

## Article

# DFA-YOLO: A Novel YOLO Model for Electric Power Operation Violation Recognition

Xiaoliang Qian <sup>1</sup>, Xinyu Ding <sup>1</sup>, Pengfei Wang <sup>1</sup>, Jungang Guo <sup>2</sup>, Hu Chen <sup>1</sup>, Wei Wang <sup>1,\*</sup> and Peixu Xing <sup>1,\*</sup>

<sup>1</sup> College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China; qxlzengli@zzuli.edu.cn (X.Q.); 332301040061@zzuli.edu.cn (X.D.); 332301050081@zzuli.edu.cn (P.W.); chenhu@zzuli.edu.cn (H.C.)

<sup>2</sup> Zhengzhou Fengjia Technology Co., Ltd., Zhengzhou 450009, China; gjg@csqc.cn

\* Correspondence: wangwei-zzuli@zzuli.edu.cn (W.W.); 2003071@zzuli.edu.cn (P.X.)

## Abstract

The You Only Look Once (YOLO) series of models, particularly the recently introduced YOLOv12 model, have demonstrated significant potential in achieving accurate and rapid recognition of electric power operation violations, due to their comprehensive advantages in detection accuracy and real-time inference. However, the current YOLO models still have three limitations: (1) the absence of a dedicated feature extraction for multi-scale objects, resulting in suboptimal detection capabilities for objects with varying sizes; (2) naive integration of spatial and channel attentions, which restricts the enhancement of feature discriminability and consequently impairs the detection performance for challenging objects in complex backgrounds; and (3) weak representation capability in low-level features, leading to insufficient accuracy for small-sized objects. To address these limitations, a novel YOLO model named DFA-YOLO is proposed, a real-time object detection model with YOLOv12n as its baseline, which makes three key contributions. Firstly, a dynamic weighted multi-scale convolution (DWMConv) module is proposed to address the first limitation, which employs lightweight multi-scale convolution followed by learnable weighted fusion to enhance feature representation for multi-scale objects. Secondly, a full-dimensional attention (FDA) module is proposed to address the second limitation, which gives a unified attention computation scheme that effectively integrates attention across height, width, and channel dimensions, thereby improving feature discriminability. Thirdly, a set of auxiliary detection heads (Aux-Heads) are introduced to address the third limitation and inserted into the backbone network to strengthen the training effect of labels on the low-level feature extraction module. The ablation studies on the EPOVR-v1.0 dataset demonstrate the validity of the proposed DWMConv module, FDA module, Aux-Heads, and their synergistic integration. Relative to the baseline model, DFA-YOLO achieves significant improvements in mAP@0.5 and mAP@0.5–0.95, by 3.15% and 4.13%, respectively, meanwhile reducing parameters and GFLOPS by 0.06M and 0.06, respectively, and increasing FPS by 3.52. Comprehensive quantitative comparisons with nine official YOLO models, including YOLOv13n, confirm that DFA-YOLO achieves superior performance in both detection precision and real-time inference, further validating the effectiveness of the DFA-YOLO model.



Academic Editors: Syed Naqvi and Md Mostafizer Rahman

Received: 18 September 2025

Revised: 28 October 2025

Accepted: 7 November 2025

Published: 11 November 2025

**Citation:** Qian, X.; Ding, X.; Wang, P.; Guo, J.; Chen, H.; Wang, W.; Xing, P. DFA-YOLO: A Novel YOLO Model for Electric Power Operation Violation Recognition. *Information* **2025**, *16*, 974. <https://doi.org/10.3390/info16110974>

**Copyright:** © 2025 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

(<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** electric power operation violation recognition (EPOVR); object detection; DFA-YOLO; dynamic weighted multi-scale convolution (DWMConv); full-dimensional attention (FDA); auxiliary detection heads (Aux-Heads)

