

Article

Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography

Bingjiang Qiu ^{1,2,3}, Jiapan Guo ^{2,3,*}, Joep Kraeima ^{1,4}, Haye Hendrik Glas ^{1,4}, Weichuan Zhang ^{5,6}, Ronald J. H. Borra ⁷, Max Johannes Hendrikus Witjes ^{1,4} and Peter M. A. van Ooijen ^{2,3}

¹ 3D Lab, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands; b.qiu@umcg.nl (B.Q.); j.kraeima@umcg.nl (J.K.); h.h.glas@umcg.nl (H.H.G.); m.j.h.witjes@umcg.nl (M.J.H.W.)

² Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands; p.m.a.van.ooijen@umcg.nl

³ Data Science Center in Health (DASH), University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands

⁴ Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands

⁵ Institute for Integrated and Intelligent System, Griffith University, Nathan, QLD 4111, Australia; zwc2003@163.com

⁶ CSIRO Data61, Epping, NSW 1710, Australia

⁷ Medical Imaging Center (MIC), University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713GZ Groningen, The Netherlands; r.j.h.borra@umcg.nl

* Correspondence: j.guo@umcg.nl



Citation: Qiu, B.; Guo, J.; Kraeima, J.; Glas, H.H.; Zhang, W.; Borra, R.J.H.; Witjes, M.J.H.; van Ooijen, P.M.A. Recurrent Convolutional Neural Networks for 3D Mandible Segmentation in Computed Tomography. *J. Pers. Med.* **2021**, *11*, 492. <https://doi.org/10.3390/jpm11060492>

Academic Editor: Peter Polverini

Received: 3 May 2021

Accepted: 28 May 2021

Published: 31 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Purpose: Classic encoder–decoder-based convolutional neural network (EDCNN) approaches cannot accurately segment detailed anatomical structures of the mandible in computed tomography (CT), for instance, condyles and coronoids of the mandible, which are often affected by noise and metal artifacts. The main reason is that EDCNN approaches ignore the anatomical connectivity of the organs. In this paper, we propose a novel CNN-based 3D mandible segmentation approach that has the ability to accurately segment detailed anatomical structures. Methods: Different from the classic EDCNNs that need to slice or crop the whole CT scan into 2D slices or 3D patches during the segmentation process, our proposed approach can perform mandible segmentation on complete 3D CT scans. The proposed method, namely, RCNNSeg, adopts the structure of the recurrent neural networks to form a directed acyclic graph in order to enable recurrent connections between adjacent nodes to retain their connectivity. Each node then functions as a classic EDCNN to segment a single slice in the CT scan. Our proposed approach can perform 3D mandible segmentation on sequential data of any varied lengths and does not require a large computation cost. The proposed RCNNSeg was evaluated on 109 head and neck CT scans from a local dataset and 40 scans from the PDDCA public dataset. The final accuracy of the proposed RCNNSeg was evaluated by calculating the Dice similarity coefficient (DSC), average symmetric surface distance (ASD), and 95% Hausdorff distance (95HD) between the reference standard and the automated segmentation. Results: The proposed RCNNSeg outperforms the EDCNN-based approaches on both datasets and yields superior quantitative and qualitative performances when compared to the state-of-the-art approaches on the PDDCA dataset. The proposed RCNNSeg generated the most accurate segmentations with an average DSC of 97.48%, ASD of 0.2170 mm, and 95HD of 2.6562 mm on 109 CT scans, and an average DSC of 95.10%, ASD of 0.1367 mm, and 95HD of 1.3560 mm on the PDDCA dataset. Conclusions: The proposed RCNNSeg method generated more accurate automated segmentations than those of the other classic EDCNN segmentation techniques in terms of quantitative and qualitative evaluation. The proposed RCNNSeg has potential for automatic mandible segmentation by learning spatially structured information.

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

1. Introduction

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

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

Alternatively, deep learning approaches have demonstrated a strong ability in automatic descriptive and detailed image feature extraction [21,22]. Ibragimov et al. [23] applied a tri-planar patch-based CNN for mandible segmentation in head and neck CT scans and achieved superior performance for most of the OAR segmentation when compared to the existing approaches. Zhu et al. [24] applied a 3D Unet that uses a loss function combining Dice score and focal loss for the training of the network. This 3D Unet-based method achieved better performances than that of the state-of-the-art approaches in OAR segmentation. Tong et al. [25] proposed a fully CNN (FCNN) with a shape representation model for mandible segmentation in CT scans and achieved better results than those of the conventional approaches. Egger et al. [26] implemented a three-step training strategy of the fully convolutional networks proposed by [27] to segment the mandibles in the CT scans on a locally acquired dataset. Qiu et al. [11] presented a 2D Unet-based mandible segmentation approach that segments mandibles on orthogonal planes of the CT scans. Liang et al. [28] proposed a multi-view spatial aggregation frame for joint localization and the segmentation of the OARs and achieved competitive segmentation performance.

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

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

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

Our major contributions are three-fold. First, we proposed a novel mandible segmentation architecture that uses recurrent neural networks to facilitate connected units between adjacent slices in an anatomical sequence of scanning. Each unit is then implemented as a classic encoder–decoder segmentation architecture for 2D slice segmentation. Second, the state-of-the-art 3D segmentation algorithms [37,38] usually down-sample the raw data to overcome memory issues. In contrast, the implementation of the RCNNSeg approach enables 3D segmentation in a way that reduces computational complexity of the model without loss of image quality. The 3D mandible segmentation can be performed for any varied length of the scan. Third, the proposed approach is able to tackle the problems of truncated inputs, which represent the main reason for the underperformance on detailed structures and regions that are corrupted by metal artifacts in CT scans. Extensive experiments were performed on two head and neck CT scans, and the proposed RCNNSeg approach outperformed the existing segmentation approaches for automatic mandible segmentation.

