IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received December 18, 2020, accepted January 11, 2021, date of publication January 18, 2021, date of current version January 25, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052224

A Method of Steel Bar Image Segmentation Based on Multi-Attention U-Net

JIE SHI[1], KUNPENG WU[1,2], CHAOLIN YANG[1,2], AND NENGHUI DENG[1]

[1]Design and Research Institute Company Ltd., University of Science and Technology Beijing, Beijing 100083, China
[2]Institute of Engineering Technology, University of Science and Technology Beijing, Beijing 100083, China

Corresponding author: Jie Shi (CindyShih1108@126.com)

ABSTRACT Due to it is difficulty to segment the steel bar image in a complex background and with external interference. In this paper, we propose a multi-attention U-Net to segment the steel bar image. First of all, in order to accurately find the steel bar region and filter the noise in the background part, we add the row mean attention module in the decoding path of the U-shaped network by using the characteristic of the constant diameter of the steel bar, which reduces the background to be segmented into steel bar. In addition, an attention branch optimization strategy based on mask attention module is designed, which uses the output of high-level semantics to filter the features of adjacent low-level semantics, which can maintain the continuity of the segmented steel bar region. Secondly, we design an improved loss function for training, which can improve the accuracy of fitting of steel bar width and better optimize the effect of steel bar segmentation. Finally, in order to improve the generalization ability of the multi-attention U-Net method and avoid over fitting, we propose a data augmentation method based on correction deformation to expand the sample database. Compared with standard U-Net, attention U-Net, R2U-Net and DUNet, the experimental results show that the multi-attention U-Net proposed in this paper has higher IoU accuracy in steel bar image segmentation, and this method has real-time performance.

INDEX TERMS Multi-attention U-Net, row mean attention module, attention branch optimization strategy, mask attention module.

I. INTRODUCTION

In recent years, customers have higher and higher requirements on the surface quality of steel bar. Therefore, more and more attention has been paid to the effective surface defect detection and recognition methods of steel bar. However, there are some problems in the surface defect detection of steel bar, such as high temperature, fast running speed, difficult manual detection and poor traceability. In order to solve these problems, scholars have begun to study the methods of steel bar surface defect detection and recognition by using machine vision in recent years. Firstly, the image acquisition equipment with linear CCD camera and laser light source is used to collect the steel bar surface image, and then the machine vision method is used to detect and recognize the defect of the steel bar surface image [1]–[3]. Wu Bin Li *et al.* proposed a new detection method based on local annular contrast (LAC) and a low envelope Weber contrast(LEWC) recognition method for steel bar surface defects [4], [5].

However, during the real rolling process, the shearing machine continuously cuts the steel bar. Under the influence of shear force, the steel bar will appear obvious shaking, which leads to the deformation of the collected steel bar surface image. For this kind of deformation images, the above mentioned steel bar surface defect detection methods can not accurately detect the longitudinal defects on the steel bar surface, such as scratches, cracks, wire drawing, etc. In order to solve this problem, we need to correct the steel bar part in the image. The most important step in the correction process is the segmentation of the steel bar part in the image.

For image segmentation, many methods have been proposed in recent years. MP Dewi *et al.* discussed the image segmentation method based on minimum spanning tree to segment digital image, the disadvantage of this method is only suitable for the image with small size and no noise [6]. Amila akagic *et al.* used Otsu threshold segmentation to segment pavement crack image [7]. Otsu algorithm has a short implementation time, but it is not effective for the process of the high signal-to-noise ratio image. In addition, watershed algorithm is also commonly used in image

The associate editor coordinating the review of this manuscript and approving it for publication was Omar Sultan Al-Kadi.

segmentation [8]–[10]. However, when watershed algorithm processes different images, it is difficult to form a general set of super parameters because of the parameters is needed to be set pertinently. For the last few years, a lot of methods based on deep learning have been proposed one after another. FCN [11], [12] is the first proposed semantic segmentation method, which can achieve end-to-end training. However, due to its neglect of the relationship between pixels in space, the segmentation effect is not fine enough. The U-Net proposed by Olaf ronneberger *et al.* is based on the encoding and decoding structure [13], In the case of a small number of samples, great success has been achieved in the field of medicine. As a result, there are many improved methods for U-Net. Zhou Zongwei *et al.* proposed a nested U-Net++ model [14]. In order to avoid a large number of redundant parameters, attention U-Net [15]–[18], bottleneck feature monitoring (BS) U-Net [19] and ResUNet [20] have also been proposed. In addition, Hao Dong *et al.* applied a 'Soft' Dice based loss function in U-Net, which has a unique advantage that is adaptive to unbalanced samples [21]. Besides, Zaiwang Gu *et al.* proposed a context coding network (CE-Net) to capture more high-level information [22]. MD zahangir Alom *et al.* proposed a recurrent residual convolution neural network (R2U-Net) [23]. Shupeng Liu *et al.* combined ResNet and U-Net, and then proposed a new method called Res-Unet [24], which can segment images with poor quality. He Tang *et al.* proposed a dual densely connected U-shaped neural network (DDU-Net) [25], which can segment lumbar spinal CT image with unclear edges. Qiangguo Jin *et al.* proposed a kind of end-to-end deformable U-Net (DUNet) [26], which can segment the retinal vessels in fundus images and has good generalization ability. Changlu Guo *et al.* proposed a structured dropout U-Net [27], which achieves accurate positioning of the up-sampling, and this method can discard some features of contiguous regions during training, thus it can alleviate the overfitting problem. Ming Zhao *et al.* proposed a semi-automatic segmentation method called snakes method [28], which is based on the U-Net structure. Ibtehaz and Rahman [29] proposed a MultiResUNet model, which can achieve good segmentation results in both 2D and 3D image datasets. Most of the existing improved U-Net methods are used for medical image segmentation, and U-Net can also be used for other image segmentation problems. Daniele Liciotti *et al.* introduces an approach to track and detect people in cases of heavy occlusions based on U-Net3 for semantic segmentation using top-view depth visual data [30], which can learn the high-level representation of image content and obtain high-precision and recall rate. Augustauskas [31] utilized residual connections, atrous spatial pyramid pooling with parallel and "Waterfall" connections, and attention gates to improve the U-Net structure, which can better extract the defect characteristics of pavement and can better segment pavement defects. And some works have applied U-Net to other tasks, such as video object segmentation [32], [33]. These works should be discussed.

In this paper, a multi-attention U-Net is proposed for steel bar image segmentation, which is based on the standard U-Net method. The contributions of this work can be summarized as follows:

- The improvements to the structure of the standard U-Net are as follows, a row mean attention module and the attention branch optimization strategy based on mask attention module are introduced in decoding path, which can reduce the influence of complex background on steel bar segmentation and ensure the continuity of the segmented steel bar.
- An improved loss function is designed in the main branch in decoding path of multi-attention U-Net. By increasing the fitting loss of steel bar width with the certain weight, the segmentation effect can be guaranteed in many aspects, and the risk of over fitting can be reduced to a certain extent.
- In this paper, a data augmentation method based on correction deformation is introduced to expand the sample database, which improves the diversity of the sample dataset and enhances the robustness of the multi-attention U-Net method.
- Experiments show that the multi-attention U-Net model can not only improve the IoU accuracy of steel bar segmentation, but also meet the real-time requirements of practical applications.

