A, Lindqvist C, Mesimäki K, Danielsson D, Westermark A, *et al*. Rapid prototyped patient specific implants for reconstruction of orbital wall defects. *J Cranio-Maxillofac Surg*. 2014;42(8):1644–1649.

5. Badiali G, Ferrari V, Cutolo F, Freschi C, Caramella D, Bianchi A, *et al*. Augmented reality as an aid in maxillofacial surgery: validation of a wearable system allowing maxillary repositioning. *J Cranio-Maxillofac Surg*. 2014;42(8):1970–1976.

6. Jiang T, Zhu M, Chai G, Li Q. Precision of a novel craniofacial surgical navigation system based on augmented reality using an occlusal splint as a registration strategy. *Sci Rep*. 2019;9:501.

7. van Dijk JD, van den Ende RPJ, Stramigioli S, Köchling M, Höss N. Clinical pedicle screw accuracy and deviation from planning in robot-guided spine surgery: robot-guided pedicle screw accuracy. *Spine*. 2015;40:E986–E991.

8. Haas OL Jr, Becker OE, de Oliveira RB. Computer-aided planning in orthognathic surgery—systematic review. *Int J Oral Maxillofac Surg*. 2015;44:329–342.

9. Lonic D, Lo L-J. Three-dimensional simulation of orthognathic surgery—surgeon's perspective. *J Formos Med Assoc*. 2016;115(6):387–388.

10. Shweel M, Amer MI, El-Shamanhory AF. A comparative study of cone-beam CT and multidetector CT in the preoperative assessment of odonto-genic cysts and tumors. *Egypt J Radiol Nucl Med*. 2013;44(1):23–32.

11. Hieu LC, Zlatov N, Vander Sloten J, Bohez E, Khanh L, Binh PH, *et al*. Medical rapid prototyping applications and methods. *Assembly Automat*. 2005;25:284–292.

12. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phys*. 2018;51:6–16.

13. Pauwels R, Araki K, Siewerdsen JH, Thongvigitmanee SS. Technical aspects of dental CBCT: state of the art. *Dentomaxillofacial Radiology*. 2014;44(1):20140224.

14. De Man B, Nuyts J, Dupont P, Marchal G, Suetens P. Metal streak arti-facts in X-ray computed tomography: a simulation study. *IEEE Trans Nucl Sci*. 1999;46(3):691–696.

15. Barrett JF, Keat N. Artifacts in CT: recognition and avoidance. *Radiographics*. 2004;24(6):1679–1691.

16. Mori I, Machida Y, Osanai M, Iinuma K. Photon starvation artifacts of X-ray CT: their true cause and a solution. *Radiol Phys Technol*. 2013;6:130–141.

17. Gjesteby L, De Man B, Jin Y, Paganetti H, Verburg J, Giantsoudi D, *et al*. Metal artifact reduction in CT: where are we after four decades? *IEEE Access*. 2016;4:5826–5849.

18. Hahn A, Knaup M, Brehm M, Sauppe S, Kachelrieß M. Two methods for reducing moving metal artifacts in cone-beam CT. *Med Phys*. 2018;45(8):3671–3680.

19. Meilinger M, Schmidgunst C, Schütz O, Lang EW. Metal artifact reduction in cone beam computed tomography using forward projected reconstruction information. *Med Phys*. 2011;21(3):174–182.

20. Diehn FE, Michalak GJ, DeLone DR, DeLone DR, Kotsenas AL, Lindell EP, *et al*. CT Dental artifact: comparison of an iterative metal artifact reduction technique with weighted filtered back-projection. *Acta Radiol Open*. 2017;6(11):205846011774327.

21. Hegazy MAA, Cho MH, Lee SY. A metal artifact reduction method for a dental CT based on adaptive local thresholding and prior image generation. *Biomed Eng Online*. 2016;15(1):119.

22. Claus BEH, Jin Y, Gjesteby L, Wang G, De Man B. Metal Artifact Reduction Using Deep Learning Based Sinogram Completion: Initial Results. *Fully3D 2017 Proceedings*. 2017:631–635.

23. Ghani MU, Karl WC. Deep learning based sinogram correction for metal artifact reduction. *Elect Imaging*. 2018;2018:472-1–4728.

24. Gjesteby L, Yang Q, Xi Y, Zhou Y, Zhang J, Wang G. Deep learning methods to guide

CT image reconstruction and reduce metal artifacts. *Proceedings Volume 10132, Medical Imaging 2017: Physics of Medical Imaging*. 2017. doi:10.1117/12.2254091.

25. Zhang Y, Yu H. Convolutional neural network based metal artifact reduction in x-ray computed tomography. *IEEE Trans Med Imaging*. 2018;37(6):1370–1381.

26. Litjens G, Kooi T, Bejnordi BE, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal*. 2017;42:60-88.

27. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. Proc Natl Acad Sci. 2018;115(2):254–259.

28. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI. 2015:234–241.

29. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. https://doi.org/10.1109/ CVPR.2016.90

30. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, *et al*. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. *NeuroImage*. 2006;31(3):1116–1128.

31. Krizhevsky A, Sutskever I, Geoffrey EH. ImageNet classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. 2012;25:1–9.

32. Yu F, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions. 2015. http://arxiv.org/abs/1511. 07122.

33. Kingma DP, Ba J, Adam: a method for stochastic optimization. ArXiv:1412.6980 [Cs]. 2015. http://arxiv.org/abs/1412.6980.

34. Anguita D, Ghelardoni L, Ghio A, Oneto L, Ridella S. The "k" in k-fold cross validation. *Paper presented at: ESANN 2012 Proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning; April 25–27, 2012; Bruges, BE,* 2012: 441–446.

35. Fan Y, Beare R, Matthews H, Schneider P, Kilpatrick N, Clement J, *et al*. Marker-based watershed transform method for fully automatic mandibular segmentation from CBCT images. *Dentomaxillofac Radiol*. 2019;48(2):20180261.

36. Qian X, Wang J, Guo S, Li Q. An active contour model for medical image segmentation with application to brain CT image: an active contour model for medical image segmentation. *Med Phys*. 2013;40(2): 021911.

37. Saadatmand-Tarzjan M. Self-affine snake for medical image segmentation. *Pattern Recogn Lett*. 2015;59:1–10.

38. Vallaeys K, Kacem A, Legoux H, Le Tenier M, Hamitouche C, Arbab-Chirani R. 3D dento-maxillary osteolytic lesion and active contour segmentation pilot study in CBCT: semi-automatic vs manual methods. *Dentomaxillofac Radiol*. 2015;44:20150079.

39. Milesial. Pytorch-UNet. *Github*. 2019. https://github.com/milesial/Pytorch-UNet.

40. pytorch. Torchvision. *Github*. 2019. https://github.com/pytorch/vision/tree/master/torchvision.

41. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. ArXiv:1502.03167 [Cs]. 2015. http://arxiv. org/abs/1502.03167.

42. Fedorov A, Beichel R, Kalpathy-Cramer J, Finet J, Fillion-Robin JC, Pujol S, *et al*. 3D slicer as an image computing platform for the quantitative imaging network. *Magn. Reson. Imag.* 2012; 30(9): 1323–1341. doi:10. 1016/j.mri.2012.05.001.

43. 3D Slicer. 2018. http://www.slicer.org.

44. Yin L, Yu K, Lin S, Song X, Yu X. Associations of blood mercury, inorganic mercury, methyl mercury and bisphenol A with dental surface restorations in the U.S. population, NHANES 2003–2004 and 2010–2012. *Ecotoxicol Environ Saf*. 2016;134P1:213–225.

45. Wang L, Gao Y, Shi F, Li G, Chen KC, Tang Z, *et al*. Automated segmentation of dental CBCT image with prior-guided sequential random forests: automated segmentation of dental CBCT image. *Med Phys*. 2015;43(1):336–346.

46. Evain T, Ripoche X, Atif J, Bloch I. Semi-automatic teeth segmentation in Cone-Beam Computed Tomography by graph-cut with statistical shape priors. *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. 2017:1197–1200. https://doi.org/10.1109/ISBI.2017.7950731

47. Lamecker H, Zachow S, Wittmers A, Weber B, Hege HC, Elsholtz B, *et al*. Automatic

5

segmentation of mandibles in low-dose CT-data. *Int J Comput Assist Radiol Surg*. 2006;1:393–395.

48. Gan Y, Xia Z, Xiong J, Zhao Q, Hu Y, Zhang J. Toward accurate tooth segmentation from computed tomography images using a hybrid level set model: accurate tooth segmentation from computed tomography images. *Med Phys*. 2016;42:14–27.

49. Li H, Xu Z, Taylor G, Studer C, Goldstein T. Visualizing the Loss Land-scape of Neural Nets. *Advances in Neural Information Processing Systems 31 (NeurIPS). 2018:6389-6399*

50. Hendriksen AA, Pelt DM, Palenstijn WJ, Coban SB, Batenburg KJ. On-the-fly machine learning for improving image resolution in tomography. *Appl Sci*. 2019;9(12):2445.

51. Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput Biol Med*. 2018;103:130–139.

5

CHAPTER 6

EFFICIENT HIGH CONE-ANGLE ARTIFACT REDUCTION IN CIRCULAR CONE-BEAM CT USING DEEP LEARNING WITH GEOMETRY-AWARE DIMENSION REDUCTION

Jordi Minnema, Maureen van Eijnatten, Henri der Sarkissian, Shannon Doyle, Juha Koivisto, Jan Wolff, Tymour Forouzanfar, Felix Lucka, Kees Joost Batenburg

ABSTRACT

High cone-angle artifacts(HCAAs) appear frequently in circular cone-beam computed tomography (CBCT) images and can heavily affect diagnosis and treatment planning. To reduce HCAAs in CBCT scans, we propose a novel deep learning approach that reduces the three-dimensional (3D) nature of HCAAs to two-dimensional (2D) problems in an efficient way. Specifically, we exploit the relationship between HCAAs and the rotational scanning geometry by training a convolutional neural network (CNN) using image slices that were radially sampled from CBCT scans. We evaluated this novel approach using a dataset of input CBCT scans affected by HCAAs and high-quality artifact-free target CBCT scans. Two different CNN architectures were employed, namely U-Net and a mixed-scale dense CNN (MS-D Net). The artifact reduction performance of the proposed approach was compared to that of a Cartesian slice-based artifact reduction deep learning approach in which a CNN was trained to remove the HCAAs from Cartesian slices. In addition, all processed CBCT scans were segmented to investigate the impact of HCAAs reduction on the quality of CBCT image segmentation. We demonstrate that the proposed deep learning approach with geometry-aware dimension reduction greatly reduces HCAAs in CBCT scans and outperforms the Cartesian slice-based deep learning approach. Moreover, the proposed artifact reduction approach markedly improves the accuracy of the subsequent segmentation task compared to the Cartesian slice-based workflow.

1. INTRODUCTION

Circular cone-beam computed tomography (CBCT) is becoming an increasingly popular imaging modality in dentistry and maxillofacial surgery due to its short scanning times, low costs and low radiation doses compared to conventional computed tomography (CT) scanners [1]. A well-known problem in circular CBCT imaging is the high cone-angle artifact (HCAA), also referred to as the cone-beam artifact. These artifacts occur because projection data acquired in a circular x-ray source trajectory are not complete: the Tuy-Smith data sufficiency condition is not satisfied [2, 3]. This implies that it is impossible to exactly reconstruct data points outside the midplane, i.e. the plane in which x-ray source and projector rotate. Therefore, approximate methods such as the Feldkamp−David−Kress(FDK) [4] algorithm are commonly used. However, these approximations typically lead to streaking artifacts and image distortions that become more severe with increasing cone-angles [5]. Figure 1 presents a schematic overview of the circular CBCT image acquisition and reconstruction process of a walnut using the FDK algorithm.

Over the past decades, various approaches have been proposed to reduce HCAAs. For example, Grass et al. developed a modified FDK algorithm that rearranges the pixels in cone-beam projections into fan-shaped projections [6]. Another popular approach is to reconstruct CBCT scans using iterative reconstruction (IR) methods [7, 8]. Although IR algorithms generally result in fewer HCAAs [9] and higher signal-to-noise ratios [10] compared to FDK reconstruction, they require much longer reconstruction times, which limits their clinical feasibility [11, 12]. Hu et al. and Zhu et al. independently proposed cone-angle artifact reduction approaches in which missing CBCT projection data were estimated [13, 14]. However, this approach resulted in residual artifacts due to incorrect estimation of high-frequency Radon space data [14]. Another way to reduce HCAAs is the two-pass approach initially developed by Hsieh [15], in which high-density structures in FDK-reconstructed CBCT scans are segmented using global thresholding, followed by a simulation and subtraction of the HCAAs from the CBCT scans. This two-pass approach was recently expanded to
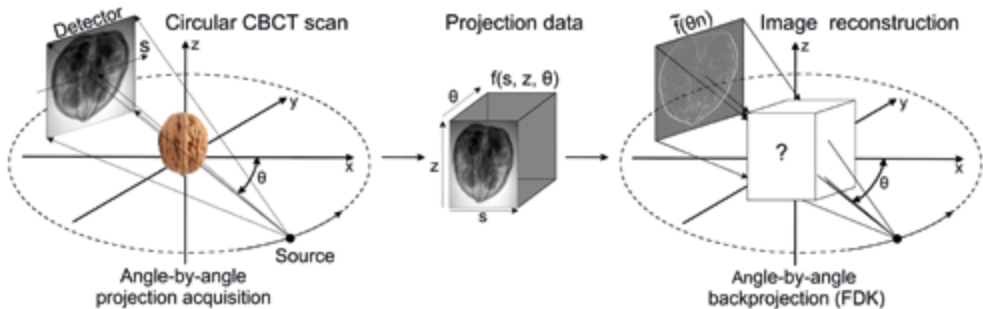

Figure 1. Schematic illustration of the CBCT image acquisition and reconstruction process with a circular x-ray source trajectory. The source and detector rotate around the object (e.g. a patient or in this case a walnut), and 2D x-ray projections of the object are acquired angle by angle. The 3D projection data volume f is indexed by the vertical and horizontal detector coordinates z and s, and angle θ. The order of operations in the FDK algorithm mirrors that of the data acquisition: each filtered 2D x-ray projection is backprojected angle by angle and added to the reconstructed CT volume.

a multi-pass approach by Han and Baek [16], who performed multiple segmentations to iteratively subtract HCAAs from the CBCT scans. However, these manually designed artifact reduction techniques require expert domain knowledge on the precise structure of the information in the measured data.

In recent years, the field of image artifact reduction has transitioned from manually designed image artifact reduction techniques reflecting expert domain knowledge to training deep neural networks. In particular convolutional neural networks (CNNs) have been employed to reduce a wide variety of CBCT image artifacts, such as metal artifacts [17–19], scatter-induced artifacts [20–22], and noise [23]. The main advantage of CNNs over alternative methods is that they can learn to automatically compute a generalized mapping between artifact-affected input images and artifact-free target images. However, training CNNs to process full high resolution three-dimensional (3D) image volumes remains computationally prohibitive and requires a large amount of suitable 3D training data. As a consequence, the feasibility of such fully 3D CNNs is extremely limited in clinical settings, where predictions are needed promptly. In order to circumvent this problem, CNNs are commonly trained on Cartesian two-dimensional (2D) image slices extracted from 3D volumes.

The reduction of HCAAs in Cartesian image slices is, however, subject to difficulties not seen with other imaging artifacts. One inherent problem is that the HCAAs in a particular Cartesian slice often originate from image structures in other slices. As a consequence, the contextual information needed to remove the artifacts is lacking. Another difficulty is the fact that the severity of HCAAs depends on the distance between the Cartesian slice and the midplane (Fig. 2) and thus greatly differs between slices. We hypothesize that this variability impairs the CNN's ability to learn a generalized mapping between input and target CBCT slices, which subsequently results in incomplete artifact reduction or secondary artifacts. One way of reducing this variability is to train separate CNNs for each Cartesian slice position. However, this is computationally extremely challenging and requires far more training data since only a single 2D training slice would be extracted from each 3D volume. Another way of reducing the variability using expert domain knowledge was presented by Han et al. [24] who used a CNN to process slices of the differentiated backprojection domain in different Cartesian slices, after which the slices were combined by a spectral blending technique.
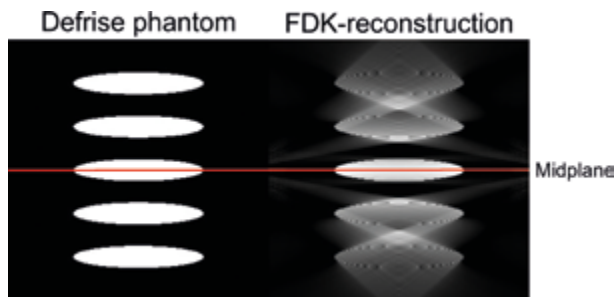

Figure 2. Illustration of the high cone-angle artifact in a Defrise phantom with five ellipses and a maximum cone-angle of 25°. Increasing the distance between the ellipse and the midplane produces stronger HCAAs in the reconstructed volume.

In this study, we propose a simpler approach to reduce high HCAAs in CBCT scans using deep learning by exploiting the most basic property of the CBCT scanning trajectory: its rotational geometry. The rotational nature of data acquisition in CBCT combined with the angle-wise backprojection performed by the FDK algorithm (Fig. 1) leads to HCAAs that share the same rotational geometry. Therefore, we suggest that this rotational geometry should also be retained when post-processing the CBCT scans. To this end, we developed a workflow in which we sample 2D radial CBCT slices (i.e., angle by angle), instead of processing conventional 2D Cartesian slices to train a CNN. Radially sampled slices exhibit much less variation in HCAAs than Cartesian slices, which simplifies the cone-angle artifact reduction task. In addition, a large number of radial slices can be extracted from a single 3D volume, which further facilitates CNN training. The proposed combination of deep learning with geometry- and context-aware dimension reduction thus allows processing of high-resolution 3D CBCT volumes by 2D CNNs with high accuracy and computational efficiency.

The main contributions of this study are as follows:

- We propose a CBCT imaging pipeline consisting of the following steps: direct FDK reconstruction of radial image slices, followed by a 2D CNN and re-sampling of the radial slices into a regular 3D volume.
- This approach was evaluated using a large, publicly available, experimental CBCT dataset specifically designed to examine cone-angle artifact reduction techniques. For each CBCT scan that was reconstructed using projection data from a single x-ray source trajectory, a high-quality, HCAA-free target scan is available that was acquired by combining the single-trajectory projection data with projection data from two additional trajectories using an IR scheme.
- The results of this study show that the proposed radial slicing approach improves the artifact reduction performance of CNNs compared to traditional Cartesian slicing in terms of quantitative error measures and visual image quality. In addition, we demonstrate that this process leads to substantial improvements in image segmentation performance, which is currently the most commonly performed image processing task in clinical settings [25].

2. MATERIALS AND METHODS

Our proposed HCCA reduction approach is based on a direct reconstruction of radial slices using the FDK algorithm. These radial slices are subsequently processed by a CNN, after which they are re-sampled into a Cartesian volume. The proposed artifact reduction workflow was compared to a Cartesian slice-based artifact reduction workflow, in which a standard FDK-reconstructed volume is sliced orthogonally (i.e., axial, sagittal or coronal) and each slice is processed by a CNN. The Cartesian slice-based workflow and the proposed artifact reduction workflow are illustrated in Figure 3.

In order to evaluate both artifact reduction workflows, a CBCT dataset was required consisting of input scans that were heavily affected by HCAAs and high-quality target scans without HCAAs. To this end, we used a publicly available CBCT data collection containing 42 scans of walnuts described

in a previous publication [26]. This data collection was specifically acquired with the purpose of benchmarking HCAA reduction algorithms. The advantage of using walnuts over anthropomorphic phantoms or manufactured objects is that they exhibit a natural variability. Moreover, since walnuts consist of a hard shell, soft core and air-filled cavities, they resemble the structure of the human skull [27].

