



CariesNet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic X-ray image

Haihua Zhu¹ · Zheng Cao² · Luya Lian¹ · Guanchen Ye¹ · Honghao Gao^{3,4} · Jian Wu⁵

Received: 15 August 2021 / Accepted: 27 October 2021

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

Dental caries has been a common health issue throughout the world, which can even lead to dental pulp and root apical inflammation eventually. Timely and effective treatment of dental caries is vital for patients to reduce pain. Traditional caries disease diagnosis methods like naked-eye detection and panoramic radiograph examinations rely on experienced doctors, which may cause misdiagnosis and high time-consuming. To this end, we propose a novel deep learning architecture called CariesNet to delineate different caries degrees from panoramic radiographs. We firstly collect a high-quality panoramic radiograph dataset with 3127 well-delineated caries lesions, including shallow caries, moderate caries, and deep caries. Then we construct CariesNet as a U-shape network with the additional full-scale axial attention module to segment these three caries types from the oral panoramic images. Moreover, we test the segmentation performance between CariesNet and other baseline methods. Experiments show that our method can achieve a mean 93.64% Dice coefficient and 93.61% accuracy in the segmentation of three different levels of caries.

Keywords Dental caries · Computer-aided diagnosis · Segmentation · Deep learning

1 Introduction

Dental caries is defined as a localized disease which affects the hard tissues of teeth caused by microorganisms in plaque [1]. It is one of the most common oral diseases.

H. Zhu and Z. Cao contribute equally.

✉ Honghao Gao
gaohonghao@shu.edu.cn

✉ Jian Wu
wujian2000@zju.edu.cn

¹ Stomatology Hospital, School of Stomatology, Zhejiang University School of Medicine, Clinical Research Center for Oral Diseases of Zhejiang Province, Key Laboratory of Oral Biomedical Research of Zhejiang Province, Cancer Center of Zhejiang University, Hangzhou 310006, China

² Real Doctor AI Research Centre, College of Computer Science and Technology, Zhejiang University, Hangzhou 310006, China

³ School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China

⁴ Gachon University, Gyeonggi-Do 461-701, South Korea

⁵ First Affiliated Hospital School of Medicine, and School of Public Health, Zhejiang University, Hangzhou 310058, China

According to the 4th Chinese National Oral Health Epidemiological Survey in 2015, the prevalence of caries in primary teeth of children aged 3–5 is as high as 70.81%. With the increase of age, the incidence rate of caries also mounts progressively. The prevalence of dental caries is 80.7% in the 65–74 age group [2]. At the same time, dental caries causes a huge social and economic burden. A global burden of diseases study showed that about 3.5 billion people worldwide suffer from oral disease and the direct cost of treating these diseases is 298 billion dollars [3].

Meanwhile, the diagnosis of dental caries depends on the subjectivity of doctors in clinic. The traditional dental caries diagnosis method is mainly discovered by the attending doctor through the visual inspection and the probe exploration, which has certain subjectivity. Given that the early caries and the hidden caries are difficult to be detected, the misdiagnosis rate is high. If not treated timely, the dental caries may expand gradually, invade the dental pulp, trigger the tooth apical inflammation, the apical abscess and other dental diseases, and eventually the teeth may fall off.

Oral panoramic radiographs (X-rays) play a critical role in the diagnoses of dental diseases like caries. As a preventative diagnostic tool, dentists can utilize oral

panoramic radiographs, a preventive diagnostic tool, to find hidden dental structure, bone loss, malignant or benign masses and cavities that cannot be examined under visual examination. Caries decay can be reflected radiographically when there is sufficient decalcification of tooth structures [1]. X-ray image of dental caries shows different gray value in different developing stages. According to [4], the shallow caries is defined as caries radiolucency in enamel or in the outer third of dentin; moderate caries is defined as caries radiolucency in the middle third of dentin; and deep caries is defined as caries radiolucency in the inner third of dentin with or without apparent pulp involvement. An example of panoramic radiograph image is shown in Fig. 1. The Boxes A, B and C are corresponding to shallow caries, moderate caries and deep caries, respectively.

