

Received June 4, 2020, accepted June 14, 2020, date of publication June 17, 2020, date of current version June 29, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003089

Defect Inspection in Tire Radiographic Image Using Concise Semantic Segmentation

ZHOUSHOU ZHENG¹, SEN ZHANG¹, BIN YU^{1,2}, QINGDANG LI³,
AND YAN ZHANG¹, (Member, IEEE)

¹College of Mechanical and Electrical Engineering, Qingdao University of Science and Technology, Qingdao 266061, China

²College of Mathematics, Qingdao University of Science and Technology, Qingdao 266061, China

³College of Sino-German Science and Technology, Qingdao University of Science and Technology, Qingdao 266061, China

Corresponding author: Yan Zhang (zy@qust.edu.cn)

This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2019MEE066 and Grant ZR2018MC007.

ABSTRACT Automated tire visual inspection plays an extraordinary important role in ensuring tire quality and driving safety. Due to the anisotropic complex multi texture and defect diversity characteristic of tire radiographic image, tire intelligent visual inspection has become one of the technical bottlenecks of intelligent manufacturing. In this work, a novel tire defect detection model using Concise Semantic Segmentation Network (Concise-SSN) is investigated for automated tire visual inspection. We perform an end-to-end pixel-wise tire defect detection by combining the power of an optimized semantic segmentation network and a compact convolutional neural network for classification. It can achieve the end-to-end pixel-wise full class defect detection and classification. The experimental results show superior performance on defect segmentation and classification tasks compared to state-of-the-art models with smaller model size and faster computation. Comparative experiments indicated that our Concise-SSN achieves the mPA score of 85.13%, the mIoU score of 77.34% on our test set. The accuracy of defect classification is 96.5% on average. Finally, we show faster computation (0.132 seconds per image) with competitive results on our dataset, which can meet the needs of online tire detection.

INDEX TERMS Intelligent defect detection, tire, radiographic image, semantic segmentation network.

I. INTRODUCTION


According to World Health Organization, about 1.3 million people die and 50 million disabled from road accidents each year, among them 40% are caused by tire failures [1]. Automated tire visual inspection plays an extraordinary important role in ensuring tire quality and driving safety. In recent years, automated vision inspection methods have obtained increasingly attention by researchers and industries. Numerous methods were proposed based on traditional methods [2]–[4], conventional computer vision based techniques [5]–[11], and deep learning methods [12]–[19].

Since the middle of last century, researchers have been focusing on applying nondestructive testing techniques including laser shearography [2], ultrasonic methods [3], electromagnetic pulse [4] etc. to tire defect detection. Laser shearography can directly show a wide range of tire internal defects, and it is easy to judge. However, it needs vacuum environment to complete the detection, and the detection

efficiency is relatively slow and cumbersome. Ultrasonic technology can also detect a variety of tire defects. The detection process is complex with low efficiency and high cost. Electric pulse technology can't detect tire defects comprehensively.

Radiography, especially X-ray radiography testing, is an important and widely used industrial non-destructive testing (NDT) method because of its cost-effective and intuitive characteristics which leads to images with a more reliable and faster interpretation of the objects to be tested. X-ray radiography was applied to tire defect detection and followed by naked-eye detection. However, naked-eye detection has a low detection efficiency, low accuracy (90%-95% on average) and is sensitive to the tiredness of the inspectors. Thus, tire automated visual inspection has become one of the technical bottlenecks of intelligent manufacturing.

To alleviate the problem, by virtue of their efficiency, accuracy and real-time performance, conventional computer vision techniques have been playing an increasingly important role in the field of NDT and automated visual inspection applications [5]–[11]. Guo and Wei [5] applied local total

The associate editor coordinating the review of this manuscript and approving it for publication was Orazio Gambino .

variation filtering for tire texture decomposition, followed by vertical mean filtering and thresholding for tire impurity defect segmentation. Zhang *et al.* systematically studied tire defect detection by image enhancement and defect feature extraction through image decomposition, image transformation combining edge operators [6]–[8]. In [8], the authors detected tire bubble and foreign matter defects by combining the total variation, curvelet transform based image enhancement and edge operator. Bubbles and defects in tires surface can be detected by combining curvelet transform and Canny operator in tire laser shearography images [6]. In [7], a tire defect detection method was proposed based on wavelet multiscale analysis in ways of local regularity analysis and scale characteristic. This framework distinguishes the defect features from the background textures. Li [9] proposed a radial tire defect detection method in radiographic images based on fuzzy edge detection. The existence of defects in tires could lead to the transformation of texture in tire radiography images, such that defect detection can be realized based on texture analysis. In [10], feature similarity was analyzed to capture the texture distortion of each pixel by weighted averaging the dissimilarity between the pixel and its neighborhoods. This method can be used to detect texture and foreign body defects in tread and sidewall. Zhao *et al.* proposed a detection method using local inverse difference moment features and received good performance on foreign object defects [11].

Compared with naked-eye detection conventional computer vision based methods provide great convenience for automatic online detection. However, except for the high computational complexity, the most crucial and indispensable steps for these techniques are feature extraction and selection. As the number of categories increases, feature extraction becomes more and more difficult. It depends on the judgment of researchers and long-term trial and error to determine which features best describe different target categories. In addition, each feature definition also needs to deal with a large number of parameters, all of which must be adjusted by performers. Moreover, artificial features provide insufficient representation for each kind of defect. For the above reasons, conventional methods mostly focus on a single texture structure detection or one particular kind of defect.

Recently, deep neural networks (DNNs) have demonstrated remarkable effectiveness and achieved the state-of-the-art performance in image analysis and recognition. They have been utilized in visual inspection applications such as fabric, weld and tire recently [12]. According to the periodicity of fabric texture, Jing *et al.* [13] divided the fabric image into blocks, and classify the blocks using the deep convolution neural network (DCNN) to find out the defective area. The variation of fabric periodic textures and block dimension would affect the generalization of the method. In [14], Ouyang *et al.* developed a fabric defect inspection system by combining image processing, fabric motif determination, candidate defect map generation, and convolutional neural networks (CNNs). Unlike fabric defect detection

problem, tire radiography has the characteristics of multi texture structure, various defect types, and changeable features of the same type of defects, which brings great challenges to tire visual inspection. To solve these difficulties, Cui *et al.* designed a multi CNNs model for tire defect classification by combining 5 CNNs trained on different dataset [15]. The model achieved satisfied classification accuracy at the cost of complex network structure and tedious training process. Zhang *et al.* [16] proposed a supervised feature embedded deep learning method for tire defects classification by combining regularization techniques to boost performance. To perform a pixel-wise detection, a tire defect detection method was introduced based on a fully convolutional network (FCN) [17] in which defect region can be segmented at pixel level.

Existing DNNs based methods show attractive defect segmentation or defect classification results and has great developmental potentialities. Nevertheless, intelligent manufacturing industry requires realizing end-to-end tire defect segmentation and classification in tire visual inspection application. It is still an unsolved challenge.

To target the aforementioned problems, in this work a novel tire visual inspection model is investigated for automated tire visual inspection. We observe that in literature many efficient implementations employ a two or multi-branch architecture [18], [19]. We perform an end-to-end tire defect detection by combining an optimized SSN for pixel-wise segmentation and a compact CNN classifier for classification. In the SSN module, we choose VGG16 as the base model and designed a novel network for defective region pixel-wise prediction. Predicted defective regions are cropped from original images and input to the compact CNN classifier module for re-targeting defect classification.

The main contributions of the paper are as follows.

1. We propose a novel and concise two-branch architecture Concise-SSN for pixel-wise tire visual inspection in radiographic images.
2. Different from the existing methods, the proposed model can achieve the end-to-end pixel-wise defect segmentation and classification.
3. We introduce an unsupervised tire texture image segmentation method based on Gabor filter and fuzzy c-means clustering.

The rest of this paper is organized as follows. Section II introduces the state-of-the-art semantic segmentation network and the proposed approach. The details of the experimental setup and performance comparison are given in Section III. The result analysis is carried out in Section IV. Finally, conclusions and further work are delivered in Section V.

