

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/356090767>

HLU2-Net: A Residual U-structure Embedded U-Net with Hybrid Loss for Tire Defect Inspection

Article in IEEE Transactions on Instrumentation and Measurement · February 2023

DOI: 10.1109/TIM.2021.3126847

CITATIONS

45

READS

387

5 authors, including:



Zhouzhou Zheng

Northwest A & F University

23 PUBLICATIONS 220 CITATIONS

[SEE PROFILE](#)



Bin Yu

Qingdao University of Science and Technology

83 PUBLICATIONS 2,267 CITATIONS

[SEE PROFILE](#)



Yan Zhang

Qingdao University of Science and Technology

50 PUBLICATIONS 770 CITATIONS

[SEE PROFILE](#)

HLU²-Net: A Residual U-Structure Embedded U-Net With Hybrid Loss for Tire Defect Inspection

Zhouzhou Zheng[✉], Huanbo Yang, Liang Zhou[✉], Bin Yu[✉], and Yan Zhang[✉], Member, IEEE

Abstract—Intelligent defect detection have been widely studied and applied in many industrial fields. However, intelligent tire defect inspection remains a challenging task due to tire radiographic images' anisotropic multi-texture background in which a variety of defects may appear with intra class dissimilarity and inter class similarity. This article addresses the problem intelligent tire defect detection using end-to-end saliency detection network. A novel end-to-end residual U-structure embedded U-Net with hybrid loss function and coordinate attention module (HLU²-Net) is proposed. In HLU²-Net, the novel residual U-structure is used to replace encode-decode block of U-Net for fusing multiscale and multilevel features, and a hybrid loss is presented to guide defect detection for complete and clean defect mask. Moreover, a coordinate attention module is introduced to highlight useful features and weaken irrelevant features. Comparative experimental results verify that our method outperforms the state-of-the-art methods on our dataset according to six evaluation metrics. Additionally, we demonstrate that the computing efficiency of our method can meet online visual detection on tire production line.

Index Terms—Defect detection, HLU²-Net, hybrid loss, residual U-structure, saliency object detection (SOD).

I. INTRODUCTION

TIRES play an important role in cushioning and shock absorption during driving. According to the statistics, 40% traffic accidents are caused by tire failure [1], consequently, tires are essential to the safety of the driver. According to the statistics of International Organization of Motor Vehicle Manufacturers, sales of new passenger car vehicles of 2019–2020 reached 53 598 846 which means that at least 100 million passenger car tires are produced and assembled

Manuscript received August 17, 2021; revised October 14, 2021; accepted October 26, 2021. Date of publication November 9, 2021; date of current version November 23, 2021. This work was supported by the Natural Science Foundation of Shandong Province under Grant ZR2019MEE066. The Associate Editor coordinating the review process was Eduardo Cabal-Yepes. (*Corresponding author: Yan Zhang.*)

Zhouzhou Zheng is with the College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao 266061 China, and also with the College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China (e-mail: zhengzz@163.com).

Huanbo Yang is with the College of Mechanical and Electronic Engineering, Northwest A&F University, Yangling 712100, China (e-mail: myheartises@163.com).

Liang Zhou is with the College of Mechanical Engineering and Automation, Northeastern University, Shenyang 110819, China (e-mail: zhouliang514@126.com).

Bin Yu is with the College of Mathematics, Qingdao University of Science and Technology, Qingdao 266061, China (e-mail: yubin@qust.edu.cn).

Yan Zhang is with the College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao 266061, China (e-mail: zy@qust.edu.cn).

Digital Object Identifier 10.1109/TIM.2021.3126847

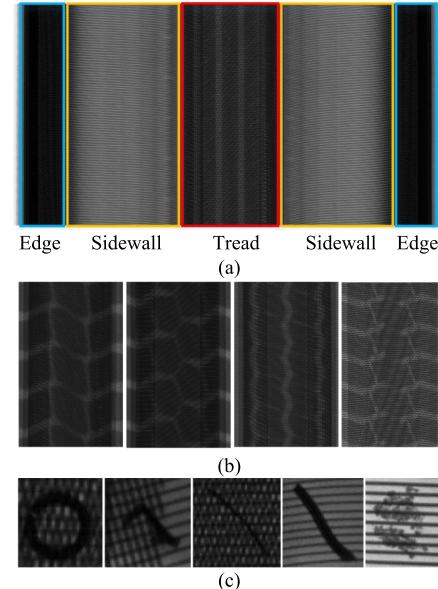


Fig. 1. Challenges faced by intelligent tire defect detection. (a) Tire X-ray image, (b) tread patterns, and (c) defect samples.

every year. However, naked-eye tire quality inspection is still widely used by tire industries with the aid of radiographic imaging system. Artificial inspection is of low efficiency and strong subjectivity which will lead to inaccurate quality assessment [2]. Particularly, due to the tread patterns and materials including steel wire, nylon cloth, and a large variety of rubbers, tire radiographic image has complicated anisotropic multi-texture low-contrast background in which a variety of defects may occur, as shown in Fig. 1. Intra-class feature variousness and inter-class feature similarity in tire defect feature space is another challenge for automated tire defect inspection [2].

In recent years, with the development of intelligent manufacturing and industrial demand, automatic visual inspection technology is becoming a research hotspot. Therefore, it is necessary to develop an automatic tire defect inspection method based on the tire radiographic images. Generally speaking, tire defect inspection includes two fundamental computer vision problems, namely, object detection and object classification. A series of work about tire defect inspection is done, which can be divided into traditional methods [2]–[10] and deep learning methods [11]–[19].

Traditional machine vision based tire defect inspection methods have been presented include statistic-based [3], [9], [10], frequency domain-based [2], [5]–[7] and model-based methods [4], [8]. A series of research is

conducted by Zhang *et al.* [2]–[4]. A Canny edge detection algorithm combined with curve transformation was utilized to improve the completeness of tire defect boundary [3]. Subsequently, wavelet-based defect detection method was proposed [2] in which multiscale local modular maximum edge scheme was adopted to reduce false alarm. They also presented a total variation model-based scheme to decompose tire image into texture and cartoon components to reduce the complexity of problem space [4]. Similarly, an image reconstruction algorithm based on inverse principal component residual transformation is used to reconstruct the tire image which strengthen defective regions and weaken the background [5]. Sparse representation was applied to tire defect detection [6] by representing global and local features by which location of defect can be detected precisely. Zheng and Pang [7] proposed an extremum filter and adaptive threshold segmentation method for tire defect detection which divides tire image into cords and backgrounds, and then applies adaptive threshold segmentation to separate defects and backgrounds. Xiang *et al.* [8] presented a tire defect detection method in which a dictionary learning model was established. The coefficient obtained by multiplying the pseudo-inverse matrix and the image block was used to determine if it was defective, based on which a detection method using similarity of tire background was proposed [9]. Moreover, Zhang *et al.* [10] presented a maximum class spacing method and achieved satisfactory results in tire foreign matter defect detection. Although plenty of traditional machine vision-based methods can be utilized to detect defect to a certain extent, tire defect detection remains an unsolved problem due to their limitations and challenges of the issue: 1) statistical-based methods have difficulty in handling low-contrast multi-texture backgrounds; 2) frequency domains-based methods are sensitive to parameters and can hardly adapt to detect different defects; and 3) model-based methods have large amount of calculation and rely too much on post-processing.

In the past decade, deep learning techniques have achieved great success in tire defect detection [11]–[19] and many other research fields [20]–[24]. Researchers have utilized deep learning for various defect inspection applications including steel defect detection [25], fabric defect detection [26], electronic device defect inspection [27], and so on. Generally, the predominant deep learning-based methods can be divided into three categories in tire defect detection fields, which are image-level classification [11]–[13], region of interest (ROI) detection [14]–[16], and pixel-wise segmentation-based methods [17]–[19]. Cui *et al.* [11] combined five convolutional neural networks to classify six different tire defects and achieved a classification accuracy of 98.47%. Tada *et al.* [12] presented a tire inner surface image functional classifier which combined a convolutional neural network and SVM. In [13], to reduce feature redundancy of CNN, a spare-coding block was designed to replace convolution kernel in CNN, by which classification accuracy was improved and detection time was greatly reduced. These image-level classification methods can only be used to classify defect types which cannot be used to perform visual inspection and defect detection tasks.

