Heliyon 10 (2024) e25892

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Semantic segmentation for tooth cracks using improved DeepLabv3+ model

Zewen Xie [a,b], Qilin Lu [a], Juncheng Guo [a], Weiren Lin [a], Guanghua Ge [c], Yadong Tang [d], Damiano Pasini [e], Wenlong Wang [a,e,*]

[a] School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, 510006, China
[b] School of Physics and Material Science, Guangzhou University, Guangzhou, 510006, China
[c] Department of Dentistry, Hospital of Guangdong University of Technology, Guangdong University of Technology, Guangzhou, 510006, China
[d] School of Biomedical and Pharmaceutical Sciences, Guangdong University of Technology, Guangzhou, 510006, China
[e] Department of Mechanical Engineering, McGill University, 817 Sherbrooke Street West, Montreal, QC H3A 0C3, Canada

ARTICLE INFO

Keywords:
Cracked teeth
Oral health
Semantic segmentation
DeepLabv3+

ABSTRACT

Objective: Accurate and prompt detection of cracked teeth plays a critical role for human oral health. The aim of this paper is to evaluate the performance of a tooth crack segmentation model (namely, FDB-DeepLabv3+) on optical microscopic images.
Method: The FDB-DeepLabv3+ model proposed here improves feature learning by replacing the backbone with ResNet50. Feature pyramid network (FPN) is introduced to fuse muti-level features. Densely linked atrous spatial pyramid pooling (Dense ASPP) is applied to achieve denser pixel sampling and wider receptive field. Bottleneck attention module (BAM) is embedded to enhance local feature extraction.
Results: Through testing on a self-made hidden cracked tooth dataset, the proposed method outperforms four classical networks (FCN, U-Net, SegNet, DeepLabv3+) on segmentation results in terms of mean pixel accuracy (MPA) and mean intersection over union (MIoU). The network achieves an increase of 11.41% in MPA and 12.14% in MIoU compared to DeepLabv3+. Ablation experiments shows that all the modifications are beneficial.
Conclusion: An improved network is designed for segmenting tooth surface cracks with good overall performance and robustness, which may hold significant potential in computer-aided diagnosis of cracked teeth.

1. Introduction

Oral diseases are among the most prevalent global health issues, recognized as a significant global public health challenge by numerous scholars [1]. These conditions place a substantial burden on both healthcare and the economy, underscoring the critical importance of maintaining oral health. Toothache, one of the most prevalent oral health issues, can result in pain, eating disturbances, and sleep disorders [2]. Cracks in teeth are one of the causes of toothaches, and the American Association of Endodontists (AAE) has categorized these cracks into five types: craze lines, fractured cusps, cracked teeth, split teeth, and vertical root fractures (VRF) [3]. Moreover, epidemiological studies have singled out cracked tooth syndrome (CTS) as one of the primary factor contributing to tooth

* Corresponding author. School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou, 510006, China
E-mail address: wlwang@gzhu.edu.cn (W. Wang).

https://doi.org/10.1016/j.heliyon.2024.e25892
Received 21 January 2024; Received in revised form 1 February 2024; Accepted 5 February 2024
Available online 10 February 2024

loss [4]. However, the identification of cracks in hard tissue, particularly within the oral cavity, represents a notable challenge at the intersection of dentistry and engineering [5] (Fig. S1 shows three challenging cases). Due to the often-vague early symptoms of cracked teeth, it is often difficult for dentists to detect cracks in the teeth [6]. If the symptom of the cracked tooth could not be detected and treated in time, the crack in the teeth would gradually propagate to the dentin layer with the use of the diseased tooth, which would lead to pulpitis, or even complete tooth loss [7]. An accurate diagnosis of cracked teeth is imperative to prevent the worsening of symptoms and tooth loss, which holds great significance for human oral health.

Over the past few decades, numerous techniques have been developed by researchers to detect cracks in human teeth, which can be broadly categorized into the following four groups: traditional methods based on patient clinical symptoms [8–11], cone-beam computed tomography (CBCT)-based methods [12–15], inspection method based on microscopy imaging [16–19], and deep learning-based methods [20–24]. Although these methods are commonly used diagnostic methods, they face three main challenges: (1) the diagnostic process may exacerbate the patient's symptoms and cause discomfort, (2) the accuracy of the diagnosis is dependent on the dentist's work status and work experience, (3) the robustness is low and susceptible to noise. More details can be referred to previously written reviews [5,6,25]. Briefly speaking, as for traditional methods, the main methods included cold stimulation [8], probing method [9], occlusion test method [10], and light *trans*-illumination method [11]. However, most of them can be painful and even aggravating for the patient at the time of diagnosis. For cone-beam computed tomography-based methods, they have been widely used to detect cracks in teeth. Nevertheless, the utilization of CBCT for the detection of cracked teeth is challenged by the presence of metal artifacts, overlapping vignettes, and noise in the obtained images, making it difficult to identify the presence of occult cracks, which could easily be overlooked by practitioners [26]. Furthermore, CBCT scans are relatively expensive and labor-intensive. For inspection method based on microscopy imaging, dentists magnify each tooth 14 to 18 times with a microscope to detect cracks and assess pathological risk levels. Although it is low-cost and does not cause radiation damage, this method often requires the assistance of methylene blue dye [17], which makes the procedure more complex. Additionally, sometimes to achieve better staining, coarse grinding and polishing with coarse pumice slurry are required [19], which could worsen the patient's condition. Moreover, some tiny cracks that are highly concealed can be easily misdiagnosed when teeth are affected by cavities, plaque, and other factors. Traditional diagnosis is also dependent on the professionalism and experience of radiologist and dentist [27]. In addition, prolonged and intensive work with the eyes can cause visual fatigue in dentists, which increases the probability of missed diagnosis and misdiagnosis.

In recent years, deep learning-based methods have garnered increasing attention in the realm of damage monitoring [28,29] and computer-aided diagnosis [30]. Transformer-based model [28] and YOLO model [29] were proposed to detect tunnel lining cracks with acceptable detection accuracy. As for medical diagnosis, deep learning has been employed for auxiliary diagnosis of various biological tissues (liver, thyroid, blood), achieving a high level of diagnostic accuracy and efficiency through convolutional neural network (CNN) such as U-Net [31,32] and DeepLabv3+ [33]. Unlike common segmentation targets, tooth cracks are usually small targets that are often accompanied by interference from other conditions such as dental plaque, dental caries, and fluoride poisoning. Furthermore, tooth cracks can vary from person to person, and imaging process (i.e. CBCT) may also generate noise, making the cracks indistinguishable from their surroundings. Therefore, the development of methods for detection of cracks in human teeth remains a challenging area.

Certain scholars have already employed deep learning methods to detect tooth cracks. For example, Hu et al. [20] utilized three deep learning networks, namely ResNet50, VGG19, and DenseNet169, to assist in diagnosing vertical root fractures based on CBCT images. Jared Vicory et al. [21] proposed a tooth microcrack detection algorithm that combined wavelet features and machine learning, which achieved better segmentation of tooth cracks on CBCT images. Sahu et al. [22] developed a model that combined signal enhancement and a CNN-based deep crack detection model to accurately detect tooth cracks using high-resolution CBCT scans. Dumbryte et al. [23] combined X-ray imaging and CNN for non-destructive and qualitative analysis of tooth microcracks. It is worth noting that X-ray-based techniques, although powerful, the presence of unwelcome factors such as metal artifacts, overlapping vignettes, and image noise in CBCT imagery have been observed to exert a detrimental effect [26]. Nevertheless, there is less attention given to the application of deep learning techniques for the detection of cracks in tooth in optical images. Guo et al. [24] introduced a novel approach for automatic tooth crack detection based on optical images, utilizing a combination of the sliding window algorithm and convolutional neural network. However, methods based on binary classification lack more detailed information about cracks, and dentists need to further locate the position and contour of cracks during detection. Hence, research on quantitative crack analysis with optical imaging is meaningful and necessary.

This paper introduces a model, namely FDB-DeepLabv3+ that targets tooth crack segmentation. It is capable of segmenting cracks on tooth surfaces on oral microscopic images for a variety of challenging situations. This model may further provide a substantial approach to assist doctors in diagnosing cracked teeth. The primary contribution of this paper can be summarized as follows:

(1) Extensive and systematic research is conducted, to explore the influence of different backbones (Xception, MobileNetV2, Inception-ResNetV2, ResNet18, ResNet50, ResNet101) on DeepLabv3+ with thorough qualitative and quantitative analyses, resulting in the selection of the optimal solution.
(2) Based on the selection of the backbone, three additional modules (FPN, Dense ASPP, BAM) are integrated to form the proposed FDB-DeepLabv3+ model, which significantly improves MPA by 11.41% and MIoU by 12.14% on self-made dataset.
(3) The application of semantic segmentation technology on optical images of human tooth cracks is explored for the first time in this paper.

The manuscript is structured as follows. Section 2 offers the related works. Section 3 presents the common methods and the proposed method. Section 4 outlines the experimental procedure. Section 5 gives the discussion. Section 6 is the conclusion.

2. Related works

With the development of computer vision technology, semantic segmentation methods have been widely used in medical image processing. In recent years, some scholars have proposed fracture detection models tailored to specific tasks by considering the unique features of fractures and background characteristics in different application scenarios, such as wrist cracks and bone fracture [34–37]. As illustrated in Fig. 1, commonly used semantic segmentation methods include FCN, U-Net, SegNet, and DeepLabv3+.