2.1. Data acquisition

The input CBCT scans were obtained by scanning all 42 walnuts using a custom-built cone-beam CT scanner (FleX-ray) [28], with a tube voltage of 40 kV and a tube current of 0.3 mA. To ensure that that the CBCT scans were affected by severe HCAAs, the vertical cone-angle was maximized by moving the samples as close as possible to the x-ray source. This resulted in a maximum cone-angle of 20°, which is comparable to cone-angles found in clinical CBCT scanners [29, 30]. In addition, the x-ray source height was chosen such that the midplane was close to the bottom part of the walnuts, which induced severe artifacts in the upper parts of the scans (Fig. 4). A total of 1201 cone-beam projections were acquired and used to reconstruct CBCT scans with a volume size of 501 × 501 × 501 and an isotropic voxel size of 0.1 mm using the FDK algorithm implemented in the ASTRA Toolbox [31]. These reconstructed CBCT scans served as input data to train the CNN in this study.

In order to acquire high-quality target CBCT scans, all 42 walnuts were also scanned at 15 and 30 mm above the initial source height. The projection data obtained from the three different source positions were used to iteratively reconstruct CBCT scans by solving a non-negativity constrained least-squares problem with 50 iterations of Nesterov accelerated gradient descent [26,32]. Since three different x-ray source trajectories were used, projections with relatively small cone-angles were available for all parts of the walnut. By combining these projections, high-quality artifact-free target CBCT scans were acquired that offered high contrast and signal-to-noise ratios. Examples of the input and target CBCT scan used in this study are presented in Figure 4.

In order to obtain the data for the proposed cone-angle artifact reduction workflow, radial CBCT slices needed to be reconstructed from the projection data. Although it is relatively straight-forward to implement an FDK algorithm that directly reconstructs radial slices [33, 34], we aimed
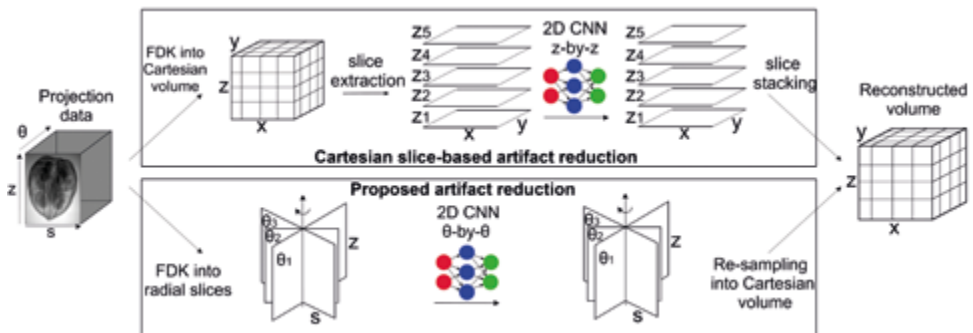

Figure 3. Schematic overview of a Cartesian slice-based workflow and the proposed cone-angle artifact reduction workflow.
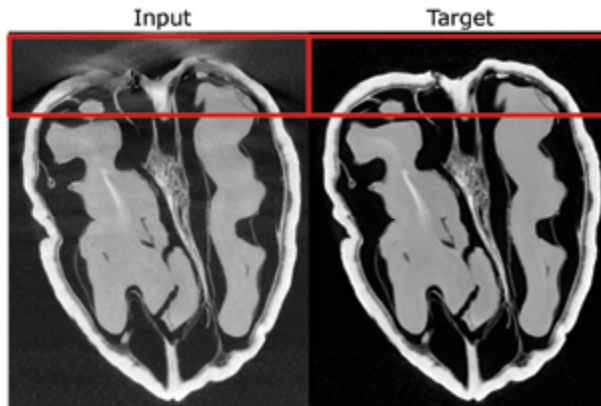
Figure 4. Examples of an input and target CBCT scan used to train the CNNs in this study. The red boxes indicate the regions of interest that were particularly affected by high HCAAs.

to first prove the concept of the proposed workflow. For this purpose, we interpolated radial slices from the Cartesian input and target volumes described above. To minimize interpolation artifacts, we only extracted slices close to the central x- and y-axes of the Cartesian volumes, i.e. for radial angles θ close to 0° and 90°. In order to acquire slices from the whole range of radial angles, we artificially rotated the scanning geometry and again extracted radial slices close to 0° and 90°. This process was repeated 24 times, such that a total of 709 radial slices were acquired per reconstructed volume. We used a larger number of radial slices compared to Cartesian slices (709 versus 501) to ensure that the radial-to-Cartesian re-sampling (see section 2.3. Experiments) performed after the cone-angle artifact reduction step (Fig. 3) would not result in secondary artifacts.

2.2. CNN architectures

Two different CNN architectures were used to evaluate the proposed artifact reduction workflow. First, we employed the widely-used U-Net. Since U-Net consists of a relatively high number of trainable parameters it tends to overfit when the training set is relatively small [35]. Therefore, we also used a mixed-scale dense convolutional neural network (MS-D Net) developed by Pelt and Sethian [35]. The MS-D networkuses densely connected convolutional layers to directly pass important feature maps to deeper layers of the network. Information from various scales is aggregated by dilating convolutions instead of using explicit up- and downsampling layers. As a result, MS-D network has far fewer trainable parameters compared to U-Net, which reduces the risk of overfitting [35].

We used publicly available U-Net and MS-D network implementations [36, 37] that are based on the deep learning framework PyTorch (v. 1.4.0) for Python (v. 3.7.3). The U-Net employed in the present study was modified in two ways compared to the original U-Net architecture proposed by Ronneberger et al. [38]. First, reflective padding was applied on all input images of which the dimensions were not divisible by 16, since the U-Net would not be able to process these images otherwise. Second, batch normalization [39] was performed before each ReLU activation function.

The MS-D network used in this study was the same as the one described by Pelt and Sethian [35] with a width of 1. Training of the network was performed using a batch size of 4 for the U-Net, and a batch size of 1 for the MS-D Net. In addition, training was performed using the default Adam [40] optimizer on a server with 192 GB RAM and one NVidia GeForce GTX 1080 Ti GPU. More detailed implementation details and justification of the chosen network parameters can be found in the Supplemental Material.

2.3. Experiments

The Cartesian slice-based artifact reduction workflows and the proposed artifact reduction workflow (Fig. 3) were evaluated in this study. For this purpose, we divided the 42 available CBCT scans into a training set comprising 28 scans, a validation set comprising 7 scans, and a test set comprising 7 scans.

The validation set was used to determine the number of epochs necessary to train the MS-D network and the U-Net. The best model performances were achieved after 40 epochs for U-Net, and 60 epochs for MS-D Net. The validation set was also used to optimize the number of layers in the MS-D network. Specifically, the number of layers was varied between 10 and 100 with a step size of 10. The number of layers, i.e. 80, was chosen such that the MS-D network achieved the highest performance on the validation set. The training process of U-Net and MS-D network took approximately 47 min and 107 min per epoch, respectively, whereas the processing of all slices in one CBCT scan using the trained networks took approximately 27 s and 34 s, respectively.

The trained CNNs were subsequently used to reduce artifacts in the seven CBCT scans of the test set. To ensure that differences in artifact reduction performance were not due to randomness in the initialization of the CNNs' trainable parameters, both the U-Net and the MS-D network were independently initialized and trained five times. To enable fair comparison between the two artifact reduction workflows, all processed radial slices were re-sampled into a Cartesian volume using linear interpolation. The code used for the radial-to-Cartesian resampling is publicly available online [41].

2.4. Performance evaluation metrics

The cone-angle artifact reduction achieved using the proposed artifact reduction workflow was compared to those of Cartesian slice-based artifact reduction workflows. More specifically, the Cartesian slice-based artifact reduction workflow was evaluated independently three times, once for each orthogonal slice orientation (i.e. axial, sagittal and coronal). As an additional benchmark, we applied the Nesterov accelerated gradient descent reconstruction algorithm to the projection data acquired using only the lowest x-ray source position (see section 2.1. Data acquisition). We will refer to this method as the IR approach. Furthermore, we included the FDK-reconstructed input CBCT scans in the comparison. The cone-angle artifact reduction performances were evaluated by computing the structural similarity (SSIM) index [42] and the mean-squared error (MSE) between the artifact-free target CBCT scans and the output of each approach. These metrics have been commonly used to assess the quality of medical images including CBCT scans [43].

In addition to these generic image quality metrics, we examined the impact of the artifact reduction on a concrete, clinically-relevant image processing task. Since CBCT scans are being

increasingly used for human skull segmentation [19] and the walnuts used in this study were chosen as proxies for the human skull, we assessed the impact of the proposed approach on the accuracy of CBCT image segmentation. For this purpose, Otsu's thresholding method [44] was used to segment all CBCT scans into three different classes: background, soft inner walnut structures and the hard outer shell. Evaluation of this segmentation step was performed by calculating the Dice similarity coefficient (DSC) between the segmented target CBCT scans and the CBCT scans segmented after cone-beam artifact reduction. Since the HCAAs were most severe in the upper parts of the CBCT scans, all three metrics (i.e. SSIM, MSE and DSC) were also computed for this specific region of interest (ROI), which is indicated by the red boxes in Figure 4. All metrics were calculated for the seven CBCT scans in the test set. The metrics obtained using the trained CNN's (i.e. U-Net and MS-D Net) were averaged over the five independent runs.

3. RESULTS

The SSIMs, MSEs and DSCs obtained using the different cone-angle artifact reduction approaches are summarized in Table 1 and Table 2. Of all evaluated approaches, the proposed geometry-aware artifact reduction workflow generally resulted in the highest mean SSIMs and mean DSCs, and the lowest mean MSEs. The FDK-reconstructed input CBCT scans always resulted in the lowest SSIMs and DSCs, and the highest MSEs. When analyzing the full CBCT scans (i.e. no ROI), the IR approach resulted in SSIMs that were comparable or higher than all other evaluated approaches, whereas the IR approach was outperformed by all deep learning approaches when analyzing the ROI in the upper part of the scans. The proposed artifact reduction workflow generally resulted in higher mean SSIMs and lower mean MSEs compared to the Cartesian slice-based workflows. However, the proposed artifact reduction workflow resulted in lower mean DSCs than the horizontal slice-based workflow.

Figure 5 shows examples of HCAA reduction in CBCT slices, and the corresponding segmentations. All CBCT scans processed with the deep learning approaches demonstrated reduced HCAAs. More specifically, the proposed artifact reduction workflow produced CBCT scans that closely resembled the target CBCT scans and generally resulted in less residual HCAAs when compared to those produced by the Cartesian slice-based artifact reduction approach. The Cartesian slice-based approaches produced secondary artifacts (Fig. 6) in the majority of CBCT scans, which significantly affected their image qualities. The IR approach generally produced high-quality CBCT scans with low noise levels, but failed to remove the HCAAs from the upper parts of the scans. Figure 7 shows an example of a radial slice and an axial slice obtained through the radial-to-Cartesian interpolation step in the proposed HCAA reduction workflow. The interpolation step did not result in any notable artifacts.

4. DISCUSSION

HCAAs are an inherent problem of circular CBCT technology. Although deep learning has recently been successfully employed to reduce various imaging artifacts, cone-angle artifact reduction remains especially challenging. Therefore, the aim of this study was to develop a novel deep learning-based artifact reduction workflow that exploits the symmetry of HCAAs.

Table 1. Structural similarity (SSIM) index (± standard deviation) and the mean-squared error (MSE) (± standard deviation) produced by the different artifact reduction approaches.

	SSIM		MSE (x10⁻⁶)	
	Full volumes	**ROI**	**Full volumes**	**ROI**
Input - FDK	0.679 ± 0.067	0.547 ± 0.079	15.12 ± 2.58	55.50 ± 8.15
IR approach	**0.963 ± 0.009**	0.870 ± 0.030	5.04 ± 1.25	27.54 ± 6.54
U-Net				
Axial	0.919 ± 0.028	0.918 ± 0.040	19.70 ± 15.21	8.56 ± 3.12
Sagittal	0.930 ± 0.039	0.945 ± 0.029	41.11 ± 43.10	13.43 ± 10.32
Coronal	0.956 ± 0.020	0.943 ± 0.031	2.61 ± 0.97	6.51 ± 2.14
Proposed	0.957 ± 0.047	**0.953 ± 0.017**	2.59 ± 0.76	6.38 ± 3.12
MS-D Net				
Axial	0.947 ± 0.029	0.920 ± 0.032	2.88 ± 1.43	8.50 ± 3.33
Sagittal	0.953 ± 0.017	0.915 ± 0.035	3.04 ± 1.26	7.93 ± 2.77
Coronal	0.942 ± 0.023	0.904 ± 0.049	3.96 ± 2.50	10.1 ± 3.65
Proposed	0.949 ± 0.021	0.932 ± 0.031	**2.56 ± 0.70**	**5.71 ± 2.37**

Table 2. Mean Dice similarity coefficients (DSC) (± standard deviation) of the segmented low-density inner structures and high-density outer structures.

	DSC – low density structures		DSC – high density structures	
	Full volumes	**ROI**	**Full volumes**	**ROI**
Input - FDK	0.904 ± 0.017	0.720 ± 0.033	0.895 ± 0.022	0.682 ± 0.027
IR approach	0.911 ± 0.015	0.772 ± 0.166	0.917 ± 0.019	0.794 ± 0.025
U-Net				
Axial	0.898 ± 0.040	0.863 ± 0.130	0.913 ± 0.034	0.936 ± 0.016
Sagittal	0.914 ± 0.039	0.897 ±0.027	0.931 ± 0.044	0.918 ± 0.087
Coronal	**0.946 ± 0.016**	**0.908 ± 0.015**	**0.956 ± 0.020**	0.946 ± 0.014
Proposed	0.933 ± 0.012	0.898 ± 0.013	0.951 ± 0.013	0.945 ± 0.017
MS-D Net				
Axial	0.941 ± 0.018	0.891 ± 0.017	0.942 ± 0.061	0.933 ± 0.018
Sagittal	0.940 ± 0.019	0.896 ± 0.014	0.950 ± 0.018	0.937 ± 0.017
Coronal	0.936 ± 0.020	0.894 ± 0.032	0.945 ± 0.018	0.925 ± 0.020
Proposed	0.933 ± 0.013	0.902 ± 0.013	0.952 ± 0.013	**0.950 ± 0.014**

The proposed geometry- and context-aware artifact reduction workflow was able to reduce HCAAs in CBCT scans more effectively than the Cartesian slice-based artifact reduction workflow (Figs. 5 and 6). Moreover, the CNNs trained in the proposed workflow always resulted in SSIMs, MSEs and DSCs that were comparable or better than those obtained using the IR approach (Tables 1 and 2). The primary reason why the proposed workflow resulted in better HCAA reduction is that radial slicing provides a more efficient dimension reduction from 3D to 2D compared to Cartesian slicing: the HCAAs exhibit much less spatial variation in radial slices compared to Cartesian slices, and radial slices offer more relevant context to remove the HCAAs. This facilitates the training process of
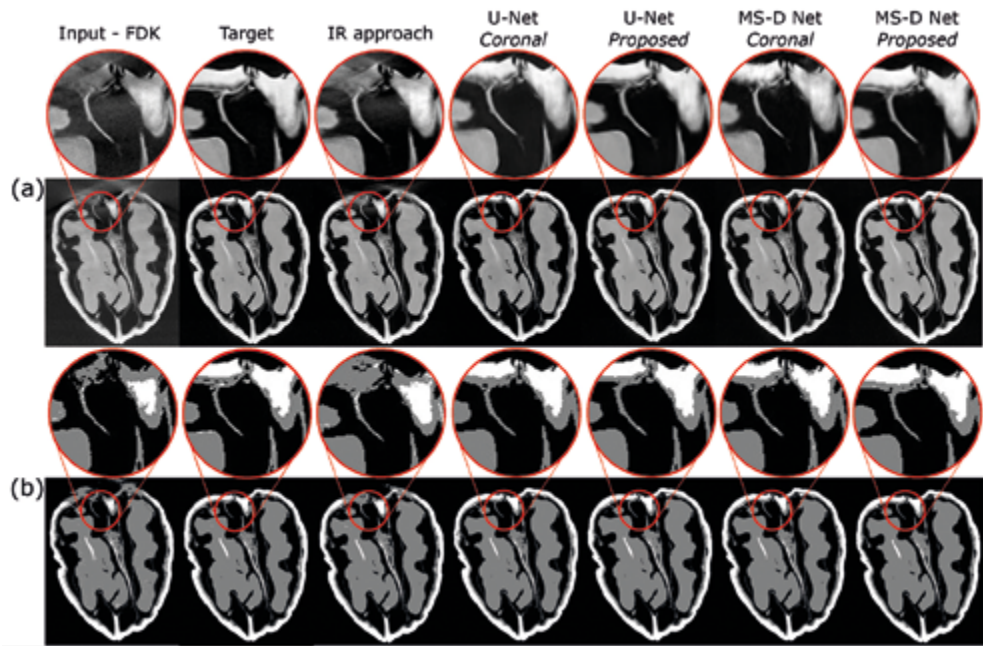
Figure 5. Example of a vertical slice of a CBCT scan processed by the HCAA reduction methods. The upper row (a) shows the output of the different artifact reduction approaches, and the bottom row (b) shows the corresponding segmented slices. Of the three evaluated Cartesian slice approaches, only the results of the coronal slice-based approach are visualized since it performed best in terms of SSIMs, MSEs and DSCs.
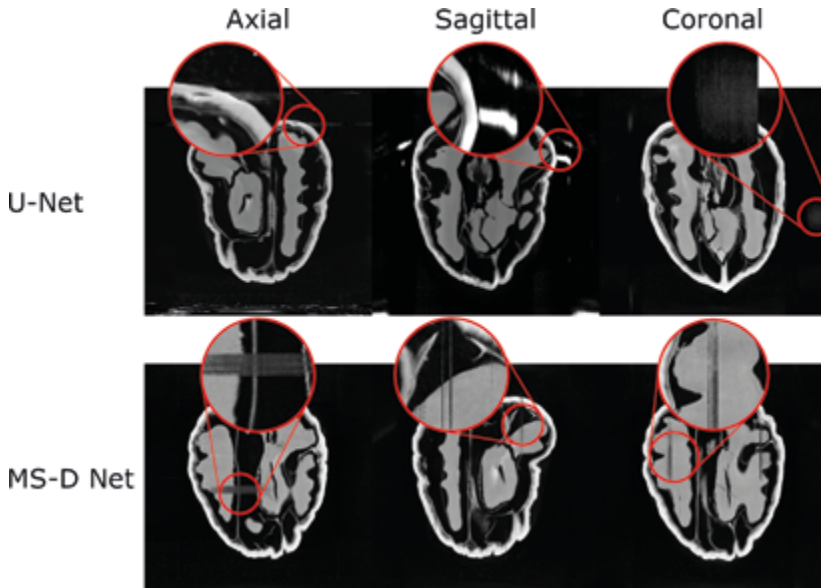

Figure 6. Examples of secondary artifacts in walnut CBCT slices produced by Cartesian slice-based workflows (i.e. axial, sagittal and coronal).
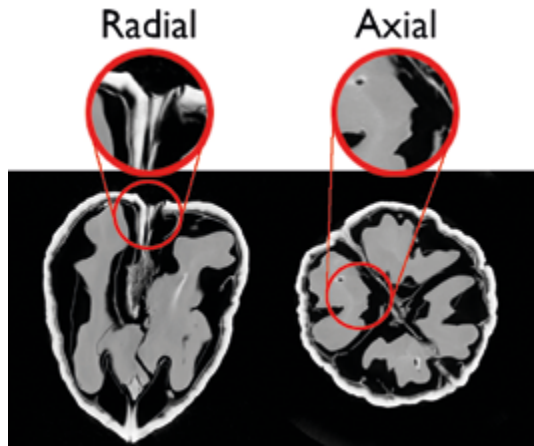
Figure 7. Example of a radial slice and an axial slice obtained through the radial-to-Cartesian interpolation step in the proposed HCAA reduction workflow.

a CNN and enables the learning of a generalized mapping between artifact-affected input scans and artifact-free target scans.

