RESEARCH ARTICLE

Power Adapter Appearance Defect Detection Based on Task Feature Decoupling YOLOv8n

JIE CHEN, YU XIE [ID], ZHENGWEI QIAN, KEQIONG CHEN [ID], MAOFA ZHEN, AND XUEYOU HU

School of Advanced Manufacturing Engineering, Hefei University, Hefei, Anhui 230601, China

Corresponding author: Yu Xie (xiey@hfuu.edu.cn)

ABSTRACT In the realm of defect detection, there are distinctions in the emphasis placed on features between the classification and localization components of the task. The classification task emphasizes semantic information from the global context, while the localization task prioritizes spatial details such as edges. Directly coupling these two subtask features can hinder model convergence and degrade performance in the appearance defect detection of power adapters. To address this issue, we proposed the Task-feature Decoupled Feature Pyramid Network (TDFPN) module based on YOLOv8n. This module enhances semantic information and fuses corresponding features to improve detection performance in both localization and classification tasks. Additionally, we introduced the EMA module to suppress redundant information, enhance the model's attention towards defects, and improve the precision rates of detection. Furthermore, we replaced CIoU with an Inner-SIoU loss function that combines Inner-IoU based on auxiliary bounding boxes with SIoU, considering the matching direction. This replacement accelerated model convergence and improved the recall rates of detection. During training, transfer learning is employed by utilizing pre-trained weights from the YOLOv8n backbone, along with frozen training, to enhance efficiency. The experimental findings indicated that our proposed approach outperforms the original YOLOv8n model, demonstrating a 3.12% enhancement in mAP@0.5 and a 14.41% improvement in mRecall@0.5.

INDEX TERMS Defect detection, deep learning, object detection, power adapter, YOLOv8.

I. INTRODUCTION

Various factors in the production process of the power adapter can lead to defects on its surface. However, some aspects of the texture and defects of the power adapter cannot be accurately distinguished, which may compromise the efficiency and accuracy of quality inspection. At present, common defect detection methods include thermal imaging detection [1], laser scanning methods [2], and manual inspection method. The non-destructive testing technology mentioned above relies on human judgment and is easily influenced by various factors, resulting in low detection accuracy, time-consuming processes, and labor-intensive tasks, making it somewhat inadequate for real-time detection tasks [3]. With the continuous advancement of computer technology,

The associate editor coordinating the review of this manuscript and approving it for publication was Szidonia Lefkovits [ID].

traditional machine learning and deep learning algorithms based on computer vision are gradually emerging in the field of defect detection.

In traditional machine learning-based methods, the features used for learning are mainly roughly extracted through traditional image processing operations such as edge detection, morphological processing, various image thresholding techniques, etc., such as HOG [4], and SIFT [5]. Then classification is performed using classifiers such as SVM [6]. The detection performance largely depends on the quality of the previously extracted features, which perform poorly in complex scenes, especially for complex defect detection. At the same time, each method is designed for specific applications and requires manual feature design, a lot of professional knowledge and experience, and cannot capture complex defect features, with poor generalization ability and robustness [7].

Detection algorithms based on deep learning do not require the manual design of feature extraction rules. They automatically extract deeper features through convolution and other operations, resulting in higher accuracy, applicability, and robustness. The object detection algorithm based on deep learning has a wide application space in remote sensing [8], traffic detection [9], [10], [11], anomaly detection [12], and industrial defect detection. With the rapid advancement of computer technology, the utilization of convolutional neural networks based on computer vision object detection algorithms for defect detection in the field is on the rise. However, there is still room for improvement in detecting multi-scale variations and tiny defects [13].

The current mainstream defect detection algorithms are divided into two main categories. One approach involves a two-stage process that relies on region suggestion, such as the Faster R-CNN [14] model. This two-stage object detection algorithm exhibits a high level of precision. However, it necessitates a substantial computational workload and demonstrates sluggish inference speeds. The other category is the single-stage object detector based on object regression, as exemplified by the YOLO [15] series, which is faster than the two-stage algorithm. The detection algorithm of the YOLO series has undergone several version iterations to ensure efficiency while improving accuracy, making it a popular choice in the field of engineering inspection. Subsequently, numerous academics suggested refining algorithms to augment the efficacy of identifying defects within models.

Zhang et al. [16] proposed an insulator defect detection algorithm based on an improved YOLOv8s model. Introduction of multi-scale large-kernel attention (MLKA) module to enhance focus on features of different scales and low-level feature maps. Adoption of an improved loss function using SIoU to optimize detection performance and enhance feature extraction capability for insulator defects. Zhong et al. [17] proposed a steel surface defect detection algorithm based on the improved YOLOv5 model. A Bidirectional Cascade Path Aggregation Network (BIC-PAN) structure is proposed to enhance feature fusion. The coordinate attention (CA) mechanism is introduced to further improve the model's detection performance. Zhou et al. [18] introduced a PCB defect detection algorithm based on MSD-YOLOv5. Introduction of attention mechanism to enhance feature extraction by highlighting important channels. Replacement of coupling detection head with decoupling detection head to extract and learn defect location and category information separately. The solution to the problem of highly coupling different information feature distributions and enhancement of model's generalization ability.

The above-mentioned algorithm mainly improved the multi-scale feature fusion, introduced attention mechanisms, decoupled the detection heads, and enhanced the model's detection performance by using better loss functions. However, decoupling only the detection head, while the features input for the classification and localization tasks are directly coupled. This structure has some issues as it ignores the spatial differences in attention points between the classification and localization subtasks. Where classification features require more semantic context information, while localization features require more object edge information. Direct coupling of features from different tasks leads to spatial misalignment, affecting network convergence [19].