The remainder of the paper is organized as follows: Section 2 introduces our proposed RCNNSeg approach for automatic mandible segmentation. In Section 3, extensive experiments on two head and neck CT datasets are presented. The proposed method is compared both qualitatively and quantitatively with the EDCNN-based approaches as well as state-of-the-art methods for mandible segmentation. In Section 4, different aspects of the proposed method are discussed. Conclusions are given in Section 5.

2. Materials and Methods

Due to extensive demands on computational resources, the encoder–decoder-based mandible segmentation methods, such as Unet [29], SegNet [39], and AttUnet [40], cannot

directly deal with the 3D medical images. Instead, three strategies are widely used for the feeding of 3D volumetric data, which are shown in Figure 1. The first two strategies in Figure 1a,b use either 2D slices or a few adjacent slices as inputs to feed into the EDCNNs, which are truncated along the z-axis from the volumetric data. The third one in Figure 1c uses 3D patches as inputs, which are cropped along all three orthogonal axes. These strategies are easy to implement and can deal with the large memory demands from deep learning algorithms; however, they cannot take into account the complete anatomical structures of the organs in CT scans that often frustrate the accurate segmentation of detailed structures.

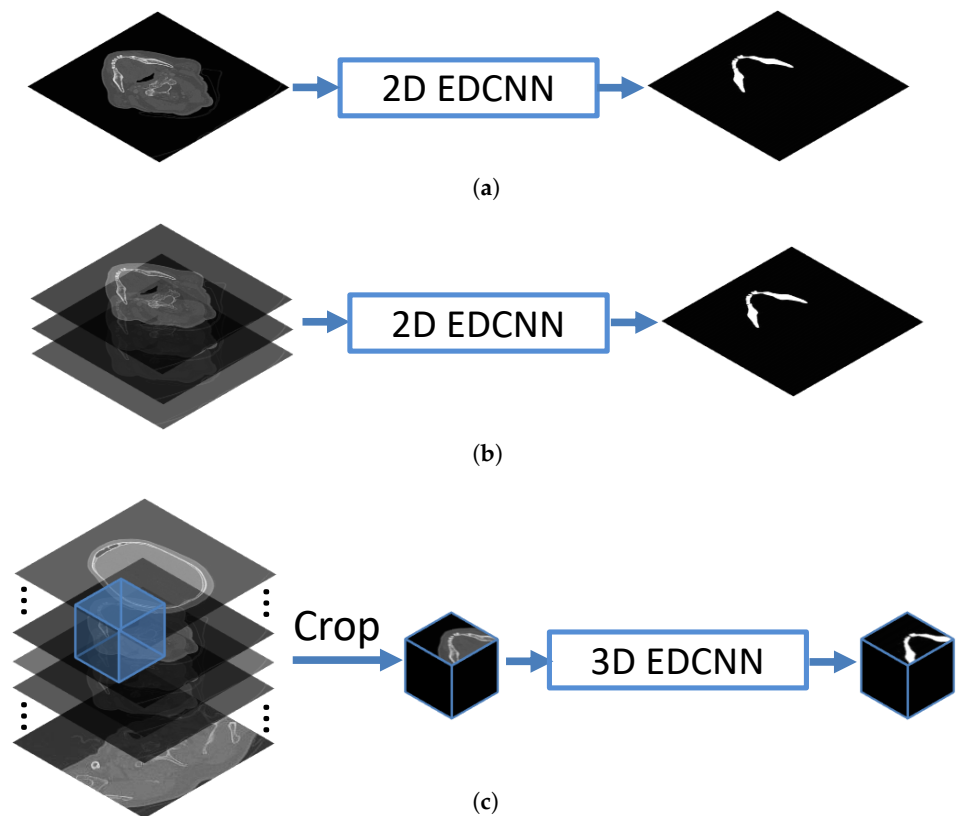


Figure 1. Illustration of three prevalent strategies for the feeding of input data to the classic encoder-decoder-based convolutional neural networks. (a) Use of 2D EDCNN network for 2D slice-based object segmentation. (b) Use of 2D EDCNN for the image segmentation based on several adjacent slices from the volumetric data. (c) Use of 3D EDCNN based on 3D patches cropped from the complete volumetric data.

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

2.1. Feedforward, Feedback, and Recurrent Networks

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

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

2.2. Recurrent Convolutional Neural Networks for Mandible Segmentation (RCNNSeg)

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

Our proposed architecture for 3D image segmentation can be observed in Figure 2a. The RCNNSeg takes a sequence of CT slices as inputs and then outputs a sequence of the corresponding segmented images of the same length. The RCNNSeg architecture can be unfolded to the structure shown in Figure 2b. RCNNSeg consists of varied lengths of successive units to be able to process sequential data of different lengths. Each of the units has a structure of the classic 2D EDCNN to evoke slice-based mandible segmentation.