## 1. Introduction

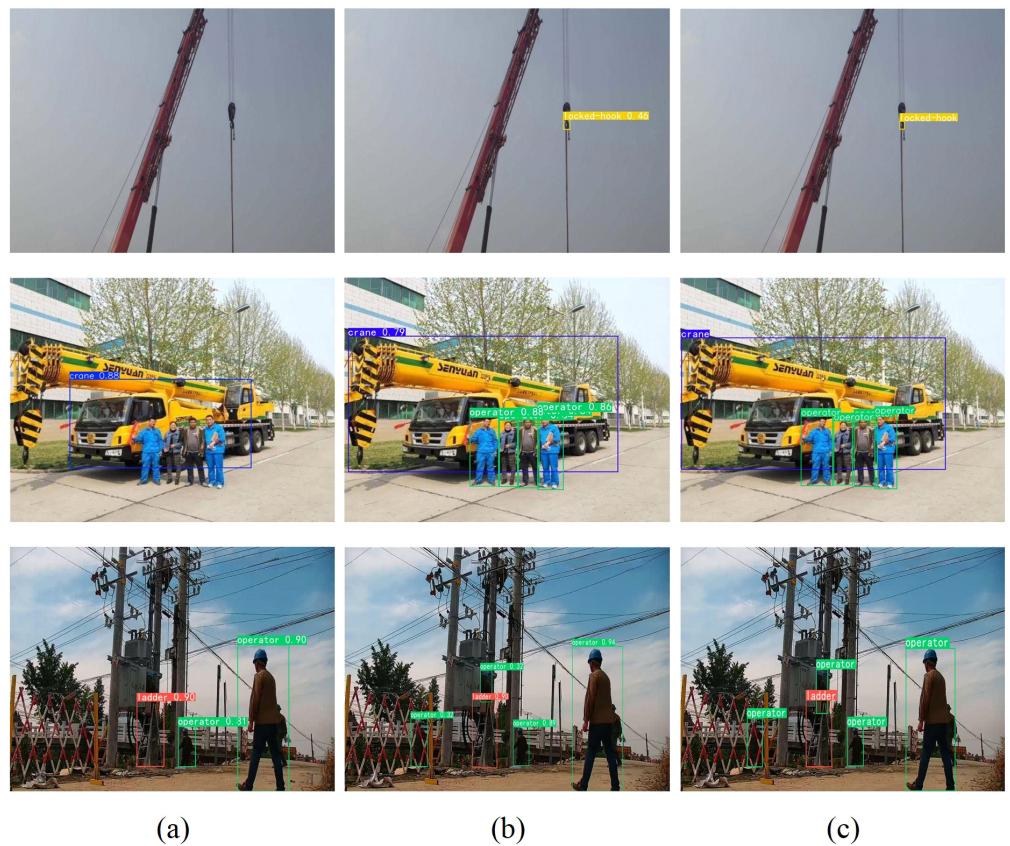
Due to the rapid advancement of power infrastructure and the growing frequency of on-site operations, safety concerns in the electric power operation process have been significantly amplified; therefore, the protection of operators' safety has become a critical challenge in the power industry. Traditional methods for electric power operation violation recognition (EPOVR) predominantly rely on manual inspection through human patrols or the back-end video monitoring, which are not only inefficient and resource-intensive but also susceptible to oversight due to human fatigue or negligence [1]. The advancement of computer vision and artificial intelligence has catalyzed the emergence of image recognition-based EPOVR technologies in recent years. Such technology can automatically identify and warn whether the operators' wearing and code of conduct are compliant, which greatly improves the efficiency of safety supervision and significantly reduces the labor cost [2].

The core of EPOVR lies in the accurate detection of operators, tools, and protective equipment at the operational site, such as safety helmets, safety belts, and work clothes, which form the foundation for subsequent violation analysis [3]. To date, significant advancements have been made in deep learning-based object detection technologies [4–8], where the You Only Look Once (YOLO) object detection models [9–19] have gained widespread adoption across various industrial and commercial applications, due to their excellent detection efficiency. The YOLO models have undergone rapid evolution since the introduction of YOLOv1 [9], with the latest iteration being YOLOv13 [20]. The YOLOv12 model, distinguished by its dual advantages in computational efficiency and detection accuracy, has been selected as the baseline for our object detection model.

Up to now, several improvements of official YOLO models have been proposed to enhance their capabilities. The YOLO-LSM model [21] leveraged shallow and multi-scale feature learning to enhance small object detection accuracy while reducing parameters. The YOLO-SSFA model [22] employed adaptive multi-scale fusion and attention-based noise suppression to enhance object detection accuracy in complex infrared environments. The YOLO-ACR [23] model introduced an adaptive context optimization module, which leveraged contextual information to enhance object detection accuracy. Ji et al. [24] introduced an enhanced YOLOv12 model for detecting small defects in transmission lines, which integrated a two-way weighted feature fusion network to improve the interaction between low-level and high-level semantic features, additionally using spatial and channel attentions to boost the feature representation of defects.

Although the YOLOv12 model demonstrates a commendable balance between object detection precision and real-time inference, it still exhibits three significant limitations in feature extraction. Firstly, the absence of a dedicated feature extraction strategy for multi-scale objects restricts its capability to detect multi-scale key targets in EPOVR. As demonstrated in the first row of Figure 1a, YOLOv12 did not detect the small-sized locked-hook, and it inaccurately localizes the large-sized crane, as shown in the second row of Figure 1a. Secondly, current methods often add an attention module in the final stage of the backbone network to enhance feature representation, and these modules usually combine spatial and channel attentions in a straightforward manner, rather than integrating them into an organic whole, thereby limiting the discriminative capability of features. This limitation is evident in the third row of Figure 1a, where YOLOv12 fails to detect two operators that are significantly affected by background interference. Thirdly, early convolutional neural networks suffered from the vanishing gradient problem during backpropagation. Although the introduction of residual connections has substantially mitigated this issue, the influence of labels still diminishes with increasing distance from the detection head of YOLO models, which results in weaker representation of low-level features in the backbone network, leading to suboptimal detection of small-sized objects.

This is another critical factor contributing to the failure of detecting the locked-hook in the first row of Figure 1a.



**Figure 1.** Detection results of the (a) YOLOv12 model and (b) our DFA-YOLO model and (c) ground truth on the EPOVR-v1.0 dataset [25].

To tackle the aforementioned mentioned challenges, a real-time object detection model called DFA-YOLO is introduced. The primary contributions are summarized as follows.

- (1) To tackle the first issue, a lightweight dynamic weighted multi-scale convolution (DWMConv) module is proposed to substitute a convolution layer in the backbone network. The DWMConv module leverages depthwise separable multi-scale convolutions to extract multi-scale features, and employs learnable adaptive weights to dynamically fuse these features. The fused features have powerful representations for multi-scale objects. As demonstrated in the first two rows of Figure 1b, the DFA-YOLO model successfully detects the small-sized locked-hook and accurately localizes the large-sized crane.
- (2) To address the second issue, a full-dimensional attention (FDA) module is proposed and inserted at the end of backbone network. The FDA module computes and integrates attention across the height, width, and channel dimensions, producing a unified feature representation that captures important information from these above three dimensions. It overcomes the limitation of conventional designs in which spatial attention and channel attention are computed independently, thereby significantly improving feature discriminability. As shown in the third row of Figure 1b, the DFA-YOLO model can detect two operators despite serious background interference.
- (3) To resolve the third issue, a set of auxiliary detection heads (Aux-Heads) is incorporated into the baseline model. The Aux-Heads are added at multiple feature extraction nodes distant from the primary object detection heads, providing additional supervision signals to enhance the training of low-level features. Note that the Aux-Heads are

only involved in training stage, thereby improving the representational capability of low-level features without compromising inference speed. Consequently, adding Aux-Heads can boost the model's capability for detecting small-sized objects, as evidenced by the accurate detection of locked-hook in Figure 1b.