To address the above-mentioned issue, a task feature decoupling YOLOv8n is proposed, which decouples the input features of the classification and localization detection heads, and integrates corresponding features for the classification and localization detection heads. The aim is to improve the detection accuracy of the model for power adapters while maintaining good real-time performance and lower parameter counts. This paper makes the following contributions: (1) Based on the YOLOv8n model, an improved algorithm was proposed, designing an effective feature fusion module called TDFPN, applying the EMA module to filter out redundant background information, and constructing a more suitable Inner-SIoU regression loss function. (2) We proposed the TDFPN module to decouple classification and localization features, introducing Global-Context Feature Interaction (GCFI) to enhance the global sensing ability of the network, and improve recall and precision in power adapter defect detection; (3) Applying the EMA attention mechanism to suppress redundant information and enhance detection precision for power adapter defects; (4) Developing the Inner-SIoU regression loss function can enhance the convergence of the model and notably enhance the recall rate for detecting power adapter defects.
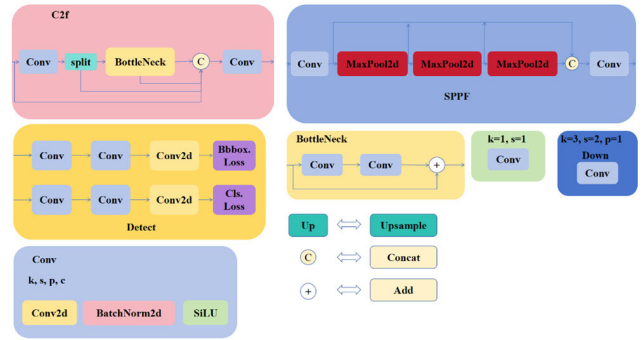
II. RELATED WORK
A. PREVIOUS RESEARCH
Xiao et al. [20] combined the transformer structure and Bi-FPN [21] structure with the YOLO-v5 model to propose the YOLOv5-Transformer-Bi-FPN (YOLOv5-TB) model. The YOLOv5-TB model is used for zinc-coated steel defect detection, which can accurately and quickly detect spangled defects on the surface of zinc-coated steel. Li et al. [22] used depth-wise separable convolution on the original YOLO network structure and enhanced the network's feature extraction by incorporating MECA (More Efficient Channel Attention). The ASPF (Atrous Fast Spatial Pyramid) module is developed by employing dilated convolutions with different expansion rates to capture a greater amount of contextual information. Utilizing dense multi-scale weighted fusion to enhance the accuracy of detection. Wu et al. [23] proposed a lightweight YOLOX surface defect detection network and introduced the Multi-scale Feature Fusion Attention Module (MFFAM). Lightweight CSP structures are used to optimize the backbone of the original network. MFFAM uses different scales of receptive fields for feature maps of different resolutions, after which features are fused and passed into the spatial and channel attention modules in parallel.
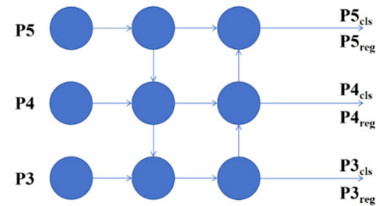
B. YOLOv8

The YOLOv8 [24] object detection algorithm is innovative and optimized based on the original YOLO series, to enhance model performance. Firstly, YOLOv8 builds on the design concept of YOLOv7 [25] to optimize the backbone network. The C3 module has been replaced by the C2f module with a richer gradient flow, improving the performance of the model. Secondly, it introduces Distribution Focal Loss [24]. Additionally, Mosaic data enhancement [26] is utilized to improve the accuracy of the model. Finally, decoupled detection heads and An Anchor-free design were adopted. These improvements have enabled YOLOv8 to achieve good performance, but there are still issues with low accuracy in the power adapter detection scene. Firstly, there are defects in the feature fusion of the neck structure of YOLOv8n, where the features obtained by the classification detection head and regression detection head are coupled, while these two subtasks have different focus points on features. Secondly, the feature extraction capability of the model for redundant and complex background information is limited. Lastly, there are many low-quality samples in the defect samples of the power adapter detection scene, and CIoU only considers aspect ratios. YOLOv8n is a lightweight version of YOLOv8. Considering the industrial application requirements for power adapter defect detection, it is essential to focus on both detection accuracy and speed. Based on the standard YOLOv8n algorithm, its neck structure and regression loss function are improved, and a defect detection algorithm for power adapters, named YOLOv8n with decoupled task features, is proposed. The model structure is shown in Figure 1 and Figure 2, and the details of the structure are presented in Sections II-A and II-B.
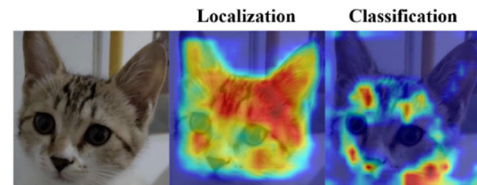

FIGURE 1. Improved YOLOv8n framework.

III. METHODS

A. BETTER FEATURE FUSION TDFPN

The YOLOv8n network utilizes the PANet [27] network structure for feature fusion, where the input features of

FIGURE 2. Improved YOLOv8n some detail of structure.

the classification and localization detection heads interact. As shown in Figure 3:

FIGURE 3. PAFPN feature fusion framework.

However, there are differences in feature focus across tasks. The network structure provides coupled features for the localization and classification heads. YOLOv8n (Detection COCO) pre-trained weights were used to visualize the Grad-Cam [28] heat map of the detected object. The gradients of the regression head and classification head backpropagation are summed respectively, and the heat maps of the classification head and regression head are generated, as shown in Figure 4 and Figure 5.

FIGURE 4. Comparison of thermal maps between classification and positioning detection heads.

The classification task differs from the localization task. Localization emphasizes edge features, whereas classification emphasizes local features within a global perspective. The direct coupling of classification features and regression features can result in spatial misalignment [19], which, in turn, impacts the convergence of the network.

The classification task is a coarse-grained task that requires richer semantic contextual information [29]. The localization task, on the other hand, is a fine-grained task that is
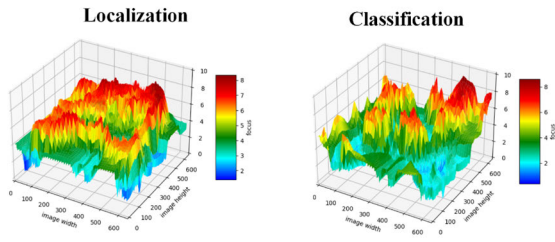
FIGURE 5. Visualization of the differences in the focus of classification and localization tasks.

more critical for providing detailed boundary information. Therefore, we need to decouple the features based on the task, improve the semantic information of the classification features, and integrate shallow features with more detailed spatial information for the localization features. The classification, localization, feature decoupling, and coupling are shown in Figure 6.
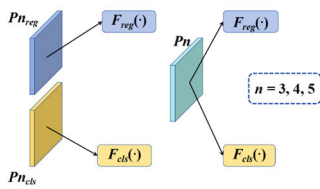

FIGURE 6. Classification, regression features decoupling, and coupling schematic diagram.

TDFPN decouples the classification and localization features, and integrates the features of interest into the subtasks. The structure is illustrated in Figure 7.
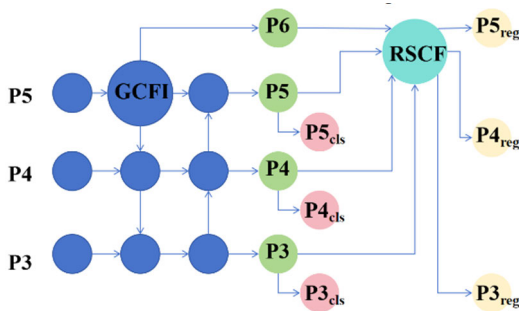

FIGURE 7. TDFPN features fusion framework.