The paper is organized as follows: Section II discusses the architecture of the proposed multi-attention U-Net model and performance evaluation metrics. Section III shows the experiments and results. The conclusion and future direction are discussed in Sec. IV.

II. METHODOLOGY

A. THE STRUCTURE OF MULTI-ATTENTION U-NET

The purpose of this study is to establish a semantic segmentation model to extract the steel bar part from the image. Steel bar image mainly includes background part and steel bar part. The steel bar part has two characteristics: one is that the steel bar part is continuous and there is at most one steel bar region appears in the image, the other is that the diameter of the steel bar is constant, so the width of the steel bar part in the image is consistent. In view of these two characteristics, this paper proposes the following improvement schemes for the structure of multi-attention U-Net in decoding path. A row mean attention module is proposed to ensure the width consistency of the segmented steel bar region, which reduces the background to be segmented into steel bars. In addition, an attention branch optimization strategy based on mask attention module is designed to ensure the integrity of segmented image by using high-level semantic features.

The structure of multi-attention U-Net is shown in Fig. 1, which is a U shape network. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box, which is represented by Ci. The size of the image is expressed as Hi × Wi. White box in the

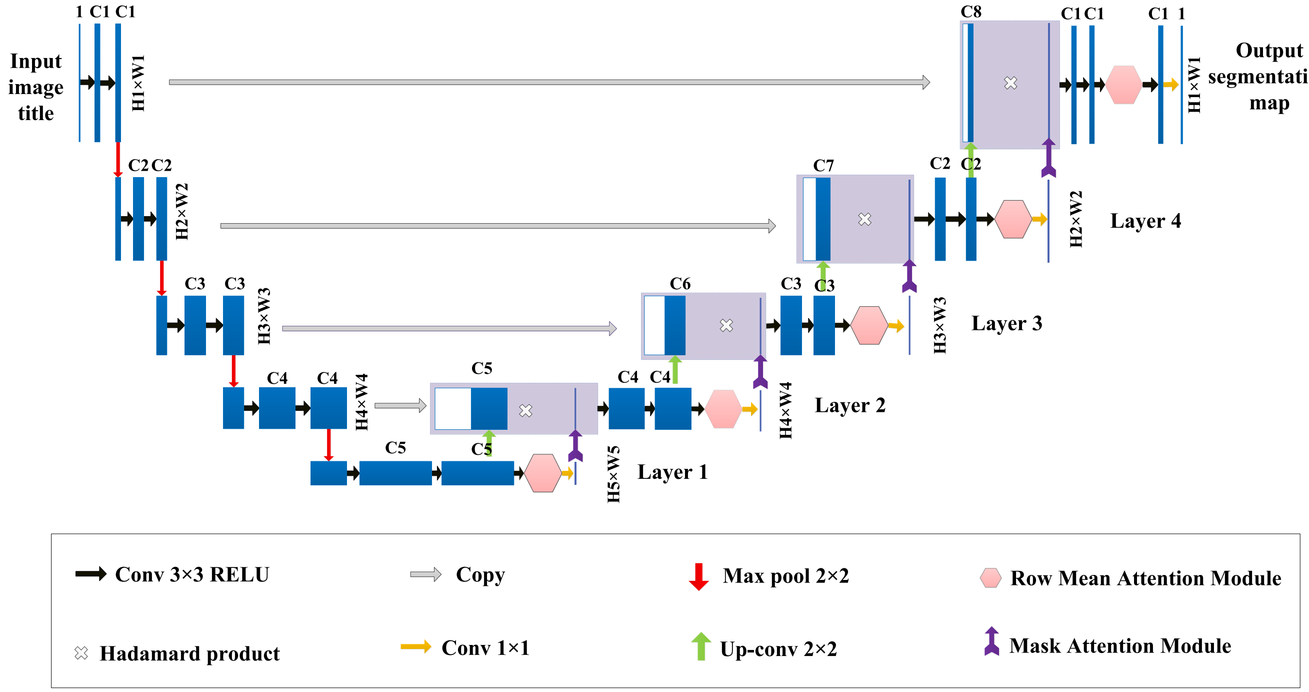


FIGURE 1. The structure of multi-attention U-Net.

decoding path is the feature map copied from the coding path for feature fusion.

In each layer of the decoding path of the multi-attention U-Net, a row mean attention module is added. The purpose of the row mean attention module is to add attention weight value to the steel bar region according to the obtained semantic features, which can increase the weight of the real bar region and reduce the impact of the background part being segmented into steel bar.

The calculation formula of row mean attention module is as follows:

$$rowMean_j = \frac{1}{W \cdot C} \sum_{k=0}^{C} \sum_{i=0}^{W} X_j\,(k, i) \tag{1}$$

$$X^{sign} = sign(X - \mathrm{row}Mean_j) \tag{2}$$

$$rowSign = \frac{1}{W \cdot C} \sum_{k=0}^{C} \sum_{i=0}^{W} X_j^{sign}\,(k, i) \tag{3}$$

$$attWeight_j = \frac{1}{sigmoid(rowSign_j)} \tag{4}$$

$$X_j = X_j attWeight_j \tag{5}$$

$$X = \sum_{j=0}^{H} X_j \tag{6}$$

In the formulas, the image feature size is $H \times W \times C$, $0 \leqq i < W, 0 \leqq j < H, 0 \leqq k < C$. $rowMean_j$ is the feature mean value of the j^{th} row. X is the feature of input. X_j (k,i) represents feature value of (k, i) coordinates in the j^{th} feature plane. X_j denotes the feature plane of the jth row. X^{sign} is the difference value after the sign function conversion. $rowSign_j$ represents the feature mean value of X^{sign} corresponding to the j^{th} row. $attWeight_j$ represents the weight information of the j^{th} row, which obtained through the row mean calculation.

The structure of multi-attention U-Net is shown in Fig. 1, which is a U shape network. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box, which is represented by *Ci*. The size of the image is expressed as *Hi* × *Wi*. White box in the decoding path is the feature map copied from the coding path for feature fusion.