## 2. Baseline Model

The YOLOv12n [26] is selected as the baseline model, which comprises three parts, i.e., the backbone network, neck network, and multi-scale detection heads. The backbone network handles feature extraction tasks, while the neck network integrates low-level and high-level features. The multi-scale detection heads are utilized to predict detection results for small-sized, medium-sized, and large-sized objects. With the exception of A2C2f modules, the remaining modules in YOLOv12 are consistent with those in YOLOv11 [19]. The A2C2f module is built upon the R-ELAN architecture [26], with its core innovation being the incorporation of an area attention (A2) module that employs a transformer mechanism to compute attention. Unlike the traditional transformer [27], which performs global attention computation with a significant computational cost, the A2 module partitions an input feature map into four vertical or horizontal regions and computes attention within the 1/4 region containing the current sampling point. Consequently, this approach reduces the computational load to one-fourth of the original while maintaining a broad visual receptive field, thereby achieving a better balance between detection precision and real-time inference.

## 3. DFA-YOLO Model

### 3.1. Overview

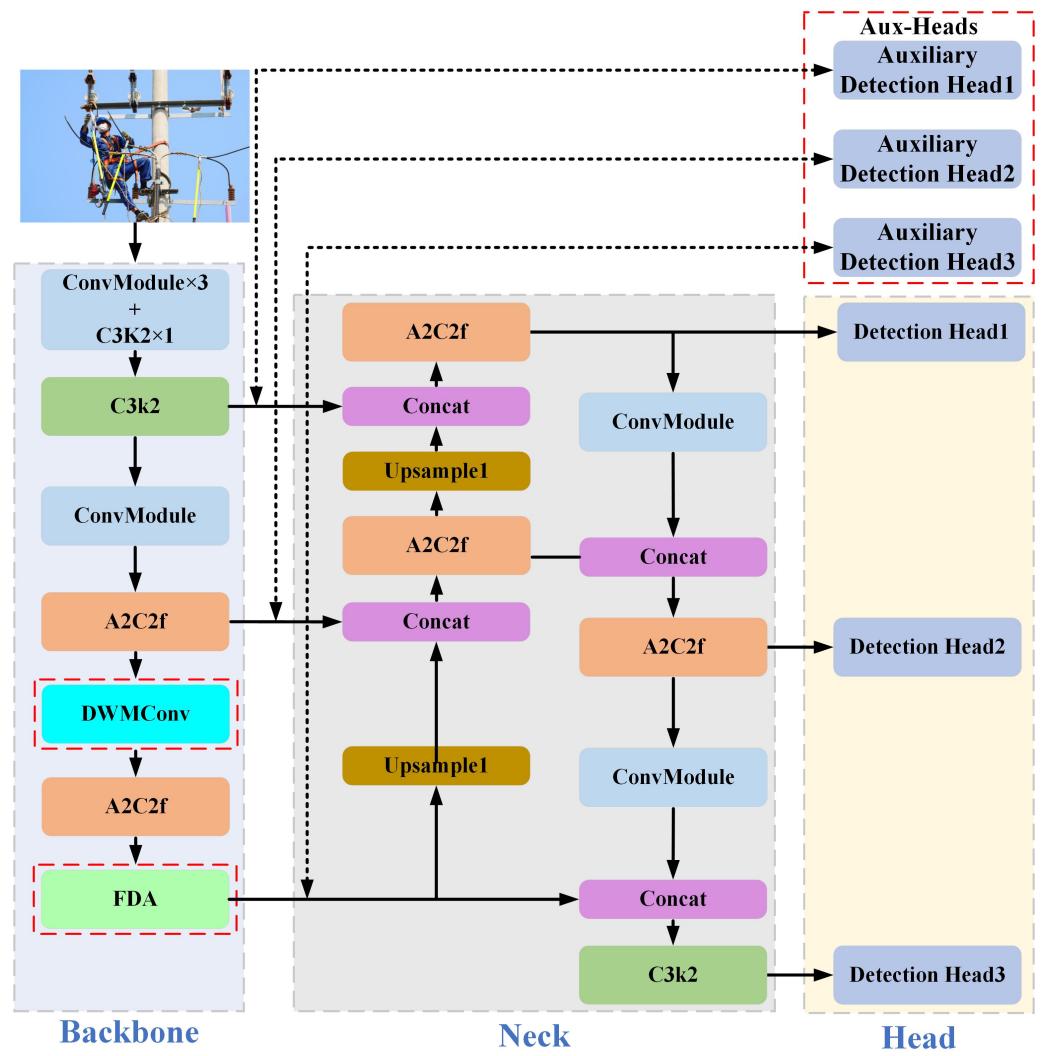
As shown in Figure 2, the proposed DFA-YOLO model uses YOLOv12n as its baseline, incorporating three key innovations. Firstly, the final convmodule of the YOLOv12n backbone is substituted by the lightweight DWMConv module introduced in this paper. This modification not only reduces the parameters and computational complexity but also improves the representation of multi-scale objects. Secondly, the proposed FDA module is added to the end of the backbone network, which effectively combines attention across the height, width, and channel dimensions, thereby improving the discriminative capability of features. Thirdly, three Aux-Heads are added at feature nodes distant from three primary detection heads, which can boost the representation capability of low-level features.