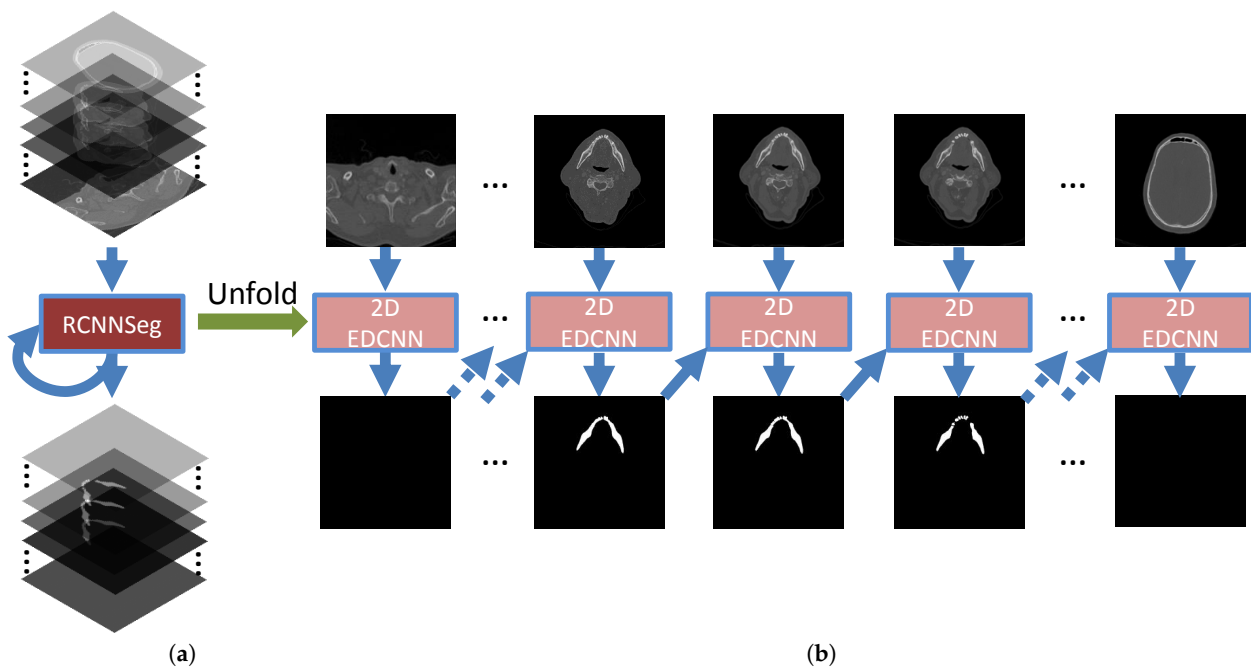


Figure 2. The overall graphic scheme of the proposed methods. The architecture of RCNNSeg with two components: (a) the RCNNSeg and its loss drawn with recurrent connections; (b) the same seen as a time-unfolded computational graph, where each node is now associated with one particular time instance.

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

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

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

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

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

$$\mathcal{L}^t = \alpha \times \mathcal{L}_{\text{BCE}}^t + \beta \times \mathcal{L}_{\text{Dice}}^t \quad (2)$$

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

Training of the model For the training of the RCNNSeg, we use the backpropagation through time (BPTT) [47] technique that begins by unfolding a recurrent neural network into successive units. As shown in Figure 2, the unfolded network contains N units, each of which takes a slice as the input in the ordered sequence and outputs the corresponding segmentation result. According to Equation (2), the loss at the t th step is denoted as \mathcal{L}^t . The total loss for a given ordered sequence of slices is the sum of the loss over all of the steps. We then optimize all of the parameters of the model based on the chain rule [48] to minimize the total loss.

2.3. Experimental Setup

2.3.1. Implementation Details

Our proposed approach was implemented in Pytorch [49]. All of our experiments were performed on a workstation equipped with Nvidia K40 GPUs with 12 GB of memory. We set the weights of the loss function to 0.5 for both the BCE loss term α and the Dice loss term β . We used Adam optimization [50] with a learning rate of $r = 10^{-4}$. The total number of epochs was 40 and 80 for the UMCG dataset and the PDDCA dataset, respectively. Moreover, an early stopping strategy was utilized if there was no improvement in the loss of the validation set for 10 epochs in order to avoid over-fitting.

2.3.2. Evaluation Metrics

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

The Dice similarity coefficient (DSC) is often used to measure the consistency between two objects [51]. Therefore, it is widely applied as a metric to evaluate the performance of image segmentation algorithms. We do not elaborate further about the DSC and refer the readers to [51].

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

$$D_{\text{ASD}}(A, B) = \frac{d(A, B) + d(B, A)}{2}, \quad (3)$$

$$d(A, B) = \frac{1}{N} \sum_{a \in A} \min_{b \in B} \|a - b\|, \quad (4)$$

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

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

$$D_{\text{HD}}(A, B) = \max(h(A, B), h(B, A)), \quad (5)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (6)$$

where $h(A, B)$ is often called the directed HD. The maximum HD is sensitive to contours. When the image is contaminated by noise or occlusion, the original Hausdorff distance is prone to mismatch [52,53]. Huttenlocher et al. [54] proposed the concept of partial Hausdorff distance. The 95HD is similar to maximum HD and selects 95% of the closest points in set B to the point in set A in Equation (6) to calculate $h(A, B)$

$$D_{95HD} = \max(h^{95\%}(A, B), h^{95\%}(B, A)), \quad (7)$$

$$h^{95\%}(A, B) = \max_{a \in A} \min_{b \in B^{95\%}} \|a - b\|. \quad (8)$$

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

3. Results

In this section, we demonstrate the effectiveness of the proposed approach for mandible segmentation in CT scans. We verify this by both quantitative and qualitative experimental results on two datasets, namely, the local UMCG dataset and the public PDDCA dataset [52], and facilitate comparisons with the state-of-the-art methods.

