Journal of
*Personalized
Medicine*

MDPI

Article

Mandible Segmentation of Dental CBCT Scans Affected by Metal Artifacts Using Coarse-to-Fine Learning Model

Bingjiang Qiu [1,2,3], Hylke van der Wel [1,4], Joep Kraeima [1,4], Haye Hendrik Glas [1,4], Jiapan Guo [2,3,*], Ronald J. H. Borra [5], 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, 9713 GZ Groningen, The Netherlands; b.qiu@umcg.nl (B.Q.); h.van.der.wel@umcg.nl (H.v.d.W.); 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, 9713 GZ 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, 9713 GZ Groningen, The Netherlands
[4] Department of Oral and Maxillofacial Surgery, University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands
[5] Medical Imaging Center (MIC), University Medical Center Groningen, University of Groningen, Hanzeplein 1, 9713 GZ Groningen, The Netherlands; r.j.h.borra@umcg.nl
* Correspondence: j.guo@umcg.nl

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

Keywords: mandible segmentation; cone-beam computed tomography (CBCT); computed tomography (CT); metal artifacts; 3D virtual surgical planning (3D VSP); convolutional neural networks

check for updates

Citation: Qiu, B.; van der Wel, H.; Kraeima, J.; Glas, H.H.; Guo, J.; Borra, R.J.H.; Witjes, M.J.H.; van Ooijen, P.M.A. Mandible Segmentation of Dental CBCT Scans Affected by Metal Artifacts Using Coarse-to-Fine Learning Model. *J. Pers. Med.* **2021**, *11*, 560. https://doi.org/10.3390/jpm11060560

Academic Editor: Jorge Luis Espinoza

Received: 21 May 2021
Accepted: 14 June 2021
Published: 16 June 2021

1. Introduction

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

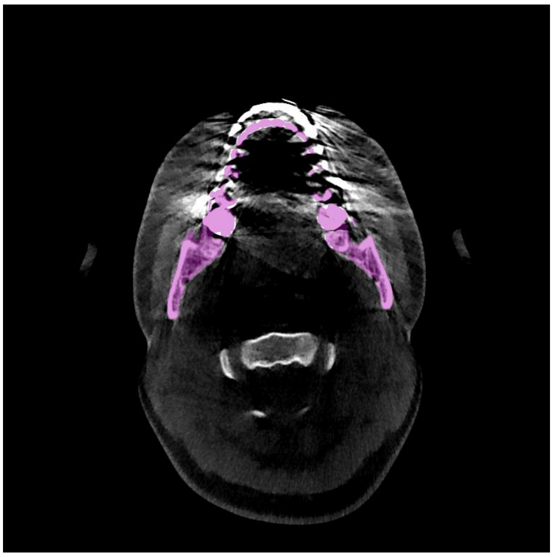

Figure 1. Example of manual annotation of the mandible in a CBCT image with strong metal artifacts.

To reduce the workload of mandible segmentation, a number of traditional segmentation methods have been developed in the past, including statistical shape model [2] as well as machine learning methods [3–6]. Sebastian et al. [2] presented a statistical shape model (SSM) based mandible segmentation approach. They introduced an optimized correspondence to their SSM model. Wang et al. [3] employed a majority voting method and combined it with random forest for mandible segmentation. Rarasmaya et al. [4] proposed a method based on histogram thresholding and polynomial fitting to segment mandibular cortical bone in CBCT scans. Oscar et al. [5] used super-voxels and graph clustering for mandible segmentation in CBCT images. Fan et al. [6] proposed an automatic approach for segmenting mandibles from CBCT using a marker-based watershed transform. However, some of these traditional techniques require mandible shape prior to initialization, and the performances of these methods are often affected by noise or metal artifacts. Furthermore, it is difficult to adjust the model parameters according to the overall characteristics of the target contour [7].

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

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

Motivated by the concept of curriculum learning [12], we propose a novel mandible segmentation approach based on a coarse-to-fine learning framework to solve the aforementioned challenges in mandible segmentation of CBCT. Curriculum learning draws upon a learning idea that follows a learning order from easy to difficult tasks [12]. Specifically, a complex task can be solved by dividing it into simple sub-tasks, then starting from the simplest sub-task and progressing to the more difficult sub-tasks. In this study, we propose a hybrid method which consists of a coarse stage and fine stage, in which the coarse stage makes use of 3D CNN for predicting the mandible-like organ and the fine stage utilizes the recurrent segmentation CNN for fine mandible segmentation in CBCT images which are mostly affected by metal artifacts. The proposed approach aims at overcoming the oversegmentation in some parts of the tooth regions and undersegmentation in the weak edges of the ramus and condyles. The coarse segmentation from the coarse stage guides the segmentation of the mandible in the fine stage, and therefore decreases the difficulty of segmenting the mandible from CBCT. Along with this coarse-to-fine segmentation (named as C2FSeg) task, we design a network by stacking a 3D SegUnet and a recurrent SegUnet. In addition, we extend the proposed segmentation network with a hybrid loss proposed by Taghanaki et al. [13], which has been demonstrated to offer superior performance in many visual applications.

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

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

2. Methodology

From the perspective of the framework, a coarse-to-fine mandible segmentation approach (C2FSeg) is proposed according to curriculum learning [12]. The C2FSeg consists of two main components: coarse stage and fine stage, in which the coarse stage obtains the mandible-like organs, while the fine stage reduces the false positive rate by embedding the overall information of the mandible-like organs and the neighboring information. In the coarse stage, potential mandible candidates are first identified, and then reduction of the false positives (FPs) within the candidates is performed in the fine stage. The coarse model identifies potential mandible candidates, and the fine model reduces the false positives (FPs) within the candidates. The overview of the mandible segmentation framework is given in Figure 2.

