MULTI-SCALE DEFECT DETECTION NETWORK FOR TIRE X-RAY IMAGES

Ren Wang[1,2], *Qiang Guo*[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

ABSTRACT

Though automatic detection method has been tremendous improved, with the development of deep learning. Defect detection in many industrial inspection is one of the remaining challenging tasks due to the diversity of products. In this work, we focus on detection tasks in tire industry and develop a *Multi-scale Defect Detection Network (MDDN)*, which contains two parallel sub-networks to capture semantic and texture features. Specifically, high-abstracted semantic features containing defect shapes and locations are mined via a *semantic-aware sub-network*, simplified by an off-the-shelf fully convolutional network. Furthermore, to complement the details filtered by the sub-sampling, a novel *texture-aware sub-network* is used to cover edge features and small defects as much as possible. Finally, pixel-wised detection results are obtained by fusing features with semantic and texture information. Extensive experiments demonstrate that MDDN can produce comparable results and achieve significantly performance improvement in small tire defects detection over existing methods.

Index Terms— Defect detection, Fully convolutional network, Semantic segmentation, Multi-scale context

1. INTRODUCTION

Automatic defect detection, used to improve quality and accelerate production, has become an indispensable part in industrial processes [1, 2]. Especially in tire manufacturing, numerous detection algorithms have been proposed [3, 4, 5] and aroused extensive attention recently. In most real-world applications, tire defect detection is first carried out by deriving the defective region from tire X-ray images, which contains various types of defects caused by unclean raw materials and undesired manufacturing facilities [6]. Then, defective products are hierarchical processed according to the location and area of defects. Due to unique properties of the tire X-ray image, for instance complexity and low-quality, illustrated in previous studies [7, 8], most inspection processes are performed by human observers, which increases the risk and reduces the efficiency. Therefore, tire defect detection remains one of the most challenging inspection tasks.

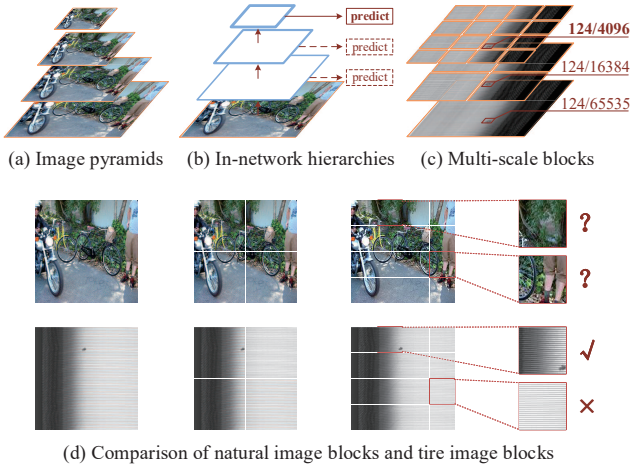

Fig. 1. (a) shows the image pyramids. (b) indicates the in-network feature hierarchies. Prediction results can be derived from each layer. (c) represents multi-size tire image blocks, which can increase the relative scale of small defects and backgrounds. (d) illustrates that the semantic information is more easily retained in tire image blocks by comparison.

At present, existing detection methods are mostly devoted to distinguishing difference between defective regions and backgrounds (defective-free regions). A key issue for such methods is feature extraction. Guo *et al*. [9] proposed a component decomposition based method to detect tire defects, which separated the background from the image by means of two designed filters. Then through an adaptive thresholding processing, defects were derived from the residual image. Besides, independent component analysis was also used for defect detection tasks [10, 11]. A major disadvantage of these fundamental methods is the limitation of the information contained in low-level clues and domain features. To address the limitation, Zhang *et al*. [7, 12] introduced mulit-scale wavelet and curvelet transform in detection tasks. Furthermore, optimized edge detection and total variation algorithm were used

