Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Tire Defect Detection Using Fully Convolutional Network

REN WANG[1,2], QIANG GUO [1,2], (Member, IEEE), SHANMEI LU[1,2], CAIMING ZHANG[3]
[1]School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China
[2]Shandong Provincial Key Laboratory of Digital Media Technology, Jinan, China
[3]Software College, Shandong University, Jinan, China

Corresponding author: Qiang Guo (e-mail: guoqiang@sdufe.edu.cn).

ABSTRACT Deep convolutional neural network has recently witnessed rapid progress due to the strong feature learning capability. In this paper, we focus on its application in the industrial field and propose a method based on fully convolutional network (FCN) for detecting defects in tire X-ray images. Owing to the capability of pixel-wise prediction of FCN, location and segmentation of defects are completed simultaneously. The network architecture used in the method mainly consists of three phases. The first phase is a traditional deep network which is used to extract the feature of tire images, and feature maps are obtained at last layer. By replacing fully connected layers into convolution layers, final feature maps retain sufficient spatial information. By adding up-sampling layers, in the second phase, outputs with the same size as the original image can be generated. After first two phases, we develop coarse segmentation results and refine them through fusing multi-scale feature maps. Experimental results show that the proposed method can accurately locate and segment defects in tire images.

INDEX TERMS Defect detection, convolutional neural network, object segmentation

I. INTRODUCTION

AUTOMATIC detection technology plays an important role in industrial quality inspection, which lowers the risk of human intervention in a hazardous environment. Compared with human inspection, automatic defect detection has high efficiency and excellent performance while reducing labor costs. Most existing defect detection methods are based on hand-crafted features, defects in images can be detected by using these low-level features. A suitable detection method can greatly improve the processing speed while ensuring accuracy.

Over the past two decades, automatic detection techniques have been widely used in industry inspection such as steel [1], titanium coated aluminum [2], solar wafers [3] and fabrics [4]–[6]. For simple fabric images with a repetitive unit-motif like wallpaper and ceramic, Li *et al.* [7] provided a feasible detection method based on combining low rank and sparse matrix analysis. For complex patterned fabric, Gao *et al.* [8] preprocessed images with Gabor filter on the basis of low rank. These methods lack adaptability although some of them have been used in industrial production. To improve overall performance, Li *et al.* [9] proposed a feature descriptor based on biological vision modeling by simulating the mechanism of biological visual perception, which is suitable for describing fabric images and superior to traditional hand-crafted feature descriptors. However, it is not suitable for applications in the tire industry, due to weak contrast and diversity of tire defects. Fig. 2 (a) shows a variety of tire defect images. It can be observed, there are certainly major difficulties in tire defect detection as follows.

1) Low visual quality. There are many uncertainties in the acquisition of tire images, due to differences in the types of machines and changes in the environment. On the other hand, the images used for automatic detection are derived from X-ray irradiation, and have some undesirable characteristics such as low contrast and low brightness.

2) Different texture structures. Generally, tire images consist of tread images and sidewall images. The tread is the part of the tire that comes in contact with the road surface, which is made up of thick rubber. The sidewall is largely rubber but reinforced with fabric or steel cords. Therefore, the texture features are completely different between tread and sidewall images. Due to the lower brightness of tread images, the defects occurring in the tread are more difficult to detect.

3) Diverse defects. There are similar types of defects distributed on the tread and sidewall, such as metallic impurities and bubbles, but some defects only exist in the tread

or sidewall images. The characteristics of various types of defects are obvious and there is a large gap between their textures. In general, impurities have sharp edges and are darker than their neighbors. For the bubble, it is brighter than their neighbors, although its texture is similar to the texture of defective-free parts. These differences in characteristics lead to challenges for tire defect detection.

As mentioned above, there are unique characteristics in tire images. Compared with defects in fabrics, the defects in the tire image are less distinguishable from the defective-free parts. Therefore, tire defect detection is more challenging. In view of the diversity of tire defects, the methods for analyzing the edges and texture features of defects have been widely used in detection tasks in recent years. They can be roughly classified into spatial domain methods and transform domain methods. Spatial domain methods usually employ low-level clues to produce regional detection results. For example, a contrast-based method was proposed in [10]. It first applies the local total variation filtering to decompose an image into structure and texture components, and then locates the defects by an adaptive thresholding operator. Guo *et al.* [11] used the local kernel regression descriptor to derive spatial texture features, and detected defects through weighted texture dissimilarity. The texture distortion degree of each pixel was estimated by weighted averaging of the dissimilarity between one pixel and its neighbors, which results in an anomaly map of the inspected image. To the best of our knowledge, it is the first work that can accurately locate the defects on tire thread images. However, due to the high computational complexity, it is not suitable for real-time detection tasks.

Besides, transform-based methods have also been applied into tire defect detection tasks. Wavelet transform, which has been proven to have excellent performance in analyzing one-dimensional signals, was introduced to tire defect detection tasks in [12]. Zhang *et al.* [13], [14] proposed a multi-scale transform based method to drive the edge information of defects. The curvelet transform was used to strengthen the edges of the image, and optimized Canny edge detection algorithm for locating defects. Furthermore, this method was improved by preprocessing the defect image with total variation. Compared with the wavelet transform, the curvelet-based method has obtained better results. However, edge detection based methods are not sensitive enough to defects in tread images. At the same time, the detection speed of the curvelet-based method is slow due to its high computation complexity. From different perspective, the projection transformation method was introduced in tire inspection tasks in [15]. Using the radon transform to perform multi-angle projection on the tire image can effectively detect the linear defects. However, due to the limitation of linear projection, it is not suitable for irregular defects. Most of the classical transform methods that use fixed transform kernels have a wide range of applications but are not targeted to tire images. In addition, a detection method based on dictionary representation was proposed in [16], which learned a dictionary from tire images. The position of the defects is detected based on the difference in distribution between the indications of defects and defect-free parts. Compared to fixed transform kernels, tire images are more accurately described by self-learning representations. In spite of this, detecting defects in the tire image, especially in tire tread images without regular texture and obvious distribution pattern, is still a challenging detection task.