3.1. The UMCG Head and Neck Dataset

The UMCG head and neck dataset was collected in the department of oral and maxillo-facial surgery at the University Medical Center Groningen. This dataset contains 109 head and neck CT scans reconstructed with a kernel of Siemens Br64, I70 h(s) or B70s. Each scan consists of 221 to 955 slices with a size of 512×512 pixels. The pixel spacing varies from 0.35 to 0.66 mm, and the slice thickness varies from 0.6 to 0.75 mm. The corresponding manual mandible segmentations were obtained by an experienced researcher using Mimics software version 20.0 (Materialise, Leuven, Belgium) and then confirmed by a clinician.

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

Quantitative Results We evaluated the proposed RCNNSeg for mandible segmentation based on the metrics mentioned in Section 2.3.2. In Table 1, we list the results on DSC, D_{ASD} , and D_{95HD} as well as the corresponding standard deviations. In general, our proposed RCNNSeg-based approaches outperformed the EDCNN-based methods on mandible segmentation. For the DSC, all RCNNSeg-based approaches performed better than the corresponding EDCNNs with different strategies, with an improvement of at least 0.79%. Our proposed RCNNSeg also achieved the minimum error on D_{ASD} . On D_{95HD} , except for RUnetSeg, which underperformed the 2.5D U-Net, all other RCNNSeg-based approaches outperformed the corresponding EDCNNs. It is also surprising that the 3D patch EDCNNs dramatically underperformed when compared with all the other approaches.

Table 1. Quantitative comparison of segmentation performance in the UMCG dataset between the proposed RCNNSeg and the classic EDCNNs. The values in the square brackets indicate the standard deviation of the corresponding measurements. We mark in bold the best performance in each metric.

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

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

Figure 4 illustrates examples of three slices in the aforementioned encircled regions in a 2D slice view. The three images in Figure 4a are the corresponding ground truth (GT) segmentations. The images shown in Figure 4b,f,j (the middle column) are results from the 2D EDCNNs. The third and fourth columns in Figure 4 show the results from 2.5D EDCNNs and 3D patch EDCNNs. The last column gives the results from the proposed RCNNSeg approaches. Pink indicates the regions that are missed by the corresponding segmentation approaches. In general, the 2D and 2.5D EDCNN approaches missed more regions than the RCNNSeg approaches in segmenting the detailed structures. The 3D patch-based EDCNNs seem to segment the mandible quite well, but they also segment other bone structures, such as the skull, as mandibles.

3.2. PDDCA Dataset

We also evaluated the proposed pipeline on the public dataset PDDCA [52]. This dataset contains 48 patient CT scans from the Radiation Therapy Oncology Group (RTOG) 0522 study with manual segmentation of the left and right parotid glands, brainstem, optic chiasm, and mandible. Each scan consists of 76 to 360 slices with a size of 512×512 pixels. The pixel spacing varies from 0.76 to 1.27 mm, and the slice thickness varies from 1.25 to 3.0 mm. We followed the same training and testing split as described in [52]. Forty out of the 48 patients in PDDCA with manual mandible annotations were used in previous studies [52,55], in which the dataset was split into training and test subsets with 25 (0522c0001–0522c0328) and 15 (0522c0555–0522c0878) cases, respectively [52]. We used the pre-trained models obtained from the UMCG dataset and fine-tuned them on the training subset of the PDDCA dataset. We evaluated the performances of the models on the test subset.

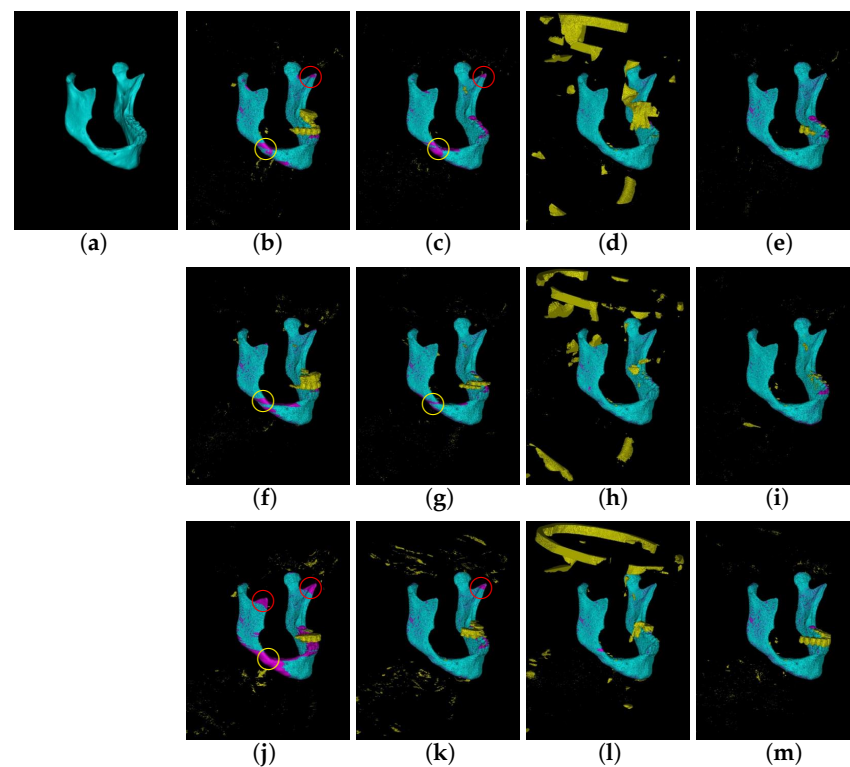


Figure 3. 3D view of a case from the UMCG head and neck dataset. (a–m) Ground truth, 2D U-Net, 2.5D U-Net, 3D U-Net, RUnetSeg, 2D SegU-Net, 2.5D SegU-Net, 3D U-Net, RSegUnetSeg, 2D Att U-Net, 2.5D Att U-Net, 3D Att U-Net, and RAttUnetSeg. We use cyan to indicate the correctly segmented mandible compared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

Quantitative results Table 2 shows the quantitative evaluation based on the average DSC, ASD, and 95HD used in the challenge [25,52]. The achieved performance indicates that our proposed RCNNSeg-based methods outperformed the corresponding EDCNN-based approaches with different input feeding strategies regarding all evaluation metrics.

