Computer Methods and Programs in Biomedicine 207 (2021) 106192

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

Comparison of convolutional neural network training strategies for cone-beam CT image segmentation

Jordi Minnema[a,*], Jan Wolff[b,c,g], Juha Koivisto[d], Felix Lucka[e,f], Kees Joost Batenburg[e], Tymour Forouzanfar[a], Maureen van Eijnatten[e]

[a] Department of Oral and Maxillofacial Surgery/Pathology, 3D Innovationlab, Amsterdam UMC and Academic Centre for Dentistry Amsterdam (ACTA), Vrije Universiteit Amsterdam, Amsterdam Movement Sciences, Amsterdam 1081 HV, theNetherlands
[b] Fraunhofer Research Institution for Additive Manufacturing Technologies IAPT, Am Schleusengraben 13, Hamburg 21029, Germany
[c] Department of Oral and Maxillofacial Surgery, Division for Regenerative Orofacial Medicine, University Hospital Hamburg-Eppendorf, Hamburg 20246, Germany
[d] Department of Physics, University of Helsinki, Helsinki 20560, Finland
[e] Centrum Wiskunde & Informatica (CWI), Amsterdam 1090 GB, the Netherlands
[f] University College London, London WC1E 6BT, United Kingdom
[g] Department of Dentistry and Oral Health, Aarhus University, Vennelyst Boulevard 9, DK-8000 Aarhus C, Denmark

ARTICLE INFO

Article history:
Received 8 June 2020
Accepted 11 May 2021

Keywords:
Medical image segmentation
Deep learning
Convolutional neural networks
Training strategies
Cone-beam computed tomography

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

* Corresponding author.
E-mail address: j.minnema@amsterdamumc.nl (J. Minnema).

https://doi.org/10.1016/j.cmpb.2021.106192

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 & 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 $\sigma = 0.2$) 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 × 1001 × 153 and 1001 × 1001 × 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.