### 3.2. DWMConv Module

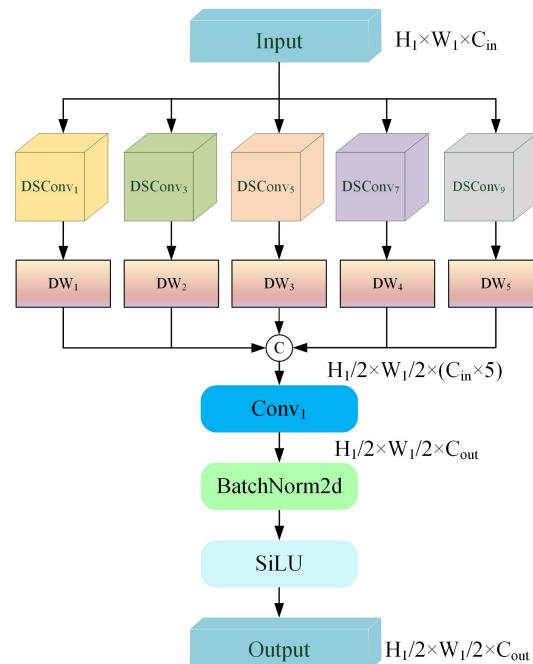
As illustrated in Figure 3, the DWMConv module comprises three distinct components. First of all, the input feature maps are processed through depthwise separable convolutions (DSConv) [28] at five different scales, which are subsequently dynamically weighted. The above process can be formulated as follows:

$$F_i = DW_i \cdot DSConv_j(X), \text{s.t. } \sum_{i=0}^5 1, i \in \{1, 2, 3, 4, 5\}, j \in \{1, 3, 5, 7, 9\} \quad (1)$$

where  $X \in \mathbb{R}^{H_1 \times W_1 \times C_{in}}$  denotes the input feature maps,  $H_1, W_1$ , and  $C_{in}$  represent its height, width, and number of channels, respectively,  $DSConv_j(\cdot)$  denotes a depthwise separable convolution with a kernel size of  $j \times j$  and a stride of two, and  $F_i \in \mathbb{R}^{H_1/2 \times W_1/2 \times C_{in}/2}$  denotes the weighted feature of the  $i$ -th scale.  $DW_i$  denotes the learnable dynamic weight corresponding to the  $i$ -th scale, and the initial weights of five  $DW_i$  are 0.2. The value of  $DW_i$  is dynamically optimized along with the minimization of overall training loss.



**Figure 2.** Architecture of our DFA-YOLO model, where DWMConv, FDA, and auxiliary detection heads are the innovation points of DFA-YOLO model and are highlighted by the red dashed boxes.



**Figure 3.** Architecture of DWMConv module.

Secondly, the weighted features of five scales are fused through following equation:

$$F_m = \text{contact}(F_1, F_2, \dots, F_5) \quad (2)$$

where  $\text{Contact}(\cdot)$  denotes the concatenation operation along channel dimension, and  $F_m \in \mathbb{R}^{H_1/2 \times W_1/2 \times (C_{in} \times 5)}$  denotes the initial multi-scale features.

Finally, the output feature of the DWMConv module, denoted as  $X_o \in \mathbb{R}^{H_1/2 \times W_1/2 \times C_{out}}$ , can be obtained through the following equation:

$$X_o = \text{SiLU}(\text{BN}(\text{Conv}_1(F_m))) \quad (3)$$

where  $\text{Conv}_1(\cdot)$  denotes a  $1 \times 1$  convolution operation with a stride of one,  $\text{BN}(\cdot)$  denotes the batch normalization operation [29],  $\text{SiLU}(\cdot)$  denotes the SiLU activation function [30], and  $C_{out}$  denotes the number of channels of output features.

In summary, the DWMConv module achieves multi-scale feature extraction through lightweight convolutions at multiple kernel scales, and it enables adaptive fusion of these features through learnable dynamic weights. Consequently,  $X_o$  has robust representation capability for objects across different scales.

### 3.3. FDA Module

As illustrated in Figure 4, the execution process of the FDA module can be categorized into three steps. First of all, three soft pooling operations are imposed on the input feature maps independently along the height, width, and channel dimensions. This operation can be represented by the following equations:

$$F_h = \text{SP}_h(Y) \quad (4)$$

$$F_w = \text{SP}_w(Y) \quad (5)$$

$$F_c = \text{SP}_c(Y) \quad (6)$$

where  $Y \in \mathbb{R}^{H_2 \times W_2 \times C_2}$  represents the input feature map,  $H_2$ ,  $W_2$ , and  $C_2$  denote its height, width, and number of channels,  $F_h \in \mathbb{R}^{H_2 \times 1 \times 1}$ ,  $F_w \in \mathbb{R}^{W_2 \times 1 \times 1}$ , and  $F_c \in \mathbb{R}^{C_2 \times 1 \times 1}$  separately denote the pooled features in the height, width, and channel dimensions, and  $\text{SP}_h(\cdot)$ ,  $\text{SP}_w(\cdot)$ , and  $\text{SP}_c(\cdot)$  separately denote soft pooling in height, width, and channel dimensions. The core idea of soft pooling is illustrated in Figure 5. Notably, compared to max pooling, soft pooling exhibits smoother behavior, while in contrast to average pooling, it assigns higher weights to features with stronger responses, thereby more effectively emphasizing foreground objects.