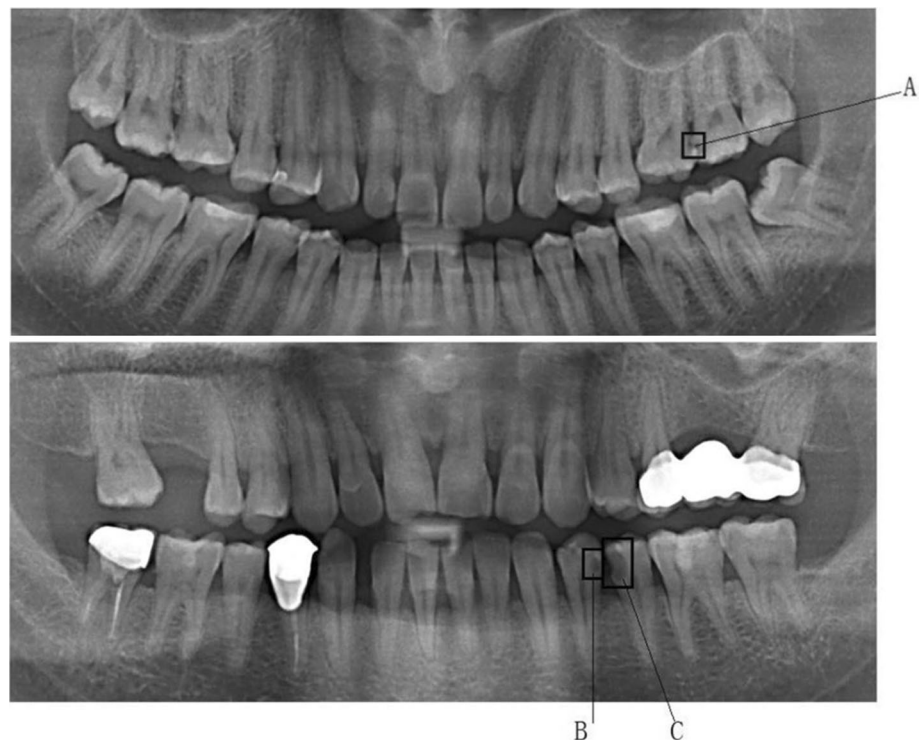
Computer-aided diagnosis system provides a more efficient method to solve the above problems. By means of the analysis and calculation ability of computer, a mathematical model of a disease diagnoses is established. Furthermore, the classification, prediction and localization of the lesions of this type of disease can greatly alleviate the burden and reduce the difficulty for clinical doctors. In recent years, with the rapid development of artificial intelligence, this technology has also gained its popularity in the field of medical imaging, while deep learning is the most used in the field of artificial intelligence and automatic learning based on large datasets of medical images

with the means of introducing a convolutional neural network (CNN) to extract image features.

In this paper, to attain accurate segmentation of caries lesions, we propose a new deep learning network called CariesNet. Inspired by the structure of U-Net [5], we build an U-shape neural network for oral panoramic image segmentation. In particular, we use the full-scale axial attention module as well as the partial encoder module to enhance the segmentation performance. To sum up, the main contributions of this work are threefold. (1) We propose a novel deep architecture CariesNet for segmenting dental caries lesions in panoramic radiograph. (2) We propose full-scale axial attention (FSSA) module to enhance the robustness for the segmentation of small size lesions. (3) The proposed CariesNet achieves an average Dice similarity coefficient (DSC) of 93.64%, and it shows effective results on the collected dataset.

The rest of this paper is organized as follows. Section 2 briefly reviews the related work. Our proposed CariesNet method is described in Sect. 3. Section 4 reports the experimental results. Finally, Section 5 concludes the work.

Fig. 1 Example of different grade caries lesions from panoramic radiograph: shallow caries in Box (a), moderate caries in Box (b) and deep caries in Box (c)



2 Related works

2.1 Computer-aided diagnosis methods for dental caries

The computer system can be used to quantify the changes of gray value in the image and realize clinical diagnoses. In recent years, deep learning has also been applied to identify and diagnose dental caries. In 2016, Anias et al. extracted 48 regions of interest from oral panoramic X-ray images by threshold segmentation and utilize a neural network combined with BP backpropagation to diagnose dental caries [6]. Ali et al. use three stacked sparse automatic encoders to extract the characteristics of apices and apply the Soft-max classifier to determine whether the teeth had caries [7]. In 2017, a linear adaptive particle swarm optimization (LA-PSO) algorithm is introduced to generate l-rate for 120 panoramic images of decayed teeth, and the classification performance of the proposed LA-PSO is evaluated by backpropagation neural model [8]. Prajapati et al. introduce migration techniques to construct a convolutional neural network-based dental caries diagnostic model, using VGG-16 to detect caries in 251 X-ray images [9]; Zhang et al. construct a computer-aided assessment system based on CBCT images to improve the accuracy of caries diagnosis [10]. In 2020, Lin et al. construct a computer-aided dental caries diagnosis system based on the depth-learning model, which shows that deep learning has good performance in detecting dental caries in root-tip X-ray images, to detect the adjacent surface caries of permanent teeth in apical X-ray images and to provide a reference for the early diagnosis of the adjacent surface caries [11]. Haghanifar et al. collect 480 oral panoramic X-ray images and propose a teeth segmentation and caries detection workflow to achieve a 90.52% caries detection accuracy [12]. However, collecting the high-quality caries dataset and building a highly efficient deep learning architecture still remain huge challenges.