2.1. FCN

FCN [38] is composed of two main components, down-sampling and up-sampling convolution operations, and utilizes VGG16 as its backbone (as shown in Fig. 1 (b)). It replaces the fully-connected layer in the convolutional neural network with a fully-convolutional layer so that the network can accept input images of arbitrary size. By incorporating the skip structure, the feature map generated by the last convolutional layer is fused with the output of the pooling layer. Ultimately, the model produces a feature map with the identical size as the input image, enabling better segmentation results by predicting each pixel of the original image.

2.2. U-Net

U-Net is a fully convolutional neural network model improved by Ronneberger et al. [39], which is named after its U-shaped architecture and is frequently used in biomedical imaging (as shown in Fig. 1 (a)). This network employs the skip connection method to concatenate feature maps from the encoder and decoder at the same level in the channel dimension. This approach enables the fusion of deep network features with contextual information from shallow network images, facilitating multi-scale feature fusion. By fusing features through the concatenation of channels, the U-Net network reduces the loss of image information and improves image segmentation performance.

2.3. SegNet

SegNet is a deep convolutional neural network model that was developed by Badrinarayanan et al. [40] It employs a symmetric encoder-decoder architecture that effectively preserved deep and abstract details (as shown in Fig. 1 (c)). The encoder is a network model that continues to use VGG16 and is primarily used for analyzing object information. In the decoder, feature information is mapped back to the original image. By reusing the maximum pooling indices during the up-sampling process, the network reduces the detail loss caused by pooling layers. The output of the decoder is finally sent to the softmax layer, generating the category probability of each pixel.

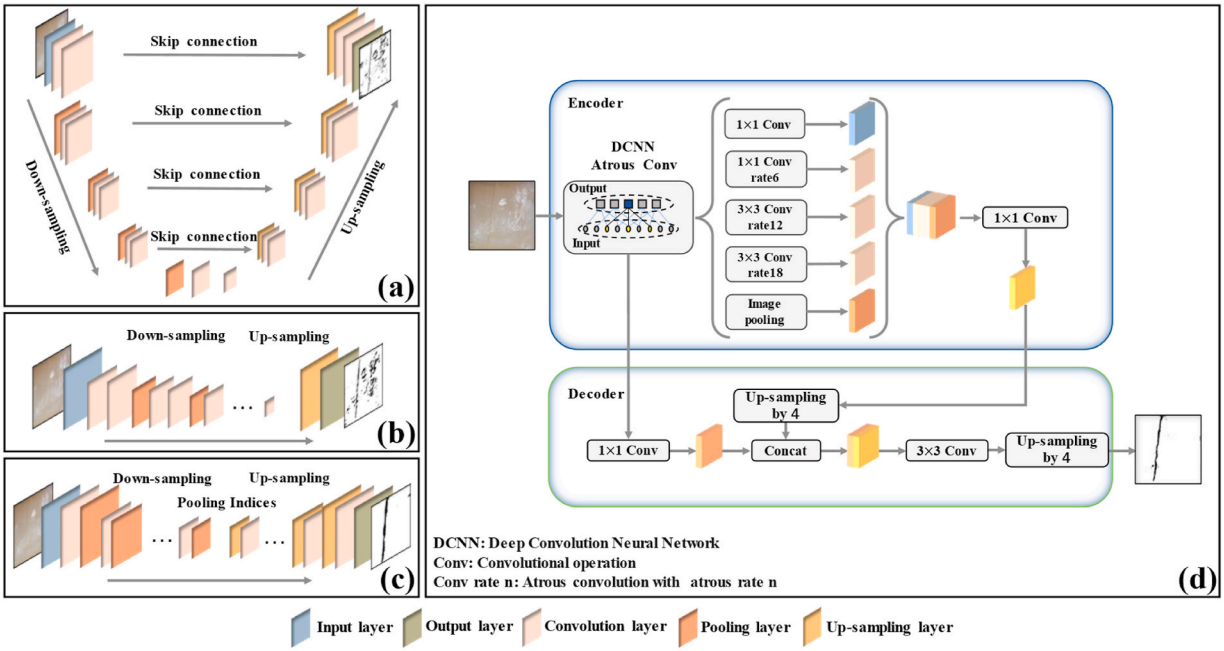

Fig. 1. The four commonly used semantic segmentation network architectures, including (a)U-Net, (b) FCN, (c) SegNet, (d) DeepLabv3+.

2.4. DeepLabv3+

Chen et al. [41] proposed the DeepLabv3+ network model to address issues such as spatial insensitivity, resolution reduction caused by down-sampling, and multiscale images in deep convolutional neural networks (as shown in Fig. 1 (d)). The network adopts an encoder-decoder structure based on the Xception backbone [42], which includes atrous convolution and spatial pyramid pooling (SPP) modules. The atrous spatial pyramid pooling (ASPP) module in DeepLabv3+ is a variant based on SPP and atrous convolution, which can capture rich contextual information by feature pooling at different resolutions and further optimizes the deep features extracted by the encoding part. The decoding section up-samples the deep feature maps and fuses them with shallow features to optimize the positional information. Through this encoder-decoder architecture, clear object boundaries can be obtained, and segmentation accuracy can be improved.

3. Method

3.1. Architecture of the proposed FDB-DeepLabv3+

In this paper, an improved semantic segmentation model is proposed based on the DeepLabv3+ architecture to further enhance the segmentation accuracy of tooth cracks in various complex scenarios (Fig. 2(a) shows the network architecture).

Modifications to the network architecture include the following three aspects:

i. Replacing the backbone with ResNet50 and introducing feature pyramid network (FPN), which enables the fusion of feature maps from the bottom to the top, thereby obtaining richer semantic and spatial information, and improving the model's ability to detect small pixel points such as tooth cracks.
ii. Combining the densely linked atrous spatial pyramid pooling (Dense ASPP), so as to reduce the ignored pixel features, maintain the integrity of feature information, and achieve an expanded receptive field without augmenting the parameter count in the original ASPP module.
iii. Following stage 4 of the ResNet50 in DeepLabv3+, the bottleneck attention module (BAM) module is introduced, so that it can pay more attention to the local information of the tooth cracks feature and further refine the detailed feature of the tooth crack.

3.1.1. ResNet50

ResNet [43] is a breakthrough architecture that introduces residual connections to tackle the gradient vanishing problem in deep convolutional neural networks (CNNs). As shown in Fig. S2, the Residual Block contains residual connections, these connections allow information to bypass certain layers and flow directly from input to output, preserving the original features and enabling the learning of residual information. Among various types of ResNet, ResNet50 is chosen as the backbone of our model for several reasons. First, ResNet50 has shown superior performance in various computer vision tasks, such as image classification, object detection, and semantic segmentation. Its depth and design make it a powerful feature extractor for complex visual patterns. Second, the residual connections in ResNet50 alleviate the gradient vanishing problem, allowing for more effective training of deep neural networks. Third, while ensuring accuracy, ResNet50 achieves a balance between training speed and network performance, unlike ResNet101 and ResNet152, which require a substantial amount of computational resources due to their higher complexity [44]. By using ResNet50,

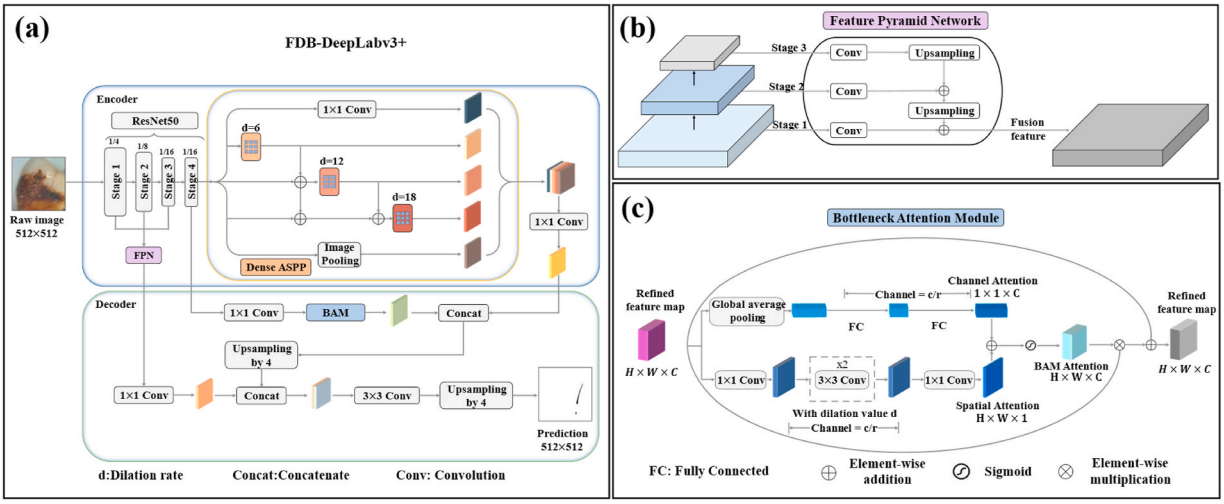

Fig. 2. Architecture of the proposed FDB-DeepLabv3+, which includes FPN, Dense ASPP, and BAM. (a) FDB-DeepLabv3+ (Dense ASPP is shown in detail), (b) FPN, (c) BAM.

the aim is to enhance our model's feature learning and overall performance in our application domain.