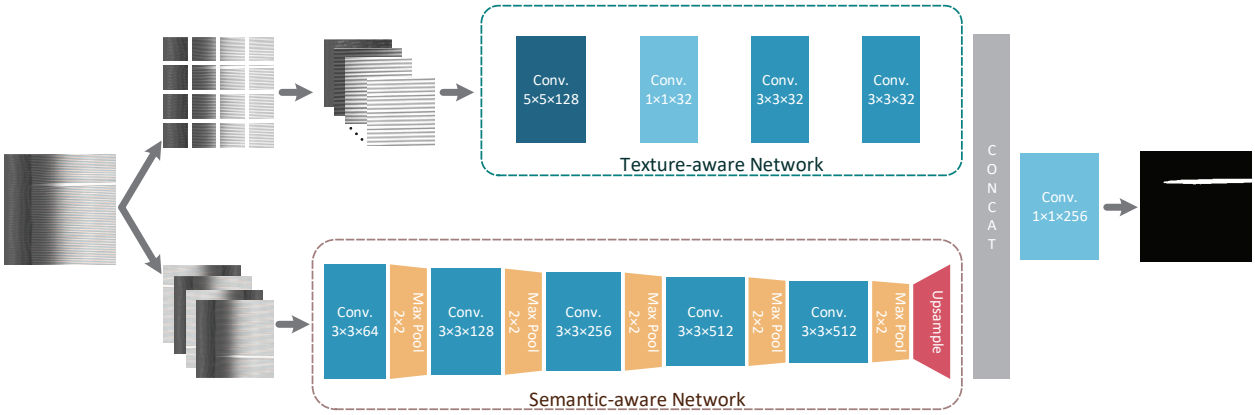

Fig. 2. Overall architecture of the proposed MDDN. A novel texture-aware network and a semantic-aware network with different preprocessing strategies are adopted to extract low-level clues and high-level features, respectively. Then, the detection results are obtained by fusing these features.

to achieve more accurate results [13]. However, the representation capability of fixed kernels is not comprehensive enough. Moreover, the transform processes are computationally expensive. Recently, Cui *et al*. [14] attempted to classify tire defects by means of deep convolutional networks (ConvNets), which have outstanding performances in the recognition and segmentation tasks of natural images. With the excellent feature extraction capability of ConvNets, Wang *et al*. [8] further implemented the detection and segmentation in tire images by a fully convolutional network (FCN) [15]. However, due to the existence of pooling layers, FCN is not sensitive to small defects and edge details, which is similar to that in dealing with natural image segmentation tasks.

To achieve better detection performance on these small objects, many methods have been proposed in benchmark datasets. Most of them are based on multi-scale strategies and can be roughly classifiified into image pyramids and in-network feature hierarchies. Image pyramids are directly scaled to get multi-scale images and extensively used in the era of hand-crafted features [16], as shown in Fig. 1(a). With the popularity of self-learning feature representation, image pyramids based methods are impractical for real applications due to the considerable increase in inference time. In-network feature hierarchies are formed by the forward propagation within deep ConvNets. Through several of sub-sampling layers, in-network hierarchies produce feature maps including different spatial resolutions, with an multi-scale and pyramid shape [17]. The Single Shot Detector (SSD) [18] is one of the first attempts at combining predictions from these features maps to detect objects of various sizes. Generally, shallow features are used to predict small objects, and deep features with large receptive fields are used to detect large objects. However, the lack of semantic information is harmful to the detection of small targets in shallow layers. Another fusing way can effectively address this problem by concatenating multi-scale features and detecting on top of the expanded feature maps, as shown in Fig. 1(b). For example, FCN defined a skip architecture to produce more accurate segmentation. Similar top-down skip architectures are popular in recent researches [19, 20]. Since the training goal of the entire network is to extract abstract semantics, there exists a basic problem that it is still not enough to mine the detail texture in these structures [21]. CrowdNet [22] combined shallow and deep networks to overcome this shortcoming. Unlike in-network feature hierarchies, the shallow network is specifically designed to retain more details by reducing the number of pooling layers. Therefore, the capability to extract detail features is further enhanced.

Inspired by CrowdNet, we construct a end-to-end network named *Multi-scale Defect Detection Network (MDDN)* that consists of a semantic-aware sub-network and a texture-aware sub-network. In texture-aware networks, pooling layers are discarded in order to completely retain detail textures and small-sized defects. Furthermore, image blocking [22, 23] was adopted as the preprocessing strategy during training. Unlike natural image blocks, defects (objects) are still significant and discernible in tire image blocks, illustrated in Fig. 1(d). In addition, as shown in Fig. 1(c), the proportion of defects in a 256×256 tire image increases from 124/65535 to 124/4096, which is advantageous for better capturing of detailed information.

