Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/mbe

MBE, 19(4): 4178–4195.
DOI: 10.3934/mbe.2022193
Received: 05 January 2022
Revised: 05 February 2022
Accepted: 11 February 2022
Published: 17 February 2022

Research article

The defect detection for X-ray images based on a new lightweight semantic segmentation network

Xin Yi, Chen Peng*, Zhen Zhang and Liang Xiao

School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200444, China

* **Correspondence:** Email: c.peng@shu.edu.cn.

Abstract: The tire factory mainly inspects tire quality through X-ray images. In this paper, an end-to-end lightweight semantic segmentation network is proposed to realize the error detection of bead toe. In the network, firstly, the texture feature of different regions of tire is extracted by an encoder. Then, we introduce a decoder to fuse the output feature of the encoder. As the dimension of the feature maps is reduced, the positions of bead toe in the X-ray image have been recorded. When evaluating the final segmentation effect, we propose a local mIoU(L-mIoU) index. The segmentation accuracy and reasoning speed of the network are verified on the tire X-ray image set. Specifically, for 512×512 input images, we achieve 97.1% mIoU and 92.4% L-mIoU. Alternatively, the bead toe coordinates are calculated using only 1.0 s.

Keywords: tire X-ray images; defect detection; image processing; lightweight semantic segmentation network; auxiliary supervision

1. Introduction

Traffic accidents are the third leading cause of death in the United States according to the Centers for Disease Control and Prevention [1]. An unnoticeable fact is that the quality of tires has caused a lot of traffic accidents. The National Highway Traffic Safety Administration reports an average of nearly 11,000 accidents related to tire failures each year [2]. Undoubtedly, the quality of tires not only affects the safety of riders, but also affects the lives and property of every road traveler. In the tire production process, quality inspection can prevent consumers from using unqualified products. Currently, the main market share is dominated by radial tires. This type of tire not only has multiple layers of belt attached to the crown, but also has a tightly wound wire coil, which significantly improves the strength and stability of the ring [3, 4]. Radial tires require a complex production process, which increases the possibility of various quality defects [5, 6]. For this type of tire, the most popular nondestructive detection scheme is to place the tire product in an X-ray room for irradiation.The resulting 2-D grayscale

image is measured according to enterprise standards. Although a large number of enterprises use artificial visual inspection, such a process has significant disadvantages in terms of efficiency and accuracy. For now, unattended defect detection based on computer vision is the trend.

To meet the needs of intelligent detection, many researchers have studied tire X-ray images. Based on traditional image processing methods, researchers have proposed corresponding research results from multiple perspectives. Guo et al. [7] firstly evaluated the pixel-level texture similarity. Impurities in the sidewall and crown were then located by threshold treatment. However, pixel-level operation is extremely challenging for computational speed, especially when we consider that the size of X-ray images is quite large. Zhang et al. [8–10] have systematically studied the detection methods to detect impurities. Combining the total variation with the Curvelet transformation, the original image was split into texture and cartoon. Ultimately, impurities were located by cartoons [8]. A method of combining Curvelet with Canny operator was designed to extract the defect feature effectively [9]. Wavelet multiscale analysis was proposed in [10]. This method included local analysis and scale feature extraction to separate defect from background. There were obvious differences between the local inverse differential moment characteristics of the normal area and the impurity area. Therefore, [11] used this principle to detect significant impurities. Different tire types are significantly different in the amount of raw material, resulting in differences in the brightness of the grayscale image after X-ray irradiation. The above traditional defect detection methods need to set relevant algorithm parameters according to the detection model. Therefore, these methods are particularly sensitive to the brightness of grayscale images. The first motivation of this paper is to design a tire defect detection mathod. Specifically, it reduces the impact of image brightness on the detection results, while taking into account the defect detection of various types of tires.

In recent years, deep learning technology has been widely used in image classification [12, 13], target recognition [14, 15], semantic segmentation [16, 17] and the other fields. Some researchers have also applied deep learning technology to a variety of visual detection tasks [18], such as fabric, welding, and tire internal defect detection. In [19], the convolution neural network based on ResNet [20] learned the texture features of the fabric, making it possible to accurately locate small defects on the fabric. Aiming at the target of automatic location of welding defects, [21] proposed an improved U-Net network. The network combines random cropping and preprocessing methods to effectively expand the data set. In [22], an end-to-end tire X-ray image defect classification network called TireNet was proposed. A new network realizes the representation of defect feature as a part of the downstream classification network module. TireNet achieved a recall rate of 94.7% on a data set composed of more than 10 defect types. A convolutional network based on supervised feature embedding was proposed in [23]. This method effectively improved the accuracy of tire X-ray image classification based on AlexNet. According to the above introduction, the application of deep learning model in visual detection tasks mainly focuses on distinguishing the types of defects. Few studies have been done to determine the level of defects. The task of defect detection in tire factory is to judge defect level on the basis of defect identification. Unqualified products are usually classified as defective products or scraps according to the degree of defect. Thus, the second research motivation of this paper is the application of deep learning network for accurate judgment of defective products and scraps.

Error defects tend to occur in the bead toe. Whether an error defect occurs is determined by the maximum pixel width and minimum pixel width of the bead toe. Moreover, in the actual detection of toe error defects, defective products and scraps need to be distinguished. Therefore, the current

method of manual visual inspection requires accurate measurement to judge the defect level after initial assessment. Certainly, it seriously affects the defect detection speed of each tire. Fortunately, semantic segmentation network can realize pixel level classification. That is, it is satisfied to extract the bead toe and measure the pixel width. Naturally, semantic segmentation network is proposed to replace manual task of defect level determination. The idea of semantic segmentation can be traced back to the proposal of FCN [24]. Later, U-Net [25] achieved very high accuracy by using a specific jump connection structure, followed by many improved forms, such as [26–29]. With the optimization of the speed of semantic segmentation network, the concept of real-time semantic segmentation network was proposed. E-Net [30] designed a relatively small coding layer and a relatively large decoding network. This kind of network was characterized by reducing the number of parameters and greatly increasing the network speed. Bisenet [31] proposed a dual path network structure, in which two paths obtained high-resolution features and enough receptive field. STDC-Net [32] abandoned the method of extracting spatial information and context information separately. Instead, single-stream networks were used. The learning of spatial information was integrated into the bottom layer of the network, which further speeded up the output of the network. However, STDC-Net still has the possibility of improving the accuracy of pixel-level segmentation. Naturally, the third research motivation of this paper is to design a lightweight semantic segmentation network. The network has tradeoffs between segmentation accuracy and inference speed. To be specific, the detection of bread toe error is superior to manual visual inspection in both speed and accuracy.