3.1.2. FPN

During the process of training ResNet50 to learn target features, the image resolution gradually decreases due to the consecutive convolutional operations, which may have resulted in recognition errors for objects with small pixel proportions in the image. FPN [45] is a method for fusing feature maps from different levels (as shown in Fig. 2(b)). In FPN, the feature maps produced by Stage 3 of the ResNet50 model are first adjusted to 256 channels through a 1×1 convolutional kernel and then up-sampled by a factor of 2 to create one branch of the FPN. The feature maps produced by Stage 2 are also adjusted to 256 channels through a 1×1 convolutional operation and use to create another branch of the FPN. These two branches are then fused by element-by-element summation, and the resulting feature map is up-sampled by a factor of 2 and fused with the 256-channel feature map produced by Stage 1. Ultimately, the fused feature map generated by FPN contains more comprehensive semantic and spatial details, thereby enhancing the capacity to identify minute pixel-based features such as tooth fractures.

3.2. Dense ASPP

In the ASPP module of the original DeepLabv3+ model, the parallel-connected atrous convolution exhibited discreteness in image space, which may result in the loss of feature information and cause discontinuity in the segmentation of small cracks. Replacing the ASPP module with the Dense ASPP module may better extract crack features. The advantages of the improved Dense ASPP includes two aspects. Firstly, it makes pixel sampling dense. Secondly, it may achieve a broader receptive field without increasing the number of parameters [46].

In the setting of the densely linked ASPP module described in this paper, the atrous convolution kernel size is set to 3, and the module consists of three atrous convolutional layers (as shown in Fig. 2(a)). The choice of a kernel size of 3 strikes a balance between computational efficiency and high performance. The same parameter selections, including the atrous rate in each branch, are also used in Refs. [41,46]. The DeepLabv3+ model uses ResNet50 as the primary feature extraction network, and the output channel number of ResNet50 to the densely linked ASPP module is 2048. The feature dimension of the output feature map of the densely linked ASPP module is set to 256. This setting reduces the dimensionality of the feature map to mitigate computational complexity while preserving critical information. Additionally, it facilitates information compression, enabling efficient representation of the extracted features. Moreover, this configuration helps maintain a balance between model capacity and available computational resources, preventing overfitting and enhancing generalization capability.

Table 1 shows the receptive fields obtained by original ASPP as well as using the dense link stack atrous convolution. In the original DeepLabv3+ model, the ASPP module works in parallel with no information sharing among the branches, and the feature maps are directly concatenated after passing through each atrous convolutional layer. As shown in Table 1, the receptive fields are 13, 25, and 37 when the atrous rates are 6, 12, and 18, respectively. However, the densely linked ASPP structure improves the reuse of feature information across different layers and correspondingly increases the receptive field. As shown, the receptive fields are 37 when the branches with atrous rates of 6 and 12 are connected, and 61 when the branches with atrous rates of 12 and 18 are connected. When the three branches with atrous rates of 6, 12 and 18 are connected, the receptive field is 73.

3.3. BAM

In the task of tooth crack segmentation, the model recognizes interference information such as the black shadow lines of the teeth outline and background as tooth cracks in the image, which ultimately affects the segmentation performance of the model. To address this issue, the attention mechanism module is introduced to make the deep learning model pay more attention to the local information where the tooth crack features are located. Both channel information and spatial information play a crucial role in refining the crack information. Therefore, in this paper, the BAM [44] is introduced in stage 4 of the ResNet50 backbone in the DeepLabv3+ model to extract both channel and spatial features from the image.

The BAM attention module consists of a channel branch and a spatial branch (as shown in Fig. 2(c)). For the input feature map, the BAM module is able to obtain a new feature map that is obtained by fusing the channel attention map and the spatial attention map [47]. The new feature map can be computed by Eq. (1):

$$F_1^{'} = F^c + F^s \tag{1}$$

Table 1
Receptive field of Densely Linked ASPP (kernel size = 3).

Type	Parameters (atrous rate)	Receptive field
ASPP	6	13
	12	25
	18	37
Dense ASPP	6,12	37
	12,18	61
	6,12,18	73

where $F_1^{'}$ represents the new feature map; F^c represents the channel attention map; F^s represents the spatial attention map.

The final feature map obtained through the BAM module can be calculated by Eq. (2):

$$F_2 = F_1 \times F_1^{'} + F_1 \tag{2}$$

where $F_1 \in R^{H\times W\times C}$ represents the input feature map and $F_2 \in R^{H\times W\times C}$ represents the final feature map.

The purpose of the channel attention branch in the BAM module is to selectively focus on the relationship between different channels, which makes the network pay more attention to the crack information. The purpose of the spatial attention branch is to emphasize the features of each spatial position so that the network can extract more spatial position information features related to the crack.

The channel attention branch of the BAM module consists of a global average pooling layer and two fully connected layers. It can aggregate features between each channel and generate a feature vector using Eq. (3).

$$F^c = FC(FC(GlobalAvgPool(F_1))) \tag{3}$$

where *FC* represents fully connected layers, F^c represents a $1 \times 1 \times C$ feature vector, and *GlobalAvgpool* represents global average pooling.

As contextual information is crucial to understanding spatial positions in the task of tooth crack segmentation, a larger receptive field is needed to effectively capture contextual information. Therefore, the spatial attention branch of the BAM module consists of two 1×1 convolution layers and two 3×3 dilated convolution layers, which can generate a feature map of size $H \times W \times 1$. In the experiments, the kernel size of the dilated convolution layers is set to 3×3 with an atrous rate of 4. The spatial attention feature map is defined in Eq. (4).

$$F^s = conv_{1\times1}(Dilated(Dilated(conv_{1\times1}(F_1)))) \tag{4}$$

where $conv_{1\times1}$ represents a 1×1 convolution layer, and *Dilated* represents a dilated convolution layer.

Following the two parallel attention branches, the dimensions of the attention maps are adjusted to the same size $H \times W \times C$ to ensure compatibility, then they are fused pixel-wise to obtain a three-dimensional feature map. After compressing the new feature map into the range of 0–1 via the sigmoid function, element-wise multiplication is performed with the input feature map. Finally, pixel-wise addition is carried out with the original input feature map, resulting in the generation of the final feature map.

3.4. Segmentation evaluation metrics

The task of tooth crack segmentation can be considered as a two-class classification problem, where the objective is to assign each pixel in the input image to either the tooth crack category or the background category. It can also be viewed as a positive and negative classification problem, where the tooth crack category is regarded as the positive category, and the background category is regarded as the negative category. By comparing the prediction results of the model with the ground truth of the dataset, four basic indicators can be obtained, namely True Positive (*TP*), False Positive (*FP*), True Negative (*TN*), and False Negative (*FN*). The True Positive represents the total number of positive pixels that are correctly predicted as positive pixels, while False Positive indicates the number of negative class pixels that are wrongly classified as positive class pixels. The True Negative denotes the number of pixels of the positive class that are incorrectly classified into the negative class, and the False Negative indicates the total number of pixels of the negative class that are correctly predicted. Using these four basic indicators, the other evaluation metrics can be calculated.

In this paper, image segmentation is evaluated using MPA and MIoU. MPA is a commonly used pixel-level accuracy metric function for evaluating the overall accuracy of a semantic segmentation model in terms of pixel classification. Its calculation formula is shown in Eq. (5).

$$MPA = \frac{1}{n}\sum_{i=1}^{n}\frac{TP+TN}{FN+FP+TP+TN} \tag{5}$$

MIoU is a mean cross-merge ratio metric function that is commonly used to evaluate the classification effectiveness of semantic segmentation models. It takes into account not only the classification accuracy of the model for each class of pixels but also the effect of pixel distribution and the number of pixels on the results, enabling a more comprehensive assessment of the model's performance. Its calculation formula is shown in Eq. (6).

$$MIoU = \frac{1}{n}\sum_{i=1}^{n}\frac{TP}{FN+FP+TP} \tag{6}$$

4. Experiment

4.1. Dataset and implementation

4.1.1. Dataset

A total of 286 fresh human molars (from 20 to 30-year-old adults) obtained from the Stomatology Hospital of Sun Yat-sen

University and the Affiliated Hospital of Guangdong University of Technology, are used in this study. Cracks are created in these teeth through thermal expansion and contraction in thermodynamics to simulate cracked teeth. The same preparation protocol for vitro simulated cracked teeth and its details can be found in Refs. [24,48,49]. It should be mentioned that the preparation of cracked teeth samples requires a significant effort because of costs and resources, with specific requirements for tooth type and integrity. In the end, we select 600 optical images that meet the aforementioned criteria. According to the literature survey, many studies that apply semantic segmentation for detecting crack have concentrated on datasets with less than 600 images [50,51]. For instance, in Refs. [50, 51], they employed 200 and 504 images, respectively. Both studies achieved commendable performance in their specific tasks. Therefore, the dataset used in this paper is not considered small within the domain of crack segmentation.