TDFPN is enhanced by incorporating the PAFPN structure and introducing RSCF (Regress Specific Context Fusion) to generate more detailed edge features for localization tasks. Additionally, GCFI (Global Context Feature Interaction) is introduced to handle high-level features with richer semantic concepts, such as P5 [30], to enhance their semantic information. The GCFI module handles its processing as follows:

$$Q = K = V = Flatten(\text{P5}) \quad (1)$$

$$\text{P5} = Reshape(FFN(MSHA(Q, K, V))) \quad (2)$$

Firstly, feature P5 is flattened, serialized, and then embedded with position information. The terms Q, K, and V denote Query, Key, and Value respectively in this context. Semantic information enhancement is then carried out by GCFI, which is a combination of MHSA (Multi-head Self-attention) [31], [32] and FFN (Feed Forward Network) [31]. Finally, the reshape operation is applied to restore the shape of P5, which is the inverse operation of Flatten. The structure of the GCFI module shown in Figure 8 is illustrated.
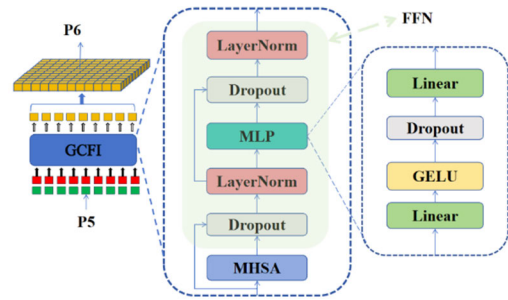

FIGURE 8. Structure of GCFI.

Thanks to the global attention ability of multi-head attention, the network's global sensing field is effectively improved without losing information about small objects, enhancing the detection ability of large-scale change defects. By introducing GCFI, the network can capture more semantic information, enhancing the global sensing ability and laying a stronger foundation for the subsequent modules to detect and recognize objects in images.

The P5 processed through GCFI handling and channel dimension reduction is called P6. P5, P4, and P3 are three effective feature layers obtained by fusing PAFPN features, and {P3, P4, P5} undergoes channel dimension reduction through convolution. If the original feature before reduction is $P_l \in \mathbb{R}^{B_l \times C_{l1} \times H_l \times W_l}$, then the reduced feature is $P_l \in \mathbb{R}^{B_l \times C_{l2} \times H_l \times W_l}$, where B represents the batch dimension, $C_{l1} = 2C_{l2}$ denotes the channel dimension, l is the index of the feature layer where $l = \{3, 4, 5, 6\}$, and H and W refer to the height and width dimensions. This process results in {P3, P4, P5, P6}. Thanks to the semantic information enhancement of the GCFI module, the {P3, P4, P5} features at this time have gained stronger semantic information, which is extracted as the three input features P3$_{cls}$, P4$_{cls}$, P5$_{cls}$ for the classification detection head. P6 will be utilized as a deep feature of P5 to contribute to the feature fusion in subsequent positioning tasks.

In regression tasks, richer spatial detail information is needed, which can usually be provided by shallow features. Generally speaking, features at the current level are more correlated with features at adjacent levels, while the correlation with features at other levels may be lower. Directly fusing features with large spans may lead to a "semantic gap" [19], where spatial details do not match semantic information, affecting network convergence. Therefore, when combining features, features from adjacent levels are usually

given priority. Thus, the RSCF module was born, as illustrated in Figure 9. The module takes {P3, P4, P5}, and P6 as input features, and through the fusion of adjacent shallow and deep features, it provides more detailed edge information for the localization head.
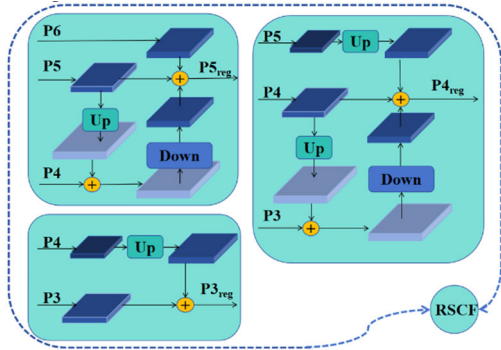

FIGURE 9. Structure of RSCF.

The process of RSCF can be expressed as follows.

$$\begin{cases} P3_{reg} = up\,(P4) + P3 \\ P4_{reg} = up\,(P5) + down\,(up\,(P5) + P3) + P4 \\ P5_{reg} = P6 + down\,(up\,(P5) + P4) + P5 \end{cases} \quad (3)$$

where, $up(\cdot)$ is the nearest interpolation upsample, $down(\cdot)$ is stride 2, and kernel size is 3×3 convolution.

B. ATTENTION MECHANISM MODULE EMA

The structure of the EMA [33] (Efficient Multi-Scale Attention) module is shown in Figure 10. The EMA module utilizes three parallel paths to extract features and generate attention weights. In the 1×1 branch, channels are encoded using two global average pooling operations, and features are aggregated using a 1×1 convolution followed by a sigmoid function. In the 3×3 branch, local cross-channel interactions are captured by employing 3×3 convolution to expand the feature space [34].

The EMA module achieves channel weighting and spatial coding by grouping the channels using a limited number of parameters with low complexity. After integrating it into the C2f module of the TDFPN, it suppresses redundant information, effectively enhancing the detection of power adapter defects. The cross-space learning component encodes the outputs of the 1×1 branch and 3×3 branch through global average pooling. The spatial attention map is obtained through a matrix dot product operation. This process effectively integrates spatial information across multiple scales in the output feature map, thereby enhancing the model's ability to selectively attend to defects at various scales.

C. BETTER REGRESSION LOSS FUNCTION INNER-SIOU

YOLOv8 classification loss uses Binary Cross Entropy Loss (BCE Loss) [35]. The regression loss utilizes Distribution Focal Loss and CIoU Loss [36]. The total loss is shown
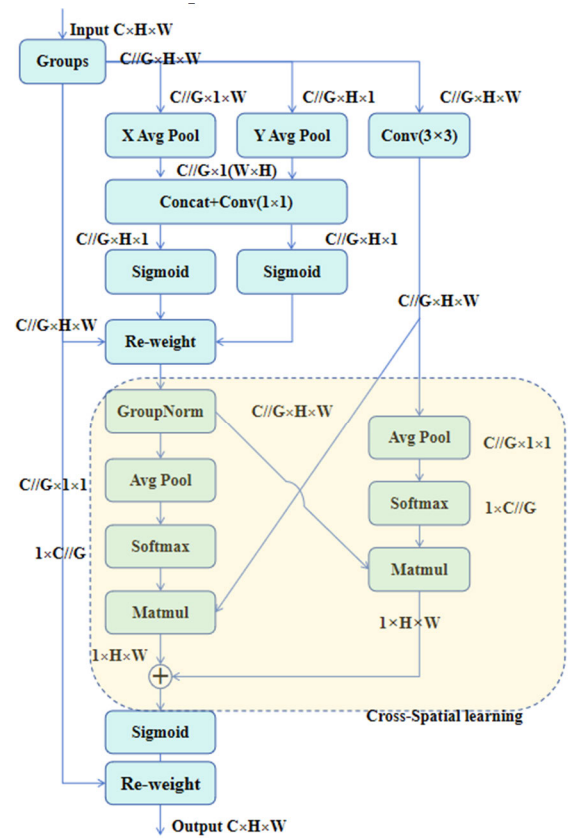