In recent years, convolutional neural network (CNN) models have been proposed. A classical CNN consists of convolutional layers, pooled layers, and fully connected layers. Owing to the excellent feature extraction capability, it is often used to solve classic computer vision problems. Specially, CNN can extract key features of a dataset by self-learning when the artificial feature extraction fails, particularly for tire images with various characteristics. CNN is also introduced into the tire defect detection by Cui *et al.* [17] and has been proven to be feasible in dealing with this challenging task. This CNN-based method identified defect types by averaging five parallel network classification results. However, the position and shape of defects are not detected, and the performance of detection in tread images is significantly worse than that of the sidewall images.

Unlike the previous methods, we propose a tire detection method based on fully convolutional network (FCN), which is a CNN with the ability to preserve spatial information of features. In the FCN, all full-connected layers used in the traditional classification network are replaced by convolutional layers. Feature information is learned by the convolution and pooling layers, and retained in feature maps. Compared with the full-connected layer, which reduces each feature map to a vector and outputs label results, convolved full-connected layers can retain spatial information of the feature map to achieve accurate pixel-wise prediction and object segmentation. In this paper, we take advantage of the powerful self-learning and segmentation capability of FCN to overcome the deficiency of the traditional tire defect detection.

The rest of this paper is organized as follows. In Section II, we first briefly introduce the proposed method. Then the implementation details of this method are described in Section III. Section IV presents experimental results and analyses. A conclusion is drawn in Section V.

II. PROPOSED APPROACH

Our architecture is partly inspired by the FCN, whose performance has been validated in semantic segmentation tasks. We simplify a full convolution segmentation network into a binary-classification and pixel-wise prediction model, and combine different scale features to refine the defection results. The overall architecture of our approach is illustrated in Fig. 1. Each pixel in input image is tagged with category label through three phases, i.e., feature extraction, map up-sampling and multi-scale fusion. As the basic framework of the network, in the first phase, the classical VGG16 is used. It consists of the repeated application of a stack of 3×3 convolution layers and 2×2 pooling layers, each con-

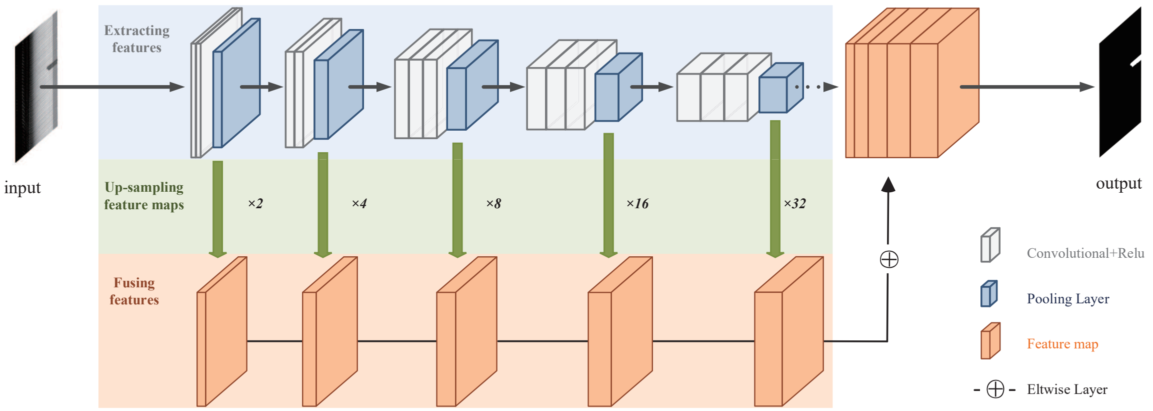

FIGURE 1. The network architecture used in this paper. In first phase, features of different scales are obtained through an original FCN based on VGG16. Then up-sampling layers are adopted to enlarge and align the feature map in up-sampling phase. Finally, multi-scale features are concatenated and fused using a simple linear strategy to obtain accurate results.

volution layer followed by rectified linear unit (ReLU) for non-linearity rectification. Then, the fully connected layers in VGG16 are convolved so that the generated feature maps can retain complete spatial information. Nevertheless, the extracted feature maps are down-sampled due to the existence of pooling layers. In the second phase, feature maps are enlarged to keep the size identical to input images through up-sampling layers. We can obtain multi-scale feature maps by feedforwarding an input image in the feature extraction module. In general, feature maps after pooling can be used to derive the detection result using class score maps obtained by the softmax layer. However, these results are generally coarse, due to feature maps after pooling layers contain insufficient texture features. Therefore, multi-scale feature maps are sequentially aligned and fused through the crop and the eltwise layer. Then refined detection results are derived by fusing different scales features. The following describes our model in details.