2. MULTI-SCALE DEFECT DETECTION NETWORK

Our network aims to improve the accuracy of tire defect detection, especially for edge details and small defects. To achieve this goal, on the one hand, the proposed MDDN combines deep layer features with low-level clues through two parallel sub-networks, illustrated in Fig. 2. Among them,

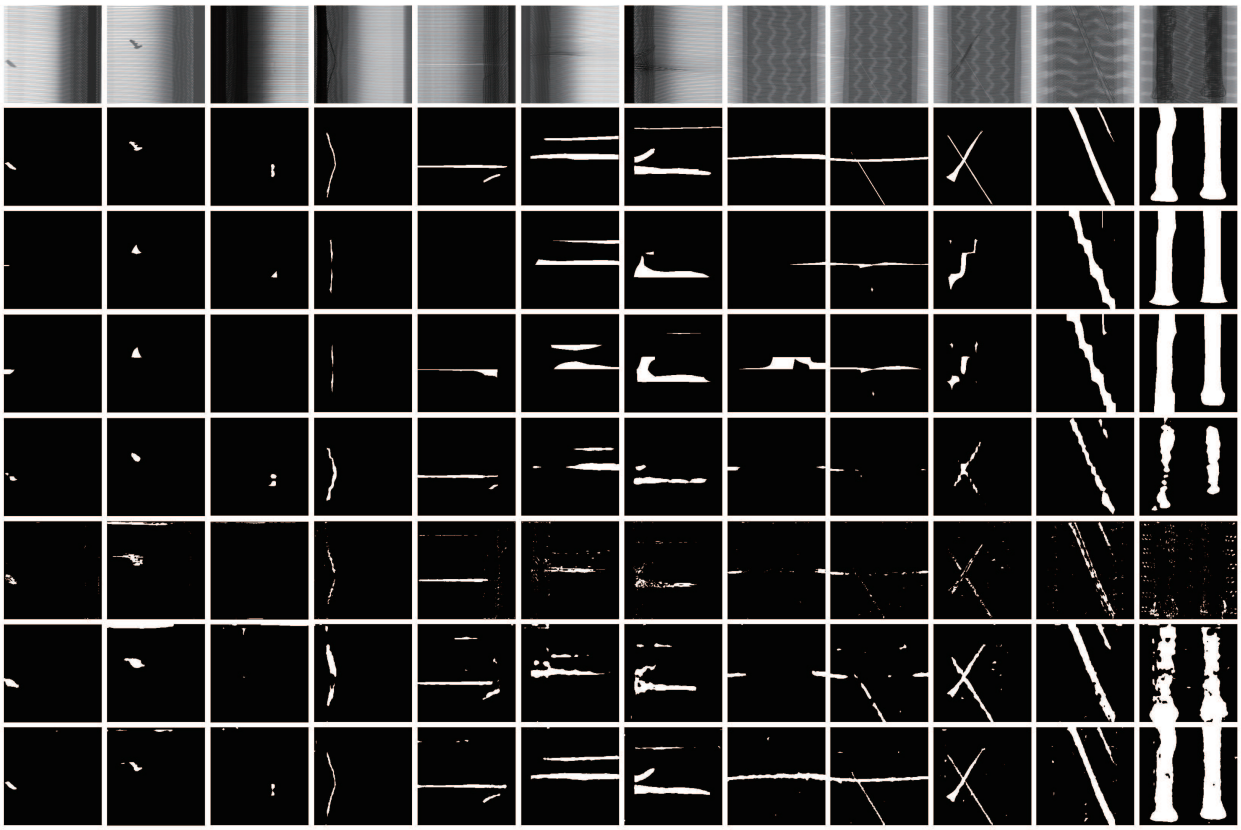

Fig. 3. Comparison of experimental results. The first to seventh columns are sidewall images with defects such as impurities, bubbles, slacks and overlaps. The last five columns are tread images with overlaps.

the semantic-aware network is adopted to mine abstract semantic features including shape and location information, which is simplified by FCN. Furthermore, in order to complement the detailed information filtered during the extraction of high-level semantics, a novel texture-aware network is developed to attempt to cover more shallow details without increasing computational complexity. On the other hand, a variety of data preprocessing methods are used to enhance performance before training the two sub-networks, owing to the characteristics of tire images, such as texture regularity in defective-free regions and diversity of various defects. In the following subsections, we describe these networks and strategies in details.

