

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

Ren Wang^{1,2}, Qiang Guo^{1,2}, Caiming Zhang³

¹School of Computer Science and Technology,
Shandong University of Finance and Economics, Jinan, China

²Shandong Provincial Key Laboratory of Digital Media Technology, Jinan, China

³Software 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 processes 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 multi-scale defect 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.

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, 3]. Especially in tire manufacturing, numerous detection algorithms have been proposed [4, 5, 6] 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 [7]. 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 study [8, 9], 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.

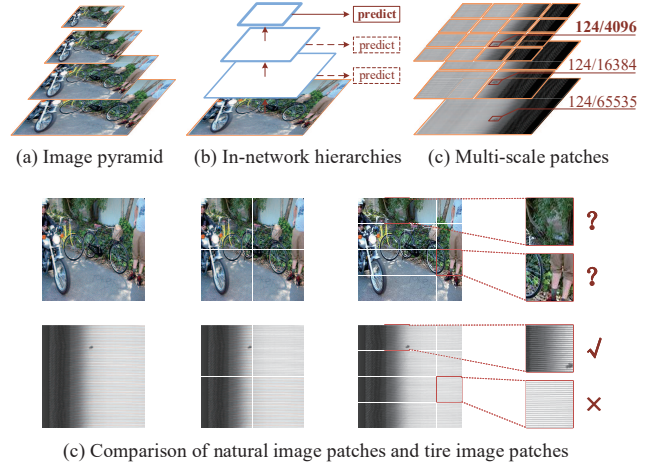


Fig. 1. (a) shows the image pyramids. (b) indicates in-network feature hierarchies. Prediction results can be derived from each layer. (c) represents multi-size tire image patches, which can increase the relative scale of small defects and backgrounds. (d) illustrates that the semantic information is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

At present, existing computer vision based detection methods are mostly devoted to distinguishing difference between defective regions and background (defective-free regions). A key issue for such methods is feature extraction. Guo *et al.* [10] 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 [11, 12]. 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.* [8, 13] introduced multi-scale wavelet and curvelet transform in detection tasks respectively. Furthermore, optimized edge detection and total variation algorithm are used to achieve more accurate results [14]. However, the repre-

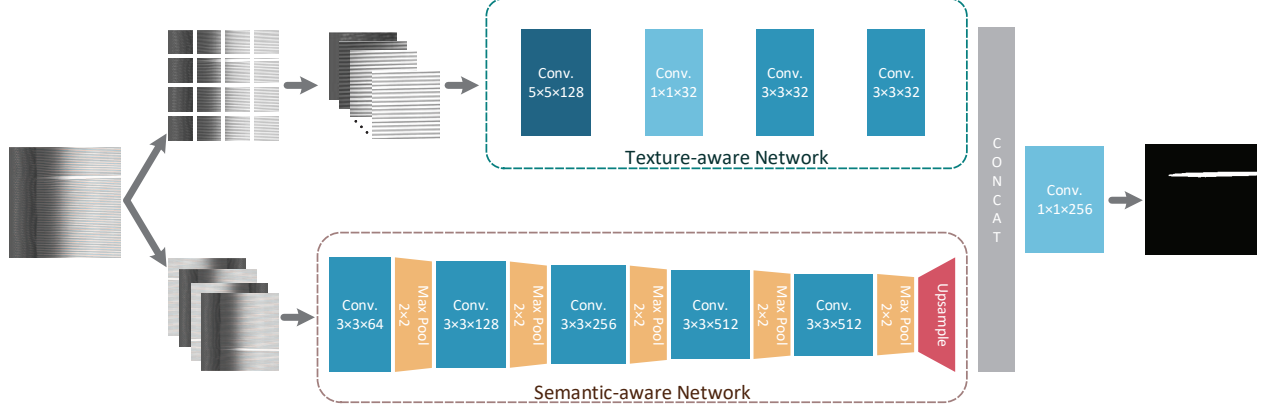


Fig. 2. Overall architecture of the proposed MDDN.

sensation capability of fixed kernels is not comprehensive enough. Moreover, transform processes are computationally expensive. Recently, Cui *et al.* [15] attempted to classify tire defects by means of deep convolutional networks (ConvNets), which has outstanding performance in the recognition and segmentation tasks of natural images. With the excellent feature extraction capability of the ConvNet, Wang *et al.* [9] further implemented the detection and segmentation in tire images by a fully convolutional network (FCN) [16]. 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 overcome these shortcomings, many methods have been proposed in benchmark datasets. Most of them are based on multi-scale strategies and can be roughly classified into image pyramids and in-network feature hierarchies. Image pyramids were directly scaled to get multi-scale images and extensively used in the era of hand-crafted features [17, 18], as shown in Figure 1(a). With the popularity of self-learning feature representation, image pyramid based methods is 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 of different spatial resolutions, with an multi-scale and pyramid shape[19]. The Single Shot Detector (SSD) [20] 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 by the red line in Figure 1b[. For example, FCN defined a skip architecture to produce more accurate segmentation. Similar top-down skip

architectures are popular in recent research[21, 22]. 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[23]. CrowdNet [24] combines shallow and deep features derived from two different structures to overcome this shortcoming. The capability to extract detail features is further enhanced by a parallel shallow subnetwork.