Secondly, the pooled features conduct convolution and activation operations in turn, which can be represented by the following equations:

$$W_h = \text{Sigmoid}(\text{Conv}_1(F_h)) \quad (7)$$

$$W_w = \text{Sigmoid}(\text{Conv}_1(F_w)) \quad (8)$$

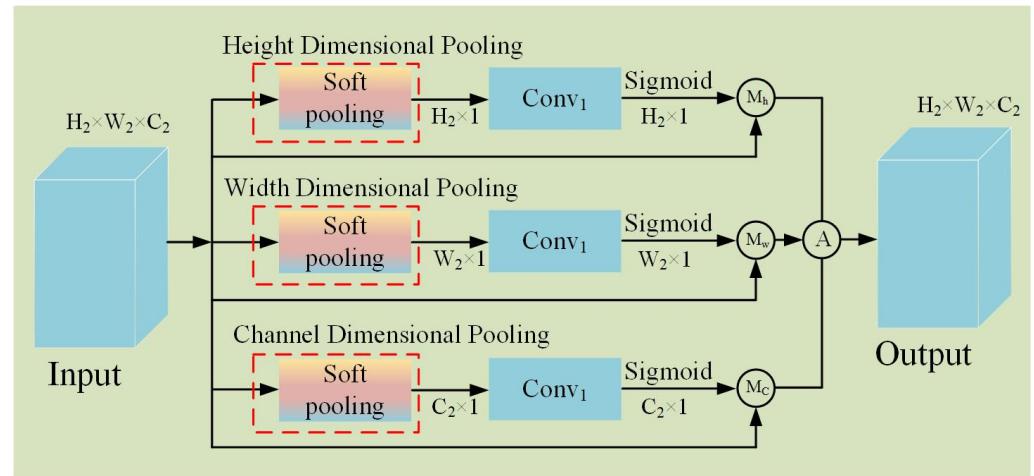
$$W_c = \text{Sigmoid}(\text{Conv}_1(F_c)) \quad (9)$$

where  $\text{Sigmoid}(\cdot)$  denotes the Sigmoid activation function [31], and  $W_h \in \mathbb{R}^{H_2 \times 1 \times 1}$ ,  $W_w \in \mathbb{R}^{W_2 \times 1 \times 1}$ , and  $W_c \in \mathbb{R}^{C_2 \times 1 \times 1}$  separately denote the attention weights of  $Y$  in height, width, and channel dimensions.

Finally, the attention weights are integrated with the input features to generate the output features of the FDA module, denoted as  $Y_o \in \mathbb{R}^{H_2 \times W_2 \times C_2}$ , which can be represented by the following equation:

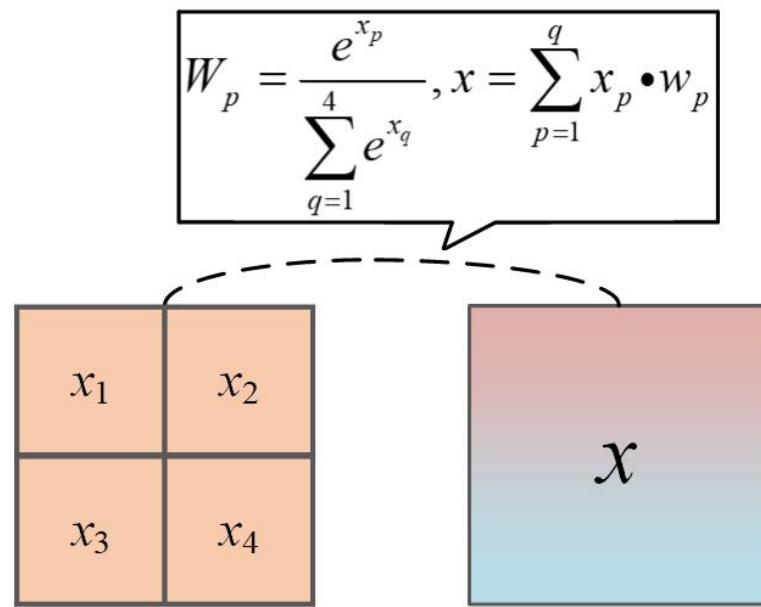
$$Y_o = Avg(M_h(Y, W_h) + M_w(Y, W_w) + M_c(Y, W_c)) \quad (10)$$

where  $Avg(\cdot)$  denotes the averaging operation, and  $M_h(\cdot)$ ,  $M_w(\cdot)$ , and  $M_c(\cdot)$  denote the element-wise multiplication operations in the height, width, and channel dimensions, respectively. For instance,  $M_c(Y, W_c)$  performs a point-wise multiplication between each element in  $W_c$  and the corresponding channel feature map in  $Y$ .



**Figure 4.** Architecture of FDA module.

In summary, the FDA module gives a unified attention scheme rather than computing spatial and channel attention independently, which achieves a comprehensive full-dimensional attention computation, thereby enhancing the discriminative capability of feature representations.



**Figure 5.** Illustration of soft pooling.

### 3.4. Auxiliary Detection Heads and Loss Function

As shown in Figure 2, three Aux-Heads are integrated into the distinct nodes of the backbone network to augment the feature representation capability of convolution modules

that are distant from the primary detection heads. Three junctions between the backbone network and the neck network are selected as the insertion nodes for these Aux-Heads in this paper, and the motivation for which is as follows. Features at these three nodes are subsequently fed into the neck network for feature fusion, and obviously, their importance is higher than that of the features at other nodes in the backbone network. Consequently, the insertion of Aux-Heads at these nodes can enhance the training effect of features at these critical junctions.

