Automated Segmentation of Dental CBCT Image Using an Improved U-Net Network

Zeyu Chen[1], Senyang Chen[2], Songming Liu[3]

[1] School of Computer Science and Artificial Intelligence, Changzhou University, Changzhou 213164, China

[2] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China

[3] Zhejiang Chuanran Intelligent Technology Group Co., Ltd. Hangzhou 310000, China

DOI: 10.32629/jcmr.v4i2.1191

Abstract: In the field of clinical dental medicine, Cone Beam Computed Tomography (CBCT) is a useful tool for the measurement of various dimensions related to the oral cavity, including height and thickness. This provides invaluable guidance and reference for risk assessment in orthodontic treatment, selection of treatment plans and implant treatment. However, segmentation of the teeth region from CBCT images is a daunting task due to complex root morphology and indistinct boundaries between the root and the alveolar bone. Manual annotation of the teeth area is resource-intensive, and deep learning-based segmentation methods are susceptible to noise, reducing their efficiency. To tackle these complexities, a multi-filter attention module is proposed in this paper, which effectively minimizes the noise in CBCT images through utilization of multiple filters and self-attention techniques. Additionally, an Improved U-Net model is proposed, where the original convolution block in the U-Net is replaced with a Double ConvNeXt block to yield better network performance. Experimentally, the proposed Improved U-Net method showed remarkable progress as it achieved a Dice Similarity Coefficient of 86.95% in oral CBCT image segmentation, surpassing existing models and affirming the effectiveness and advancedness of the proposed model and method.

***Keywords*:** deep learning, images segmentation, medical image processing, convolutional neural networks, image denoising

1. Introduction

Cone-beam computed tomography (CBCT) is a relatively new technology that has been in clinical use since the turn of the century. CBCT provides more accurate 3D images of dental structures compared to traditional dental X-rays [1]. In the clinic, CBCT image segmentation and the construction of 3D dental models enables more precise and comprehensive analysis of patients' oral conditions by dentists. CBCT can be used to assist in the diagnosis of bite problems, oral tumors, and facial injuries. Additionally, it assists dental surgeons in accurately planning dental and orthodontic surgeries, as well as developing detailed digital dental correction schemes [2]. Translation: But the segmentation of CBCT image data is a complex task, and the traditional dental CBCT image segmentation method require professional doctors to perform annotation. The annotation results depend on the doctors' subjective judgments and also require a lot of time, which greatly limits efficiency [3].

In recent years, deep learning technology has been extensively applied to medical image segmentation [4]. Deep learning is an artificial neural network structure that can automatically extract features from large-scale and complex data through hierarchical learning and provide higher-precision solutions to various problems through data analysis and classification. Compared with traditional computer vision technology, deep learning has higher automation and robustness [5]. Currently, deep learning technology is gradually being applied to the analysis and diagnosis of medical images, especially in medical image segmentation, where deep learning technology has achieved significant advantages. By using deep learning technology, the performance of image recognition, pathology analysis, accurate detection, segmentation, and quantitative analysis can be optimized during the medical image segmentation process. Deep learning techniques have also been gradually used in dental cbct image segmentation tasks[6-10], Minnema et al. proposed a mixed-scale dense (MS-D) convolutional neural network to reduce the metal artifacts on the segmentation accuracy of the image, and accurately segment the bone structure in the image [11]. Wang et al. modified the MS-D network to make it suitable for multi-class segmentation of jaw, teeth, and background in CBCT scans [12]. Zheng et al. proposed a novel anatomically constrained dense U-Net, which combines oral anatomical knowledge with a data-driven dense U-Net to improve computational efficiency and segmentation accuracy [13]. Duan et al. proposed a two-stage dental pulp segmentation network based on U-Net [14], and Duong et al. also proposed a framework for alveolar bone segmentation based on U-Net and noise reduction techniques [15]. Jang et al. proposed a deep learning-based hierarchical multi-step model to segment individual teeth automatically [16]. Awari et al. introduced a novel 3D dental

image segmentation and classification method using deep learning with Sac Swarm Optimization (3DDISC-DLTSA) model [17], conducted extensive experimental analysis to demonstrate the effectiveness of the model. Deep learning technology can be used to separate the teeth efficiently and accurately from the CBCT images, which will have an important role in improving the application effectiveness and case analysis in the medical imaging field [18].