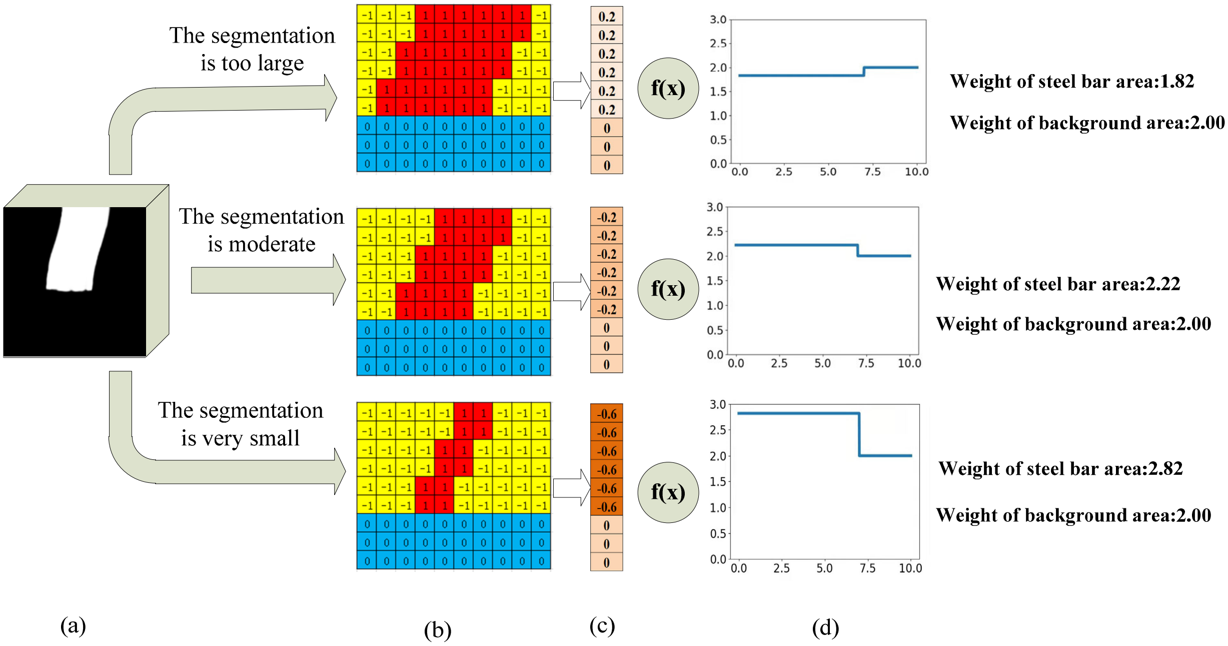
In each layer of the decoding path of the multi-attention U-Net, a row mean attention module is added. The purpose of the row mean attention module is to add attention weight value to the steel bar region according to the obtained semantic features, which can increase the weight of the real bar region and reduce the impact of the background part being segmented into steel bar.

The calculation formula of row mean attention module is as follows:

Step 1: For feature X with the size of H × W × C, to calculate the mean value of each row according to formula (1), a matrix of size H × 1 is obtained.

Step 2: Using feature X to subtract the corresponding $rowMean_j$ to obtain the normalized feature, and then perform sign function operation on it.

Step 3: Use formula (3) to calculate the row mean value of the result obtained by step 2. By performing this step, we can ensure that the background region can get constant weight in different segmentation situations. This step is of great significance for the segmentation of steel bar head images and tail images.


The segmentation is too large
The segmentation is moderate
The segmentation is very small
f(x)
f(x)
f(x)
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0.0
2.5
5.0
7.5
10.0
Weight of steel bar area:1.82
Weight of background area:2.00
Weight of steel bar area:2.22
Weight of background area:2.00
Weight of steel bar area:2.82
Weight of background area:2.00
(a)
(b)
(c)
(d)

−1	−1	−1	1	1	1	1	1	1	−1
−1	−1	−1	1	1	1	1	1	1	−1
−1	−1	1	1	1	1	1	1	−1	−1
−1	−1	1	1	1	1	1	1	−1	−1
−1	1	1	1	1	1	1	−1	−1	−1
−1	1	1	1	1	1	1	−1	−1	−1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

0.2
0.2
0.2
0.2
0.2
0.2
0
0
0

−1	−1	−1	−1	1	1	1	1	−1	−1
−1	−1	−1	−1	1	1	1	1	−1	−1
−1	−1	−1	1	1	1	1	−1	−1	−1
−1	−1	−1	1	1	1	1	−1	−1	−1
−1	−1	1	1	1	1	−1	−1	−1	−1
−1	−1	1	1	1	1	−1	−1	−1	−1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

-0.2
-0.2
-0.2
-0.2
-0.2
-0.2
0
0
0

−1	−1	−1	−1	−1	1	1	−1	−1	−1
−1	−1	−1	−1	−1	1	1	−1	−1	−1
−1	−1	−1	−1	1	1	−1	−1	−1	−1
−1	−1	−1	−1	1	1	−1	−1	−1	−1
−1	−1	−1	1	1	−1	−1	−1	−1	−1
−1	−1	−1	1	1	−1	−1	−1	−1	−1
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

-0.6
-0.6
-0.6
-0.6
-0.6
-0.6
0
0
0

FIGURE 2. The weight obtaining process of the row mean attention module.

Step 4: Use formula (4) to perform sigmoid operation on step 3, and then take the reciprocal of the result to get the weight of corresponding to the jth row. The value is normalized by sigmoid function to prevent the weight from being too large or too small to keep it in a relatively stable range.

Step 5: The final weighted output image features are obtained by using formula (5) and formula (6).

The weight obtaining process of the row mean attention module is shown in Fig. 2. Fig. 2(a) shows the feature image. As can be seen from the Fig. 2(b), the values in the figures are obtained from step 1 of the row mean attention module. After the second steps, the image is divided into three parts by executing the sign function operation on the normalized features. Among them, the blue part is the background area without steel bar in the horizontal direction, and each value of this area is 0. The red part is the segmented steel bar area and each value of this area is 1. The yellow part is the background area where there exist steel bar in the horizontal direction, each value of this area is -1. After the third step, the mean value of each row in the image is obtained, as shown in Fig. 2(c). After performing step 4, the row mean weight of each row in the image is obtained, as shown in Fig. 2(d). The horizontal coordinates represents the row number, and the vertical coordinates represents the row mean weight.

In view of the three different situations that may occur in steel bar segmentation, the row mean attention module can automatically process the weight of the steel bar region. We use the image with size of 10×10 as an example in the Fig. 2. In the first case, when the large background part is segmented into steel bar part, the weight is reduced to 1.8. In the second case, when the segmented steel bar part is moderate, the weight is also moderate, and the calculation result is 2.22. The third case, when the recall rate of steel bar part is low, the weight of steel bar region is increased to 2.82. In these three cases, a constant weight of 2 is obtained for the background region without steel bar in the horizontal direction, which is the blue area shown in Fig. 2(b). This weight is close to that of steel bar when the segmentation is moderate, so that the weight of each pixel is similar when the segmentation effect is good. Aiming at the problem of steel bar image segmentation, the applicability of row mean attention module is strong, which can effectively control the consistency of bar area width. Aiming at the problem of steel bar image segmentation, row mean attention module has strong applicability.

Furthermore, when the steel bar image is segmented, if the width value of each row of the segmented steel bar region in an image is inconsistent, the weight can be changed by the row mean attention module, which can optimize the width consistency of the segmented steel bar in the horizontal direction.