2.2 Deep learning methods for image segmentation

In dentistry, many methods have been proposed for computer-assisted image segmentation (see [13, 14] for comprehensive reviews of such methods). Similar to the processing in the natural image, deep learning has been widely applied in computer vision tasks such as image classification and object detection [15]. Recently, an increasing number of deep learning-based methods were developed for image segmentation. One typical method is fully connected networks (FCNs) can perform end-to-end segmentation and be effective in diverse imaging

applications (e.g., semantic segmentation [16, 17], video object detection [18, 19], multi-modality classification [20]). However, due to its fully connected structure, FCN uses plenty of parameters, thus incurring obstacles in model training. SegNet [21] was presented with an encoder–decoder architecture for accelerating the training process. Based on FCNs and SegNet, an improved network, U-Net [5], employed an encoder–decoder architecture and used skip connections between the upsampling and downsampling layers to combine high-resolution features with the upsampled output. Some variants of U-Net have also been proposed to enhance performance, such as 3D U-Net [22], V-Net [23], UNet++ [24], SE-ResUnet [25] and attention U-Net [26]. In particular, Fan et al. [27] proposed the efficient network PraNet to balance the inference speed and segmentation performance.

Besides the known general image segmentation frameworks mentioned above, some dedicated deep learning models have also been developed. Specifically, for segmenting X-ray images, multiple deep learning-based models have been devised. Al-Antari et al. used the DeepLab directly to segment live [28]. Blain et al. proposed a modified U-Net network to detect COVID-19 infections from chest X-ray image [29]. Moeskops et al. utilized different image modalities to train a multi-task segmentation model [30]. Trullo et al. introduced the structure of a conditional random field module as RNN into FCN [31]. Moreover, deep learning methods have been significantly successful in other medical image segmentation tasks, such as segmentation of cells [32], head and neck (HaN) [33], liver [34], brain [35] and optic disk [36].

3 Materials and methods

3.1 Overview

In this section, we demonstrate the workflow of the proposed CariesNet. We first explain the collection of a comprehensive oral panoramic X-ray image dataset. Next we introduce the CariesNet architecture as well as the full-scale axial attention (FSSA) module. Finally we explain the loss function and the model training details.

3.2 Dataset preparation

Most related studies in the field of dental problem detection using X-rays lack a sufficient number of images in their datasets. Large datasets let the models have more sophisticated architectures, including more parameters. Hence, developed models can handle more complicated features and detect subtle abnormalities that appeared in the tooth texture, like dental caries in the early stages. Annotation is

an essential but time-consuming part that needs to be performed by the field specialists, e.g., dentists or radiologists.

To address the issues of data lacking, we try to build a high-quality oral panoramic dataset. A set of 1159 panoramic images originating from dental treatment and routine care are collected by the Affiliated Stomatology Hospital, Zhejiang University School of Medicine, from 2015 to 2020. Data collection was ethically approved by the Chinese Stomatological Association ethics committee. Only panoramic images of permanent teeth were included in the dataset. Panoramic images of primary teeth or those where any assessment is deemed impossible were excluded. Most of the data were generated using radiographic machines from the manufacturer Dentsply Sirona (Bensheim, Germany), mainly Orthophos XG. On all panoramic images, each tooth was segmented and labeled using the FDI scheme by three dentist and checked by a fourth dentists. From 1159 oral panoramic images, 3217 caries regions are labeled as shallow caries, moderate caries or deep caries. The detail of our caries dataset is shown in Table 1.

3.3 CariesNet overall architecture

Generally, the size of the oral panoramic radiograph is large, while the target caries region is small. It is a challenge to find and delineate the overall architecture of the network as shown in Fig. 2. We design CariesNet inspired by the overall architecture from the PraNet [27], which is based on reverse attention mechanism [37]. As is shown in Fig. 2, CariesNet is a general U-shape encoder–decoder framework, which can aggregate the features extracted from multi-level convolution networks. Traditional U-Net simply passes the feature to each decoder layer, and some high-level contextual information may lose in the decoder. Similar as introduced in [27], we use the partial decoder to aggregate more high-level features in CariesNet. In this paper, we utilize Res2Net [38] as an efficient backbone. We concatenate three high-level feature maps in backbone to the partial decoder, and it predicts the initial saliency map for dental caries, which is labeled as a global map in Fig. 2. Then both the backbone feature and partial decoder

feature are concatenated to the attention module. In CariesNet, we replace the reverse attention (RA) module with the full-scale axial attention (FSAA) module, and the detail of FSAA is described in Sect. 3.4. Next, the feature map is passed through a 1×1 convolution layer and added with the previous FSAA global map. Besides, in each high-level layer, the feature map obtained from FSAA in the previous layer and the feature map from the backbone is concatenated as the input of FSAA as well. We use three consequent FSAAs to compute the high-level saliency map. In the end, A 4 times bi-linear upsampling transformation with a sigmoid function is used to obtain the final output from the global feature map.

The CariesNet network is efficient to segment slight dental caries regions from oral panoramic X-ray images. By aggregating the features in three high-level layers in the partial decoder, the contextual information can be effectively extracted from the global map, which means the target dental caries lesions can be placed at the initial guidance area (global map). The full-scale axial attention module can further mine the boundary cues of the output segmentation result. To sum up, the overall architecture shows that the Res2Net backbone features are forwarded to the partial decoder to generate the initial global map, and the full-scale axial attention module can reconstruct accurate dental caries segmentation results.