An interesting finding in this study was that the Cartesian slice-based artifact reduction workflows resulted in unexpectedly severe secondary artifacts in the majority of the CBCT scans (Fig. 6). This phenomenon was observed when using both U-Net and MS-D Net. Since the secondary artifacts manifested as clearly distinguishable streaks in the Cartesian planes, it is likely that the CNNs failed to correctly process the corresponding Cartesian slices of those CBCT scans. A possible explanation for this finding is that the CNNs tried to reduce HCAAs in all slices of the CBCT scan, even if those slices were not affected by HCAAs. After all, since the CNNs had no contextual information on the position of the 2D slices within the CBCT scan, it might have been difficult to distinguish artifact-affected slices from artifact-free slices.

A major advantage of using CNNs for cone-angle artifact reduction is the low computational time required to process CBCT scans compared to more traditional methods such as the IR approach. More specifically, the time needed to process one CBCT scan in this study was approximately 27 s for U-Net and 34 s for MS-D Net, whereas the IR approach took approximately 20 min to reconstruct a CBCT scan ($512 \times 512 \times 512$). The shorter computational times make it much easier to incorporate these CNNs approaches in clinical workflows.

Another advantage of the proposed HCAA artifact reduction workflow is its broad applicability. Since our workflow exploits basic geometrical properties of circular CBCT scanners, it is expected that this workflow is capable of accurately reducing HCAAs in CBCT scans obtained using any kind of circular CBCT scanner. Moreover, the proposed deep learning workflow may provide promising avenues to reduce HCAAs in limited angle CBCT systems. Nevertheless, further research is necessary to determine whether CNNs trained on scans from a specific CBCT scanner can generalize to scans acquired using different CBCT scanners.

A technical challenge of employing our novel deep learning approach was the radial-to-Cartesian resampling step that is necessary after processing the radial CBCT slices. Although Figure 7 shows that no notable interpolation artifacts were introduced in this study, the re-sampling step remains a 2D task that is performed for each z-slice independently, which could possibly introduce minor interpolation artifacts in the CBCT scans [45]. One way of avoiding such interpolation artifacts could be to use a deep learning-based interpolation method [46]. Because such methods are able to memorize local structures in CBCT scans, they can recognize geometric variations and thus produce more accurate pixel estimations.

The next step towards employing the proposed artifact reduction workflow is to translate the learning procedure and the necessary data acquisition to clinical practice. A possible way of achieving this would be to perform a phantom or cadaver study and acquire high-quality target CBCT scans following the methodology of the present study, i.e. by combining projection data from three different circular x-ray trajectories. Another option would be to use helical CT scans as the high-quality target, since those are not affected by HCAAs. However, since patients are not commonly scanned using both CBCT and helical CT, it might be necessary to apply semi-supervised learning approaches such as generative adversarial networks [47]. Another way of facilitating the application of the proposed approach in clinical settings might be through transfer learning. In such a transfer learning scheme, the CNN models trained on walnut CBCT scans could be used as the initial model, instead of randomly initializing a new model. Since the models do not need to be trained from scratch, fewer clinical CBCT scans would be necessary to learn the cone-angle artifact reduction. By translating the proposed deep learning-based cone-angle artifact reduction approach into the clinic, medical practitioners will be able to establish accurate diagnoses and create feasible treatment plans based on circular CBCT images of the patient. In addition, the proposed workflow might allow clinicians to scan larger volumes by increasing the cone-angles, which would markedly improve the applicability of circular CBCT technology.

5. CONCLUSION

This study presents a novel pipeline to reduce HCAAs in CBCT scans with deep learning. By designing a tailored dimension reduction scheme that reflects the rotational symmetry of CBCT scans, we were able to efficiently reduce 3D cone-angle artifacts in CBCT scans with 2D CNNs. The proposed HCAA reduction workflow showed to be more robust than the Cartesian slice-based workflows, while it consistently outperformed the IR approach. In addition, we showed that the cone-angle artifact reduction leads to considerable improvements when segmenting the CBCT scans. The results of this study will hopefully motivate clinicians and medical engineers to adopt the proposed artifact reduction workflow in clinical settings, thereby opening up promising new avenues for a large-scale use of CBCT imaging in a wide variety of medical disciplines.

ACKNOWLEDGMENTS

support by Holland High Tech through the PPP allowance for research and development in the HTSM topsector.

SOFTWARE AVAILABILITY

To increase the reproducibility of this study, we provide the Python scripts that were used for training data preparation, model training and model evaluation. These scripts will be made publicly available on Github (https://github.com/Jomigi/Cone_angle_artifact_reduction). Running the scripts requires the open-source software packages ASTRA toolbox and Pytorch, and a Graphics Processing Unit(GPU).

CONFLICTS OF INTEREST

6

The authors have no conflicts of interest to disclose.

REFERENCES

1. Venkatesh E, Venkatesh Elluru S. Cone beam computed tomography: basics and applications in dentistry. *J. Istanbul Univ. Fac. Dent.* 2017; 51(3 Suppl 1): S102–S121

2. Tuy HK. An inversion formula for cone-beam reconstruction. *SIAM J. Appl. Math.* 1983;43(3):546–52

3. Smith BD. Image reconstruction from cone-beam projections: necessary and sufficient conditions and reconstruction methods. *IEEE Trans. Med. Imaging* 1985;4(1):14–25

4. Feldkamp LA, Davis LC, Kress JW. Practical cone-beam algorithm. *J. Opt. Soc. Am.* 1984;1(6):612–619

5. Scarfe WC, Farman AG. What is cone-beam CT and how does it work? *Dental Clin. North Am.* 2008;52(4);707–730

6. Grass M, Köhler T, Proksa R. 3D cone-beam CT reconstruction for circular trajectories. *Phys. Med. Biol.* 2000;45(2):329–347

7. Sidky EY, Pan X. Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization. *Phys. Med. Biol.* 2008;53(17):4777–4807

8. Karimi D, Ward RK. Image reconstruction in computed tomography using variance-reduced stochastic gradient descent. *IEEE 14th Int. Symp. on Biomed. Imaging.* 2017:543–547

9. Maass C, Dennerlein F, Noo F, Kachelriess M. Comparing short scan CT reconstruction algorithms regarding cone-beam artifact performance. *IEEE Nuclear Science Symp. & Med. Imaging Conf.* 2010:2188–2193

10. 10. Gardner SJ, Mao W, Liu C, Aref I, Elshaikh M, Lee JK, *et al.* Improvements in CBCT image quality using a novel iterative reconstruction algorithm: a clinical evaluation. *Adv. Radiat. Oncol.* 2019;4(2):390–400

11. Hsieh J, Nett B, Yu Z, Sauer K, Thibault J-B and Bouman CA. Recent advances in CT image reconstruction. *Curr. Radiol. Rep.* 2013;1:39–51

12. Tang X, Krupinski EA, Xie Hl, Stillman AE. On the data acquisition, image reconstruction, cone beam artifacts, and their suppression in axial MDCT and CBCT—a review. *Med. Phys.* 2018;45(9):e761–e782

13. Hu H. An improved cone-beam reconstruction algorithm for the circular orbit. *Scanning.* 2006;18(8):572–581

14. Zhu L, Starman J, Fahrig R. An efficient estimation method for reducing the axial intensity drop in circular cone-beam CT. *Int. J. Biomed. Imaging.* 2008:242841

15. Hsieh J. Two-pass algorithm for cone-beam reconstruction *Proc. SPIE 3979. Medical Imaging.* 2000:533–540

16. Han C, Baek J. Multi-pass approach to reduce cone-beam artifacts in a circular orbit cone-beam CT system. *Opt. Express.* 2019;27(7):10108–10126

17. Claus BEH, Jin Y, Gjesteby L, Wang G, De Man B. Metal-artifact reduction using deep-learning based sinogram completion: initial results. *Fully3D.* 2017:631–635

18. Gjesteby L, Yang Q, Xi Y, Shan H, Claus B, Jin Y, *et al.* Deep learning methods for CT image-domain metal artifact reduction. *Proc. SPIE.* 2017:103910W

19. Minnema J, Eijnatten M, Hendriksen AA, Liberton NPTJ, Pelt DM, Batenburg KJ, *et al.* Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med. Phys.* 2019;46(11):5027–5035

20. Maier J, Sawall S, Knaup M, Kachelrieß M. Deep scatter estimation (DSE): accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network. *J. Nondestruct. Eval.* 2018;37:57

21. Jiang Y, Yang C, Yang P, Hu X, Luo C, Xue Y, *et al.* Scatter correction of cone-beam CT using a deep residual convolution neural network (DRCNN). *Phys. Med. Biol.* 2019;64(14):145003

22. Griner D, Garrett JW, Li Y, Li K, Chen G-H. Correction for cone beam CT image artifacts via a deep learning method. *Medical Imaging 2020: Physics of Medical Imaging.* 2020:162

23. Burger HC, Schuler CJ, Harmeling S. Image denoising with multi-layer perceptrons: I. Comparison with existing algorithms and with bounds. 2012. arXiv:1211.1544

24. Han Y, Kim J, Ye JC. Differentiated backprojection domain deep learning for conebeam artifact removal. *IEEE Trans. Med. Imaging.* 2020;39(11):3571-3582

25. Maier-Hein L, Eisenmann M, Reinke A, Onogur S, Stankovic M, Scholz P, *et al.* Why rankings of biomedical image analysis competitions should be interpreted with care. *Nat Commun.* 2018; 9(1): 5217.

6

26. Der Sarkissian H, Lucka F, van Eijnatten M, Colacicco G, Coban SB and Batenburg KJ. A cone-beam x-ray computed tomography data collection designed for machine learning. *Sci. Data*. 2019;6:215

27. Hämäläinen K, Harhanen L, Kallonen A, Kujanpää A, Niemi E, Siltanen S. Tomographic x-ray data of a walnut. 2015. arXiv:1502.04064

28. Coban SB, Lucka F, Palenstijn WJ, Van Loo D and Batenburg KJ. Explorative imaging and its implementation at the FleX-ray laboratory. *J. Imaging*. 2020;6(4):18

29. Li Y, Garrett JW, Li K, Wu Y, Johnson K, Schafer S, *et al*. Time-resolved C-arm cone beam CT angiography (TRCBCTA) imaging from a single short-scan C-arm cone beam CT acquisition with intra-arterial contrast injection. *Phys. Med. Biol*. 2018;63(7):075001

30. Sheth NM, De Silva T, Uneri A, Ketcha M, Han R, Vijayan R, *et al*. A mobile isocentric C-arm for intraoperative cone–beam CT: Technical assessment of dose and 3D imaging performance. *Med. Phys*. 2020; 47(3):958–974

31. van Aarle W, Palenstijn W J, Cant J, Janssens E, Bleichrodt F, Dabravolski A, De Beenhouwer J, Joost Batenburg K and Sijbers J 2016 Fast and flexible x-ray tomography using the ASTRA toolbox Opt. Express 24 25129–47

32. Chambolle A, Pock T. An introduction to continuous optimization for imaging. *Acta Numer*. 2016;25:161–319

33. Buurlage J-W, Kohr H, Palenstijn WJ, Batenburg KJ. Real-time quasi-3D tomographic reconstruction. *Meas. Sci. Technol*. 2018;29:064005

34. Vanrompay H, Buurlage J, Pelt DM, Kumar V, Zhuo X, Liz-Marzán LM, *et al*. Real-time reconstruction of arbitrary slices for quantitative and in situ 3D characterization of nanoparticles. *Part. Part. Syst. Charact*. 2020;37(7): 2000073

35. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl Acad. Sci*. 2018;115(2):254–259

36. Hendriksen AA. 2019. ahendriksen/msd_pytorch: v0.7.2 Zenodo https://zenodo.org/record/3560114

37. Milesial. 2019. Pytorch-UNet https://github.com/milesial/Pytorch-UNet

38. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention — MICCAI*. 2015:234–241

39. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. ArXiv:1502.03167 [Cs]. 2015. http://arxiv. org/abs/1502.03167 (accessed June 21, 2019).

40. Kingma DP, Ba J, Adam: a method for stochastic optimization. ArXiv:1412.6980 [Cs]. 2015. http://arxiv.org/abs/1412.6980.

41. Minnema J and Lucka F. Cone_angle_artifact_reductionGithub. 2020. https://github.com/Jomigi/Cone_angle_artifact_reduction

42. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process*. 2004;13(4):600–612

43. Venkat Narayana Rao T, Govardhan A . Assessment of diverse quality metrics for medical images including mammography. *IJCA*. 2013;83(4):42–47

44. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern*. 1979;9(1):62–66

45. Aganj I, Yeo BTT, Sabuncu MR, Fischl B. On removing interpolation and resampling artifacts in rigid image registration. *IEEE Trans. Image Process*. 2013;22:816–827

46. Zhou W, Li X, Reynolds DS. Nonlinear image interpolation via deep neural network. *51st Asilomar Conf. on Signals, Systems, and Computers*. 2017:228–232

47. Maspero M, Houweling AC, Savenije MHF, van Heijst TCF, Verhoeff JJC, Kotte ANTJ, *et al*. A single neural network for cone-beam computed tomography-based radiotherapy of head-and-neck, lung and breast cancer. *Phys. Imaging Radiat. Oncol*. 2020;14:24–31

SUPPLEMENTAL MATERIAL: CNN IMPLEMENTATION DETAILS

This supplemental material provides a detailed description of the implementation of the CNN architectures in the present study. In addition, the challenges faced when training these architectures and the importance of choosing correct training parameters are addressed.

In the present study, two CNN architectures were implemented, namely U-Net and a mixed-scale dense convolutional neural network (MS-D Net). A U-Net consists of a down-sampling path and an up-sampling path (Fig. S1). The down-sampling path reduces the dimensions of the input image in order to extract (i.e., encode) the most relevant image features at different scales. In turn, the up-sampling path uses these image features to reconstruct the desired image output. Both paths are inter-connected with skip-connections that allow the network to combine image feature at various dimensional scales.

An MS-D network consist of multiple densely-connected convolutional layers (Fig. S2). This means that the output of each convolutional layer, i.e., the feature map, is passed to all subsequent convolutional layers in the network, instead of only passing the output to next convolutional layer. Furthermore, instead of using down- or up-sampling paths, the MS-D network uses dilated convolutions to recognize important features at different scales. In a dilated convolution, a convolutional kernel is expanded by a dilation factor s and filled with zeros at distances that are not a multiple of s voxels from the kernel center. In the present study, The dilation factor was set as 1 for the first convolutional and was doubled after each subsequent layer. After the fifth convolutional layer (i.e., a dilation factor of 16), the dilations factor was reset to 1, and the process was repeated.
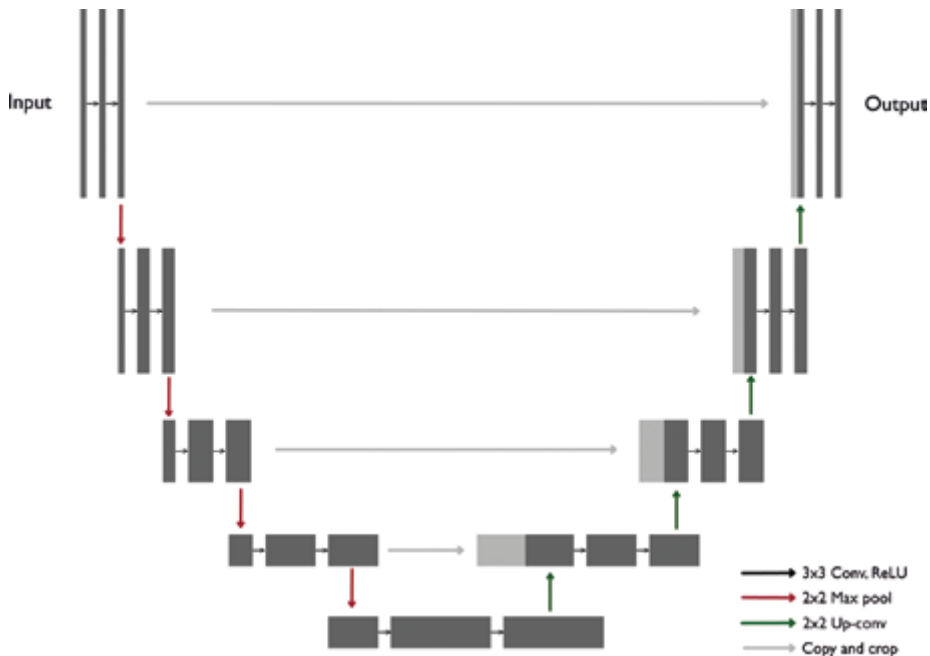
6

Figure S1. Schematic representation of the U-Net architecture used in the present study.
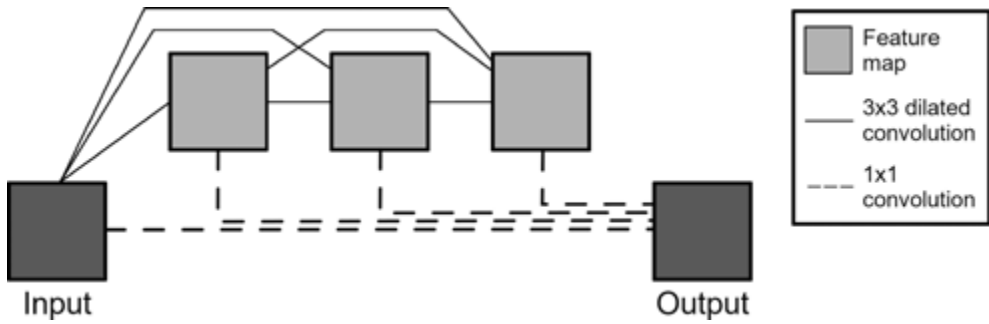
Figure S2. Example of an MS-D network architecture with a depth of 3 convolutional layers and a width of 1.

6

The implementation of the U-Net and MS-D network used in our study was based on publicly available code [1, 2]. The source code is built in the deep learning framework PyTorch (v. 1.4.0) in Python (v. 3.7.3). The U-Net employed in the present study comprised two small modifications compared to the original U-Net proposed by Ronneberger et al. [3]. First, we applied reflective padding on all input images of which the dimensions were not divisible by 16, since the U-Net would not be able to process these images otherwise. Second, batch normalization [4] was performed before each ReLU activation function. The MS-D network used in this study was the same as the one described by Pelt and Sethian [5] with a width of 1.

Training of the CNNs was initially performed using the default parameters defined in the original source code. This means that both networks were trained with a batch size of 1, an L2 loss function, and a default Adam optimizer (i.e., learning rate = 0.01, ε=1e-8). Training of the U-Net was performed for 40 epochs since convergence of the L2 loss was observed, and overfitting started to occur when trained for longer. The MS-D network was trained for 60 epochs since the validation loss did not improve when trained longer. A depth of 80 layers was chosen for the MS-D network since this resulted in the highest artifact reduction performance on the validation set.

Although these CNN training parameters generally resulted in satisfying artifact reduction performances, it must be noted that the U-Net trained on radial slices occasionally created erroneous air cavities in the soft structures of the walnuts in the CBCT scans (Fig. S3).

Since such mistakes cannot be afforded in clinical settings, we further investigated the training process of the U-Net. In this context, we found that the U-Net exhibited unstable behavior during training: the training and validation losses strongly oscillated (Fig. S4). The oscillating behavior of the training and validation losses made us believe that the batch size should be increased when training the U-Net. Moreover, we hypothesized that changing the learning rate or the loss function could help the U-Net to converge more easily. To this end, multiple U-Net models were trained with various batch sizes, learning rates and loss functions. The different settings are summarized in Table S1.

Based on these experiments, we found that configuration ID 5 resulted in the most satisfying training and validation loss curves (i.e., reduced oscillating behavior). Hence, only the batch size was increased compared to the default values; the learning rate and loss function were kept the same.