In Table 3, we also compare our proposed approach with the state-of-the-art methods on the PDDCA dataset. We mark in bold the best three performances for each evaluation metric. In general, our RCNNSeg-based approaches outperformed the other existing approaches. Only the RAttUnetSeg model underperformed when compared with some of the existing approaches [56–58] on DSC and 95HD.

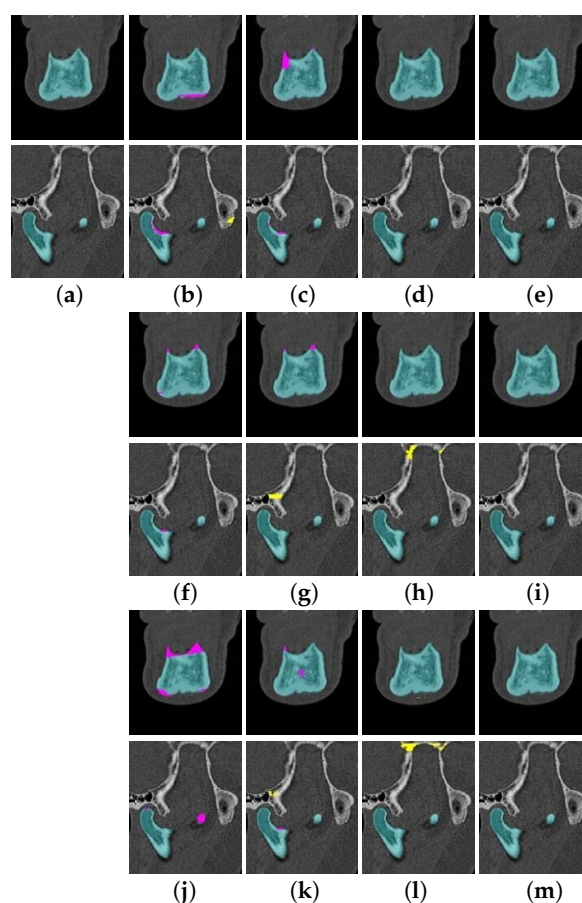


Figure 4. Examples of the automatic segmentation of mandibles in the UMCG dataset. (a) Ground truth segmentation on the three examples. (b–m) Segmentation results obtained on the example slices from 2D U-Net, 2.5D U-Net, 3D U-Net, RUnetSeg, 2D SegU-Net, 2.5D SegU-Net, 3D SegU-Net, RSegUnetSeg, 2D Att U-Net, 2.5D Att U-Net, 3D Att U-Net, and RAttUnetSeg. Cyan indicates the correctly segmented mandible compared to the ground truth. Pink represents the regions that were missed by the algorithms, while yellow indicates the segmented non-mandible regions.

Table 2. Quantitative comparison of the segmentation performance between the proposed RCNNSeg-based approaches and the EDCNN-based methods on the PDDCA dataset. The values in the square brackets indicate the standard deviation of the corresponding measurements. We mark in bold the best performance in each metric.

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

Table 3. Comparison of segmentation performance between the state-of-the-art methods and our proposed RCNNSeg approach; bold font indicates the best three performers for each measurement.

	DSC (%)	D _{ASD} (mm)	D _{95HD} (mm)
Multi-atlas [13]	91.7 [±2.34]	-	2.4887 [±0.7610]
AAM [14]	92.67 [±1]	-	1.9767 [±0.5945]
ASM [15]	88.13 [±5.55]	-	2.832 [±1.1772]
CNN [23]	89.5 [±3.6]	-	-
NLGM [59]	93.08 [±2.36]	-	-
AnatomyNet [24]	92.51 [±2]	-	6.28 [±2.21]
FCNN [25]	92.07 [±1.15]	0.51 [±0.12]	2.01 [±0.83]
FCNN+SRM [25]	93.6 [±1.21]	0.371 [±0.11]	1.5 [±0.32]
CNN+BD [57]	94.6 [±0.7]	0.29 [±0.03]	-
HVR [56]	94.4 [± 1.3]	0.43 [± 0.12]	-
Cascade 3D U-Net [58]	93 [±1.9]	-	1.26 [±0.5]
Multi-view [28]	94.1 [±0.7]	0.28 [±0.14]	-
RUNetSeg	94.71 [±1.35]	0.1353 [±0.0614]	1.4098 [±0.8573]
RSegUnetSeg	95.10 [±1.21]	0.1367 [±0.0382]	1.3560 [±0.4487]
RAttUnetSeg	93.87 [±1.29]	0.1773 [±0.0515]	1.6397 [±0.6219]

4. Discussion

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

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

Quantitative and qualitative evaluation on two datasets demonstrates that our proposed approach is robust in mandible segmentation, as illustrated in Figures 3 and 4 and Tables 1–3. Comparisons with the classic EDCNN approaches and state-of-the-art approaches illustrate that the RCNNseg approach significantly improves segmentation of the mandible. It is worth noting that RCNNSeg is very robust for the segmentation of the weak and blurry boundaries (for instance, the coronoids of mandibles).