II. RELATED WORK

A. PRINCIPLE OF TIRE RADIOGRAPHICAL IMAGING AND INSPECTION

In tire industry, radiographic imaging is widely used for visual inspection. Due to the difference of tire tread, sidewall,

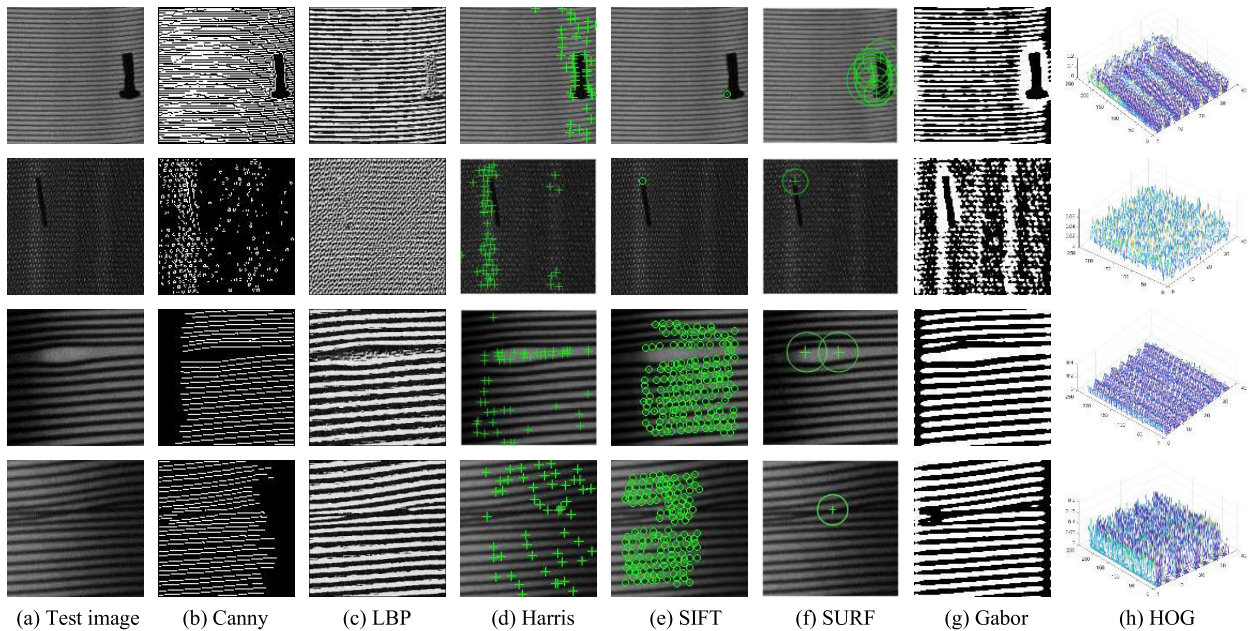


FIGURE 1. Conventional local feature descriptions for tire defects.

edge in materials, density and thickness, the different absorption degree of X-ray in different positions of tire makes its imaging present different intensity. Compared with other visual inspection applications like fabrics, paper and IC etc., conventional inspection methods based on statistics, template, texture, and motif-based etc. are not satisfactory for intelligent manufacturing.

There are certain major difficulties in intelligent tire defect detection due to the different texture structures and varied forms defect features of tire radiographic images [7]. Firstly, to perform flexible manufacturing a variety of tires with different tread depths and patterns should be produced via the same production line, meaning that intelligent visual inspection should be robust to different textures of different types of tires as well as texture difference between sidewall and tread. However, visual inspection system should not be too sensitive to normal deformation of texture caused by viscoelasticity of composite polymer. Secondly, tire defect detection is particularly challenging due to the large number of defect categories which located in complex anisotropic textures. To make it worse, different from natural scene understanding, color space provides guarantee for feature extraction and expression, tire radiographic images are 8-bit gray level, not to mention the features of the same kind of defects are in varied forms as shown in Fig. 5.

Therefore, the intelligent tire defect detection problem is to automatically detect and recognize the possible defects of multiple types from complex feature space in anisotropic multi texture image. For conventional computer vision techniques, it is an ill-posed problem that the overall system becomes complicated, expensive, and may not be robust to flexible manufacturing. Fig. 1 shows the representation of

conventional features on tire anisotropic multi texture and various defects. The conventional artificial features of different types of defects varies significantly such that conventional methods need different features for different defects, which is not conducive to automatic detection.

B. SEMANTIC SEGMENTATION NETWORKS

As one of the key problems in the field of computer vision, semantic segmentation is playing an increasingly important role in the complete understanding of the scene in a variety of applications including object recognition, autonomous driving, virtual reality, human-computer interaction etc. Since 2015, the state-of-the-art semantic segmentation networks, e.g. FCN [20], SegNet [21], U-Net [22], Mask R-CNN [35] and their successors, have been demonstrated for its effectiveness in end-to-end prediction and have been successfully applied to real-world applications.

Considering real-world applications, runtime is becoming an essential factor in semantic segmentation architectures. Among the state-of-the-art semantic segmentation DCNNs, SegNet utilizes the joint encoder-decoder module architecture and pooling indices strategy to reduce network parameters and the computational cost. Derived from simple classification DCNN method, e.g. VGG16 [23], the encoder module is topologically identical to its convolutional layers to extract DCNN features. The fully connected layers of VGG or ResNet are removed to make it smaller. The decoders use the pooling indices strategy received from the corresponding encoder to perform non-linear up-sampling of their input feature maps. The reduction of parameters enabled end-to-end training.

To perform real-time online tire inspection, efficacy and efficiency in terms of memory and computational time are two important requirements for the system. SegNet, as one of the most efficient semantic segmentation models, achieves high scores for road scene understanding on large and well-known datasets. Most importantly, it only stores the max-pooling indices of the feature maps and uses them in its decoder network such that it is more efficient than those architectures which store the encoder network feature maps in full to achieve good performance.

Real-time semantic segmentation has recently enjoyed significant gain in popularity in real-world applications because of its performance [24]–[26]. Although all the tricks adapted by semantic segmentation networks improve performance on challenging benchmarks, it is unfortunately difficult to disentangle the key design factors necessary for real-life real-time applications to achieve satisfactory performance even with the help of high-end GPUs. For most such applications, how to keep efficient inference speed and high accuracy with high resolution images is a critical question. Another problem faced by applications, such as medical diagnosis, industrial visual inspection etc., is that samples from a real-life application may not always evenly distribute among classes, in other words, they are brimming with imbalanced datasets or there are not enough samples for a certain class in the neural network learning process. SegNet was designed to perform efficient road and indoor scene understanding with evenly distributed nature classes. In the following section, we will reveal the inappropriateness of conventional SegNet model in tire visual inspection. For the major difficulties mentioned above in intelligent tire inspection and real-time semantic segmentation applications, we therefore cope with the intelligent tire inspection problem by construct a novel and concise DCNN architecture based on semantic segmentation and reveal its pros and cons. Different from the existing methods, we probe into employing a two-step lightweight network architecture to provide sufficient receptive field for real-time end-to-end tire defect inspection at pixel level.

III. PROPOSED METHOD

A. OVERVIEW

Inspired by the SegNet, in this work we propose a concise encoder-decoder architecture for tire radiographic image feature learning. The problem space is different from road and indoor scene understanding such that we probe into designing a more compact and effective network structure by removing layers that have little influence on tire radiographic image feature learning to perform the efficiency and accuracy of real-time detection. The encoder network of SegNet consists of 13 convolutional layers which correspond to the first 13 convolutional layers in the VGG16 network to obtain the feature maps. The fully connected layers from the VGG16 network which consists of about 90% of the parameters of their entire network are discarded to be able to train the network using the relevant training set using

SGD optimization. This would therefore significantly reduce memory consumption and improve inference time without sacrificing performance. Since input resolution and network depth are main factors for runtime, to further improve model efficiency to make it suitable for such type of applications, a shallower and more efficient network structure is constructed.

In this section, we first describe our preprocessing method for simplifying problem space and then detail the specific network architecture and configurations of Concise-SSN in Section III C.

B. TIRE RADIOGRAPHIC IMAGE PREPROCESSING METHOD

The structure of tire, especially radial tire, is complex, which is composed of multi composite materials which determine the multi texture distribution structure of tire radiographic image. This multi texture background and high resolution of input images brings challenges to the design of a unified robust and automatic tire visual inspection algorithm. To alleviate this problem, we first propose a texture segmentation method based on Gabor filter and fuzzy c-means clustering to segment different regions of tire radiographic images. Segmented image blocks are then input to the proposed tire visual inspection model for further detection. This trick would however significantly reduce the complexity and computation of the problem space in this application.

C. CONCISE-SSN NETWORK ARCHITECTURE

The proposed Concise-SSN network architecture is designed to be efficient for intelligent real-time tire visual inspection at pixel level. We employ a two-step lightweight network architecture, namely pixel-wise defective region semantic segmentation module and defect classifier module considering the characteristics of such applications. Fig. 2 presents the proposed Concise-SSN network architecture for defect inspection in tire radiographic image. In defective region semantic segmentation module, a concise encoder network and a corresponding decoder network is applied for pixel wise defective region prediction. The encoder network produces these low-resolution representations of input tire gray level images. The role of the decoder network is to map the low-resolution encoder feature maps to full input resolution feature maps for pixel-wise prediction. And thus, the pixel-wise defective region semantic segmentation module solves a lower complexity binary classification problem. Defective regions are then cropped as input for further classification in a shallow CNN.