However, there are still limitations to using deep learning technology for the segmentation of dental CBCT images. Firstly, objects in the oral cavity such as teeth and soft tissues have various shapes and structures, requiring segmentation algorithms to consider the complexity of object morphology and the irregularity of boundaries. Additionally, blind spots that occur when collecting CBCT data may result in partial volume effects, making it difficult to distinguish the surface and boundary of different objects and thus increasing the difficulty of segmentation. Secondly, CBCT images may be affected by factors such as X-ray beam angle, scanning parameters, noise, ghosting, and the intersection of multiple objects. These interference factors may result in artifacts and blurring in the images, thereby affecting the accuracy of segmentation. In summary, the deep learning-based segmentation of CBCT images is a challenging task that requires comprehensive and in-depth considerations and research into network structures.

The main contributions of this paper are summarized as follows.

We propose a multi-filter attention (MFA) module, which effectively alleviates the influence of noise and artifacts on segmentation accuracy.

Referring to the improvement of the original ResNet[29] by ConvNeXt[30], we proposed a Double ConvNeXt (DCN) block to replace the ordinary convolution module in U-Net, and combine with MFA module to propose an improved U-Net network.

Our proposed Improved U-Net Network obtained 86.95% DSC in dental CBCT segmentation task, outperforming the existing models.

2. Related Works

In the field of medical image segmentation, U-Net is the classic and most commonly used network, which was proposed by Ronneberger et al. in MICCAI conference 2015 [19]. U-Net consists of two parts: Encoder and Decoder, which can simultaneously learn position information and regional semantic information, achieving highly accurate image segmentation. The Encoder part uses the VGG network [20], which uses convolutional layers and pooling layers to reduce the resolution of the original image. In the Decoder part, up-convolutional layers are used for upsampling, restoring the resolution, and shallow feature information and deep feature information are fused through skip-connections, resulting in precise segmentation.

In the field of medical imaging, U-Net is widely used for the segmentation of medical images such as CT and MRI, including specific regions such as liver [21], lung [22], and heart [23]. In addition, there are some studies that combine U-Net with other neural networks to further improve the performance of the model, such as Attention UNet [24], TransUNet [25], and Swin-UNet [26]. Attention UNet's main improvement is the addition of a attention mechanism. Attention UNet uses attention gate to weigh feature maps and focus on a range of features, from low-level to high-level and everything in between, improving the network's segmentation accuracy. TransUNet and Swin-Unet mainly enhance the model's receptive field and macro-expression ability of features by introducing the Transformer, allowing the model to better capture global dependency relationships in the image, thereby improving U-Net's performance in image segmentation tasks.

SegNet[27] and DeepLab V3+[28], in addition to U-Net and its related networks, are also frequently used image segmentation networks in both medical and natural image segmentation tasks. These two networks, SegNet and DeepLabv3+, are similar to U-Net, all of which have encoder-decoder architecture models. SegNet is characterized by its shallow network structure, as well as the combination of drastic downsampling and upsampling, fewer parameters, and higher accuracy. DeepLabv3+ is based on an atrous convolutional network, focusing on extracting and expressing multi-scale semantic information at different receptive fields on the same feature map to achieve accurate image segmentation.

All of the above traditional segmentation models utilize the encoder-decoder and skip connection structures, which can combine features of different receptive fields, effectively locate, and preserve detailed features, and prevent gradient diffusion. Nevertheless, traditional networks are difficult to handle the random noise and metal artifacts in dental CBCT images, which significantly affects the segmentation quality of the model. Meanwhile, the unclear boundary between the tooth root and alveolar bone presents a challenge for the model's performance. Therefore, further research on deep learning based dental CBCT image segmentation is needed.