A. EXTRACTING FEATURES

As mentioned above, the features of tire images are firstly extracted by a convolutional network. In general, a classical CNN with the excellent feature extraction capability and less training parameters is usually expected. Here we use VGGNet [18] as our backbone architecture. VGGNet is based on AlexNet [19] by repeatedly stacking 3×3 small convolutional kernels and 2×2 maximum pooling layers, and has achieved good recognition performance in segmenting natural images. A stack of two 3×3 convolution layers has an effective receptive field of 5×5, and three such layers can replace 7×7. The configuration of convolution layers is the same in each stack. Therefore, the strong feature extraction capability is realized by increasing layer depth without reducing receptive field in VGGNet. In Section IV, we compare the performance of AlexNet and VGGNet in detecting tire defects.

Furthermore, we also compare VGGNet with different configurations and depths, such as VGG11, VGG13, VGG16. Due to its strong feature extraction capability, VGG16 is selected as the backbone. More specifically, VGG16 consists of 13 convolutional layers with a ReLU and 3 fully connected layers. Some convolutional layers are followed by a non-overlapping maximum pooling layer for filtering noisy features. The softmax layer is finally used for category prediction. In order to extract features, we completely preserve the convolutional and pooling layers, where the pooling layers are used to filter noisy features by abstracting features in a receptive field with a single representative value. The maps after each pooling layer involve different scale features. Besides, the fully connected layers are regarded as special convolutions with large receptive fields. The features with spatial information are remained through these special convolutional layers, which are beneficial for accurate detection.

B. UP-SAMPLING FEATURE MAPS

The feature maps with spatial information are derived from the feature extraction phase. However, these feature maps are down-sampled after the five pooling layers in the first phase. For instance, a 2×2 max-pooling layer with stride 2, which is designed to help classification by retaining only robust features, can reduce the image size by half. Due to the size reduction, the final output of the network can not indicate the probability of each pixel that belongs to one of the predefined classes. Therefore, feature maps are up-sampled in second phase so that it keeps the same size as the input images. In this paper, we use bilinear interpolation strategy for obtaining the enlarged feature map, which is initialized in network construction and updated during backpropagation. Bilinear interpolation can effectively reduce parameters without reducing the accuracy.

C. FUSING MULTI-SCALE FEATURES

The pooled feature maps can be directly up-sampled to obtain the pixel-wise predictions corresponding to original images. However, the standard pooling layers lose detailed textures while retaining high-level semantic information. These lost details are critical for accurately detecting defects. To address such issue, we fuse multi-scale feature maps to reduce the negative impact of the detail loss and refine the detection results. More concretely, feature maps obtained by the each pooling layer are sequentially up-sampled, as shown in the up-sampling phase of Fig. 1. These enlarged feature maps describe different scales information. Local details are involved in shallow layers and semantic information is involved in the deep layers. Due to up-sampling and padding, the size of cross layer maps are not consistent. Before fusing these maps, the enlarged feature maps must be aligned by cropping. More details about cropping are discussed in the next section. Then we fuse these maps through the simple element-wise operations, which has been proven to be valid in [20], and also use the softmax classifier as the end of the network for pixel-wise prediction.

III. IMPLEMENTATION DETAILS

Although a standard FCN performs well in processing natural images, the detail loss caused by pooling operation makes the segmentation results of small objects unsatisfactory. For instance, it is not sensitive to small defects in tire images, for example, bubbles. Meanwhile, the amount of training data is limited, and these lost information cannot be effectively recovered from other data samples. To overcome this issue, on one hand, we modify the traditional multi-classification neural network into a two-classification network. The reduction of categories can also prevent over-fitting and enhance the robustness of the network. On the other hand, multi-scale features derived from pooled layers of different depths are aligned and fused to complement the detailed texture. In this section, we discuss the implementation details of data alignment and fusion.

A. PIXEL ALIGNMENT

In order to derive dense prediction output from the feature map, we adopt the softmax classifier, which has been widely used in multi-classification and segmentation tasks. The calculation of the softmax loss function requires ground truths and feature maps of the same size. Meanwhile, the multi-scale feature map fusion in third phase is essentially an element-wise operation. Therefore, for aligning the corresponding pixels, we analyze the effect of different network components on the size of feature maps.

In convolutional layers, the size of the output feature map depends on the layer configuration. Specifically, writing $z_{i,j}$ for the pixel value at location (i,j) in convolved image, $x_{i,j}$ be the value at location (i,j) in image before convolution, and the convolutional mapping used in the proposed approach can be formulated as follows

$$z_{i,j} = g\Big(\sum_{m=0}^{k}\sum_{n=0}^{k} \omega_{m,n} x_{i+m,j+n} + \omega_b\Big), \tag{1}$$

where g is the activation function (ReLU), k is the convolutional kernel size, $\omega_{m,n}$ denotes the weight of pixel value at location (m,n). For a $W \times W$ input image, the size N of convolved image is calculated by

$$N = (W - k + 2p)/s + 1, \tag{2}$$

where p and s represent the padding and stride size in the convolution operation, respectively. In general, convolution layers have no influence on the image size, when the sizes of the convolution kernel and padding are 3 and 1.

With the help of up-sampling layers, the image size is enlarged for fusing multi-scale features. In this paper, we use bilinear interpolation method to up-sample feature maps and update its weight parameters during training. Simple bilinear interpolation computes each output y_{ij} from the nearest four inputs by a linear map that depends only on the relative positions of the input and output cells. The output y_{ij} can be written as

$$y_{ij} = \sum_{\alpha,\beta=0}^{1} |1-\alpha-\{i/f\}|\,|1-\beta-\{j/f\}|\, x_{\lfloor i/f\rfloor+\alpha,\lfloor j/f\rfloor+\beta}, \tag{3}$$

where f is the up-sampling factor, and $\{\cdot\}$ denotes the fractional part. In the concrete implementation, up-sampling layers are initialized using bilinear interpolation, then performed in-network for end-to-end learning through backpropagation from the pixel-wise loss. With some layer configurations, the feature map size is enlarged by non-integer multiples after up-sampling and convolved fully connected layers, for instance, the convolution kernel size and padding size are set to 7 and 0, respectively. However, a standard pooling layer strictly reduces the feature map size by half. Therefore, we add an cropping layer after each up-sampling layers to ensure the same size between feature maps and the ground truth.