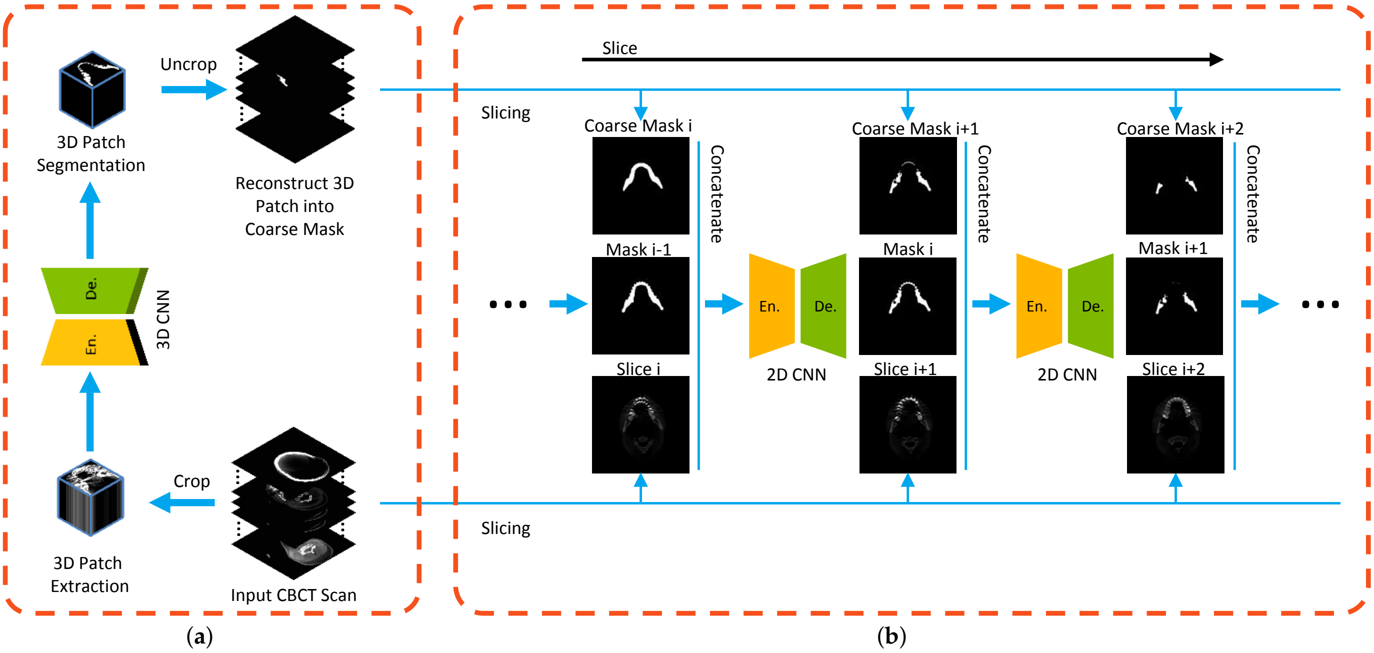

Figure 2. Overview of the proposed method consisting of (**a**) a 3D CNN for rough mandible segmentation and (**b**) a 2D recurrent segmentation network for further accurate mandible segmentation. The implementation of stage (**a**) is as follows: input scan → 3D patch extraction → 3D CNN (3D SegUnet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

2.1. Curriculum Learning in Mandible Segmentation

Curriculum learning describes a type of learning in which tasks can start with simple tasks before the number of difficult tasks is gradually increased. This learning method is proposed by [12]. It assumes that the curriculum learning can improve the convergence speed of the training process and find a better local minimum [12]. To elaborate upon the proposed C2FSeg approach, we first formulate a segmentation model that can be generalized to both coarse stage and fine stage; we will customize the segmentation model to these two stages in Sections 2.2 and 2.3, respectively.

Let $X = \{x^1, \dots, x^t, \dots, x^n\}$ be the head and neck scan volume, where X belongs to the CBCT image domain, $X \in \Omega = R^{n \times w \times h}$, where n, w and h represent slice number, width and height, respectively. The corresponding ground truth is $Y = \{y^1, \dots, y^t, \dots, y^n\} \in \{0,1\}^{n \times w \times h}$, where t denotes the t-th slice of the CT scan. Let $\hat{Y} = \{\hat{y}^1, \dots, \hat{y}^t, \dots, \hat{y}^n\} \in S \in [0,1]^{n \times w \times h}$ denote the predicted segmentation ($\Omega \to S$). We denote a segmentation task by an operator F, i.e., $\hat{Y} = F(X, \theta)$, where θ indicates model parameters. Specifically, in a CNN model with L layers and parameters $\theta = \{w^1, w^2, \dots, w^L; b^1, b^2, \dots, b^L\}$, $\{w^1, w^2, \dots, w^L\}$ is a set of weights and $\{b^1, b^2, \dots, b^L\}$ is a set of biases. According to the concept of curriculum learning, in which a task can be divided into several simple sub-tasks, the task is defined as $F = F_1, \dots, F_s$, where s represents the number of sub-tasks. Therefore, the predicted segmentation $\hat{Y} = F(X, \theta)$ can be rewritten as $\hat{Y} = F_s(F_{s-1}, \dots, F_2(F_1(X, \theta_1), X, \theta_2), \dots, X, \theta_{s-1}), X, \theta_s)$. Although this structure can improve the performance of the task and reduce the difficulty of the task, this model exponentially increases required computing resources, resulting in low efficiency. A small s is sufficient to handle the problem of mandible segmentation. We use $s = 2$ in this study, i.e.,

$$\hat{Y} = F_2(F_1(X, \theta_1), X, \theta_2), \tag{1}$$

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

2.2. Coarse Stage: Mandible-Like Organ Prediction

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

$$\hat{Y}_1^c = F_1(X^c, \theta_1). \tag{2}$$

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

As illustrated in Figure 3a, the coarse stage of our proposed C2FSeg network for capturing the mandible-like candidates is based on the 3D SegUnet, which is the original SegUnet [15] expanded from 2D to 3D. The 3D SegUnet consists of an encoder and a decoder, each of which has four convolutional blocks follwed by 3D maxpooling or 3D uppooling layers. The 3D convolutional block includes two convolution operations with a kernel of 3 × 3 × 3, each of which is followed by a batch normalization [16] and a rectified linear unit (ReLU) [17] activation function. The number of filters in the encoder starts at 32 and increases by a factor of 2 after every 3D convolutional block, while it declines by a factor of 2 in the decoder path. The number of feature maps is listed on the left of each convolutional block, and the convolutional layers are represented in Figure 3. In addition to the encoder and the decoder, we also use a cross connection to bridge the short-cut connection between the low-level and high-level layers, and we use a cross connection to transfer coordinates from maxpooling to uppooling. In the forward phase, the low-level feature maps extracted from the encoder are directly concatenated to the high-level feature maps, which can improve fine-scaled segmentation [18]. As for the backward phase, the high-level feature maps can be propagated backward through the connections. This approach can prevent the network from enduring gradient vanishing, which will hinder the convergence of the network in the training process [19]. The output of the 3D SegUnet is obtained by applying a convolutional layer with a kernel of 1 × 1 × 1 followed by the sigmoid function.