3. Methodologies

3.1 Network Structure

The improved U-Net model structure we proposed is shown in Figure 1. The main part of the model is the same as U-Net in structure, both using U-shaped structure, and utilizing skip connection to connect the encoder with the decoder. The difference is that we added MFA module before the main part of the network, which is used to reduce the impact of noise on the quality of CBCT images. We also replaced the original convolution block with DCN block and modified the upsampling and downsampling methods to improve the performance of the model and the final segmentation effect. Both the encoder and decoder include four stages, with each stage consisting of a DCN block and an upsampling or downsampling.

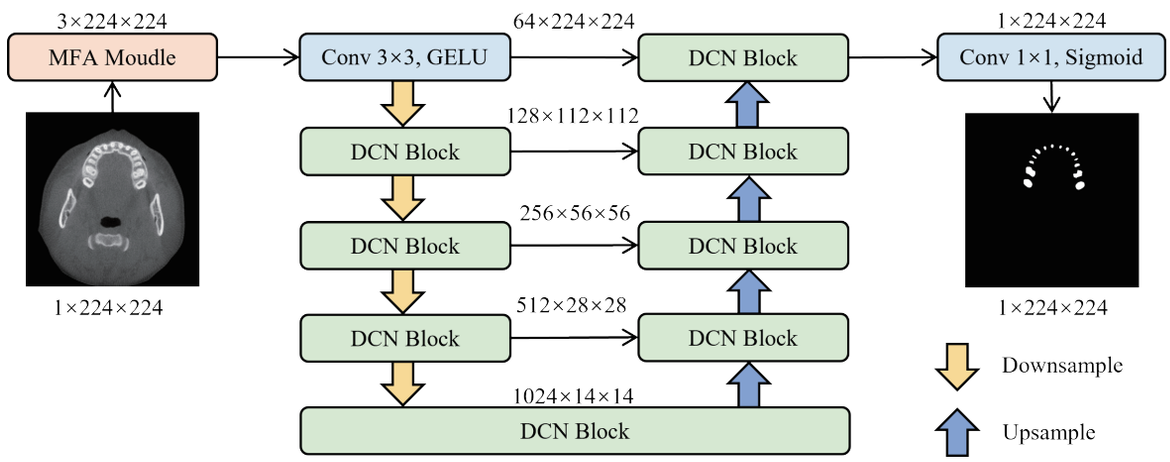

Figure 1. Proposed improved U-Net structure

U-Net uses global max-pooling layer as the downsampling method, and 1x1 convolution combined with nearest neighbor interpolation as the upsampling method. The improved U-Net replaces the global max-pooling with the convolution operation that uses a 2x2 kernel and a stride of 2. Convolution operation has learnable parameters, which can be adjusted based on data to make the downsampling effect more suitable for actual needs, while global max-pooling has no learnable parameters and cannot be adjusted. Therefore, using convolution for downsampling preserves more details of the original image and avoids the loss of information that may occur in global max-pooling. Similar to downsampling, using interpolation for upsampling also has similar problems as global max-pooling. So, we use transposed convolution with a kernel size of 2x2 and a stride of 2 to replace the interpolation operation. Additionally, before upsampling and downsampling, we used layer normalization to normalize the data and ensure training stability during the process. The implementation details of upsampling and downsampling are shown in Figure. 2.