Aiming at the above three research motivations, this paper proposes a lightweight semantic segmentation network to realize the detection of bead toe error. In this paper, the shallow and deep texture information is stored in the feature map of each stage. We fuse multi-scale feature information in the decoding module. Finally, using size extensions, the output is decided by sufficient spatial and contextual information. Besides, an auxiliary supervision structure is added to improve the precision of class boundary segmentation without increasing model parameters.

The main contributions of this paper are summarized as follows:

1). A semantic segmentation network based on encoder and decoder structure is proposed. In the encoder, a weighted dense connection (WSTDC) module is proposed. In the decoder, a feature fusion structure using a chained residual pooling (CRP) is proposed. This structure uses pooling operation and small convolution kernel to replace the large convolution kernel. Spatial and contextual information is then expressed.

2). We design an auxiliary boundary supervision method. The labels composed of border and background are designed based on the original three categories of labels. The auxiliary loss function calculates the difference between the feature maps and the boundary labels in the coding stage to correct the attention of the coding operation to the boundary. The main loss and auxiliary loss are super imposed to complete the network training in the end.

3). Based on the calculation of mIoU, a more accurate index to measure the class boundary accuracy is proposed, that is, L-mIoU. On a self-made tire X-ray dataset, a result of the lightweight semantic segmentation network is impressive. To be more precise, 92.4% L-mIoU and 97.1% mIoU are achieved at 1.0 s.

4). A deep learning method is applied to bead toe defect detection in this paper. It is the first method to identify whether an error defect occurs by calculating the boundary of the bead toe. The effect of on-line detection was simulated on a data set composed of defect samples and normal samples.

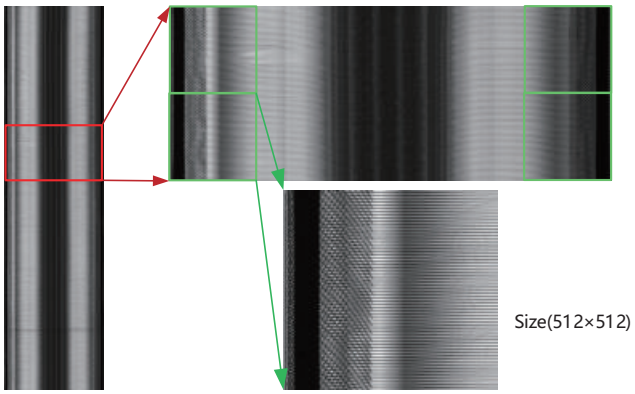

Figure 1. From the original tire X-ray image to the input image of a lightweight semantic segmentation network.

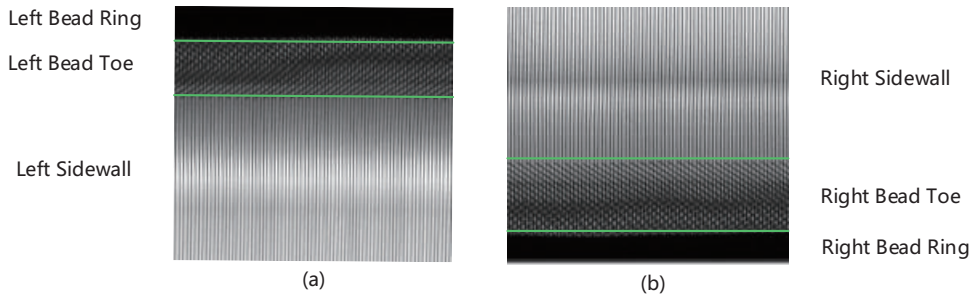

Figure 2. Tire images with resolution 512 × 512. (a) A part of the X-ray image on the left side of the crown. (b) A part of the X-ray image on the right side of the crown.

2. Proposed method

Inspired by STDC-Net, in this work, we propose a lightweight semantic segmentation network. The height of a complete tire X-ray image is usually thousands of pixels and the width is 2469 pixels. If the original image is directly fed into the network, on the one hand, the operation speed is very bad, on the other hand, large image is more difficult to label. In order to complete the tire toe defect detection and reduce the computational cost, the key is to cut the area containing the left tire toe and the right tire toe. Considering that the width of the bead toe varies with different types, the 512 × 512 image resolution is finally determined to ensure that the clipped subgraph includes bead ring, bead toe and sidewall, as shown in Figure 1. Obviously, each input to the network is a subgraph. This lightweight semantic segmentation network learns the texture feature of each area of the tire X-ray subgraph. As a result, the original input X-ray image is divided into three parts, as shown in Figure 2, representing the bead ring, bead toe, and sidewall respectively. The architecture of the network is mainly composed of an encoder, a decoder and an auxiliary boundary supervision module, as shown in Figure 3. In the network structure, the basis is the encoder, including multiple WSTDCs to achieve feature representation. Each WSTDC consists of a stack of multiple 3 × 3 convolution layers followed

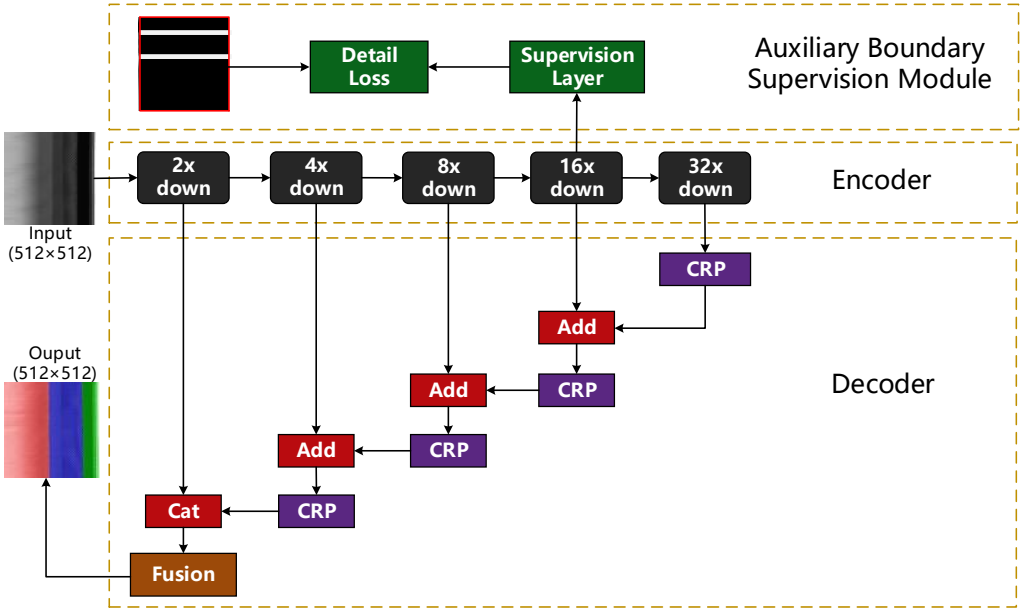

Figure 3. A lightweight semantic segmentation network designed in this paper.