2.3. Fine Stage: False Positive Reduction

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

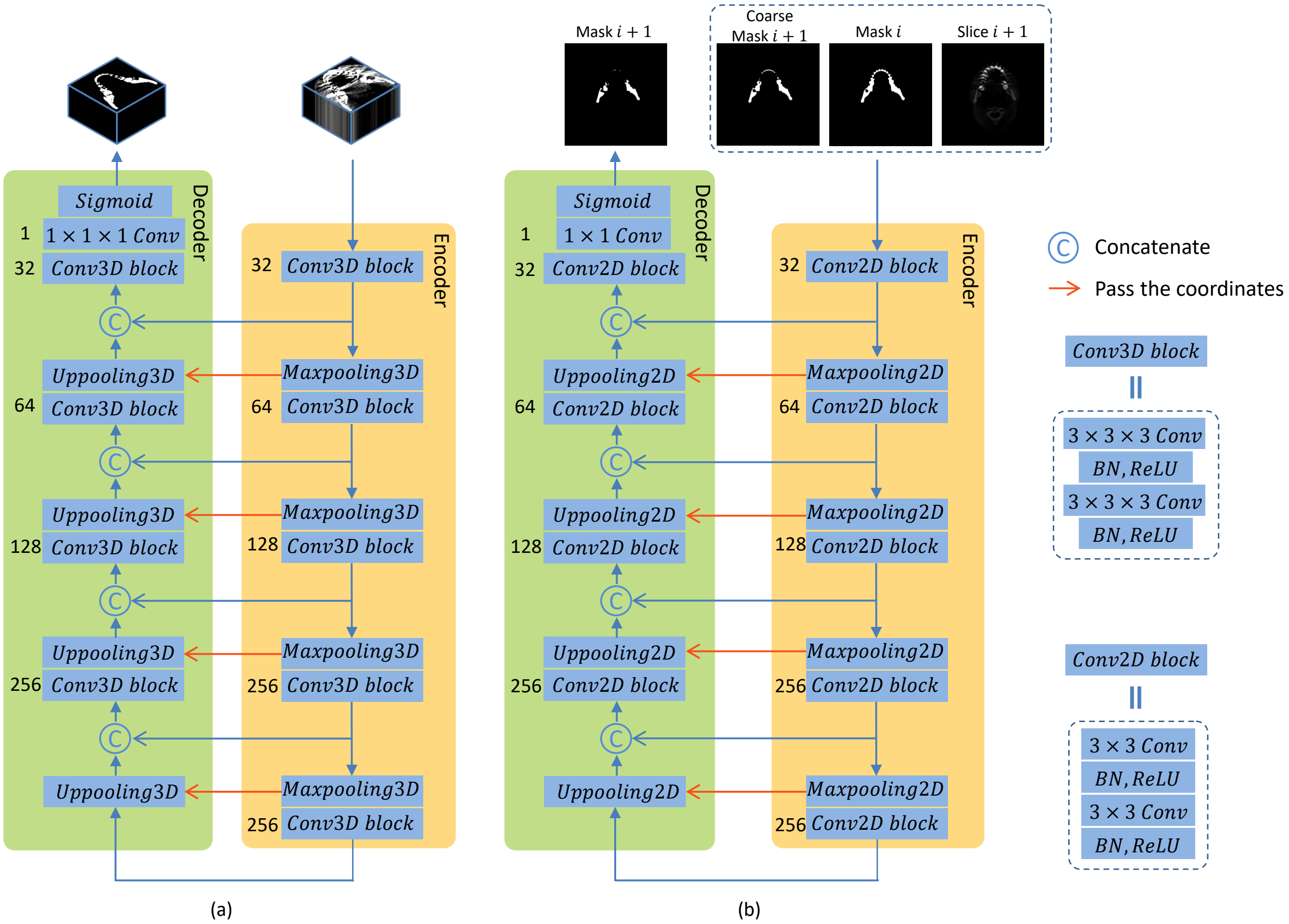

Figure 3. The detailed architecture setting of the proposed method consists of a 3D CNN (**a**) and a 2D recurrent segmentation network (**b**).

To further utilize the information obtained from the prediction of the previous neighborhood slice, we use RSegUnet [11] to accurately segment the mandible in the fine stage. The sequential design of RSegUnet allows the network to learn anatomical structure continuity in 3D form. In the fine stage, RSegUnet, F_2, processes each slice sequentially. The input of RSegUnet is sampled from the scan volume X and the rough predictions from $\hat{Y}_1$ of the coarse stage. Here, $\hat{Y}_2 = \{\hat{y}_2^1, \ldots, \hat{y}_2^t, \ldots, \hat{y}_2^n\} \in S_2 \in \{0,1\}^{n \times w \times h}$ is the binary segmentation map generated from the fine stage ($\{\Omega, S_1\} \rightarrow S_2$). In general, $\hat{Y}_2 = F_2(\hat{Y}_1, X, \theta_2)$. In this task, RSegUnet maps a sequence input slice $(x^t, \hat{y}_1^t, \hat{y}_2^{t-1})$ to a sequence output $\hat{y}_2^t$ of the same length, i.e., the output of the unfolded RSegUnet after t steps is represented as:

$$\hat{y}_2^t = F_2(\hat{y}_2^{(t-1)}, x^t, \hat{y}_1^t). \tag{3}$$

All in all, the fine stage of the proposed C2FSeg framework is illustrated in Figure 3b.