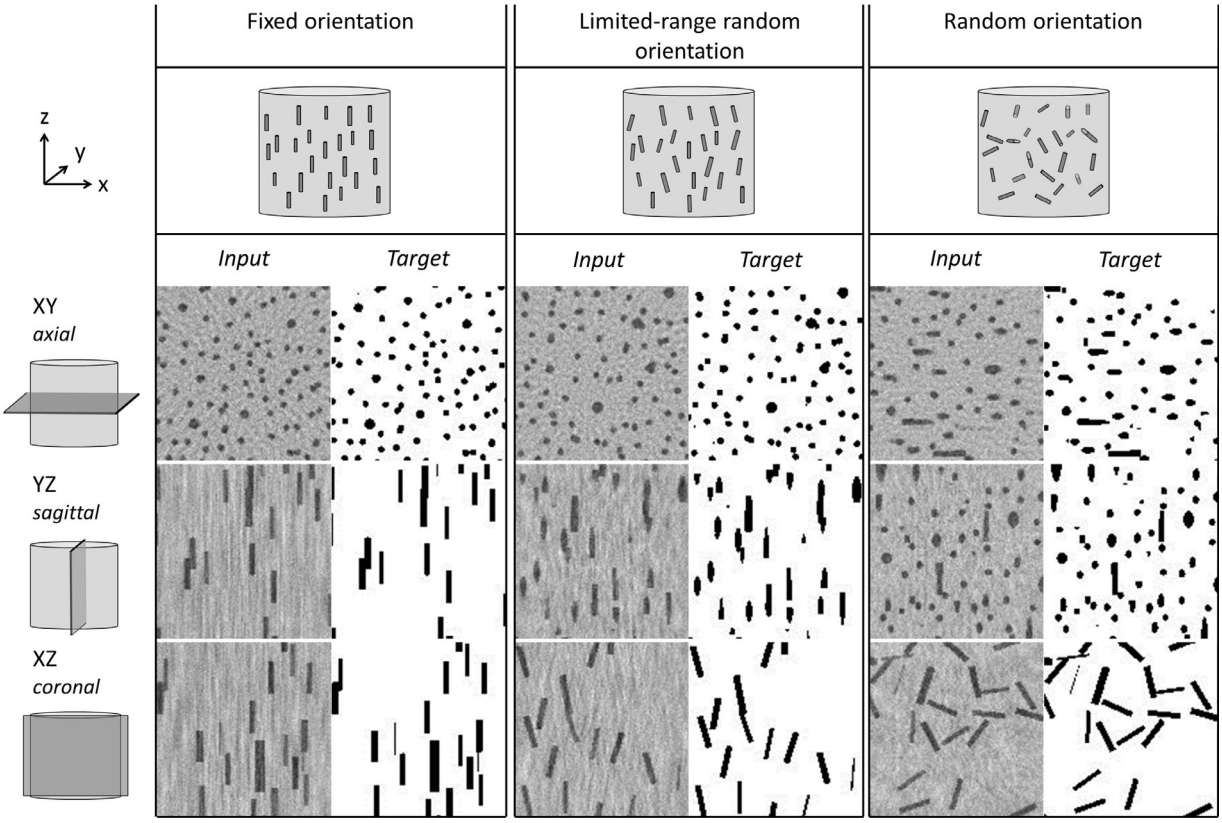

Fig. 1. Schematic representation of the three different simulation phantoms used in this study (top) and magnified examples (4x) of the central slices of the resulting CBCT images (input and gold standard segmentation labels).

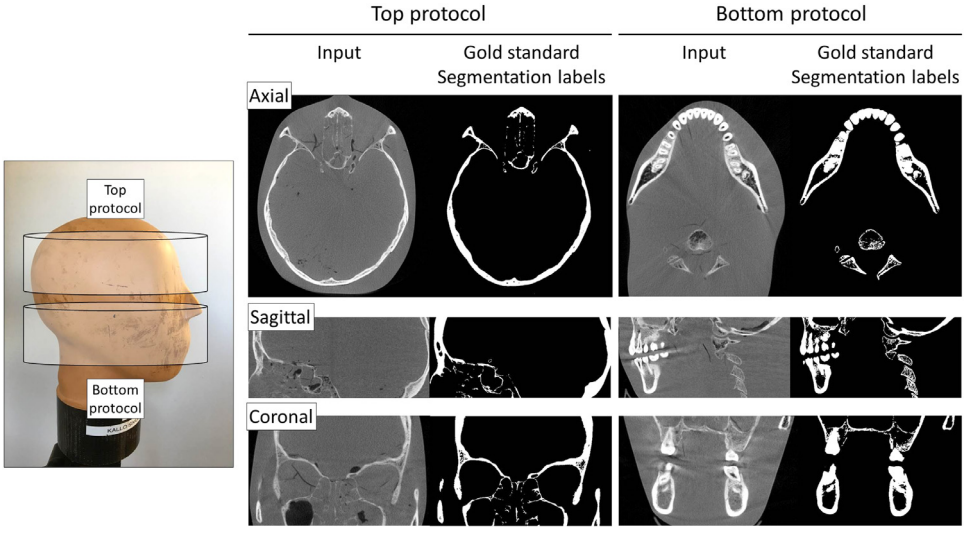

Fig. 2. Example of one anthropomorphic phantom head with the corresponding input CBCT slices and gold standard segmentation labels.

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).