FIGURE 10. Structure of EMA.

in Equation (4).

$$Loss_{TOTAL} = Loss_{BOX} + Loss_{CLS} + Loss_{DFL} \quad (4)$$

CIoU only considers the difference in aspect ratio, and the loss calculation method does not differ for samples with high IoU and low IoU. In industrial defect detection scenarios, samples with low Intersection over Union (IoU) values are more prevalent than those with high IoU values. The penalties of distance and aspect ratio enhance the impact on samples with low Intersection over Union (IoU), which not only affects the convergence of the network but also reduces the generalization performance of the model [37].

The SIoU [38] method incorporates the angle loss into the bounding box regression loss function by considering the influence of the angle between the anchor box and the ground truth (GT) box on bounding box regression. This is mathematically defined as presented in Equation (5).

$$L_{SIoU} = 1 - IoU + \frac{(\Omega + \Delta)}{2} \quad (5)$$

The angular loss represents the minimum angle between the connection of the GT box and the center point of the Anchor box [39], as shown in Equation (6).

$$\Lambda = sin\left(2sin^{-1}\frac{\min(\left|x_c^{gt} - x_c\right|, \left|y_c^{gt} - y_c\right|)}{\sqrt{\left(x_c^{gt} - x_c\right)^2 + \left(y_c^{gt} - y_c\right)^2 + \in}}\right) \quad (6)$$

Angular loss is designed to adjust the anchor box to the nearest axis and prioritize proximity to either the X-axis or Y-axis based on the angle change. After considering the angular cost, the distance loss is redefined as follows:

$$\Delta = \frac{1}{2} \sum_{t=2,h} \left(1 - e^{-\gamma \rho t}\right), \quad \gamma = 2 - \Lambda \tag{7}$$

$$\begin{cases} \rho x = \left(\frac{b_x - b_x^{gt}}{w^c}\right)^2 \\ \rho y = \left(\frac{b_y - b_y^{gt}}{h^c}\right)^2 \end{cases} \tag{8}$$

Shape loss primarily refers to the difference in size between the Ground Truth (GT) box and the Anchor box, with the value of θ determining the importance of the shape cost [40]. The range of this parameter is from 2 to 6. The definition is as follows.

$$\Omega = \frac{1}{2} \sum_{t=w,h} \left(1 - e^{wt}\right)^\theta, \quad \theta = 4 \tag{9}$$

$$\begin{cases} w_w = \frac{|w - w_{gt}|}{max(w, w_{gt})} \\ w_h = \frac{|h - h_{gt}|}{max(h, h_{gt})} \end{cases} \tag{10}$$

Inner-IoU [41] is used to enhance existing bounding box regression methods. The method accelerates the regression process by using auxiliary bounding boxes and scale factor ratios. The parameters are illustrated in Figure 11, and the GT box and the Anchor box are represented as B^{gt} and B, respectively. The centroids of the GT box and the inner GT box are denoted by (x_c^{gt}, y_c^{gt}), while (x_c, y_c) represents the centroids of the Anchor box and the inner Anchor box. The width and height of the GT box are denoted by w^{gt} and h^{gt}, respectively, and those of the Anchor box are denoted by w and h. The variable-ratio corresponds to the scaling factor, which is usually in the range of [0.5, 1.5]. The derivation formula for Inner-IoU is as follows.

$$\begin{cases} b_l^{gt} = y_c^{gt} - \frac{w^{gt} * ratio}{2} \\ b_r^{gt} = y_c^{gt} - \frac{w^{gt} * ratio}{2} \end{cases} \quad \begin{cases} b_t^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2} \\ b_b^{gt} = y_c^{gt} - \frac{h^{gt} * ratio}{2} \end{cases} \tag{11}$$

$$\begin{cases} b_l = x_c - \frac{w * ratio}{2}, \quad b_r = x_c - \frac{w * ratio}{2} \\ b_t = y_c - \frac{h * ratio}{2}, \quad b_b = y_c - \frac{h * ratio}{2} \end{cases} \tag{12}$$

$$inter = \left(\min\left(b_r^{gt}, b_r\right) - \max\left(b_l^{gt}, b_l\right)\right) * \left(\min\left(b_b^{gt}, b_b\right) - \max\left(b_t^{gt}, b_t\right)\right) \tag{13}$$

$$union = \left(w^{gt} * h^{gt}\right) * (ratio)^2 + (w * h) * (ratio)^2 - inter \tag{14}$$

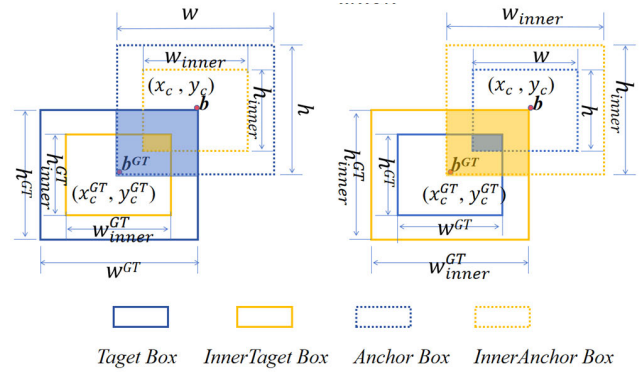$$IoU^{inner} = \frac{inter}{union} \tag{15}$$

FIGURE 11. Diagram of Inner-IoU parameters.

Applying the Inner-IoU loss to the SIoU is shown in equation (17):

$$Loss_{Inner-SIoU} = Loss_{SIoU} + IoU - IoU^{inner} \tag{16}$$

The total loss of replacing Inner-SIoU with CIoU is as follows:

$$Loss_{now} = Loss_{Inner-SIoU} + Loss_{CLS} + Loss_{DFL} \tag{17}$$

By introducing angle constraints and shape loss, SIoU accelerates the convergence speed of bounding box regression, allowing the model to learn the appearance defects of power adapters more quickly and improving detection performance. Inner-IoU enhances the model's detection capabilities for high and low IoU samples by calculating losses using auxiliary bounding boxes of different scales without introducing additional loss terms. Combining SIoU with Inner-IoU forms Inner-SIoU, which significantly improves the model's convergence speed, enhances the model's learning ability for quality-imbalanced IoU samples, strengthens the model's generalization ability and overall performance, and significantly increases the recall rate of power adapter defect detection. Experimental results demonstrating these improvements are presented in Section IV-D.