4.1.2. Implementation details

After a thorough examination by an experienced dentist, the dataset is randomly divided into two groups (training, and testing sets) with a ratio of 5:1, and the same strategy is also employed in the research works of [50,52]. The dataset contains images of normal, exposed, dark light, dental fluorosis, dental plaque, and dental caries, with up to 450 images for normal teeth. Dental plaque and dental caries have the least representation, with 15 and 20 images, respectively. More details can be found in Table 2. Due to the limited representation of dental plaque and dental caries images, this paper conducts further qualitative analysis on these two categories. As these images are of different sizes, they are resized to a uniform resolution of 512 × 512 pixels to complete the dataset. Through the application of transfer learning technique, pre-training weights from ImageNet are used as the initialization of our model, which improve the network's performance, training efficiency, and effectiveness in the task of tooth crack segmentation. In addition, in order to reduce the risk of overfitting, batch normalization is introduced after each convolutional layer [53] and L2 regularization is applied for additional protection against overfitting [54]. Fig. S3 illustrates the validation loss and training loss plots for various DeepLabv3+ networks with different backbone architectures. By examining the figure, it can observe that all the models exhibit relative stability during the training process, with both the training loss and validation loss showing synchronous reduction and eventual convergence. Additionally, there is minimal fluctuation, and the values of training and validation losses remain consistently low, indicating that the models do not demonstrate significant overfitting.

The experimental platform uses Windows 10 operating system, Intel Xeon E5-2678 V3 processor, 32G memory capacity, and NVIDIA GeForce RTX 3090 24 GB graphics card, and the training and testing are all performed on MATLAB@2022a. During the training process, stochastic gradient descent with momentum (sgdm) is utilized as the optimizer. The batch size is set to 8, and the initial learning rate is established as 0.001. The model was trained for a total of 50 epochs. The data set is manually annotated using the Image Labeler toolbox in MATLAB@2022a. It is noteworthy that a professional dentist, who is one of the authors of this article, conducts the final review of all image annotation.

4.2. Comparison of commonly used models

To better validate the performance of the FDB-DeepLabv3+ model, a rigorous experimental design is employed as follows.

a. Comparative experiments are conducted to evaluate the performance of several commonly used image segmentation models, including FCN, U-Net, SegNet, and DeepLabv3+ (Sections 4.2.1 and 4.2.2). The results demonstrate that DeepLabv3+ outperforms the other models in terms of Mean Pixel Accuracy (MPA) and Mean Intersection over Union (MIoU).
b. Building on the previous results, to enhance the performance of the DeepLabv3+ model, experiments are conducted with different backbones, including ResNet18, ResNet50, ResNet101, Xception, MobileNetV2, and Inception-ResNetV2 (Sections 4.2.1 and 4.2.2). Among these models, it is revealed that the best performance is achieved by the ResNet50-DeepLabv3+ network.
c. Subsequently, a detailed comparative analysis is conducted between the proposed FDB-DeepLabv3+ and ResNet50-DeepLabv3+ to evaluate the performance of the FDB-DeepLabv3+ model in image segmentation tasks (Section 4.3). Through this comparative experiment, the strengths and limitations of FDB-DeepLabv3+ relative to ResNet50-DeepLabv3+ and other models are accurately assessed, further validating the performance of FDB-DeepLabv3+.

4.2.1. Quantitative results of commonly used models

In this section, the currently used semantic segmentation models are compared, including FCN, U-Net, SegNet, and DeepLabv3+ models based on six different backbones (ResNet18, ResNet50, ResNet101, Xception, MobileNetV2, Inception-ResNetV2). The same

Table 2
Proportion of different categories of images in the dataset.

Categories	Training set	Test set	Total
Exposure	24	6	30
Darker light	33	7	40
Dental fluorosis	38	7	45
Dental plaque	10	5	15
Dental caries	15	5	20
Normal	380	70	450
Total	500	100	600

dataset is used for both training and testing in the experiments to ensure comparability and consistency in the independent variable. Table 3 displays the quantitative comparison results of the eight models. It is observed that ResNet50-DeepLabv3+ exhibites superior performance in the comparisons on MPA and MIoU with values of 72.67 and 72.06, respectively. Notably, when compared with Xception-DeepLabv3+(baseline), ResNet50-DeepLabv3+ shows significant improvements of 8.56% and 9.13% in MPA and MIoU, respectively. Furthermore, it should be noted that the performance of ResNet101-DeepLabv3+ exhibits slightly inferior results compared to ResNet50-DeepLabv3+, suggesting that in our crack segmentation task, deeper network cannot further enhance feature learning and image segmentation capabilities. Similar results can also be found in Refs. [55–57]. These results further demonstrate the dominance of ResNet50-DeepLabv3+ in the task of image segmentation. Among the nine models evaluated, FCN demonstrats the lowest performances.

4.2.2. *Visual segmentation results of commonly used models*

Fig. 3 provides the segmentation results of different models on tooth cracks under normal circumstances. It can be observed that the segmentation results of FCN-8s and U-Net are coarse, with serious segmentation accuracy issues. When segmenting, slight reflections and uneven areas on the tooth surface are easily misrecognized as cracks. SegNet suffers from serious over-segmentation of the detailed contours of tooth cracks. It can be seen that the DeepLabv3+ series models exhibit significant segmentation advantages. As shown in Fig. 3(a) and (b), the segmentation accuracy of each model is relatively high, and the segmented crack contours are closer to the ground truth. For Fig. 3(c) and (d), each model tends to recognize the boundary between the tooth contour and the background as a crack, resulting in inaccurate segmentation. As for Fig. 3(e) and (f), when the cracks on the tooth surface are fine and the contrast with the tooth surface is low, the DeepLabv3+ series models with different backbones may not accurately find the contour details of tooth cracks, resulting in over-segmentation and coarse segmentation results. Among these models, DeepLabv3+ with ResNet as its backbone delivers the best performance, yielding a more comprehensive outline of segmented tooth cracks. However, noticeable differences in segmentation quality emerge when comparing various ResNet versions. In contrast to ResNet50-DeepLabv3+ and ResNet101-DeepLabv3+, ResNet18-DeepLabv3+ is capable of delineating crack outlines effectively, but it demonstrates a reduced proficiency in classifying interfering pixels within each image. Both ResNet50-DeepLabv3+ and ResNet101-DeepLabv3+ exhibit strong segmentation capabilities for tooth crack pixels. However, ResNet101-DeepLabv3+ lags in terms of the continuity of crack segmentation, while ResNet50-DeepLabv3+ demonstrates the most comprehensive advantage, excelling in both the continuity of crack segmentation and the handling of interfering pixels.

Fig. 4 shows the segmentation results of various models under the influence of factors such as severe exposure, dental plaque, dental fluorosis, and cavities. Although the overall performance of the DeepLabv3+ series models in segmenting tooth cracks is satisfactory, these models still have some segmentation accuracy issues. In Fig. 4(a), when the tooth crack is in an image with severe exposure, the DeepLabv3+ models with Xception, MobileNetv2, and Inception-ResNetv2 as the backbone incorrectly classify some pixels in the overexposed areas as cracks, resulting in over-segmentation. In Fig. 4(b), when the tooth crack is covered by dental plaque, all models recognize some of the dental plaque area as cracks, resulting in inaccurate segmentation. For Fig. 4(c) and (d), when cracks appear on teeth with dental fluorosis or other lesions, the DeepLabv3+ models cannot precisely capture the intricate characteristics of the crack edges, resulting in segmentation errors in the lesion area. For teeth with cracks affected by dental caries, such as Fig. 4(e) and (f), even the relatively good performing ResNet50-DeepLabv3+ model and ResNet101-DeepLabv3+ model can cause missing edge features of tooth cracks. Among the tested models, the ResNet50-DeepLabv3+ model and the ResNet101-DeepLabv3+ model both show promising performance, but it is still affected to some extent by factors such as image illumination, dental fluorosis, and dental caries. Considering the performance of the model's performance as well as the computational resources, we further tend to improve DeepLabv3+ based on the ResNet50-DeepLabv3+ model. In order to verify the performance of the proposed model (FDB-DeepLabv3+), a series of comparison experiments between it and ResNet50-DeepLabv3+ are conducted.

4.3. *Comparison between FDB-DeepLabv3+ and ResNet50-DeepLabv3+*

The quantitative results and visual segmentation outcomes of commonly used models have been discussed in the previous sections, confirming the effectiveness of backbone network replacement from DeepLabv3+ to ResNet50. In this section, the training loss curves of DeepLabv3+ with different backbones are first examined, followed by a comparison of the training loss curves between FDB-DeepLabv3+ and ResNet50-DeepLabv3+. Subsequently, a quantitative analysis of the performance of FDB-DeepLabv3+ and

Table 3
Quantitative results among the eight methods on self-made hidden cracked tooth dataset.