In addition, an attention branch optimization strategy based on mask attention module is designed in the decoding path of the multi-attention U-Net structure, which is the auxiliary optimization branch in the Layer 1 to the Layer 4 of the decoding path. As can be seen from Fig. 1, in each layer, the output of semantic segmentation of different scales is obtained through a 1×1 convolution after row mean attention module. Due to the size of the segmented image obtained by using high-level semantic features is relatively small, it can obtain better global information rather than detail features, which is very important for the steel bar segmentation. Therefore, we can get the basic steel bar region in the high-level

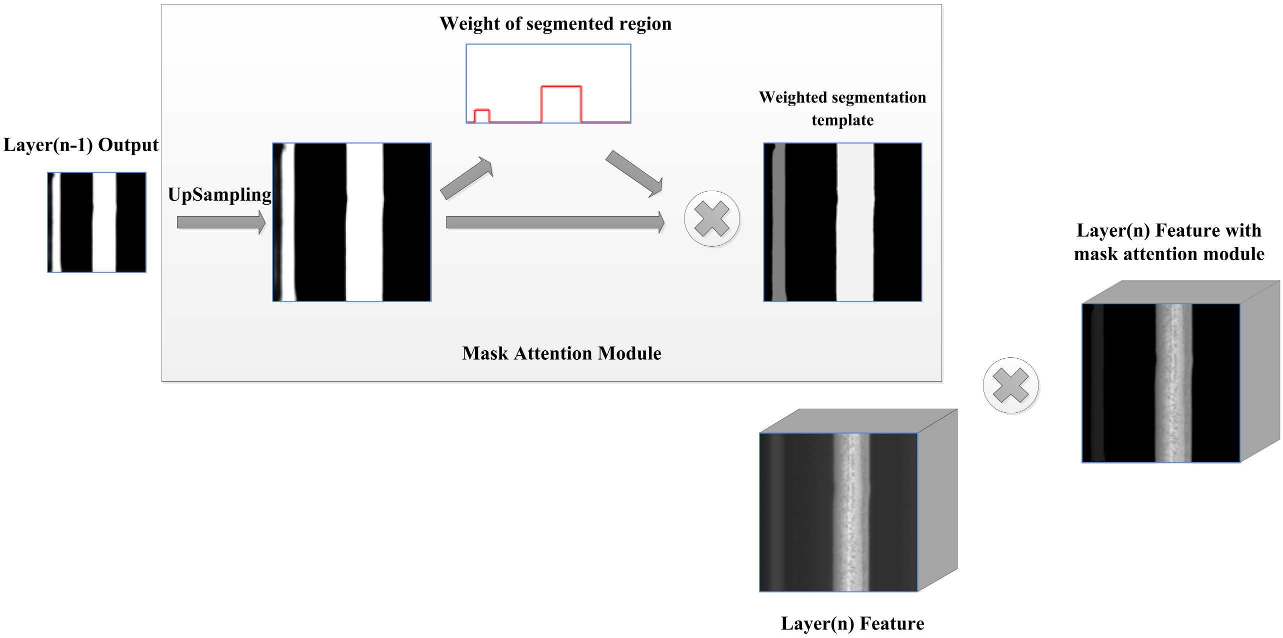

Weight of segmented region
Weighted segmentation template
Layer(n-1) Output
UpSampling
Layer(n) Feature with mask attention module
Mask Attention Module
Layer(n) Feature

FIGURE 3. The principle of the attention branch optimization strategy based on mask attention module.

semantic output layer, and then use this region as a template to filter the features of fusion in the next layer, which can well suppress the background interference.

The principle of attention branch optimization strategy based on mask attention module is shown in Fig. 3, the purpose of which is to pre judge the region of the steel bar in the up sampling feature, and ensure the continuity of the steel bar region. Firstly, the output image of Layer (n − 1) is performed the up sampling operation, the width and height of the obtained semantic segmentation image are consistent with the fusion feature of layer (n). Secondly, the mask attention module is executed. The main idea of this module is as follows: in the segmented image, there may be many regions may be segmented as possible steel bar regions, but there is only one real steel bar region. We judge whether it is a real steel bar region according to the following characteristics, the larger size of the region is, the more likely it is to be the real bar region. Therefore, the weight is calculated according to the size of each region, the larger the region is, the higher the weight is. Then the product of weight and segmented image is used to obtain the weighted segmentation template. Finally, in each layer of the decoding path, the Hadamard product is performed between the weighted segmentation template image and the skip connection feature of the current layer. It can effectively reduce the interference of background in the fusion features, and the integrity of the steel bar region is ensured to the maximum extent.

B. IMPROVED LOSS FUNCTION

In the auxiliary optimization branch, we use the original binary cross entropy loss function, the expression is

$$loss = \widehat{y}\log(y) + (1-\widehat{y})\log(1-y) \tag{7}$$

On the optimization of the main branch, in order to obtain steel bar segmentation effect better, we design an improved loss function in the output layer, which combine the binary cross entropy loss function and mean-squared-error loss function to recall the width of each row of the steel bar region in the image.

The expression of the improved loss function is

$$\begin{aligned} loss = &-\varphi\left(\widehat{y}\log(y) + (1-\widehat{y})\log(1-y)\right) \\ &-(1-\varphi) * \frac{1}{2}\left\|\widehat{y}_{r,m} - y_{r,m}\right\|^2 \end{aligned} \tag{8}$$

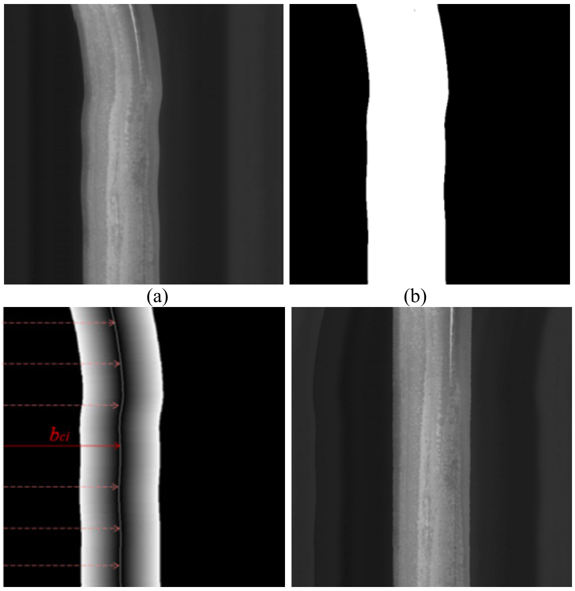
where $\widehat{y}$ represents the predicted value, y represents the true value. $\widehat{y}_{(r,m)}$ denotes the width value of the steel bar in each row of the predicted image, $y_{(r,m)}$ denotes the steel bar width value of each row of the real segmented image. φ is the ratio of binary cross entropy loss function to the whole loss function.