3.4 Full-scale axial attention module

Generally, the delineation of the target dental caries lesion includes two steps for experienced doctors. First, a coarse region that may contain a target lesion is located. And the second step is to annotate the accurate boundary of the target area. Since the rough saliency map is obtained from the partial decoder, we propose the FSSA module that can mine the boundary cues. The above-mentioned module can extract fine-grained feature maps which have both high-level semantic information and low-level detail information.

As is shown in Fig. 3, the input high-level backbone feature map and the upsampled location map are concatenated firstly. Different from the normal axial attention module, in order to enable the module to integrate more

Table 1 Dental caries dataset description

	Training set	Valid set	Test set	Overall
Oral panoramic radiograph	900	135	124	1159
Shallow caries	709	125	112	946
Moderate caries	669	141	124	934
Deep caries	992	177	168	1337
Positive region	2370	443	404	3217

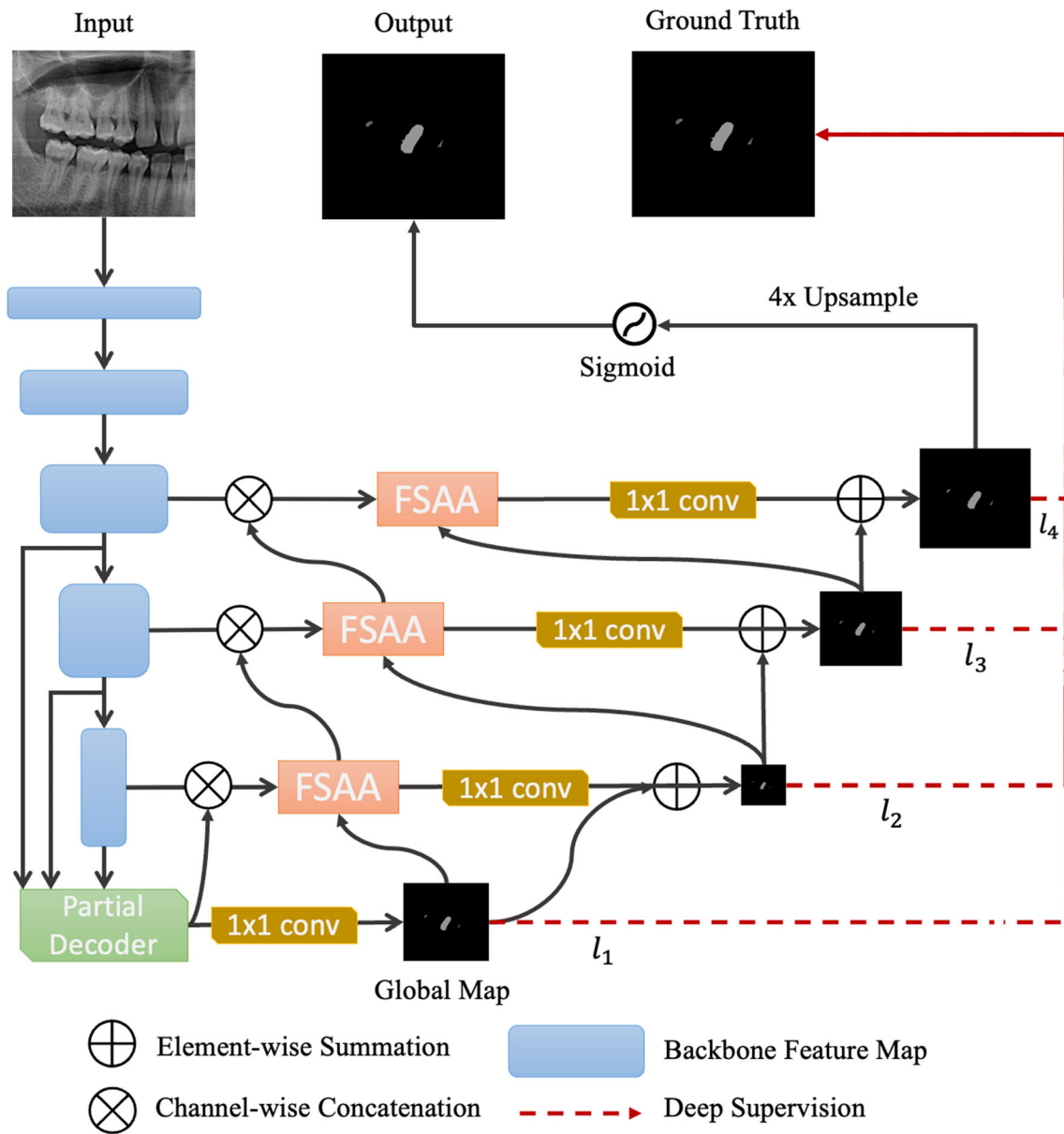


Fig. 2 Schematic diagram of the architecture of CariesNet, which consists of three full-scale axial attention modules with a partial decoder