The U-Net trained with these configurations was able to adequately reduce cone-angle artifacts in CBCT scans without producing erroneous air cavities.
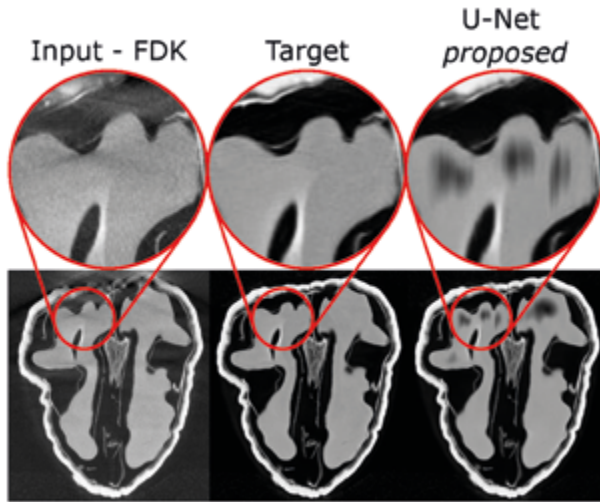
Figure S3. Example of a CBCT scan in which a U-Net trained on radial slices (proposed) erroneously created air cavities within the soft inner structures of the walnut.
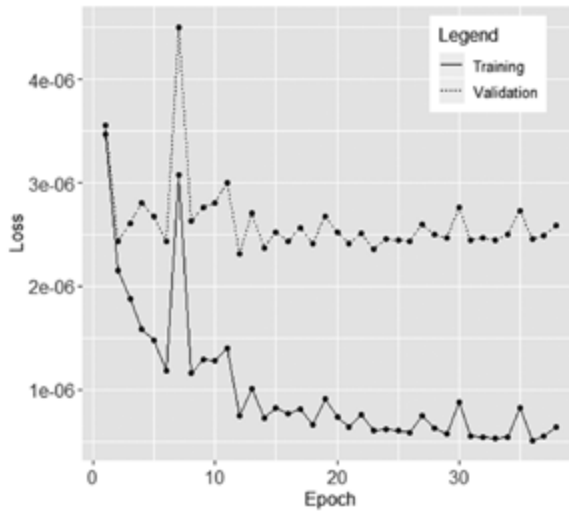

Figure S4. Example of training and validation losses of a U-Net trained with a batch size of 4.

Table S1. Overview of the configurations used to train U-Net.

Configuration ID	Batch size	Learning rate	Loss function
1 (default)	1	0.001	L2
2	1	0.001	L1
3	1	0.0001	L2
4	1	0.005	L2
5	4	0.001	L2
6	4	0.0001	L2
7	4	0.005	L2

6

REFERENCES

1. Hendriksen AA. 2019. ahendriksen/msd_pytorch: v0.7.2 Zenodo https://zenodo.org/record/3560114

2. Milesial. 2019. Pytorch-UNet https://github.com/milesial/Pytorch-UNet

3. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention—MICCAI.* 2015:234–241

4. Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. ArXiv:1502.03167 [Cs]. 2015. http://arxiv. org/abs/1502.03167 (accessed June 21, 2019).

5. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl Acad. Sci.* 2018;115(2):254–259

6

CHAPTER 7

A REVIEW ON THE APPLICATION OF DEEP LEARNING FOR CT RECONSTRUCTION, BONE SEGMENTATION AND SURGICAL PLANNING IN ORAL AND MAXILLOFACIAL SURGERY

Jordi Minnema, Anne Ernst, Maureen van Eijnatten, Ruben Pauwels, Tymour Forouzanfar, Kees Joost Batenburg, Jan Wolff

ABSTRACT

Computer-assisted surgery (CAS) allows clinicians to personalize treatments and surgical interventions and has therefore become an increasingly popular treatment modality in maxillofacial surgery. The current maxillofacial CAS consists of three main steps: (1) CT image reconstruction, (2) bone segmentation, and (3) surgical planning. However, each of these three steps can introduce errors that can heavily affect the treatment outcome. As a consequence, tedious and time-consuming manual post-processing is often necessary to ensure that each step is performed adequately. One way to overcome this issue is by developing and implementing neural networks (NNs) within the maxillofacial CAS workflow. These learning algorithms can be trained to perform specific tasks without the need for explicitly defined rules. In recent years, an extremely large number of novel NN approaches have been proposed for a wide variety of applications, which makes it a difficult task to keep up with all relevant developments. This study therefore aimed to summarize and review all relevant NN approaches applied for CT image reconstruction, bone segmentation, and surgical planning. After full text screening, 76 publications were identified: 32 focusing on CT image reconstruction, 33 focusing on bone segmentation and 11 focusing on surgical planning. Generally, convolutional NNs were most widely used in the identified studies, although the multilayer perceptron was most commonly applied in surgical planning tasks. Moreover, the drawbacks of current approaches and promising research avenues are discussed.

1. INTRODUCTION

Spatial information embedded in medical three-dimensional (3D) images is being increasingly used to personalize treatments by means of computer-assisted surgery (CAS). This novel image-based treatment modality enables clinicians to perform patient-specific virtual operations, 3D-print personalized medical constructs and perform robot-guided surgery [1]. Moreover, CAS offers the unique possibility to conduct a limitless amount of different surgical simulations (osteotomies, grafts, implants, etc.) prior to surgery in a stress-free environment [1], and to predict surgical outcomes with minimal risk to the patient. Furthermore, such virtual simulations have proven to be useful for patient communication and medical education [2,3]. As a result, CAS is currently being employed in multiple surgical branches involving the musculoskeletal system. In particular, CAS has advanced in the area of maxillofacial surgery [4].

The current maxillofacial CAS workflow consists of different steps that are illustrated in Figure 1. The first step in the workflow is image acquisition. To date, numerous imaging modalities have become available on the market, including CT and MRI. CT imaging modalities are most commonly used to visualize bony structures due to their superior hard tissue contrast. CT scanners acquire X-ray projections of the patients' anatomy from multiple angles. These projection data can be subsequently reconstructed into a 3D image using a wide variety of reconstruction methods. After the CT image acquisition step, image processing is necessary to convert the CT scans into a virtual 3D model in the standard tessellation language (STL) file format. This file format is supported by all FDA-approved medical software packages that are currently used for computer-aided design (CAD) and computer-aided manufacturing. The most important step in this CT-to-STL conversion is image segmentation (Fig. 1), in which clinicians define and delineate anatomies of interest such as bone. In the final step of the CAS workflow, the acquired STL models are exported to dedicated medical CAD software packages and used for surgical planning by virtually designing patient-specific implants, surgical guides and radiotherapy boluses [5].

Each of the aforementioned steps (i.e. CT image reconstruction, bone segmentation and surgical planning) is a potential source of errors which can lead to inaccuracies in the final STL models and impair the treatment outcome [6]. For example, imaging noise, metallic structures and patient movements can heavily affect the CT image quality after reconstruction. For image segmentation, the segmentation technique can have a considerable effect on the accuracy of the resulting model [6]. Furthermore, the surgical planning step currently relies on extensive domain expertise and manual software input, which often hampers its reproducibility.
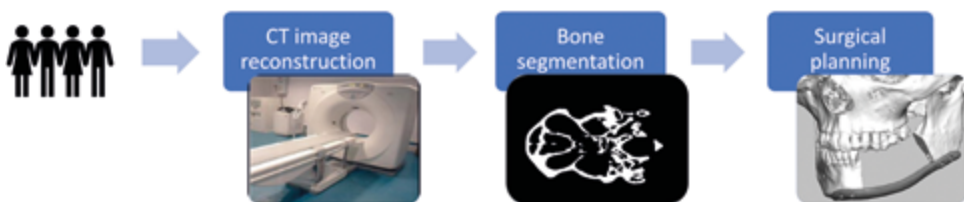

Figure 1. Schematic overview of the maxillofacial CAS workflow.

One way to overcome these limitations is to employ neural networks (NNs) during the different steps of this maxillofacial CAS workflow. These learning algorithms are different from traditional computer methods in that they can be trained to find characteristic features and patterns in data, without the need for explicit rules specified by domain experts. The most common NN is the multilayer perceptron (MLP), which consists of an input layer, several hidden layers and an output layer. Each of these layers comprises several computational building blocks called neurons. Within an MLP, neurons are connected to neurons in subsequent layers. The output of each neuron is the product of its input with a learned set of weights plus a learned bias. Finally, a non-linear activation function is applied [7]. It can be proven that MLPs can approximate any continuous function (universal approximation theorem [8]), which gives them the ability to infer descriptive functions from data.

In the training phase, the weights and biases of a NN are learned from training data. During this training process, a large amount of input data is propagated through the NN to predict the values of the output layer. The goal of training is to minimize the difference between the NN prediction and the desired output by iteratively updating the weights and biases of the network. After optimizing these trainable parameters, NNs can be used to automatically perform specific tasks of the maxillofacial CAS workflow.

Recent advances in computational power and the development of novel NN algorithms have brought about a paradigm shift in the CAS workflow [9]. A wide variety of advanced NN architectures have been successfully employed for various tasks such as image reconstruction and segmentation. For example, convolutional neural networks (CNNs) are especially useful for processing image data. These networks apply convolutions instead of a set of multiplications to compute the output of the individual layers. Another important type of neural networks is the recurrent neural network (RNN), which can handle temporal dynamic data. However, due to the rapidly increasing number of studies published in the field, maxillofacial surgeons and medical engineers have been facing the difficult task of keeping up with all developments. Therefore, this scoping review aims to provide an overview of the different types of NN approaches that have been used during the three main steps required in the CAS workflow, i.e. CT image reconstruction, bone segmentation and surgical planning. Furthermore, the secondary goal of this review paper is to identify the current bottlenecks and possible next research steps regarding the application of NNs in the maxillofacial CAS workflow.

2. MATERIALS AND METHODS

Existing literature on the application of NNs in the maxillofacial CAS workflow was obtained using Pubmed, Embase, Scopus, Web of Science, and Google Scholar. An initial database was generated with the following search terms:
- (CT OR CBCT OR computed tomography OR cone-beam computed tomography) AND (image reconstruction OR image processing OR image analysis OR image segmentation) AND (artificial intelligence OR deep learning OR neural network)

- (bone OR bones OR bony) AND ((implant OR prosthesis OR virtual model) AND (design OR planning OR construct OR model)) AND (artificial intelligence OR deep learning OR neural network)

It must be noted that there was no specific focus on maxillofacial surgery when defining the search terms. The reason for this is that we believe that many of the techniques and methods used in other fields can also be of relevance to maxillofacial CAS. Choosing more generic search terms thus allowed us to identify a wider variety of literature relevant that was potentially relevant to maxillofacial CAS.

Publications were only included in the initial database if the search terms one or two were found in their title, abstract or keywords. After removing duplicates and adding literature found from references, a database of 6994 papers was acquired. The title and abstract of these publications were screened, resulting in 248 publications that were eligible for full-text reading. In order to assess whether papers were eligible for inclusion, the following inclusion and exclusion criteria were used:

Inclusion criteria:
- Described the development or implementation of a NN.
- Performed at least one of the three main steps required in the maxillofacial CAS workflow, i.e. CT image reconstruction, bone segmentation and surgical planning. Although surgical planning is an extremely broad field, this review specifically focused on the designing and optimization of implants and virtual models.
- Evaluated on medical data sets or artificial medical phantoms.

Exclusion criteria:
- Used non-CT based imaging modalities.

The study selection process of the present study is shown in Figure 2.

3.   RESULTS AND DISCUSSION

This review aimed to identify NN architectures, training strategies and workflows that can potentially benefit CT image reconstruction, bone segmentation or surgical planning, since these steps are pivotal in the CAS workflow. In total, 76 studies were included in this review: 32 focusing on CT image reconstruction, 33 focusing on bone segmentation and 11 focusing on surgical planning. All studies are summarized in Table 1. In addition, Figure 3 shows the most popular NN approaches used in the reviewed studies.

Of the 27 reviewed studies for CT image reconstruction, 15 studies used simulated CT data to train and test their NN approach, 14 studies used clinical CT data, and 2 studies used CT data of physical phantoms. In contrast, all 26 bone segmentation studies used clinical datasets to train and test the NN approaches, while surgical planning tasks were always performed based on simulated data. Details of the datasets used to train the NN approaches are provided in Figure 3.

Training and testing of the NNs with clinical data was performed with a mean of 33 ± 42 CT volumes and 12 ± 16 CT volumes, respectively. The mean ratio between the amount of training and
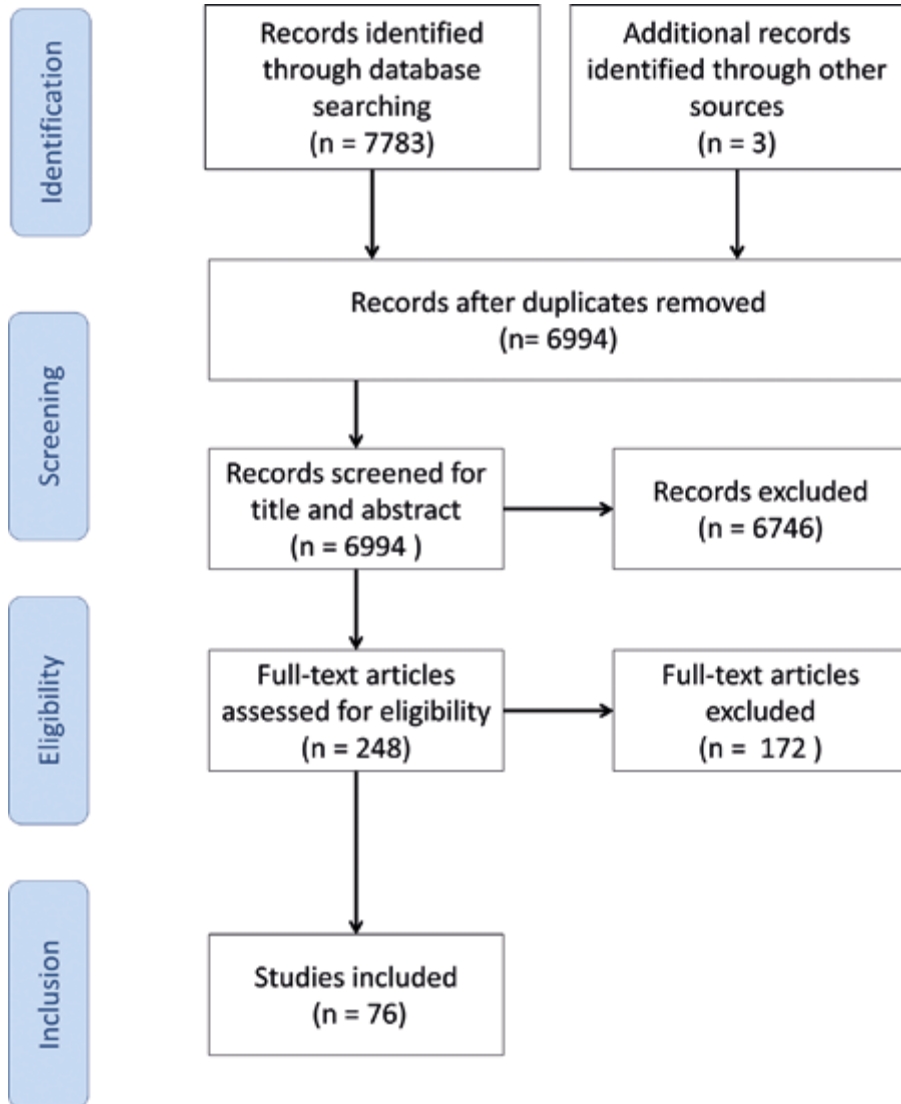
Figure 2. Overview of the study selection process of the present study.

testing data was approximately 9:2. Furthermore, 13 of the 76 reviewed studies employed a leave-k-out testing strategy to improve validation of the NNs. In this cross-validation strategy, the available data are split into k folds, where one fold is alternately used as testing data, and the remaining folds serve as training data for the NN. This process is repeated k times such that all folds have been used for testing.

Quantitative evaluation of the NNs' performances for CT image reconstruction tasks was most commonly performed using the peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) (Fig. 5a). The bone segmentation NNs were most commonly evaluated using

the dice similarity coefficient (DSC) (Fig. 5b). No consistency was observed in the performance metrics used to evaluate the surgical planning step (Fig. 5c). Moreover, 3 of the 11 surgical planning studies did not include quantitative performance evaluations.

In the following subsections we elaborate on the NN approaches that have been employed in each of the three steps of maxillofacial CAS.
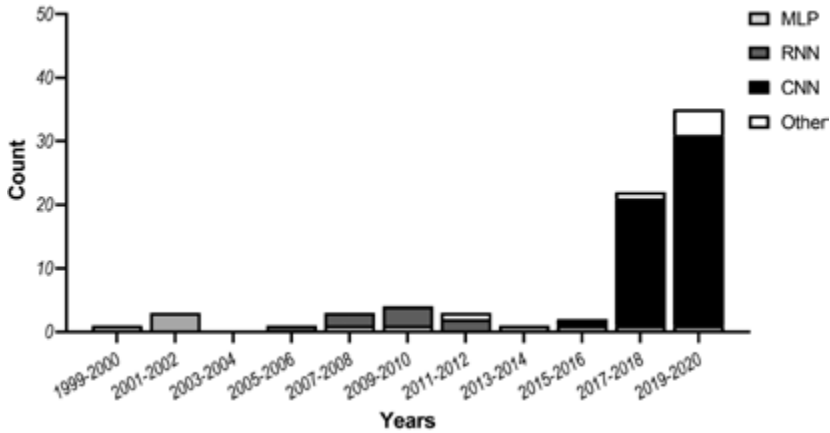

Figure 3. Most popular NN approaches in the maxillofacial CAS workflow over the past two decades. CAS, computer-assisted surgery; CNN, convolutional neural network; MLP, multilayer perceptron; NN, neural network; RNN, recurrent neural network.
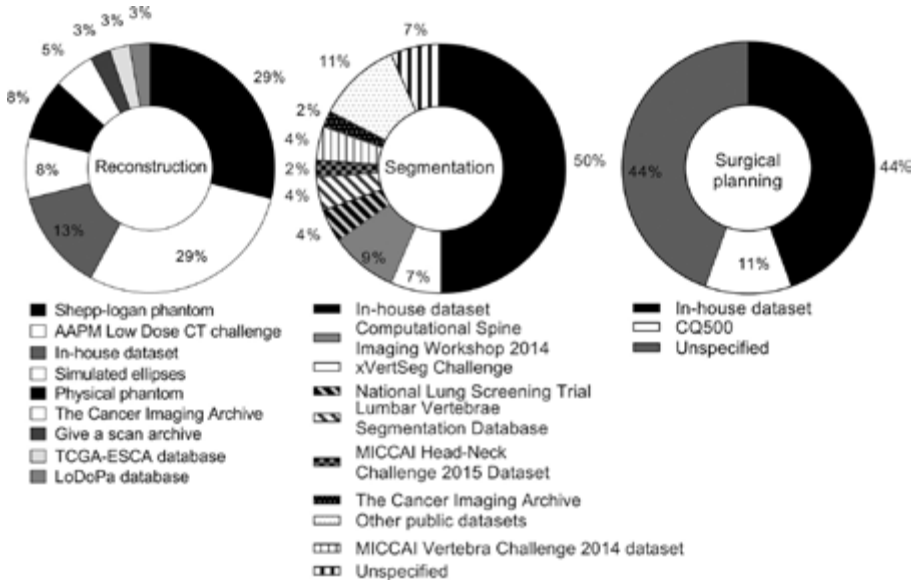

Figure 4. The data sets that were used to train and test the NN approaches in the reviewed studies. NN, neural network.
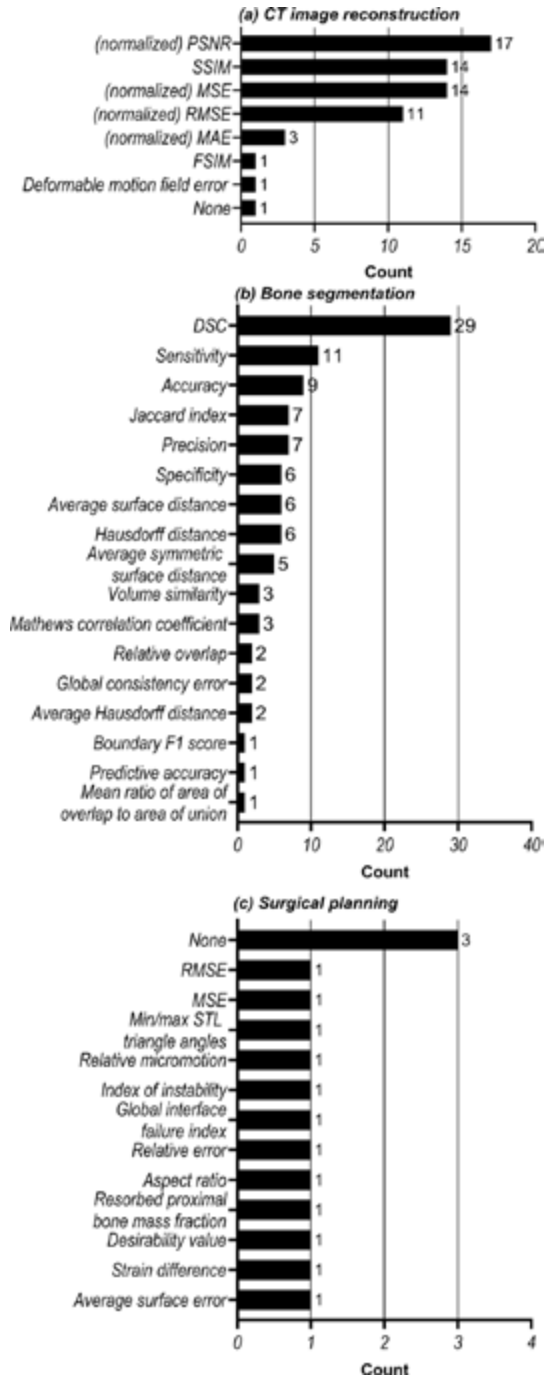
Figure 5. Analysis of the evaluation metrics used to quantify the performance of NN approaches in (a) CT image reconstruction, (b) bone segmentation and (c) surgical planning. DSC, dice similarity coefficient; RMSE, root meat-squared error; MSE, mean-squared error; PSNR, peak signal-to-noise ratio; SSIM, structural similarity index measure; STL, standard tessellation language.