by a nonlinear activation unit (Relu) and a batch standardization unit (BN). It is worth noting that the output of each convolution operation in WSTDC will be recorded. The output of this module is the combination of all the convolution operation records. After going through multiple WSTDCs, the image is eventually reduced to 1/32 of its original size. Each time the size is reduced, a new feature map is generated. The encoder outputs a number of feature maps of different sizes to form a complete spatial information semantic information. In the decoder, the expansion mechanism solves the problem of how to fuse different size feature maps from the encoder. The deep feature maps are merged with the shallow feature maps after passing through 5×5 pooling layers and 1×1 convolution layers. Crucially, the representation ability of deep spatial features affects the segmentation effect of boundary details. Based on this observation, we propose an auxiliary boundary supervision module to guide the deep module in the encoding stage by learning boundary details. The details of each module in the network are shown in the following description.

2.1. Encoder

2.1.1. An improved STDC module

According to the steps described above, the extraction of texture feature requires the use of advanced down-sampling modules. The recently proposed STDC module has gained our attention due to its strong extraction ability and small number of training parameters. So many improved STDC modules are stacked to form a forward feature learning path. There are two WSTDC modules involved in the path, referred to as WSTDC1 and WSTDC2, as shown in Figure 4. Including a convolutional layer with 1×1 filters and 4 layers with 3×3 filters are the same points of WSTDC1 and WSTDC2. The output image size of WSTDC1 is consistent with the input image size, while the output size of WSTDC2 is half of the input size. In this paper, WSTDC1 and WSTDC2 are combined to form

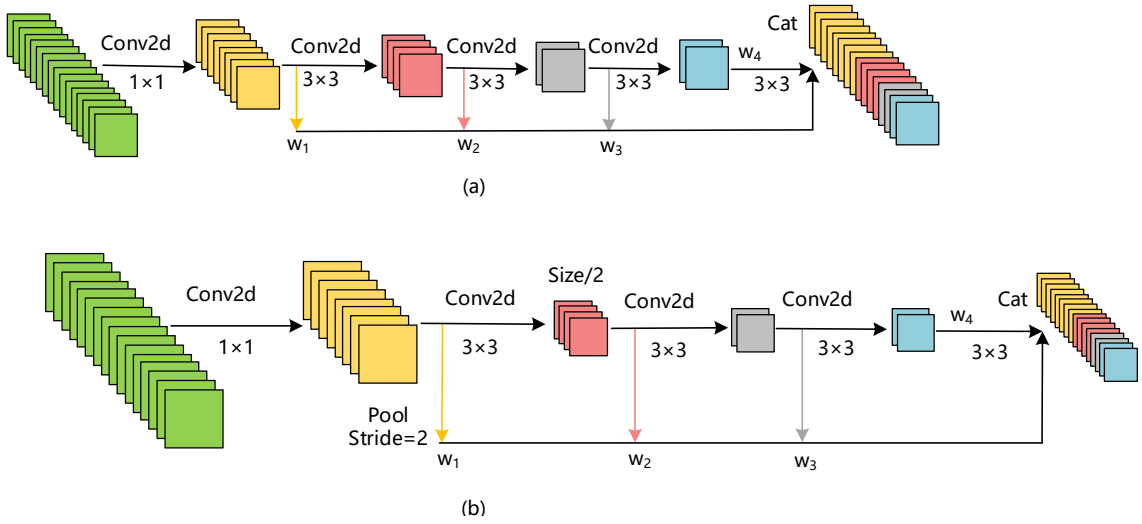

Figure 4. A lightweight semantic segmentation network designed in this paper. (a) represents the structure of WSTDC1(the output size is the same as the input size). (b) represents the structure of WSTDC2(the output size is reduced to half of the input size).

a standard down-sampling module called double feature extraction blocks (DFEB). The key to this semantic segmentation network is that shallow blocks contain a lot of spatial information. As for deep feature maps, they mainly represent contextual information. To speed up the convergence speed of the training parameters in the DFEB and reduce redundant parameters, the output of each convolutional layer is fused to the output layer through jump connections. The weights of the intermediate feature maps are learnable. That is to say, the weight parameters change dynamically during the training process.The source of the output is shown in the following equation:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & w_3 & 0 \\ 0 & 0 & 0 & w_4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \tag{2.1}$$

$$y_{out} = f(y_1, y_2, y_3, y_4) \tag{2.2}$$

where w_1, w_2, w_3, w_4 denotes the weights of the intermediate feature maps (i.e., x_1, x_2, x_3, x_4) from left to right, y_{out} denote WSTDC modules output, f indicates that the fusion operation is completed by concatenation.

Therefore, the advantages of the WSTDC structure are mainly reflected in the fewer parameters and expression multi-scale features. The concatenation of multiple intermediate feature maps constitutes the output channel of the WSTDC module. Compared with the conventional channel transform, less convolution computation is required under the requirement of the same number of output channels. Convolution layers of different depths extract texture information of different sizes, that is, the texture contrast of each region is more distinct.

Table 1. The structure of each stage in the encoder.