Inspired by CrowdNet, we construct a end-to-end network named Multi-scale Defect Detection Network (MDDN) consisting of a semantic-aware sub-network and a texture-aware sub-network. To detect small tire defects, image patches [24, 25] are fed in the texture-aware sub-network. Unlike natural image patches, defects (objects) are still significant and discernible in the tire image patches, illustrated in Figure 1(d). In addition, as shown in Figure 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 deep network aims to improve the accuracy of tire defect detection, especially for edge detail 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. Among them, the semantic-aware network is adopted to mine abstract semantic features including shape and location information, which is simplified by FCN [16], a deep network that has outstanding performance in solving the semantic segmentation problems with natural images. Furthermore, in order to supplement the detailed information filtered during the extraction of high-level semantics, a novel texture-aware network is developed to attempt to cover more shallow detail information without increasing computational complexity. On the other hand, owing to the

characteristics of the tire images, a variety of data preprocessing methods are used to enhance performance for the two sub-networks. In the following subsections, we describe these networks and strategies in detail.

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. At first, a stack of convolution layers 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. Moreover, we normalize the parameters of convolutional layers, such as both padding and stride are set to 1 and the convolution kernel size is set to 3×3 , and remove the crop layers to reduce noise and increase the detection efficiency.

Although the learned filter 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 having the capability to mine global semantic information. Therefore, the FCN is not sensitive to the edge details detection, especially for small-sized defects.

2.2. Texture-aware sub-network

As mentioned above, to complement the details filtered by pooling layers, a novel texture-aware network, in parallel with the semantic-aware network, is developed to obtain detailed clues in shallow layers. The texture-aware network employs 4 convolutional layers inspired by FSRCNN[26], a network that is expert in handling precise visual tasks such as super-resolution. The first layer of convolution uses a 5×5 convolution kernel to extract the features of the input image. These features are directly used by later layers to increase computational complexity. 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 convolutional layer uses a 1×1 convolution kernel to reduce computational cost and redundancy information, called shrinking. The latter two 3×3 convolutions are adopted to improve representation capability. Without the existence of pooling layers, detailed features can be completely retained. Meanwhile, the down-sampling capability of the shrinking layer in the channel dimension can both reduce redundant information and retain the spatial resolution.

In order to make the texture-aware network more robust to scale variations, images are preprocessed through multi-scale blocking and scaling strategies. Generally, in natural image segmentation tasks, the blocking operation first obtains object proposals[], and then finely segmentation 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.

2.3. Combination and training

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 combined and fused in the channel dimension via a *concat* layer and a 1×1 convolution layer.

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 training texture-aware network parameters. As described in the previous subsections, owing to the characteristics of tire image textures, original images can be directly input during the test phase. And the end-to-end test results can be obtained. The experiment proves the effectiveness of this means.

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 are 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 is implemented on the public FCN code, and the parameters of the texture-aware network are randomly initialized.