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

This strategy takes advantage of shape prior information, which considers the segmentation result of the previous slice as the shape prior to the segmentation of the current slice. Therefore, the proposed approach can help to identify the continuous structure of

the mandible in 2D segmentation networks. In addition, the proposed approach utilizes the RNN module that helps the extraction of spatial information of the object based on the collection of context and shape information. This strategy can support further research on 3D image segmentation. It can also help alleviate memory issues for 3D medical image segmentation as well as segmentation tasks on sequential data, such as video images.

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

5. Conclusions

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

Author Contributions: Conceptualization, all authors; methodology, B.Q. and J.G.; validation, B.Q. and J.G.; formal analysis, B.Q., J.G., and W.Z.; investigation, B.Q., J.K., P.M.A.v.O., and J.G.; data curation, J.K., H.H.G., and B.Q.; writing—original draft preparation, B.Q. and J.G.; writing—review and editing, all authors; visualization, B.Q. and J.G.; supervision, all authors; project administration, J.K., M.J.H.W., and P.M.A.v.O. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (or Ethics Committee) of METC of UMCG.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The Public Domain Database for Computational Anatomy Dataset (PDDCA) is available at <https://www.imagenglab.com/newsite/pddca/> (accessed on 24 January 2019). Unfortunately, for reasons of ethics and patient confidentiality, we are not able to provide the sequencing data in a public database. The data underlying the results presented in the study are available from the corresponding author.

Acknowledgments: The author is supported by a joint PhD fellowship from the China Scholarship Council. The authors acknowledge Erhan Saatcioglu for the training data preparation. The authors would like to thank Gregory C Sharp (Harvard Medical School (MGH), Boston) and his group for providing and maintaining the Public Domain Database for Computational Anatomy Dataset (PDDCA). The authors would also like to thank the Center for Information Technology of the University of Groningen for their support and for providing access to the Peregrine high performance computing cluster.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Edge, S.B.; Byrd, D.R.; Carducci, M.A.; Compton, C.C.; Fritz, A.; Greene, F. *AJCC Cancer Staging Manual*; Springer: New York, NY, USA, 2010; Volume 649.
2. Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R.L.; Torre, L.A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Cancer J. Clin.* **2018**, *68*, 394–424. [[CrossRef](#)]
3. Shah, J.P.; Gil, Z. Current concepts in management of oral cancer—surgery. *Oral Oncol.* **2009**, *45*, 394–401. [[CrossRef](#)]
4. Ciocca, L.; Mazzoni, S.; Fantini, M.; Persiani, F.; Marchetti, C.; Scotti, R. CAD/CAM guided secondary mandibular reconstruction of a discontinuity defect after ablative cancer surgery. *J. Cranio-Maxillo-Facial Surg.* **2012**, *40*, 511–515. [[CrossRef](#)] [[PubMed](#)]
5. Kraeima, J.; Dorgelo, B.; Gulbitti, H.A.; Steenbakkens, R.J.H.M.; Schepman, K.P.; Roodenburg, J.L.N.; Spijkervet, F.K.L.; Schepers, R.H.; Witjes, M.J.H. Multi-modality 3D mandibular resection planning in head and neck cancer using CT and MRI data fusion: A clinical series. *Oral Oncol.* **2018**, *81*, 22–28. [[CrossRef](#)] [[PubMed](#)]
6. Succo, G.; Berrone, M.; Battiston, B.; Tos, P.; Goia, F.; Appendino, P.; Crosetti, E. Step-by-step surgical technique for mandibular reconstruction with fibular free flap: Application of digital technology in virtual surgical planning. *Eur. Arch. Oto-Rhino-Laryngol.* **2014**, *272*, 1491–1501. [[CrossRef](#)] [[PubMed](#)]
7. Wallner, J.; Mischak, I.; Egger, J. Computed tomography data collection of the complete human mandible and valid clinical ground truth models. *Sci. Data* **2019**, *6*, 190003. [[CrossRef](#)]
8. Jeanneret-Sozzi, W.; Moeckli, R.; Valley, J.F.; Zouhair, A.; Ozsahin, E.M.; Mirimanoff, R.O.; SASRO. The reasons for discrepancies in target volume delineation. *Strahlenther. Onkol.* **2006**, *182*, 450–457. [[CrossRef](#)]
9. Hoang Duc, A.K.; Eminowicz, G.; Mendes, R.; Wong, S.L.; McClelland, J.; Modat, M.; Cardoso, M.J.; Mendelson, A.F.; Veiga, C.; Kadir, T.; et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. *Med. Phys.* **2015**, *42*, 5027–5034. [[CrossRef](#)]
10. Huff, T.J.; Ludwig, P.E.; Zuniga, J.M. The potential for machine learning algorithms to improve and reduce the cost of 3-dimensional printing for surgical planning. *Expert Rev. Med. Devices* **2018**, *15*, 349–356. [[CrossRef](#)]
11. Qiu, B.; Guo, J.; Kraeima, J.; Glas, H.H.; Borra, R.J.; Witjes, M.J.; van Ooijen, P.M. Automatic segmentation of the mandible from computed tomography scans for 3D virtual surgical planning using the convolutional neural network. *Phys. Med. Biol.* **2019**, *64*, 175020. [[CrossRef](#)]
12. Gollmer, S.T.; Buzug, T.M. Fully automatic shape constrained mandible segmentation from cone-beam CT data. In Proceedings of the IEEE 9th International Symposium on Biomedical Imaging, Barcelona, Spain, 2–5 May 2012; pp. 1272–1275. [[CrossRef](#)]
13. Chen, A.; Dawant, B. A multi-atlas approach for the automatic segmentation of multiple structures in head and neck CT images. In Proceedings of the Head and Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 9 October 2015.
14. Mannion-Haworth, R.; Bowes, M.; Ashman, A.; Guillard, G.; Brett, A.; Vincent, G. Fully automatic segmentation of head and neck organs using active appearance models. In Proceedings of the Head and Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 9 October 2015.
15. Albrecht, T.; Gass, T.; Langguth, C.; Lüthi, M. Multi atlas segmentation with active shape model refinement for multi-organ segmentation in head and neck cancer radiotherapy planning. In Proceedings of the Head and Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 9 October 2015.
16. Torosdagli, N.; Liberton, D.K.; Verma, P.; Sincan, M.; Lee, J.; Pattanaik, S.; Bagci, U. Robust and fully automated segmentation of mandible from CT scans. In Proceedings of the IEEE 14th International Symposium on Biomedical Imaging, Melbourne, Australia, 18–21 April 2017; pp. 1209–1212. [[CrossRef](#)]
17. Chuang, Y.J.; Doherty, B.M.; Adluru, N.; Chung, M.K.; Vorperian, H.K. A Novel Registration-Based Semiautomatic Mandible Segmentation Pipeline Using Computed Tomography Images to Study Mandibular Development. *J. Comput. Assist. Tomogr.* **2017**, *42*, 306–316. [[CrossRef](#)] [[PubMed](#)]
18. Blaschke, T.; Burnett, C.; Pekkarinen, A. Image segmentation methods for object-based analysis and classification. In *Remote Sensing Image Analysis: Including the Spatial Domain*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 211–236. [[CrossRef](#)]
19. Bankman, I. *Handbook of Medical Image Processing and Analysis*; Elsevier: Amsterdam, The Netherlands, 2008. [[CrossRef](#)]
20. Yuheng, S.; Hao, Y. Image segmentation algorithms overview. *arXiv* **2017**, arXiv:1707.02051.
21. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
22. Simonyan, K.; Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. In Proceedings of the 3rd International Conference on Learning Representations, ICLR 2015, Conference Track Proceedings, San Diego, CA, USA, 7–9 May 2015.
23. 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. [[CrossRef](#)]
24. Zhu, W.; Huang, Y.; Tang, H.; Qian, Z.; Du, N.; Fan, W.; Xie, X. AnatomyNet: Deep 3D Squeeze-and-excitation U-Nets for fast and fully automated whole-volume anatomical segmentation. *arXiv* **2018**, arXiv:1808.05238.
25. Tong, N.; Gou, S.; Yang, S.; Ruan, D.; Sheng, K. Fully automatic multi-organ segmentation for head and neck cancer radiotherapy using shape representation model constrained fully convolutional neural networks. *Med. Phys.* **2018**, *45*, 4558–4567. [[CrossRef](#)]