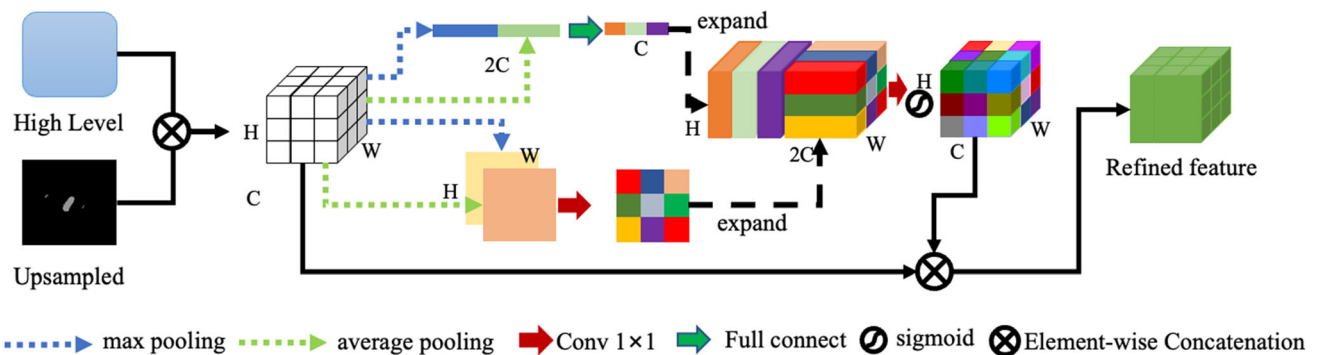


Fig. 3 Details of full-scale axial attention (FSAA) module used in CariesNet

layers of characteristic information, we consider average pooling and maximum pooling at the same time. The extracted channel domain features are mapped to the same dimension with the number of channels as the original feature image again through the full connection layer, while the spatial domain features are mapped through the convolution layer of element-wise convolution kernel to obtain the single-channel feature with the same size as the spatial feature. We parallel extract the attention features from the channel domain and the spatial domain and then allow the network to aggregate both of them through the element-wise convolution layer. In order to get a smoother attention feature map, we utilize a sigmoid layer after the fusion layer. FSAA eventually outputs an attention feature map that represents the contextual information from a global view.

3.5 Learning process and implementation details

Loss Function The binary cross-entropy (BCE) is usually employed as the loss function, which can be formulated as follows:

$$L_{BCE} = -\frac{1}{f} \sum_{j=1}^k n_j \log m_j + (1 - n_j) \log(1 - m_j) \quad (1)$$

where f is the number of pixels, and m_j and n_j , respectively, show the predicted value and its corresponding groundtruth value. However, the resulting inefficient optimization requires the adaptive loss function due to the high susceptibility of the cross-entropy loss function to class imbalance. Therefore, Dice loss is used as the loss function in our model as follows:

$$L_{dice} = 1 - \frac{\sum_{j=1}^k m_j n_j + \delta}{\sum_{j=1}^k (m_j + n_j + \delta)} - \frac{\sum_{j=1}^k m_j n_j + \delta}{\sum_{j=1}^k (2 - m_j - n_j + \delta)} \quad (2)$$

where m_j is the predicted value, and n_j is the corresponding ground-truth value.

We combine the BCE loss and Dice loss in CariesNet, and the final loss function can be expressed as:

$$L = \sum_{i=1}^4 (L_{BCE}^i + L_{Dice}^i) \quad (3)$$

Implementation Details We train CariesNet with 200 epochs on all the 512×512 size oral panoramic images in the training data. We use Adam as the optimizer with the initial learning rate of $1e-4$, which decays at 80, 120 and 150 epochs. Our experiments are performed on a workstation platform with Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz, 256GB RAM and 8x NVIDIA GeForce 2080Ti

GPU with 12GB GPU memory. The code is implemented with PyTorch 1.3.1 in Ubuntu 18.04.

4 Experiments and discussion

4.1 Evaluation metrics

Some evaluation metrics, including the Dice coefficient, accuracy, precision and recall, are adopted to compare the performances of CariesNet and other methods to compare the performances of different methods. The Dice coefficient measures the overlapping pixels between the automatic and manual segmentation of dental caries, which is calculated as follows:

$$Dice = \frac{2 \times TP}{2 \times TP + FP + FN} \quad (4)$$

where TP, FP, TN and FN represent true-positive, false-positive, true-negative and false-negative prediction, respectively. Accuracy is the entire accuracy of the dental caries types and background segmentation, which is described as the following:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

Precision is the proportion of dental caries area that are classified as true-positive areas concerning all pixels of caries lesions that are classified by automatic segmentation, which is delimited as follows:

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

The recall represents the proportion of the true-positive pixels of in dental caries that are classified by automatic segmentation versus the pixels of caries lesions that are classified by manual segmentation, which is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