Table 1. Overview of the studies included in this review.

CAS step	Year	CT imaging modality	Anatomy	Neural network architecture	Authors
CT image reconstruction	2005	Fan-beam	Abdomen	Radial basis-function NN	Hu [10]
	2006	Parallel-beam	Shepp-Logan phantom	Radial basis-function NN	Guo [11]
	2008	Parallel-beam	Shepp-Logan phantom	RNN	Cierniak [12]
	2008	Parallel-beam	Shepp-Logan phantom	RNN	Cierniak [13]
	2009	Fan-beam	Shepp-Logan phantom	RNN	Cierniak [14]
	2010	Parallel-beam	Shepp-Logan phantom	RNN	Cierniak [15]
	2010	Parallel-beam	Shepp-Logan phantom	RNN	Cierniak [16]
	2011	Parallel-beam; Fan-beam	Shepp-Logan phantom	RNN	Cierniak [17]
	2012	Parallel-beam	Shepp-Logan phantom	RNN	Cierniak and Lorent [18]
	2016	Parallel-beam, Fan-beam	Chest	CNN	Würfl et al. [19]
	2017	Fan-beam	Shepp-Logan phantom, anthropomorphic phantom head	CNN	Adler and Öktem [20]
	2018	Parallel-beam, Fan-beam	Ellipse phantom, Shepp-Logan phantom Human phantoms	CNN	Adler and Öktem [21]
	2018	Multi-detector row	Abdomen	CNN	Chen et al. [22]
	2018	Multi-detector row	Abdomen; Chest; Rat brain	CNN	Gupta et al. [23]
	2018	Multi-detector row	Abdomen	CNN	Han et al. [24]
	2018	Fan-beam	Chest	CNN	Liang et al. [25]
	2018	Cone-beam	Abdomen	CNN	Würfl et al. [26]
	2019	Fan-beam	Skull	CNN	Dong et al. [27]
	2019	Parallel-beam	Chest	CNN	Fu and de Man [28]
	2019	Multi-detector row	Torso	CNN	He et al. [29]
	2019	Multi-detector row	Chest	CNN	Lee et al. [30]
	2019	Fan-beam	Skull	GAN	Li et al. [31]
	2019	Multi-detector row	Chest	CNN	Shen et al. [32]
	2019	Fan-beam	Abdomen	CNN	Wu et al. [33]
	2019	-	Shepp-Logan phantom	CNN	Zhang and Zuo [34]
	2019	Cone-beam	Chest	CNN	Zhang et al. [35]
	2020	Fan-beam	Prostate	CNN	Chen et al. [36]
	2020	Translational CT	Chest	CNN	Wang et al. [37]
	2020	Parallel-beam	Chest	CNN	Baguer et al. [38]
	2020	Parallel-beam	Chest	CNN	Ma et al. [39]

Table 1. (continued)

CAS step	Year	CT imaging modality	Anatomy	Neural network architecture	Authors
	2020	Fan-beam; Cone-beam	Chest	CNN	Wang et al. [40]
	2020	Cone-beam	Breast	GAN	Xie et al. [41]
Bone segmentation	2002	Multi-detector row	Chest, Skull	MLP	Zhang and Valentino [42]
	2008	Multi-detector row	Phalanx	MLP	Gassman et al. [43]
	2013	-	Jaw, mouth, nose, eye, brain	MLP	Kuo et al. [44]
	2017	Multi-detector row	Femur	CNN	Chen et al. [45]
	2017	-	Mandible, Spinal cord	CNN	Ibragimov and Xing [46]
	2017	Multi-detector row	Femoral head, bladder, intestine, colon	CNN	Men et al. [47]
	2018	Multi-detector row	Whole body	CNN	Klein et al. [48]
	2018	Multi-detector row	Vertebrae	CNN	Lessman et al. [49]
	2018	Multi-detector row	Skull	CNN	Minnema et al. [50]
	2018	-	Vertebrae	Deep-belief network	Qadri et al. [51]
	2018	Multi-detector row	Mandible	CNN	Yan et al. [52]
	2018	Multi-detector row	Vertebrae	CNN	Zhou et al. [53]
	2019	-	Vertebrae	CNN	Dutta et al. [54]
	2019	-	Teeth	CNN	Gou et al. [55]
	2019	Multi-detector row	Whole-body	CNN	Klein et al. [56]
	2019	Multi-detector row	Orbital bones	CNN	Lee et al. [57]
	2019	Multi-detector row	Vertebrae	CNN	Lessmann et al. [58]
	2019	Cone-beam	Mandible, teeth	CNN	Minnema et al. [4]
	2019	Multi-detector row	Vertebrae	CNN	Rehman et al. [59]
	2019	Cone-beam	Skull	CNN	Torosdagli et al. [60]
	2019	-	Vertebrae	CNN	Vania et al. [61]
	2019	Multi-detector row	Pelvic bones	CNN	Wang et al. [62]
	2019	Micro-CT	Teeth	CNN	Yazdani et al. [63]
	2020	Cone-beam	Teeth	CNN	Lee et al. [64]
	2020	-	Temporal bone	CNN	Li et al. [65]
	2020	-	Vertebrae	CNN	Yin et al. [66]
	2020	Multi-detector row	Whole-body	CNN	Noguchi et al. [67]
	2020	Multi-detector row	Vertebrae	CNN	Bae et al. [68]
	2020	Multi-detector row	Cochleae	CNN	Heutink et al. [69]
	2020	Cone-beam	Skull	CNN	Zhang et al. [70]
	2020	-	Vertebrae	Cascaded CNN	Xia et al. [71]
	2020	Cone-beam	Teeth	CNN	Chen et al. [72]
	2020	Cone-beam	Teeth	CNN	Rao et al. [73]

Table 1. (continued)

CAS step	Year	CT imaging modality	Anatomy	Neural network architecture	Authors
Surgical planning	2000	-	Skull	MLP based on legendre polynomials	Hsu and Tseng [74]
	2001	-	Skull	MLP based on legendre polynomials	Hsu and Tseng [75]
	2001	Multi-detector row	Radius	Bernstein Basis function network	Knopf and Al-Naji [76]
	2010	n.a.	Femur	MLP	Hambli [77]
	2012	n.a.	Femur	MLP	Campoli et al. [78]
	2012	n.a.	Bone microstructure	Meshing growing neural gas (MGNG)	Fischer and Holdstein [79]
	2016	n.a.	Femur	MLP	Chanda et al. [80]
	2018	n.a.	Dental implant	MLP	Roy et al. [81]
	2019	n.a.	Spinal implant	MLP	Biswas et al. [82]
	2020	-	Skull	GAN	Kodym et al. [83]
	2020	-	Mandible	GAN	Liang et al. [84]

NN: neural network; RNN: recurrent neural network; CNN: convolutional neural network; MLP: multi-layer perceptron; GAN: generative adversarial network; n.a.: not applicable; - : not specified

3.1. Image reconstruction

Historically, CT image reconstruction (Fig. 6) has been a notoriously difficult task, which aims to compute the density of objects or anatomical structures based on the attenuation of X-rays. To date, two different reconstruction methods have been predominantly employed in clinical settings: filtered backprojection (FBP) and iterative reconstruction (IR). FBP is an analytical method in which measured projection data are uniformly distributed across the CT scan with an angle that corresponds to the acquisition of the projection data. A filter is subsequently applied to reduce blurring in the CT scan. By using projection data acquired at multiple angles with respect to the patient, a 3D CT scan can be reconstructed. IR approaches start similar to the FBP in that they use the measured projection data to reconstruct an initial CT scan. Based on this initial scan, a forward operation is performed to create artificial projection data. The artificial projection data are then compared to the measured projection data, which are used to update the initial CT scan. The forward operation and the scan update are repeated until the quality of the CT scan is satisfactory or for a fixed number of iterations. To date, a wide variety of different IR algorithms have been developed, including the algebraic reconstruction technique (ART), the simultaneous iterative reconstruction technique (SIRT) and model-based iterative reconstruction (MBIR).

Over the last decade, NNs have opened up a wealth of opportunities in the field of CT image reconstruction, offering CT images with higher quality than with FBP, while requiring shorter
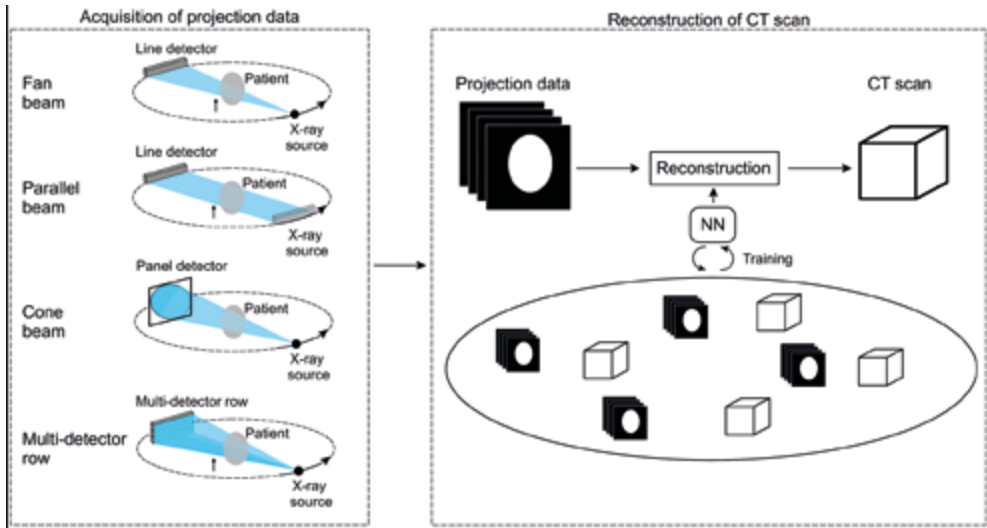
7

Figure 6. Schematic overview of the CT image reconstruction step. First, CT projection data are acquired with one of the four commonly used CT geometries (i.e., fan-beam, parallel-beam, cone-beam and multi-detector row). The acquired projection data are used to reconstruct a CT scan. This reconstruction step can be improved by training a NN.

reconstruction times than current IR reconstruction approaches. One of the first efforts in developing NNs for medical CT image reconstruction can be traced back to 2005, when Hu et al proposed two different NN-based approaches [10]. The first approach aimed to reconstruct 2D CT images using a Radial Basis Function NN (RBF-NN), which is a similar to the classical MLP, but uses a radial basis function as non-linear activation function. The input of this RBF-NN consisted of CT projection data, and the desired target consisted of previously reconstructed CT scans. In the second approach, the RBF-NN was employed to iteratively estimate the intensities of the voxels in the CT scan. More specifically, an IR scheme was employed in which the RBF-NN was trained to update the voxel values in the CT scans. Although both RBF-NN-based approaches were initially used to reconstruct small 32 × 32 images with only 8–16 projection angles, image sizes were increased to 128 × 128 in a later study by Guo et al [11].

In a series of publications between 2008 and 2012 [12–18], Cierniak et al developed multiple RNNs to iteratively reconstruct CT scans after using traditional backprojection to create an initial CT scan. The proposed RNNs were essentially used as learnable filters for the FBP reconstruction method, replacing the fixed filters in FBP [17]. Cierniak demonstrated that the proposed RNNs improved CT image quality compared to FBP. Moreover, it was shown that the proposed method was able to reconstruct projection data acquired from various CT scanning geometries such as fan-beam and parallel-beam (Fig. 6).

Although the aforementioned RBF-NN and RNNs initially demonstrated promising results, they have been rapidly surpassed by CNNs. CNNs have the ability of capturing spatially oriented patterns in imaging data, which makes them particularly suited to reconstruct CT scans. An example of such

a CNN approach was proposed by Würfl et al., who demonstrated that the traditional FBP method can be expressed in terms of CNNs [19]. Their CNN consisted of a single convolutional layer to mimic the filtering of FBP, and a fully connected layer to learn the backprojection step. They found that the CNN achieved comparable results to traditional FBP while markedly reducing the computational complexity required to perform the reconstruction. In addition, they showed that their framework can be extended to mimic the Feldkamp David and Kress (FDK) algorithm that is commonly used to reconstruct CT scans acquired with the cone-beam geometry (Fig. 6) [26].

Another way of using CNNs for CT image reconstruction is to incorporate them within IR algorithms. For example, Adler and Öktem replaced forward operations of the iterative gradient descent reconstruction algorithm [20] and the primal-dual hybrid gradient algorithm [21] by partially trainable CNNs. In addition, various CNN approaches have been developed to improve reconstruction quality [34,36,37] and computational efficiency [33] of IR algorithms.

A different strategy was taken by Chen et al, who developed a CNN-based framework to find a direct mapping between CT projection data and reconstructed CT scans [22]. Their framework used 50 iteration-inspired layers that each consisted of three learned convolutional operations. This framework significantly outperformed state-of-the-art reconstruction approaches. Moreover, they showed that this approach can be effectively used on incomplete projection data, which is a common problem in clinical practice, since radiation dose often needs to be reduced in order to comply with the 'as low as reasonably achievable' (ALARA) principle. The CNN-based IR framework was further improved by Xie et al. [41], who implemented a learnable back-projection step that was previously fixed.

A different way of finding a mapping between projection data and reconstructed CT scans was developed by Fu and De Man [28]. However, instead of finding a direct mapping, they proposed a hierarchical CNN in which the difficult reconstruction problem was split up into multiple intermediate steps that can be easily learned. This approach does not only improve the quality of reconstructed CT scans, but also gives clinicians an insight into the intermediate steps learned by the CNN. Similarly, Wang et al [40], developed two coupled CNNs to convert sinograms into CT images. The first CNN takes sinograms as input and converts them to data that are better suitable for the FBP or FDK algorithms. The output of the FBP or FDK algorithms (in the image domain) is subsequently fed to the second CNN, which further improves the reconstructed image quality. Ma et al. [39] also aimed to reconstruct CT images directly from the projection data. However, instead of breaking down the reconstruction challenge into multiple steps, they applied a combination of fully connected layers and convolutional layers to reduce the memory space requirement. They showed that their proposed approach results in substantially better image quality than standard FBP.

Ever since it was introduced in 2015, the U-Net has become a popular CNN architecture for medical image analysis [85]. The U-Net consists of two convolutional paths. The downsampling path (i.e. encoder) creates a low-dimensional representation of the input data in order to capture local patterns. The upsampling path (i.e. decoder) subsequently captures global patterns through a series of upsampling steps. Both paths of the U-net are interconnected with skip-connections that allow the network to combine learned patterns at various scales. Applications of U-Nets for medical CT image reconstruction include the estimation of incomplete CT projection data (e.g. sparse-

7

view and limited-angle CT) [27,37,38] and improving IR, specifically the projected gradient descent reconstruction algorithm, by replacing the forward projector with a U-Net [23]. Furthermore, variants of U-Net such as dual-frame and tight-frame U-Net have been proposed to reduce streaking and blurring artifacts in sparse-view CT scans during post-processing [24]. Finally, U-Net has also been employed to predict deformation vector fields used to reconstruct 4DCT images [35].

In summary, a wide variety of NN approaches have been proposed for CT image reconstruction (Table 1). In particular, CNNs have shown to be an incredibly interesting research topic with many new recent publications. Three CNN approaches were identified that are particularly interesting for the current maxillofacial CAS workflow. The first CNN approach aims to replace the computationally demanding forward operations within current IR methods [20,21]. This markedly reduces the time constraint of applying IR methods in the maxillofacial CAS workflow. The second interesting CNN approach identified in this study is capable of reconstructing CT scans from incomplete projection data [27,31,37]. Such approaches would enable clinicians to acquire high-quality CT scans of patients using low dose protocols. The third CNN approach replaces the total reconstruction process [22,28], which means that the CNN is trained to directly reconstruct CT scans from raw projection data. However, it must be noted that such fully learned reconstruction approaches are computationally expensive and require large amounts of annotated training data, which are not always available.

3.2. Image segmentation

Image segmentation refers to the task of labelling voxels of an image as a particular class (Fig. 7). In the context of maxillofacial surgery, this image segmentation step is typically used to distinguish bony structures from soft tissues or air. Although a large variety of statistical methods have been developed for bone segmentation, they usually require the intervention of a medical professional in order to produce an accurate output, mainly due to the lack of reliable Hounsfield units and the limited signal-to-noise ratio of CBCT scans. It is therefore desirable to automate this task as far as possible, thereby relieving the medical professional from this labor-intensive and time-consuming task, while also increasing the accuracy and consistency of the segmentation results. We therefore reviewed the NNs used to automate CT bone segmentation tasks.
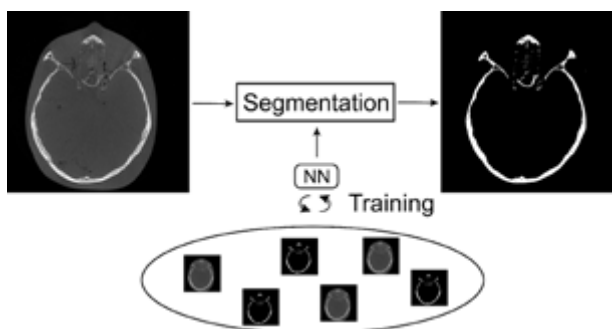

Figure 7. Schematic representation of the bone segmentation task required for maxillofacial CAS. A NN can be trained to automatically perform this segmentation task.

The first study describing the segmentation of bone in CT scans using NNs was published in 2002 [42]. In this study, a hierarchy of MLPs was trained on small patches of head and chest CT scans. The trained MLPs subsequently classified the center pixels of the patches and combined all separate pixel classifications to create a segmented image. Although similar MLPs were adopted by Gassman et al. [43] (2008) and Kuo et al. [44] (2013) to segment the phalanges and the nasal septum, respectively, different input data were used to train the MLPs. Namely, Gassman et al provided spherical co-ordinates, probabilities and intensities of individual pixels as input for their MLP, whilst Kuo et al used single rows of CT scans.

Segmentation using NNs took a significant leap after the ground-breaking performance of a novel CNN architecture (AlexNet) developed by Krizhevsky et al. in 2012 [86]. Similar to the aforementioned MLPs, this CNN architecture was trained using a patch-based approach in which the CNN aimed to classify the center voxels of small image patches. Inspired by the performance of this patch-based CNN, researchers have shown that such CNNs can achieve similar and in some cases superior performances compared to state-of-the art statistical methods when segmenting the mandible [46,52], the spine [51,61] and the skull [50] in CT scans. Nevertheless, the clinical application of patch-based CNNs for segmentation has been limited since many redundant convolution operations are necessary to classify all image voxels, which significantly slows down training and increases segmentation times.