Detection head1, detection head2, and detection head3 are separately used to detect small-sized, medium-sized, and large-sized targets. Analogously, Aux-Head1, Aux-Head2, and Aux-Head3, positioned from lower to higher levels in the backbone network, are separately used for the auxiliary training of small-sized, medium-sized, and large-sized object detection heads.

Aux-Heads are only utilized during the training stage and are omitted in the inference stage. The labels and loss functions of Aux-Head1, Aux-Head2, and Aux-Head3 are the same as those of detection head1, detection head2, and detection head3, respectively. Therefore, the overall loss of the DFA-YOLO model, denoted as  $L$ , can be formulated as follows:

$$L = L_{cls} + \alpha \cdot L_{cls_{aux}} + L_{reg} + \beta \cdot L_{reg_{aux}} \quad (11)$$

where  $L_{cls}$  and  $L_{reg}$  represent the losses for classification and regression loss of three corresponding primary detection heads,  $L_{cls_{aux}}$  and  $L_{reg_{aux}}$  separately denote the classification and regression losses of three Aux-Heads, and the hyper parameters  $\alpha$  and  $\beta$  are used for controlling the relative weights of  $L_{cls_{aux}}$  and  $L_{reg_{aux}}$ , respectively.

## 4. Experimental Results and Analysis

### 4.1. Dataset

The performance of the DFA-YOLO model is evaluated on the EPOVR-v1.0 dataset [25]. This dataset comprises 1200 images, with resolutions spanning from  $1920 \times 1080$  to  $3840 \times 2160$ , and includes nine object categories: crane, throwing, operator, supervision, unlocked-hook, no-throwing, ladder, locked-hook, and LHRM [25]. The training, validation, and testing subsets separately include 840, 120, and 240 images. It is worth noting that the EPOVR-v1.0 dataset is derived from our team's previous practical project, and all images were taken at actual power operation sites.

To further evaluate the performance of the DFA-YOLO model, we also made a comparison on the Alibaba Tianchi dataset [32]. This dataset comprises 2352 images, with resolutions spanning from  $1968 \times 4144$  to  $3472 \times 4624$ , and includes four object categories: badge, offground person, ground person, and safebelt. The training, validation, and testing subsets separately include 1645, 235, and 472 images. Obviously, the object categories of this database do not overlap with those of the EPOVR-v1.0 dataset, which can better verify the generalization ability of the DFA-YOLO model.

### 4.2. Experimental Details

The input images are resized to  $640 \times 640$ . The number of training epochs, batch size, and learning rate are 600, 32, and 0.01, respectively. The optimizer selects stochastic gradient descent (SGD), with a momentum and weight decay of 0.937 and 0.0005, respectively. The intersection-over-union (IoU) threshold in non-maximum suppression (NMS) is 0.7. The software configuration is Ubuntu 22.04, Python 3.11, PyTorch 2.2.2, and CUDA 12.4. The hardware configuration comprises an Intel(R) Xeon(R) E5-2650 v4 CPU and single NVIDIA GeForce RTX 2080 Ti GPU (11 GB memory). Notably, the above setting is consistent with the default configuration of YOLOv12 [26] except for the hardware configuration.

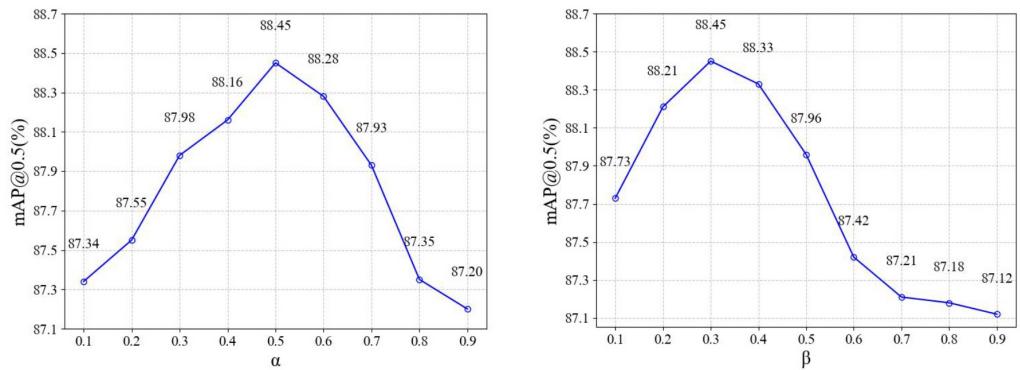
#### 4.3. Evaluation Metrics

The evaluation metrics include mAP@0.5, mAP@0.5–0.95 [33], parameters, GFLOPS, and FPSs. The mAP@0.5 denotes the mean average precision with an IoU threshold of 0.5, while mAP@0.5–0.95 denotes the average of mean average precision values across IoU thresholds ranging from 0.5 to 0.95. Parameters indicate the total number of parameters for the inference model. GFLOPs denote the number of billions of floating-point operations per second, reflecting the computation complexity of the inference model. FPSs measure the number of images inferenced per second, providing an intuitive evaluation of inference speed.