In the early stage of training, the loss of segmentation of the steel bar contour dominates the process of the parameter optimization. With the gradual reduction of segmentation loss, the width recall loss of the steel bar begins to occupy a dominant position, and the optimization of the model continues. The loss function designed in this paper reduces the risk of over fitting in the training process, and improves the training accuracy in the limited epochs.

C. DATA AUGMENTATION BASED ON CORRECTION DEFORMATION

In order to prevent over fitting, we do data augmentation before training. Based on the traditional data amplification methods, such as image transformation, rotation, scaling, truncation, etc., a data augmentation method based on correction deformation is designed. The background texture of the sample images generated by this method change obviously, which is good for improving the robustness of the multi-attention U-Net.

The process of deformation correction of the steel bar is shown in Fig. 4. Fig. 4(a) is the original image of the steel



(a) (b)

(c) (d)

FIGURE 4. The process of steel bar outline correction.

bar. Fig. 4(b) shows the segmented steel bar image, and then marks the central axis of the steel bar part in Fig. 4(b), as shown in Fig. 4(c). Finally, the steel bar image is corrected by using the steel bar correction formula (9), and the corrected steel bar image is obtained as shown in Fig. 4(d).

The steel bar correction formula is as follows,

$$gray_{new}(i,j) = \begin{cases} gray_{old}(i,j+offset-w), & when((j+offset)\geq w) \\ gray_{old}(i,j+offset), & when(0\leq(j+offset)\leq w) \\ gray_{old}(i,j+offset+w), & when((j+offset)\leq 0) \end{cases} \tag{9}$$

where $gray_{old}(i,j)$ represents the pixel value of the position (i,j) before image correction, $gray_{new}(i,j)$ represents the pixel value of the position (i,j) after image correction, w is the width of the image, and offset is calculated by the formula (10).

$$offset = b_{ci} - \frac{1}{2}w \tag{10}$$

where b_{ci} represents the pixel position of the center point of the ith bar area on the image.

Finally, we put the original images and the corrected images into the training dataset, which increases the diversity of sample data and reduces the risk of model over fitting.

D. PERFORMANCE EVALUATION METRICS

We used the following metrics to evaluate our model:

- Accuracy (PA)

$$PA = \frac{TP+TN}{TP+TN+FP+FN} \tag{11}$$

- Recall (Recall)

$$\text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

TABLE 1. The number of different types of images.

The type of images	Number
steel bar images in complex background	203
steel bar images with water mark interference	19
head and tail images of steel bar	104
steel bar images in clean background	54

- Intersection over Union (IoU)

$$Iou = \frac{TP}{TP+FP+FN} \tag{13}$$

- Because the steel bar image segmentation process does not need high accuracy of detail information, this paper evaluates the image segmentation effect of steel bar by calculating the proportion of image with IoU > 0.95, which is more in line with the artificial subjective evaluation metrics. The percentage calculation formula is as follows,

$$Percent(IoU > 0.95) = \frac{Num_{IoU>0.95}}{Num_{total}} \tag{14}$$

where TP represents the number of the true positive samples; TN stands for the number of the true negative samples; FP means the number of the false positive samples; FN means the number of the false negative samples. $Num_{(IoU>0.95)}$ is the number of images with IoU greater than 0.95 in the test dataset, Num_{total} is the total number of images in the test dataset

III. EXPERIMENTS AND RESULTS

In this paper, the images of steel bar collected from the steel bar production line of a steel mill are used as the dataset. A total of 380 steel bar images were marked, among them, 265 sample images are used as training dataset, and 115 sample images are used as test dataset. Table 1 shows the number of different types of images in the 380 marked steel bar images. We compare the multi-attention U-Net proposed in this paper with the standard U-Net [13], Attention U-Net [15], R2U-Net [23], D-Uet [26].

During the experiment, the training epochs of each method were set to 100, and the best training model was selected for testing. The multi-attention U-Net method is divided into five stages in the training process. In the first stage, the output of Layer1 is trained in the first 10 epochs. In the second stage, continue training for 10 epochs to optimize the output of Layer1 + Layer2. In the third stage, another 10 epochs are trained to optimize the output of Layer1 + Layer2 + Layer3. In the fourth stage, once again 10 epochs are trained to optimize the output of Layer1 + Layer2 + Layer3 + Layer4. In the final stage, the remaining 60 epochs are used to train the final output layer. According to many experiments, we set φ in the improved loss function as 0.6.

During the process of steel bar rolling, the background of some steel bar images is affected by light, many longitudinal stripes are generated which are consistent with the direction of steel bar in the image, which makes the background part easily be segmented into steel bar part. Fig. 5 shows the results of steel bar segmentation under complex background.