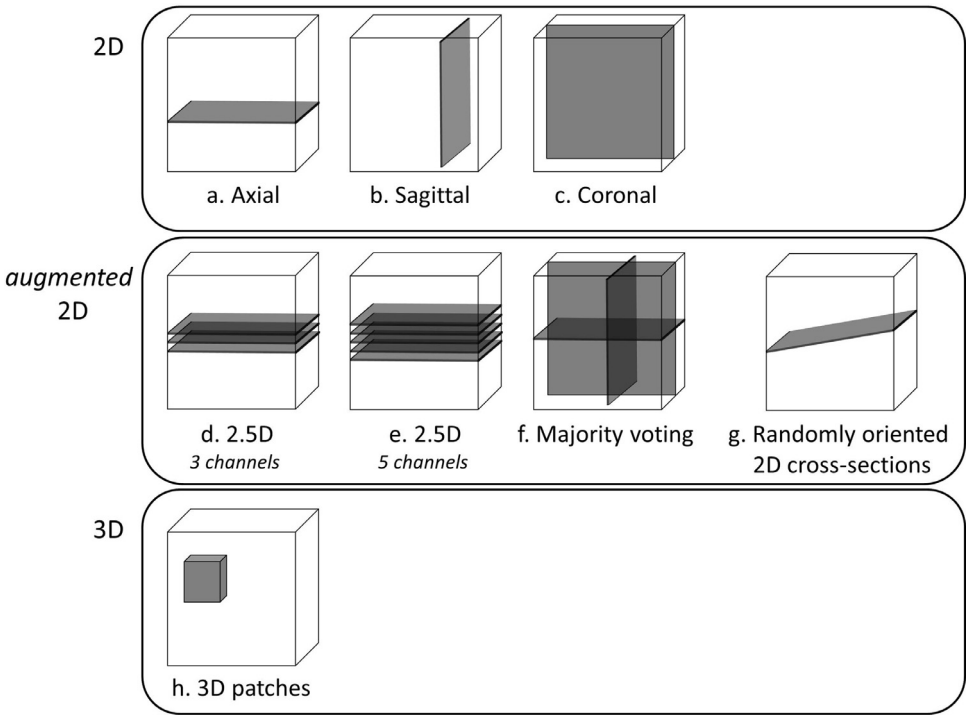

Fig. 3. Schematic overview of the eight CNN training strategies evaluated in the present study.

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) 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 dimen-

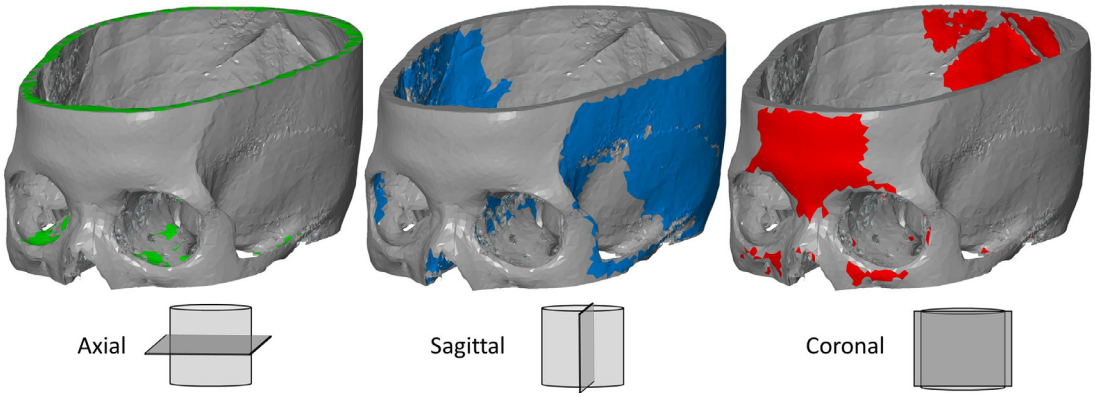

Fig. 4. Visual representation of the bony structures that were separately analyzed for each orthogonal plane.

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.

Training strategy	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.

sions 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, ε=1e-08) 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

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

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 calcu-

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

Training strategy	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.

lated 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 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.