B. FUSION STRATEGIES

As described in Section II-C, we fuse multi-scale features to obtain accurate prediction results. The feature maps obtained by each pooling layer are up-sampled to the same size and fused through a pixel-wise operation. Although this strategy is feasible, the concatenated features are stored in memory during fusion. In order to reduce memory overhead, in the experiment, we fuse score maps [20] corresponding to the features rather than fusing feature maps. These score maps are derived from feature maps by a 1×1 convolutional layer before each up-sample layer. Due to parameters of the convolutional layer are obtained by learning, the score fusion and the feature fusion can replace each other. Meanwhile, both of these two fusion strategies are linear operations, and the fusion of score maps has fewer parameters. Therefore, this trick is used to overcome the issue of insufficient memory during training.

IV. RESULTS AND DISCUSSION

Our proposed method has been implemented on the public FCN code [1], which was coded with Python 3.5 in the Caffe

[1] Available at https://github.com/shelhamer/fcn.berkeleyvision.org.

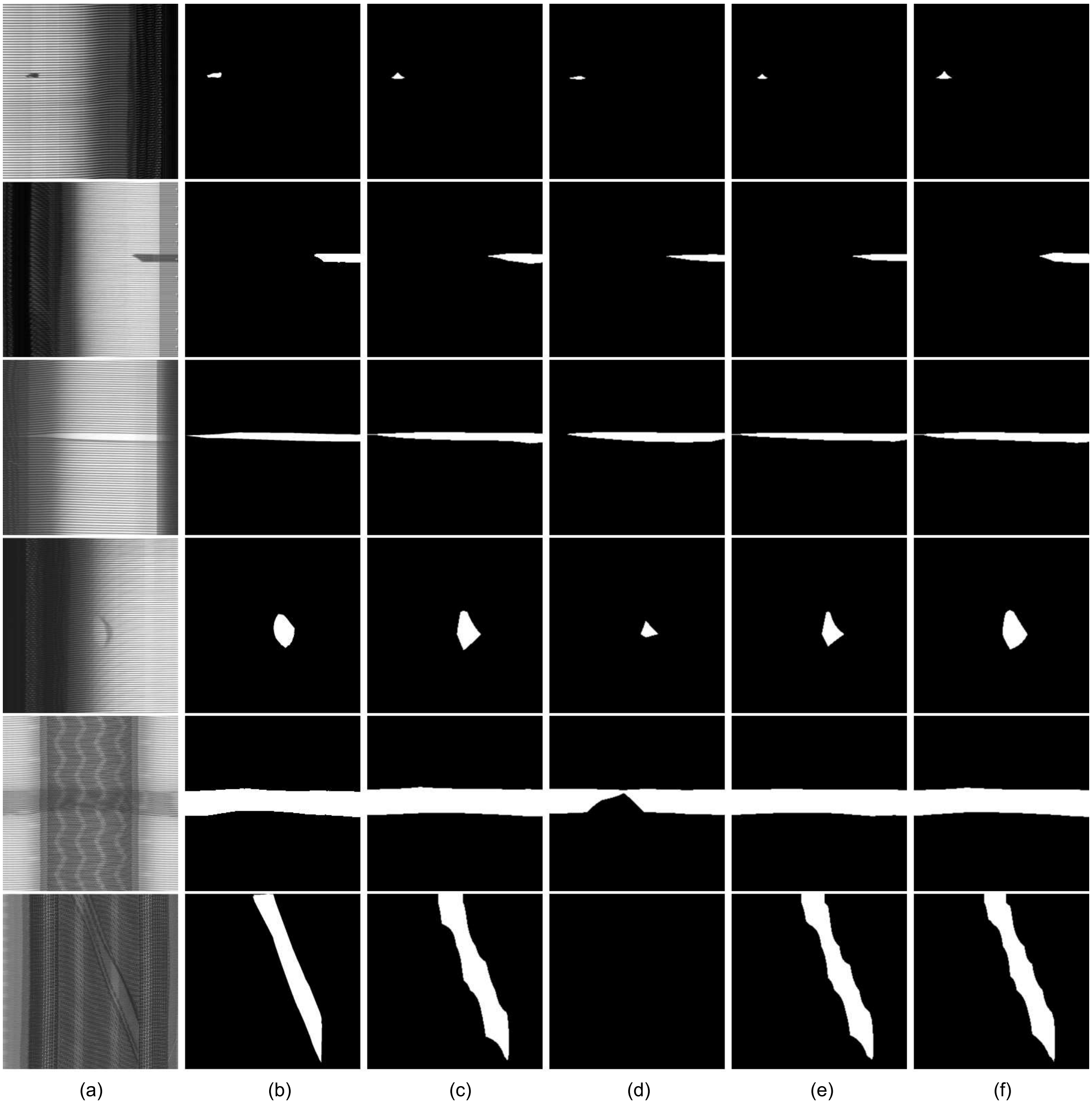

FIGURE 2. Detection results of several benchmark architecture. (a) shows input tire images with different defects. From top to bottom, the first four are tire sidewall images, which involve following defect types: impurity, overlap, slack, bubble. The last two are tire tread images, which involve overlaps. (b) indicates ground truths obtained by manual marking. (c),(d),(e) and (f) are detection results using AlexNet, VGG11, VGG13 and VGG16 as the basic architecture, respectively.

framework. A GTX-1080 GPU and Intel Xeon-E5 3.40GHz CPU are used for both training and testing. During training, we set the momentum parameter to 0.99 and the weight decay to 0.0005. The total number of iteration is set to 200k, and the proposed model is tested on the validation set every 2k iterations. Experimental results of the network with different configurations are reported and compared as follows.