Model	Backbone/Type	MPA (%)	MIoU (%)
FCN	8s	56.91	54.52
U-Net	VGG16	58.67	56.80
SegNet	VGG16	59.98	58.36
DeepLabv3+	ResNet18	70.15	69.30
DeepLabv3+	**ResNet50**	**72.67**	**72.06**
DeepLabv3+	ResNet101	72.24	71.16
DeepLabv3+	Xception	64.11	62.93
DeepLabv3+	MobileNetv2	65.38	64.34
DeepLabv3+	InceptionResNetv2	69.98	69.12

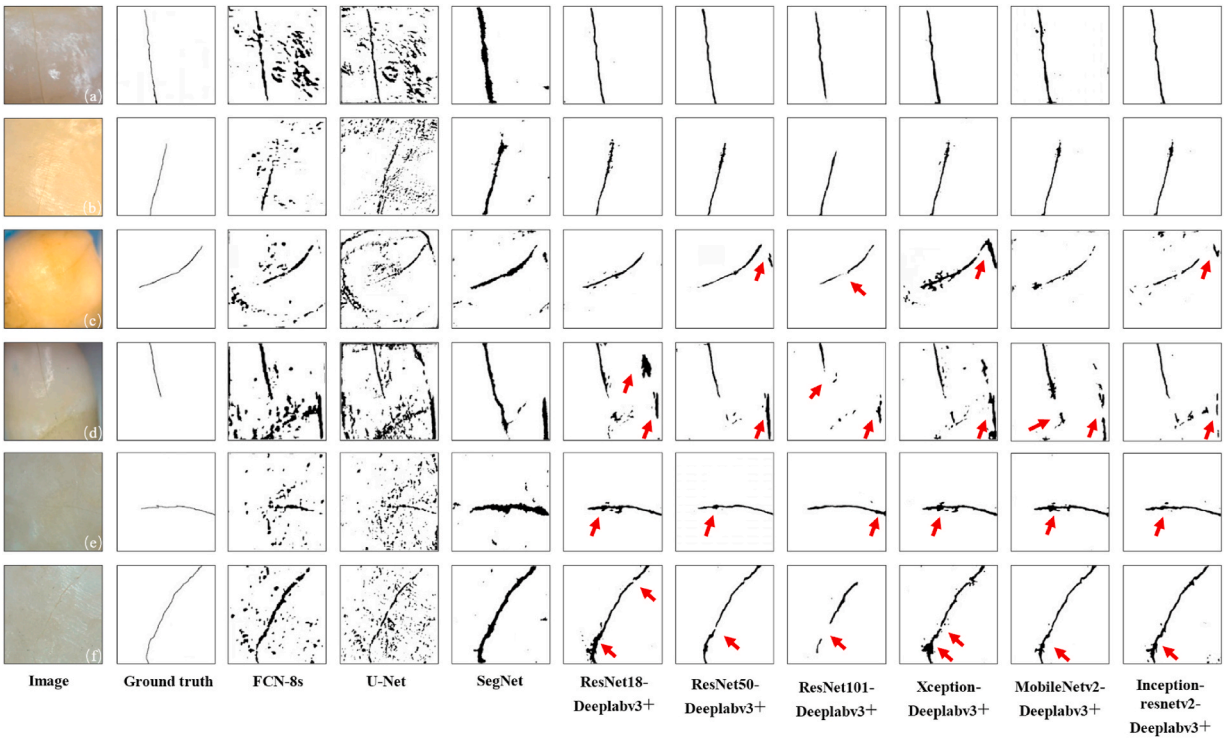

Fig. 3. Segmentation comparison of eight models when handling normal tooth surface, including (a)–(d) Adequate light, (e)–(f) Darker light and tinny cracks.

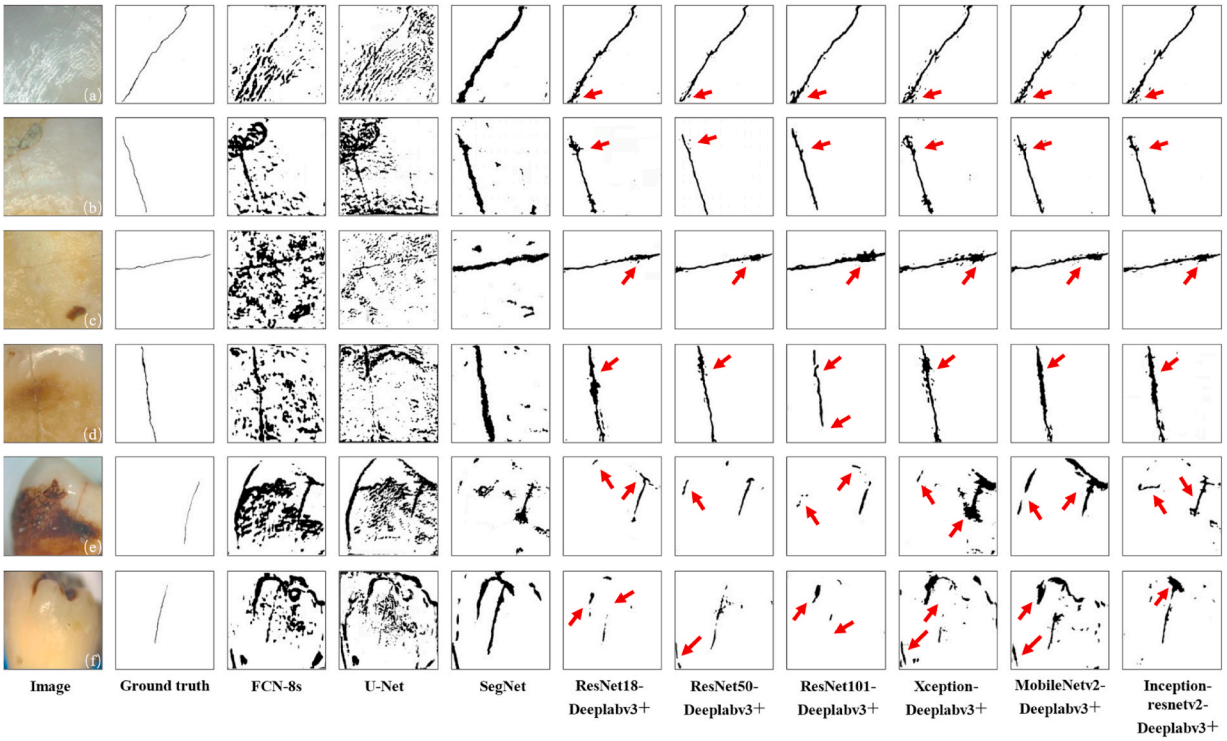

Fig. 4. Segmentation comparisons of eight models when handling challenging tooth surfaces, including (a) Exposure, (b) Dental plaque, (c)–(d) Dental fluorosis, (e)–(f) Dental caries.

ResNet50-DeepLabv3+ is conducted based on MPA and MIoU metrics. Finally, the visualization of tooth crack segmentation results achieved by FDB-DeepLabv3+ and ResNet50-DeepLabv3+ is compared.

4.3.1. Quantitative analysis of FDB-DeepLabv3+

Fig. 5 shows the training loss curves of the DeepLabv3+ model based on different backbones and the proposed FDB-DeepLabv3+. As shown in Fig. 5(a), the ResNet50-DeepLabv3+ model performs the best and has the lowest loss value compared to DeepLabv3+ with other backbone (ResNet18, ResNet101, Xception, MobileNetV2, Inception-ResNetV2). Specifically, the Xception-DeepLabv3+ model exhibits the poorest performance with the highest loss value. Although the MobileNetV2-DeepLabv3+ model shows slightly better results than Xception-DeepLabv3+, it still falls short of achieving satisfactory outcomes. Moreover, both the ResNet18-DeepLabv3+, ResNet101-DeepLabv3+ and Inception-ResNetV2-DeepLabv3+ models are close in performance, but neither of them represents the optimal solution. This suggests that using ResNet50 as the backbone can result in higher accuracy and a lower loss rate due to its residual structure and deeper network architecture. On the other hand, Fig. 5(b) indicates that the proposed FDB-DeepLabv3+ model, which improved on the ResNet50-DeepLabv3+ model, has a faster convergence rate during the initial training period. After approximately 700 iterations, the loss curve gradually stabilizes and remains around 0.05, and its loss value is lower than that of the ResNet50-DeepLabv3+ model. This suggests that the FDB-DeepLabv3+ model may effectively reduce the loss of feature information and improve the segmentation effect of tooth cracks.

Table 4 presents the quantitative comparison results between the ResNet50-DeepLabv3+ and FDB-DeepLabv3+ models. The results indicate that the proposed FDB-DeepLabv3+ model achieves a 2.85% higher MPA and 3.01% higher MIoU than the ResNet50-DeepLabv3+ model. These results demonstrate the proposed method has improved the detection capability of small pixel points, such as tooth cracks, by better extracting crack features and enhancing the attention to local information where tooth crack features reside.

By further combining the results from Table 3, it is found that, compared to the original DeepLabv3+ model, the MPA of the FDB-DeepLabv3+ model is improved by 11.41%, and the MIoU is improved by 12.14%. This indicates that the FDB-DeepLabv3+ model demonstrates higher overall accuracy and better classification performance in pixel-level categorization. Considering these results collectively, it can be concluded that the FDB-DeepLabv3+ model exhibits superior performance in pixel-level classification tasks, providing enhanced capabilities for the accurate detection of fine structures such as tooth cracks.

4.3.2. Qualitative analysis of FDB-DeepLabv3+

To validate the segmentation performance of the improved DeepLabv3+ model on tooth crack images, some difficult-to-segment tooth crack images are selected from the test set for comparative analysis. By visualizing and comparing the segmented results, it is evident that the FDB-DeepLabv3+ model indeed improves the segmentation performance of tooth cracks.