2.1. Semantic-aware sub-network

The proposed approach captures the desired high-level semantics of defects using an architectural design similar to the well-known FCN-VGG16 (a FCN with VGG16 as the basic framework), which has been proven to be viable in tire defect detection tasks [8]. At first, a stack of convolution and max pooling layers are used repeatedly to obtain the most representative information in tire images. Each fully connected layer in the VGG16 is then replaced by a special convolutional layer to retain sufficient spatial information. Finally, prediction results with the same size as input images are derived by up-sampling these spatial feature maps. We simplify an off-the-shelf FCN-VGG16 into a binary-classification and pixel-wise detection model. Specifically, both padding and stride are set to 1 and the size of convolution kernels are set to 3×3 in each convolution layer. The crop layers are removed to reduce noise and increase the detection efficiency. Although with these settings, small-sized input images cannot be processed after five pooling layers. In real-world applications, the size of the input x-ray image is fixed through the processing of x-ray devices and scaling operations.

Although the filter learned in VGG network are excellent generic visual descriptors, it is originally trained for the purpose of object classification tasks. The existence of the five max pooling layers allows it to filter out the essential details while mining global semantic information. Therefore, the FCN-VGG16 is not sensitive to the edge details detection, especially for small-sized defects.

2.2. Texture-aware sub-network

As mentioned above, with several pooling layers, especially the max-pooling layers, deep networks can extract high-level abstract features and semantic information, while detail clues are unavoidable filtered. Taking Fcn-vgg16 as an example, Figure ** indicates the impact of pooling layers on the extracted features. However, directly dropping pooling layers from existing deep networks will bring a series of training problems, such as parameter explosion and overfitting. Therefore, our shallow texture-aware network was cautiously designed with only four layers of convolutional layers, where the shrinking layer is utilized to indirectly reduce the parameter dimension. A similar structure is used in FSRCNN [24] to handling precise visual tasks such as super-resolution. On the one hand, the representation of the texture does not require the capture of high-level semantics. On the other hand, the down-sampling capability of the shrinking layer in the channel dimension can both reduce redundant information and retain the spatial resolution. As shown in **, the first convolution layer uses 5×5 convolution kernels to extract the features of the input images. Although these feature maps carrying a large amount of valuable information can be directly mapped by next layers, this leads to an increase in computational complexity. Therefore, the second convolution layer uses a 1×1 convolution kernel to reduce computational cost and redundancy information, called shrinking layer. The latter two convolution layers with 3×3 kernels are adopted to improve representation capability. Without pooling layers, detailed features are completely retained through the texture-aware network.

2.3. Combination of semantic and texture

Although features extracted by the texture-aware network contain the essential details, the detection results derived from these features usually have a large amount of noise and a high false positive rate without the guidance of global semantics. Therefore, the fusion of semantic-aware networks and texture-aware networks is necessary. Semantic and texture feature maps are concatenated and fused in the channel dimension via a *concat* layer and a convolution layer with 1×1 kernels. Considering the semantic network as a guide and texture network as a supplement, final predictions are automatically learned by training an additional convolution layer.

2.4. Pre-processing and training

In order to make the texture-aware network more robust to scale variations, input images are preprocessed through multi-scale blocking and scaling strategies. Generally, in natural image segmentation tasks, the blocking operation first obtains object proposals [25, 26], and then finely segment in these regions. In contrast, tire images can be cropped evenly while retaining defect semantic information, due to the similarity of background textures. As shown in Figure 1c, the cropping and scaling essentially change the proportion of defects in training images. In this paper, we crop 4 and 16 blocks without overlap from each image, and consider scales of 0.5 and 1.5.

Since the two sub-networks use different input data and preprocessing methods, we first train the texture-aware network using the block data, and then train the entire network with fixed texture-aware network parameters. As described in the previous subsections, owing to the characteristics of tire image textures, original images can be directly fed to obtain end-to-end results during test phases. The feasibility of this strategy are proved via next ablation experiments.

3. EXPERIMENTS

3.1. Implementation details

Dataset. Our experimental dataset consists of 914 tire images including both sidewall and tread images. These images involve various defects such as metal impurities, bubbles, and overlaps. For semantic-aware networks, we consider flipping and mirroring to enhance the data. For texture-aware networks, cropping and scaling are used, as mentioned above.