#### 4.4. Parameter Analysis

##### 4.4.1. Parameter Analysis for $\alpha$ and $\beta$

As shown in Equation (7), the DFA-YOLO model has two hyper parameters, i.e.,  $\alpha$  and  $\beta$ . This section conducts a quantitative parameter analysis of  $\alpha$  and  $\beta$  for evaluating their impact on the DFA-YOLO model. As shown in Figure 6, the optimal values are set to be  $\alpha = 0.5$  and  $\beta = 0.3$ , as these settings yield the maximum mAP@0.5 value.



**Figure 6.** Quantitative parameter analysis of  $\alpha$  and  $\beta$  on the EPOVR-v1.0 dataset.

##### 4.4.2. Parameter Analysis for DWMConv Module

This section provides a quantitative parameter analysis of convolution kernel sizes used in the DWMConv module. Firstly, starting from 1 and 3 with an interval of 2, two groups of kernel sizes were selected, i.e., (1, 3, 5, 7, 9) and (3, 5, 7, 9, 11). Then, starting from 1 with an interval of 4, a third group of kernel sizes (1, 5, 9, 13, 17) was chosen. As shown in Table 1, all five metrics achieve the best performance when the kernel sizes are set to (1, 3, 5, 7, 9).

**Table 1.** Quantitative parameter analysis of kernel sizes in DWMConv module.

Model	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
Baseline	85.30	60.10	2.51	5.83	94.34
Baseline+DWMConv (1, 3, 5, 7, 9)	86.75	61.94	2.40	5.74	98.53
Baseline+DWMConv (3, 5, 7, 9, 11)	86.14	61.37	2.42	5.75	98.11
Baseline+DWMConv (1, 5, 9, 13, 17)	85.67	60.66	2.46	5.92	92.13

#### 4.5. Ablation Study

This section presents the ablation experiments of the DWMConv module, FDA module, and Aux-Heads to verify their validity. The corresponding experimental results are shown in Table 2.

Table 2 demonstrates that the introduction of the DWMConv module alone leads to a 1.45% improvement in mAP@0.5 and a 1.84% gain in mAP@0.5–0.95, while reducing

the parameter quantity by 0.11M and GFLOPs by 0.09 and increasing FPSs by 4.19. This indicates the DWMConv module not only enhances object detection precision but also enhances inference speed, while the FDA module is proposed independently, mAP@0.5 and mAP@0.5–0.95 are separately improved by 1.23% and 1.79%, with a slight increase in parameters and GFLOPs, and a minor decrease in FPSs, confirming the effectiveness of the FDA module. The standalone introduction of Aux-Heads separately results in a 1.14% and 1.73% improvement in mAP@0.5 and mAP@0.5–0.95, without any change in parameters, GFLOPs, and FPSs, thereby validating the effectiveness of Aux-Heads. The pairwise combinations of the DWMConv module, FDA module, and Aux-Heads further enhance mAP@0.5 and mAP@0.5–0.95, with parameters and GFLOPs either decreasing or remaining stable, and FPSs either increasing or staying constant, which verifies the effectiveness of these pairwise integrations. The optimal experimental results are achieved when all three components are separately integrated, with mAP@0.5 and mAP@0.5–0.95 increasing by 3.15% and 4.13%, relative to the baseline model, while parameters and GFLOPs decrease by 0.06 M and 0.06, and FPSs increases by 3.52. These results demonstrate that the DFA-YOLO model achieves simultaneous improvements in both accuracy and real-time object detection over the baseline model, thereby verifying the effectiveness of the integration of all three components.

**Table 2.** Ablation experimental results of DFA-YOLO model.

DWMConv	FDA	Aux-Heads	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
✗	✗	✗	85.30	60.10	2.51	5.83	94.34
✓	✗	✗	86.75	61.94	2.40	5.74	98.53
✗	✓	✗	86.53	61.89	2.55	5.84	93.86
✗	✗	✓	86.44	61.83	2.51	5.83	94.34
✓	✓	✗	87.75	62.72	2.45	5.77	97.86
✓	✗	✓	87.34	62.47	2.40	5.74	98.53
✗	✓	✓	87.04	62.36	2.55	5.84	93.86
✓	✓	✓	88.45	64.23	2.45	5.77	97.86

#### 4.6. Comprehensive Comparison Between FDA Module and Other Popular Attention Modules

As shown in Table 3, compared with other popular attention modules, the FDA module achieves the highest scores in both mAP@0.5 and mAP@0.5–0.95 on the EPOVR-v1.0 dataset, while the differences among the other three metrics are relatively small. This demonstrates that the overall performance of the proposed FDA module outperforms the other three attention modules.

**Table 3.** Comprehensive quantitative comparison results of DFA-YOLO model and other official YOLO models. Bold fonts denote the best results.

Model	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
Baseline	85.30	60.10	2.51	5.83	94.34
Baseline+ECA [34]	85.75	60.73	2.53	5.83	94.34
Baseline+CBAM [35]	86.05	61.18	2.58	5.84	93.86
Baseline+SE [36]	86.11	61.34	2.52	5.83	94.34
Baseline+FDA	86.53	61.86	2.55	5.84	93.86

#### 4.7. Comprehensive Comparison Between DWMConv Module and SAHI Solution

As shown in Table 4, whether compared separately or with the addition of the FDA module and Aux-Heads, the detection accuracy of the DWMConv module is comparable to that of the slicing-aided hyper inference (SAHI) solution [37], but it significantly outperforms the SAHI solution in terms of computational cost and inference speed.