With the emergence of object detection network, defect detection deep networks were proposed. In [14], a lightweight ZFNet was chose as backbone network of Faster R-CNN which can detect tire foreign matter and bubble defects. Chang *et al.* [15] proposed an integrated hybrid network combining ConvNets and Faster R-CNN which achieved good detection results in detecting tire bubble defects. An end-to-end network named TireNet [16] was proposed in which Siamese was used as a part of the classifier to capture tire defect features. Compared with the state-of-art object detection networks such as Faster R-CNN, SSD, and YOLO, TireNet reached a missing rate of 0.17%. However, object detection network-based methods can only provide the location of defect without exact shape and size information.

To further address this problem, Wang *et al.* [17] introduced a fully convolutional neural network for tire defect segmentation for six kinds of defects. In [18], a concise semantic segmentation network combining CNN classification network was proposed for tire defect segmentation and classification, which achieved 85.13% in mPA and 77.34% in mIoU. Zheng *et al.* [19] presented a two-stage network combining YOLOv3 and an improved PSPNet with pyramid feature fusion and convolutional block attention module (CBAM) which further achieved an AP of 91.39 and mIoU 87.86%, respectively. The aforementioned two-stage methods cannot realize end-to-end defect detection and segmentation but train the separate two-stage networks, which will cause cumulative errors.

The aforementioned deep learning-based inspection methods received relatively good results compared to traditional machine learning-based ones. However, there are still challenges in the field of intelligent defect inspection that cannot be ignored. Specifically, the aforementioned detection methods are limited by the size of the receptive field of the convolution layer, the low-level features lack global information, and the high-level features cannot accurately predict the details such as the target edge due to the low resolution of multiple pooling operations. On the other hand, attention mechanism is not applied to improve network's information processing ability, inspection speed, and effect with less parameters. Obtaining these feature maps from high-resolution images requires a lot of computation. To densely cover many different spatial positions, scales, and aspect ratios, evaluation is required for exhaustive classification. Moreover, the cross-entropy loss function is insensitive to the object boundary and thus has difficulty in obtaining clear boundary.

To address the above challenges, we propose an end-to-end defect saliency detection network with application to tire defect inspection. We present a residual U-structure-embedded encode-decode block of U-Net with a hybrid loss function and a coordinate attention module. As shown in Fig. 2, first, test samples are fed into the first residual U structure on the encode stage (En_1) to obtain defect boundary information. Second, after the feature extraction and enhancement of residual U structure and coordinate attention module on encode stage, four decode blocks are utilized to unsampled feature and output prediction mask. Furthermore, the hybrid loss function is used to narrow the gap between the prediction mask and the

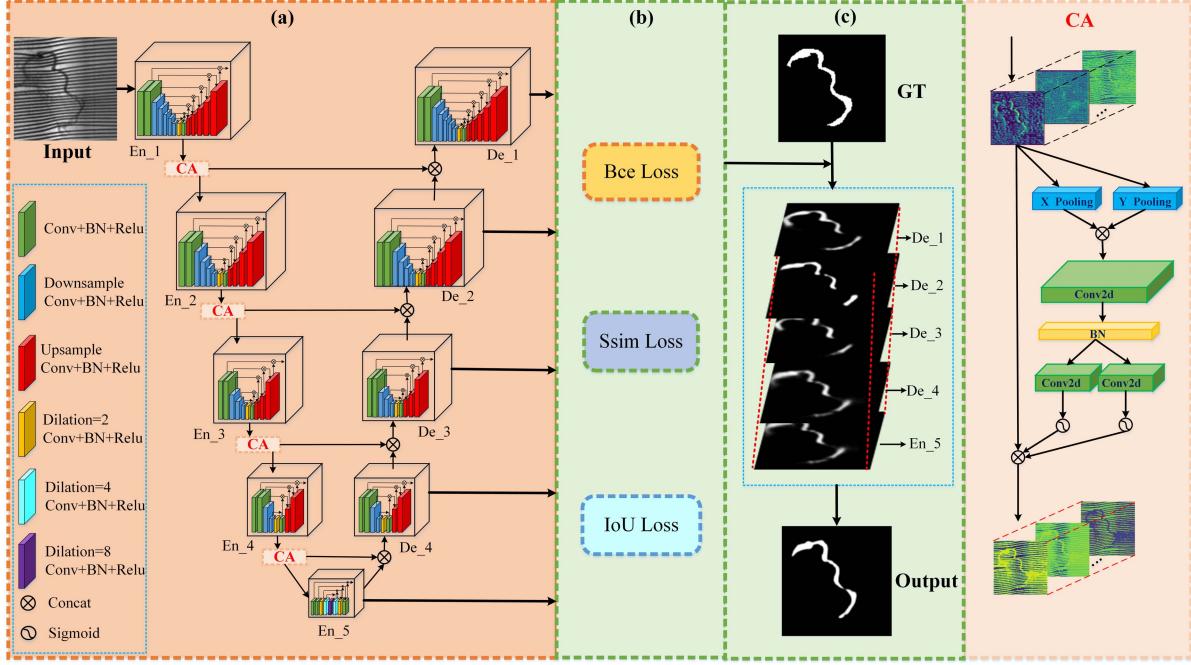


Fig. 2. Architecture of our proposed HLU²-Net. (a) Network architecture, (b) train loss, and (c) mask generation.

label. Finally, tire defect detection result is obtained by fusing five refined prediction masks.

In summary, our contributions are concluded as follows.

- 1) A residual U structure is introduced and embedded to encode-decode block of U-Net to enlarge receptive field and fuse features of different scales.
- 2) We apply a hybrid loss to guide network training and enhance joint decision-making for tire defect detection.
- 3) A coordinate attention is introduced to highlight useful features and weaken irrelevant features.
- 4) We established a tire defect detection dataset covering 583 defects with pixel-level accurate labels.

The rest of this article is organized as follows. We introduce related work on salient object detection in Section II. Section III describes our proposed HLU²-Net in detail. Section IV discusses comparative experimental results. Finally, Section V concludes this article.

II. RELATED WORK

Human vision system focuses only on the information-rich area in the scene to deepen the understanding of the area. Inspired by this mechanism, researchers have been paying close attention to salient object detection (SOD). SOD aims to strengthen ROI and weaken the background and has achieved great success in various tasks. Typically, there are two types of SOD methods: traditional and deep learning-based methods.

A. Traditional Methods

Following cognitive theories and early attention models, e.g., Koch and Ullman [28] and Itti *et al.* [29], hundreds of computational saliency models have been proposed [30] which laid the foundation for SOD. In traditional SOD methods, feature gradient changes including color, brightness feature,

and so on between object and background are utilized to get saliency map. Specifically, Liu *et al.* [31] formulated this problem as a binary labeling task in which features including multiscale contrast, center-surround histogram, and color spatial distribution are used to describe a salient object locally, regionally, and globally. Achanta *et al.* [32] utilized low-level color feature and brightness feature to detect saliency area in complex scene images. He further designed frequency tuned approach based on color feature for SOD [33]. However, these methods have difficulties in detecting saliency object from images with low contrast between the object and the background. To solve the issue, researchers begin to consider starting from the background to achieve SOD. Subsequently, a series of studies were carried out, for example, Jiang *et al.* [34] introduced the regional background vector to describe the background of the saliency map, and regarded it as the counterpart of the object descriptor to detect object with low contrast. A background measurement algorithm for regional connectivity was proposed by Zhu *et al.* [35] to locate the object's location by searching for inconsistent areas of the background. Additionally, some classic approaches based on region comparison (RC) [36] and context-aware algorithm (CA) [37] and so on have been applied to saliency detection.