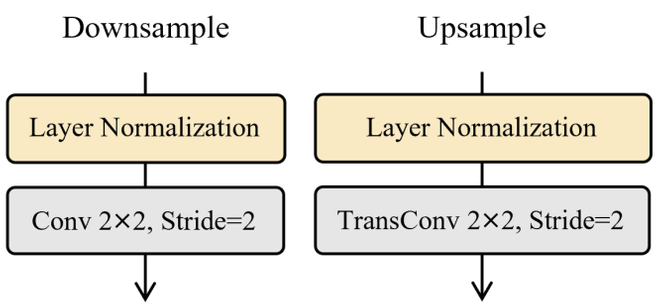

Figure 2. The implementation details of upsampling and downsampling

Additionally, we referred to ConvNeXT, replaced the original convolution blocks with DCN blocks. A DCN block contains two groups of basic blocks, each of which includes a depth convolution with a convolution kernel size of 7×7, and two convolutions with a convolution kernel size of 1×1. This structure can significantly reduce the network parameters. Furthermore, replacing batch normalization with layer normalization, and ReLU with GELU have both improved the performance of the network to a certain extent. The designed Global Response Normalization (GRN) encourages mutual competition of features among different channels, thereby strengthening the expressive ability of the network.

The implementation details of this module is shown in the Figure 3.

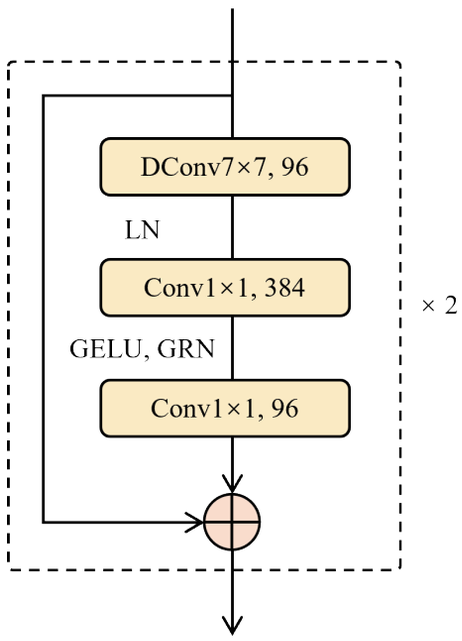

Figure 3. The implementation details of the DCN block

3.2 Multi-Filter Attention Module

Traditional image denoising methods often use filters to reduce noise in images, such as median and Gaussian filters. However, single filters have limitations. For example, median filters are mainly aimed at shot noise, while Gaussian filters are designed to target Gaussian noise. Using a single filter to denoise CBCT images may blur the image further. Therefore, this paper proposes a method that integrates multiple filtering methods and uses an attention mechanism to model the correlation between the original image and the denoised image. The proposed LDM is shown in the Figure 4.