2.4. Loss

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

$$\mathcal{L} = \omega_1 \times \mathcal{L}_{BCE} + \omega_2 \times \mathcal{L}_{Dice}, \tag{4}$$

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

$$\mathcal{L}_{BCE}(\hat{y}, y) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) + (1 - y_i)\log(1 - \hat{y}_i), \tag{5}$$

$$\mathcal{L}_{Dice}(\hat{y}, y) = 1 - \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \hat{y}_i}. \tag{6}$$

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

According to Equation (5), the term $(1 - y_i)\log(1 - \hat{y}_i)$ penalizes false positives (FPs), as it is 0 when the prediction probability is correct, and $y_i \log(\hat{y}_i)$ penalizes false negatives (FNs) [13]. Therefore, the BCE term is able to control the trade-off between FPs and FNs in the pixelwise segmentation task. In spite of that, the networks with only BCE as loss function are often prone to generate more false positives in the segmentation [20]. The study from [21] has proven that Dice loss yields better performance for one-target segmentation and is able to predict the fine appearance features of the object. Dice loss is based on the Dice coefficient metric, which measures the proportion of overlap between the resulting segmentation and the ground truth. Thus, the combination of loss functions can control the penalization of both FPs and FNs by the BCE term and simultaneously boost the model parameters out of local minima via Dice term.

The training procedure of RSegCNN is the same as that of the traditional CNN, where the trainable weights are updated with the backpropagation through time (BPTT) algorithm [22]. According to Equation (4), the loss for the t-th step with prediction $\hat{y}$ with respect to ground truth y is:

$$\mathcal{L}^t(\hat{y}^t, y^t) = \omega_1 \times \mathcal{L}_{BCE}(\hat{y}^t, y^t) + \omega_2 \times \mathcal{L}_{Dice}(\hat{y}^t, y^t), \tag{7}$$

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

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

$$\begin{aligned}\frac{\partial \mathcal{L}^t}{\partial \hat{y}_j^t} =& \omega_1 \frac{\partial \mathcal{L}_{BCE}(\hat{y}^t, y^t)}{\partial \hat{y}_j^t} + \omega_2 \frac{\partial \mathcal{L}_{Dice}(\hat{y}^t, y^t)}{\partial \hat{y}_j^t} \\ =& -\frac{\omega_1}{N}\left(\frac{y_j^t}{\hat{y}_j^t} - \frac{1 - y_j^t}{1 - \hat{y}_j^t}\right) - \omega_2 \frac{2y_j^t}{\sum_{i=1}^{N} y_i^t + \hat{y}_i^t} + \\ & \omega_2 \frac{2\sum_{i=1}^{N} y_i^t \hat{y}_i^t}{\left(\sum_{i=1}^{N} y_i^t + \hat{y}_i^t\right)^2}\end{aligned} \tag{8}$$

2.5. Evaluation Metrics

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

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

$$\text{Dice} = \frac{2\sum_{i=1}^{N} y_i \hat{y}_i}{\sum_{i=1}^{N} y_i + \hat{y}_i}. \tag{9}$$

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

$$\mathrm{ASD}(A,B) = \frac{d(A,B) + d(B,A)}{2}, \tag{10}$$

$$d(A,B) = \frac{1}{N}\sum_{a\in A} \min_{b\in B} \|a-b\|, \tag{11}$$

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

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

$$\mathrm{HD}(A,B) = \max(h(A,B), h(B,A)) \tag{12}$$

$$h(A,B) = \max_{a\in A} \min_{b\in B} \|a-b\| \tag{13}$$

where $h(A,B)$ denotes the directed HD. The maximum HD is sensitive to contours. When the image is contaminated by noise or occluded, the original Hausdorff distance is prone to mismatch [25,26]. Thus, Huttenlocher proposed the concept of partial Hausdorff distance in 1933 [24]. The 95HD metric is similar to maximum HD. In brief, 95HD selects 95% of the closest points in set B to the point in set A in Equation (13) to calculate $h(A,B)$:

$$95\mathrm{HD} = \max(h^{95\%}(A,B), h^{95\%}(B,A)) \tag{14}$$

$$h^{95\%}(A,B) = \max_{a\in A} \min_{b\in B^{95\%}} \|a-b\| \tag{15}$$

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

3. Experiments

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

3.1. Datasets

3.1.1. CBCT Dataset

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

3.1.2. CT Dataset

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

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

3.2. Implementation Details

We implement all the experiments based on the PyTorch [28] platform developed by Facebook. The experiments are trained on a workstation equipped with an Nvidia P6000 or Tesla V100 GPU. For the data pre-processing, we simply truncated the raw intensity values to be within $[-1000, 2000]$, and then normalized each raw CT case to $[0, 1]$ to decrease the data variance caused by physical considerations of the medical device. Note that different CBCT/CT cases have different physical resolutions. As described in Section 3.1, we maintain their resolutions in a unified resolution of 512 × 512. The weights of the BCE loss term ω_1 and the Dice loss term ω_2 in the loss function are both set to 0.5. We use Adam optimization with a learning rate of $r = 10^{-4}$.

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

3.3. Results

3.3.1. Experiments on the CBCT Dataset

We compare our methods with numerous standard segmentation architectures such as Unet [18], Segnet [29], SegUnet [15], AttUnet [30] and RSegUnet [11]. Table 1 shows the performance comparison as well as the corresponding standard deviation for the mandible segmentation. The average $Dice$, D_{ASD} and D_{95HD} values of the proposed method are 95.31%, 1.2827 mm and 3.1258 mm, respectively. From the Table 1, it can be observed that the proposed method outperforms the existing approaches with respect to $Dice$, D_{ASD} and D_{95HD}. These experimental results indicate that our proposed model with the C2FSeg learning method performs significantly better and achieves the highest overall Dice scores compared to other segmentation methods. According to Table 1, the proposed method also outperforms most other methods, with the second-lowest ASD and 95HD scores.

Table 1. Quantitative comparison of segmentation performance for the CBCT dataset between the proposed method and the state-of-the-art methods.

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

To better demonstrate the performance of the presented approach, several 2D and 3D view examples of the different algorithms are depicted in Figures 4 and 5. Figure 4 shows some examples of ground truth (GT), Unet [18], Segnet [29], SegUnet [15], AttUnet [30], RSegUnet [11] and the proposed method. As shown in the first two rows of Figure 4, the other methods fail to obtain satisfactory results for the main mandible body, while the results from our proposed approach are much better. The third row in Figure 4 show that the proposed algorithm achieves better performances when the upper jaw teeth and lower jaw teeth appear within the same slice. The final row in Figure 4 illustrates that the proposed method can deal with the ambiguity and blurred boundaries common to CBCT scans of the condyles area.