26. Egger, J.; Pfarrkirchner, B.; Gsaxner, C.; Lindner, L.; Schmalstieg, D.; Wallner, J. Fully Convolutional Mandible Segmentation on a valid Ground Truth Dataset. In Proceedings of the Annual International Conference IEEE Engineer Medical Biology Society, Honolulu, HI, USA, 18–21 July 2018; pp. 656–660.
27. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [\[CrossRef\]](#)
28. Liang, S.; Thung, K.; Nie, D.; Zhang, Y.; Shen, D. Multi-view Spatial Aggregation Framework for Joint Localization and Segmentation of Organs at risk in Head and Neck CT Images. *IEEE Trans. Med. Imaging* **2020**, *39*, 2794–2805. [\[CrossRef\]](#)
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [\[CrossRef\]](#)
30. Ye, J.C.; Sung, W.K. Understanding Geometry of Encoder-Decoder CNNs. In Proceedings of the 36 th International Conference on Machine Learning, PMLR 97, Long Beach, CA, USA, 9–15 June 2019.
31. Ger, R.B.; Craft, D.F.; Mackin, D.S.; Zhou, S.; Layman, R.R.; Jones, A.K.; Elhalawani, H.; Fuller, C.D.; Howell, R.M.; Li, H.; et al. Practical guidelines for handling head and neck computed tomography artifacts for quantitative image analysis. *Comput. Med. Imaging Graph.* **2018**, *69*, 134–139. [\[CrossRef\]](#)
32. Minnema, J.; van Eijnatten, M.; Hendriksen, A.A.; Liberton, N.; Pelt, D.M.; Batenburg, K.J.; Forouzanfar, T.; Wolff, J. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network. *Med. Phys.* **2019**, *46*, 5027–5035. [\[CrossRef\]](#)
33. Yu, Q.; Xia, Y.; Xie, L.; Fishman, E.K.; Yuille, A.L. Thickened 2D Networks for Efficient 3D Medical Image Segmentation. *arXiv* **2019**, arXiv:1904.01150.
34. Chen, J.; Yang, L.; Zhang, Y.; Alber, M.; Chen, D.Z. Combining fully convolutional and recurrent neural networks for 3d biomedical image segmentation. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 3036–3044.
35. Bai, W.; Suzuki, H.; Qin, C.; Tarroni, G.; Oktay, O.; Matthews, P.M.; Rueckert, D. Recurrent neural networks for aortic image sequence segmentation with sparse annotations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 586–594.
36. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [\[CrossRef\]](#)
37. Çiçek, Ö.; Abdulkadir, A.; Lienkamp, S.S.; Brox, T.; Ronneberger, O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Athens, Greece, 17–21 October 2016; pp. 424–432. [\[CrossRef\]](#)
38. Milletari, F.; Navab, N.; Ahmadi, S.A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision, Stanford, CA, USA, 25–28 October 2016; pp. 565–57. [\[CrossRef\]](#)
39. Kamal, U.; Tonmoy, T.I.; Das, S.; Hasan, M.K. Automatic Traffic Sign Detection and Recognition Using SegU-Net and a Modified Tversky Loss Function With L1-Constraint. *IEEE Trans. Intell. Transp. Syst.* **2020**, *21*, 1467–1479. [\[CrossRef\]](#)
40. Oktay, O.; Schlemper, J.; Folgoc, L.L.; Lee, M.; Heinrich, M.; Misawa, K.; Mori, K.; McDonagh, S.; Hammerla, N.Y.; Kainz, B.; et al. Attention U-Net: Learning where to look for the pancreas. *arXiv* **2018**, arXiv:1804.03999.
41. Abbott, L.F.; Dayan, P. *Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems*; The MIT Press: Cambridge, MA, USA, 2001.
42. Liang, M.; Hu, X. Recurrent convolutional neural network for object recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3367–3375. [\[CrossRef\]](#)
43. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Proceedings Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015.
44. Lai, S.; Xu, L.; Liu, K.; Zhao, J. Recurrent convolutional neural networks for text classification. In Proceedings of the AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–30 January 2015; pp. 2267–2273. [\[CrossRef\]](#)
45. Sudre, C.H.; Li, W.; Vercauteren, T.; Ourselin, S.; Cardoso, M.J. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 240–248. [\[CrossRef\]](#)
46. Taghanaki, S.A.; Zheng, Y.; Zhou, S.K.; Georgescu, B.; Sharma, P.; Xu, D.; Comaniciu, D.; Hamarneh, G. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Comput. Med. Imaging Graph.* **2019**, *75*, 24–33. [\[CrossRef\]](#)
47. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [\[CrossRef\]](#)
48. Magin, R.L. *Fractional Calculus in Bioengineering*; Begell House: Danbury, CT, USA, 2006.
49. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–14 December 2019; pp. 8024–8035.
50. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
51. Ghafoorian, M.; Karssemeijer, N.; Heskes, T.; Uden, I.W.; Sanchez, C.I.; Litjens, G.; Leeuw, F.E.; Ginneken, B.; Marchiori, E.; Platel, B. Location sensitive deep convolutional neural networks for segmentation of white matter hyperintensities. *Sci. Rep.* **2017**, *7*, 1–12. [\[CrossRef\]](#)