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

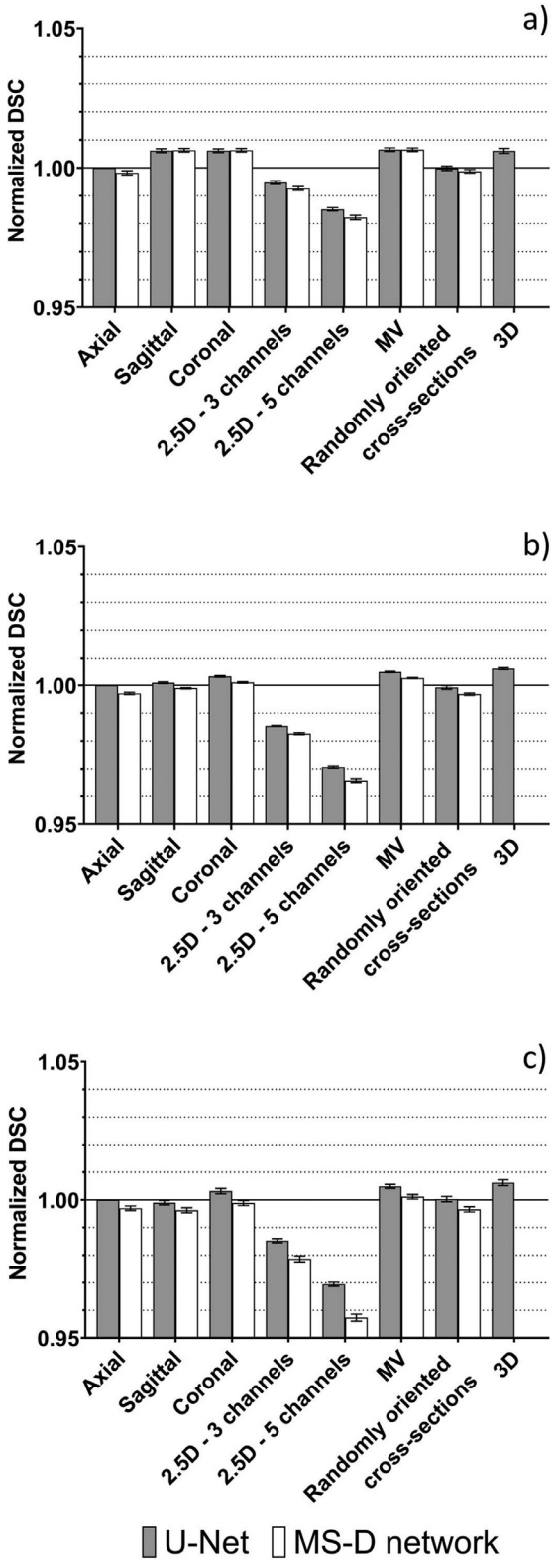

Fig. 5. Mean normalized Dice similarity coefficients (DSCs) of the segmented simulated CBCT images comprising cylinders with (a) fixed orientation; (b) limited-range random orientation; and (c) random orientation. DSCs were normalized by dividing them by the DSC obtained using U-Net trained on axial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

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.

Training strategy	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.

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
2.5D 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* 14,787,844	*3 x* 11,631	*3 x* 377.8 ± 0.4	*3 x* 260.2 ±1.3	*3 x* 25.6 ± 0.8	*3 x* 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.

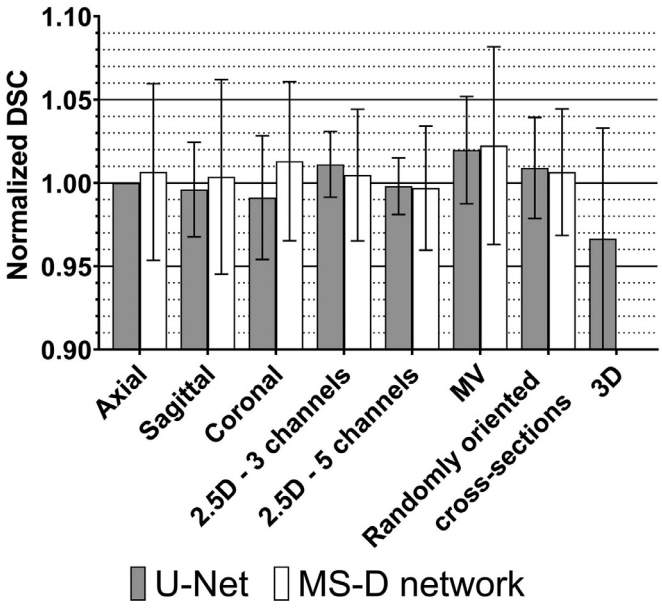

Fig. 6. Mean normalized Dice similarity coefficients (DSC) of the segmented experimental CBCT images. DSCs were normalized by dividing them by the DSC obtained using U-Net trained on axial slices.