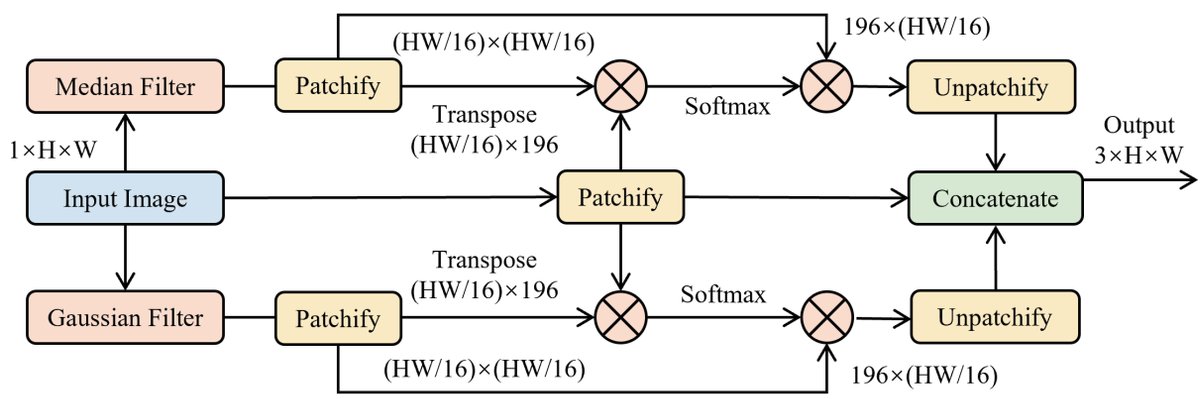

Figure 4. The implementation details MFA module

This module is divided into three paths: the median filtering path, the Gaussian filtering path, and the original image path. In the non-original image path, the original image is first denoised using a filter to obtain a denoised image that is the same size as the original image. Secondly, with reference to the Self-attention mechanism, the dependence relationship between the denoised image and the original image is calculated. Specifically, the original image denoised image is embedded into multiple patches, the patch size is 16×16, then multiplied by the original image to obtain a similarity matrix. Calculate the weight using the Softmax function and then multiply the weight by patches obtain the result. Finally, the original image, the result from the median filtering path, and the result from the Gaussian filtering path are combined to obtain the output of the MFA module. This module effectively utilizes multiple filters and attention mechanism to obtain denoised images while also preventing information loss in the original images.

4. Experiments

4.1 Dataset and experiment details

In this paper, we collected and annotated 45 volumes of dental CBCT scans, with a total of 7,224 labeled CBCT images. The original size of the dental CBCT images was 1×536×536, but they were resized to 1×224×224 and normalized to grayscale values between 0 and 1 for dataset construction.

The experiment is based on the PyTorch framework, which is developed based on the Python language, inspired by Torch, and released in 2016. It has been widely used in the fields of machine learning and deep learning, and is widely used in training deep neural networks, generative adversarial networks, natural language processing, computer vision and many other tasks.

The operating system is Windows 10, the CPU is an Intel Core i7-12700KF, and the GPU is NVIDIA GeForce GTX 3070TI. When training, the optimization method uses random gradient descent. Its momentum parameter is set to 0.9, and the weight decay rate is 1×10−4. The learning rate uses the initial learning rate learning rate 6×10−4 multiplied by (1 – current iteration number / max iteration number)0.9. The number of experimental training epoch is 30, the batch size is 16, and the number of total iterations is 10,710.

Due to the relatively small proportion of the tooth area in the dental CBCT image, evaluation metric such as accuracy, precision, and recall cannot effectively compare the differences in the segmentation quality of the models. Therefore, this paper uses the Dice similarity coefficient (DSC) as the evaluation metric and uses the Dice Loss as the loss function. Where DSC is shown in the following equation:

$$\mathrm{DSC}(X,Y)=\frac{2|X\cap Y|}{|X|+|Y|} \tag{1}$$

Where X is the segmentation result of the network and Y is the ground truth. Dice Loss can be calculated from 1-DSC(X, Y).

4.2 Comparison with Existing Models

To demonstrate the effectiveness of the Improved U-Net proposed by us, we trained it on the CBCT image segmentation dataset and compared it with existing networks. The loss curve during the training process is shown in the Figure 5.

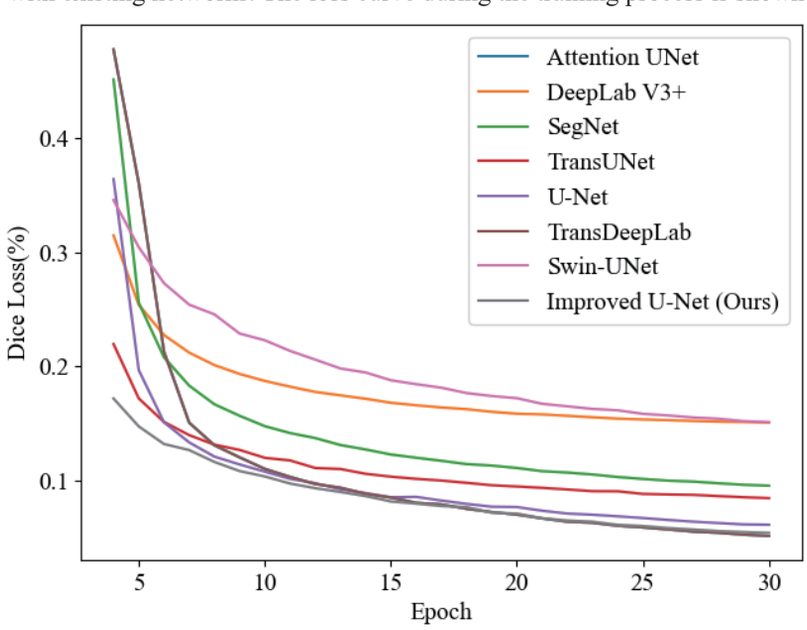

Figure 5. Train loss curve

From the training curve, it can be seen that under the same training parameters, Improved U-Net can converge faster and reduce the Dice Loss to a lower level compared to other compared networks, indicating that Improved U-Net has more stable performance and higher convergence speed. The segmentation results of each network are shown in Table 1.