Due to the different adaptability and characterization of different kinds of defects to the network architecture, less convolution usually leads to worse semantics and more noise, while too many convolutions result in lower resolution and poor perception of details. An appropriate network structure is essential to improve the efficiency of defect segmentation. Too many network layers may lead to too much computation and feature overlearning. For the defective region semantic segmentation module, three network architectures

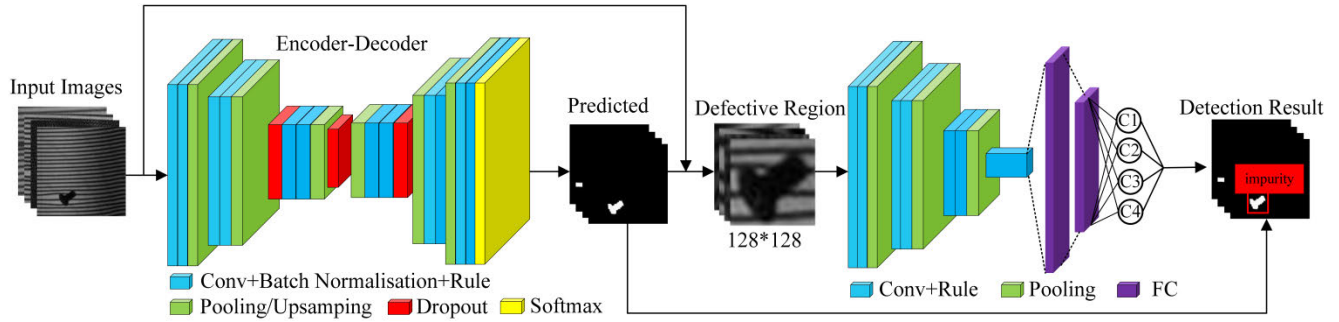


FIGURE 2. The proposed Concise-SSN Network architecture for defect inspection in tire radiographic image.

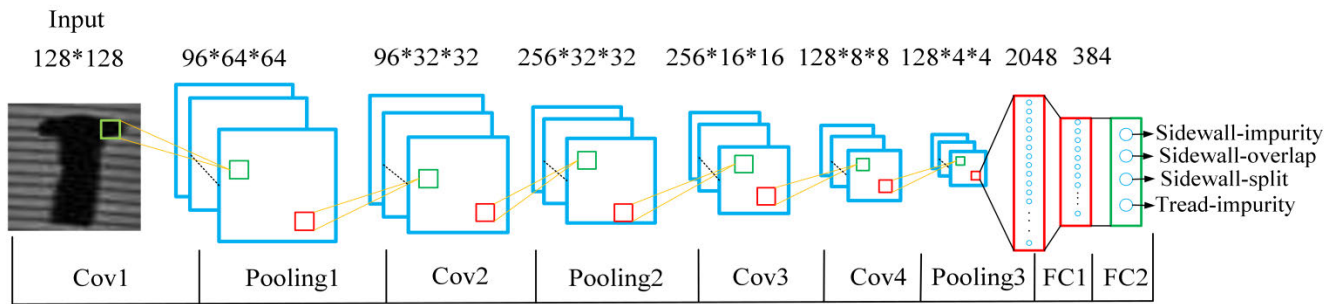


FIGURE 3. The detailed architecture of the CNN classifier module.

are designed for comparison. The parameter settings are given in Table 1. The encoder network performs convolution with a filter bank to produce a set of feature maps. Identical to SegNet in which the encoder network consists of 13 convolutional layers, 5 pooling layers which correspond to the VGG16 network, we apply the first 7 convolutional layers as in VGG16 and followed with 3 pooling layers aiming at allowing for real-time computation. The 3 max-pooling layers are used to achieve more translation invariance for robust classification. Similarly, fully connected layers of VGG16 are removed to make the encoder smaller and easier to train.

The decoder is a symmetric network of the encoder network. It is used to upsample its input feature maps using the memorized max-pooling indices from the corresponding encoder feature maps. The feature maps are convolved with a trainable decoder filter bank to produce dense feature maps from sparse feature maps.

The dense feature maps of the decoder are input into a trainable soft-max layer as a classifier which classifies each pixel independently. As shown in Fig. 2, considering the goal of the defective region semantic segmentation module is to predict defective regions, such that the output of the softmax classifier is a two-channel image of probabilities, namely the output of semantic segmentation module are defect-free and defective classes.

For further defect category classification, in the following compact CNN classifier module predicted defective regions are cropped from original images. If the input resolution is too large, the number of neurons in the full connection layer will increase, and the network parameters will increase; on the

other hand, if the input resolution is too small, it will affect the field of vision and defect feature extraction of the network. We have made statistics on the geometric dimensions of tire defects in the data set. The result shows that the defect width and height are within the range of 20 pixels to 120 pixels [15]. Therefore, in order to maintain the vision field and network performance, we mapped into them to a fixed-size of 128×128 resolution as input for re-targeting defect classification. In the classifier module, 4 convolution layers, 3 pooling layers and 2 full connection layers are employed. The detailed architecture of the CNN classifier module is illustrated in Fig. 3.

IV. EXPERIMENTS AND DISCUSSION

In this section, we perform experiments on tire radiographic image dataset and present experimental results to validate the effectiveness of the proposed model. We investigate the performance of our proposed Concise-SSN Network with the state-of-the-art segmentation networks and DCNNs. In the following, we first introduce the implementation details, the tire radiographic image dataset, the training and implementation details the proposed architecture, then conduct the pixel accuracy (PA), mean pixel accuracy (mPA), intersection over union (IoU), mean intersection over union (mIoU) and classification accuracy (CA) results on the same dataset compared with the existing algorithms.

A. IMPLEMENTATION DETAILS

In this work, all the experiments were implemented on the MATLAB 2019a and conducted with Intel(R) Core (TM)

TABLE 1. Parameters for three segmentation modules.

Model	Conv1	Pool1	Conv2	Pool2	Conv3	Pool3	tConv1	tConv2	tConv3	Pixel Class
Model A	3*3*64	2*2					4*4*64			2
Model B	3*3*64	2*2	3*3*128	2*2			4*4*64	4*4*128		2
Model C	3*3*64	2*2	3*3*128	2*2	3*3*256	2*2	4*4*64	4*4*128	4*4*256	2

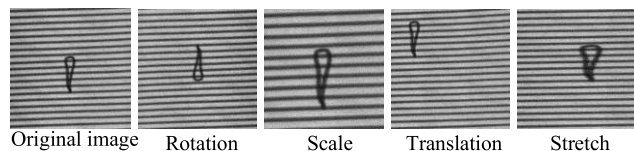
i7-8700 CPU @ 3.20GHz 3.19GHz processor, 8.0 GB RAM and NVIDIA GEFORCE GTX 1080ti GPU. The initial learning rate is set to 0.001. For quantitative evaluation, three commonly used image segmentation evaluation metrics, namely PA, IoU, and CA, are applied to evaluate the segmentation effect of each compared algorithm over all classes. PA is the ratio of the correct segmented pixel to the total pixel, which reflects the accuracy of single category defect segmentation. As an evaluation index of semantic segmentation, IoU is the ratio of intersection and union of ground truth and predicted segmentation. CA is used as the evaluation standard of classification network.

B. DATA AUGMENTATION, PREPROCESSING AND DATASET DESCRIPTION

The original tire radiographic images are of high resolution (11400 × 2469 pixels) with 8-bit gray levels captured from a variety of different tires in a real-life production line. Preprocessing stage is composed by three main operations including down-sampling, texture segmentation and image blocking. Firstly, image down-sampling is implemented to reduce computational complexity and improve efficiency. Secondly, texture segmentation is performed using Gabor wavelet and fuzzy clustering method. Thirdly, test images are divided into blocks based on texture segmentation results. The complexity of problem space is reduced such that theoretically it helps to achieve better inspection results.

Dataset is one of the most critical aspects of successful deep learning applications. Normally, limited labeled data leads to overfitting, which means the model will not be able to generalize to unseen examples. This can be mitigated by data augmentation, which effectively increases the amount and diversity of data seen by the network. Limited by industrial production conditions, the dataset for semantic segmentation architecture is limited. To improve the generalization ability and detection accuracy of the network, data augmentation techniques are applied including random resizing between 0.2 to 4, rotation, scale, translation and stretch as shown in Fig. 4. After data augmentation, the dataset reached 3234 image samples with high label quality.