F1 score is used to quantify the weighted average of dental caries lesions between the precision and recall rate, with a value in $[0, 1]$, and is calculated as follows:

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

4.2 Comparative experiments

The results obtained from the caries dataset are reported in Table 2. In each test case, we split the oral panoramic image two parts, the left and the right. In regard of the test

Table 2 Comparison of OARs segmentation results with different methods

Algorithm	DSC(%)	Accuracy	F1 score	Precision	Recall
U-Net [5]	80.87	80.71	88.92	86.89	70.31
DeepLabV3+ [39]	81.90	81.68	87.43	89.76	71.72
Res-UNet [40]	81.24	81.15	88.93	90.16	72.80
Attention-UNet [26]	90.91	90.38	91.66	90.42	83.13
PraNet [27]	93.52	93.24	92.35	93.85	85.28
CariesNet	93.64	93.61	92.87	94.09	86.01

Bold values indicate the best results

result evaluation, joint segmentation results of the two parts are merged. DeepLab is a widely used pixel-wise segmentation tool [39], which also uses an encoder–decoder structure. Here, we use U-Net and DeepLabV3+ as baseline models. We use Res2Net as a backbone in Res-UNet [40], which is implemented for the ablation experiments as a backbone method as well. All the deep learning models are tested on the same validation set, and the results of DSC, accuracy, F1 score, precision and recall are shown in Table 2, respectively. It is obvious that the PraNet and CariesNet model performs well in localizing the target lesions with partial decoder module, and CariesNet improves the overall performance with a large margin. The normal U-Net and DeepLabv3 perform similar to the backbone methods Res-UNet. Attention-UNet has a much better segmentation results, and it proves the attention mechanism can significantly improve the model performance. Meanwhile, CariesNet outperforms the state-of-the-art method PraNet because the full-scale axial attention module can capture wide and efficient contextual information. CariesNet achieves a DSC of 93.64% eventually.

4.3 Ablation study

Apart from the above comparison with the state-of-the-art methods, we also conduct extensive ablation experiments to validate the effectiveness of our method, including the partial encoder module, full-scale axial attention module, BCE/Dice loss function and deep supervision strategy. As shown in Table 3, the Dice coefficient of each model is reported when segmenting three types of dental caries. It is clear that FSAA can significantly improve the model performance. Besides, it is noticed that the performance on the

Table 3 Ablation study of the CariesNet segmentation performance (DSC) on three dental caries types

Model	Deep caries	Moderate caries	Shallow caries	Average
Backbone	0.860	0.653	0.765	0.812
+ Partial encoder	0.892	0.642	0.801	0.870
+ FSAA	0.947	0.673	0.916	0.919
+ BCE/Dice loss	0.955	0.687	0.924	0.921
+ Deep supervision	0.971	0.694	0.939	0.936

Bold values indicate the best results

moderate caries delineation is relatively low, because either the boundaries between deep caries and moderate caries or the boundaries between shallow caries and moderate caries are relatively blurred, and the models tend to misclassify moderate caries as shallow caries or deep caries. Although CariesNet shows limited performance on moderate caries over the backbone, it helps improve about 11.1% DSC on deep caries and 12.4% DSC on shallow caries.

4.4 Results visualization

Figure 2 shows the segmentation results. CariesNet can kindly find the small dental caries lesions from oral panoramic radiographs. We mark the deep dental caries lesions as yellow part in Fig. 4. The moderate caries and the shallow caries area are marked as blue and green parts, respectively. To compare the performance between the methods clearly, we select a part to enlarge the display. The segmentation results of CariesNet, PraNet, U-Net, DeepLabv3 and Res-U-Net are shown in Fig. 4. Compared with other methods, CariesNet has a smoother and more accurate boundary.

5 Conclusion

In conclusion, we developed an automated system for caries diagnosis. Experiments demonstrate that the deep learning model can effectively segment the dental caries lesions from the oral panoramic X-ray image. In particular, we developed a state-of-the-art segmentation network CariesNet, implementing the partial encoder module and