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.

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,

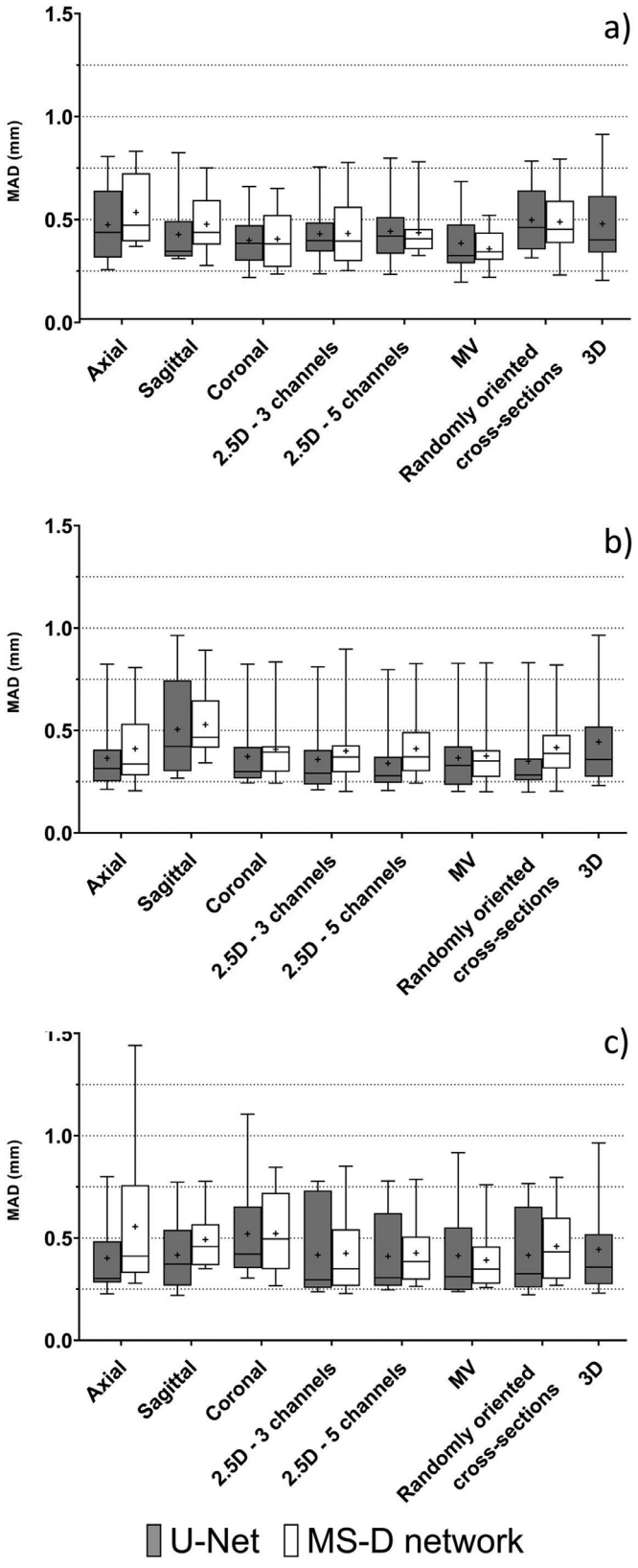

Fig. 7. Box and whisker plot of the mean absolute surface deviations (MADs) between the gold standard STL models and the STL models acquired using the eight different CNN training strategies. MADs were calculated separately for bony structures oriented in the axial (a), sagittal (b) and coronal (c) plane. 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.

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 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 ($512 \times 512 \times 512$) 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] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B.V. Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88 https://doi.org/10.1016/j.media.2017.07.005.

[2] D. Ciresan, A. Giusti, L.M. Gambardella, J. Schmidhuber, 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] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, H. Larochelle, Brain tumor segmentation with deep neural networks, Med. Image Anal. 35 (2016) 18–31 https://doi.org/10.1016/j.media.2016.05.004.