52. Raudaschl, P.F.; Zaffino, P.; Sharp, G.C.; Spadea, M.F.; Chen, A.; Dawant, B.M.; Albrecht, T.; Gass, T.; Langguth, C.; Lüthi, M.; et al. Evaluation of segmentation methods on head and neck CT: Auto-segmentation challenge 2015. *Med. Phys.* **2017**, *44*, 2020–2036. [[CrossRef](#)]
53. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 1–28. [[CrossRef](#)]
54. Huttenlocher, D.P.; Rucklidge, W.J.; Klanderman, G.A. Comparing images using the Hausdorff distance under translation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, USA, 15–18 June 1992; pp. 654–656. [[CrossRef](#)]
55. Ren, X.; Xiang, L.; Nie, D.; Shao, Y.; Zhang, H.; Shen, D.; Wang, Q. Interleaved 3D-CNNs for joint segmentation of small-volume structures in head and neck CT images. *Med. Phys.* **2018**, *45*, 2063–2075. [[CrossRef](#)]
56. Wang, Z.; Wei, L.; Wang, L.; Gao, Y.; Chen, W.; Shen, D. Hierarchical vertex regression-based segmentation of head and neck CT images for radiotherapy planning. *IEEE Trans. Image Process.* **2017**, *27*, 923–937. [[CrossRef](#)]
57. Kodym, O.; Španěl, M.; Herout, A. Segmentation of Head and Neck Organs at Risk Using CNN with Batch Dice Loss. In Proceedings of the German Conference on Pattern Recognition, Stuttgart, Germany, 9–12 October 2018; pp. 105–114. [[CrossRef](#)]
58. Wang, Y.; Zhao, L.; Song, Z.; Wang, M. Organ at Risk Segmentation in Head and Neck CT Images by Using a Two-Stage Segmentation Framework Based on 3D U-Net. *arXiv* **2018**, arXiv:1809.00960.
59. Orbes-Arteaga, M.; Pea, D.; Dominguez, G. Head and neck auto segmentation challenge based on non-local generative models. In Proceedings of the Head and Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 9 October 2015.
60. Castro-Núñez, J.; Shelton, J.M.; Snyder, S.; Sickels, J. Virtual surgical planning for the management of severe atrophic mandible fractures. *Craniofacial Trauma Reconstr.* **2018**, *11*, 150–156. [[CrossRef](#)]
61. Hoving, A.M.; Kraeima, J.; Schepers, R.H.; Dijkstra, H.; Potze, J.H.; Dorgelo, B.; Witjes, M.J. Optimisation of three-dimensional lower jaw resection margin planning using a novel Black Bone magnetic resonance imaging protocol. *PLoS ONE* **2018**, *13*, e0196059.