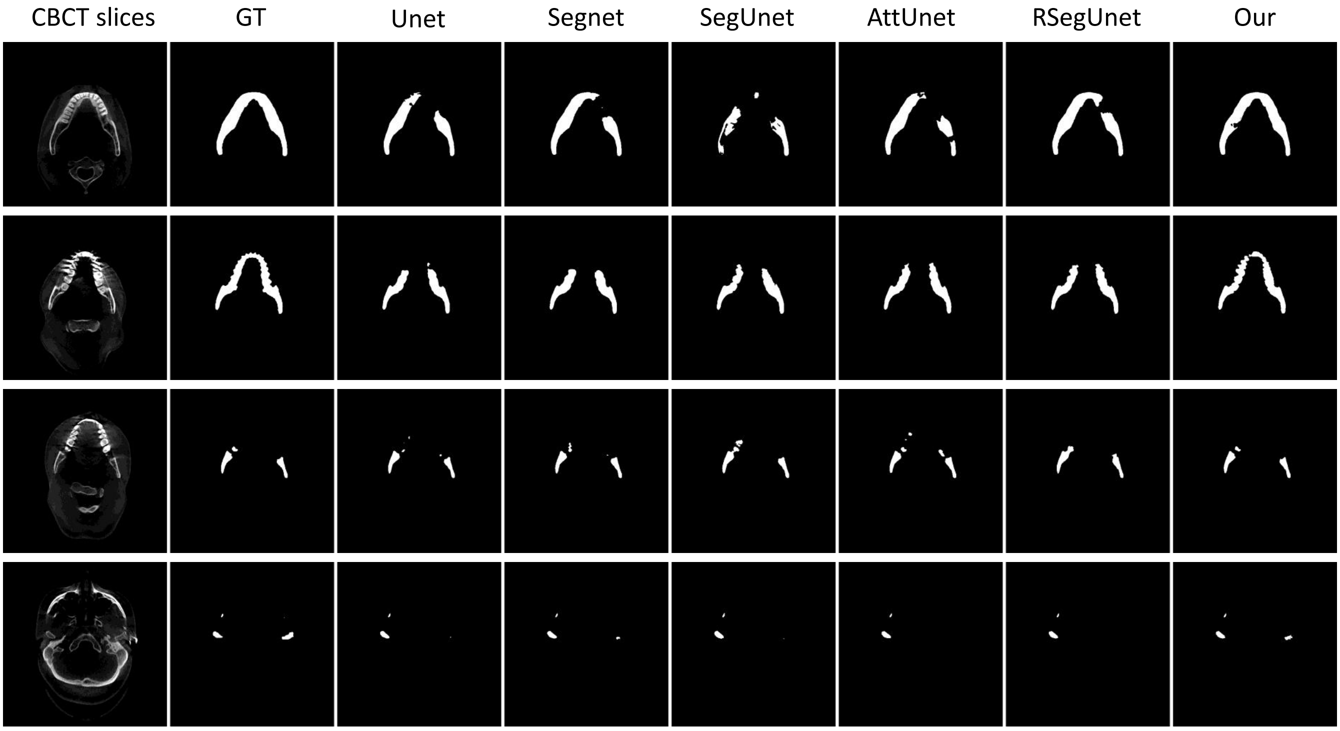

Figure 4. 2D examples from CBCT dataset. From left to right: Original CT slice, Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

Figure 5 illustrates two 3D view examples (the 1st, and 3rd rows) with the corresponding post-processed examples (the 2nd, and 4th rows) obtained from Unet [18], Segnet [29], SegUnet [15], AttUnet [30], RSegUnet [11] and the proposed method, respectively. The first case shown in Figure 5 demonstrate that the proposed method can effectively segment the angle area of the mandible. The second examples shown in Figure 5 show that the ramus, the coronoid process area and the teeth are missed by the other methods while the proposed method can tackle the thin parts of the mandible, which are almost always challenging mandible segmentation tasks.

Table 1, Figures 4 and 5 indicate that the proposed approach is quite accurate in segmenting mandibles affected by metal artifacts in CBCT.