Among them the 3234 test images used in this work, there are 738 tread-impurity samples, 1011 sidewall-impurity samples, 468 sidewall-overlap samples, 549 sidewall-split, 234 defect-free samples for tread and sidewall respectively, in which 70% are randomly chosen as training set, 20% are randomly chosen as test set and 10% are randomly chosen

**FIGURE 4.** Sample data augmentation techniques implied.

as validation set. High-quality pixel labels for the training set and validation set are available.

For texture segmentation, 40 groups of Gabor filters in 8 directions are designed to filter the test image, from which the filter directions and parameters with strong class representation ability are preselected, as shown in Fig. 6. Six groups of filter banks with $\sigma = 1.0, \sigma = 1.5, \sigma = 2.0, \sigma = 2.5, \sigma = 3.0$ and $\sigma = 4.0$; $\omega = 1.0, \omega = 1.5, \omega = 2.0, \omega = 2.5, \omega = 3.0$ and $\omega = 3.5$; Gabor filter directions $0, \pi/4, \pi/2$ and $3\pi/4$ were used for cross comparison. Experimental results on our dataset indicated that with Gabor filter directions $0, \pi/4, \pi/2$ and $3\pi/4$, the Gabor filter banks have strong class representation ability and can achieve better segmentation effect with fewer filters. These parameters can thus effectively reduce the Gabor feature dimension. A detailed illustration on texture segmentation can be found in Appendix A. Fig. 11 shows the segmentation results of test images in the test set based on Gabor filter texture analysis using different parameters. Fig. 12 shows texture segmentation result of the proposed method.

C. TRAINING AND IMPLEMENTATION DETAILS

In the later comparative experiment, the networks U-net [27], SegNet [28], Faster R-CNN [29], Mask R-CNN [35] and FCN [17] used for comparison follow the same training strategy as our method. The networks were trained using Adam optimizer with mini-batch of size 5. Our proposed segmentation network is migrated from convolutional pooling layer parameters of vgg16 pre-trained on ImageNet as the initial parameters of our encoder network. The same migration strategy is used in the feature extraction part of the networks for comparison.

The cross-entropy loss function is used as the objective function to train the proposed network. In order to improve the training speed and give consideration to the quality of training, we achieve the benefits of decaying the learning rate during training. The decay rate is set to 0.1, and the initial learning rate is to 0.001 to accelerate convergence. When the

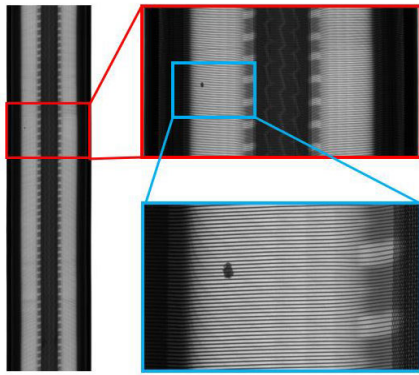


FIGURE 5. Defective samples, from top to bottom: sidewall-impurity, tread-impurity, sidewall-split and sidewall-overlap.

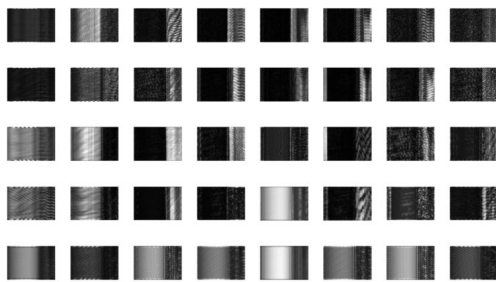
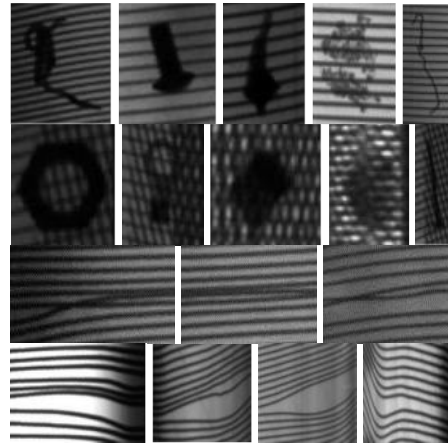


FIGURE 6. Experimental results on test image using 40 Gabor filters with different scales and orientations.

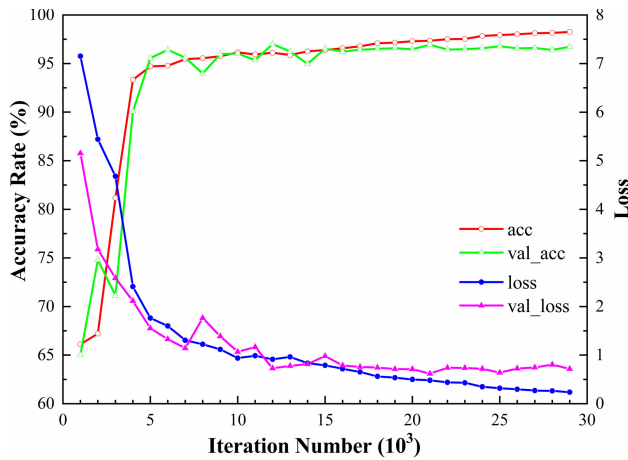


FIGURE 7. Schedules for the learning rate during the training process.

loss does not decrease for 3000 iterations, the new learning rate is updated as $0.1 \times lr$. At this time, when the *val-loss* of the network does not decrease for 5000 iterations, stop network training. Fig. 7 shows schedules for the learning rate during the training process.

It is worth noting that, in the comparative experiment, label settings of the SegNet [28] and Mask R-CNN [35] are different from that of our method. The former sets five types

TABLE 2. Training result of the state-of-the-art networks and ours.

Method	Model Size(MB)	Train Time(h)	Detection Speed(s/image)
U-net [27]	95	4.6	0.156
FCN [16]	139	5.5	0.18
SegNet [28]	176	6.1	0.318
Mask R-CNN [35]	268	10.2	0.415
Faster R-CNN [17]	276	8.2	0.512
Our method	83	4.2	0.132

of labels, namely, impurity, sidewall split, sidewall overlap, tread and sidewall background. In Faster R-CNN [29], four kinds of rectangular box labels, namely tread impurity, sidewall impurity, sidewall split and sidewall overlap, are used. We show the training result in Table 2.

According to the training results shown in Table 2, the proposed network has less network parameters and shorter training time than the others. The detection time for tire radiographic test set images is 0.132s per image on average, which can meet the needs of online tire detection. In order to verify the effect of texture segmentation preprocessing on defect semantic segmentation, we compared the samples after texture segmentation with the original samples without texture segmentation. 200 original tire images are selected as test samples, which contain 50 defects of 4 types. Experimental results show that the average test time for a single tire is 3.76s in the dataset without preprocessing however, for the preprocessed dataset, the average test time of a single tire is only 2.85s. Moreover, the mPA of the proposed method on the dataset preprocessed by texture segmentation is 1.21% higher than that of the dataset without preprocessing. These results indicate that the preprocessing method can reduce the computational complexity and, improve the detection efficiency and network performance.

TABLE 3. Comparative experimental results of the three architectures.

Defect	Test image	Ground truth	Model A	Model B	Model C
Sidewall impurity					
Tread impurity					
Sidewall split					
Sidewall overlap					

For the defect classifier module, the classification network uses random normal distribution to initialize the network weight. We set iteration to 1000 times. In the network training process, batch size, as the most important parameter of network, affects the training and performance of network. With a small batch size, the training is difficult to converge. With the increase of batch size, the speed of processing is faster. However, the larger batch size increases the calculation burden and it easily falls into the local optimal solution, which affects the detection accuracy of the model. To explore the optimal batch size, we use different batch sizes to train the network. The change of test set CA with batch size is shown in Fig. 8. According to the experimental results, we choose $batchsize = 32$. To prevent network over fitting, we set $Dropout = 0.5$.

D. EXPERIMENTAL RESULTS

We evaluate overall performance of the proposed scheme on our test set. We validate the architecture of defective region semantic segmentation module through comparison experiment. Comparison experiments with the state-of-the-art DCNN and semantic segmentation methods are performed.