Fig. 6 illustrates the segmentation results of the DeepLabv3+ model before and after improvement on tooth crack images. Fig. 6(a) shows a tooth surface image under exposure influence, where the ResNet50-DeepLabv3+ model produces rough crack boundaries due to the exposure effect on the crack area. In contrast, the FDB-DeepLabv3+ model can effectively refine the crack boundaries, but some segmentation errors still occur in the exposure area. Fig. 6(b) depicts image distortion, where the ResNet50-DeepLabv3+ model produces discontinuities in the segmented tooth cracks, while the improved model maintains the continuity of cracks. Furthermore, several examples (tooth decay, dental plaque, fluorosis) have been showcased, as some images in the dataset are heavily contaminated, which can serve as a better test of the model's performance. Fig. 6(c–f) shows images of lesioned tooth surfaces, where the ResNet50-DeepLabv3+ model produces segmentation errors in some lesion areas during tooth crack segmentation, leading to missing the edge features of tooth cracks. In contrast, the FDB-DeepLabv3+ model effectively improves this phenomenon and demonstrates robustness. In summary, the improved FDB-DeepLabv3+ model effectively reduces the impact of factors such as plaque, dental caries and image

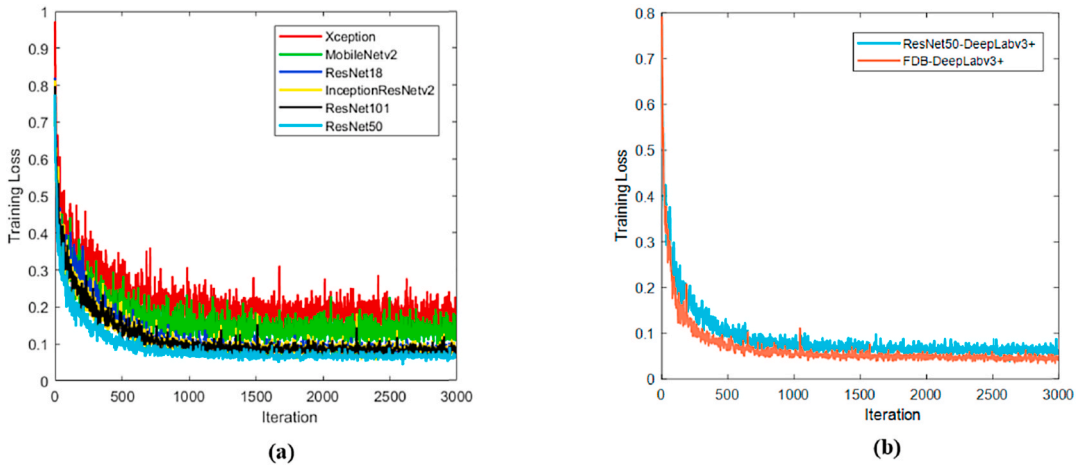

Fig. 5. Loss curves of different models on self-made hidden cracked tooth dataset (a) DeepLabv3+ model based on different backbones, (b) ResNet50-DeepLabv3+ and FDB-DeepLabv3+.

Table 4
Quantitative results among the ResNet50-DeepLabv3+ and FDB-DeepLabv3+ on self-made hidden cracked tooth dataset.

Model	MPA (%)	MIoU (%)
ResNet50-DeepLabv3+	72.67	72.06
FDB-DeepLabv3+	75.52	75.07

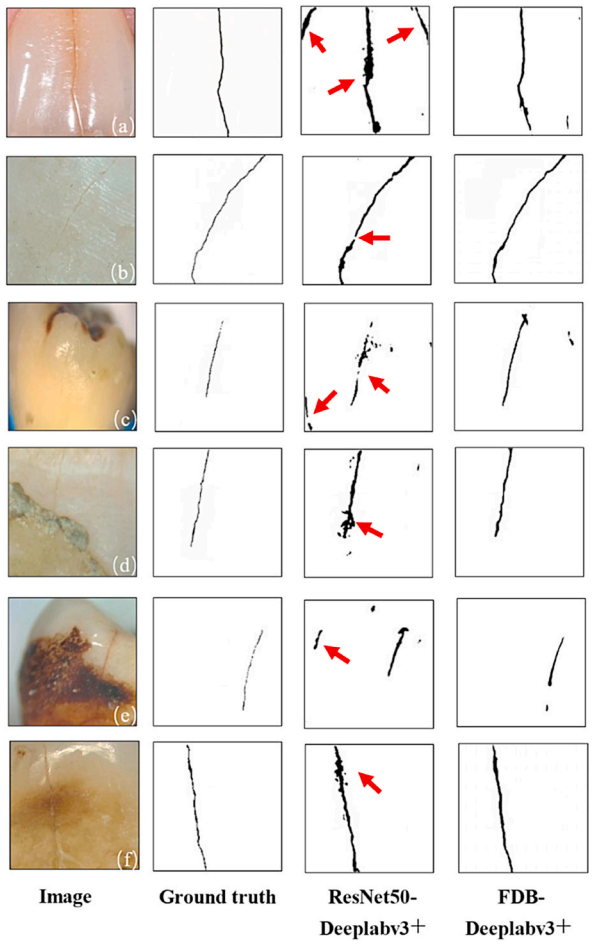

Fig. 6. Segmentation comparisons of the model before and after improvement when handling challenging cases, including (a) Exposure, (b) Image distortion, (c)–(f) Diseased teeth.

exposure on the segmentation accuracy of the model. Moreover, the crack contours segmented by the model are more closely aligned with the annotated ground truth in various scenarios, and improves the segmentation results of the tooth boundary.

Considering the limited number of images depicting dental plaque and dental caries in the dataset, this paper undertakes a comprehensive comparison of these two image types. In Fig. 7(a), it can observe that when plaque is severe, DeepLabv3+ models with non-ResNet backbones (including Xception, MobileNetv2, and InceptionResNetv2) exhibit relatively poor segmentation performance. Specifically, Xception performs weakest in plaque classification, while MobileNetv2 requires further improvement in distinguishing reflective pixels. Although InceptionResNetv2 performs the best among these three, it still occasionally misclassifies boundary pixels of plaque as cracks. In contrast, DeepLabv3+ models with ResNet as the backbone perform significantly better. Among these, ResNet18 shows some segmentation errors, while ResNet50 and ResNet101 excel in plaque boundary pixel classification. FDB-DeepLabv3+ outperforms all of them without the aforementioned shortcomings. In Fig. 7(b), all models face challenges in fully segmenting cracks. However, DeepLabv3+ models with ResNet as the backbone demonstrate better error segmentation, while FDB-DeepLabv3+ excels by not erroneously classifying any non-crack pixels. For Fig. 7(c–d), given the large area of dental caries all models find it challenging to fully segment cracks. Nevertheless, in terms of error segmentation, ResNet50, ResNet101, and FDB-DeepLabv3+ stand out, with FDB-DeepLabv3+ demonstrating the most outstanding performance.

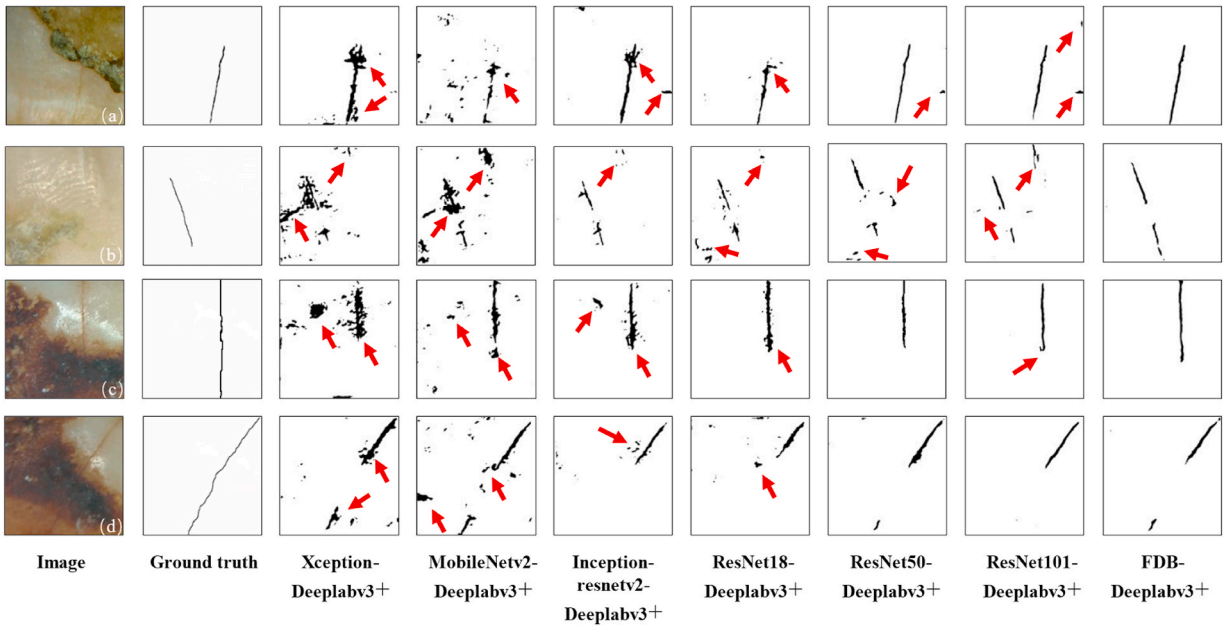

Fig. 7. Different DeepLabv3+ models for segmenting crack in tooth with dental plaque and dental caries.

4.4. Ablation analysis

Ablation experiments are conducted to verify the impact of the improved modules on the overall performance of the proposed model. The semantic segmentation results of tooth crack images with different combination methods are shown in Table 5.