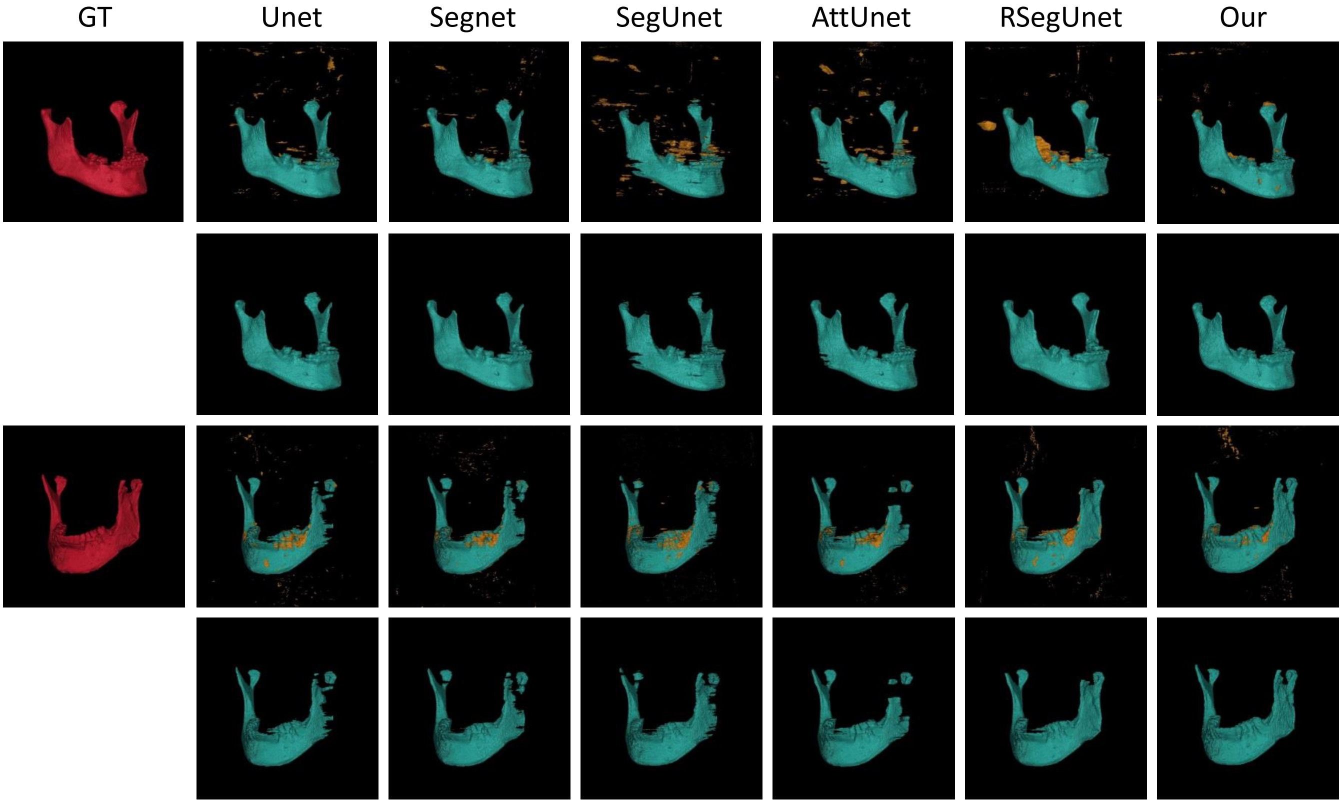

Figure 5. Visual 3D examples of final segmentations from the CBCT dataset. From left to right: Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

3.3.2. Experiments on the CT Dataset

We also test the proposed method on a CT dataset. To quantitatively compare the proposed approach with other methods, we compute the Dice scores, ASD and 95HD values of the five methods. Table 2 lists the average Dice, ASD and 95HD, as well as the corresponding standard deviation. In general, the average values of these metrics obtained from our proposed method are better than those of the other methods. As shown in Table 2, it can be observed that our method yields the highest mean Dice score and the smallest mean errors in 95HD.

Table 2. Quantitative comparison of segmentation performance for the CT dataset between the proposed method and the state-of-the-art methods.

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

Figure 6 shows some examples from the original CT slice, the ground truth (GT) and the results obtained from Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method. As shown in the first two rows of Figure 6, the other methods fail to obtain satisfactory results for the main mandible body and some parts of teeth, while the results from our proposed approach are much better. The three row in Figure 6 shows that the proposed algorithm achieves better performances when the upper jaw teeth and lower jaw teeth appear within the same slice. The final row in Figure 6 illustrates that the proposed method can process the condyles area, which is often ambiguous due to blurred boundaries in CT scans.

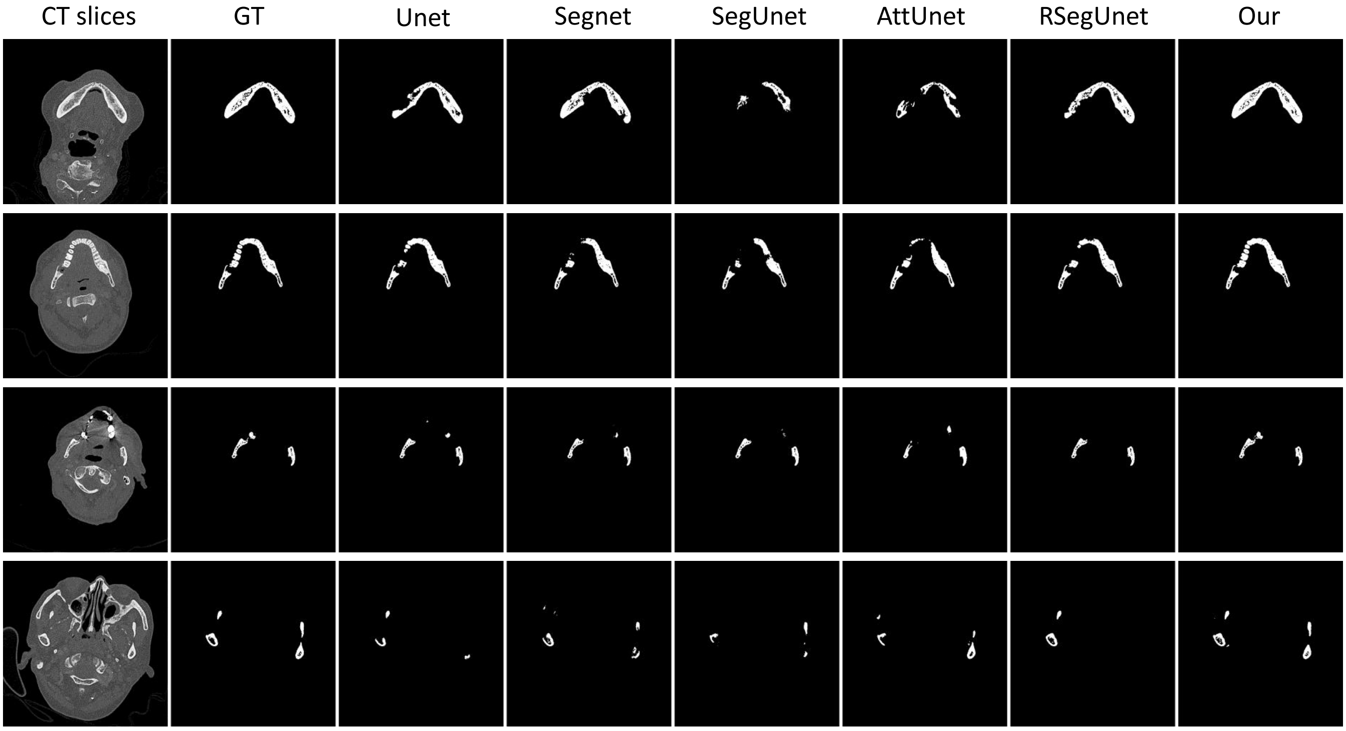

Figure 6. 2D examples from CT dataset. From left to right: Original CT slice, Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

Moreover, two cases of the automatic segmentation (the 1st, and 3rd rows) and the corresponding post-processed examples (the 2nd, and 4th rows) in the 3D views of Unet, Segnet, SegUnet, AttUnet, RSegUnet and the proposed method are displayed in Figure 7. The first case shown in Figure 7 indicate that the proposed method can effectively segment the ramus area and body of the mandible. The second examples shown in Figure 7 show that the angle area and the teeth in the case are missed by the other methods, while the proposed method can tackle the thin parts of the mandible, which are almost always challenging mandible segmentation tasks. The conventional methods usually lead to erroneous segmentation within the whole mandible, as shown in Figure 7. The visual comparison of the automatic segmentation with manual segmentation demonstrates the effectiveness of our method with respect to automatic mandible segmentation. To summarize, Figures 6 and 7 and Table 2 indicate that the proposed approach also works well with respect to the other datasets.