IV. EXPERIMENT
A. DATASET AND DEFECT CLASSES
In this study, a dataset comprising 235 images depicting various types of power adapter appearance defects was employed. Published at http://ieee-dataport.org/12647. These defects were categorized into five classes: label, mark, scratch, smudge, and spill, with each class containing 32, 38, 58, 80, and 32 images, respectively. To augment the dataset, various techniques such as vertical and horizontal flipping, mirroring, rotation, and panning were applied, resulting in a total of 4840 images. The dataset was then partitioned into training, validation, and test sets in an 8:1:1 ratio. Figure 12 shows a detailed image of label and mark defects, Figure 13 shows a detailed image of scratch and spill defects, and Figure 14 shows a detailed image of smudge defects.
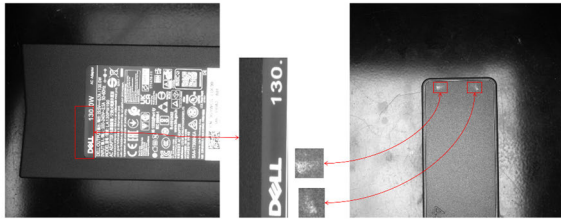
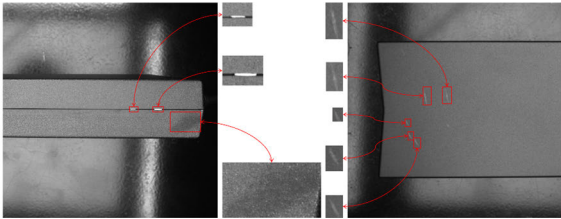**FIGURE 12.** Detail diagram of label and mark defect.
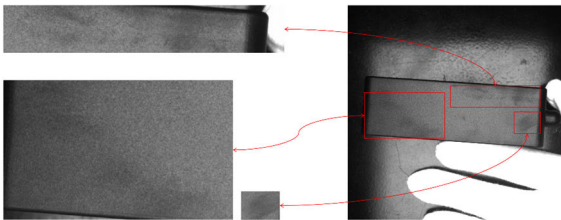

FIGURE 13. Detail diagram of scratch and spill defect.

FIGURE 14. Detail diagram of smudge defect.

TABLE 1. Comparison of FPS, and parameter with added improvement.

Method	FPS(Images/s)	Params(M)
YOLOv8n	88.875	3.012
+TDFPN	63.146	4.254
+EMA	76.993	3.013
+Inner-SIoU	88.875	3.012
+Add all three	54.851	4.255

B. EXPERIMENTAL ENVIRONMENT CONSTRUCTION AND EVALUATION METRICS

The experimental setup includes the utilization of a Windows 10 operating system, PyTorch 1.7.1 deep learning framework, an Intel i5-11400K CPU, and an RTX 3080 GPU. Input image size is 640×640, with 100 training epochs, and transfer learning utilizing the pre-trained weights of the model backbone. To safeguard the pre-trained weights of the backbone model from perturbation during the initial training phases, the weights are kept frozen for the first 50 epochs, allowing only model fine-tuning. The batch size is set at 16 during the frozen stage and 8 during the unfreezing stage. The model's initial learning rate is 0.01, with a momentum value of 0.937, using the SGD optimizer. The cosine learning rate scheduler is used, and a weight decay of 0.005 is applied.
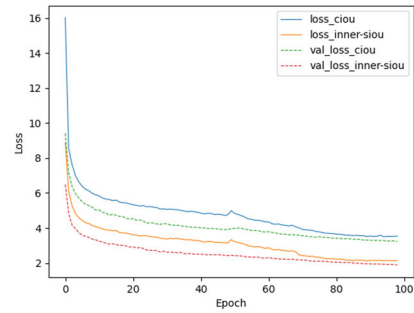

FIGURE 15. Change diagram of the loss function (removed the first value).

To assess the enhanced model's efficacy, its performance is measured across various metrics including precision (P), recall (R), average precision (AP), mean average precision (mAP), mean precision across all defect classes (mP), mean recall across all defect classes (mR), and mean average precision across all defect classes (mAP).

C. RESULTS AND ANALYSES

TDFPN is denoted as Method 1, EMA as Method 2, and Inner-SIoU as Method 3. The experiments are conducted on a defective dataset of power adapters, and the effectiveness of the improvement is determined by evaluation metrics such as mP@0.5, mR@0.5, mAP@0.5, (@0.5 represents IoU $= 0.5$), FPS, and Param.

The label on the defect type indicates a long tube shape. This type of feature usually requires a large receptive field with detailed edge information. The smudge shows significant scale changes, limited texture information, and considerable morphological changes. Therefore, for the smudge defect model, attention must be paid to weak edge details, local texture information, and the need for a large receptive field. Method 1 decouples the features and then enhances the P5 features with rich semantic information. This approach avoids feature coupling, which can cause the model to misinterpret blemish edges and result in leakage detection. Meanwhile, Method 1 enhances the global sensing ability of the network through the global contextual interaction of P5 features, thereby improving the accuracy of defect classification. Experimental results demonstrate that Method 1 increases the precision rate by 1.44%, the recall rate by 2.67%, and the mAP@0.5 by 0.72% compared to the original model. Furthermore, Method 1 improves the AP of the label and smudge defects classes by 2.71% and 1.37%, respectively, confirming the effectiveness of the enhancement. Method 2 focuses on the feature layer of Cf2 fusion to reduce redundant information and enhance the model's focus on defects. This consequently improves the precision of the network. Due to the smaller scale of marks, spills, and scratches, which bear some resemblance to the background noise, there was a decrease in R by 2.24%, 1.83%, and 4.80%, respectively. On the other hand, the larger scale of label and smudge improved R by 1.69% and 1.91%, respectively.

TABLE 2. Comparison of average precision with added improvement.

Method	label	mark	scratch	smudge	spill	mAP@0.5(%)
YOLOv8n	87.04	93.10	91.77	96.71	98.32	93.39
+TDFPN	89.75	93.27	91.33	98.08	98.11	94.11(+0.72)
+EMA	90.44	92.37	92.70	97.09	96.96	93.91(+0.52)
+Inner-SIoU	93.88	91.07	94.59	99.65	98.50	95.54(+2.15)
+Add all three	97.19	94.28	94.28	99.19	97.63	96.51(+3.12)

TABLE 3. Comparison of recall with added improvement.