These models have been successfully applied in many applications such as object detection and recognition and so on [38]. However, although traditional SOD methods have achieved satisfactory results on some typical datasets, they still face great challenges in solving the tire defect saliency detection problem because of its low-contrast, complex anisotropic multi-texture, gray-level background.

B. Deep-Learning-Based Detection Methods

Inspired by the performance of deep learning methods in vision tasks like image classification and object detection [39],

deep convolutional neural networks (CNNs) are also studied in recent works [40]–[49] and have demonstrated the effectiveness of data-driven feature integration and parameter self-optimization capabilities. Many state-of-the-art methods were proposed including multilevel feature-aggregation-based methods [40]–[42], edge-enhancement-based methods [43]–[46], and external-knowledge-base-based methods [47]–[49].

1) *Multilevel Feature-Aggregation-Based Methods*: Multilevel feature aggregation saliency networks, e.g., Amulet [40], etc., integrate feature maps of different horizontal resolutions to maximizing the use of features and object details in the prediction stage. Pang *et al.* [41] designed an aggregation interaction module to integrate the features from adjacent levels. A feature aggregation module (FAM) [42] was proposed to make full use of feature and thus to improve detection accuracy by fusing high-level semantic feature and low-level boundary feature. However, multilevel feature-aggregation-based methods improve detection accuracy at a cost of increasing the amount of network parameters and reducing the detection efficiency. This will affect the real-time performance of the networks.

2) *Edge-Enhancement-Based Methods*: To solve the problem of blurred edges and glitches of predicted mask, an edge guidance network (EGNet) [43] was designed to get salient edge feature by integrating boundary information and global object information. Wu *et al.* [44] performed object boundary repair by stacking cross refinement unit (CRU) to make use of target features and edge features of different levels, which achieved good detection results. Song *et al.* [45] introduced a refinement network to the encode-decode network in which the channels weighted block and the residual decoder block were adopted alternatively to integrate the spatial features. Su *et al.* [46] proposed a novel boundary-aware network in which the feature selectivity at boundaries is enhanced by incorporating a boundary localization stream. These edge-enhancement-based methods increase the models' understanding of the edge through post-processing or introducing external knowledge base of the object edge to improve segmentation quality. However, edge post-processing makes it impossible for the model to be trained in an end-to-end manner, and the external edge knowledge base requires an additional edge dataset which intensifies the workload.

3) *Attention-Mechanism-Based Methods*: To improve network accuracy, researchers proposed some complex deep networks, which result in parameter growth and thus affect the detection efficiency. A plug-and-play attention module has attracted widespread attention from researchers. This module, while guaranteeing the prominent features, retrenches the amount of parameters. For example, Zhao *et al.* [47] proposed a pyramid feature attention network in which context-aware pyramid feature extraction (CPFE) module was designed for multiscale and high-level feature mapping to obtain rich contextual features. Liu *et al.* [48] proposed a pixel-wise contextual attention network (PiCANet) for saliency detection by generating an attention weight map. In [49], a recurrent attention module was introduced to generate more accurate saliency features for SOD. The above work has improved the accuracy of SOD to a certain extent and indicated the

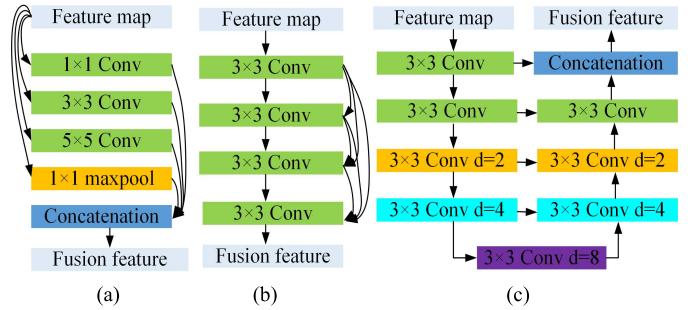


Fig. 3. Frameworks of typical and the proposed feature fusion methods. (a) Multiscale feature fusion, (b) multilevel feature fusion, and (c) our proposed residual U-structure feature fusion(En_5).

promising direction in the future. There is still great potential in reducing network parameters in the future.

III. HLU²-NET

A. Overview of Network Architecture

The U-Net encoder-decoder framework has achieved great success in semantic segmentation and other fields. As shown in Fig. 2, based on this architecture we propose a new HLU²-Net for tire defect inspection by incorporating residual U structure for multiscale and multilevel feature fusing, hybrid loss function and coordinate attention for SOD. Our framework consists of five encoder blocks and four decoder blocks. A hybrid loss was utilized to narrow the gap between predicted mask and ground truth, and accelerate model convergence.

To overcome the above issue, inspired by Qin *et al.* [53], a U-shape feature connection method called residual U-structure is used to replace dense connections to simplify the complexity of feature connections. The structure can reduce parameters and improve detection efficiency.

B. Residual U-Structure

For SOD, it is important to make the most efficient use of multiple features of different scales and levels. Typically, researchers focus on aggregating multiscale and multilevel features by a series of feature fusion strategies, such as pyramid feature fusion [19], residual connection [50], inception module [51], dense connection [52], etc., as shown in Fig. 3(a) and (b); these techniques can be divided into two categories, namely multiscale-based feature fusion and multilevel-based feature fusion. Among them, multilevel-based feature fusion methods achieve effective utilization of low-level and high-level features by dense connection. However, they also greatly increase the amount of parameters and thus affect detection efficiency. Multiscale-based feature fusion methods enlarge receptive field with convolution kernels of different sizes. Similarly, this will also improve model parameters when using 5×5 convolution kernel and affect the real-time performance.

Moreover, we integrate multiscale and multilevel features by nesting the residual U-structure in which unequal dilation rate and U connection are used to enlarge receptive field and fuse different features, as shown in Fig. 3(c).

C. Hybrid Loss Function

SOD is binary classification task in which cross-entropy loss function is used to measure error between the ground truth and the predicted mask. Researchers have proposed a series of SOD methods [54] using cross-entropy loss function. However, cross-entropy loss function is insensitive to the object boundary and thus will cause blurring boundaries. This issue has been raised and studied and recently some solid progress has been reported [45], [55]. In SOD tasks, especially defect detection, the boundary information plays an important role in industrial quality inspection. In SOD tasks, especially defect detection, the boundary information plays an important role in industrial quality inspection. In this sense, inspired by [45], [55], a hybrid loss function is built in this work for a robust defect detection model.

The hybrid loss function (Hy_{loss}) consists of three parts including binary cross entropy loss function (Bce_{loss}) [56], structural similarity loss function ($\text{Ssim}_{\text{loss}}$) [57], and boundary intersection over union loss function (IoU_{loss}) [58], respectively.

Among them, Bce_{loss} is widely used in pixel-level binary classification, semantic segmentation, and SOD. It can be defined as

$$\begin{aligned} \text{Bce}_{\text{loss}} = -\frac{1}{H * W} \sum_x^H \sum_y^W & [G(x, y) \log(S(x, y)) \\ & + (1 - G(x, y)) \log(1 - S(x, y))] \end{aligned} \quad (1)$$

where $G(x, y)$ and $S(x, y)$ are ground truth and predicted salient object, respectively. $G(x, y) \in \{0, 1\}$ and $S(x, y) \in [0, 1]$ denote the probabilities of being predicted as salient object.

$\text{Ssim}_{\text{loss}}$ is a structural loss function, which reflects mask similarity between ground truth and salient object. It is integrated into loss function to reduce mask structural error. Assume that x and y are the predicted saliency map and ground truth mask, respectively. Let $x = \{\chi_j : j = 1, \dots, N^2\}$ and $y = \{y_j : j = 1, \dots, N^2\}$, the SSIM of x and y can be defined as

$$\text{Ssim}_{\text{loss}} = 1 - \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (2)$$

To acquire clear boundary and more precise masks, IoU_{loss} is introduced to guide pixel-level classification. It is defined as

$$\begin{aligned} \text{IoU}_{\text{loss}} &= 1 - \frac{\sum_{x=1}^H \sum_{y=1}^W S(x, y) G(x, y)}{\sum_{x=1}^H \sum_{y=1}^W [S(x, y) + G(x, y) - S(x, y) G(x, y)]} \end{aligned} \quad (3)$$

where $G(x, y)$ and $S(x, y)$ are ground truth and predicted salient object, respectively. $G(x, y) \in \{0, 1\}$ and $S(x, y) \in [0, 1]$ denote the probabilities of being predicted as salient object.