1) SEMANTIC SEGMENTATION MODULE PARAMETER VALIDATION

Based on the baseline model, we first perform a comparison experiment on the three semantic segmentation module architectures that designed for comparison to evaluate the impact

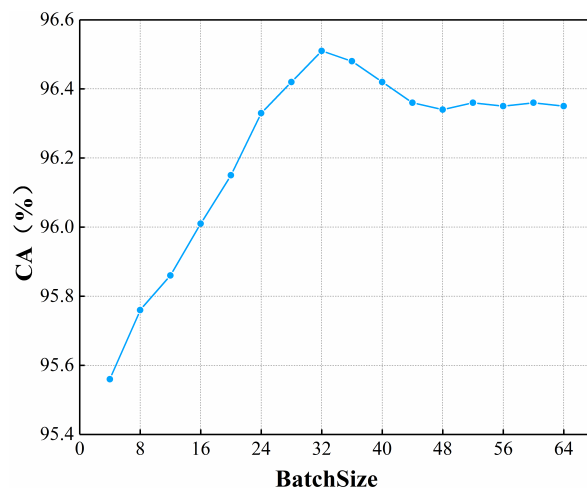


FIGURE 8. The CA of test set with different batch Size.

of each choice in the segmentation performance, as shown in Table 1.

We use the same data set to train the three models respectively. We evaluate and investigate the impact of different modules on each type tire defects respectively. Table 3 shows the comparative experimental results of the three architectures and ground truth segmentation. We can observe that the completeness of defective regions segmented by Model A is poor for sidewall impurity and sidewall split defects. Most importantly, as the features extracted are not enough,

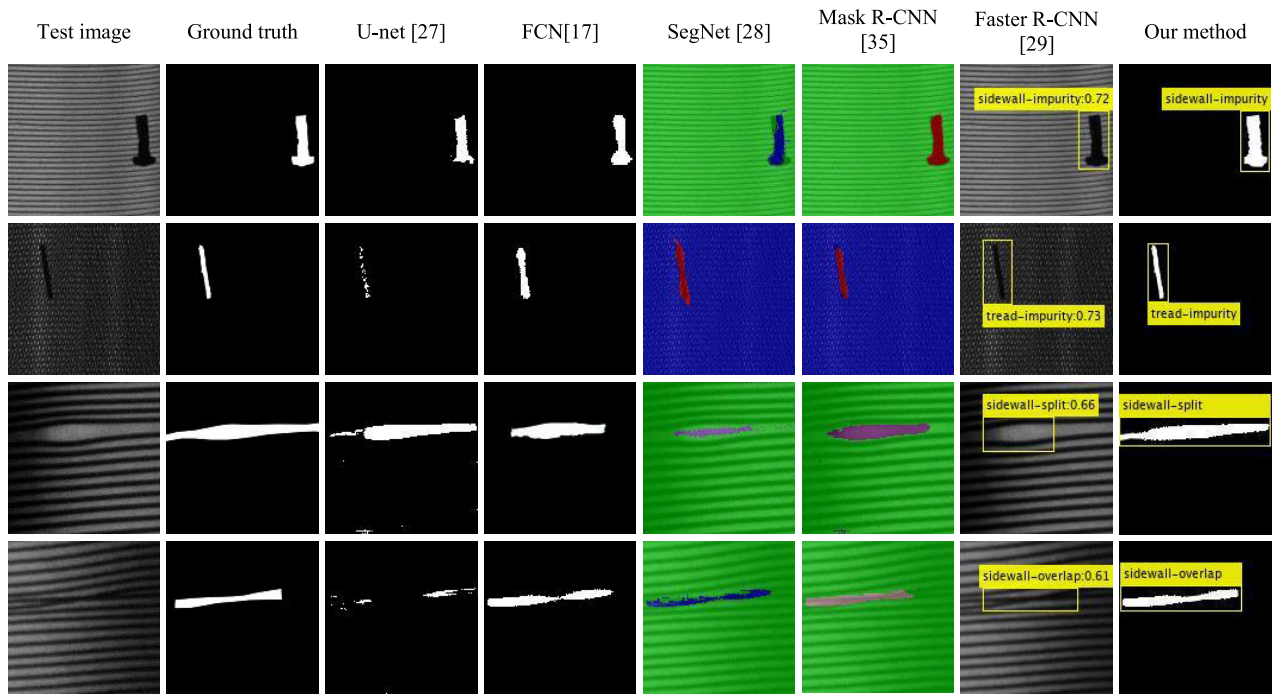


FIGURE 9. The result of comparative experiments between the proposed Concise-SSN and the state-of-the-art deep learning methods. From top to bottom: Result on sidewall impurity, tread impurity, sidewall split and sidewall overlap defects respectively.

TABLE 4. Semantic segmentation accuracy.

Defect	ModelA		ModelB		ModelC	
	PA%	IoU%	PA%	IoU%	PA%	IoU%
Sidewall impurity	72.4	70.8	85.6	73.2	92.32	75.92
Tread impurity	63.7	62.6	86.4	80.12	90.13	83.79
Sidewall split	65.1	61.5	64.5	61.7	85.90	83.45
Sidewall overlap	56.5	53.9	62.1	55.2	72.15	66.21

Model A does not detect the defects of tread impurities and sidewall overlap defects. On the other hand, the completeness of defective regions segmented by Model B is also poor for two kinds of cord defects. Model C has a good detection effect for all defect categories, the impurity defects segmented are slightly oversaturated than the ground truth. To comprehensively conclude our results, it could be regarded that Model C outperforms the other two models in tire defective region feature detection and extraction for semantic segmentation.

As shown in Table 4, segmentation accuracy between ground truth and the results of semantic segmentation module based on Model C is assessed by employing comparison metrics of PA and IoU.

The module reaches satisfactory results in PA, especially for impurity defect it achieves the highest scores as compared to other defects. It can be seen from the metrics that the defective semantic segmentation module can satisfy real-life tire defect segmentation application.

2) COMPARATIVE STUDY

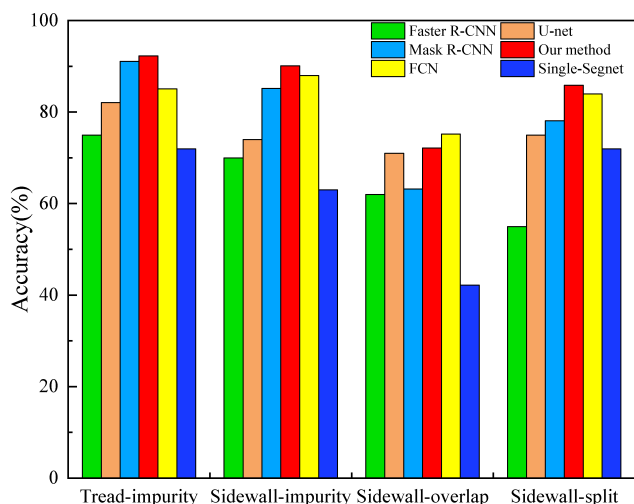
In order to evaluate the defect segmentation and classification performance of the proposed Concise-SSN, we compare the semantic segmentation and classification performance of our method with that of existing methods respectively.

The result of comparative experiments between the proposed Concise-SSN and conventional state-of-the-art deep learning methods (U-net [27], FCN [17], SegNet [28], Mask R-CNN [35], Faster R-CNN [29]) is shown in Fig. 9.

As can be seen from Fig. 9, the effect of end-to-end defect segmentation with single SegNet is poor which is because the gray value of impurity defects is close to the tread background. In the process of pixel semantic expression with a single SegNet, two kinds of defects, i.e. sidewall impurity and sidewall overlap, were mistakenly classified as tread background, resulting in the wrong classification of defects. Even though U-net has a good performance in the segmentation of medical images [27], the effect of using it in the detection of tire defects is unsatisfactory, especially for the

TABLE 5. Comparison results of tire defect classification.

Method	BP	HOG+SVM	KPCA+BP	Wavelet+ANN	Alexnet	Our method
CA%	86.2	68.9	89.1	91.3	94.6	96.5

**FIGURE 10.** The comparative experimental results using PA as evaluation index.

tread impurities defects and sidewall overlap defects. Faster R-CNN based detection method uses the region recommendation network (RPN) to recommend candidate regions, and then using the bounding box regression to modify the anchors to obtain accurate proposals. Through the ROI pooling layer, the input feature maps and proposals are collected. After synthesizing these information, the proposed feature maps are extracted and sent to the subsequent full connection layer to determine the category. As shown in Fig. 9, Faster R-CNN has a good effect on tire defect target detection, and can achieve the classification of defects in the target area. However, its detection accuracy is low, and most importantly, it cannot achieve the pixel-wise segmentation of defects. Compared with SegNet, Mask R-CNN has better result on all defect categories and background segmentation. Even though FCN and Mask R-CNN have better segmentation results than U-net and SegNet, further investigation is need for this real-life application.