Method	label	mark	scratch	smudge	spill	mR@0.5(%)
YOLOv8n	63.25	82.84	74.39	84.08	89.42	78.80
+TDFPN	73.50	80.60	75.00	91.72	86.54	81.47(+2.67)
+EMA	64.94	80.60	72.56	85.99	84.62	77.74(-1.06)
+Inner-SIoU	88.03	96.30	93.90	97.45	95.19	94.17(+15.37)
+Add all three	90.60	90.30	89.63	99.36	96.15	93.21(+14.41)

TABLE 4. Comparison of precision with added improvement.

Method	label	mark	scratch	smudge	spill	mP@0.5(%)
YOLOv8n	87.06	98.23	94.57	94.96	97.89	94.54
+TDFPN	90.53	96.43	96.09	97.96	98.90	95.98(+1.44)
+EMA	93.83	98.18	97.54	95.07	97.78	96.48(+1.94)
+Inner-SIoU	85.12	93.60	92.55	95.12	96.19	92.52(-2.02)
+Add all three	90.60	96.80	91.30	95.71	96.15	94.11(-0.43)

Although the mR@0.5 decreased by 1.06%, the model's precision mP@0.5 improved by 1.94%, and the mAP@0.5 increased by 0.52%. Method 3 resulted in a significant increase in the network's recall for label, scratch, spill, smudge, and mark by 27.35%, 7.46%, 15.24%, 15.28%, and 6.73%, respectively. Although mP@0.5 decreased by 2.02%, mAP@0.5 increased by 2.15%. TABLE 1 illustrates that the enhancement increased the complexity of the network, resulting in a decrease in the speed of inference. However, it still fulfills the criteria for real-time detection. It can be seen from TABLE 2, TABLE 3, and TABLE 4 that the evaluation indicators of the three methods on the defect dataset have all improved, proving the effectiveness of the model improvement.

D. LOSS FUNCTION COMPARISON

From Figure 15, it can be seen that $Loss_{CIoU}$ is replaced by $Loss_{Inner-SIoU}$. The total loss is reduced, and convergence occurs more quickly.

E. ABLATION STUDIES

The above experiments demonstrated the effectiveness of each method. To confirm the compatibility of the proposed methods, ablation experiments were conducted for each method. mAP@0.5 and FPS were used as evaluation metrics. Methods 1, 2, and 3 are noted as A, B, and C, respectively. The experimental findings indicate a high level of compatibility among different methodologies, and the model that integrates

all methods has improved the mAP@0.5 for defects by 3.12% according to TABLE 5.

TABLE 5. Results of the ablation experiment.

A	B	C	mAP@0.5(%)	FPS(Images/s)
	√	√	95.91	76.993
√		√	96.23	63.146
√	√		94.75	54.851
√	√	√	96.51	54.851

F. MODEL COMPARISON

The approach outlined in this study is contrasted with prevalent object detection algorithms like Faster R-CNN, EfficientDet, YOLOv5, YOLOX [42], YOLOv7, etc. YOLOv5, YOLOX, and YOLOv7 utilize the same data augmentation and training hyperparameters as YOLOv8n, along with the same training strategy. Small Object Detection Head (SODH) is added to YOLOv8n for enhancing small object detection.

The results in TABLE 6 indicate that, compared to the YOLO algorithm series, the enhanced YOLOv8n algorithm demonstrates significant superiority in mAP@0.5 and has a lower parameter count. Although there is a slight decrease in real-time performance, it still meets the requirements for industrial detection.

G. EFFECT COMPARISON

To assess the detection performance before and after improving the YOLOv8n model, comparative experiments were
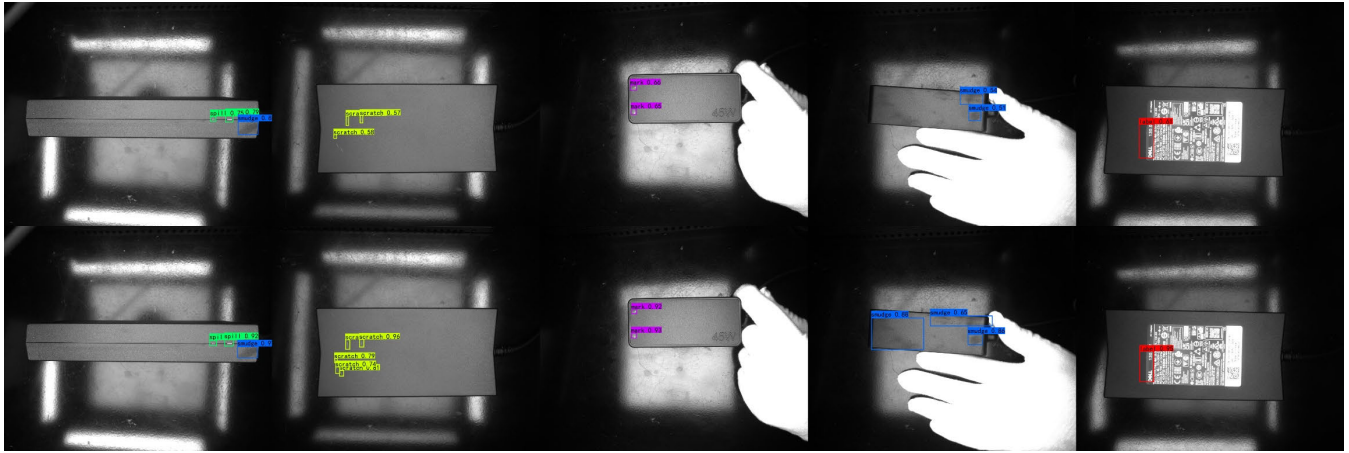
FIGURE 16. Comparison of detection effect before and after improvement (above is before improvement, below is after improvement).

TABLE 6. Comparison of experimental results of models.

Model	mAP@0.5(%)	FPS(Images/s)	Params(M)
Faster R-CNN	50.87	11.847	28.316
EfficientDetD0	61.96	30.121	3.876
YOLOv5n	32.22	102.314	1.781
YOLOv5s	73.17	97.420	7.074
YOLOXn	90.18	65.325	0.912
YOLOXtiny	94.76	75.534	5.056
YOLOv7L	81.47	49.520	37.620
YOLOv8n	93.39	88.875	3.152
SODH	95.72	71.813	3.354
Ours	96.51	54.851	4.255

conducted on samples containing spill, label, scratch, mark, and smudge for comparison. The enhanced model results in improved detection accuracy, as illustrated in Figure 16.

H. BAIDU PADDLE ALUMINUM BLADE SURFACE INDUSTRIAL DEFECT DETECTION EXPERIMENT

The source of this dataset is the open-source dataset "Industrial Defect Object Detection on Aluminum Sheet Surface" provided by Baidu Paddle AI Studio. https://aistudio.baidu.com/datasetdetail/135640. (accessed on 22 February 2024) It includes categories such as pinhole (zhen_kong), abrasion (ca_shang), fold (zhe_zhou), and smudge (zang_wu). The images are sized at 640×480, totaling 412 images, and the dataset has been augmented to include 8,400 images. The dataset partitioning strategy remains consistent. Keep the dataset partitioning strategy consistent. Increase the training hyperparameter batch size by two times, while keeping the other parts consistent. The model comparison experiments are shown in TABLE 7.