Stage	Operation	Output size	Kernel size	Stride
stage1	Conv2d	$64 \times 256 \times 256$	3	2
stage2	Conv2d	$128 \times 128 \times 128$	3	2
stage3	Double Conv2d	$256 \times 128 \times 128$	3	1
stage4	WSTDC1/WSTDC2	$256 \times 64 \times 64$	3	2/1
stage5	WSTDC1/WSTDC2	$512 \times 32 \times 32$	3	2/1
stage6	WSTDC1/WSTDC2	$1024 \times 16 \times 16$	3	2/1

2.1.2. Design of the encoder

Considering the relationship between the richness of texture feature, the number of DFEBs is set to 3. The encoder structure is shown in Table 1. On the whole, X-ray image processing with a resolution of $1 \times 512 \times 512$ consists of 6 stages. Stages 1 and 2 use a structure that only contains a convolutional layer followed by a batch normalization and a non-linear activation unit. Such a simple design shows favorable feature extraction capabilities in the following experiments. Stage 3 includes 2 convolutional layers that behave as channel expansions without size changes. From stages 4 to 6, three DFEBs are used to refine the original input feature map, so that the image size is reduced to 1/8, 1/16 and 1/32 in turn. In the end, the number of channels eventually expands to 1024 in stage 6. In the experiment, we found that the combination of WSTDC1 and WSTDC2 can meet the needs of texture expression. More importantly, this structure has more advantages in terms of controlling the number of parameters.

2.2. Decoder

2.2.1. CRP module

Inspired by [33], the optimization of the chained residual pooling module is shown in the reduction of the convolution kernel size. Specifically, a 3×3 convolution kernel is replaced with consideration for computational speed. Experiments show that the 1×1 convolution does not significantly lose the texture information contained in the feature maps. In fact, the CRP module is a stack of 5×5 pooling operation and 1×1 convolution operation, as shown in Figure 5. It should be noted that the number of stacked layers is set to 3 in this paper. The 1×1 convolution kernel instead of the 3×3 convolution kernel loses the local receptive field to a certain extent. However, the use of the large pooling kernel and the jump connection weakenes this kind of loss.

2.2.2. Design of the decoder

The output feature maps of down-sampling encoding stages 1 to 6 are the input source of the up-sampling fusion operation. We recorded the decoder details as shown in Table 2. In summary, the output feature maps of the current stage is fused with the feature maps of the previous stage, that is, the up-sampling decoding operation is realized by constructing a U-shaped structure. It should be mentioned that the output feature maps of the two stages need to be transformed through CRP module after being merged. In our structure, only the output feature maps of stage 1 is not fused by pixel-level addition operation. Instead, pixel-level concatenation operation is used. Add operation and Cat operation

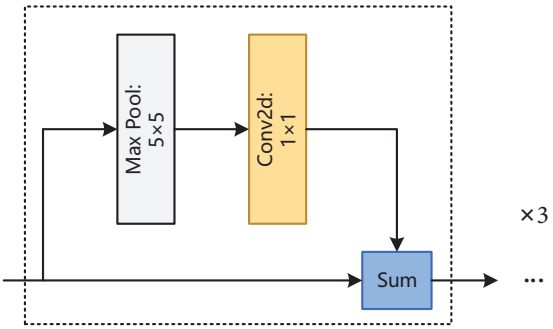

Figure 5. A new CRP module.

Table 2. Decoding structure embedded in the lightweight network. The two input branches come from the output feature maps of two adjacent encoding stages.

Input size_1	Operation_1	Input size_2	Operation_2	Fusion	Output size
1024 × 16 × 16	1 × 1 Conv2d CRP Upsample	512 × 32 × 32	1 × 1 Conv2d	Add/CRP 1 × 1 Conv2d Upsample	256 × 64 × 64
256 × 64 × 64	1 × 1 Conv2d	256 × 64 × 64	*	Add/CRP 1 × 1 Conv2d Upsample	128 × 128 × 128
256 × 128 × 128	1 × 1 × Conv2d	128 × 128 × 128	*	Add/CRP 1 × 1 Conv2d Upsample	64 × 256 × 256
64 × 256 × 256	*	64 × 256 × 256	*	Cat 3 × 3 Conv2d Upsample	32 × 512 × 512
Pixel-level classifier					

represent pixel-level addition and concatenation, respectively. Also, a double-layer 3 × 3 convolution is placed behind the feature fusion layer. When it comes to the benefits of the concatenation, the direct advantage is that the number of channels can be rapidly expanded. Obviously,the higher dimensional space is suitable for separating the details of texture differences. Finally, the pixel-level classifier makes the number of channels to 3.

2.3. An auxiliary boundary supervision module

In order to improve the accuracy of class boundary supervision in semantic segmentation tasks, many contributors put forward their own ideas. These methods generally train a separate network for boundary supervision. Inspired by MSFNet [34], a module to supervise the background and boundary of the encoder is designed. In detail, an auxiliary supervision module is placed behind the stage 5, where the feature maps have compressed spatial information. Before using the auxiliary boundary supervision module, we use sobel operator to generate labels composed of boundary and background according to the region marker of three textures, as shown in Figure 6. The vertical boundary is concerned while ignoring the horizontal Sobel operator. Subsequently, all boundary distribution maps are obtained. Finally, the value 0.9 is used as the threshold value for transforming boundary a distribution

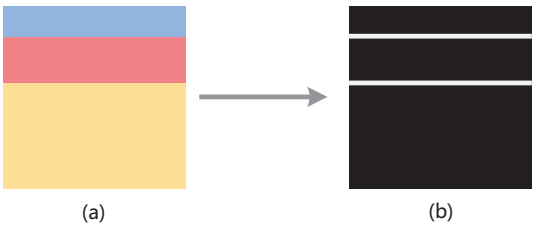

Figure 6. (a) Position marker for three textures of 512 × 512 tire image. (b) A label consisting of a boundary and a background that contains two categories.

map into a new label, that is, this kind of label containing boundaries and backgrounds can be constructed. When a 64-channel feature map is converted into a single-channel boundary detail map, the two-layer 3 × 3 convolution in the auxiliary boundary supervision model is used. In order to optimize the boundary prediction capability of the encoder, the cross-entropy loss and DICE loss are calculated based on the true boundary and the predicted boundary. What needs to be emphasized is that the auxiliary boundary supervision greatly improves the bias of deep network to boundary difference.

2.4. Loss function

The output of the encoder at stage 5 is supervised by the auxiliary loss function to enhance the spatial boundary information in the deep network. In addition, the main loss function supervises the output of the decoder. In detail, the cross entropy is adopted in the main loss function, while the auxiliary loss includes DICE loss and cross entropy loss. As shown in Eq (2.3), a weight coefficient is set to balance the main loss and auxiliary loss. In this paper, a good semantic segmentation effect can be observed when $\beta = 0.96$.

$$loss = L_m\left(y_1, y_1^{'}\right) + \beta L_a\left(y_2, y_2^{'}\right) \tag{2.3}$$

where y_1 and $y_1^{'}$ represent the labels of three categories and the output of the decoder respectively. The other term represents the border label and the output of the auxiliary supervision module.