Metrics. We adopt 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

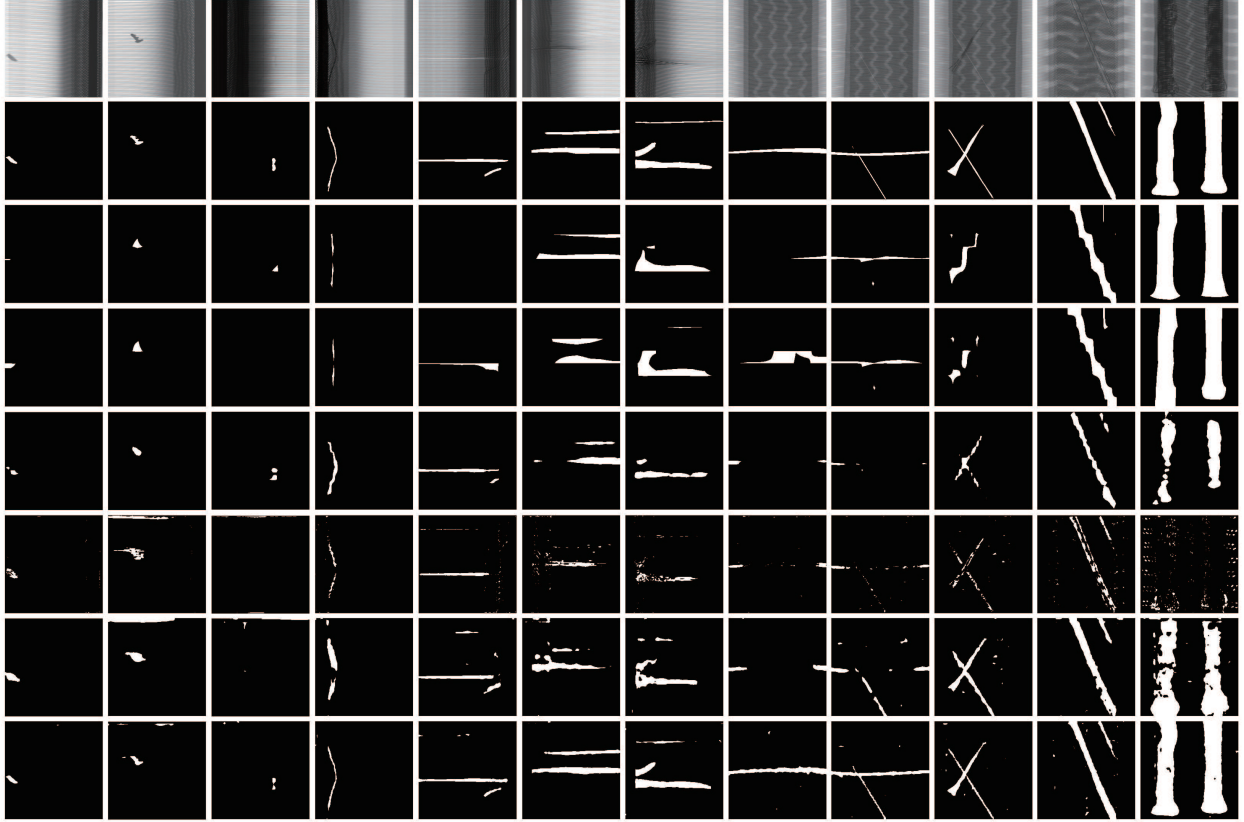


Fig. 3. Comparison of experimental results.

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 proposed MDDN method, we conduct 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 are 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 use the blocks and the original images for testing. The experimental results verify 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.

4. CONCLUSION

In this paper, we proposed a MDDN model for detailed tire defect detection tasks. Through combining a semantic-aware network and a novel texture-aware network, MDDN can preserve the necessary detail features while mining the semantic information hidden in deep layers. We showed that an increase in the proportion of defective areas is critical for small defect detection. In addition, we have 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] 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, 2012.
- [4] 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.
- [5] 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.
- [6] C.D. Jones, A.B. Smith, and E.F. Roberts, "A dictionary-based method for tire defect detection," in *ICIA*. IEEE, 2014, pp. 519–523.
- [7] 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, 2016.
- [8] 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.
- [9] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire defect detection using fully convolutional network," *IEEE Access*, vol. 7, pp. 43502–43510, 2019.
- [10] Q. Guo and Z. W. 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] X. Cui, Y. Liu, and C. Wang, "Defect automatic detection for tire x-ray images using inverse transformation of principal component residual," in *AIPR*. IEEE, 2016, pp. 1–8.
- [12] X. H. Cui, Y. Liu, C. X. 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.
- [13] 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.
- [14] 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.
- [15] 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. 04, pp. 1850011, 2018.
- [16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.
- [17] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *IJCV*, vol. 60, pp. 91–110, 2004.
- [18] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005.
- [19] T. Y. Lin, P. Dollr, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *CVPR*, 2017, pp. 2117–2125.
- [20] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *ECCV*. Springer, 2016, pp. 21–37.
- [21] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *ECCV*. Springer, 2016, pp. 483–499.
- [22] G. Ghiasi and C. C. Fowlkes, "Laplacian pyramid reconstruction and refinement for semantic segmentation," in *ECCV*. Springer, 2016, pp. 519–534.
- [23] P. Zhou, B. Ni, C. Geng, J. Hu, and Y. Xu, "Scale-transferrable object detection," in *CVPR*, 2018, pp. 528–537.
- [24] L. Boominathan, S. S. Kruthiventi, and R. V. Babu, "Crowdnet: A deep convolutional network for dense crowd counting," in *ACM international conference on Multimedia*. ACM, 2016, pp. 640–644.
- [25] 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.
- [26] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *EC-CV*. Springer, 2016, pp. 391–407.