A. DATASET

Our experimental dataset consists of 914 tire images [2], including both sidewall and tread images. Among them, 700 images are randomly selected as the training set, and the remaining 214 are used to form the testing set. These images involve various defects such as metal impurities, bubbles, and overlaps. All types of defects are treated as detection targets. Although images without defects are not included in our dataset to improve accuracy, there exists a small amount of defect-free pixels which are detected as defective. In practice, this problem can be solved by the post-processing step, such as the local variance analysis method [11]. During training, the data fed into the network contains corresponding ground truths in addition to the original tire images, where ground truths are labeled manually. More specifically, defective regions in images are regarded as objects, and defective-free regions as background. Their pixel values are marked

[2] Raw images are provided by Linglong Tyre Co. Ltd.

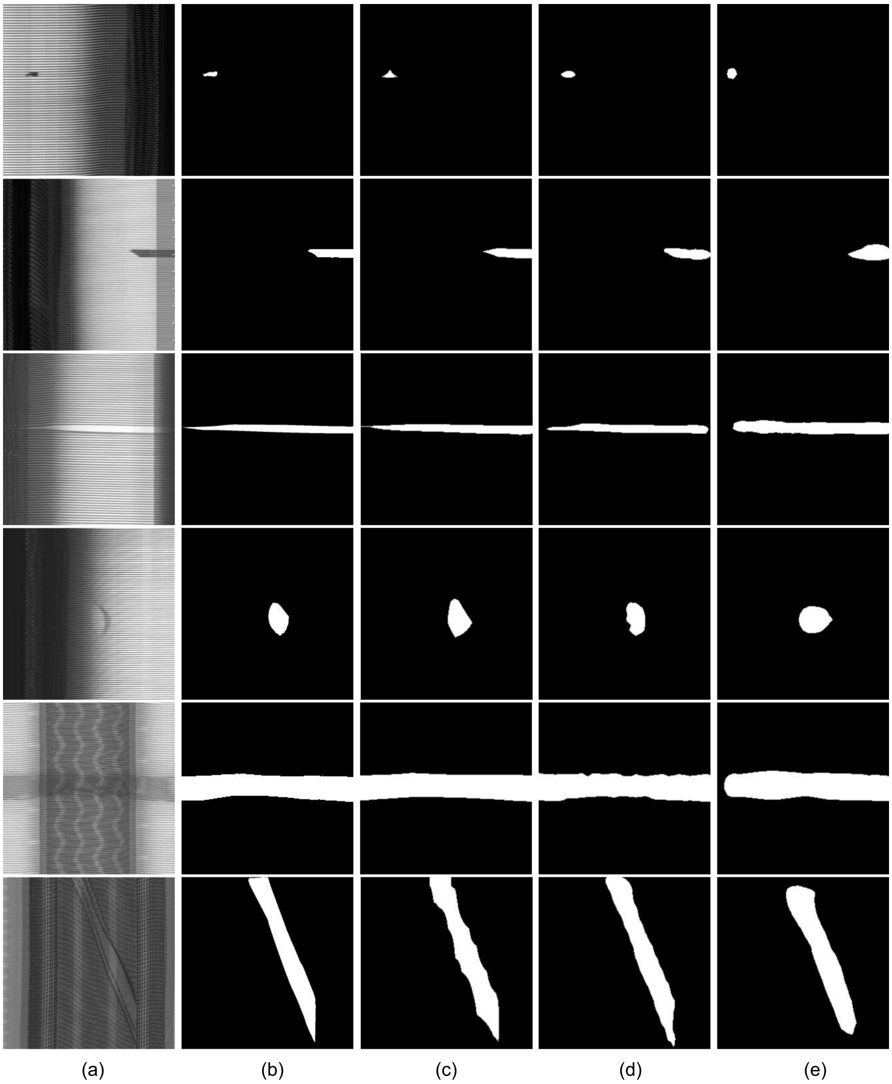

FIGURE 3. Results of ablation experiments on the optimal fusion layer number. (a), (b) indicate input tire images and ground truths as Fig. 2, respectively. (c) shows detection results using FCN without fusion. (d) shows detection results that fuse three scale features, which are derived from pool5, pool4, and pool3. (e) shows detection results that fuse four scale features, which are derived from pool5, pool4, pool3 and pool2.

as 1 and 0, respectively. In order to improve the detection accuracy, each ground truth is independently labeled by three persons, and finally determined by the voting strategy. For the testing set, its components are the same as the training set except that the ground truth is not used as a network input. In addition, to simplify the alignment operation, we scale all original images and ground truths to 256 × 256 before feeding into the network, although our network is not sensitive to the input image size.