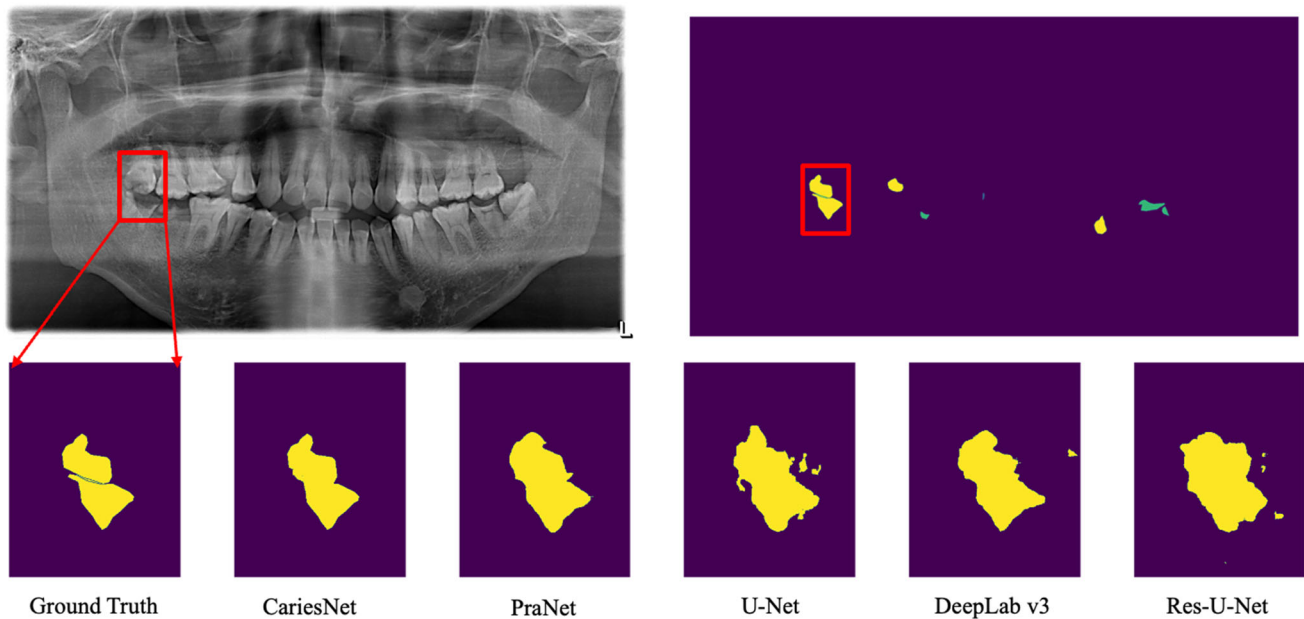


Fig. 4 The visualization of segmentation results from CariesNet, PraNet, U-Net, DeepLabv3 and Res-U-net. Deep caries, moderate caries and shallow caries masks are labeled as yellow, blue and green, respectively

the full-scale axial attention module into the common encoder–decoder U-shape structure. We conducted experiments on the dataset and the validation and test studies showed the capability of our new approach for this segmentation task. Comparison and ablation experiments also suggested that our new CariesNet architecture yields very good performance in segmenting slight lesions from large X-ray images.

Acknowledgements The research was supported by Natural Science Foundation of Zhejiang Province under grant LZY21F030002. It was also partially supported by the National Key R&D Program of China under Grant No. 2019YFB1404802, the Zhejiang University Education Foundation under Grants No.K18-511120-004, No.K17-511120-017 and No.K17-518051-02, the Zhejiang Public Welfare Technology Research Project under Grant No.LGF20F020013, the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2019R01007), the Wenzhou Bureau of Science and Technology of China (No. Y2020082) and the Key Laboratory of Medical Neurobiology of Zhejiang.

Declaration

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Krithiga R (2016) A survey: segmentation in dental X-ray images for diagnosis of dental caries. *Int J Control Theory Appl* 9(40):941
2. Lu HX, Tao DY, Lo ECM, Li R, Wang X, Tai BJ, Hu Y, Lin HC, Wang B, Si Y et al (2018) The 4th national oral health survey in

- the mainland of China: background and methodology. *Chin J Dent Res* 21(3):161
3. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A et al (2018) Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories. *The Lancet* 392(10159):1789
4. Lian L, Zhu T, Zhu F, Zhu H (2021) Deep learning for caries detection and classification. *Diagnostics* 11(9):1672
5. Ronneberger O, Fischer P, Brox T (2015) Squeeze-and-excitation networks, in *International Conference on Medical Image Computing and Computer-Assisted Intervention (Springer)*, pp. 234–241
6. ALbahbah AA, El-Bakry HM, Abd-Elgahany S (2016) Detection of caries in panoramic dental X-ray images using back-propagation neural network. *Int J Electron Commun Comput Eng* 7(5):250
7. Ali R.B, Ejbali R, Zaied M (2016) Detection and classification of dental caries in x-ray images using deep neural networks, in *International Conference on Software Engineering Advances (ICSEA)*, p. 236
8. Sornam M, Prabhakaran M (2017) In 2017 IEEE Int Conf Power. Signals and Instrumentation Engineering (ICPCSI) (IEEE, Control, pp 2698–2703
9. Prajapati S.A, Nagaraj R, Mitra S (2017) Classification of dental diseases using CNN and transfer learning, in 2017 5th International Symposium on Computational and Business Intelligence (ISCBI) (IEEE), pp. 70–74
10. Zhiling Z (2017) Study on the sensitivity of computer-aided detection of adjacent caries in cone-beam CT images. *Chinese J Stomatol* 52(002):103
11. Xiujiao L, Dong Z, Minyi H, Hui C, Hao Y (2020) Study on the sensitivity of computer-aided detection of adjacent caries in cone-beam CT images. *Chinese J Stomatol* 55(09):654
12. Haghanifar A, Majdabadi M.M, Ko, S.B (2020) Paxnet: Dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier, arXiv preprint [arXiv:2012.13666](https://arxiv.org/abs/2012.13666)