To improve the accuracy of pixel-level classification of salient object, we consider their respective advantages in supervising network training. Bce_{loss} can be used in most

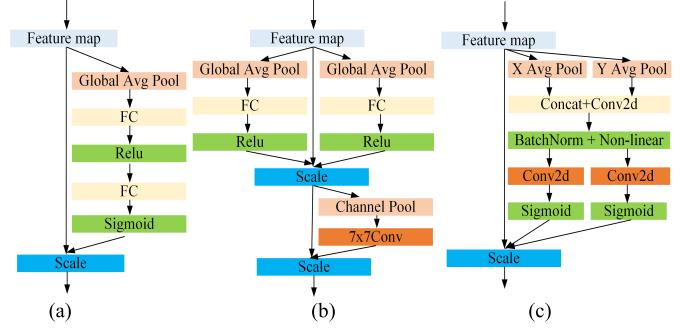


Fig. 4. Different attention mechanism blocks. (a) SE block, (b) CBAM block, and (c) CA block.

image segmentation tasks. However, when the number of current scene pixels is far less than that of background pixels, that is, the number of $G(x, y) \in 0$ is much larger than that of $G(x, y) \in 1$, the component of $G(x, y) \in 0$ will dominate, making the model seriously biased toward the background, thus resulting in poor performance. $\text{Ssim}_{\text{loss}}$ is used to measure the structural similarity between two images, and IoU_{loss} is used to measure the intersection between the target and the label. To alleviate the shortcomings of Bce_{loss} and narrow the gap between the ground truth of the predicted mask, we introduce Hy_{loss} as the loss function of this article.

To be specific, Hy_{loss} is defined as

$$\text{Hy}_{\text{loss}} = \sum_{k=1}^K (\text{Bce}_{\text{loss}}^k + \text{Ssim}_{\text{loss}}^k + \text{IoU}_{\text{loss}}^k) \quad (4)$$

where k is the number of output salient object map. As shown in Fig. 2, the output includes of DE_1, DE_2, DE_3, DE_4, and En_5, hence, in this work we set $k = 5$.

D. Attention Machine Module

Normally, traditional convolution neural networks treat features of each channel equally; therefore, they cannot make use of more meaningful features effectively. To address this issue, researchers have carried out a lot of research on attention mechanism [59].

As show in Fig. 4(a), in SENet, channel attention mechanism was proposed to focus on useful channel feature by generating channel weight. However, it pays close attention to the importance of channel feature while ignores spatial feature. To further solve the problem, as shown in Fig. 4(b), a convolution block attention module (CBAM) [60] including channel and spatial attention was proposed to focus on spatial information. However, this method increases the amount of parameters because of cascading channels and spatial attention mechanisms. Inspired by [61], we introduce coordinate attention (CA) to capture positional information and channel-wise relationships to further enhance feature representation ability. As shown in Fig. 4(c), CA block includes coordinate information embedding and coordinate attention generation. First, the $(H, 1)$ and $(1, W)$ pooling kernels are used to encode the channel along the horizontal and vertical coordinates,

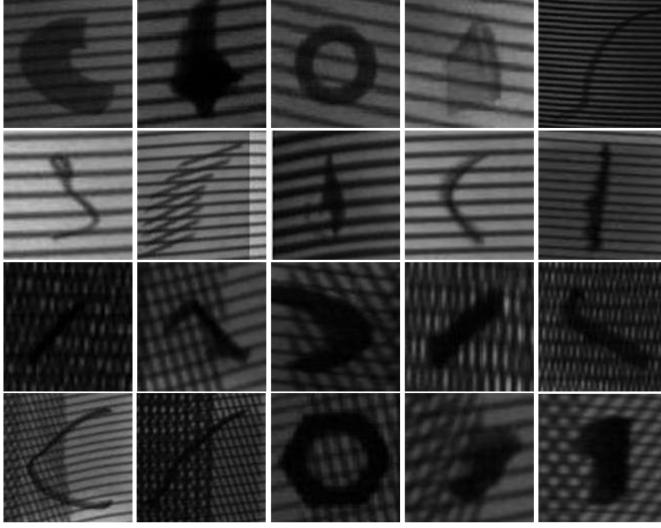


Fig. 5. Defective samples of our dataset.

respectively. The c th channel can be expressed as

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} \chi_c(h, i) \quad (5)$$

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} \chi_c(j, w) \quad (6)$$

where w , h , and c are width, height, and channel of feature map, respectively.

We concatenate $z_c^h(h)$ and $z_c^w(w)$, and the 1×1 convolution function F_1 is used to transform them.

$$f = \delta(F_1(\text{Concat}[\mathbf{z}^h, \mathbf{z}^w])) \quad (7)$$

where δ is nonlinear activation function, f is an intermediate feature map that encodes spatial information in the horizontal and vertical direction. The new feature map $y(i, j)$ obtained by CA can be expressed as

$$y(i, j) = x(i, j) \times \sigma(\text{Conv}(f^w))(i) \times \sigma(\text{Conv}(f^h))(j) \quad (8)$$

where $x(i, j)$ is original feature map.

IV. EXPERIMENTAL RESULTS

A. Datasets and Implementation Details

Our dataset is built on tire radiographic images obtained from real-life production lines. As shown in Fig. 5, 583 defective images are used in our work and the corresponding pixel-wise high-quality SOD labels are made using the Labelme [71] tool which can label images in the form of lines and points. Specifically, the pixels of defective and defect-free regions are marked as 1 and 0, respectively. To avoid the influence of subjective factors, our labels are independently marked by three professional quality inspector, and then made a decision by voting strategy. The dataset is randomly divided into 80% training set and 20% test set. Adam optimizer and proposed hybrid loss are used to narrow and reflect error between ground truth and predicted masks. We set batch size and learning rate as 4 and 0.001 to accelerate network

training. All the experiments are conducted on Windows 10 platform with single GTX 1080 GPU (8GB) and Intel core i7-8700 3.2GHz CPU. Traditional methods are conducted on MATLAB 2019b and deep learning methods are performed using Pytorch 1.2.

B. Evaluation Metrics

In our work, six evaluation metrics are used to evaluate the effectiveness of our model including precision-recall (PR) curves, F-measure, mean absolute error (MAE) [62], weight F-measure (WF) [63], structure measure (SM) [64], and enhanced-alignment measure (EM) [65], respectively.

1) *F-Measure*: F_β is used to comprehensively evaluate precision and recall, it is defined as

$$F_\beta = \frac{(1 + \beta^2) \times \text{Precision} \times \text{Recall}}{\beta^2 \times \text{Precision} + \text{Recall}} \quad (9)$$

where β^2 is set as 0.3.

2) *MAE*: MAE is utilized to calculate pixel-wise average error between ground truth and predicted salient map which is normalized to [0,1]

$$\text{MAE} = \frac{1}{W \times H} \sum_{x=1}^w \sum_{y=1}^H |P(x, y) - G(x, y)| \quad (10)$$

where W and H are width and height of $P(x, y)$ and $G(x, y)$, $P(x, y)$ is predicted salient map and $G(x, y)$ is ground truth. The smaller the MAE, the better the saliency detection effect.

3) *WF*: weight ω was introduced to WF based on FM to consider the neighborhood information, it is defined as

$$\text{WF} = \frac{(1 + \beta^2) \times \text{Precision}^\omega \times \text{Recall}^\omega}{\beta^2 \times \text{Precision}^\omega + \text{Recall}^\omega}. \quad (11)$$

4) *SM*: SM is used to measure regional and object-level structural similarity between ground truth and predicted salient map

$$S_\alpha = \alpha * S_r + (1 - \alpha) * S_o \quad (12)$$

where S_r and S_o are regional-based and object-based structural similarity of predicted salient map, α is always set as 0.5.