Fig. 10 shows the comparative experimental results using PA as evaluation index. Consider that Faster R-CNN can't achieve pixel level target segmentation such that it can't be evaluated with PA. In Fig. 10, the data of Faster R-CNN method is the correct detection accuracy of defects. Among these methods, single SegNet method achieves worst the segmentation results in tread impurity, sidewall impurity and sidewall overlap, it achieves 42.2% for sidewall overlap defects. Mask R-CNN can achieve the correct semantic expression of the background texture and defects. In the detection of sidewall impurity, the PA value of Mask R-CNN

reaches 91.1% and the mPA value reaches 80.13%, but the segmentation performance is not as good as our method, the mPA value is 4.9% lower than our method. Faster R-CNN's performance is slightly better than that of single SegNet method. U-net performs slightly better than the previous two approaches and achieves 82.12% for tread impurity defects. Our method, by contrast, has the best performance among the five methods, it achieves 92.32% in PA for tread impurity defects and 90.13% for sidewall impurity, 72.15% for sidewall overlap and 85.90% for sidewall split defects respectively, and therefore it achieves the mPA score of 85.13%, the mIoU score of 77.34%. FCN achieves good detection accuracy (75.2%) on sidewall overlap defect segmentation, which is 3.05% higher than our method. However, for the other three kind of defects, the segmentation accuracies are not as good as that of ours, and its class average mPA is 2.12% lower than our method.

From the above results, we can also find that compared with the above methods, the segmentation accuracy of impurity defects is higher than that of cord defects. The overall segmentation accuracy of the proposed method is better than that of other methods.

To evaluate the classification performance of the compact CNN classifier module proposed in this paper, we compared the method with the commonly used machine learning based image classification methods, including BP neural network (BP) [30], support vector machine based on hog features (HOG + SVM) [31], kernel principal component analysis and neural network (KPCA + BP) [32], artificial neural network based on wavelet decomposition features [33] (Wavelet + ANN) and AlexNet [34]. Table 5 shows the comparison results of classification accuracy. As shown in Table 5, although traditional methods use artificial features to represent prior knowledge and then combine with machine learning to classify defects, they have lower classification accuracies. The reason is that, as previously described, there are many kinds of tire defects, and the features of various defects are quite different within and between categories, such that it is difficult to achieve a better classification effect with a single artificial feature. The proposed compact CNN classifier module in this paper realizes end-to-end defect classification and achieves a classification accuracy as high as 96.5%. More experimental results of the proposed method are shown in Fig. 13 in Appendix B.

V. CONCLUSION

In this paper, we proposed a novel defect inspection model Concise-SSN which is a concise two-step approach

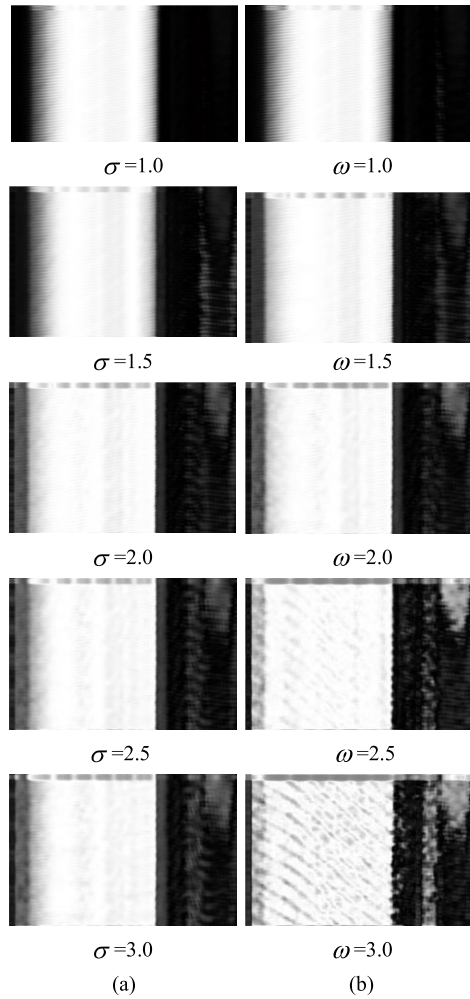


FIGURE 11. Clustering experimental results on test image using Gabor filters with different parameters. (a) Experimental results using different σ ($\omega = 1.75$); (b) Experimental results using different ω ($\sigma = 2$).

combining semantic segmentation module and lightweight CNN defect classifier module for intelligent tire defect inspection application. Considering the anisotropic and complex multi texture characteristics of tire radiographic image, for preprocessing, we proposed a texture segmentation method based on Gabor filter and fuzzy c-means clustering which significantly reduce the complexity and computation of the problem space in this application. The concise semantic segmentation module used an encoder-decoder framework for pixel wise defective region prediction, in which the encoder network produces these low-resolution representations of input tire gray level images. The corresponding decoder map the feature maps to full input resolution feature maps for pixel-wise predication. The proposed compact CNN classifier module classifies defective regions. From the experiments on our dataset, the proposed model achieved satisfactory performance compared to the existing state-of-the-art models. The results of comparative experiments also demonstrated the efficiency and effectiveness of our model to provide a real-time reliable automatic tire defect inspection scheme.

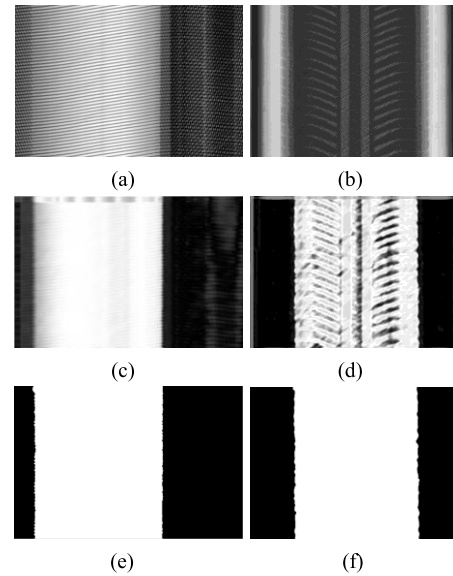


FIGURE 12. Experimental results on test images using the proposed method. (a) and (b) Test image; (c) and (d) Feature image filtered by Gabor filters; (e) and (f) Final segmentation result after clustering.

APPENDIX A

Two-dimensional Gabor function $g(x, y)$ can be defined as

$$g(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left(-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j\omega x \right) \quad (1)$$

For a given gray image $I(x, y)$, the two-dimensional Gabor wavelet transform can be defined as

$$W_{mn} = \iint I(x, y) g_{mn}^*(x - x_1, y - y_1) dx dy \quad (2)$$

where $*$ denotes the conjugate complex, (x_1, y_1) is a pixel in the image. The non-orthogonality of Gabor wavelet determines the information redundancy among the filtered images. To reduce redundancy, multi-channel filtering is designed. Let U_l and U_h be the range of center frequency from low to high, the ratio of the amplitude growth of the filter's two consecutive spectrum half peaks is a , the half peak amplitude of the minimum filter is t , then the width between U_l and U_h is

$$u_h - u_l = t + 2at + 2a^2t + \dots + 2a^{S-2}t + a^{S-1}t \quad (3)$$

namely

$$u_h - u_l = \frac{a + 1}{a - 1} (a^{S-1} - 1)t \quad (4)$$

where K denotes the number of orientations, S is the number of scales used in multi-scale decomposition.

For a Gaussian function with given σ , its half peak value is $\sigma\sqrt{2\ln 2}$, then the maximum half peak value of Gabor filter can be expressed as

$$a^{S-1}t = \sigma_u\sqrt{2\ln 2} \quad (5)$$

We have

$$u_h = a^{S-1}u_l \quad (6)$$

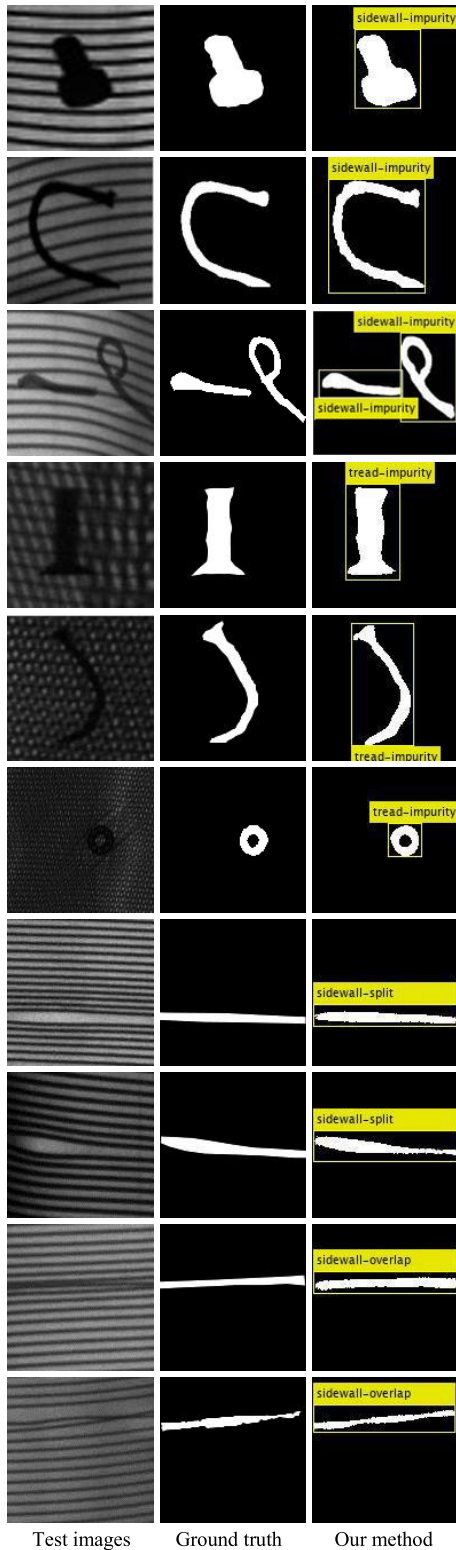


FIGURE 13. From top to bottom: Result on sidewall impurity, tread impurity, sidewall split and sidewall overlap defects respectively.