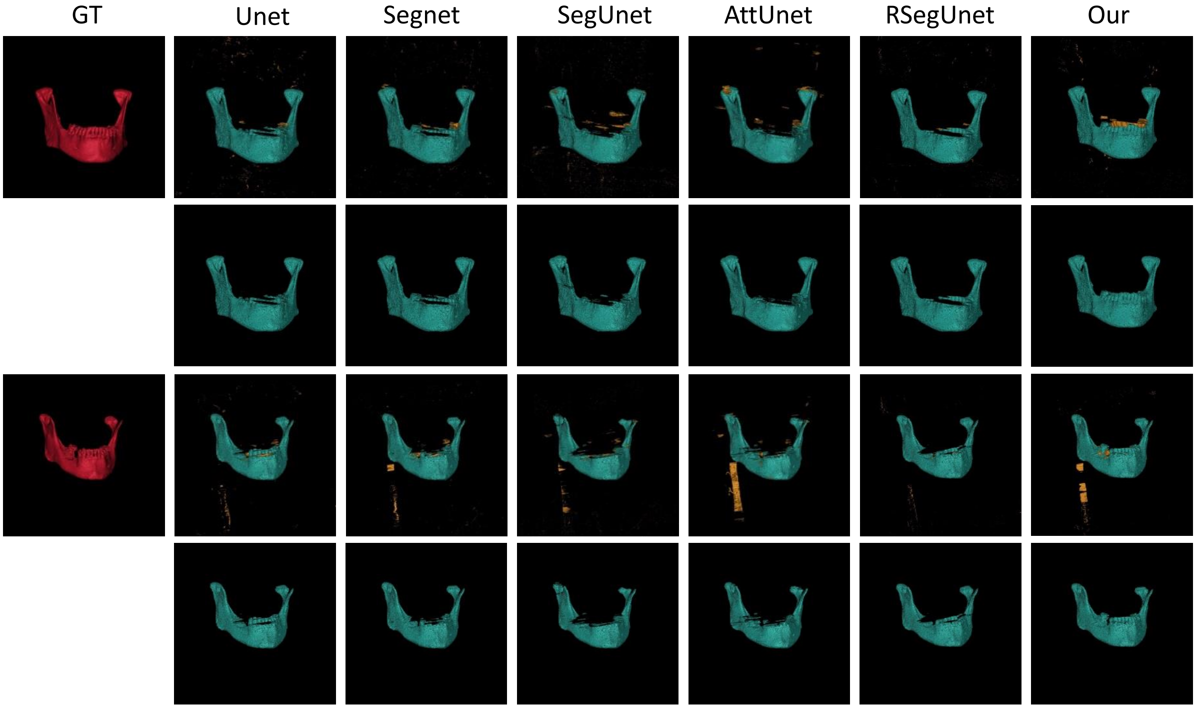

Figure 7. Visual 3D examples of final segmentations from the CT dataset. From left to right: Ground truth (GT), Unet, Segnet, SegUnet, AttUnet, RSegUnet, and the proposed method.

To further investigate the proposed method, it is preferable to test it with respect to a public dataset and measure the performance of the proposed method on the dataset. Here, we compare our proposed method with several state-of-the-art methods with respect to the PDDCA dataset. Table 3 also lists Dice, ASD and 95HD used in the Challenge paper [10,25]. According to Table 3, the performance of the proposed model surpasses the majority of the other methods. The proposed method outperforms other methods, with the third-highest mean Dice score, the lowest ASD and the lowest 95HD. For Dice score results, the segmentation result of our method is only slightly worse than RSegUnet [10], while it is better than RSegUnet in terms of ASD and 95HD.

Table 3. Quantitative comparison of segmentation performance for the PDDCA dataset between the proposed method and the state-of-the-art methods.

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

4. Discussion

In this paper, we present a novel C2FSeg method for mandible segmentation that utilizes the curriculum learning strategy [12] of separating the task into two simpler sub-tasks (coarse-to-fine). We apply 3D SegUnet to look for mandible-like organs in the coarse stage. In the fine stage, recurrent SegUnet is then employed to finely segment the mandible based on the results from the coarse stage. Quantitative evaluation results shown in Tables 1–3 demonstrate that our proposed approach outperforms the state-of-the-art methods for mandible segmentation. In addition, qualitative visual inspection in Figures 4–7 illustrates that our automatic segmentation approach performs quite well in comparison with the ground truth. The direct comparison in PDDCA, as shown in Table 3, illustrates that the proposed method significantly improves mandible segmentation. Remarkably, we found that this segmentation architecture is very robust for weak and blurry edge segmentation. For instance, the networks can segment both condyles and ramus of the mandible quite well, even under the influence of strong metal artifacts. In addition, the results based on the two CT datasets indicate that our proposed approach offers excellent generalization ability, since the images in the two datasets were produced by different imaging technologies.

This method takes advantage of the concept of curriculum learning to simplify the difficult task to several easy sub-tasks. Therefore, the proposed approach can help to learn the rough mandible structure in the coarse stage. Furthermore, the proposed approach utilizes a recurrent network in the fine stage to extract spatial information of objects based on mandible-like candidates from the first stage. The experimental results show that the proposed approach is feasible and effective in 3D mandible segmentation and that it can also be applied to other segmentation tasks. This method can support further research on the 3D image segmentation. It can also help overcome the disadvantage of cropping volume for the 3D network due to the high memory consumption, as well as accomplish 3D segmentation tasks. Furthermore, in this study, 3D SegUnet is used for searching mandible-like organs, and many other networks which have the similar ability for seeking mandible candidates can be used to replace the 3D SegUnet in the coarse stage.