3. Experiment

In this section, we conduct experiments on real tire X-ray image data. The main work is to compare the performance of this lightweight network with several classic networks in terms of speed and accuracy. Considering that the accuracy of the bead toe boundary measurement depends entirely on the positioning effect of the toe, semantic segmentation index indirectly reflects the detection effect of toe error defects. Essentially, the output of the semantic segmentation network is pixel-level markup of bead ring, toe and sidewall. As shown in Figure 7, the output subgraphs belonging to the same original tire X-ray image are joined together, where the number of subgraphs is represented by N, thus the coordinates of the left and right toe boundaries can be calculated. When the toe error defect exists, the coordinate difference will be abrupt, as shown in Figure 8. Toe error defects are considered to occur when the coordinate difference between position A and position B is greater than the set standard. It

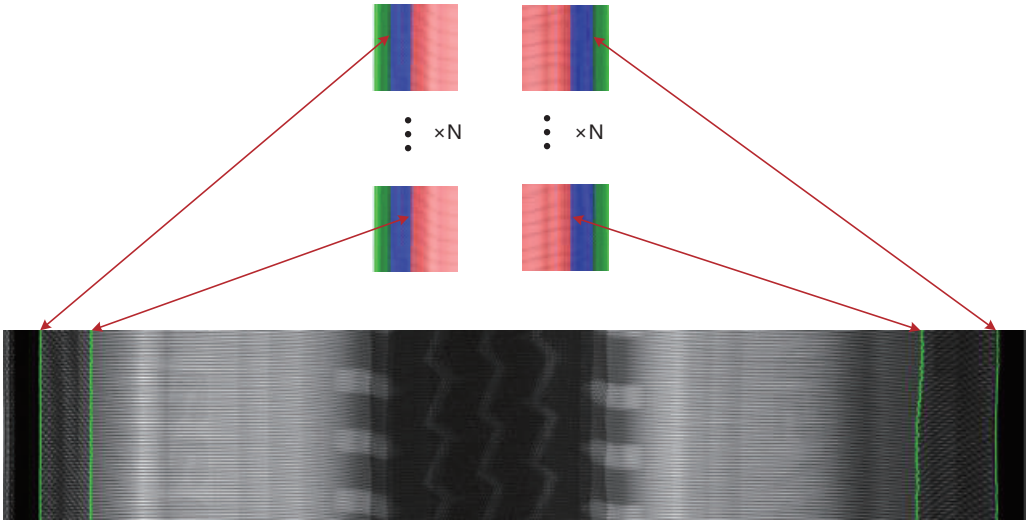

Figure 7. Method for constructing bead toe boundary coordinates.

should be noted that the standard of defects is expressed in millimeters, so pixel coordinates are converted to millimeters according to the scale bar. According to the standard, it can be further determined whether the defect level is a defective product or a scrap. Besides, bead toe error defects may occur on both the left and right side of the tire.

3.1. Network running environment

In this work, all hardware environments are Intel(R) Core (TM) i7-8700K CPU @ 3.70GHz, 16.0 GB RAM and NVIDIA GEFORCE GTX 2080 GPU. Each network participating in training uses a stochastic gradient descent (SGD) with a momentum of 0.9, a batch size of 4, and a weight decay of 1e-5. The initial learning rate is 1e-4. Crucially, the goal of semantic segmentation task is to label the

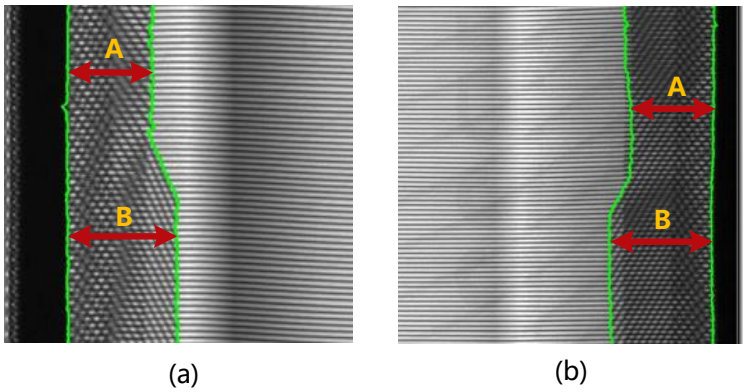

Figure 8. Calculation principle of toe error defect.

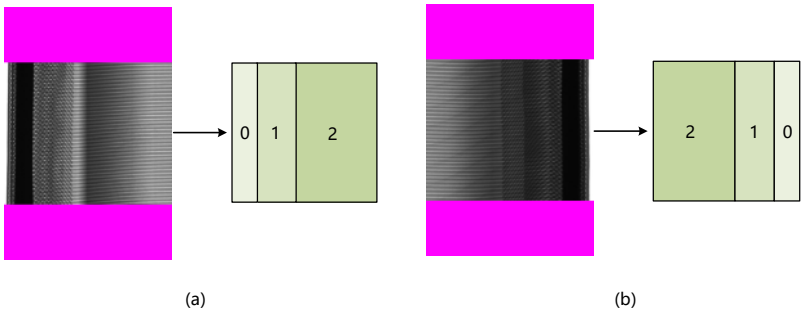

Figure 9. (a) Input image to be marked on the left. (b) Input image to be marked on the right.

input image into three categories, so it is sufficient that the total training epoch is set to 30.

3.2. A data set

The data set is divided into training set and test set, accounting for 1218 and 854, respectively. In the test set, 554 images are used for the evaluation of semantic segmentation metrics. Accordingly, the remaining 300 images are used to measure the accuracy of defect diagnosis. In detail, the left bead toe and the right bead toe each account for one-half of the images. The ground-truth of input image contains three parts, in which the bead ring, the bead toe and the sidewall are represented by class name 0, 1 and 2 respectively. We know that the height of the X-ray image reaches 512 pixels, so using the original image with a resolution of 512×512 will cause difficulty in marking the top and bottom when making the label. To this end, 200 redundant pixels were filled at the top and bottom of the original image in the sample making process, as shown in Figure 9. Once the marking is complete, the step of eliminating redundant pixels is very easy.