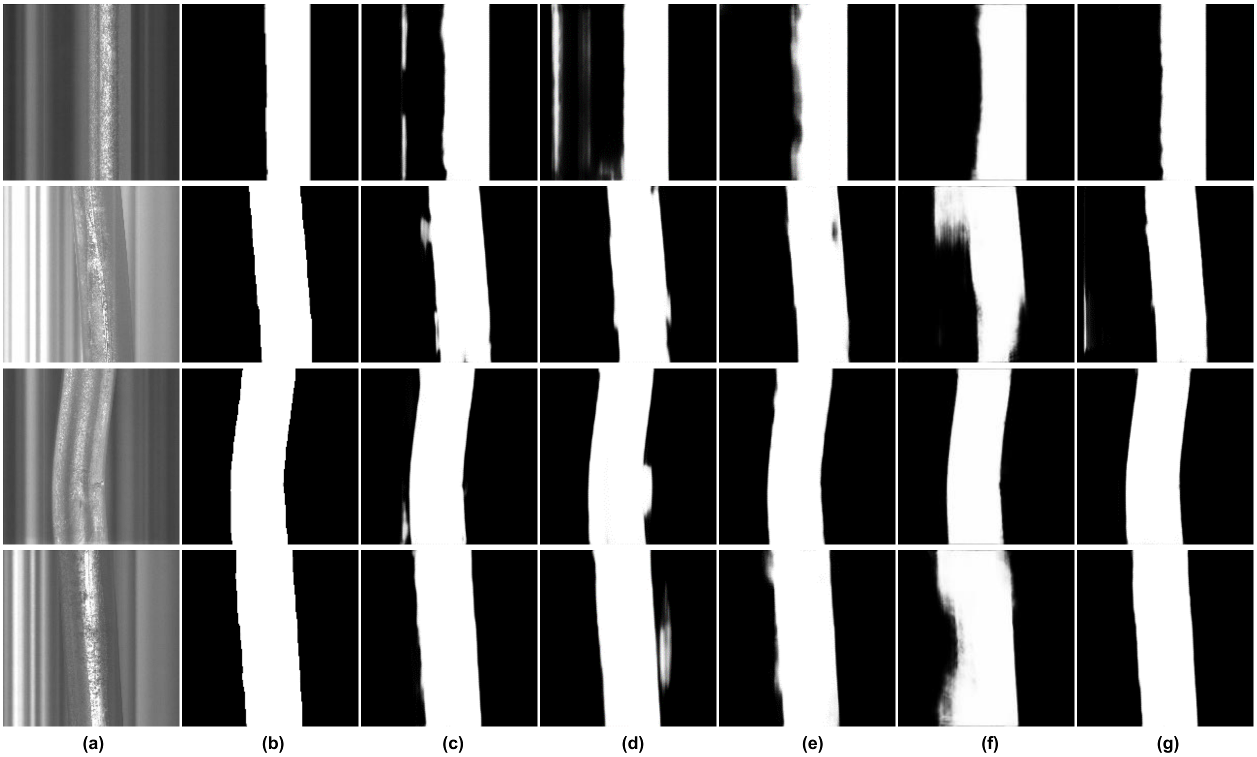


FIGURE 5. Image segmentation results of steel bar image in complex background, (a) Original image, (b) Ground truth, (c) U-Net, (d) Attention U-Net, (e) R2Unet, (f) DUnet, (g) Multi-attention U-Net.

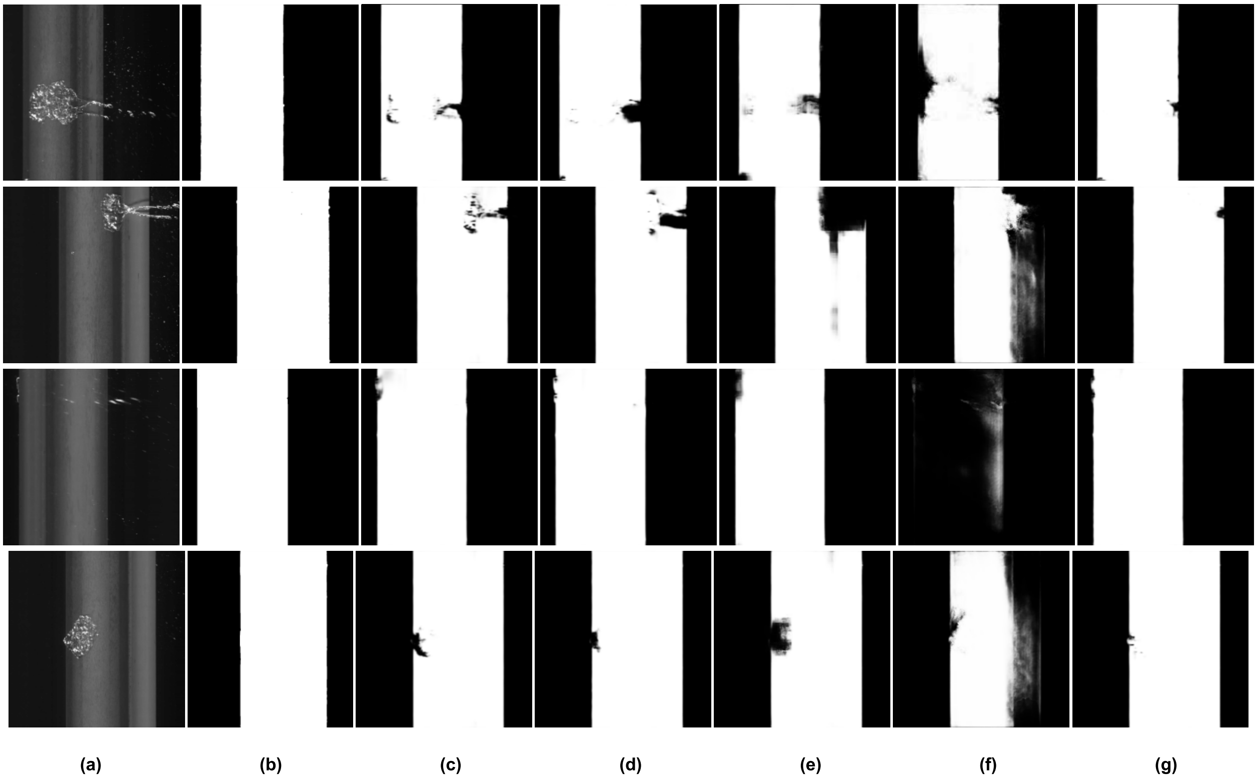


FIGURE 6. Image segmentation results of steel bar image with water mark interference, (a) Original image, (b) Ground truth, (c) U-Net, (d) Attention U-Net, (e) R2Unet, (f) DUnet, (g) Multi-attention U-Net.

It can be seen from the figure that when the image background is complex, the multi-attention U-Net method can segment the steel bar better and effectively filter out the interference of the background part.