B. ACCURACY COMPARISONS OF SEVERAL BENCHMARK ARCHITECTURE

In this experiment, we choose VGG16 as the basic architecture of the network. To verify the comprehensive performance in dealing with tire detection task, we compare candidate networks AlexNet, VGG11, VGG13 and VGG16. The structure and parameters of various networks are listed in Table 1. Obviously, a five-layer convolution and pooling structure are used in both AlexNet and VGGNet. Latter uses 3×3 convolution kernels and expands each layer into two or three sub-layers. Compared to AlexNet, VGGNet has the excellent extraction capability of detailed features without increasing parameters. Therefore, the detection of impurities, bubbles and other small defects is more effective in tire images. For clearer comparisons, convolved fully connected, softmax and loss layers are all kept unchanged in candidate networks. Then these networks are tested separately, and experimental results with different basic architectures are shown in Fig. 2. As can be seen, with the same number of iterations, VGG16 shows the strong feature extraction capability. Compared with other networks, VGG16 is more effective for representing tire images.

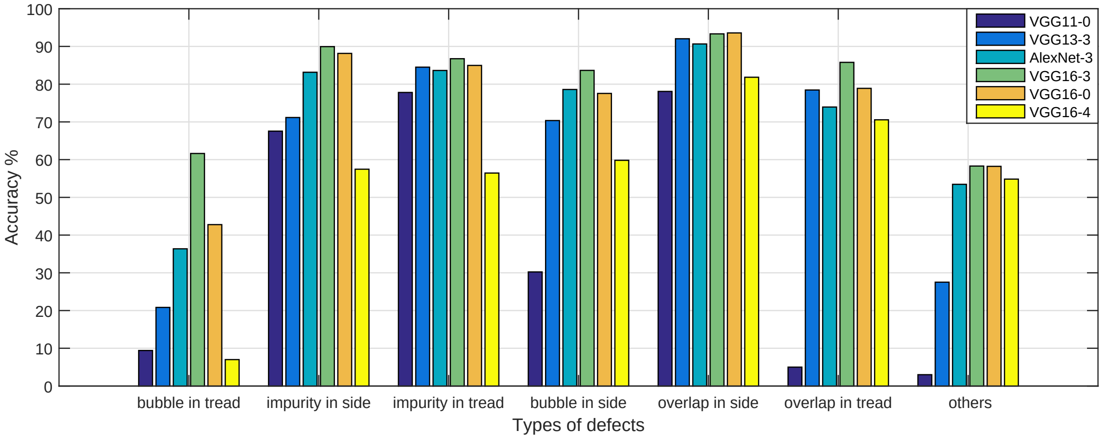

FIGURE 4. Accuracy (PA) of our approach on different types of tire defects. "-number" indicates the number of fusion layers. For instance, "VGG16-3" indicates VGG16 fused features from pool5, pool4 and pool3.

TABLE 1. Configurations and parameters of convolution and pooling layers in different basic networks

Net Configuration/Parameters			
AlexNet	**VGG11**	**VGG13**	**VGG16**
input images (256 × 256 × 1)			
conv×1 (11×11×96)	conv×1 (3×3×64)	conv×2 (3×3×64)	conv×3 (3×3×64)
maxpool (2 × 2)			
conv×1 (5×5×256)	conv×1 (3×3×128)	conv×2 (3×3×128)	conv×3 (3×3×128)
maxpool (2 × 2)			
conv×1 (3×3×384)	conv×2 (3×3×256)	conv×2 (3×3×256)	conv×3 (3×3×256)
	maxpool (2 × 2)		
conv×1 (3×3×384)	conv×2 (3×3×512)	conv×2 (3×3×512)	conv×3 (3×3×512)
	maxpool (2 × 2)		
conv×1 (3×3×256)	conv×2 (3×3×512)	conv×3 (3×3×512)	conv×3 (3×3×512)
maxpool (2 × 2)			

C. ABLATION EXPERIMENTS ON OPTIMAL FUSION LAYER NUMBER

As discussed above, FCN is not sensitive to small defects and edge details due to the information loss. We fuse different scale feature maps to complement lost details. To investigate the optimal number of fused layers, we conducted several ablation experiments. First, feature maps obtained by five pooling layers (pool5) are directly up-sampled and fed in the softmax classifier to derive score maps. In Fig. 3 (c), the predictions without multi-scale features are shown as a comparison. After that, the predictions from pool2, pool3 and pool4 are fused and shown in Fig. 3 (d) - (f), respectively. As can be seen from Fig. 3, VGG16-3 (VGG16 fused features from pool5, pool4 and pool3) has a positive effect on the overall results, it makes detection results more accurate in defect edge regions. However, the computational cost of the network increases as the number of layers increases. Moreover, fusing too many layers can increase the computational complexity and cause the segmentation result too smooth. To make a tradeoff, we adopt VGG16-3 in tire defects detection task.

D. QUANTITATIVE ANALYSIS

We use PA (pixel accuracy) as a standard metric to evaluate the accuracy of the proposed model. PA is a common strategy in the field of image segmentation, which indicates the ratio of the correct labeled pixels to the total pixels. Specifically, the assumptions are as follows: There are a total of k+1 classes in images. The p_{ij} represents the number of pixels that belong to class i but are predicted to be class j, and p_{ii} represents the correct number. Then PA can be defined as

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}}. \tag{4}$$

We select the several representative types of defects in the tire field, and PA values of different models are compared as shown Fig. 4. For common defect types, the VGG16-3 has a significant improvement than other structures, even small size defects, whose PA values are inferior to other defect types. The detection performance in the sidewall is better than that in the tire tread. As comparisons, quantitative results of detection accuracy are shown in Table 2. We can observe that the VGG16 with three fused layers can not only accurately detect different scales defects but also generate more precise prediction results in different defect types.