Comparing the results based on (1) and (2), it can be observed that replacing the backbone from Xception with ResNet50 resulted in an increase of 8.56% in MPA and 9.13% in MIoU values. This indicates that the ResNet block effectively enhances the model's feature learning capability. Based on the results comparison of (2) and (3), it is seen that by fusing the feature maps of three different levels of the ResNet50 backbone through the FPN structure and using a layer-by-layer up-sampling method for concatenation and fusion as an important branch of the input decoding layer, the overall performance of the model is slightly improved, with MPA and MIoU values increasing by 0.56% and 0.62% respectively. According to the results comparison of (2) and (4), it is seen that after changing the original ASPP module to the densely linked ASPP module, the model performs better, with MPA and MIoU values increasing by 1.11% and 1.19% respectively. Based on the results comparison of (2) and (5), it is seen that after introducing the attention branch, the value of MPA and MIoU increase by 1.52% and 1.62% respectively which indicates that this method effectively improves the accuracy of feature extraction. Based on the result of approach (6), it is observed that the incorporation of all three modules concurrently leads to a significant enhancement in the overall performance of the model. The resultant MPA and MIoU values are found to be 75.52% and 75.07%, respectively.

5. Discussion

The accurate diagnosis of cracked teeth presents a significant challenge due to the presence of early symptoms that are often vague and easily confounded with other dental conditions. This can result in both misdiagnosis and missed diagnosis. Research has demonstrated that AI-based deep learning methods, such as wavelet-guided deep models, can outperform human observers in terms of diagnostic accuracy [21,22]. While certain deep learning methods have been shown to provide a relatively high diagnosis rate (exceeding 90% using binary classification neural networks [20,24] as well as combined U-Net and YOLOv4 network [37]), issues such as extended detection time and limited crack information suggest that semantic segmentation methods may offer more detailed and

Table 5
Semantic segmentation results of tooth crack images using different combination methods.

approach	ResNet50	FPN	Dense ASPP	BAM	MPA (%)	MIoU (%)
(1)					64.11	62.93
(2)	✓				72.67	72.06
(3)	✓	✓			73.23	72.68
(4)	✓		✓		73.78	73.25
(5)	✓			✓	74.19	73.68
(6)	✓	✓	✓	✓	75.52	75.07

specific information. This can facilitate a more comprehensive understanding of the crack structure and aid in the formulation of subsequent treatment plans. The utilization of optical imaging in this study is noteworthy, as it not only reduces costs but also effectively avoid the limitation of the U-Net learning framework in tooth crack segmentation with CBCT imaging. The smoothing filters used in HR-CBCT may disrupt subtle features in the images [22,25].

In traditional small object detection, DeepLabv3 demonstrates significant advantages compared to FCN and U-Net model [58]. With improvements to the ASPP module and the fully connected conditional random field module, DeepLabv3+ has achieved even better results in terms of mandibular fractures detection [37], particularly in detecting smaller targets such as cracks in terms of over-segmentation and precision. Based on our experiments DeepLabv3+ performs best in tooth crack segmentation compared to other commonly used networks. However, the results of the image segmentation show that segmenting cracks in segmentation detail and under other pathologies are flawed and can be further optimized. Our study aims to enhance the performance based on DeepLabv3+ model. ResNet50 is replaced as the backbone of the model, owing to its advantages in terms of depth, parameter sharing, expressive feature representation, and advanced architectural design. Numerous studies have indicated that this modification leads to improved performance of the model [59–61]. Feature Pyramid Network (FPN) is added to effectively integrate low-level, mid-level, and high-level features from the backbone, compensating for the loss of feature information. Similar improvement can also be found in reference quotation [62]. ASPP module is integrated to connect crack features in a more densely manner, ensuring strong correlations between segmentation objective and maintaining the integrity of objective segmentation [63]. What's more, BAM module is incorporated into our model based on the evidence from Ref. [64], which demonstrates that BAM module can focus on both spatial and channel aspects of the image, enhancing the weight allocation of crack features and further improving the details of the crack features. Other attention mechanisms, such as ECA and CBAM, come with certain limitations in handling global context dependencies and channel-wise relationships, or they have higher complexity and require more computational resources. Considering factors like execution efficiency and performance improvements (compared to residual networks), BAM is integrated into our model. Ablation experiments show that each module has a positive impact on the detection and that the inclusion of all modules is effective.

Currently, research on the use of artificial intelligence to assist dentists in detecting tooth fractures is rare. In the field of dentistry, especially in the diagnosis of diseases related to tooth cracks, the potential of artificial intelligence remains largely untapped. The novelty of this research lies not only in the exploration of using artificial intelligence to detect cracks in human teeth with optical pictures, but also in the investigation of the impact of various backbones on the performance of the DeepLabv3+ model. Most importantly, we further integrated and enhanced the traditional DeepLabv3+ model, resulting in the FDB-DeepLabv3+ model. This model significantly improves the numerical performance of the conventional DeepLabv3+ model in the task of semantic segmentation of cracked teeth, as evidenced by higher MPA and MIoU evaluation metrics on self-made cracked tooth dataset.

Traditional methods (patient clinical symptoms-based method, microscopy imaging-based method, CBCT-based methods etc.) may have limitations in diagnosis, but the application of semantic segmentation techniques through deep learning models holds the promise of enhancing the accuracy of fracture detection, providing more reliable auxiliary tools for dentists. This approach has the potential to bring about new breakthroughs in dental healthcare, improving early diagnosis rates of tooth fractures and thereby enhancing patient treatment and overall health outcomes.

The proposed method should be further performed to investigate the diagnostic value in real clinical intraoral diagnostics, especially in areas such as occlusal surfaces and dental implantations. Furthermore, reducing model complexity to facilitate deployment on embedded devices in real clinical assistive diagnostic environments is another direction that needs further investigation.

6. Conclusion

This paper has presented an in-depth investigation into AI-based semantic segmentation method for detecting crack in human teeth with optical images. Based on deep learning, the DeepLabv3+ model has been improved by incorporating knowledge from the semantic segmentation field and the experience of medical experts. A microscopic image-based tooth crack segmentation algorithm named FDB-DeepLabv3+ has been proposed for the task of segmentation of cracks in the presence of various interfering factors. This method has demonstrated superior performance in terms of MPA and MIoU metrics compared to classical DeepLabv3+ model, with an MPA that is 11.41% higher and an MIoU that is 12.14% higher. Overall, the proposed model contributes to improving the detection of tooth crack in optical image and providing a potential solution for subsequent diagnoses.

Ethics approval and consent to participate

Not applicable.

Funding

This work is supported by National Natural Science Foundation of China (Grant No. 82001983) as well as by Guangdong Basic and Applied Basic Research Foundation (Grant No. 2019A1515111202), Science and Technology Program of Guangzhou (No. 202002030269), Guangdong Science and Technology Innovation Strategy Special Funds (No. pdjh2024b302), and Student Innovation Training Program of Guangzhou University (No. XJ202311078127). Additionally, financial support is acknowledged from the program of China Scholarships Council program (No. 202108440270).

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

CRediT authorship contribution statement

Zewen Xie: Writing – original draft, Visualization, Software, Investigation. **Qilin Lu:** Writing – original draft, Data curation. **Juncheng Guo:** Investigation, Conceptualization. **Weiren Lin:** Validation. **Guanghua Ge:** Validation, Resources. **Yadong Tang:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Damiano Pasini:** Supervision, Project administration, Formal analysis. **Wenlong Wang:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank the staff of the Hospital of Stomatology at Sun Yat-sen University and the staff of Hospital at Guangdong University of Technology for their technical support and the tooth sample.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e25892.