[4] R. Vivanti, A. Ephrat, L. Joskowicz, N. Lev-Cohain, O.A. Karaaslan, J. Sosna, Automatic 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] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R.M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE Trans. Med. Imaging 35 (2016) 1285–1298, doi:10.1109/TMI.2016.2528162.

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

[7] Ö. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-net: learning dense volumetric segmentation from sparse annotation, in: S. Ourselin, L. Joskowicz, M.R. Sabuncu, G. Unal, W. Wells (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI, Springer International Publishing, Cham, 2016 pp. 424–432https://doi.org/10.1007/978-3-319-46723-8_49.

[8] Q. Dou, H. Chen, L. Yu, L. Zhao, J. Qin, D. Wang, V.C.T. Mok, L. Shi, P.A. Heng, Automatic detection of cerebral microbleeds from MR images via 3D convolutional neural networks, IEEE Trans. Med. Imaging 35 (2016) 1182–1195, doi:10.1109/TMI.2016.2528129.

[9] F. Milletari, N. Navab, S.A. Ahmadi, 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, pp. 565–571, doi:10.1109/3DV.2016.79.

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

[11] K. Kamnitsas, C. Ledig, V.F.J. Newcombe, J.P. Simpson, A.D. Kane, D.K. Menon, D. Rueckert, B. Glocker, Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation, Med. Image Anal. 36 (2017) 61–78, doi:10.1016/j.media.2016.10.004.

[12] X. Zhou, K. Yamada, R. Takayama, X. Zhou, T. Hara, H. Fujita, S. Wang, T. Kojima, 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 p. 83, doi:10.1117/12.2295178.

[13] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, M. Nielsen, 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, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013 pp. 246–253https://doi.org/, doi:10.1007/978-3-642-40763-5_31.

[14] H.R. Roth, L. Lu, A. Seff, K.M. Cherry, J. Hoffman, S. Wang, J. Liu, E. Turkbey, R.M. Summers, A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations, Med. Image Comput. Comput. Assist. Interv. 17 (2014) 520–527.

[15] A. Ben-Cohen, I. Diamant, E. Klang, M. Amitai, H. Greenspan, Fully convolutional network for liver segmentation and lesions detection, in: G. Carneiro, D. Mateus, L. Peter, A. Bradley, J.M.R.S. Tavares, V. Belagiannis, J.P. Papa, J.C. Nascimento, M. Loog, Z. Lu, J.S. Cardoso, J. Cornebise (Eds.), Deep Learning and Data Labeling for Medical Applications, Springer International Publishing, Cham, 2016 pp. 77–85https://doi.org/, doi:10.1007/978-3-319-46976-8_9.

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

[17] X. Zhou, T. Ito, R. Takayama, S. Wang, T. Hara, H. Fujita, Three-dimensional CT image segmentation by combining 2D fully convolutional network with 3D majority voting, in: G. Carneiro, D. Mateus, L. Peter, A. Bradley, J.M.R.S. Tavares, V. Belagiannis, J.P. Papa, J.C. Nascimento, M. Loog, Z. Lu, J.S. Cardoso, J. Cornebise (Eds.), Deep Learning and Data Labeling for Medical Applications, Springer International Publishing, Cham, 2016 pp. 111–120https://doi.org/, doi:10.1007/978-3-319-46976-8_12.

[18] Z. Sobhaninia, S. Rezaei, A. Noroozi, M. Ahmadi, H. Zarrabi, N. Karimi, A. Emami, S. Samavi, Brain tumor segmentation using deep learning by type specific sorting of images, ArXiv:1809.07786 [Cs, Eess]. (2018). http://arxiv.org/abs/1809.07786.

[19] R. Cheng, N. Lay, F. Mertan, B. Turkbey, H.R. Roth, L. Lu, W. Gandler, E.S. McCreedy, T. Pohida, P. Choyke, M.J. McAuliffe, R.M. Summers, 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, Melbourne, Australia, 2017, pp. 749–753, doi:10.1109/ISBI.2017.7950627.

[20] S. Banerjee, S. Mitra, B.U. Shankar, Multi-planar spatial-ConvNet for segmentation and survival prediction in brain cancer, in: A. Crimi, S. Bakas, H. Kuijf, F. Keyvan, M. Reyes, T. van Walsum (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer International Publishing, Cham, 2019 pp. 94–104https://doi.org/, doi:10.1007/978-3-030-11726-9_9.