3.3. Accuracy evaluation index—L-mIoU

mIoU is to evaluate the prediction accuracy of semantic segmentation network for all output pixels. It has a good effect on the evaluation of conventional pixel-level classification tasks. However, the lightweight semantic segmentation network proposed here is intended to locate the boundaries of the bead toe. In other words, we are more concerned with the classification effect of pixels around the toe boundary. The bead ring is particularly different from the toe and sidewall, which can be distinguished from the gray value. Therefore, pixel-level segmentation of bead ring is not challenging for most semantic segmentation networks. However, the difference in pixel gray levels between the sidewall and the bead toe is no longer obvious, that is, the gray level information cannot be simply considered. In this case, the segmentation accuracy of the toe and sidewall boundary is the key to evaluating the semantic segmentation network. Our local mIoU came into being, it only pays attention to the local real label with a width of 60 pixels. Local real labels come from two boundary regions as does network prediction output. Figure 10 shows the recording rules for the local prediction output.

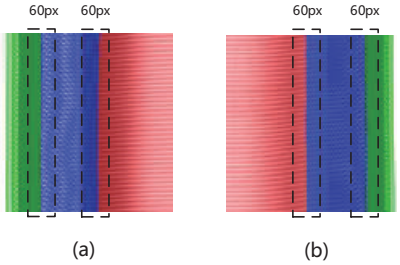

Figure 10. (a) The prediction of the left bead toe by the semantic segmentation network. (b) The prediction of the right bead toe.

Table 3. Comparisons with other popular networks on the test set.

Network	mIoU (%)	L-mIoU (%)	Parameters (MB)	Running time (s)
U-Net [25]	96.8	91.9	65.87	1.8
STDC-Net [32]	92.5	81.9	52.96	0.8
RefineNet-LW-50 [33]	96.9	91.8	104.18	1.2
Our method **(without supervision)**	**96.1**	**90.6**	**55.58**	**1.0**

3.4. Comparison of network speed and accuracy

U-Net is a high-precision semantic segmentation network, but the training parameters are more complicated. STDC-Net is a network structure oriented to real time semantic segmentation scenes. It compresses parameters and adversely affects the segmentation effect of the target boundary. We compared the pixel-level classification accuracy of the U-Net, STDC-Net, RefineNet and our lightweight network on the test set. It can be found from Table 3 that the mIoU value is significantly higher than the L-mIoU value in each model. For example, the mIoU in the U-Net model reaches 96.8%, while the L-mIoU is 91.9%. It shows that for the evaluation of class boundaries, L-mIoU is closer to the true capabilities of the network. As a real-time semantic segmentation network, STDC-Net has certain deficiencies in the segmentation of boundaries. Without auxiliary boundary supervision, our model is lower than the U-Net model in both mIoU and L-mIoU. Refinenet-LW-50 is a lightweight semantic segmentation network designed on the basis of RefineNet, which has a good effect on the expression of spatial information. In the network design of this paper, we use the feature fusion module of Refinenet-LW-50 for reference.

In addition, the size and operating speed of the lightweight network are also evaluated. The running time of the network is averaged by testing 554 X-ray images with a resolution of 8000×2469. As for the running time, the original image is first segmented into subgraphs containing the bead toe. Then predicting them at the pixel level. Finally, the boundary coordinates of the bead toe are recorded. As shown in Table 3, U-Net has excellent performance in the accuracy mentioned above, but the large number of parameters leads to the longer running time. Despite the STDC-Net is very fast, the accuracy of pixel-level prediction is not up to our requirements. RefineNet-LW-50 is inferior to the lightweight network proposed in this paper in terms of speed and accuracy. With training parameters of 55.58 MB,

Table 4. Position test of auxiliary boundary supervision module. S1, S2, S3, S4 and S5 respectively indicate that the auxiliary supervision module is placed after the output of the encoder from stage 1 to stage 5. In the three classical networks, the auxiliary supervision module is added to the encoder after the feature map with a resolution of 32×32.

Network	Input resolution	mIoU(%)	L-mIoU(%)
UNet + Boundary loss	$1 \times 512 \times 512$	97.0	92.2
STDC-Net + Boundary loss	$1 \times 512 \times 512$	92.7	82.3
RefineNet-LW-50 + Boundary loss	$1 \times 512 \times 512$	97.1	92.0
Our method + Boundary loss(S1)	$1 \times 512 \times 512$	96.6	91.6
Our method + Boundary loss(S2)	$1 \times 512 \times 512$	96.5	91.1
Our method + Boundary loss(S3)	$1 \times 512 \times 512$	96.5	91.4
Our method + Boundary loss(S4)	$1 \times 512 \times 512$	96.9	91.6
Our method + Boundary loss**(S5)**	$1 \times 512 \times 512$	**97.1**	**92.4**

this excellent lightweight semantic segmentation network is 40% faster than U-Net. In summary, speed and accuracy can be traded off using a lightweight semantic segmentation network.

In order to continue to enhance the performance of the new method, as shown in Table 4, the auxiliary boundary supervision module is added. The addition of auxiliary supervision modules to three classic semantic segmentation networks can help improve the pixel-level segmentation effect. However, the classical network does not design a special module for boundary extraction, resulting in a weaker improvement effect than the network structure proposed in this paper. Our method (S5) means that the auxiliary boundary supervision is placed after 1/16 down sampling. At this time, L-mIoU reaches 92.4%, which exceeds the pixel-level segmentation accuracy of U-Net. In the experiment, we test the effect of the auxiliary boundary supervision function at different positions. Adding auxiliary supervision modules after stage 1 to stage 5 will help advance the accuracy of boundary segmentation. Actually, for the encoder, the resolution of the shallow feature maps is relatively large, so that the loss of spatial information is less. Instead, the spatial information in the deep feature maps has severe loss. It is necessary to capture both shallow spatial information and deep semantic information at the same time. Therefore, this paper designs a decoding structure with preference for boundary information. The experimental results shows that the two branches of feature fusion (shallow feature maps and deep feature maps) by deep auxiliary supervision have rich spatial information.

Of course, the visual effect comparison after adding auxiliary boundary supervision module is indispensable. As shown in Figure 11, overwhelmingly, the results show that after adding the auxiliary supervision module, the segmentation effect of various algorithms in the boundary area has been greatly improved. There are encouraging signs that segmentation effect of our method near the boundary is almost consistent with the marked map.