The maximum filter parameter σ_u can be given by

$$\sigma_u = \frac{a - 1}{a + 1} \frac{u_h}{\sqrt{2\ln 2}} \quad (7)$$

The Gabor filter bank $\theta = \pi/K$ can be obtained by rotating and scaling the Gabor function $g(x, y)$, K indicates the total number of directions. And the response of Gabor filter banks can contact each other without overlapping in the upper half of the spectrum, thus we have

$$\frac{(u - u_h)^2}{2\ln 2 \sigma_u^2} + \frac{v^2}{2\ln 2 \sigma_v^2} = 1 \quad (8)$$

and

$$\text{tg} \frac{\theta}{2} = \frac{u}{v} \quad (9)$$

The maximum filter parameters σ_v can be given as follows

$$\sigma_v = \text{tg} \frac{\theta}{2} \sqrt{\frac{u_h^2}{2\ln 2} - \sigma_u^2} \quad (10)$$

For (7) and (10), the standard deviation of the m th filter can be obtained by changing u_h to $u_m (m = 1, 2, \dots, S)$

$$\sigma_{mu} = \frac{a - 1}{a + 1} \frac{a^{m-1} u_l}{\sqrt{2\ln 2}} \quad (11)$$

$$\sigma_{mv} = \text{tg} \frac{\theta}{2} \sqrt{\frac{(a^{m-1} u_l)^2}{2\ln 2} - \sigma_{mu}^2} \quad (12)$$

In (1), the relationship between the standard deviation σ of Gaussian function along two coordinate axes and the central frequency ω of filter is

$$W_t \sqrt{2}\sigma \approx \sqrt{2}\pi/\omega \quad (13)$$

where W_t is Gabor wavelet time window. When the parameter σ is fixed, the center frequency of frequency domain is inversely proportional to the window width of time domain. If the real domain window is small, the center frequency of the Gabor filter is large (the frequency domain window is large). The filter can extract the high-frequency features to highlight the details of the image such that it is vulnerable to noise interference and vice versa.

Considering the arrangement of steel cord of radial tire, the complexity of calculation and the feature of image texture direction, the optimal parameters of Gabor filter bank corresponding to the segmented image are obtained through experiments. The direction parameters of Gabor filter are determined as $0, \pi/4, \pi/2$ and $3\pi/4$, that is, the direction parameter $k = 4$. After obtaining the texture image features in all directions through filter banks, the image features are classified by fuzzy clustering method, and then the segmentation results are obtained.

Assuming that every sample x_j belongs to a certain class, its membership function is equivalent to

$$\begin{aligned} & \hat{P}(\omega_i | x_k, \hat{\theta}) \\ &= \frac{(x_k | \omega_i, \hat{\theta}_i) \hat{P}(\omega_i)}{\sum_{j=1}^c P(x_k | \omega_j, \hat{\theta}_j) \hat{P}(\omega_j)} \\ &= \frac{|\hat{\Sigma}_i|^{1/2} \exp\left[-\frac{1}{2}(x_k - \hat{\mu}_i)^t \hat{\Sigma}_i^{-1} (x_k - \hat{\mu}_i)\right] \hat{P}(\omega_i)}{\sum_{j=1}^c |\hat{\Sigma}_j|^{1/2} \exp\left[-\frac{1}{2}(x_k - \hat{\mu}_j)^t \hat{\Sigma}_j^{-1} (x_k - \hat{\mu}_j)\right] \hat{P}(\omega_j)} \end{aligned} \quad (14)$$

where

$$\hat{P}(\omega_i) = \frac{1}{n} \sum_{k=1}^n \hat{P}(\omega_i|x_k, \hat{\theta}) \quad (15)$$

where $\theta = (\theta_1, \dots, \theta_c)^t$ is the unknown parameter vector, $\hat{\theta}$ is the maximum likelihood estimation of θ , $P(\omega_i)$ is the prior probability of the class ω_i , $\hat{P}(\omega_i)$ is the maximum likelihood estimation of $P(\omega_i)$, μ_i is the unknown mean vector of the class ω_i , and $\hat{\mu}_i$ is the maximum likelihood estimation of μ_i .

The goal of fuzzy c-means clustering algorithm is to minimize the global cost function

$$J_{fuz} = \sum_{i=1}^c \sum_{j=1}^n \left[\hat{P}(\omega_i|x_j, \hat{\theta}) \right]^2 \|x_j - \mu_i\|^2 \quad (16)$$

where b is a free parameter used to control the mixing degree of different categories. When $b = 0$, J_{fuz} is just the sum of squares error criterion, where each sample belongs to only one cluster. When $b > 0$, the criterion allows each sample to belong to multiple clusters.

The clustering membership function of each sample point is normalized

$$\sum_{i=1}^c (\omega_i|x_j) = 1, \quad j = 1, \dots, n \quad (17)$$

Let \hat{P}_j denote the prior category probability $\hat{P}(\omega_i)$. when solving (J_{fuz} reaches the minimum), there are

$$\partial J_{fuz} / \partial \mu_i = 0 \quad (18)$$

$$\partial J_{fuz} / \partial P_j = 0 \quad (19)$$

Therefore, the solution is as follows

$$\mu_i = \frac{\sum_{j=1}^n [\hat{P}(\omega_i|x_j)]^b x_j}{\sum_{j=1}^n [\hat{P}(\omega_i|x_j)]^b} \quad (20)$$

$$\hat{P}(\omega_i|x_j) = \frac{(1/d_{ij})^{1/(b-1)}}{\sum_{r=1}^c (1/d_{rj})^{1/(b-1)}} \quad (21)$$

$$d_{ij} = \|x_j - \mu_i\|^2 \quad (22)$$

Compared with the traditional clustering methods, the fuzzy c-means clustering algorithm improves the convergence performance.

Fig. 11 shows the segmentation results of test images in the test set based on Gabor filter texture analysis using different parameters. It is shown that when $\sigma = 2$ and $\omega = 1.75$ Gabor filter banks have strong texture class representation ability, and can achieve satisfactory segmentation effect with fewer filters, thus effectively reducing the Gabor feature dimension.

It can be seen from Fig. 12 (c) and (d) that Gabor filter banks can accurately segment different textures of tire radiographic image using the selected parameters. As shown in Fig. 12 (e) and (f), after fuzzy c-means clustering a binary segmentation can be achieved.

APPENDIX B

See Figure 13.