Table 1. Performances of different network

Network	DSC (%)
U-Net	85.97
Attention UNet	85.52
SegNet	81.99
DeepLab V3+	79.48
TransDeepLab	74.94
TransUNet	85.93
Swin-UNet	74.00
Ours	86.95

The segmentation results in the table above show that our proposed Improved U-Net model achieved a DSC score of 86.95%, indicating superior performance compared to existing models. Among the existing models, while Attention U-Net, TransUNet, and U-Net have demonstrated good segmentation performance, Swin-U-Net has performed poorly. Despite the fact that all of these networks are members of the U-Net family, Swin-U-Net has a pure transformer architecture. The transformer's challenging training, due to the lower amount of data in the dental CBCT image dataset, caused the weaker segmentation performance of Swin-U-Net. In a similar way, TransDeepLab also exhibited decreased performance in comparison to DeepLab V3+. The visualization of the segmentation results is shown in Figure 6.

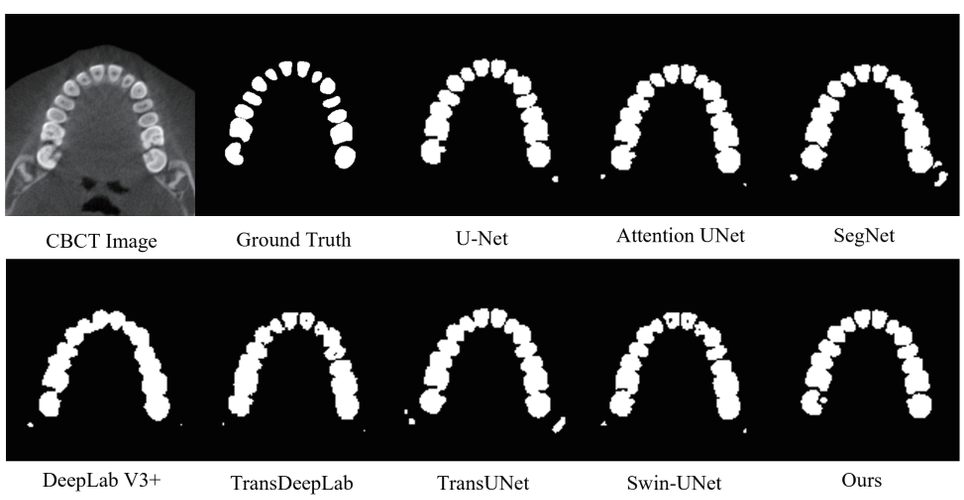

Figure 6. The visualization of the segmentation results

4.3 Ablation studies

Table 2. Ablation study segmentation results

NO.	DCNB	MFA module			DSC (%)
		3×3	5×5	7×7	
1	×	×	×	×	85.97
2	√	×	×	×	86.43
3	√	√	×	×	86.95
4	√	×	√	×	86.68
5	√	×	×	√	85.39

The results of the ablation studies are shown in the Table 2. In experiments 1 to 2, to demonstrate the effectiveness of DCNB, we trained the original U-Net network (baseline) and U-Net with DCNB replacing the convolution blocks. The results showed that U-Net with DCNB replacing the convolution blocks improved performance by 0.46% compared to the baseline, proving the effectiveness of DCNB in improving network performance.