E. COMPARISON WITH TRADITIONAL METHODS

The defect detection task has been around for a long time. To verify the effectiveness of the proposed method, we also compare two traditional methods: a wavelet transform-based

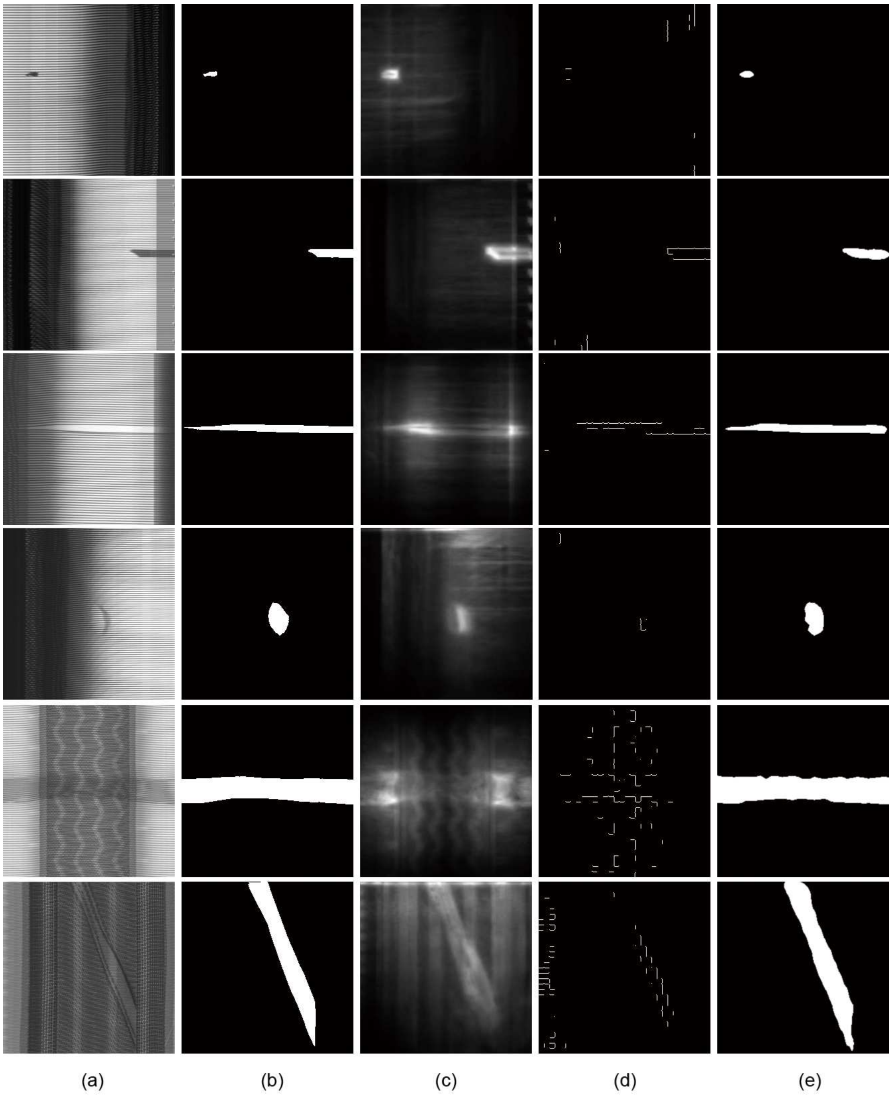

FIGURE 5. Result comparisons with traditional methods. (a), (b) indicate input tire images and ground truths as Fig. 2, respectively. (c) shows detection results using the context-aware saliency detection method [21]. (d) shows detection results using the transformation-based detection method [12]. (e) shows detection results using the proposed method (VGG16-3).

TABLE 2. The average detection accuracy of different basic networks with different fuse strategies.

Basic Net	Alex	VGG11	VGG13	VGG16	VGG16	VGG16
Fusion Layers	3	0	0	0	3	4
Accuracy	71.40	37.58	63.55	74.87	**78.91**	54.42

(WT) method [12] and a saliency detection (SD) method [21] based on context-aware. The former uses local regularity analysis and scale characteristics to represent the tire defects, where the optimal threshold parameters are selected by a defect edge measurement model. Then wavelet multi-scale analysis are used to separate the defects from the background textures. The latter combines local low-level and global clues to detect salient objects, which can be used to detect defects. Fig. 5 shows experimental results of these two methods used in the tire defect detection. As can be seen, the WT method is suitable for defects with the significant edge in sidewall images, like impurities and overlaps. For tire tread images, it has unsatisfactory results due to the interference of defective-free region textures. The SD method focuses on saliency regions rather than defects. Therefore, the method has high missing and inaccurate detection rates. Compared with the SD method, the proposed method can obtain more accurate results. In addition, our method is superior to traditional methods of comparison in detecting challenging type defects, especially in the tread image.

V. CONCLUSION

This paper explores the solution for the tire defect detection using FCN, which has outstanding performance in solving segmentation problems. With the feature extraction ability, VGG16 is constructed as the basic architecture to represent tire images. We fine-tune the parameters and structure of FCN to obtain coarse detection results, and refine results by a fusion strategy. Experiments show that the proposed method is applicable to more types of defects compared with traditional methods. Unlike the existing learning based method [17] in the tire industry, our algorithm can directly segment defects, and is valid for both the sidewall and tread images.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (61873145, 61572286, 61472220), NSFC Joint Fund with Zhejiang under Key Project (U1609218), Natural Science Foundation of Shandong Province for Excellent Young Scholars (ZR2017JL029), and Fostering Project of Dominant Discipline and Talent Team of Shandong Province Higher Education Institutions. The authors would like to thank Linglong Tyre Co. Ltd for providing raw tire images.