[21] A. Hänsch, M. Schwier, T. Morgas, J. Klein, H.K. Hahn, T. Gass, B. Haas, Comparison of different deep learning approaches for parotid gland segmentation from CT images, in: K. Mori, N. Petrick (Eds.), Medical Imaging 2018: Computer-Aided Diagnosis, SPIE, Houston, United States, 2018 p. 44https://doi.org/, doi:10.1117/12.2292962.

[22] J. Chen, L. Yang, Y. Zhang, M. Alber, D. Chen, Combining fully convolutional and recurrent neural networks for 3D biomedical image segmentation, (2016).

[23] F. Isensee, P.F. Jaeger, P.M. Full, I. Wolf, S. Engelhardt, K.H. Maier-Hein, Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features, in: M. Pop, M. Sermesant, P.M. Jodoin, A. Lalande, X. Zhuang, G. Yang, A. Young, O. Bernard (Eds.), Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges, Springer International Publishing, Cham, 2018 pp. 120–129https://doi.org/, doi:10.1007/978-3-319-75541-0_13.

[24] K. Vallaeys, A. Kacem, H. Legoux, M.L. Tenier, C. Hamitouche, R. Arbab-Chirani, 3D dento-maxillary osteolytic lesion and active contour segmentation pilot study in CBCT: semi-automatic *vs* manual methods, Dentomaxillofac. Radiol. 44 (2015) 20150079, doi:10.1259/dmfr.20150079.

[25] R.M.G. dos Santos, J.M. De Martino, L.A. Passeri, R.R.d.F. Attux, F.H. Neto, Automatic repositioning of jaw segments for three-dimensional virtual treatment planning of orthognathic surgery, J. Cranio Maxillofac. Surg. 45 (2017) 1399–1407, doi:10.1016/j.jcms.2017.06.017.

[26] G. Todorov, N. Nikolov, Y. Sofronov, N. Gabrovski, M. Laleva, T. Gavrilov, Computer aided design of customized implants based on CT-Scan data and virtual prototypes, in: V. Poulkov (Ed.), Future Access Enablers for Ubiquitous and Intelligent Infrastructures, Springer International Publishing, Cham, 2019 pp. 339–346https://doi.org/, doi:10.1007/978-3-030-23976-3_30.

[27] P.J. Verhelst, E. Shaheen, K.d.F. Vasconcelos, F.V.d. Cruyssen, S. Shujaat, W. Coudyzer, B. Salmon, G. Swennen, C. Politis, R. Jacobs, 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] R. Schulze, U. Heil, D. Groβ, D. Bruellmann, E. Dranischnikow, U. Schwanecke, E. Schoemer, Artefacts in CBCT: a review, Dentomaxillofac. Radiol. 40 (2011) 265–273, doi:10.1259/dmfr/30642039.

[29] J. Minnema, M. Eijnatten, A.A. Hendriksen, N.P.T.J. Liberton, D.M. Pelt, K.J. Batenburg, T. Forouzanfar, J. Wolff, Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network, Med. Phys. 46 (2019) 5027–5035, doi:10.1002/mp.13793.

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

[31] D.M. Pelt, foam_ct_phantom, n.d. https://github.com/conda-forge/foam_ct_phantom-feedstock.