REFERENCES

- [1] S. Erdogan, "Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of turkey," *J. Saf. Res.*, vol. 40, no. 5, pp. 341–351, Oct. 2009.
- [2] C. Cai and L. He, "Improved Mach-Zehnder interferometer-based shearography," *Opt. Lasers Eng.*, vol. 50, no. 12, pp. 1699–1705, Dec. 2012.
- [3] H. L. M. Dos Reis and K. A. Warmann, "Acousto-ultrasonic non-destructive evaluation of fatigue damage in steel-belted radial tires," *Int. J. Fatigue*, vol. 18, no. 3, p. 216, Apr. 1996.
- [4] L. E. Roemer and N. Ida, "Location of wire position in tyre belting using Bayesian analysis," *NDT E Int.*, vol. 24, no. 2, pp. 95–97, Apr. 1991.
- [5] Q. Guo and Z.-W. Wei, "Tire defect detection using image component decomposition," *Res. J. Appl. Sci., Eng. Technol.*, vol. 4, no. 1, pp. 41–44, Jan. 2012.
- [6] Y. Zhang, T. Li, and Q. Li, "Defect detection for tire laser shearography image using curvelet transform based edge detector," *Opt. Laser Technol.*, vol. 47, pp. 64–71, Apr. 2013.
- [7] Y. Zhang, D. Lefebvre, and Q. Li, "Automatic detection of defects in tire radiographic images," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1378–1386, Jul. 2017.
- [8] Y. Zhang, T. Li, and Q. Li, "Detection of foreign bodies and bubble defects in tire radiography images based on total variation and edge detection," *Chin. Phys. Lett.*, vol. 30, no. 8, p. 84205, Aug. 2013.
- [9] F. Y. Li, "The study of an improved fuzzy edge detection algorithm in the radial tire quality detection," *Adv. Mater. Res.*, vols. 317–319, pp. 968–971, Aug. 2011.
- [10] Q. Guo, C. Zhang, H. Liu, and X. Zhang, "Defect detection in tire X-ray images using weighted texture dissimilarity," *J. Sensors*, vol. 2016, pp. 1–12, Feb. 2016.
- [11] G. Zhao and S. Qin, "High-precision detection of defects of tire texture through X-ray imaging based on local inverse difference moment features," *Sensors*, vol. 18, no. 8, p. 2524, Aug. 2018.
- [12] R. Ren, T. Hung, and K. C. Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Trans. Cybern.*, vol. 48, no. 3, pp. 929–940, Mar. 2018.
- [13] J. Jing, H. Ma, and H. Zhang, "Automatic fabric defect detection using a deep convolutional neural network," *Coloration Technol.*, vol. 135, no. 3, pp. 213–223, Jun. 2019.
- [14] W. Ouyang, B. Xu, J. Hou, and X. Yuan, "Fabric defect detection using activation layer embedded convolutional neural network," *IEEE Access*, vol. 7, pp. 70130–70140, Apr. 2019.
- [15] X. Cui, Y. Liu, Y. Zhang, and C. Wang, "Tire defects classification with multi-contrast convolutional neural networks," *Int. J. Pattern Recognit. Artif. Intell.*, vol. 32, no. 04, Apr. 2018, Art. no. 1850011.
- [16] Y. Zhang, X. Cui, Y. Liu, and B. Yu, "Tire defects classification using convolution architecture for fast feature embedding," *Int. J. Comput. Intell. Syst.*, vol. 11, no. 1, pp. 1056–1066, Aug. 2018.
- [17] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire defect detection using fully convolutional network," *IEEE Access*, vol. 7, pp. 43502–43510, Apr. 2019.
- [18] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. ECCV*, 2018, pp. 325–341.
- [19] R. Poudel, U. Bonde, S. Liwicki, and S. Zach, "ContextNet: Exploring context and detail for semantic segmentation in real-time," in *Proc. BMVC*, vol. 1805, Nov. 2018, p. 04554.
- [20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
- [21] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.
- [23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: <http://arxiv.org/abs/1409.1556>
- [24] C.-M. Lin, C.-Y. Tsai, Y.-C. Lai, S.-A. Li, and C.-C. Wong, "Visual object recognition and pose estimation based on a deep semantic segmentation network," *IEEE Sensors J.*, vol. 18, no. 22, pp. 9370–9381, Nov. 2018.

- [25] B. Lei, S. Huang, R. Li, C. Bian, H. Li, Y.-H. Chou, and J.-Z. Cheng, "Segmentation of breast anatomy for automated whole breast ultrasound images with boundary regularized convolutional encoder-decoder network," *Neurocomputing*, vol. 321, pp. 178–186, Dec. 2018.
- [26] E. Aguilar, B. Remeseiro, M. Bolaños, and P. Radeva, "Grab, pay, and eat: Semantic food detection for smart restaurants," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3266–3275, Dec. 2018.
- [27] Y. Hiasa, Y. Otake, M. Takao, T. Ogawa, N. Sugano, and Y. Sato, "Automated muscle segmentation from clinical CT using Bayesian U-Net for personalized musculoskeletal modeling," *IEEE Trans. Med. Imag.*, vol. 39, no. 4, pp. 1030–1040, Apr. 2020.
- [28] S. Alqazzaz, X. Sun, X. Yang, and L. Nokes, "Automated brain tumor segmentation on multi-modal MR image using SegNet," *Comput. Vis. Media*, vol. 5, no. 2, pp. 209–219, Jun. 2019.
- [29] X. Sun, J. Gu, R. Huang, R. Zou, and B. Giron Palomares, "Surface defects recognition of wheel hub based on improved faster R-CNN," *Electronics*, vol. 8, no. 5, p. 481, Apr. 2019.
- [30] W. K. Wong, C. W. M. Yuen, D. D. Fan, L. K. Chan, and E. H. K. Fung, "Stitching defect detection and classification using wavelet transform and BP neural network," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3845–3856, Mar. 2009.
- [31] S.-H. Lee, M. Bang, K.-H. Jung, and K. Yi, "An efficient selection of HOG feature for SVM classification of vehicle," in *Proc. Int. Symp. Consum. Electron. (ISCE)*, Jun. 2015, pp. 1–2.
- [32] Y. Jie et al., "Classification of multispectral remote sensing image using kernel principal component analysis and neural network," in *Proc. MIPPR*, vol. 7496, Oct. 2009, Art. no. 74961N.
- [33] P. John, "Brain tumor classification using wavelet and texture based neural network," *Int. J. Sci. Eng. Res.*, vol. 3, no. 10, pp. 1–7, 2012.
- [34] W. Nawaz, S. Ahmed, A. Tahir, and H. A. Khan, "Classification of breast cancer histology images using alexnet," in *Proc. ICIAR*, 2018, pp. 869–876.
- [35] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," 2017, *arXiv:1703.06870*. [Online]. Available: <http://arxiv.org/abs/1703.06870>



ZHOUSHOU ZHENG received the B.E. degree in mechanical design and manufacturing from Shandong Yingcai University, Jinan, China, in 2018. He is currently pursuing the M.E. degree in mechanical and electrical engineering with the Qingdao University of Science and Technology, Qingdao, China.

His research interests include image processing, computer vision, deep learning, and nondestructive testing.



SEN ZHANG received the B.E. degree in mechanical and electronic engineering from Heze University, Shandong, China, in 2018. He is currently pursuing the M.E. degree in mechanical engineering from the Qingdao University of Science and Technology, Qingdao, China.

His research interests include image processing, computer vision, deep learning, and nondestructive testing.



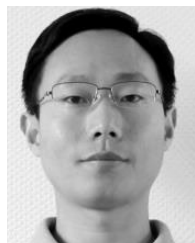
BIN YU received the M.Sc. degree in computational mathematics from Shanghai University, in 2005. He is currently pursuing the Ph.D. degree in bioinformatics with the University of Science and Technology of China.

From 2018 to 2019, he was a Visiting Professor with the Department of Biochemistry and Molecular Biology, Medical Genetics, and Oncology, University of Calgary, Canada. He is currently an Associate Professor with the College of Mathematics and Physics, Qingdao University of Science and Technology. He has published about 60 research articles in international journals and proceedings. His research interests include machine learning, image processing, and data mining.



QINGDANG LI received the B.Sc. degree in electronic engineering from the Wuhan University of Technology, in 1997, the Dipl.-Ing. degree in mechanical engineering from the University of Paderborn, Germany, and the Ph.D. degree from the University of Kassel, Germany, in 2006.

He has been with the Department of Technical Electronics, University of Kassel, since 2006. He serves as a Full Professor and a Specialist with the Qingdao University of Science and Technology. He is also an Advisor on industrial Internet for the Qingdao Regional Government, a Distinguished Expert of overseas Taishan scholars, a member of Chinese knowledge alliances for manufacturing innovation and the technical innovation consortium for the Shandong youth, a Speaker at the World Internet + Industry Conference in Qingdao, and a Co-Ordinator of cooperation activities with other faculties and institutes, in China and Germany.



YAN ZHANG (Member, IEEE) received the B.S. degree in computer science and technology from Northwestern Polytechnical University, Xi'an, China, in 2004, and the M.S. and Ph.D. degrees in automatic control from the Qingdao University of Science and Technology, Qingdao, China, in 2009 and 2014, respectively.

He was a Visiting Scholar with the Research Group on Electrical Engineering and Automatic Control (GREAH), Faculty of Sciences, Normandy University, Le Havre, France, during 2014 and 2015. He visited Shanghai Jiao Tong University, Shanghai, China, in 2018. He is currently an Associate Professor in electrical and computer engineering with the Qingdao University of Science and Technology. His research interests include artificial intelligence, digital image processing, pattern recognition, and nondestructive testing.

Dr. Zhang was awarded the Excellent Doctor Degree Dissertation of Shandong, China.

...