Despite the promising results, there are a few limitations in this study. First, in the experiment, we use 59 orthodontic CBCT scans, 109 CT scans and a public dataset (PDDCA) for the training and the validation of the proposed approach. This is because the collection of CBCT scans is limited. For future work, we will focus on the validation of the C2FSeg approach to experiment on a large number of CBCT scans in order to prove the feasibility of the approach in the clinical setting of 3D VSP. Second, the proposed C2FSeg is a two-stage approach in which the two stages are trained separately. This increases the training duration of the model. In the future, we also aim at improving the efficiency of the approach.

5. Conclusions

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

Author Contributions: Conceptualization, all authors; methodology, B.Q. and J.G.; validation, B.Q. and J.G.; formal analysis, B.Q.; investigation, B.Q., H.v.d.W., J.K., H.H.G. and J.G.; data curation, H.v.d.W., J.K., H.H.G. and B.Q.; writing—original draft preparation, B.Q., H.v.d.W. 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: For the Public Domain Database for Computational Anatomy Dataset (PDDCA) is available in 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 into 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 China Scholarship Council. The authors acknowledge Erhan Saatcioglu for the training data preparation. The authors would like to acknowledge the support received from NVIDIA by providing a GPU as part of their GPU grant program. 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 also would 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. Kraeima, J. Three Dimensional Virtual Surgical Planning for Patient Specific Osteosynthesis and Devices in Oral and Maxillofacial Surgery. A New Era. Ph.D. Thesis, University of Groningen, Groningen, The Netherlands, 2019.
2. Gollmer, S.T.; Buzug, T.M. Fully automatic shape constrained mandible segmentation from cone-beam CT data. In Proceedings of the 2012 9th IEEE International Symposium on Biomedical Imaging (ISBI), Barcelona, Spain, 2–5 May 2012; pp. 1272–1275.
3. Wang, L.; Gao, Y.; Shi, F.; Li, G.; Chen, K.C.; Tang, Z.; Xia, J.J.; Shen, D. Automated segmentation of dental CBCT image with prior-guided sequential random forests. *Med. Phys.* **2016**, *43*, 336–346. [CrossRef] [PubMed]
4. Indraswari, R.; Arifin, A.Z.; Suciati, N.; Astuti, E.R.; Kurita, T. Automatic segmentation of mandibular cortical bone on cone-beam CT images based on histogram thresholding and polynomial fitting. *Int. J. Intell. Eng. Syst.* **2019**, *12*, 130–141. [CrossRef]
5. Linares, O.C.; Bianchi, J.; Raveli, D.; Neto, J.B.; Hamann, B. Mandible and skull segmentation in cone beam computed tomography using super-voxels and graph clustering. *Vis. Comput.* **2019**, *35*, 1461–1474.
6. Fan, Y.; Beare, R.; Matthews, H.; Schneider, P.; Kilpatrick, N.; Clement, J.; Claes, P.; Penington, A.; Adamson, C. Marker-based watershed transform method for fully automatic mandibular segmentation from CBCT images. *Dentomaxillofac. Radiol.* **2019**, *48*, 20180261. [CrossRef] [PubMed]
7. 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]
8. 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]
9. 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.
10. 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]
11. Qiu, B.; Guo, J.; Kraeima, J.; Glas, H.H.; Borra, R.J.; Witjes, M.J.; van Ooijen, P.M. Recurrent convolutional neural networks for mandible segmentation from computed tomography. *arXiv* **2020**, arXiv:2003.06486.
12. Bengio, Y.; Louradour, J.; Collobert, R.; Weston, J. Curriculum learning. In Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, QC, Canada, 14–18 June 2009; pp. 41–48.
13. 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]

14. 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 (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
15. 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.* **2019**, *21*, 1467–1479. [CrossRef]
16. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv* **2015**, arXiv:1502.03167.
17. Nair, V.; Hinton, G.E. Rectified linear units improve restricted boltzmann machines. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 807–814.
18. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-ASSISTED Intervention, Munich, Germany, 5–9 October 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.
19. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings, Sardinia, Italy, 13–15 May 2010; pp. 249–256.
20. Abulnaga, S.M.; Rubin, J. Ischemic stroke lesion segmentation in ct perfusion scans using pyramid pooling and focal loss. In Proceedings of the International MICCAI Brainlesion Workshop, Granada, Spain, 16 September 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 352–363. [CrossRef]
21. 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]
22. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436. [CrossRef]
23. 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*, 5110. [CrossRef] [PubMed]
24. Huttenlocher, D.P.; Rucklidge, W.J.; Klanderman, G.A. Comparing images using the Hausdorff distance under translation. In Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, USA, 15–18 June 1992; pp. 654–656. [CrossRef]
25. 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] [PubMed]
26. Taha, A.A.; Hanbury, A. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Med. Imaging* **2015**, *15*, 29. [CrossRef] [PubMed]
27. Ren, X.; Xiang, L.; Nie, D.; Shao, Y.; Zhang, H.; Shen, D.; Wang, Q. Interleaved 3D-CNN s for joint segmentation of small-volume structures in head and neck CT images. *Med. Phys.* **2018**, *45*, 2063–2075. [CrossRef]
28. 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.
29. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [CrossRef]
30. 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.
31. 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 Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 5–9 October 2015.
32. 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 Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 5–9 October 2015.
33. 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 Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 5–9 October 2015.
34. Orbes-Arteaga, M.; Pea, D.; Dominguez, G. Head and neck auto segmentation challenge based on non-local generative models. In Proceedings of the Head Neck Auto-Segmentation Challenge (MICCAI), Munich, Germany, 5–9 October 2015.
35. Kodym, O.; Španěl, M.; Herout, A. Segmentation of Head and Neck Organs at Risk Using CNN with Batch Dice Loss. *arXiv* **2018**, arXiv:1812.02427.
36. 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] [PubMed]
37. 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.

38. 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] [PubMed]
39. Qiu, B.; van der Wel, H.; Kraeima, J.; Glas, H.H.; Guo, J.; Borra, R.J.; Witjes, M.J.H.; van Ooijen, P. Robust and Accurate Mandible Segmentation on Dental CBCT Scans Affected by Metal Artifacts Using a Prior Shape Model. *J. Pers. Med.* **2021**, *11*, 364. [CrossRef]