3.5. Defect diagnosis effect verification

Artificial judgment results only distinguish between normal and defect types. However, artificial evaluation result need to further distinguish defect levels on the basis of identifying defects. For a dataset of 300 tire x-ray images, the normal sample is exactly two thirds. Scrap and defective products

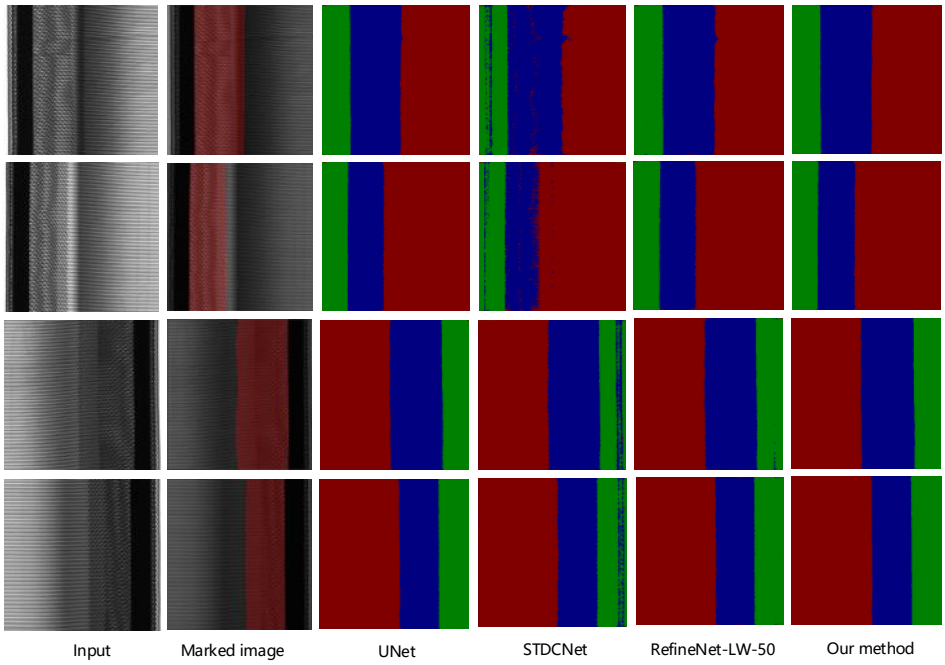

Figure 11. Comparison of visual effects between our method and popular methods.

occupy 20 and 80 respectively.

$$CR = \frac{HJ \cap AJ}{S} \tag{3.1}$$

$$PR = \frac{HLE \cap ALE}{S} \tag{3.2}$$

$$RMR = \frac{N}{S_d} \tag{3.3}$$

In Eq (3.1), CR is defined as correct rate. HJ and AJ represent the artificial judgement result and the algorithm judgement result for the input image, respectively. S stands for the total number of input images. In Eq (3.2), PR is defined as Precision ratio. HLE represent the artificial evaluation result for two defect levels. ALE stands for the evaluation result of the algorithm for defect levels. In Eq (3.3), RMR is defined as rate of missing repor. N represents the number of defect samples that were not identified. S_d stands for the number of defect samples.

After the position of each region is predicted according to the semantic segmentation network, the coordinates of the border of the toe can be obtained. Then, calculated by the ratio of the distance on the image to the actual distance, the defect level is judged against the defect standard. Due to the known error defect distribution of tire X-ray images, the advantages of our proposed algorithm in automatic detection can be directly verified. As shown in Table 5, in the three evaluation indexes of bread toe error detection, the algorithm proposed in this paper is at an advanced level, which are 91.3, 90.4 and 1.4%, respectively.

Table 5. Comparison of toe error detection results.

Network	CR (%)	PR (%)	RMR(%)
UNet + Boundary loss	89.7	88.3	1.8
STDC-Net + Boundary loss	75.4	71.6	3.5
RefineNet-LW-50 + Boundary loss	88.5	87.9	2.2
Our method + Boundary loss**(S5)**	**91.3**	**90.4**	**1.4**

4. Conclusions

In this paper, we have rediscussed the classical semantic segmentation network, U-Net and real-time semantic segmentation network, STDC-Net. On this basis, the shortcomings of U-Net in speed and STDC-Net in accuracy have been optimized. A new lightweight semantic segmentation network has been proposed, which is similar to U-Net in precision and at the same time similar to STDC-Net in speed. Our encoder is the fine tuning of STDC-Net. Talking about the idea of the decoder, it is to fuse the deep feature maps with the shallow feature maps, so that the feature maps contain multi-scale spatial information. It is key to search the boundary of the bead toe for error defect detect. Therefore, the pixel-level boundary segmentation capability is an inevitable requirement for the network. An auxiliary supervision method on the vertical direction has been proposed to enhance the pixel-level classification performance for the class boundary. Ultimately, the bead toe boundary coordinates have been located by running our lightweight semantic segmentation network. By comparing the difference between the edge coordinates and the standard, it can be judged whether there is a toe error defect and the defect level.

Conflict of interest

The authors declare there is no conflict of interest.

References

1. *Centers for Disease Control and Prevention, Leading Causes of Death*, National Center for Health Statistics, 2020. Available from: https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm.
2. *National Highway Traffic Safety Administration, Safety and Savings Ride on Your Tires*, Always Perform Proper Maintenance, 2020. Available from: https://www.nhtsa.gov/es/tires/safety-and-savings-ride-your-tires.
3. L. Sun, L. He, C. Hai, X. Han, Z. Gui, M. Yang, Design of imaging system and tomography detection method for radial tires structure under X-ray short-scan mode, *IEEE Trans. Instrum. Meas.*, **70** (2021), 1–12. https://doi.org/10.1109/TIM.2021.3118098
4. G. Fortunato, V. Ciaravola, A. Furno, M. Scaraggi, B. Lorenz, B. N. Persson, Dependency of rubber friction on normal force or load: theory and experiment, *Tire Sci. Technol.*, **45** (2017), 25–54. https://doi.org/10.2346/tire.17.450103