In order to overcome this challenge, Ronneberger et al. published a variation of the traditional CNN architecture, known as U-Net [85]. U-Net can directly provide a segmented image as output (i.e. semantic segmentation) which increases its computational efficiency compared to patch-based CNNs. As a result, this U-Net architecture has ever since been applied for several CT bone segmentation tasks. For example, Klein et al. [48,56] and Noguchi et al. [67] applied U-net to segment bone in whole-body CT scans and reported that the U-Net performed significantly better than the standard segmentation procedure, i.e. global thresholding combined with morphological operations. Furthermore, U-Net was used to segment vertebrae [54,59,68,71], teeth [55,72], pelvic bones [62], orbital bones [57], cochleae [69] and cranial bones [70]. In a different study [45], a CNN architecture very similar to the U-Net, namely SegNet [87], was used to perform edge detection and multiscale segmentation of the femur in CT scans. Finally, Lessmann et al. extended the standard U-Net architecture in order to both segment and identify an a priori unknown number of vertebrae [49,58]. Their proposed architecture was able to automatically identify the individual vertebrae, whilst having comparable segmentation performance as the standard U-Net.

Since high segmentation performances have been achieved by U-net across a large number of studies, U-Net is currently considered the state-of-the-art for CT bone segmentation. Nevertheless, alternative CNN architectures for medical image segmentation have also been widely employed in the reviewed papers. For example, Men et al. proposed a deep dilated CNN (DDCNN), in which the first and last layers perform dilated convolutions in order to extract multiscale features [47]. Such a dilated convolution refers to the inflation of a convolution kernel while leaving sparse spaces between its elements. This dilation thus increases the receptive field of the kernel without increasing the number of model parameters. Torosdagli et al. [60] segmented the mandible using a fully convolutional DenseNet, which is comparable to U-Net, but uses dense blocks instead of regular

convolutional layers. These dense blocks comprise multiple densely connected convolutional layers, which means that each convolutional layer within the dense block is connected to all other layers in the block. Similarly, a UDS-Net [64] has been proposed, consisting of a U-Net with a dense block and spatial dropout, to segment teeth. Furthermore, 3D-DSD net has been developed which consists of a U-Net with a dense block and additional skip connections [65]. The use of dilated convolutions and dense connections was further exploited in the mixed-scale dense CNN (MS-D network) [88], which was used to segment the mandible in cone-beam CT scans [4]. In an MS-D network, each convolutional layer performs a dilated convolution and is densely connected to all other layers of the network. It was found that these properties allow the MS-D network to achieve comparable segmentation performances as U-Net, while using far fewer trainable parameters [4]. Finally, Zhou et al. developed the so-called N-Net, which is similar to U-Net but has an additional stream of downsampling layers [53], whereas Rao et al. modified the U-Net by replacing the normal convolutions with so-called Deep Bottleneck Architectures [73].

In this section, different NN approaches used for bone segmentation in CT scans were reviewed. Similar to the NN approaches used for CT image reconstruction, bone segmentation seems to be increasingly performed using CNNs in favor of alternative NNs. The CNN approaches identified in this review can be roughly categorized into patch-based approaches and semantic segmentation approaches. The patch-based approaches allow extracting a large number of patches from relatively few CT scans, which facilitates CNN training. Semantic segmentation approaches, on the other hand, annotate each voxel of an image during a single forward pass through the CNN, which is typically far more computationally efficient than patch-based approaches. As a result, current state-of-the-art CNN approaches usually perform semantic segmentation. Examples of such widely used CNN architectures are U-Net [85], ResNet [89] and MS-D network [88].

3.3. Surgical planning

The final step in the maxillofacial CAS workflow is surgical planning. This step typically involves a combination of computer-simulated bone reconstruction and subsequent designing of appropriate patient-specific implants. For example, a simple method to reconstruct fractured bones in the skull is to mirror the bony structures from the contralateral healthy side [90]. However, this technique is often constrained to small defects on one side of the skull. For larger defects, a complex procedure involving various 3D-modelling software packages is required. The success of NNs in image reconstruction and image segmentation calls for the application of similar techniques during surgical planning and implant design. In this section, we therefore review different NN approaches that have been applied for surgical planning.

An interesting area for automated surgical planning is the reconstruction of skull plates because of the simplicity of the local anatomical geometry. Already in 2001, efforts were made to automatically design skull implants [74,75]. The authors of these papers used a single-layer MLP in order to reconstruct cranial bone from CT images of patients with skull defects. In order to optimize training of the MLP, an approach relying on orthogonal functions (Legendre polynomials) was employed. The presented MLP was able to significantly speed up and improve the design of skull implants. However, while mathematically interesting, this particular approach is unlikely to

generalize to larger defects on the side of the skull, since Legendre polynomials are insufficient to correctly approximate such complex defects. A similar approach was proposed by Knopf and Al-Naji [76], which also approximates anatomical features using an MLP with a single layer of polynomial functions, specifically Bernstein polynomials. After being trained on a set of segmented CT slices, the network was able to reproduce the anatomical structure of healthy bone. The main advantage of this approach is that the output of the MLP is in the shape of curves that describe the bony structures, which can be easily loaded into CAD software. Nevertheless, it must be noted that most approaches based on curve fitting are currently restricted to the neurocranium where the geometry of the skull can be reasonably approximated by smooth curves.

A different way of employing NNs for surgical planning is to optimize parametrized implant designs. The advantage of working with parametrized designs is that the NNs do not require imaging data. In addition, since implant designs can often be described using a few parameters, relatively simple NN architectures can be used to optimize the parametrization of the design. This approach was, e.g. taken by Chanda et al. [80], who optimized parameterized hip implants based on the effects of initial micromotion, stress shielding, and interface stress. Using a combination of a single-hidden-layer MLP, a genetic algorithm and finite element analysis (FEA), the authors deduced that the standard implant design can be significantly improved. A similar combination of an MLP, a genetic algorithm and FEA was also used in a different study in order to find the best combination of material properties and geometry to generate patient-specific dental molar implants [81]. This approach was also used by Biswas et al. [82], who optimized the design of patient-specific spine implants based on the bone condition, body weight and implant diameter.

Another well-attended problem in surgical planning is to predict how bone adapts to different loads. This bone adaptation can be caused by surgical implants and may lead to complications after surgery [91]. To date, bone adaptation prediction has been commonly performed using FEA. However, researchers recently found that a combination of the well-established FEA method and an MLP can significantly speed up the computations [77]. The inverse problem has also been studied, specifically the estimation of load parameters for a given bone porosity inferred from CT images. For example, Campoli et al. [78] showed how a single-layer MLP can solve this inverse problem and compute the load for a femur. The network was specific to a single femur and was trained using simulated data. After the introduction of noise to the training data, the network was still able to successfully estimate the load parameters for the femur.

An interesting NN approach that does not rely on the geometric properties of bone was employed by Morais et al. [92]. They trained a CNN, specifically a deep convolutional autoencoder, to reconstruct fractured or missing parts of the skull. This autoencoder learned a representation of a healthy skull and was able to subsequently reconstruct a portion of the skull that was artificially removed. However, the accuracy and applicability of this deep learning model to high-resolution data remains challenging due to the computing power necessary for training and validation. Moreover, the proposed method was only validated on MRI data and not yet on CT data.

A relatively new approach towards reconstructing tissue morphologies is to apply generative adversarial networks (GANs). GANs consist of two networks: an generative network for generating new, fictive images based on input images, and a discriminator for distinguishing the generated

7

images from real images. In the context of morphology reconstruction, a GAN can be used to generate images of healthy tissues based on images of fractured or diseased tissues. The generated images can then be hardly, if not at all, distinguished from real images of healthy tissues. From a clinical perspective, such generated healthy images can be extremely useful as they provide a surgeon a view of what the result should resemble. Furthermore, generating healthy images can be particularly helpful when constructing patient-specific implants. An example of a GAN applied in such a setting can be found in the study by Liang et al. [84], who developed a GAN to reconstruct the morphology of the mandible based on CT images of patients suffering from ameloblastoma or gingival cancer. Similarly, Kodym et al. [83] used a GAN to reconstruct the shape of defective skulls.

Even though interesting approaches have been developed, relatively few studies have described NNs for surgical planning (Table 1). The majority of surgical planning studies included in this review implemented MLPs for applications such as dental implant design [80] and prediction of bone adaptation [77,78]. A possible explanation for the relatively few studies describing NN-based surgical planning might be that it is extremely difficult to develop a single network to account for all variations that clinicians face during surgical planning. For example, there are numerous possibilities of designing implants or surgical tools, and choosing an adequate design heavily depends on the available software tools on the market, the type of imaging data used, and the personal preferences of medical engineers. As a consequence, it is almost impossible to effectively train a NN to design implants if no constraints are imposed. Although a few studies solved this problem by using a parametrized implant design to reduce the degrees of freedom [80,81], this leads to more generic and less patient-specific implants. Hence, automating personalized surgical planning and implant design using NNs remains difficult. Nevertheless, the field is still currently active, as demonstrated by the recent AutoImplant Challenge [93] that was created to motivate participants to develop automated methods for cranial implant design.

4. CURRENT CHALLENGES AND FUTURE RESEARCH

In order to further develop and validate NN approaches for the three main steps of the maxillofacial CAS workflow, challenges remain that need to be overcome. One of these challenges is the quantitative performance evaluation. To date, the SSIM and MSE have been commonly used to assess image quality in the CT image reconstruction step, whereas the DSC is commonly used to assess segmentation performances (Fig. 4). These generic metrics, however, do not always represent clinical relevance. For example, maxillofacial surgeons often assess the surface of the bony structures in order to create a treatment plan and design implants. Therefore, surface-based performance metrics might be preferred over the generic metrics. One possible way of evaluating segmentation performance would be to convert segmented CT scans into virtual 3D models and subsequently calculating geometrical distances between a gold-standard virtual model and the NN-based virtual models [94]. This approach enables the quantification of the surface quality of bony structures, and also allows the visualization and interpretation of the differences between the two virtual models.

Another well-known limitation of most NN approaches is the need for large amounts of paired training data, i.e. input and target [95]. Although an extremely large number of medical images are

acquired on a daily basis that can be used as input to train a NN, they commonly lack appropriate annotations. Annotating such medical images requires a high level of domain-specific expertise, in contrast to natural images that can be easily annotated through crowdsourcing. In order to avoid the challenge of acquiring annotated target images, an interesting research direction might be to develop semi- and unsupervised NN training approaches, which do not depend on annotated data sets to learn.

To date, the development of NN approaches has been mainly performed in academic settings. Although many different NNs have shown to improve the efficiency, accuracy and consistency in which clinical tasks can be performed, none of these NN approaches has, to the best of our knowledge, been approved for maxillofacial CAS by the United States Food & Drug Administration. As a consequence, the application of NNs in routine clinical tasks remains very limited. In order to allow for large-scale use of NNs in clinical settings, additional research is necessary that focus on the robustness of NNs when faced with large anatomical variations and different imaging characteristics. For example, two recent studies have already shown that a single CNN is able to accurately segment CT scans acquired using various CT scanners [50,68].

The reconstruction of CT scans in the maxillofacial CAS workflow is typically optimized for visual interpretation by clinicians. This optimization is to be expected since visual assessment is often fundamental to establish correct diagnoses. However, a CT scan optimized for human interpretation might not be the ideal input for CNNs to perform the subsequent segmentation and surgical planning tasks. Namely, CNNs solely extract patterns from the CT scans without taking visual aspects into account. A possible way of improving the quality of the CAS workflow may therefore be to jointly train two CNNs to simultaneously reconstruct and segment a CT scan. Such joint training approaches have already shown promising results in lung nodule detection [96], and may open up promising avenues when used in the maxillofacial CAS workflow.

One of the limitations that was encountered in the reviewed publications is that the heterogeneity of the data used to train the NNs makes it particularly challenging to draw generic conclusions. NN models have been trained on CT images of various different anatomical regions or phantoms that have been acquired using a plethora of different scanners, or through simulation. In order to validate a methodology, benchmark data sets should be developed which would enable a better comparison between the NN approaches. Furthermore, any work performed on simulated data and/or phantoms should be validated on clinical data, as it is challenging to assess the clinical applicability of NN approaches that have been solely validated on synthetic data.

Finally, it must be noted that research on artificial intelligence, and deep learning in particular, is evolving at an incredible rate due to the unprecedented interest and resource investment these fields. As a consequence, the state-of-the-art on this topic will naturally evolve rapidly. Nevertheless, this does not mean that a review cannot be useful. In fact, literature reviews may be arguably even more important when considering such rapidly evolving topics, as they help to identify bottlenecks, and to suggest new lines of research to advance the field.

7

5. CONCLUSION

This scoping review describes different NN approaches used in the three main steps of the maxillofacial CAS workflow, i.e. CT image reconstruction, bone segmentation and surgical planning. In recent years, CNNs have rapidly become the most popular NN approach for CT image reconstruction and image segmentation, whereas MLPs remain the most common approach in surgical planning. Although CT image reconstruction and bone segmentation have been widely explored fields of research, additional research is required on the application of NNs for surgical planning. In order to reach the full potential of NNs for maxillofacial CAS, future research should focus on overcoming the challenges addressed in this review.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. Swennen GRJ, Mollemans W, Schutyser F. Three-dimensional treatment planning of orthognathic surgery in the era of virtual imaging. *J Oral Maxillofac Surg*. 2009; 67(10): 2080–2092.

2. Harris BT, Montero D, Grant GT, Morton D, Llop DR, Lin W-S. Creation of A 3-dimensional virtual dental patient for computerguided surgery and CAD-CAM interim complete removable and fixed dental prostheses: A clinical report. *J Prosthet Dent*. 2017; 117(2): 197–204:

3. Berman NB, Durning SJ, Fischer MR, Huwendiek S, Triola MM. The role for virtual patients in the future of medical education. 2016; 91(9):1217-1222

4. Minnema J, van Eijnatten M, Hendriksen AA, Liberton N, Pelt DM, Batenburg KJ, *et al*. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med Phys*. 2019; 46(11): 5027–5035.

5. Su S, Moran K, Robar JL. Design and production of 3D printed bolus for electron radiation therapy. *J Appl Clin Med Phys*. 2014; 15(4): 4831.

6. van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phys*. 2018; 51: 6–16.

7. Litjens G, Kooi T, Bejnordi BE, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal*. 2017;42:60-88.

8. Kratsios A. The Universal Approximation Property: Characterizations, Construction, Representation and Existence. *Ann Math Artif Intell*. 2021; 89:435-469.

9. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K, *et al*. A guide to deep learning in healthcare. *Nat Med*. 2019; 25: 24–29.

10. Hu M, Guo P, Lyu MR. COMPARATIVE STUDIES ON THE CT IMAGE RECONSTRUCTION BASED ON THE RBF NEURAL NETWORK. *11th Joint International Computer Conference – JICC*. 2005: 948–951.

11. Guo P, Hu M, Jia Y. RBF Network image Representation with Application to CT Image Reconstruction. *International Conference on Computational Intelligence and Security;* 2006: 1865–1868. 10.1109/ ICCIAS.2006.295389.

12. Cierniak R. A 2D approach to tomographic image reconstruction using A hopfield-type neural network. *Artif Intell Med*. 2008; 43: 113–125.

13. Cierniak R. A new approach to image reconstruction from projections using A recurrent neural network. *International Journal of Applied Mathematics and Computer Science*. 2008; 18: 147–157.

14. Cierniak R. New neural network algorithm for image reconstruction from fan-beam projections. *Neurocomputing*. 2009; 72: 3238– 3244. https://doi.org/10.1016/j.neucom.2009.02.005

15. Cierniak R. A Statistical Tailored Image Reconstruction from Projections Method. *Advances in Intelligent Decision Technologies*. 2010: 181–190.

16. Cierniak R. A Statistical Approach to Image Reconstruction from Projections Problem Using Recurrent Neural Network. *ICANN*. 2010:138–141.

17. Cierniak R. Neural network algorithm for image reconstruction using the "grid-friendly" projections. *Australas Phys Eng Sci Med*. 2011; 34: 375–389.

18. Cierniak R, Lorent A. A Neuronal Approach to the Statistical Image Reconstruction from Projections Problem. *ICCCI*. 2012:344–353.

19. Würfl T, Ghesu FC, Christlein V, Maier A. Deep Learning Computed Tomography. *MICCAI*. 2016:432–440.

20. Adler J, Öktem O. Solving ill-posed inverse problems using iterative deep neural networks. *Inverse Problems*. 2017; 33(12): 124007.

21. Adler J, Oktem O. Learned primal-dual reconstruction. *IEEE Trans Med Imaging*. 2018; 37(6): 1322–1332

22. Chen H, Zhang Y, Chen Y, Zhang J, Zhang W, Sun H, *et al*. LEARN: learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE Trans Med Imaging*. 2018; 37: 1333–1347.

23. Gupta H, Jin KH, Nguyen HQ, McCann MT, Unser M. CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE Trans Med Imaging* 2018; 37: 1440–1453.

24. Han Y, Ye JC. Framing U-net via deep convolutional framelets: application to sparse-view CT. *IEEE Trans Med Imaging*. 2018; 37: 1418–1429.

25. Liang K, Xing Y, Yang H, Kang K, Chen G-H, Lo JY, *et al*. Improve angular resolution for sparse-view CT with residual convolutional neural network. *Physics of Medical Imaging SPIE*. 2018:105731K

26. Wurfl T, Hoffmann M, Christlein V, Breininger K, Huang Y, Unberath M, *et al*. Deep learning computed tomography: learning projection-domain weights from image domain in limited angle problems. *IEEE Trans Med Imaging*. 2018; 37: 1454–1463.

7

27. Dong J, Fu J, He Z. A deep learning reconstruction framework for X-ray computed tomography with incomplete data. *PLoS ONE*. 2019; 14(11): e0224426.

28. Fu L, De Man B, Matej S, Metzler SD. A hierarchical approach to deep learning and its application to tomographic reconstruction. *The Fifteenth International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*. 2019:1107202.

29. He J, Yang Y, Wang Y, Zeng D, Bian Z, Zhang H, *et al*. Optimizing a parameterized plug-and-play ADMM for iterative low-dose CT reconstruction. *IEEE Trans Med Imaging*. 2019; 38: 371– 82.

30. Lee D, Choi S, Kim H-J. High quality imaging from sparsely sampled computed tomography data with deep learning and wavelet transform in various domains. *Med Phys*. 2019; 46: 104–115.

31. Li Z, Cai A, Wang L, Zhang W, Tang C, Li L, *et al*. Promising generative adversarial network based sinogram inpainting method for ultra-limited-angle computed tomography imaging. *Sensors (Basel)*. 2019;19(18):3941

32. Shen L, Zhao W, Xing L, Bosmans H, Chen G-H, Gilat Schmidt T. Harnessing the power of deep learning for volumetric CT imaging with single or limited number of projections. *Physics of Medical Imaging*. 2019: 1094826.

33. Wu D, Kim K, El Fakhri G, Li Q. Computational-efficient cascaded neural network for CT image reconstruction. *Physics of Medical Imaging*. 2019:109485Z

34. Zhang J, Zuo H. Iterative CT image reconstruction using neural network optimization algorithms. *Physics of Medical Imaging*. 2019:1094863

35. Zhang Y, Huang X, Wang J. Advanced 4-dimensional cone-beam computed tomography reconstruction by combining motion estimation, motion-compensated reconstruction, biomechanical modeling and deep learning. *Vis Comput Ind Biomed Art*. 2019;2(1):23.

36. Chen G, Hong X, Ding Q, Zhang Y, Chen H, Fu S, *et al*. AirNet: fused analytical and iterative reconstruction with deep neural network regularization for sparse-data CT. *Med Phys*. 2020; 47: 2916–2930.