I. ALIYUN TIAN CHI ALUMINUM PROFILE SURFACE DEFECT DETECTION EXPERIMENTS

This dataset is derived from the preliminary open-source dataset of the 2018 Guangdong Industrial Intelligent

TABLE 7. Comparison of test results of aluminum sheet surface defect dataset.

Model	mAP@0.5(%)	mR@0.5(%)	mP@0.5(%)
YOLOv8n	98.60	97.52	99.17
Ours	99.01	98.77	99.22

Manufacturing Big Data Innovation Competition - Intelligent Algorithm Competition, specifically focusing on "Aluminum Profile Surface Defect Identification". https://tianchi.aliyun.com/dataset/148297. (accessed on 22 February 2024) Image size is 2560×1920, totaling 1887 pictures. The categories include: dent (aoxian), non-conductivity (budaodian), rubbing flower (cahua), orange peel (jupi), leakage of bottom (loudi), bruise (pengshang), pitting (qikeng), convex powder (tufen), coating cracking (tucengkailie), and smudge spots (zangdian). The dataset has been expanded to 9435 images, and the dataset division strategy and hyperparameter configuration remain consistent with the experiments in Section IV-H. A comparison of the model results is shown in TABLE 8.

TABLE 8. Comparison of test results of aluminum profile surface defect dataset.

Model	mAP@0.5(%)	mR@0.5(%)	mP@0.5(%)
YOLOv8n	69.22	46.04	70.02
Ours	74.83	59.97	91.28

V. CONCLUSION

We have implemented enhancements based on the YOLOv8n algorithm to improve its performance in detecting appearance defects in power adapters.

(1) Introducing the TDFPN feature fusion structure aims to provide relevant features for classification and localization tasks, enhance the network's global sensing ability, and improve the recognition rate of defects. (2) Applying the EMA attention mechanism helps suppress redundant

information, thereby enhancing the detection accuracy of the network with only a slight increase in parameters. (3) Adopting Inner-SIoU instead of CIoU can accelerate model convergence while significantly improving the recall rate of detection. (4) Enhanced algorithms were employed to test datasets containing power adapter defects, aluminum foil surface defects from Baidu Paddle, and aluminum profile surface defects from Aliyun Tian Chi. The results indicated a respective increase in mAP@0.5 of 3.12%, 0.41%, and 5.61% when compared to YOLOv8n. These findings demonstrate the effectiveness of the enhancements made and highlight the model's ability to generalize.

This article enhances the accuracy of the model based on YOLOv8n, which has demonstrated some effectiveness. In industrial scenarios, there is often a lack of defect data, and fully supervised learning necessitates a substantial amount of annotated data for training. This process is time-consuming and labor-intensive. In future research, semi-supervised learning methods can be adopted by utilizing a combination of a large amount of unlabeled data and labeled data to further enhance the practicality and generalization ability of the model.

REFERENCES

[1] Z. Zhang, N. Dong, C. Zhao, and W. Yin, "Metal surface defect detection system based on semiconductor laser and infrared thermal imaging," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2019, pp. 1–5, doi: 10.1109/I2MTC.2019.8826839.

[2] Y. Jiang, H. Wang, G. Tian, S. Chen, and L. Zhang, "Visual detection of AM surface defects based on laser ultrasound technology," in *Proc. Far East NDT New Technol. Appl. Forum (FENDT)*, Jun. 2019, pp. 6–10, doi: 10.1109/FENDT47723.2019.8962808.

[3] W. Ming, C. Cao, G. Zhang, H. Zhang, F. Zhang, Z. Jiang, and J. Yuan, "Review: Application of convolutional neural network in defect detection of 3C products," *IEEE Access*, vol. 9, pp. 135657–135674, 2021, doi: 10.1109/ACCESS.2021.3116131.

[4] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[5] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004, doi: 10.1023/b:visi.0000029664.99615.94.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/bf00994018.

[7] Q. Ling and N. A. M. Isa, "Printed circuit board defect detection methods based on image processing, machine learning and deep learning: A survey," *IEEE Access*, vol. 11, pp. 15921–15944, 2023, doi: 10.1109/ACCESS.2023.3245093.

[8] M. Salem, A. Gomaa, and N. Tsurusaki, "Detection of earthquake-induced building damages using remote sensing data and deep learning: A case study of Mashiki Town, Japan," in *Proc. IGARSS - IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2023, pp. 2350–2353, doi: 10.1109/igarss52108.2023.10282550.

[9] A. Gomaa, M. M. Abdelwahab, and M. Abo-Zahhad, "Real-time algorithm for simultaneous vehicle detection and tracking in aerial view videos," in *Proc. IEEE 61st Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2018, pp. 222–225, doi: 10.1109/MWSCAS.2018.8624022.

[10] A. Gomaa, M. M. Abdelwahab, and M. Abo-Zahhad, "Efficient vehicle detection and tracking strategy in aerial videos by employing morphological operations and feature points motion analysis," *Multimedia Tools Appl.*, vol. 79, nos. 35–36, pp. 26023–26043, Sep. 2020, doi: 10.1007/s11042-020-09242-5.

[11] A. Gomaa, T. Minematsu, M. M. Abdelwahab, M. Abo-Zahhad, and R.-I. Taniguchi, "Faster CNN-based vehicle detection and counting strategy for fixed camera scenes," *Multimedia Tools Appl.*, vol. 81, no. 18, pp. 25443–25471, Jul. 2022, doi: 10.1007/s11042-022-12370-9.

[12] Y. Chang, Z. Tu, W. Xie, B. Luo, S. Zhang, H. Sui, and J. Yuan, "Video anomaly detection with spatio-temporal dissociation," *Pattern Recognit.*, vol. 122, Feb. 2022, Art. no. 108213, doi: 10.1016/j.patcog.2021.108213.

[13] G. Liu, "Surface defect detection methods based on deep learning: A brief review," in *Proc. 2nd Int. Conf. Inf. Technol. Comput. Appl. (ITCA)*, Dec. 2020, pp. 200–203, doi: 10.1109/ITCA52113.2020.00049.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: 10.1109/CVPR.2016.91.

[16] L. Zhang, B. Li, Y. Cui, Y. Lai, and J. Gao, "Research on improved YOLOv8 algorithm for insulator defect detection," *J. Real-Time Image Process.*, vol. 21, no. 1, p. 22, Feb. 2024, doi: 10.1007/s11554-023-01401-9.