13. Guo Y, Liu Y, Georgiou T, Lew MS (2018) A review of semantic segmentation using deep neural networks. *Int J Multimed Inform Retriev* 7(2):87
14. Corbella S, Srinivas S, Cabitza F (2020) Applications of deep learning in dentistry, *Oral Surgery, Oral Med, Oral Pathol Oral Radiol*
15. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks, *Adv Neural Inform Process Syst* pp. 1097–1105
16. Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 3431–3440
17. Li Y, Qi H, Dai J, Ji X, Wei Y (2017) Fully convolutional instance-aware semantic segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 2359–2367
18. Wang W, Shen J, Shao L (2017) Video salient object detection via fully convolutional networks. *IEEE Trans Image Process* 27(1):38
19. Caelles S, Maninis KK, Pont-Tuset J, Leal-Taixé L, Cremers D, Van Gool L (2017) One-shot video object segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* pp. 221–230
20. Cao Z, Sun C, Wang W, Zheng X, Wu J, Gao H (2021) Multi-modality fusion learning for the automatic diagnosis of optic neuropathy. *Patt Recognit Lett* 142:58
21. Badrinarayanan V, Kendall A, Cipolla R (2017) DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Patt Anal Mach Intell* 39(12):2481
22. Çiçek Ö, Abdulkadir A, Lienkamp S.S, Brox T, Ronneberger O (2016) 3D U-Net: Learning dense volumetric segmentation from sparse annotation. *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), pp. 424–432
23. Milletari F, Navab N, Ahmadi S.A (2016) V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)* (IEEE), pp. 565–571
24. Zhou Z, Siddiquee M.M.R, Tajbakhsh N, Liang J, (2018) UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* (Springer), pp. 3–11
25. Cao Z, Yu B, Lei B, Ying H, Zhang X, Chen DZ, Wu J (2021) Cascaded SE-ResUnet for segmentation of thoracic organs at risk. *Neurocomputing* 453:357
26. Oktay O, Schlemper J, Folgoc L.L, Lee M, Heinrich M, Misawa K, Mori K, McDonagh S, Hammerla N.Y, Kainz B, et al. (2018) Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*
27. Fan D.P, Ji G.P, Zhou T, Chen G, Fu H, Shen J, Shao L, (2020) Prant: Parallel reverse attention network for polyp segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), pp. 263–273
28. Al-Antari MA, Al-Masni MA, Choi MT, Han SM, Kim TS (2018) A fully integrated computer-aided diagnosis system for digital X-ray mammograms via deep learning detection, segmentation, and classification. *Int J Medi Inform* 117:44
29. Blain M, Kassin MT, Varble N, Wang X, Xu Z, Xu D, Carrafiello G, Vespro V, Stellato E, Ierardi AM et al (2021) Determination of disease severity in COVID-19 patients using deep learning in chest X-ray images. *Diagnost Intervent Radiol* 27(1):20
30. Moeskops P, Wolterink J.M, van der Velden B.H, Gilhuijs K.G, Leiner T, Viergever M.A, Išgum I (2016) Deep learning for multi-task medical image segmentation in multiple modalities. In *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer), pp. 478–486
31. Trullo R, Petitjean C, Ruan S, Dubray B, Nie D, Shen D, (2017) Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (IEEE), pp. 1003–1006
32. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K et al (2019) U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods* 16(1):67
33. Zhu W, Huang Y, Zeng L, Chen X, Liu Y, Qian Z, Du N, Fan W, Xie X (2019) AnatomyNet: Deep learning for fast and fully automated whole-volume segmentation of head and neck anatomy. *Med Phys* 46(2):576
34. Song L, Geoffrey K, Kaijian H (2020) Bottleneck feature supervised U-Net for pixel-wise liver and tumor segmentation. *Exp Syst Appl* 145:113131
35. Mehta R, Sivaswamy J, (2017) M-Net: A convolutional neural network for deep brain structure segmentation. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)* (IEEE), pp. 437–440
36. Sevastopolsky A (2017) Optic disc and cup segmentation methods for glaucoma detection with modification of U-Net convolutional neural network. *Patt Recognit Image Anal* 27(3):618
37. Chen S, Tan X, Wang B, Hu X (2018) Reverse attention for salient object detection. *Proceedings of the European Conference on Computer Vision (ECCV)* pp. 234–250
38. Gao S, Cheng M.M, Zhao K, Zhang X.Y, Yang M.H, Torr P.H (2019) Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*
39. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2017) DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Patt Anal Mach Intell* 40(4):834
40. He K, Zhang X, Ren S, Sun J (2016) U-Net: Convolutional networks for biomedical image segmentation. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 770–778

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.