37. Wang J, Liang J, Cheng J, Guo Y, Zeng L. Deep learning based image reconstruction algorithm for limited-angle translational computed tomography. *PLoS ONE*. 2020; 15(1): e0226963

38. Baguer DO, Leuschner J, Schmidt M. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Problems*. 2020; 36: 094004.

39. Ma G, Zhu Y, Zhao X. Learning image from projection: A full-automatic reconstruction (FAR) net for computed tomography. *IEEE Access*. 2020; 8: 219400–414.

40. Wang W, Xia X-G, He C, Ren Z, Lu J, Wang T, *et al*. An end-to-end deep network for reconstructing CT images directly from sparse sinograms. *IEEE Trans Comput Imaging*. 2020; 6: 1548–60.

41. Xie H, Shan H, Cong W, Liu C, Zhang X, Liu S, *et al*. Deep efficient end-to-end reconstruction (DEER) network for few-view breast CT image reconstruction. *IEEE Access*. 2020; 8: 196633–196646.

42. Zhang D, Sonka M, Fitzpatrick JM, Valentino DJ. Segmentation of anatomical structures in x-ray computed tomography images using artificial neural networks. *Medical Imaging*. 2002:4684.

43. Gassman EE, Powell SM, Kallemeyn NA, DeVries NA, Shivanna KH, Magnotta VA, *et al*. Automated bony region identification using artificial neural networks: Reliability and validation measurements. *Skeletal Radiol*. 2008; 37:313–319.

44. Kuo C-FJ. Three-dimensional Reconstruction System for Automatic Recognition of Nasal Vestibule and Nasal Septum in CT Images. *J Med Biol Eng*. 2014; 34:574-580.

45. Chen F, Liu J, Zhao Z, Zhu M, Liao H. Three-dimensional feature-enhanced network for automatic femur segmentation. *IEEE J Biomed Health Inform*. 2019; 23: 243–252.

46. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017; 44: 547–557.

47. Men K, Dai J, Li Y. Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks. *Med Phys*. 2017; 44: 6377–6389.

48. Klein A, Warszawski J, Hillengaß J, Maier-Hein KH. Towards Whole-body CT Bone Segmentation. *Bildverarbeitung für die Medizin*. 2018:204–209.

49. Lessmann N, Išgum I, Ginneken B. Iterative convolutional neural networks for automatic

vertebra identification and segmentation in CT images. *Medical Imaging*. 2018: 1057408.

50. Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput Biol Med*. 2018; 103: 130–139

51. Furqan Qadri S, Ai D, Hu G, Ahmad M, Huang Y, Wang Y, *et al*. Automatic deep feature learning via patch-based deep belief network for vertebrae segmentation in CT images. *Applied Sciences*. 2018; 9: 69.

52. Yan M, Guo J, Tian W, Yi Z. Symmetric convolutional neural network for mandible segmentation. *Knowledge-Based Systems*. 2018; 159: 63–71.

53. Zhou W, Lin L, Ge G. N -net: 3D fully convolution network-based vertebrae segmentation rom CT spinal images. *Int J Patt Recogn Artif Intell*. 2019; 33: 1957003.

54. Dutta S, Das B, Kaushik S, Bak PR, Chen P-H. Assessment of optimal deep learning configuration for vertebrae segmentation from CT images. *Imaging Informatics for Healthcare, Research, and Applications*. 2019:109541A

55. Gou M, Rao Y, Zhang M, Sun J, Cheng K. Automatic Image Annotation and Deep Learning for Tooth CT Image Segmentation. *ICIG*. 2019: 519–528

56. Klein A, Warszawski J, Hillengaß J, Maier-Hein KH. Automatic bone segmentation in whole-body CT images. *Int J CARS*. 2019; 14:21–29.

57. Lee MJ, Hong H, Shim KW, Park S -M.N. Orbital Bone Segmentation from Head and Neck CT Images Using Multi-Gray level fully Convolutional Networks. *ISBI*. 2019: 692–695.

58. Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical Image Analysis*. 2019; 53: 142–155.

59. Rehman F, Ali Shah SI, Riaz MN, Gilani SO, R. F. A region-based deep level set formulation for vertebral bone segmentation of osteoporotic fractures. *J Digit Imaging*. 2019; 33: 191–203.

60. Torosdagli N, Liberton DK, Verma P, Sincan M, Lee JS, Bagci U. Deep Geodesic Learning for Segmentation and Anatomical Landmarking. *IEEE Trans Med Imaging*. 2019; 38: 919–931.

61. Vania M, Mureja D, Lee D. Automatic spine segmentation from CT images using convolutional

neural network via redundant generation of class labels. *Journal of Computational Design and Engineering*. 2019; 6: 224–232.

62. Wang C, Connolly B, Oliveira Lopes PF, Frangi AF, Smedby Ö. Pelvis Segmentation Using Multi-pass U-Net and Iterative Shape Estimation. *MSKI*. 2019: 49–57.

63. Yazdani A, Stephens NB, Cherukuri V, Ryan T, Monga V. Domain-Enriched Deep Network for Micro-CT Image Segmentation. *Asilomar Conference on Signals, Systems, and Computers*. 2019: 1867–1871.

64. Lee S, Woo S, Yu J, Seo J, Lee J, Lee C. Automated CNN-based tooth segmentation in cone-beam CT for dental implant planning. *IEEE Access* 2020; 8: 50507–50518.

65. Li X, Gong Z, Yin H, Zhang H, Wang Z, Zhuo L. A 3D deep supervised densely network for small organs of human temporal bone segmentation in CT images. *Neural Netw*. 2020; 124: 75–85.

66. Yin X, Li Y, Shin B. Automatic Segmentation of Human Spine with Deep Neural Network. *Advances in Computer Science and Ubiquitous Computing*. 2020: 202–207.

67. Noguchi S, Nishio M, Yakami M, Nakagomi K, Togashi K. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Comput Biol Med*. 2020; 121: 103767.

68. Bae H-J, Hyun H, Byeon Y, Shin K, Cho Y, Song YJ, *et al*. Fully automated 3D segmentation and separation of multiple cervical vertebrae in CT images using a 2D convolutional neural network. *Comput Biol Med*. 2020; 184: 105119.

69. Heutink F, Koch V, Verbist B, van der Woude WJ, Mylanus E, Huinck W, *et al*. Multi-scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high resolution CT images. *Comput Methods Programs Biomed*. 2020; 191: 105387.

70. Zhang J, Liu M, Wang L, Chen S, Yuan P, Li J, *et al*. Context-guided fully convolutional networks for joint craniomaxillofacial bone segmentation and landmark digitization. *Med Image Anal*. 2020; 60: 101621.

71. Xia L, Xiao L, Quan G, Bo W. 3D cascaded convolutional networks for multi-vertebrae segmentation. *Curr Med Imaging*. 2020; 16: 231–240.

7

72. Chen Y, Du H, Yun Z, Yang S, Dai Z, Zhong L, *et al*. Automatic segmentation of individual tooth in dental CBCT images from tooth surface map by a multi-task FCN. *IEEE Access*. 2020; 8: 97296–97309.

73. Rao Y, Wang Y, Meng F, Pu J, Sun J, Wang Q. A Symetric Fully Convolutional Residual Network with DCRF for Accurate Tooth Segmentation. *IEEE Access*. 2020; 8:92028-92038.

74. Hsu J-H, Tseng C-S. Application of orthogonal neural network to craniomaxillary reconstruction. *J Med Eng Technol*. 2000; 24: 262–266.

75. Hsu J-H, Tseng C-S. Application of three-dimensional orthogonal neural network to craniomaxillary reconstruction. *Comput Med Imaging Graph*. 2001; 25: 477–82.

76. Knopf GK, Al-Naji R. Adaptive reconstruction of bone geometry from serial cross-sections. *Artificial Intelligence in Engineering*. 2001; 15: 227–239.

77. Hambli R. Application of neural networks and finite element computation for multiscale simulation of bone remodeling. *J Biomech Eng*. 2010; 132(11): 114502.

78. Campoli G, Weinans H, Zadpoor AA. Computational load estimation of the femur. *J Mech Behav Biomed Mater*. 2012; 10: 108–119.

79. Fischer A, Holdstein Y. A Neural Network Technique for Remeshing of Bone Microstructure. *Computer-Aided Tissue Engineering*. 2012:135–141.

80. Chanda S, Gupta S, Pratihar DK. Effects of interfacial conditions on shape optimization of cementless hip stem: an investigation based on a hybrid framework. *Struct Multidisc Optim*. 2015; 53: 1143–1155.

81. Roy S, Dey S, Khutia N, Roy Chowdhury A, Datta S. Design of patient specific dental implant using FE analysis and computational intelligence techniques. *Applied Soft Computing*. 2018; 65: 272–279.

82. Biswas JK, Dey S, Karmakar SK, Roychowdhury A, Datta S. Design of patient specific spinal implant (pedicle screw fixation) using FE analysis and soft computing techniques. *CMIR*. 2020; 16: 371–382.

83. Kodym O, Španěl M, Herout A. Skull shape reconstruction using cascaded convolutional networks. *Comput Biol Med*. 2020; 123:103886.

84. Liang Y, Huan J, Li J-D, Jiang C, Fang C, Liu Y. Use of artificial intelligence to recover mandibular morphology after disease. *Sci Rep*. 2020; 10: 16431.

85. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention – MICCAI*. 2015;234–241.

86. Krizhevsky A, Sutskever I, Geoffrey EH. ImageNet classification with deep convolutional neural networks. *NIPS'12*. 2012; 1: 1097–1105.

87. Badrinarayanan V, Kendall A, Cipolla R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell*. 2017; 39: 2481–2495.

88. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc Natl Acad Sci USA*. 2018; 115(2): 254–259.

89. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016: 770–78.

90. Scolozzi P. Maxillofacial reconstruction using polyetheretherketone patient-specific implants by "mirroring" computational planning. *Aesthetic Plast Surg*. 2012; 36: 660–665.

91. Comenda M, Quental C, Folgado J, Sarmento M, Monteiro J. Bone daptation impact of stemless shoulder implants: a computational analysis. *J Shoulder Elbow Surg*. 2019; 28: 1886–1896.

92. Morais A, Egger J, Alves V. Automated Computer-aided Design of Cranial Implants Using a Deep Volumetric Convolutional Denoising Autoencoder. *New Knowledge in Information Systems and Technologies*. 2019: 151–160.

93. Egger J, Li J, Chen X, Schäfer U, Campe GO, Krall M, *et al*. Towards the automatization of cranial implant design in cranioplasty. *Lecture Notes in Computer Science*. 2020;12439.

94. van Eijnatten M, Koivisto J, Karhu K, Forouzanfar T, Wolff J. The impact of manual threshold selection in medical additive manufacturing. *Int J Comput Assist Radiol Surg*. 2017; 12: 607–615.

95. Maier A, Syben C, Lasser T, Riess C. A gentle introduction to deep learning in medical image processing. *Z Med Phys*. 2019; 29(2): 86–101.

96. Wu D, Kim K, Dong B, Fakhri GE, Li Q. End-to-End Lung Nodule Detection in Computed Tomography. *Machine Learning in Medical Imaging*. 2018: 37–45.

CHAPTER

GENERAL DISCUSSION

8

Computer-assisted surgery (CAS) has become an integral part of personalized medicine [1]. However, the current CAS process remains laborious and time-consuming and requiring a plethora of manual steps. These steps often go hand in hand with subjectivity and fatigue among medical engineers [2], which can lead to inaccuracies in the final treatment plan. One possible way of circumventing such tedious manual tasks is by using deep learning algorithms. In particular, convolutional neural networks (CNNs) offer new possibilities for automating the current CAS workflow. This thesis focused on optimising the current CAS workflow by developing CNNs for two main applications: computed tomography (CT) image segmentation and image artifact correction. The following sections will discuss these two applications. Furthermore, in the final section, current challenges and future research perspectives will be discussed.

1. CT IMAGE SEGMENTATION

One of the most important tasks in the maxillofacial CAS workflow that would benefit from CNNs is medical image segmentation. The major advantage of using CNNs for segmentation tasks is that they are capable of automatically segmenting medical scans by learning patterns between the gray-values of voxels. Through these learned patterns, CNNs are able to separate various anatomical structures of interest. In order to demonstrate the potential of applying CNNs for medical image segmentation in maxillofacial CAS, a feasibility study was conducted in **Chapter 2**. In this study, a CNN was trained using clinical CT scans from patients who had undergone skull surgery. The CT scans were acquired using various scanner types and imaging protocols. Interestingly, the CT scanner type and imaging protocol did not seem to affect the CNN's segmentation performances. This finding is likely due to the fact that the developed CNN was able to automatically learn the characteristic features of the bony structures in multiple CT scans. This chapter demonstrates that CNNs are capable of correctly segmenting bony structures in images acquired from a large variety of different CT scanners currently on the market.

In **Chapter 2**, the CNN segmentation performance was validated on a single region of interest, namely bone. However, in clinical settings maxillofacial surgeons commonly need to assess multiple anatomical regions of interest. Thus, simultaneous segmentation of multiple regions of interest, also referred to as multi-class segmentation, is a challenging task, as CNNs need to learn complex patterns and features corresponding to each region of interest. At this stage of the study, it was unclear whether a single CNN was able to accurately segment multiple relevant regions of interest.

Based on the questions that arose in **Chapter 2**, a new study was initiated (**Chapter 3).** In this study, a CNN was developed and implemented to segment CBCT scans into three regions of interest: 1) the jaw (i.e., mandible and maxilla), 2) the teeth and 3) background. For this purpose, a mixed-scale dense convolutional neural network (MS-D network) [3] was developed. The MS-D network was trained to segment the three aforementioned regions of interest, and the results were compared to an MS-D network that was trained on two regions of interest (i.e., jaw or teeth, and background). Interestingly, both MS-D networks (i.e., multi-class segmentation vs. binary segmentation) achieved statistically equivalent ($p<0.05$) segmentation performances. The findings of this study are, however, different than those reported by Saood and Hatem [4]. In their study, they

8

reported worse segmentation performances when simultaneously segmenting multiple regions of interest in the lungs of patients with COVID-19, compared to consecutively segmenting a single region of interest. A possible explanation for these contradicting findings might be that structures in the lungs share many similar features and patterns, making it extremely difficult for a CNN to simultaneously segment these structures compared to the jaw and the maxilla. The performance of multi-class segmentation approaches therefore seems to depend on the properties of the imaging data at hand and the anatomy of interest. In order to establish a better understanding of the influence of the anatomical region of interest, future studies should focus on comparing multi-class segmentation to binary segmentation for a wide range of different anatomies.

Another open research question in the field of CT image segmentation is which training strategy is best suited to train a CNN. To date, axial slices have been commonly used to train CNNs since these slices typically offer the highest in-plane resolution [5]. However, evidence in the current literature suggests that axial slices do not always yield the highest segmentation performances [6–13]. Although a wide variety of different training strategies have been proposed in recent years, it remains unclear which of these training strategies offers the best segmentation performance for maxillofacial CAS. In order to answer this question, a new study was conducted (**Chapter 4**).

In **Chapter 4**, eight different CNN training strategies were quantitatively compared and evaluated using two different CNN architectures (MS-D network and U-Net). The eight training strategies used in this study can be divided into 2D strategies, augmented 2D strategies, and 3D strategies. CNNs trained using 2D strategies aim to segment single 2D slices (e.g., axial, sagittal, coronal), whereas CNNs trained using augmented 2D strategies generally segment a combination of 2D cross-sections. 3D strategies are employed to segment an entire 3D volume or a 3D patch. The results of this study demonstrated that one of the augmented 2D strategies performed best. The best performing augmented 2D strategy used a CNN that was independently trained three times, either on axial slices, sagittal slices, or coronal slices. The segmentation results of the three CNNs trained on axial, sagittal, and coronal slices, were subsequently combined through majority voting. However, this majority voting strategy can be computationally demanding because it requires a CNN to be trained three times. If applying this strategy is infeasible, then it is recommended to train a single CNN on 2D slices that are perpendicular to the predominant orientation of the anatomical structure of interest. Finally, it is interesting to note that the training strategies evaluated in this study affected the CNN's segmentation performance more than the chosen architectures (i.e., MS-D network vs. U-Net), which highlights the importance of employing an optimal training strategy [14].

2. IMAGE ARTIFACT CORRECTION

The aforementioned studies on CT image segmentation demonstrated the potential of CNNs to segment bony structures in (CB)CT scans (**Chapters 2, 3**) and helped gain novel insights into different CNN training strategies (**Chapter 4**). However, each of these studies were conducted using high-quality CT scans, which are not always available in clinical settings. More specifically, (CB)CT scans are often affected by imaging artifacts, which can impair the image processing steps required in the CAS workflow. Such imaging artifacts are caused by disparities between the measured (CB)

CT projection data and the underlying mathematical assumptions of the image reconstruction process. This can lead to distortions in the shape and intensity of the anatomical regions of interest. Correction of such artifacts is paramount to ensure the overall quality of the maxillofacial CAS workflow.

One possible approach to overcome this challenge is to train a CNN to segment CT scans affected by artifacts. In **Chapter 5**, a feasibility study was initiated to assess whether CNNs are capable of segmenting such artifact-affected CBCT scans. In this study, three different CNNs (i.e., MS-D network, U-Net and ResNet) were trained to segment bony structures in CBCT scans that were heavily affected by metal artifacts. Interestingly all three CNNs developed in this study outperformed a snake evolution segmentation algorithm that is widely-used in clinical settings. These findings suggest that CNNs are indeed capable of learning the characteristic patterns of metal artifacts. Moreover, the CNNs were able to correctly distinguish relevant bony structures from high-density objects such as retainers and amalgam fillings. The results described in **Chapter 5** are also in agreement with previous studies that have shown that CNNs can accurately segment the prostate [15], the mandible[16], and the teeth [17] in artifact-affected (CB)CT scans. These insights thus demonstrate that CNNs do not only offer accurate segmentations of high-quality CBCT scans, but are also capable of processing artifact-affected scans.

8

Although the segmentation-based artifact correction approach described in **Chapter 5** offered good results, it is not applicable to all scenarios in the maxillofacial CAS workflow. More specifically, maxillofacial surgeons regularly need to inspect the native (CB)CT scans to establish accurate diagnoses. In such cases, one would ideally want to remove the artifacts from the artifact-affected (CB)CT scans, instead of segmenting the scans. To date, two main CNN-based approaches have been proposed to perform such CT artifact correction tasks: image-domain approaches and projection-domain approaches. In image-domain approaches, a CNN is trained to convert low-quality scans into high-quality scans. Image-domain approaches are thus generally applied as post-processing methods to improve the quality of reconstructed (CB)CT scans. In contrast, projection-domain approaches aim to correct the raw (CB)CT projection data. This requires an in-depth understanding of the CT reconstruction process, but can potentially lead to highly accurate artifact correction. Both CNN-based artifact correction approaches have shown promising performances for various tasks, including metal artifact correction [18–23] and noise reduction [24].

Even though most imaging artifacts can be accurately removed using current artifact correction approaches, the high cone-angle artifact remains a challenge for current CNNs. This imaging artifact is inherent to the circular CBCT imaging modality, and is characterized by streaks and blurring in the peripheral regions of the CBCT scans. This means that the severity of high cone-angle artifacts differs greatly between CBCT slices, which makes it difficult for CNNs to learn a generalized way of removing the artifacts.

One possibility for effectively training a CNN to correct high cone-angle artifacts would be to employ the symmetry-aware deep learning workflow described in **Chapter 6.** In this study, a CNN was trained using radially sampled 2D CBCT slices instead of orthogonal 2D slices (i.e., axial, sagittal or coronal). Since radially sampled slices exhibit much less variation in cone-angle artifacts than orthogonal 2D slices, CNN training can be facilitated. The results of **Chapter 6** confirm the hypothesis

that training a CNN using radial CBCT slices improves its artifact correction performance. In addition, the proposed radial slicing approach resulted in markedly fewer residual artifacts than the CNNs trained on axial slices. The insights gained in **Chapter 6** emphasize the importance of tailoring the CNN training process to the task at hand. In fact, according to the results presented by Isensee et al. [14], non-architectural model choices can have a stronger influence on the CNN's performance than architectural changes. These findings will hopefully inspire fellow researchers to carefully consider the training process when applying CNNs to automate the image processing tasks of the maxillofacial CAS workflow.