[17] H. Zhong, B. Wu, X. Zhang, Y. Hu, T. Zhou, and Y. Zhang, "Steel surface defect detection based on an improved YOLOv5 model," in *Proc. 5th Int. Conf. Intell. Control, Meas. Signal Process. (ICMSP)*, May 2023, pp. 51–55, doi: 10.1109/ICMSP58539.2023.10170867.

[18] G. Zhou, L. Yu, Y. Su, B. Xu, and G. Zhou, "Lightweight PCB defect detection algorithm based on MSD-YOLO," *Cluster Comput.*, pp. 1–15, Oct. 2023, doi: 10.1007/s10586-023-04156-x.

[19] J. Zhuang, Z. Qin, H. Yu, and X. Chen, "Task-specific context decoupling for object detection," 2023, *arXiv:2303.01047*.

[20] D. Xiao, F. T. Xie, Y. Gao, Z. N. Li, and H. F. Xie, "A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLO-v5," *Int. J. Adv. Manuf. Technol.*, vol. 128, nos. 1–2, pp. 937–951, Sep. 2023, doi: 10.1007/s00170-023-11963-4.

[21] M. Tan, R. Pang, and Q. V. Le, "EfficientDet: Scalable and efficient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.

[22] G. Li, S. Zhao, M. Zhou, M. Li, R. Shao, Z. Zhang, and D. Han, "YOLO-RFF: An industrial defect detection method based on expanded field of feeling and feature fusion," *Electronics*, vol. 11, no. 24, p. 4211, Dec. 2022, doi: 10.3390/electronics11244211.

[23] R. Wu, F. Zhou, N. Li, H. Liu, N. Guo, and R. Wang, "Enhanced you only look once X for surface defect detection of strip steel," *Frontiers Neurorobot.*, vol. 16, pp. 1042780–1042791, Nov. 2022.

[24] D. Reis, J. Kupec, J. Hong, and A. Daoudi, "Real-time flying object detection with YOLOv8," 2023, *arXiv:2305.09972*.

[25] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475, doi: 10.1109/CVPR52729.2023.00721.

[26] C.-Y. Wang, A. Bochkovskiy, and H. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 13024–13033, doi: 10.1109/CVPR46437.2021.01283.

[27] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, "Path aggregation network for instance segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8759–8768, doi: 10.1109/CVPR.2018.00913.

[28] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626, doi: 10.1109/ICCV.2017.74.

[29] L. Huang, W. Huang, H. Gong, C. Yu, and Z. You, "PEFNet: Position enhancement faster network for object detection in roadside perception system," *IEEE Access*, vol. 11, pp. 73007–73023, 2023, doi: 10.1109/ACCESS.2023.3292881.

[30] Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "DETRs beat YOLOs on real-time object detection," 2023, *arXiv:2304.08069*.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Long Beach, CA, USA, 2017, pp. 6000–6010.

[32] T. Ye, J. Zhang, Y. Li, X. Zhang, Z. Zhao, and Z. Li, "CT-Net: An efficient network for low-altitude object detection based on convolution and transformer," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022, doi: 10.1109/TIM.2022.3165838.

[33] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang, "Efficient multi-scale attention module with cross-spatial learning," 2023, arXiv:2305.13563.

[34] L. Nannan and W. Liuai, "Research on helmet wearing detection based on improved YOLOv8 algorithm," in Proc. IEEE 11th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), Dec. 2023, pp. 910–917, doi: 10.1109/ITAIC58329.2023.10408800.

[35] C. Yu and Y. Shin, "An efficient YOLO for ship detection in SAR images via channel shuffled reparameterized convolution blocks and dynamic head," ICT Exp., vol. 10, pp. 28–33, Feb. 2024, doi: 10.1016/j.icte.2024.02.007.

[36] Z. Zheng, P. Wang, D. Ren, W. Liu, R. Ye, Q. Hu, and W. Zuo, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," IEEE Trans. Cybern., vol. 52, no. 8, pp. 8574–8586, Aug. 2022, doi: 10.1109/TCYB.2021.3095305.

[37] X. Zhang, Y. Wang, Y. Zhang, Y. Lin, G. Gui, O. Tomoaki, and H. Sari, "Data augmentation aided few-shot learning for specific emitter identification," in Proc. IEEE 96th Veh. Technol. Conf. (VTC-Fall), Sep. 2022, pp. 1–5, doi: 10.1109/VTC2022-Fall57202.2022.10012809.

[38] Z. Gevorgyan, "SIoU loss: More powerful learning for bounding box regression," 2022, arXiv:2205.12740.

[39] G. Han, T. Li, Q. Li, F. Zhao, M. Zhang, R. Wang, Q. Yuan, K. Liu, and L. Qin, "Improved algorithm for insulator and its defect detection based on YOLOX," Sensors, vol. 22, no. 16, p. 6186, Aug. 2022, doi: 10.3390/s22166186.

[40] P. Mo, D. Li, M. Liu, J. Jia, and X. Chen, "A lightweight and partitioned CNN algorithm for multi-landslide detection in remote sensing images," Appl. Sci., vol. 13, no. 15, p. 8583, Jul. 2023, doi: 10.3390/app13158583.

[41] H. Zhang, C. Xu, and S. Zhang, "Inner-IoU: More effective intersection over union loss with auxiliary bounding box," 2023, arXiv:2311.02877.

[42] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, arXiv:2107.08430.

ZHENGWEI QIAN received the bachelor's degree from Bengbu University, Bengbu, China, in 2021. He is currently pursuing the master's degree with the School of Advanced Manufacturing Engineering, Hefei University, Hefei, China. His research interests include atmospheric remote sensing, machine learning, and air pollution monitoring.

KEQIONG CHEN received the B.S. degree in automation, the M.S. degree in control theory and control engineering, and the Ph.D. degree in electrical engineering from Hefei University of Technology, China, in 2008, 2011, and 2017, respectively. She is currently a Lecturer with Hefei University, China. Her research interests include image processing and pattern recognition for control engineering.

MAOFA ZHEN received the master's degree from Hefei University of Technology. He is currently a Professor of engineering with the School of Advanced Manufacturing Engineering, Hefei University. His research interest includes control science and engineering.

JIE CHEN received the bachelor's degree from Hefei University, Hefei, China, in 2022, where he is currently pursuing the master's degree with the School of Advanced Manufacturing Engineering. His research interests include computer vision and machine learning.

YU XIE received the Ph.D. degree from Nanjing University of Science and Technology, Nanjing, China, in 2015. He is currently an Associate Professor with the School of Advanced Manufacturing Engineering, Hefei University, Hefei, China. His current research interests include evolutionary computation, machine learning, and air pollution monitoring.

XUEYOU HU received the Ph.D. degree from Anhui University, Hefei, China, in 2011. He is currently a Professor with the School of Advanced Manufacturing Engineering, Hefei University, Hefei. His current research interests include signal processing and pattern recognition.

• • •