5) *EM*: EM combines local pixel values with the image-level mean value and captures image-level statistics and local pixel matching information

$$Q_s = \frac{1}{W \times H} \sum_{x=1}^w \emptyset(x, y) \quad (13)$$

where $\emptyset(x, y)$ is enhanced alignment matrix, and W and H are width and height of salient map.

C. Ablation Study

In this section, we validate the effectiveness of our model using U-structure, hybrid loss, and CA, respectively. Based on our datasets, ablation study was performed to validate the effectiveness of U-Structure by comparing the embedded U-structure ($U^2\text{-Net}_{\text{BCE}_{\text{loss}}}$) and the U-Net ($U\text{-Net}_{\text{BCE}_{\text{loss}}}$) without embedded U-structure.

TABLE I
ABLATION STUDY ON DIFFERENT CONFIGURATION

Configurations	SM	WF	MAE
Baseline (U-Net_BCE _{loss})	0.801	0.731	0.032
U-Net +U-Structure (U ² -Net_BCE _{loss})	0.906	0.820	0.020
U ² -Net+SSIM _{loss}	0.902	0.857	0.017
U ² -Net+IoU _{loss}	0.904	0.873	0.016
U ² -Net+Bce _{loss} +SSIM _{loss}	0.910	0.865	0.016
U ² -Net+IoU _{loss} +SSIM _{loss}	0.911	0.841	0.016
U ² -Net+Hy _{loss}	0.911	0.853	0.015
U ² -Net+ Hy _{loss} +CA(HLU ² -Net)	0.915	0.872	0.014

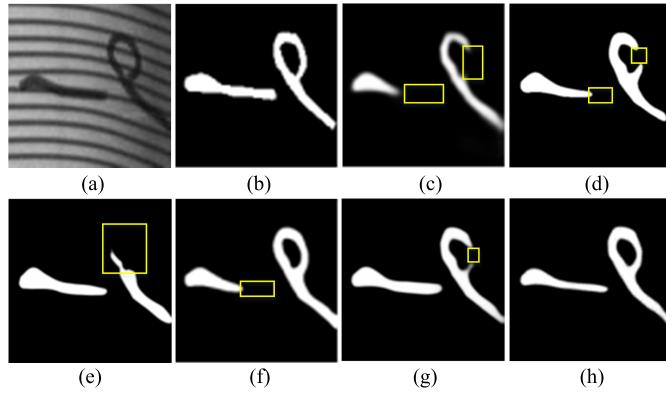


Fig. 6. Saliency maps obtained with different hybrid loss. (a) Test image, (b) ground truth, (c) Bce_{loss}, (d) IoU_{loss}, (e) Ssim_{loss}, (f) Bce_{loss} + Ssim_{loss}, (g) Ssim_{loss} + IoU_{loss}, and (h) our hybrid loss.

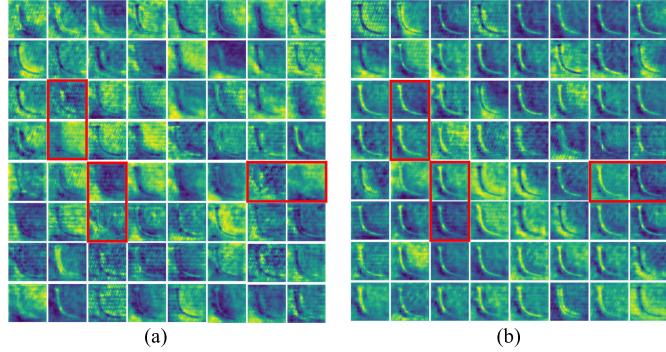


Fig. 7. Visual feature map after En_4. (a) and (b) are feature heatmap without and with CA, respectively.

As shown in Table I, experimental results indicate that compared with U-Net, the SM and WF of U²-Net are improved by 13.1% and 12.2%, respectively, and the MAE is dropped by 60%. To demonstrate the effectiveness of hybrid loss, we sequentially add Bce_{loss}, Ssim_{loss}, IoU_{loss} functions on U²-Net. The quantitative evaluation results also show that with the introduced loss function, the detection accuracy improves gradually on testing set. The SM and WF of U²-Net with our hybrid loss function are improved by 0.6% and 4.0% than that of with Bce_{loss} function, respectively, and the MAE is dropped by 33.3%. As one of the most commonly used loss function in segmentation tasks, Bce_{loss} is used to narrow the gap between the prediction mask and ground truth. Ssim_{loss} is a patch-level measure, which considers a local neighborhood of each pixel [55]. Ssim_{loss} assigns higher weight to the contour

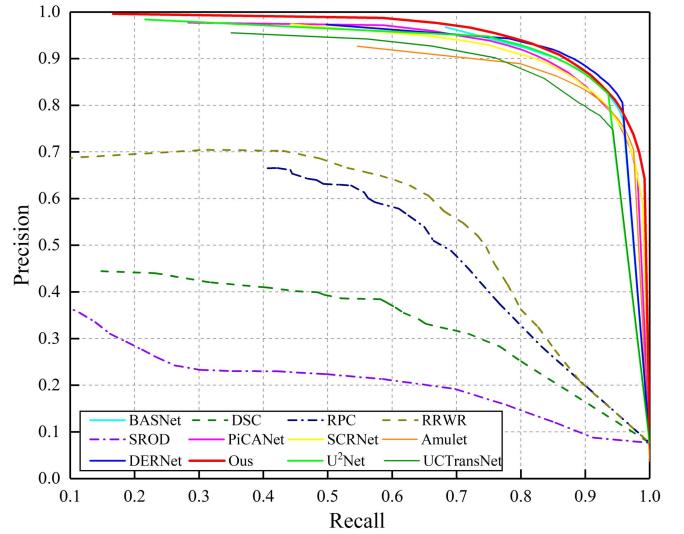


Fig. 8. PR curves of our method and other state-of-the-art method on testing set.

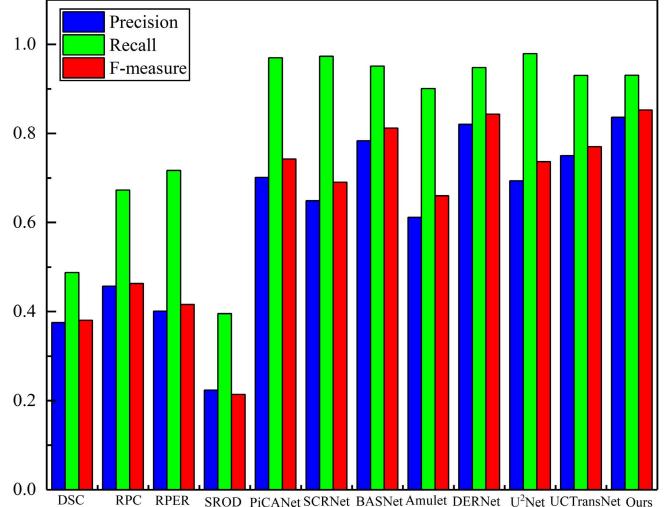


Fig. 9. Precision, Recall, and F-measure of our method and other state-of-the-art method on testing set.

profile of saliency object to obtain a finer boundary. IoU_{loss} has scale invariance and is utilized to increase coincidence between predicted mask and ground truth. As shown in Fig. 6, qualitative evaluation results show that with the Ssim_{loss} and IoU_{loss} function, the completeness of the predicted masks has more clear boundary and are more consistent with the ground truth. To verify the effectiveness of CA module, it is found that the SM and WF with CA are improved by 0.4% and 2.2% than that without CA, respectively. Moreover, as shown in Fig. 7, qualitative analysis of a feature map with and without CA module is conducted, it can be seen that in the feature heat map with CA useful features are highlighted and irrelevant features are suppressed. The yellow area of the heat map is the feature focus area, and the red boxes in feature map after En_4 clearly indicate the difference of feature map with and without CA. Representative features can be highlighted.