In experiments 2-5, to demonstrate the effectiveness of the MFA module, and to verify the effect of different filter sizes on the module, we added MFA modules using 3×3, 5×5, and 7×7 to the improved U-Net, and compared them with the results of experiment 2. The results showed that using MFA module can effectively improve network performance, and the

best effect was achieved using the MFA module with a filter size of 3×3. The effect was slightly worse when a filter size of 5×5 was used, and the effect of using the MFA module with a filter size of 7×7 was basically the same as without the MFA module, which may be because the overlarge filter made the image blurry and lost too much detail information, thus reducing the effectiveness. The visualization of the segmentation results is shown in Figure 7.

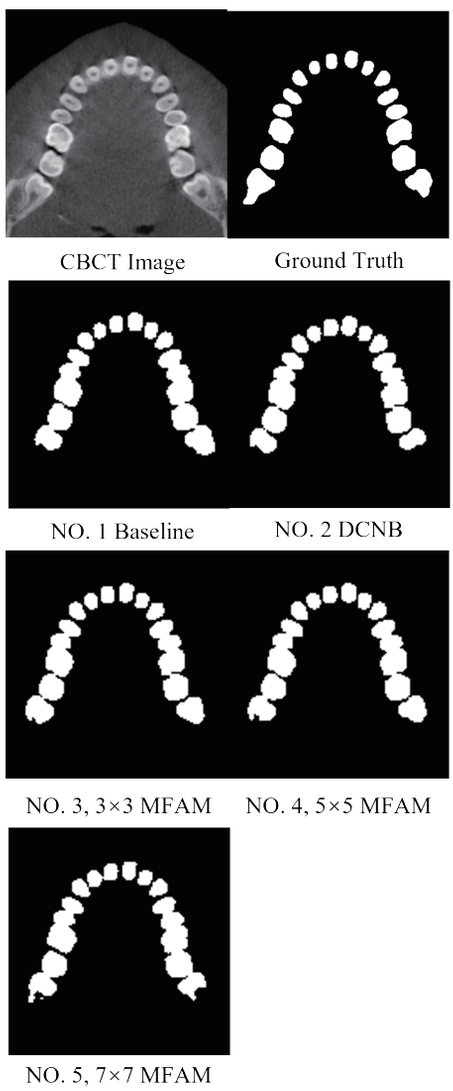

Figure 7. The visualization of the ablation study segmentation results

5. Conclusions

This paper delves into the constraints of dental CBCT image segmentation methods, both traditional and deep learning-based. We collected and annotated CBCT data from 45 patients, and constructed a new dental CBCT image segmentation dataset. We proposed an Improved U-Net, in which an MFA module is added to the input part of the network to reduce noise in the CBCT images. Based on the original U-Net network structure, we modified the upsample and downsampling layers and replaced the original convolutional blocks with DCN blocks, which improved the network performance and prevented the loss of feature information.

The experimental results are encouraging, with the proposed Improved U-Net model outperforming traditional models, obtaining an 86.95% DSC in dental CBCT image segmentation tasks. This finding is of particular significance for a wide range of applications, including orthodontics and implant dentistry.

Data Availability

The data are not publicly available due to patient privacy concerns.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] Loubele M, Bogaerts R, Van Dijck E, et al. Comparison between effective radiation dose of CBCT and MSCT scanners for dentomaxillofacial applications[J]. European journal of radiology, 2009, 71(3): 461-468.

[2] Mah J K, Yi L, Huang R C, et al. Advanced applications of cone beam computed tomography in orthodontics[C]//Seminars in Orthodontics. WB Saunders, 2011, 17(1): 57-71.

[3] Hwang J J, Jung Y H, Cho B H, et al. An overview of deep learning in the field of dentistry[J]. Imaging science in dentistry, 2019, 49(1): 1-7.

[4] Liu X, Song L, Liu S, et al. A review of deep-learning-based medical image segmentation methods[J]. Sustainability, 2021, 13(3): 1224.

[5] Hesamian M H, Jia W, He X, et al. Deep learning techniques for medical image segmentation: achievements and challenges[J]. Journal of digital imaging, 2019, 32: 582-596.