REFERENCES

[1] S. Ghorai, A. Mukherjee, M. Gangadaran, and P. K. Dutta, "Automatic defect detection on hot-rolled flat steel products," IEEE Trans. on Instrumentation and Measurement, vol. 62, no. 3, pp. 612–621, 2013.

[2] M. Win, A. Bushroa, M. Hassan, N. Hilman, and A. Ide-Ektessabi, "A contrast adjustment thresholding method for surface defect detection based on mesoscopy," IEEE Trans. on Industrial Informatics, vol. 11, no. 3, pp. 642–649, 2015.

[3] D. Tsai, S. Wu, and W. Chiu, "Defect detection in solar modules using ICA basis images," IEEE Trans. on Industrial Informatics, vol. 9, no. 1, pp. 122–131, 2013.

[4] A. Kumar, "Computer-vision-based fabric defect detection: A survey," IEEE Trans. on Industrial Electronics, vol. 55, no. 1, pp. 348–363, 2008.

[5] H. Ngan, G. Pang, and N. Yung, "Automated fabric defect detection—a review," Image and Vision Computing, vol. 29, no. 7, pp. 442–458, 2011.

[6] Y. Li, W. Zhao, and J. Pan, "Deformable patterned fabric defect detection with Fisher criterion-based deep learning," IEEE Trans. on Automation Science and Engineering, vol. 14, no. 2, pp. 1256–1264, 2017.

[7] P. Li, J. Liang, X. Shen, M. Zhao, and L. Sui, "Textile fabric defect detection based on low-rank representation," Multimedia Tools and Applications, pp. 1–26, 2017.

[8] G. Gao, D. Zhang, C. Li, Z. Liu, and Q. Liu, "A novel patterned fabric defect detection algorithm based on GHOG and low-rank recovery," in Proc. IEEE 13th International Conference on Signal Processing, pp. 1118–1123, 2016.

[9] C. Li, G. Gao, Z. Liu, M. Yu, and D. Huang, "Fabric defect detection based on biological vision modeling," IEEE Access, vol. 29, no. 7, pp. 27659–27670, 2018.

[10] Q. Guo and Z. Wei, "Tire defect detection using image component decomposition," Research Journal of Applied Sciences, Engineering and Technology, vol. 4, no. 1, pp. 41–44, 2012.

[11] Q. Guo, C. Zhang, H. Liu, and X. Zhang, "Defect detection in tire X-ray images using weighted texture dissimilarity," Journal of Sensors, vol. 2016, pp. 868–880, 2016.

[12] Y. Zhang, D. Lefebvre, and Q. Li, "Automatic detection of defects in tire radiographic images," IEEE Trans. on Automation Science and Engineering, vol. 14, no. 3, pp. 1378–1386, 2017.

[13] Y. Zhang, T. Li, and Q. Li, "Defect detection for tire laser shearography image using curvelet transform based edge detector," Optics & Laser Technology, vol. 47, pp. 64–71, 2013.

[14] 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," Chinese Physics Letters, vol. 30, no. 8, pp. 084205, 2013.

[15] C. Zhang, X. Li, Q. Guo, X. Yu, and C. Zhang, "Texture-invariant detection method for tire crack," Journal of Computer-Aided Design & Computer Graphics, vol. 25, no. 6, pp. 809–816, 2013. (in Chinese)

[16] Y. Xiang, C. Zhang, and Q. Guo, "A dictionary-based method for tire defect detection," in Proc. IEEE International Conference on Information and Automation, pp. 519–523, 2014.

[17] X. Cui, Y. Liu, Y. Zhang, and C. Wang, "Tire defects classification with multi-contrast convolutional neural networks," International Journal of Pattern Recognition and Artificial Intelligence, vol. 32, no. 4, Article 1850011, 2018.

[18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. International Conference of Learning Representation, 2015.

[19] A. Krizhevsky, I. Sutskever, and G. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in Neural Information Processing Systems, pp. 1097–1105, 2012.

[20] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 3431–3440, 2015.

[21] G. Stas and Z. Lihi and T. Ayellet, "Context-aware saliency detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 34, no. 10, pp. 1915–1926, 2012.

REN WANG received the B.S. degree in computer science and technology from Ludong University, Yantai, China, in 2017. He is currently pursuing the M.S. degree in computer application with Shandong University of Finance and Economics. His research interests lie in the domain of image classification and segmentation, computer vision, and machine learning.

QIANG GUO received the BS degree from Shandong University of Technology, Zibo, China, in 2002, the MS and PhD degrees from Shanghai University, Shanghai, China, in 2005 and 2010, respectively. From 2012 to 2015, he was a postdoctoral fellow with Shandong University, Jinan, China. He is currently an associate professor in the School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan, China. His research interests include image restoration, sparse representation, and object detection. He is a member of the IEEE.

SHANMEI LU received the B.S. degree in digital media technology from Shandong University of Finance and Economics in 2018. She is currently pursuing the M.S. degree in computer application technology with Shandong University of Finance and Economics. She is currently an aspiring Computer Vision Researcher. Her research interests lie in the domain of saliency detection using convolutional neural networks and machine learning.

CAIMING ZHANG received the BS and MS degrees from Shandong University, Jinan, China, in 1982 and 1984, respectively, and the PhD degree from the Tokyo Institute of Technology, Tokyo, Japan, in 1994. From 1998 to 1999, he was a postdoctoral fellow with the University of Kentucky, Lexington. He is currently a professor with Shandong University, and a distinguished professor with Shandong University of Finance and Economics. His research interests include computer vision, computer aided geometric design, computer graphics, information visualization, and medical image processing.