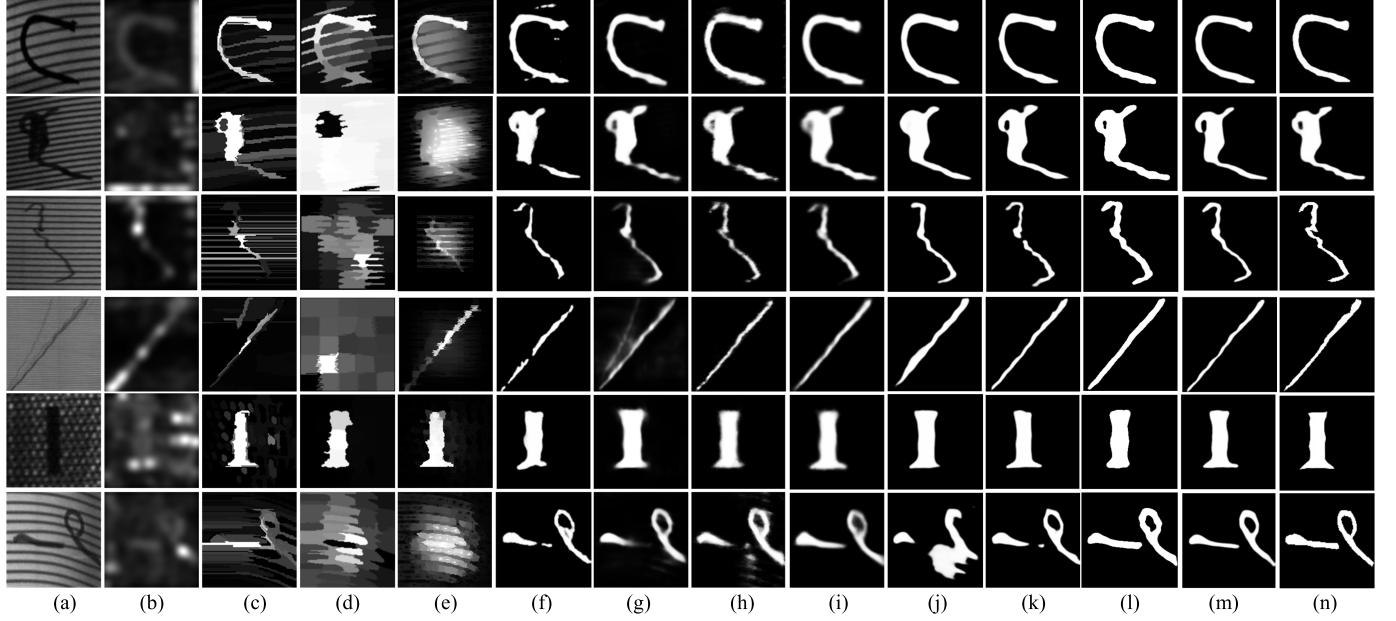


Fig. 10. Qualitative evaluation of saliency maps. (a) Test image, (b) SROD, (c) RPC, (d) DSC, (e) PPWR, (f) Amulet, (g) SCRNet, (h) PiCANet, (i) U^2 -Net, (j) BASNet, (k) DERNet, (l) UCTransNet, (m) HLU 2 -Net, and (n) ground truth.

Hence, ablation study indicates that the tricks proposed in this study are effective in getting robust models with application to tire defect detection.

D. Comparative Experiments

Comparative experiments are conducted to further validate the efficiency of the proposed scheme. We compare our model with 11 state-of-the-art salient detection models among which four models are traditional methods including DSC [66], RRWR [67], SROD [68], and RPC [69], and seven models are deep learning methods including Amulet [40], SCRNet [44], U^2 Net [53], PiCANet [48], BASNet [55], UCTransNet [70], and DERNet [45]. All experiments are conducted under the same environment and all deep learning methods are performed on the same training set and with the default hyper parameters.

1) *Quantitative Evaluation*: As shown in Fig. 8, it is shown that the precision-recall curves of deep learning-based methods are obviously higher than that of the traditional-based methods. Additionally, among all the methods in comparison, our method has satisfactory performance. Fig. 9 shows the Precision, Recall, and F-measure of the comparison experiments. Although the recall of our method is slightly lower than the others, the precision and comprehensive evaluation indicator F-measure outperform that of the state-of-the-art.

In addition, we show EM, SM, WF, F-measure, and MAE scores in Table II. Except the WF metrics in which DERNet has the highest score, our proposed method shows promising results in EM, SM, FM over the state-of-the-art methods.

2) *Qualitative Evaluation*: Fig. 10 shows qualitative evaluation results which demonstrate the superiority of our HLU 2 -Net over the state-of-the-art methods in terms of both subjective defect integrity and quantitative metrics. Specifically, the saliency map of our method has cleaner borders and

TABLE II
FIVE QUANTITATIVE EVALUATION METRICS OF DIFFERENT SALIENT DETECTION METHODS

Method	Emeasure	Smeasure	Weighted_F	Fmeasure	MAE
DSC[66]	0.606	0.515	0.299	0.381	0.351
RRWR[67]	0.639	0.664	0.355	0.416	0.136
SROD[68]	0.598	0.510	0.128	0.214	0.193
RPC[69]	0.714	0.685	0.416	0.463	0.092
Amulet[40]	0.872	0.783	0.607	0.660	0.044
SCRNet[44]	0.857	0.881	0.739	0.690	0.025
UCTransNet[70]	0.888	0.847	0.775	0.756	0.036
U^2 Net[53]	0.900	0.908	0.825	0.737	0.021
PiCANet[48]	0.905	0.895	0.812	0.743	0.024
BASNet[55]	0.949	0.907	0.864	0.812	0.017
DERNet[45]	0.965	0.914	0.882	0.843	0.014
HLU 2 -Net	0.970	0.915	0.872	0.853	0.014

Red, green and blue represent the best, second best and third best performance, respectively.

retains more edge details than that of the others. For defects with low contrast, e.g., the defects in the fourth and fifth row, our method can generate complete and clear saliency maps which intuitively supported the aforementioned quantitative results. For samples with multiple defects in a single test image, e.g., the sixth row in Fig. 10, our method can still maintain the integrity of the defects while the other methods have, more or less, incompleteness or false alarm phenomenon in map integrity, or both.

3) *Detection Efficiency*: Detection efficiency denotes an important evaluation metric of model performance. Table III shows the average detection time of our method and the 11 state-of-the-art methods on our test set. Experimental results indicate that our method has the highest detection efficiency (0.064 s/image) compared to other algorithms, which can meet real-time detection on production line. It is also shown that although the parameter amount of SCRNet

TABLE III
DETECTION TIME OF DIFFERENT SALIENT DETECTION METHODS

Method	DSC	RRWR	SROD	RPC	Amulet	SCRNet	U ² Net	PiCANet	BASNet	DERNet	UCTransNet	HLU ² -Net
Time(s)	0.678	0.237	3.938	0.186	0.098	0.069	0.077	0.265	0.066	0.078	0.273	0.064
Size(MB)	—	—	—	—	126	96.7	168	180	332	150	760.6	97.8
Code	Matlab	Matlab	Matlab	Matlab	Pytorch	Pytorch	Pytorch	Pytorch	Pytorch	Pytorch	Pytorch	Pytorch

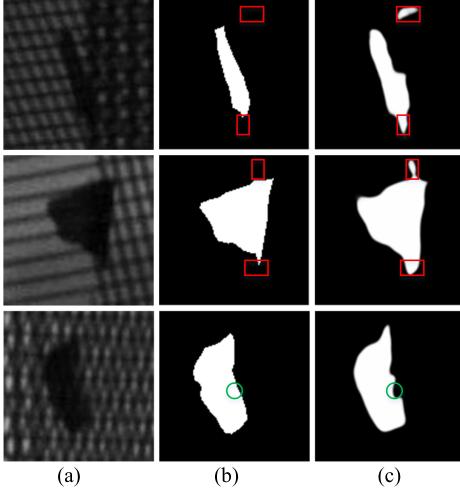


Fig. 11. Failed prediction sample. (a) Test image, (b) ground truth, and (c) predicted mask of HLU²-Net.