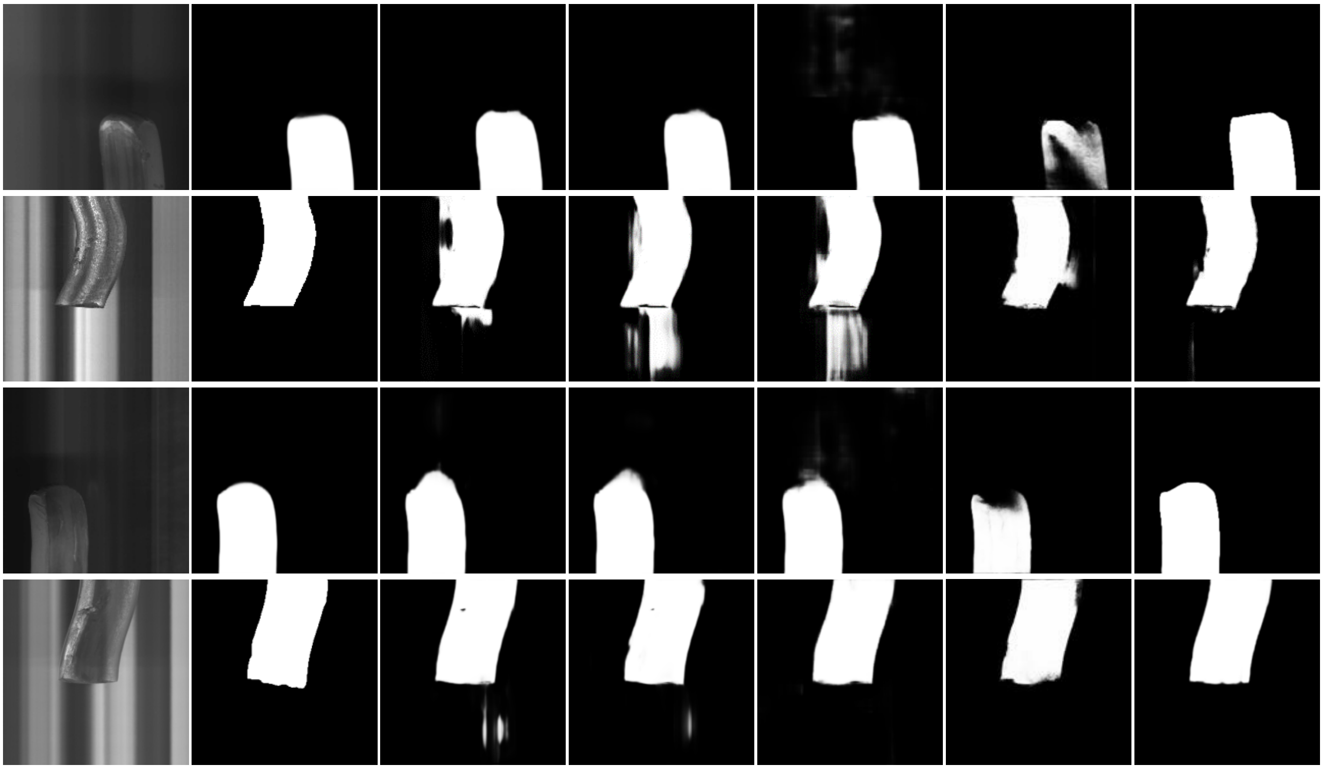


FIGURE 7. Image segmentation results of head and tail images of steel bar, (a) Original image, (b) Ground truth, (c) U-Net, (d) Attention U-Net, (e) R2Unet, (f) DUnet, (g) Multi-attention U-Net.

TABLE 2. The results of the performance evaluation of each method.

Method	PA	Recall	IoU	Percent(IoU>0.95)
U-Net[13]	0.9894	0.9816	0.9605	0.8526
Attention U-Net[15]	0.9829	0.9907	0.9521	0.8152
R2Unet[23]	0.9856	0.9881	0.9682	0.8421
DUnet[26]	0.9434	0.9455	0.9212	0.7522
Multi-attention U-Net	0.9903	0.9814	0.9718	0.9263

TABLE 3. The average number of images processed per second of each method.

Method	FPS
U-Net[13]	45
Attention U-Net[15]	41
R2Unet[23]	16
DUnet[26]	12
Multi-attention U-Net	41

In addition, due to the external interference on the surface of steel bar during rolling process, some steel bar surface images with good background have water mark, which will affect the image segmentation of steel bar. Fig. 6 shows the results of steel bar image segmentation under the condition of water mark interference. As can be seen from the figure, the multi-attention U-Net method can segment the steel bar part better, which will not segment the water mark to the steel bar part, and which can deal with the occlusion of water mark.

In addition to the above two cases, the proportion of the head and tail of the steel bar in the image is relatively small. When the background is complex, the steel bar head image and steel bar tail image are easier to segment the background into the steel bar part. When the background is clean, it is easy to segment the steel bar part into the background. Fig. 7 shows the segmentation results of the head and tail of the steel bar images. It can be seen from the figure that the multi-attention U-Net method has the best segmentation effect for this kind of image.

In order to verify the performance of the multi-attention U-Net method proposed in this paper, we use the performance evaluation metrics mentioned in the previous chapter to evaluate the performance of each method. The performance evaluation results are shown in Table 2. As can be seen from Table 1, the accuracy rate of multi-attention U-Net reaches to 99.03%, the value of IoU reaches to 0.9718, and the result of the percent (IoU $>$ 0.95) reaches 0.9263. This paper focuses on the performance evaluation results of percent (IoU $>$ 0.95), because it is more in line with the artificial subjective evaluation standard. The results show that the performance of multi-attention U-Net method is better than other methods. In addition, the data augmentation method based correction deformation proposed in this paper can improve the accuracy by 1% to 1.5%.

In the actual rolling process of steel bar, the production speed of steel bar is faster, so the real-time performance of the method is also required. Through testing on NVIDIA GetForce RTX2060 graphics card, we calculate the processing speed of each method. As shown in Table 3, the speed of the multi-attention U-Net methods proposed in this paper can reach 40 images per second, which can meet the real-time requirements of steel bar production.

IV. CONCLUSION

In this paper, a multi-attention U-Net method is proposed for steel bar image segmentation. In this method, the row mean attention module is proposed, which gives dynamic weight to the features to control the filtering of the background and steel bar recall rate. In addition, an attention branch optimization strategy based on mask attention module is designed, which aims to filter the low-level segmentation features by using the semantic features of the high-level, which can improve the integrity of segmented steel bar region.

The improved loss function is special designed to improve the processing effect of the network on the steel bar segmentation. The diversity of the steel bar image sample dataset is increased and the robustness of the network is improved by the correction deformation data amplification method.

Compared with the standard U-Net, attention U-Net, R2U-Net and DUNet, the IoU value of the proposed multi-attention U-Net method in steel bar image segmentation is increased to 0.9718, which is higher than that of other methods. In addition, according to the results of the subjective performance evaluation metrics of percent(IoU $>$ 0.95), we can see that the value of the percent(IoU $>$ 0.95) obtained by our method is as high as 92.63%, which is also much higher than other methods. Through experiments, the speed of the proposed method is 41 FPS, which can meet the real-time requirements of the segmentation of the steel bar image.

Our method can segment the steel bar image well, which paves the way for detection and recognition of the steel bar image surface defects.

REFERENCES

[1] W. B. Li, C. H. Lu, and J. C. Zhang, "Analysis of steel bar surface defects based on machine vision," *Adv. Mater. Res.*, vol. 549, pp. 1017–1020, Jul. 2012.

[2] J. C. Zhang, W. B. Li, and C. H. Lu, "Design of automatic detection device for steel bar surface defects," *Adv. Mater. Res.*, vols. 532–533, pp. 390–393, Jun. 2012.

[3] W. B. Li, Q. Z. Zhang, J. L. Sun, L. Liu, and S. L. He, "Review of vision inspection technology for surface defect of steel bar," *Appl. Mech. Mater.*, vol. 740, pp. 543–546, Mar. 2015.

[4] W.-B. Li, C.-H. Lu, and J.-C. Zhang, "A local annular contrast based real-time inspection algorithm for steel bar surface defects," *Appl. Surf. Sci.*, vol. 258, no. 16, pp. 6080–6086, Jun. 2012.

[5] W.-B. Li, C.-H. Lu, and J.-C. Zhang, "A lower envelope weber contrast detection algorithm for steel bar surface pit defects," *Opt. Laser Technol.*, vol. 45, pp. 654–659, Feb. 2013.

[6] J. M. P. Dewi, A. Armiati, and S. Alvini, "Image segmentation using minimum spanning tree," in *Proc. IOP Conf., Mater. Sci. Eng.*, vol. 335, 2018, Art. no. 012135.

[7] C. A. Akagic, E. Buza, S. Omanovic, and A. Karabegovic, "Pavement crack detection using Otsu thresholding for image segmentation," in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO)*, vol. 5, May 2018, pp. 1092–1097.

[8] P. PratimAcharjya and D. Ghoshal, "A modified watershed segmentation algorithm using distances transform for image segmentation," *Int. J. Comput. Appl.*, vol. 52, no. 12, pp. 46–50, Aug. 2012.