[32] A.A. Hendriksen, D.M. Pelt, W.J. Palenstijn, S.B. Coban, K.J. Batenburg, On-the-fly machine learning for improving image resolution in tomography, Appl. Sci. 9 (2019) 2445, doi:10.3390/app9122445.
[33] W. van Aarle, W.J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. De Beenhouwer, K. Joost Batenburg, J. Sijbers, Fast and flexible X-ray tomography using the ASTRA toolbox, Opt. Express 24 (2016) 25129, doi:10.1364/OE.24.025129.
[34] W. van Aarle, W.J. Palenstijn, J.D. Beenhouwer, T. Altantzis, S. Bals, K.J. Batenburg, J. Sijbers, The Astra toolbox: a platform for advanced algorithm development in electron tomography, Ultramicroscopy 157 (2015) 35–47, doi:10.1016/j.ultramic.2015.05.002.
[35] L.A. Feldkamp, L.C. Davis, J.W. Kress, Practical cone-beam algorithm, J. Opt. Soc. Am. A 1 (1984) 612, doi:10.1364/JOSAA.1.000612.
[36] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J.V. Miller, S. Pieper, R. Kikinis, 3D slicer as an image computing platform for the quantitative imaging network, Magn. Reson. Imaging 30 (2012) 1323–1341, doi:10.1016/j.mri.2012.05.001.
[37] 3D Slicer, (2018). http://www.slicer.org.
[38] X. Feng, K. Qing, N.J. Tustison, C.H. Meyer, Q. Chen, Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images, Med. Phys. 46 (2019) 2169–2180, doi:10.1002/mp.13466.
[39] L. Chen, Y. Wu, A.M. DSouza, A.Z. Abidin, A. Wismüller, C. Xu, MRI tumor segmentation with densely connected 3D CNN, in: E.D. Angelini, B.A. Landman (Eds.), Medical Imaging 2018: Image Processing, SPIE, Houston, United States, 2018 p. 50https://doi.org/, doi:10.1117/12.2293394.
[40] N. Lessmann, B. van Ginneken, P.A. de Jong, I. Išgum, Iterative fully convolutional neural networks for automatic vertebra segmentation and identification, Med. Image Anal. 53 (2019) 142–155, doi:10.1016/j.media.2019.02.005.
[41] B. Qiu, J. Guo, J. Kraeima, R.J.H. Borra, M.J.H. Witjes, P.M.A.V. Ooijen, 3D segmentation of mandible from multisectional CT scans by convolutional neural networks, ArXiv:1809.06752 [Cs]. (2018). http://arxiv.org/abs/1809.06752.
[42] A. Klein, J. Warszawski, J. Hillengaß, K.H. Maier-Hein, Automatic bone segmentation in whole-body CT images, Int. J. CARS 14 (2019) 21–29, doi:10.1007/s11548-018-1883-7.
[43] D.M. Pelt, J.A. Sethian, A mixed-scale dense convolutional neural network for image analysis, Proc. Natl. Acad. Sci. 115 (2018) 254–259, doi:10.1073/pnas.1715832114.
[44] A. Hendriksen, ahendriksen/msd_pytorch: v0.7.2, Zenodo, 2019. https://doi.org/10.5281/ZENODO.3560114.
[45] A. Hendriksen, On the fly, Github, n.d. https://github.com/ahendriksen/on_the_fly.
[46] S. Ioffe, C. Szegedy, 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] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, ArXiv:1412.6980 [Cs]. (2014). http://arxiv.org/abs/1412.6980.
[48] P. Mlynarski, H. Delingette, H. Alghamdi, P.Y. Bondiau, N. Ayache, Anatomically consistent CNN-based segmentation of organs-at-risk in cranial radiotherapy, J. Med. Imaging 7 (2020) 1, doi:10.1117/1.JMI.7.1.014502.
[49] A.D. Desai, G.E. Gold, B.A. Hargreaves, A.S. Chaudhari, Technical considerations for semantic segmentation in MRI using convolutional neural networks, ArXiv:1902.01977 [Cs, Eess]. (2019). http://arxiv.org/abs/1902.01977.
[50] M.H. Vu, G. Grimbergen, T. Nyholm, T. Löfstedt, 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] Y. Zhang, Q. Liao, J. Zhang, 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] P. Mlynarski, H. Delingette, A. Criminisi, N. Ayache, 3D convolutional neural networks for tumor segmentation using long-range 2D context, Comput. Med. Imaging Gr. 73 (2019) 60–72, doi:10.1016/j.compmedimag.2019.02.001.
[53] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P.F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert, K.H. Maier-Hein, nnU-Net: self-adapting framework for U-Net-based medical image segmentation, ArXiv:1809.10486 [Cs]. (2018). http://arxiv.org/abs/1809.10486.