References

[1] Marco A. Peres, et al., Oral diseases: a global public health challenge, Lancet 394 (10194) (2019) 249–260.
[2] Carlos Zaror, et al., Assessing oral health-related quality of life in children and adolescents: a systematic review and standardized comparison of available instruments, Clin. Oral Invest. 23 (2019) 65–79.
[3] William Kahler, The cracked tooth conundrum: terminology, classification, diagnosis, and management, Am. J. Dent. 21 (5) (2008) 275.
[4] Werner Geurtsen, Thomas Schwarze, Huesamettin Günay, Diagnosis, therapy, and prevention of the cracked tooth syndrome, Quintessence Int. 34 (2003) 6.
[5] Juncheng Guo, et al., A perspective on the diagnosis of cracked tooth: imaging modalities evolve to AI-based analysis, Biomed. Eng. Online 21 (1) (2022) 36.
[6] J.E.N.S.C. Türp, John P. Gobetti, The cracked tooth syndrome: an elusive diagnosis, J. Am. Dent. Assoc. 127 (10) (1996) 1502–1507.
[7] Sung-Eun Yang, et al., Analysis of the characteristics of cracked teeth and evaluation of pulp status according to periodontal probing depth, BMC Oral Health 17 (2017) 1–6.
[8] Thomas J. Hilton, et al., Associations of types of pain with crack-level, tooth-level and patient-level characteristics in posterior teeth with visible cracks: findings from the National Dental Practice-Based Research Network, J. Dent. 70 (2018) 67–73.
[9] Kadandale Sadasiva, et al., Cracked tooth syndrome: a report of three cases, J. Pharm. BioAllied Sci. 7 (Suppl 2) (2015) S700.
[10] Byoung-Duck Roh, Young-Eun Lee, Analysis of 154 cases of teeth with cracks, Dent. Traumatol. 22 (3) (2006) 118–123.
[11] Paul Abbott, N. Leow, Predictable management of cracked teeth with reversible pulpitis, Aust. Dent. J. 54 (4) (2009) 306–315.
[12] P. Wang, et al., Detection of dental root fractures by using cone-beam computed tomography, Dentomaxillofacial Radiol. 40 (5) (2011) 290–298.
[13] Mitchell Edlund, Madhu K. Nair, Umadevi P. Nair, Detection of vertical root fractures by using cone-beam computed tomography: a clinical study, J. Endod. 37 (6) (2011) 768–772.
[14] Kıvanç Kamburoğlu, et al., Detection of vertical root fracture using cone-beam computerized tomography: an in vitro assessment, Oral Surg. Oral Med. Oral Pathol. Oral Radiol. Endod. 109 (2) (2010) e74–e81.
[15] Metska, Maria Elissavet, et al., Detection of vertical root fractures in vivo in endodontically treated teeth by cone-beam computed tomography scans, J. Endod. 38 (10) (2012) 1344–1347.
[16] David J. Clark, Cherilyn G. Sheets, Jacinthe M. Paquette, Definitive diagnosis of early enamel and dentinal cracks based on microscopic evaluation, J. Esthetic Restor. Dent. 15 (7) (2003) 391–401.
[17] Sebeena Mathew, et al., Diagnosis of cracked tooth syndrome, J. Pharm. BioAllied Sci. 4 (Suppl 2) (2012) S242.
[18] John S. Mamoun, Napoletano Donato, Cracked tooth diagnosis and treatment: an alternative paradigm, Eur. J. Dermatol. 9 (2) (2015) 293–303.
[19] As van, A. Glenn, Evaluation of enamel and dentinal cracks using methylene blue dye and the operating microscope, Inside Dent. 3 (2007) 7.
[20] Ziyang Hu, et al., Diagnosis of in vivo vertical root fracture using deep learning on cone-beam CT images, BMC Oral Health 22 (1) (2022) 382.
[21] Jared Vicory, et al., Dental microfracture detection using wavelet features and machine learning, in: Medical Imaging 2021: Image Processing, vol. 11596, SPIE, 2021.
[22] Pranjal Sahu, et al., Wavelet guided 3D deep model to improve dental microfracture detection, in: Applications of Medical Artificial Intelligence: First International Workshop, AMAI 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings, Springer Nature Switzerland, Cham, 2022.
[23] Irma Dumbryte, et al., Revelation of microcracks as tooth structural element by X-ray tomography and machine learning, Sci. Rep. 12 (1) (2022) 22489.
[24] Juncheng Guo, et al., Automatic detection of cracks in cracked tooth based on binary classification convolutional neural networks, Appl. Bionics Biomech. 2022 (2022).

[25] Erinne B. Lubisich, Thomas J. Hilton, Ferracane Jack, Cracked teeth: a review of the literature, J. Esthetic Restor. Dent. 22 (3) (2010) 158–167.
[26] Ernst-Peter Rührnschopf, Klaus Klingenbeck, A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. Part 1: scatter compensation approaches, Med. Phys. 38 (7) (2011) 4296–4311.
[27] Sebeena Mathew, et al., Diagnosis of cracked tooth syndrome, J. Pharm. BioAllied Sci. 4 (Suppl 2) (2012) S242.
[28] Zhong Zhou, Junjie Zhang, Chenjie Gong, Hybrid semantic segmentation for tunnel lining cracks based on Swin Transformer and convolutional neural network, Comput.-Aided Civil Infrastruct. Eng. (2023).
[29] Zhong Zhou, et al., Real-time tunnel lining crack detection based on an improved You Only Look Once version X algorithm, Georisk 17 (1) (2023) 181–195.
[30] Shankargouda Patil, et al., Artificial intelligence in the diagnosis of oral diseases: applications and pitfalls, Diagnostics 12 (5) (2022) 1029.
[31] Jinke Wang, et al., SAR-U-Net: squeeze-and-excitation block and atrous spatial pyramid pooling based residual U-Net for automatic liver segmentation in Computed Tomography, Comput. Methods Progr. Biomed. 208 (2021) 106268.
[32] Chen Chu, Jihui Zheng, Yong Zhou, Ultrasonic thyroid nodule detection method based on U-Net network, Comput. Methods Progr. Biomed. 199 (2021) 105906.
[33] M. Roy Reena, P.M. Ameer, Localization and recognition of leukocytes in peripheral blood: a deep learning approach, Comput. Biol. Med. 126 (2020) 104034.
[34] Jahangir Jabbar, et al., Deep learning based classification of wrist cracks from X-ray imaging, CMC-Comput. Mater. Continua 73 (1) (2022) 1827–1844.
[35] Yangling Ma, Yixin Luo, Bone fracture detection through the two-stage system of crack-sensitive convolutional neural network, Inf. Med. Unlocked 22 (2021) 100452.
[36] Chunmei Xia, et al., Multi-channel residual neural network based on squeeze-and-excitation for osteoporosis diagnosis, in: Clinical Image-Based Procedures: 11th Workshop, CLIP 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18, 2022, Proceedings, Springer Nature Switzerland, Cham, 2023.
[37] Dong-Min Son, et al., Combined deep learning techniques for mandibular fracture diagnosis assistance, Life 12 (11) (2022) 1711.
[38] Jonathan Long, Evan Shelhamer, Trevor Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[39] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-net: convolutional networks for biomedical image segmentation, in: 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.
[40] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Trans. Pattern Anal. Mach. Intell. 39 (12) (2017) 2481–2495.
[41] Liang-Chieh Chen, et al., Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018.
[42] François Chollet, Xception: deep learning with depthwise separable convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[43] Kaiming He, et al., Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[44] Shengyi Zhao, Jizhan Liu, Shuo Wu, Multiple disease detection method for greenhouse-cultivated strawberry based on multiscale feature fusion Faster R_CNN, Comput. Electron. Agric. 199 (2022) 107176.
[45] Tsung-Yi Lin, et al., Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[46] Huixuan Fu, et al., Bridge crack semantic segmentation based on improved DeepLabv3+, J. Mar. Sci. Eng. 9 (6) (2021) 671.
[47] Jongchan Park, et al., BAM: bottleneck attention module, in: British Machine Vision Conference (BMVC), British Machine Vision Association (BMVA), 2018.
[48] Chunliang Zhang, et al., A method of crack detection based on digital image correlation for simulated cracked tooth, BMC Oral Health 21 (1) (2021) 1–10.
[49] M. Yuan, et al., Using Meglumine Diatrizoate to improve the accuracy of diagnosis of cracked teeth on Cone-beam CT images, Int. Endod. J. 53 (5) (2020) 709–714.
[50] Wooram Choi, Young-Jin Cha, SDDNet: real-time crack segmentation, IEEE Trans. Ind. Electron. 67 (9) (2019) 8016–8025.
[51] Chuncheng Feng, et al., Automatic pixel-level crack detection on dam surface using deep convolutional network, Sensors 20 (7) (2020) 2069.
[52] Yahui Liu, et al., DeepCrack: a deep hierarchical feature learning architecture for crack segmentation, Neurocomputing 338 (2019) 139–153.
[53] Sergey Ioffe, Christian Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, PMLR, 2015.
[54] Xiangyu Zhao, et al., D2A U-Net: automatic segmentation of COVID-19 CT slices based on dual attention and hybrid dilated convolution, Comput. Biol. Med. 135 (2021) 104526.
[55] Çinare Oğuz, Mete Yağanoğlu, Detection of COVID-19 using deep learning techniques and classification methods, Inf. Process. Manag. 59 (5) (2022) 103025.
[56] Vílchez, Richar Fernández, David Mauricio, Bullet impact detection in silhouettes using mask R-CNN, IEEE Access 8 (2020) 129542–129552.
[57] M. Emin Sahin, et al., Detection and classification of COVID-19 by using faster R-CNN and mask R-CNN on CT images, Neural Comput. Appl. 35 (18) (2023) 13597–13611.
[58] Yandan Wang, et al., Automatic detection and recognition of nasopharynx gross tumour volume (GTVnx) by deep learning for nasopharyngeal cancer radiotherapy through magnetic resonance imaging, Radiat. Oncol. 18 (1) (2023) 76.
[59] Qiwen Qiu, Denvid Lau, Real-time detection of cracks in tiled sidewalks using YOLO-based method applied to unmanned aerial vehicle (UAV) images, Autom. ConStruct. 147 (2023) 104745.
[60] Yuhao Zhang, Zhongwei Wang, Concrete surface crack recognition based on coordinate attention neural networks, Comput. Intell. Neurosci. 2022 (2022).
[61] Amit Dipankar, Sanjeev Kumar Suman, Pavement crack detection based on a deep learning approach and visualisation by using GIS, Int. J. Pavement Eng. 24 (1) (2023) 2173754.
[62] Hao Zhang, Jiaxiu Dong, Ziqiao Gao, Automatic segmentation of airport pavement damage by AM-Mask R-CNN algorithm, Eng. Rep. (2023) e12628.
[63] Shuangcai Yin, et al., SD-UNet: a novel segmentation framework for CT images of lung infections, Electronics 11.1 (2022) 130.
[64] Xuezhi Xiang, Yuqi Zhang, Abdulmotaleb El Saddik, Pavement crack detection network based on pyramid structure and attention mechanism, IET Image Process. 14 (8) (2020) 1580–1586.