3. CURRENT CHALLENGES AND FUTURE PERSPECTIVES

The results of this thesis demonstrate that CNNs are promising algorithms for the automation of the maxillofacial CAS workflow. In particular the CT image segmentation and image artifact correction steps can benefit from CNNs. Nevertheless, several challenges still need to be overcome in order to translate the promising results into medically certified products that can be incorporated in the maxillofacial CAS workflow. These challenges are discussed in the literature review of **Chapter 7**.

8

One of the main problems observed in **Chapter 7** was that the current CAS workflow lacks cohesiveness when it comes to the different tasks that have to be performed. Each task is performed independently from other tasks in the workflow, which can lead to sub-optimal outcomes. For instance, the acquisition and reconstruction of (CB)CT scans is optimized for visual inspection by clinicians, but not for the subsequent image processing steps required in CAS. Similarly, image processing software packages are not optimized for surgical planning. A potential solution to this particular problem may be to jointly train CNNs to perform all tasks in the maxillofacial CAS workflow. In such joint training approaches, multiple CNNs are simultaneously trained towards the end-goal of the workflow (e.g., adequate surgical planning), instead of solely focussing on specific tasks. Joint training approaches have already shown promising results for lung nodule detection [25], and may also open up new avenues for the maxillofacial CAS workflow.

Another trend observed in **Chapter 7** is that the majority of studies on deep learning-based image processing methods have focused on developing novel CNN architectures, instead of evaluating their reliability. Assessing the reliability of CNNs is of the utmost importance in clinical settings, as CNNs need to perform well on imaging datasets acquired from both different scanning devices and patients. I therefore strongly believe that future research should focus on the reproducibility and reliability of CNNs. This includes performing studies on the reproducibility of the CNN training process, re-training known CNN architectures on new imaging datasets, and conducting multi-centre studies using datasets acquired from multiple scanning devices or patient populations. In addition to investigating the reproducibility of CNNs, prospective studies should also be performed where CNNs are compared to standard clinical segmentation methods. Such studies will help to explore the reliability of CNNs in a wide variety of scenarios in the maxillofacial CAS workflow and facilitate the adoption of the technology in clinical routine.

A general problem with CNNs that affects their applicability in the current maxillofacial CAS workflow is the availability of high-quality datasets. More specifically, the majority of current CNN approaches require paired training images (i.e., input images and corresponding target images), which are difficult to acquire in clinical settings. A possible solution to this problem is to apply data augmentation in order to increase the number of available training images. Such data augmentation techniques commonly consist of simple operations that are applied on training images, such as random rotations, random erasing, cropping, geometric transformations and kernel filters [26]. It is also possible, however, to employ deep learning to generate training images. For example, generative adversarial networks (GANs) [27] are capable of generating artificial target data from input images. This deep learning-based data augmentation approach has already been shown to be successful in various clinical and non-clinical image processing tasks [26, 28, 29], and might offer promising avenues for maxillofacial CAS.

Finally, there remains a mystery behind the CNNs' reasoning and decision-making process. CNNs are currently often perceived as *black boxes*. Demystifying the CNN's decision-making process is essential, especially if the CNN's output contradicts the prediction of a medical specialist. Moreover, it may help engineers to correct the errors made by CNNs. To date, only a paucity of studies have tried to interpret the reasoning of CNNs by visualizing their intermediate layers [30, 31]. This interpretation remains extremely difficult and dataset dependent. Additional research is therefore still necessary to interpret the reasoning of CNNs.

In conclusion, CNNs have proven to be capable of automating the image processing step in the CAS workflow. The studies and results presented in this thesis clearly show the potential of CNNs and demonstrate that they can outperform the traditional image processing methods currently used in maxillofacial CAS. However, when applying new methods, such as CNNs, in clinical settings, it is not only necessary that high performances are achieved but also that the advantages and limitations of the method at hand are fully understood. Future studies should therefore shift their focus from improving CNN performances towards improving the robustness of CNNs and fully understanding their decision-making processes. This understanding will help to incorporate CNNs in the maxillofacial CAS workflow, and reduce the workload among medical specialists.

8

REFERENCES

1. Swennen GRJ, Mollemans W, Schutyser F. Three-Dimensional Treatment Planning of Orthognathic Surgery in the Era of Virtual Imaging. *Journal of Oral and Maxillofacial Surgery*. 2009; 67(10): 2080-2092.

2. Greenspan H, van Ginneken B, Summers RM. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging*. 2016; 35(5): 1153-1159. doi: 10.1109/TMI.2016.2553401.

3. Pelt DM, Sethian JA. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci. Unit. States Am*. 2018; 115: 254–259. doi:10.1073/pnas.1715832114.

4. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: UNET vs. SegNET. *BMC Medical Imaging*. 2021; 21:19

5. Litjens G, Kooi T, Bejnordi BE, et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal*. 2017;42:60-88.

6. Sobhaninia Z, Rezaei S, Noroozi A, Ahmadi M, Zarrabi H, Karimi N, *et al*. Brain tumor segmentation using deep learning by type specific sorting of images. 2018. ArXiv:1809.07786 [Cs, Eess].

7. Çiçek Ö, Abdulkadir A, Lienkamp SS, Brox T, Ronneberger O. 3D U-net: learning dense volumetric segmentation from sparse annotation. *Med. Image Comput. Comput. Assist. Interv. MICCAI*. 2016: 424–432.

8. Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. *Lecture Notes in Computer Science*. 2013: 246–253.

9. Cheng R, Lay N, Mertan F, Turkbey B, Roth HR, Lu L, *et al*. Deep learning with orthogonal volumetric HED segmentation and 3D surface reconstruction model of prostate MRI. in: Proceeding of the IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017), IEEE. 2017: 749–753.

10. Banerjee S, Mitra S, Shankar BU. Multi-planar spatial-ConvNet for segmentation and survival prediction in brain cancer. *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*. 2019: 94–104. doi: 10.1007/978-3-030-11726-9_9 .

11. Hänsch A, Schwier M, Morgas T, Klein J, Hahn HK, Gass T, *et al*. Comparison of different deep learning approaches for parotid gland segmentation from CT images. *Medical Imaging 2018: Computer- Aided Diagnosis, SPIE*. 2018: 44.

12. Chen J, Yang L, Zhang Y, Alber M, Chen D. Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation. *Conference on Neural Information Processing Systems (NIPS 2016)*. 2016.

13. Isensee F, Jaeger PF, Full PM, Wolf I, Engelhardt S, Maier-Hein KH. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*. 2018: 120–129.

14. Isensee F, Petersen J, Klein A, Zimmerer D, Jaeger PF, Kohl S, *et al*. nnU-Net: self-adapting framework for U-Net-based medical image segmentation. ArXiv:1809.10486 [Cs]. 2018. http://arxiv.org/abs/1809.10486.

15. Shahedi M, Halicek M, Dormer JD, Schuster DM, Fei B. Deep learning-based three-dimensional segmentation of the prostate on computed tomography images. *J. Med. Imag. 2019;6(2)*:025003

16. Ibragimov B, Xing L. Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Med Phys*. 2017; 44: 547–557.

17. Chung M, Lee M, Hong J, Park S, Lee J, Lee J, *et al*. Pose-aware instance segmentation framework from cone beam CT images for tooth segmentation. *Computers in Biology and Medicine*, 2020;120:103720.

18. Zhu L, Han Y, Li L, Xi X, Zhu M, Yan B. Metal Artifact Reduction for X-Ray Computed Tomography Using U-Net in Image Domain. *IEEE Access*. 2019;7:98743–98754.

19. Zhang C, Xing Y. CT artifact reduction via U-net CNN. *Medical Imaging 2018: Image Processing*. 2018:62.

20. Šerifović Trbalić A, Trbalić A, Demirović D, Skejić E, Gleich D. CT Metal Artefacts Reduction Using Convolutional Neural Networks. *nternational Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 2019: 251–255

21. Ghani MU, Karl WC. Fast Enhanced CT Metal Artifact Reduction Using Data Domain Deep Learning. *IEEE Trans. Comput. Imaging*. 2020; 6: 181–193.

8

22. Hegazy MAA, Cho MH, Lee SY. A metal artifact reduction method for a dental CT based on adaptive local thresholding and prior image generation. *Biomed Eng Online*. 2016;15(1):119.

23. Ghani MU, Karl WC. Deep learning based sinogram correction for metal artifact reduction. *Elect Imaging*. 2018;2018:472-1–4728.

24. Chen H, Zhang Y, Zhang W, Liao P, Li K, Zhou J, *et al*. Low-dose CT denoising with convolutional neural network. *IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. 2017: 143–146.

25. Wu D, Kim K, Dong B, Fakhri GE, Li Q. End-to-End Lung Nodule Detection in Computed Tomography. *Machine Learning in Medical Imaging*. 2018: 37–45.

26. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019; 6(1): 60.

27. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, *et al*. Generative Adversarial Nets. *Advances in Neural Information Processing Systems 27*. 2014: 2672–2680.

28. Sandfort V, Yan K, Pickhardt PJ, Summers RM. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Sci Rep*. 2019; 9(1): 16884.

29. Lu CY, Arcega Rustia DJ, Lin TT. Generative Adversarial Network Based Image Augmentation for Insect Pest Classification Enhancement. *IFAC-PapersOnLine*. 2019; 52(30): 1–5

30. van Molle P, de Strooper M, Verbelen T, Vankeirsbilck B, Simoens P, Dhoedt B. Visualizing Convolutional Neural Networks to Improve Decision Support for Skin Lesion Classification. *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*. 2018; 11038: 115–123.

31. Rajaraman S, Silamut K, Hossain A, Ersoy I, Maude RJ, Jaeger S, *et al*. Understanding the learned behavior of customized convolutional neural networks toward malaria parasite detection in thin blood smear images. *J. Med. Imag*. 2018; 5(3): 034501.

8

CHAPTER

ENGLISH AND DUTCH SUMMARY

9

SUMMARY

Computer-assisted surgery (CAS) is a novel treatment modality that allows clinicians to create personalized treatment plans and virtually design implants, surgical guides, and radiotherapy boluses. However, even though CAS is a powerful means to optimize surgery, it currently requires a lot of tedious and time-consuming manual work by experienced medical engineers to ensure the quality of the CAS workflow. This thesis therefore focuses on developing artificial intelligence algorithms, specifically convolutional neural networks (CNNs), to automate two main image processing tasks required in maxillofacial CAS: CT image segmentation and computed tomography (CT) artifact correction.

Chapter 2 describes a CNN approach to segment bony structures in 20 different CT scans. All CT were acquired from patients that had previously undergone craniotomy, thus posing a significant challenge for the CNN to correctly recognize the pathological shape of the bones. In addition, the CT scans were acquired using different CT scanners and imaging protocols in order to represent the variability that is commonly found amongst clinical CT datasets. Even though this segmentation tasks was validated on a challenging dataset, only two regions of interest (i.e., bone or background) were segmented. In order to demonstrate the ability of CNN to take more regions into account, a CNN was developed in **Chapter 3** to segment cone-beam computed tomography (CBCT) scans into the jaw, the teeth and background.

To date, many different training strategies have been proposed in literature to train CNNs on medical images. However, it remains unclear which strategy yields the best segmentation performances. To elucidate on this topic, eight different CNN training strategies were evaluated and compared in **Chapter 4**. The CNNs described in this chapter were trained to segment anatomical structures in simulated and experimental cone-beam CT scans. These experiments demonstrated that the best strategy is generally to train three separate CNN and combine their segmentation results through majority voting.

In **Chapter 5**, an approach is described to deal with strong metal artifacts in CBCT scans during the image segmentation step required during CAS. In particular, a mixed-scale dense convolutional neural network (MS-D network) was implemented to segment the bony structures of the mandible and the maxilla. The MS-D network described in this chapter clearly outperformed a widely-used clinical segmentation method (i.e., snake evolution), and produced comparable results to alternative deep learning benchmarks.

A novel symmetry-aware deep learning approach is proposed in **Chapter 6** to reduce high cone-angle artifacts in CB CT images. In this approach, a CNN was trained using radial cone-beam CT slices to exploit the symmetry of high cone-angle artifacts. This allowed training a CNN to reduce the complex 3D cone-angle artifacts using only 2D slices as input. In this chapter, it is demonstrated that this symmetry-aware dimensionality reduction improves the performance and robustness of CNNs when reducing high cone-angle artifacts in cone-beam CT scans.

Deep learning, and specifically CNNs, have found remarkable successes in many different image processing tasks. This success is especially emphasized by the rapidly increasing number of studies that have been published in recent years. However, because of the high number of studies, it has

9

become a difficult task to keep up with all relevant developments. Therefore, the goal of **Chapter 7** was to review all studies that have been published in which neural network approaches were developed for at least one of following three specific tasks of the CAS workflow: CT image reconstruction, bone segmentation, and surgical planning. Although various neural network approaches were identified, the majority of studies (66%) applied CNNs. Interestingly, all of these CNNs were published from 2016 onwards, indicating the rapid paradigm shift this field has undergone. Nevertheless, much research is still required to make deep learning an integral part of the CAS workflow. Even though deep learning has been increasingly used for CT image reconstruction and segmentation, its application for surgical planning remains in its infancy.

In conclusion, this thesis contributes to enhance the understanding of CNN approaches for medical image processing. Nevertheless, many interesting challenges and questions remain to incorporate CNN's as an integral part of the maxillofacial CAS routine. I therefore hope that this thesis will inspire fellow researchers to take on these exciting challenges.

9

SAMENVATTING

Computerondersteunde chirurgie (CAS) is een nieuwe vorm van behandeling die clinici de mogelijkheid geeft om gepersonaliseerde behandelplannen te creëren en om implantaten, chirurgische mallen en radiotherapiebolussen virtueel te ontwerpen. Hoewel CAS een effectieve manier is om operaties te optimaliseren, is er op dit moment nog veel langdradig en tijdrovend werk nodig door ervaren medische ingenieurs om de kwaliteit van de CAS workflow te waarborgen. Deze thesis richt zich daarom op het ontwikkelen van kunstmatige intelligentie algoritmes, met name convolutionele neurale netwerken (CNN s), om twee veelvoorkomende beeldverwerkingstaken uit de maxilofaciale CAS workflow te automatiseren: medische beeldsegmentatie en computer tomografie (CT) artefactreductie.

Hoofdstuk 2 beschrijft een CNN methode om botstructuren in 20 verschillende computer tomografie (CT) beelden te segmenteren. Alle CT beelden werden gemaakt van patiënten die in een eerder stadium craniotomie hadden ondergaan. Dit zorgde voor een aanzienlijke uitdaging voor het CNN om de contouren van de pathologische botstructuren op de juiste manier te herkennen. Bovendien waren de beelden verkregen met verschillende CT scanners en beeldvormingsprotocollen om de variabiliteit van typische klinische datasets na te bootsen. Echter, hoewel deze segmentatietaak op een uitdagende dataset was gevalideerd, werden alleen 2 verschillende relevante anatomische regio s gesegmenteerd (namelijk bot en achtergrond). Om aan te tonen dat CNN s ook meerdere anatomische regio s kunnen segmenteren, werd een CNN ontwikkeld in **Hoofdstuk 3** om de kaak, tanden en achtergrond te segmenteren in cone-beam computer tomografie (CBCT) beelden.

Tot op heden zijn er veel verschillende trainingsstrategieën gepubliceerd in de literatuur voor het trainen van CNN's op medische beelden. Het is echter nog onduidelijk welke strategie de beste segmentatieprestaties oplevert. Om dit onderwerp verder te bestuderen, werden er acht verschillende CNN training strategieën geëvalueerd en vergeleken in **Hoofdstuk 4**. De CNN's die beschreven zijn in dit hoofdstuk werden getraind om anatomische structuren te segmenteren in zowel gesimuleerde als experimentele cone-beam CT beelden. Deze experimenten toonden aan dat de beste training was om drie CNN's onafhankelijk te trainen en de resultaten te combineren.

In **Hoofdstuk 5** is een aanpak beschreven om met sterke metaalartefacten om te kunnen gaan tijdens de segmentatietaak. Om specifiek te zijn, was er een gemengde-schaal dichtgebonden convolutioneel neuraal netwerk (MS-D netwerk) geïmplementeerd om botstructuren van de maxilla en de mandibula te segmenteren Het MS-D netwerk dat beschreven is in dit hoofdstuk presteerde aanzienlijk beter dan een veelgebruikte klinische segmentatie methode en produceerde vergelijkbare resultaten met alternatieve deep learning benchmarks.

Een nieuwe deep learning methode werd ontwikkeld in **Hoofdstuk 6** om grote cone hoek artefacten te verminderen in CBCT beelden. Bij deze aanpak was een CNN getraind met radiale cone-beam doorsnedes zodat de symmetrie van de cone hoek artefacten uitgebuit kan worden. Dit zorgde ervoor dat het CNN getraind kon worden om de complexe 3D artefacten te verminderen terwijl alleen 2D doorsnedes werden gebruikt als input. In dit hoofdstuk wordt aangetoond dat

9

deze symmetriebewuste dimensionaliteitreductie de prestaties en de robuustheid van CNN's sterk verbetert bij het verwijderen van cone hoek artefacten in CBCT beelden.

Deep learning en met name CNN's zijn erg succesvol geweest bij het uitvoeren van verschillende beeldbewerkingstaken. Dit succes wordt vooral benadrukt door het grote aantal studies dat de afgelopen jaren is gepubliceerd. Vanwege dit grote aantal studies is het echter een moeilijke taak geworden om bij te blijven met alle relevante ontwikkelen binnen dit vakgebied. Daarom was het doel van **Hoofdstuk 7** om 62 verschillende studies te evalueren waarin neurale netwerken zijn ontwikkeld voor ten minste één van de volgende drie taken in de CAS workflow: CT beeldreconstructie, botsegmentatie en chirurgische planning. Hoewel verschillende neurale netwerk methoden waren geïdentificeerd, werden er in de meerderheid van de studies (66%) CNN s toegepast. Interessant was dat al deze CNN's werden gepubliceerd in de periode vanaf 2016 tot heden, wat aangeeft wat aangeeft wat voor revolutie het vakgebied heeft ondergaan. Toch is er nog veel onderzoek nodig om de CAS workflow volledig te automatiseren met deep learning. Hoewel CT beeldreconstructie en botsegmentatie inmiddels veelvuldig is bestudeerd, staat chirurgische planning met deep learning namelijk nog in de kinderschoenen.

Tot slot draagt deze thesis bij aan het verbeteren van het inzicht om CNN's toe te passen op medische beeldbewerkingstaken. Desalniettemin blijven er veel interessante uitdagingen en vraagstukken om CNN's te kunnen integreren in de maxillofaciale CAS routine. Ik hoop daarom dat deze thesis mede-onderzoekers zal inspireren om deze boeiende uitdagingen aan te gaan.

9

CHAPTER | 10

PHD PORTFOLIO

AMSTERDAM MOVEMENT SCIENCES

PhD student	Jordi Minnema
Department	Oral and maxillofacial surgery
PhD Period	September 2017 – December 2021
PhD supervisors	prof. dr. T. Forouzanfar
	prof. dr. K.J. Batenburg
	prof. dr. J.E.H. Wolff
	dr. M.A.J.M. van Eijnatten

	Year	Workload (ECTS)
Courses		
Data science R basics	2020	0.57
AMS – Writing a Scientific article	2020	3.0
AMS – Scientific integrity	2019	2.0
Research data handling	2017	0.25
Introduction to scientific data analysis in Python	2017	0.25
Unix Shell and Task Automation	2017	0.25
Version control and collaboration with Git and Github	2017	0.25
Advanced medical image processing	2016	6.0
International conferences and abstracts		
Additive manufacturing meets medicine	2019	2.0
Medical imaging with deep learning	2019	1.0
Lectures and talks		
Deep learning for medical additive manufactured skull implants	2019	1.0
Metal artifact removal using deep learning	2019	1.0
Supervision		
Tutoring medical bachelor students	2018-2020	10.0
Research visits		
Fraunhofer institute Hamburg	2019	3.0

10

CHAPTER | 11

ACKNOWLEDGEMENTS

11