(96.7 MB) is 1.14% lower than that of our method (97.8 MB), the detection efficiency of our method (0.064 s/image) is 7.81% faster than that of SCRNet (0.069 s/image). Specifically, in SCRNet, an effective CRU is proposed to achieve information interaction between the task of salient object detection and edge detection, which leads to dense connections and reduces detection efficiency. Moreover, too many CRUs stacked horizontally in SCRNet also affect the real-time detection.

E. Failure Sample Analysis

Failure samples of the proposed method are shown in Fig. 11, false positive (red box) and false negative (green box) indicate wrongly predicted regions, respectively. It can be seen that false positive region always appears at the junction between sidewall and tread, the interlaced textures increase the difficulty of network learning, causing some nondefective backgrounds to be incorrectly predicted as defects. Moreover, due to the low contrast between the anisotropic texture background and defects on the tread, and there is no obvious pixel gradient change between them, which leads to the loss of some edge target predictions.

In this work, since the boundary refinement module is not applied, the boundary of the predicted mask still needs further improvement. In addition, although the probability of existence is relatively small, there are some other types of tire defects such as bubbles, cords defects, etc. which are not considered in this work.

In the future work, we consider introducing self-attention mechanisms, such as transformer, to further strengthen

network learning and deepen understanding of complex features to reduce the probability of false positives and false negatives. Moreover, a weakly supervised network for few-shot learning is worthy of our consideration.

V. CONCLUSION

In this work, a novel end-to-end HLU²-Net is proposed for automated tire defect detection. A U-Net embedded with residual U-structure and hybrid loss are integrated for multiscale multilevel feature fusion, guiding network training and joint decision-making. Ablation study shows that residual U-structure can make full use of multiscale, multilevel features. The hybrid loss can preserve the integrity of the defect details. Moreover, in HLU²-Net, coordinate attention is introduced to further improve detection effect by highlighting useful features and weakening irrelevant features. Both qualitative and quantitative experimental evaluation results demonstrate the superiority of our HLU²-Net over other state-of-the-art methods in terms of both subjective defect integrity and six quantitative evaluation metrics on tire saliency dataset. Additionally, experimental results also show that the computing efficiency of our method can meet the requirement of online visual detection on tire production line.

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] 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.
- [3] 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.
- [4] 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, Aug. 2013, Art. no. 084205.
- [5] X. Cui, Y. Liu, and C. Wang, “Defect automatic detection for tire X-ray images using inverse transformation of principal component residual,” in *Proc. 3rd Int. Conf. Artif. Intell. Pattern Recognit. (AIPR)*, Sep. 2016, pp. 1–8.
- [6] Y. Xiang, “Tire defect detection using local and global features,” *Comput. Aided Drafting, Des. Manuf.*, vol. 23, no. 4, pp. 49–52, 2013.
- [7] X. Zheng, J. Ding, Z. Pang, and J. Li, “Detection of impurity and bubble defects in tire X-ray image based on improved extremum filter and locally adaptive-threshold binaryzation,” in *Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC)*, Dec. 2018, pp. 360–365.
- [8] Y. Xiang, C. Zhang, and Q. Guo, “A dictionary-based method for tire defect detection,” in *Proc. IEEE Int. Conf. Inf. Autom. (ICIA)*, Jul. 2014, pp. 519–523.
- [9] Q. Guo, C. Zhang, H. Liu, and X. Zhang, “Defect detection in tire X-ray images using weighted texture dissimilarity,” *J. Sensors*, vol. 2016, Mar. 2016, Art. no. 4140175.
- [10] M. Zhang, Q. Guo, and X. Yang, “Tire defect detection on impurities,” in *Proc. CADDM*, 2014, vol. 24, no. 1, pp. 32–35.
- [11] 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. 4, Apr. 2018, Art. no. 1850011.