Parameter Setting. The proposed MDDN was coded with Python 3.5 in the Caffe framework. A GTX-1080 GPU and Intel Xeon-E5 3.40GHz CPU were used for both training and testing. The momentum parameter and weight decay were set to 0.99 and 0.0005 during training, respectively. Moreover, semantic-aware network was implemented on the public FCN code, and the parameters of the texture-aware network were randomly initialized.

Metrics. We adopted the widely used metrics in instance segmentation community, including intersection over union (IOU) and pixel accuracy (PA). The former can represent the accuracy of location. And the latter indicates the ratio of the correct labeled pixels to the total pixels, which reflects the pixel-level accuracy of the detection results.

3.2. Ablation experiment

In order to evaluate the parameter effectiveness of our MDDN, we conducted two groups of ablation experiments on the same data set. In one group, features learned by the semantic-aware network and the texture-aware network were used alone to detect defects. As the results shown in the figure **. The semantic-aware network can roughly detect the location and shape of the defect. The texture-aware network has a high false positive rate due to the large amount of noise in the extracted features. In another group, for the same network trained with blocks, we used the blocks and the original images for testing. The experimental results verified that tire blocks still has relatively complete semantic information.

3.3. Comparison with state-of-the-art methods

Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods. Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods. Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods. Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.Comparison with state-of-the-art methods.

4. CONCLUSION

In this paper, we proposed a MDDN model for tire defect detection tasks. Through combining a semantic-aware network and a novel texture-aware network, MDDN can obtain the essential detailed features while mining the semantic information hidden in deep layers. In addition, we experimentally verified that the blocking strategy can effectively enhance the dataset and retain detailed information, in tire images. The experiments demonstrate that our MDDN has significantly improved over the existing tire defect detecting methods, and can produce more accurate small defect detection results. The future work includes reducing noises in the texture-aware network and increasing detection speed.

5. REFERENCES

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

[2] 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, 2016.

[3] 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.

[4] Y. Zhang, X. Cui, Y. Liu, and B. Yu, "Tire defects classification using convolution architecture for fast feature embedding," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 1056–1066, 2018.

[5] Y. Xiang, C. Zhang, and Q. Guo, "A dictionary-based method for tire defect detection," in *ICIA*, 2014, pp. 519–523.

[6] 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.

[7] 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.

[8] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire defect detection using fully convolutional network," *IEEE Access*, vol. 7, pp. 43502–43510, 2019.

[9] 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.

[10] X. Cui, Y. Liu, and C. Wang, "Defect automatic detection for tire x-ray images using inverse transformation of principal component residual," in *AIPR*, 2016, pp. 1–8.

[11] X. Cui, Y. Liu, C. Wang, and H. Li, "A novel method for feature extraction and automatic recognition of tire defects using independent component analysis," *DEStech Transactions on Materials Science and Engineering*, , no. icimm, 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, 2015.

[13] Z. Yan, L. Tao, and L. Qing-Ling, "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.

[14] X. Cui, Y. Liu, Y. Zhang, and C. Wang, "Tire defects classification with multi-contrast convolutional neural networks," *IJPRAI*, vol. 32, no. 04, pp. 1850011, 2018.

[15] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.

[16] D. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.

[17] T. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 2117–2125.

[18] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. Berg, "Ssd: Single shot multibox detector," in *ECCV*, 2016, pp. 21–37.

[19] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *ECCV*, 2016, pp. 483–499.

[20] G. Ghiasi and C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *ECCV*, 2016, pp. 519–534.

[21] P. Zhou, B. Ni, C. Geng, J. Hu, and Y. Xu, "Scale-transferrable object detection," in *CVPR*, 2018, pp. 528–537.

[22] L. Boominathan, S. Kruthiventi, and R. Babu, "Crowdnet: A deep convolutional network for dense crowd counting," in *ACM international conference on Multimedia*, 2016, pp. 640–644.

[23] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Sodmtgan: Small object detection via multi-task generative adversarial network," in *ECCV*, 2018, pp. 206–221.

[24] C. Dong, C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *ECCV*, 2016, pp. 391–407.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *CVPR*, 2014, pp. 580–587.

[26] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *NIPS*, 2015, pp. 91–99.