5. J. J. Castillo Aguilar, J. A. C. Carrillo, A. J. G. Fernández, S. P. Pozo, Optimization of an optical test bench for tire properties measurement and tread defects characterization, *Tire Sci. Technol.*, **17** (2017), 707. https://doi.org/10.3390/s17040707
6. X. Cui, Y. Liu, Y. Zhang, C. Wang, Tire defects classification with multi-contrast convolutional neural networks, *Int. J. Pattern Recogn.*, **32** (2018), 1850011. https://doi.org/10.1142/S0218001418500118
7. Q. Guo, C. Zhang, H. Liu, X. Zhang, Defect detection in tire X-ray images using weighted texture dissimilarity, *J. Sens.*, **32** (2016), 2016. https://doi.org/10.1155/2016/4140175
8. Y. Zhang, T. Li, Q. Li, Detection of foreign bodies and bubble defects in tire radiography images based on total variation and edge detection, *Chin. Phys. Lett.*, **30** (2013), 084205. https://doi.org/10.1088/0256-307X/30/8/084205
9. Y. Zhang, T. Li, Q. Li, Defect detection for tire laser shearography image using curvelet transform based edge detector, *Opt. Laser Technol.*, **47** (2013), 64–71. https://doi.org/10.1016/j.optlastec.2012.08.023
10. Y. Zhang, D. Lefebvre, Q. Li, Automatic detection of defects in tire radiographic images, *IEEE Trans. Autom. Sci. Eng.*, **14** (2015), 1378–1386. https://doi.org/10.1109/TASE.2015.2469594
11. G. Zhao, S. Qin, High-precision detection of defects of tire texture through X-ray imaging based on local inverse difference moment features, *Sensors*, **18** (2018), 2524. https://doi.org/10.3390/s18082524
12. S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, S. Yu, A survey: Deep learning for hyperspectral image classification with few labeled samples, *Neurocomputing*, **448** (2021), 179–204. https://doi.org/10.1016/j.neucom.2021.03.035
13. K. Lan, G. Li, Y. Jie, R. Tang, L. Liu, S. Fong, Convolutional neural network with group theory and random selection particle swarm optimizer for enhancing cancer image classification, *Math. Biosci. Eng.*, **18** (2021), 5573–5591. https://doi.org/10.3934/mbe.2021281
14. Y. Liu, P. Sun, N. Wergeles, Y. Shang, A survey and performance evaluation of deep learning methods for small object detection, *Expert Syst. Appl.*, **172** (2021), 114602. https://doi.org/10.1016/j.eswa.2021.114602
15. H. Ni, M. Wang, L. Zhao, An improved Faster R-CNN for defect recognition of key components of transmission line, *Math. Biosci. Eng.*, **18** (2021), 4679–4695. https://doi.org/10.3934/mbe.2021237
16. Q. Zhou, X. Wu, S. Zhang, B. Kang, Z. Ge, L. J. Latecki, Contextual ensemble network for semantic segmentation, *Pattern Recognit.*, **122** (2022), 108290. https://doi.org/10.1016/j.patcog.2021.108290
17. W. Lu, J. Chen, F. Xue, Using computer vision to recognize composition of construction waste mixtures: A semantic segmentation approach, *Resour. Conserv. Recycl.*, **178** (2022), 106022. https://doi.org/10.1016/j.resconrec.2021.106022
18. R. Ren , T. Hung, K. C. Tan, A generic deep-learning-based approach for automated surface inspection, *IEEE Trans. Cybern.*, **48** (2018), 929–940. https://doi.org/10.1109/TCYB.2017.2668395

19. W. Lu, J. Chen, F. Xue, Using computer vision to recognize composition of construction waste mixtures: A semantic segmentation approach, *Resour. Conserv. Recycl.*, **178** (2022), 106022. https://doi.org/10.1016/j.resconrec.2021.106022

20. K. He, X. Zhang, S. Ren, J. Su, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, (2016), 770–778. https://doi.org/10.1109/cvpr.2016.90

21. L. Yang, H. Wang, B. Huo, F. Li, Y. Liu, An automatic welding defect location algorithm based on deep learning, *NDT E Int.*, **120** (2021), 102435. https://doi.org/10.1016/j.ndteint.2021.102435

22. Y. Li, B. Fan, W. Zhang, Z. Jiang, TireNet: A high recall rate method for practical application of tire defect type classification, *Future Gener. Comput. Syst.*, **125** (2021), 1–9. https://doi.org/10.1016/j.future.2021.06.009

23. Y. Zhang, X. Cui, Y. Liu, B. Yu, Tire defects classification using convolution architecture for fast feature embedding, *Int. J. Comput.*, **11** (2018), 1056–1066. https://doi.org/10.2991/ijcis.11.1.80

24. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, (2015), 3431-3440. https://doi.org/10.1109/cvpr.2015.7298965

25. O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, (2015), 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

26. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: A deep convolutional encoder-decoder architecture for image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.*, **39** (2017), 2481–2495. https://doi.org/10.1109/tpami.2016.2644615

27. C. Peng, X. Zhang, G. Yu, G. Luo, J. Sun, Large kernel matters–improve semantic segmentation by global convolutional network, in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, (2017), 4353-4361. https://doi.org/10.1109/CVPR.2017.189

28. G. GhiasiEmail, C. C. Fowlkes, Laplacian pyramid reconstruction and refinement for semantic segmentation, in *European Conference on Computer Vision*, Springer, (2017), 519-534. https://doi.org/10.1007/978-3-319-46487-9_32

29. V. Badrinarayanan, A. Kendall, R. Cipolla, A nonlocal deep image prior model to restore optical coherence tomographic images from gamma distributed speckle noise, *J. Mod. Opt.*, **68** (2021), 1002–1017. https://doi.org/10.1080/09500340.2021.1968052

30. A. Paszke, A. Chaurasia, S. Kim, E. Culurciello, Enet: A deep neural network architecture for real-time semantic segmentation, preprint, arXiv:1606.02147.

31. C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, N. Sang, Bisenet: Bilateral segmentation network for real-time semantic segmentation, in *European Conference on Computer Vision*, Springer, (2018), 325–341. https://doi.org/10.1007/978-3-030-01261-8_20

32. M. Fan, S. Lai, J. Huang, X. Wei, Z. Chai, J. Luo, et al., Rethinking BiSeNet for real-time semantic segmentation, in *IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, (2015), 9716–9725. https://doi.org/10.1109/cvpr46437.2021.00959

33. V. Nekrasov, C. Shen, I. Reid, Light-weight refinenet for real-time semantic segmentation, preprint, arXiv:1810.03272.

34. H. Si, Z. Zhang, F. Lv, G. Yu, F. Lu, Real-time semantic segmentation via multiple spatial fusion network, preprint, arXiv:1911.07217.

AIMS Press