- [12] H. Tada and A. Sugiura, "Defect classification on automobile tire inner surfaces with functional classifiers," *Trans. Inst. Syst., Control Inf. Eng.*, vol. 34, no. 1, pp. 1–10, 2021.
- [13] Z. Zheng *et al.*, "Tire defect classification using a deep convolutional sparse-coding network," *Meas. Sci. Technol.*, vol. 32, no. 5, 2021, Art. no. 055401.
- [14] Z. Wu *et al.*, "Tire defect detection based on faster R-CNN," in *Proc. ICRRRI*, 2020, pp. 203–218.
- [15] C.-Y. Chang, K. Srinivasan, W.-C. Wang, G. P. Ganapathy, D. R. Vincent, and N. Deepa, "Quality assessment of tire shearography images via ensemble hybrid faster region-based ConvNets," *Electronics*, vol. 9, no. 1, p. 45, Dec. 2019.
- [16] Y. Li, B. Fan, W. Zhang, and Z. Jiang, "TireNet: A high recall rate method for practical application of tire defect type classification," *Future Gener. Comput. Syst.*, vol. 125, pp. 1–9, Dec. 2021.
- [17] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire defect detection using fully convolutional network," *IEEE Access*, vol. 7, pp. 43502–43510, 2019.
- [18] Z. Zheng, S. Zhang, B. Yu, Q. Li, and Y. Zhang, "Defect inspection in tire radiographic image using concise semantic segmentation," *IEEE Access*, vol. 8, pp. 112674–112687, 2020.
- [19] Z. Zheng, S. Zhang, J. Shen, Y. Shao, and Y. Zhang, "A two-stage CNN for automated tire defect inspection in radiographic image," *Meas. Sci. Technol.*, vol. 32, no. 11, Nov. 2021, Art. no. 115403.
- [20] M. Zhao *et al.*, "Faster mean-shift: GPU-accelerated clustering for cosine embedding-based cell segmentation and tracking," *Med. Image Anal.*, vol. 71, Jul. 2021, Art. no. 102048.
- [21] W. Chu, P. S. Ho, and W. Li, "An adaptive machine learning method based on finite element analysis for ultra low-K chip package design," *IEEE Trans. Compon., Packag., Manuf. Technol.*, vol. 11, no. 9, pp. 1435–1441, Sep. 2021, doi: [10.1109/TCMT.2021.3102891](https://doi.org/10.1109/TCMT.2021.3102891).
- [22] Y. Zhang, Y. Lu, W. Chen, Y. Chang, H. Gu, and B. Yu, "MSMANet: A multi-scale mesh aggregation network for brain tumor segmentation," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107733.
- [23] M. Zhao *et al.*, "VoxelEmbed: 3D instance segmentation and tracking with voxel embedding based deep learning," 2021, *arXiv:2106.11480*.
- [24] B. Jin, L. Cruz, and N. Goncalves, "Deep facial diagnosis: Deep transfer learning from face recognition to facial diagnosis," *IEEE Access*, vol. 8, pp. 123649–123661, 2020.
- [25] Y. He, K. Song, Q. Meng, and Y. Yan, "An end-to-end steel surface defect detection approach via fusing multiple hierarchical features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, pp. 1493–1504, Apr. 2020.
- [26] G. Hu, J. Huang, Q. Wang, J. Li, Z. Xu, and X. Huang, "Unsupervised fabric defect detection based on a deep convolutional generative adversarial network," *Textile Res. J.*, vol. 90, nos. 3–4, pp. 247–270, Feb. 2020.
- [27] M. Chang, B.-C. Chen, J. L. Gabayno, and M.-F. Chen, "Development of an optical inspection platform for surface defect detection in touch panel glass," *Int. J. Optomechatron.*, vol. 10, no. 2, pp. 63–72, 2016.
- [28] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," in *Matters of Intelligence*. Dordrecht, The Netherlands: Springer, 1987, pp. 115–141.
- [29] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998.
- [30] A. Borji, M.-M. Cheng, H. Jiang, and J. Li, "Salient object detection: A benchmark," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5706–5722, Dec. 2015.
- [31] T. Liu *et al.*, "Learning to detect a salient object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 2, pp. 353–367, Feb. 2011.
- [32] R. Achanta *et al.*, "Salient region detection and segmentation," in *Proc. ICVS*, 2018, pp. 66–75.
- [33] R. Achanta, S. Hemami, F. Estrada, and S. Sussstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1597–1604.
- [34] H. Jiang, J. Wang, Z. Yuan, Y. Wu, N. Zheng, and S. Li, "Salient object detection: A discriminative regional feature integration approach," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2083–2090.
- [35] W. Zhu, S. Liang, Y. Wei, and J. Sun, "Saliency optimization from robust background detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2814–2821.
- [36] M.-M. Cheng, N. J. Mitra, X. Huang, P. H. S. Torr, and S.-M. Hu, "Global contrast based salient region detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 569–582, Mar. 2015.
- [37] S. Goferman, L. Zelnik-Manor, and A. Tal, "Context-aware saliency detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 10, pp. 1915–1926, Oct. 2012.
- [38] Z. Ren, S. Gao, L.-T. Chia, and I. W.-H. Tsang, "Region-based saliency detection and its application in object recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 24, no. 5, pp. 769–779, May 2013.
- [39] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.
- [40] P. Zhang, D. Wang, H. Lu, H. Wang, and X. Ruan, "Amulet: Aggregating multi-level convolutional features for salient object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 202–211.
- [41] Y. Pang, X. Zhao, L. Zhang, and H. Lu, "Multi-scale interactive network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9413–9422.
- [42] J.-J. Liu, Q. Hou, M.-M. Cheng, J. Feng, and J. Jiang, "A simple pooling-based design for real-time salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3917–3926.
- [43] J. Zhao, J.-J. Liu, D.-P. Fan, Y. Cao, J. Yang, and M.-M. Cheng, "EGNet: Edge guidance network for salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8779–8788.
- [44] Z. Wu, L. Su, and Q. Huang, "Stacked cross refinement network for edge-aware salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7264–7273.
- [45] G. Song, K. Song, and Y. Yan, "EDRNet: Encoder-decoder residual network for salient object detection of strip steel surface defects," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9709–9719, Dec. 2020.
- [46] J. Su, J. Li, Y. Zhang, C. Xia, and Y. Tian, "Selectivity or invariance: Boundary-aware salient object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 3799–3808.
- [47] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3085–3094.
- [48] N. Liu, J. Han, and M.-H. Yang, "PiCANet: Learning pixel-wise contextual attention for saliency detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3089–3098.
- [49] Y. Piao, W. Ji, J. Li, M. Zhang, and H. Lu, "Depth-induced multi-scale recurrent attention network for saliency detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 7254–7263.
- [50] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [51] W. Shi, F. Jiang, and D. Zhao, "Single image super-resolution with dilated convolution based multi-scale information learning inception module," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 977–981.
- [52] F. Iandola, M. Moskewicz, S. Karayev, R. Girshick, T. Darrell, and K. Keutzer, "DenseNet: Implementing efficient ConvNet descriptor pyramids," 2014, *arXiv:1404.1869*.
- [53] X. Qin, Z. Zhang, C. Huang, M. Dehghan, O. R. Zaiane, and M. Jagersand, "U2-net: Going deeper with nested U-structure for salient object detection," *Pattern Recognit.*, vol. 106, Oct. 2020, Art. no. 107404.
- [54] Y. Li and J. Li, "An end-to-end defect detection method for mobile phone light guide plate via multitask learning," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–13, 2021.
- [55] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan, and M. Jagersand, "BASNet: Boundary-aware salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7479–7489.
- [56] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Oper. Res.*, vol. 134, no. 1, pp. 19–67, 2005.
- [57] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. ACSSC*, Nov. 2003, pp. 1398–1402.
- [58] M. Rahman and Y. Wang, "Optimizing intersection-over-union in deep neural networks for image segmentation," in *Proc. ISVS*, 2016, pp. 234–244.
- [59] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
- [60] S. Woo *et al.*, "CBAM: Convolutional block attention module," in *Proc. ECCV*, 2018, pp. 3–19.

- [61] Q. Hou, D. Zhou, and J. Feng, "Coordinate attention for efficient mobile network design," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13713–13722.
- [62] R. Margolin, L. Zelnik-Manor, and A. Tal, "How to evaluate foreground maps?" in *Proc. CVPR*, 2014, pp. 248–255.
- [63] F. Perazzi, P. Krahenbuhl, Y. Pritch, and A. Hornung, "Saliency filters: Contrast based filtering for salient region detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 733–740.
- [64] D.-P. Fan, M.-M. Cheng, Y. Liu, T. Li, and A. Borji, "Structure-measure: A new way to evaluate foreground maps," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4548–4557.
- [65] D.-P. Fan, C. Gong, Y. Cao, B. Ren, M.-M. Cheng, and A. Borji, "Enhanced-alignment measure for binary foreground map evaluation," 2018, *arXiv:1805.10421*.
- [66] X. Wu, X. Ma, J. Zhang, A. Wang, and Z. Jin, "Salient object detection via deformed smoothness constraint," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2815–2819.
- [67] C. Li, Y. Yuan, W. Cai, Y. Xia, and D. Dagan Feng, "Robust saliency detection via regularized random walks ranking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2710–2717.
- [68] H. Tang, C. Chen, and X. Pei, "Visual saliency detection via sparse residual and outlier detection," *IEEE Signal Process. Lett.*, vol. 23, no. 12, pp. 1736–1740, Dec. 2016.
- [69] J. Lou, M. Ren, and H. Wang, "Regional principal color based saliency detection," *PLoS ONE*, vol. 9, no. 11, Nov. 2014, Art. no. e112475.
- [70] H. Wang, P. Cao, J. Wang, and O. R. Zaiane, "UCTransNet: Rethinking the skip connections in U-Net from a channel-wise perspective with transformer," 2021, *arXiv:2109.04335*.
- [71] B. C. Russell, A. Torralba, K. P. Murphy, and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 157–173, 2008.



Zhouzhou Zheng received the M.S. degree from the College of Electromechanical Engineering, Qingdao University of Science and Technology, Qingdao, China. He is currently pursuing the Ph.D. degree with the College of Mechanical and Electrical Engineering, Northwest A&F University, Yangling, China.

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



Huanbo Yang was born in Shangdong, China, in 1995. He received the M.S. degree from Shandong University of Technology, Zibo, China, in 2020. He is currently pursuing the Ph.D. degree with the College of Mechanical and Electrical Engineering, Northwest A&F University, Yangling, Shaanxi, China.

His research interests include image processing and computer vision.



Liang Zhou was born in Henan, China, in 1991. He received the B.S. degree from Jiamusi University, Jiamusi, China, in 2015, and the M.S. degree from Harbin University of Science and Technology, Harbin, Heilongjiang, China, in 2018. Now he is pursuing the Ph.D. degree with the School of Mechanical Engineering and Automation, Northeastern University, Shenyang, Liaoning, China.

His research interests include image processing and cutting technology.



Bin Yu received the M.Sc. degree in computational mathematics from Shanghai University, Shanghai, China, in 2005.

He is currently an Associate Professor with the College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao, China. From 2018 to 2019, he was a Visiting Professor with the University of Calgary, Calgary, AB, Canada. His research interests include machine learning, image processing, and data mining.



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 Qingdao University of Science and Technology, Qingdao, China, in 2009 and 2014, respectively.

He is an Associate Professor of Electrical and Computer Engineering with Qingdao University of Science and Technology. He was a Visiting Scholar with the Research Group on Electrical Engineering and Automatic Control (GREA), Faculty of Sciences, Normandy University, Le Havre, France, from 2014 to 2015. He visited the Shanghai Jiaotong University, Shanghai, China, in 2018. He has published more than 70 research articles in peer-reviewed international journals and proceedings. His research interests include computer vision, image processing, and nondestructive testing.

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