[9] A. S. Kavitha, P. Shivakumara, G. H. Kumar, and T. Lu, "A new watershed model based system for character segmentation in degraded text lines," *AEU-Int. J. Electron. Commun.*, vol. 71, pp. 45–52, Jan. 2017.

[10] A. Das and D. Ghoshal, "Human skin region segmentation based on chrominance component using modified watershed algorithm," *Procedia Comput. Sci.*, vol. 89, pp. 856–863, 2016.

[11] C. Sun, S. Guo, H. Zhang, J. Li, M. Chen, and S. Ma, "Automatic segmentation of liver tumors from multiphase contrast-enhanced CT images based on FCNs," *Artif. Intell. Med.*, vol. 83, no. 6, pp. 58–66, Nov. 2017.

[12] J. J. Ma, Y. Li, K. Du, F. Zheng, L. Zhang, and Z. Gong, "Segmenting ears of winter wheat at flowering stage using digital images and deep learning," *Comput. Electron. Agricult.*, vol. 168, Jan. 2020, Art. no. 105159.

[13] C. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351. Bäch, Switzerland: Springer, pp. 234–241, Nov. 2015.

[14] C. Z Zhou, Md M.R Siddiquee, N Tajbakhsh, et.al, "UNet++: A nested U-Net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045. Cham, Switzerland: Springer, Sep. 2018, pp. 3–11.

[15] J. O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," May 2018, *arXiv:1804.03999*. [Online]. Available: https://arxiv.org/abs/1804.03999

[16] S. Lian, Z. Luo, Z. Zhong, X. Lin, S. Su, and S. Li, "Attention guided U-Net for accurate iris segmentation," *J. Vis. Commun. Image Represent.*, vol. 56, pp. 296–304, Oct. 2018.

[17] C. Li, Y. Tan, W. Chen, X. Luo, Y. He, Y. Gao, and F. Li, "ANU-net: Attention-based nested U-net to exploit full resolution features for medical image segmentation," *Comput. Graph.*, vol. 90, pp. 11–20, Aug. 2020.

[18] C. Shun Zhao, T. Liu, B. Liu, and K. Ruan, "Attention residual convolution neural network based on U-net (AttentionResU-Net) for retina vessel segmentation," in *Proc. IOP Conf.: Earth Environ. Sci.*, vol. 440, Mar. 2020, Art. no. 032138.

[19] S. Li, G. K. F. Tso, and K. He, "Bottleneck feature supervised U-Net for pixel-wise liver and tumor segmentation," *Expert Syst. Appl.*, vol. 145, May 2020, Art. no. 113131.

[20] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.

[21] C. H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks," in *Medical Image Understanding and Analysis* (Communications in Computer and Information Science), vol. 723. Cham, Switzerland: Springer, pp. 506–517 Jun. 2017.

[22] Z. Gu, J. Cheng, H. Fu, K. Zhou, H. Hao, Y. Zhao, T. Zhang, S. Gao, and J. Liu, "CE-net: Context encoder network for 2D medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 38, no. 10, pp. 2281–2292, Oct. 2019.

[23] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, p. 1, Mar. 2019.

[24] S. Liu, Y. Li, J. Zhou, J. Hu, N. Chen, Y. Shang, Z. Chen, and T. Li, "Segmenting nailfold capillaries using an improved U-net network," *Microvascular Res.*, vol. 130, Jul. 2020, Art. no. 104011.

[25] H. Tang, X. Pei, S. Huang, X. Li, and C. Liu, "Automatic lumbar spinal CT image segmentation with a dual densely connected U-Net," *IEEE Access*, vol. 8, pp. 89228–89238, May 2020.

[26] Q. Jin, Z. Meng, T. D. Pham, Q. Chen, L. Wei, and R. Su, "DUNet: A deformable network for retinal vessel segmentation," *Knowl.-Based Syst.*, vol. 178, pp. 149–162, Aug. 2019.

[27] C. Guo, M. Szemenyei, Y. Pei, Y. Yi, and W. Zhou, "SD-UNet: A structured dropout U-Net for retinal vessel segmentation," in *Proc. IEEE 19th Int. Conf. Bioinf. Bioeng. (BIBE)*, Athens, Greece, Oct. 2019, pp. 439–444.

[28] J. Ming Zhao, Y. Wei, Y. Lu, and K. L. K. Wong, "A novel U-Net approach to segment the cardiac chamber in magnetic resonance images with ghost artifacts," *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105623.

[29] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.

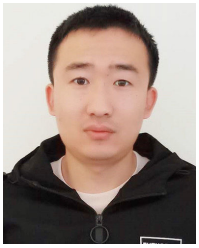
[30] C. H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Large Kernel matters–improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Honolulu, HI, USA, Nov. 2017, pp. 1743–1751.

[31] R. Augustauskas and A. Lipnickas, "Improved pixel-level pavement-defect segmentation using a deep autoencoder," *Sensors*, vol. 20, no. 9, p. 2557, Apr. 2020.

[32] C. X. Lu, W. Wang, M. Danelljan, T. Zhou, J. Shen, and G. L. Van, "Video object segmentation with episodic graph memory networks," in *Computer Vision–(ECCV)*, vol. 12348. Cham, Switzerland: Springer, Dec. 2020, pp. 661–679.

[33] X. Lu, W. Wang, J. Shen, Y.-W. Tai, D. Crandall, and S. C. H. Hoi, "Learning video object segmentation from unlabeled videos," Mar. 2020, *arXiv:2003.05020*. [Online]. Available: https://arxiv.org/abs/2003.05020

JIE SHI received the B.S. and M.S. degrees in computer science and technology from Tiangong University, in 2013 and 2017, respectively. She is currently working as an Image Algorithm Engineer with the Detection Technology Department, Design and Research Institute Company Ltd., University of Science and Technology Beijing. Her research interests include image segmentation, target detection, target recognition, and deep learning.



KUNPENG WU received the bachelor's degree from the University of Science and Technology Beijing, China, in 2014, where he is currently pursuing the master's degree in mechanical engineering. He worked as an Algorithm Engineer with the Detection Technology Department, Design and Research Institute Company Ltd., University of Science and Technology Beijing. His research interests include deep learning, image segmentation, surface defect detection and recognition, character recognition, and image presentation.

CHAOLIN YANG received the B.S. and M.S. degrees in mechanical and electronic engineering from the University of Science and Technology Beijing, in 2000 and 2003, respectively. He is currently an Assistant Research Fellow with the Institute of Engineering Technology, University of Science and Technology Beijing, Beijing, China. His current research interests include medical machine vision and deep learning.

NENGHUI DENG received the M.S. degree in mechanical and electronic engineering from the University of Science and Technology Beijing, in 2013. He is currently the Director with the Testing Technology Department, Design and Research Institute Company Ltd., University of Science and Technology Beijing. His research interests include machine vision and deep learning.