[6] Shaheen E, Leite A, Alqahtani K A, et al. A novel deep learning system for multi-class tooth segmentation and classification on cone beam computed tomography. A validation study[J]. Journal of Dentistry, 2021, 115: 103865.

[7] Minnema J, van Eijnatten M, Hendriksen A A, et al. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network[J]. Medical physics, 2019, 46(11): 5027-5035.

[8] Wang X, Meng X, Yan S. Deep learning-based image segmentation of cone-beam computed tomography images for oral lesion detection[J]. Journal of Healthcare Engineering, 2021, 2021: 1-7.

[9] Ma J, Yang X. Automatic dental root CBCT image segmentation based on CNN and level set method[C]//Medical Imaging 2019: Image Processing. SPIE, 2019, 10949: 668-674.

[10] Corbella S, Srinivas S, Cabitza F. Applications of deep learning in dentistry[J]. Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology, 2021, 132(2): 225-238.

[11] Minnema J, van Eijnatten M, Hendriksen A A, et al. Segmentation of dental cone-beam CT scans affected by metal artifacts using a mixed-scale dense convolutional neural network[J]. Medical physics, 2019, 46(11): 5027-5035.

[12] Wang H, Minnema J, Batenburg K J, et al. Multiclass CBCT image segmentation for orthodontics with deep learning[J]. Journal of dental research, 2021, 100(9): 943-949.

[13] Zheng Z, Yan H, Setzer F C, et al. Anatomically constrained deep learning for automating dental CBCT segmentation and lesion detection[J]. IEEE Transactions on Automation Science and Engineering

[14] Duan W, Chen Y, Zhang Q, et al. Refined tooth and pulp segmentation using U-Net in CBCT image[J]. Dentomaxillofacial Radiology, 2021, 50(6): 20200251.

[15] Duong D Q, Nguyen K C T, Kaipatur N R, et al. Fully automated segmentation of alveolar bone using deep convolutional neural networks from intraoral ultrasound images[C]//2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, 2019: 6632-6635.

[16] Jang T J, Kim K C, Cho H C, et al. A fully automated method for 3D individual tooth identification and segmentation in dental CBCT[J]. IEEE transactions on pattern analysis and machine intelligence, 2021, 44(10): 6562-6568.

[17] Awari H, Subramani N, Janagaraj A, et al. Three-dimensional dental image segmentation and classification using deep learning with tunicate swarm algorithm[J]. Expert Systems, 2022: e13198.

[18] Lahoud P, Diels S, Niclaes L, et al. Development and validation of a novel artificial intelligence driven tool for accurate mandibular canal segmentation on CBCT[J]. Journal of dentistry, 2022, 116: 103891.

[19] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[20] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[21] Liu Z, Song Y Q, Sheng V S, et al. Liver CT sequence segmentation based with improved U-Net and graph cut[J]. Expert Systems with Applications, 2019, 126: 54-63.

[22] Chen K, Xuan Y, Lin A, et al. Lung computed tomography image segmentation based on U-Net network fused with dilated convolution[J]. Computer Methods and Programs in Biomedicine, 2021, 207: 106170.

[23] Diniz J O B, Ferreira J L, Cortes O A C, et al. An automatic approach for heart segmentation in CT scans through image processing techniques and Concat-U-Net[J]. Expert Systems with Applications, 2022, 196: 116632.

[24] Oktay O, Schlemper J, Folgoc L L, et al. Attention u-net: Learning where to look for the pancreas[J]. arXiv preprint arXiv:1804.03999, 2018.

[25] Chen J, Lu Y, Yu Q, et al. Transunet: Transformers make strong encoders for medical image segmentation[J]. arXiv preprint arXiv:2102.04306, 2021.
[26] Cao H, Wang Y, Chen J, et al. Swin-unet: Unet-like pure transformer for medical image segmentation[J]. preprint arXiv:2105.05537, 2021.
[27] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.
[28] Chen L C, Zhu Y, Papandreou G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.
[29] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[30] Woo S, Debnath S, Hu R, et al. ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders[J]. arXiv preprint arXiv:2301.00808, 2023.