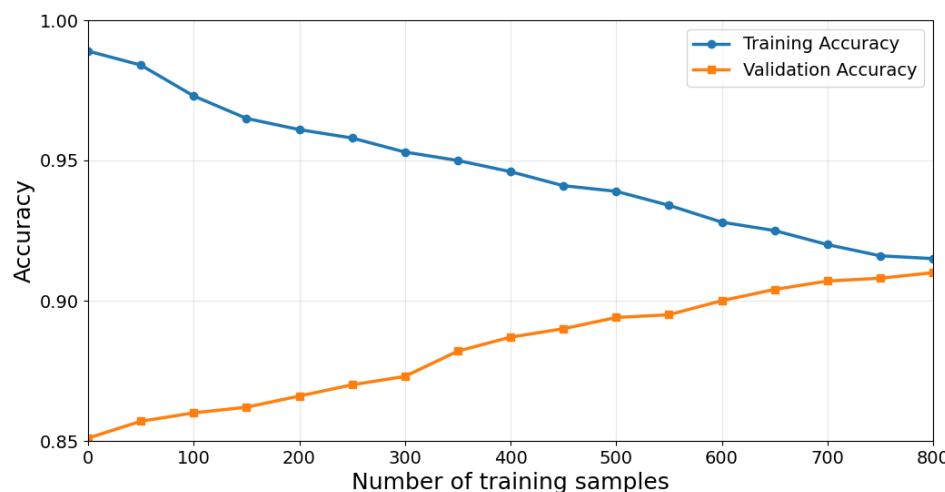
**Table 4.** Quantitative comparison between DWMConv module and SAHI solution.

Model	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
Baseline+SAHI	85.30	60.10	2.51	5.83	94.34
Baseline+DWMConv	86.75	61.94	2.40	5.74	98.53
Baseline+FDA+Aux-Heads+SAHI	86.53	61.89	2.55	5.84	93.86
Baseline+FDA+Aux-Heads+DWMConv	88.45	64.23	2.45	5.77	97.68

#### 4.8. Evaluation of Training Performance

As shown in Table 5, compared to the baseline model, the training time increased by 1.18 h and the peak memory increased by 0.75 G after adding the Aux-Heads. Compared to the baseline+Aux-Heads, the training time of DFA-YOLO decreased by 0.16 h and the peak memory decreased by 0.26 G, because using the DWMConv module instead of the traditional convolution module can reduce the parameters and GFLOPs. Note that the Aux-Heads do not participate in the inference stage, so they have no impact on the inference speed of the DFA-YOLO model.

As shown in Figure 7, the training accuracy and validation accuracy are both high, and they tend to be consistent as the number of training samples gradually increases, indicating that the DFA-YOLO model has good training performance, i.e., there are no overfitting and underfitting problems.

**Figure 7.** Learning curve of our DFA-YOLO model on the EPOVR-v1.0 dataset.**Table 5.** Quantitative analysis in terms of training time and peak memory.

Model	Training Time (H)	Peak Memory (G)
Baseline (YOLOv12n)	4.87	7.44
Baseline+Aux-Heads	6.05	8.19
DFA-YOLO	5.89	7.93

#### 4.9. Comprehensive Comparison with Other YOLO Models

In order to rigorously evaluate the performance of the DFA-YOLO model, our model is quantitatively compared with other official YOLO models. All of models share the same configuration for fair comparison. The results are shown in Tables 6 and 7 as the mean  $\pm$  std dev over  $\geq 3$  runs.

**Table 6.** Comprehensive quantitative comparison results of DFA-YOLO model and other official YOLO models on the EPOVR-v1.0 dataset. The values of mAP@0.5, mAP@0.5–0.95, and FPS are shown in the form of ‘Mean $\pm$ STD’. \* indicates that the Aux-Heads are used for ensemble prediction in the inference stage, and <sup>†</sup> denotes that the resolution of input images is changed to  $1280 \times 1280$ .

Model	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
YOLOv5n [13]	$83.19 \pm 0.13$	$57.32 \pm 0.14$	2.52	7.08	$88.73 \pm 2.1$
YOLOv6n [14]	$82.48 \pm 0.12$	$56.94 \pm 0.13$	4.18	12.12	$83.37 \pm 1.8$
YOLOv7t [15]	$84.34 \pm 0.07$	$58.48 \pm 0.16$	6.03	13.13	$79.68 \pm 1.7$
YOLOv8n [16]	$85.00 \pm 0.06$	$59.47 \pm 0.15$	3.04	8.07	$86.25 \pm 1.5$
YOLOv9t [17]	$82.33 \pm 0.12$	$57.66 \pm 0.13$	2.1	7.68	$87.87 \pm 1.3$
YOLOv10n [18]	$81.87 \pm 0.11$	$56.83 \pm 0.13$	2.28	6.53	$89.74 \pm 2.1$
YOLOv11n [19]	$84.32 \pm 0.13$	$59.03 \pm 0.15$	2.58	6.31	$90.33 \pm 1.4$
YOLOv12n [23]	$85.30 \pm 0.09$	$60.10 \pm 0.12$	2.51	5.83	$94.34 \pm 1.3$
YOLOv13n [20]	$84.08 \pm 0.08$	$59.88 \pm 0.11$	2.45	6.22	$91.34 \pm 1.6$
DFA-YOLO *	$87.59 \pm 0.04$	$62.53 \pm 0.05$	2.58	6.38	$90.15 \pm 1.2$
DFA-YOLO <sup>†</sup>	$89.87 \pm 0.05$	$66.38 \pm 0.06$	2.45	23.10	$24.21 \pm 1.3$
DFA-YOLO (Ours)	$88.45 \pm 0.03$	$64.23 \pm 0.04$	2.45	5.77	$97.86 \pm 1.1$

As shown in Tables 6 and 7, the DFA-YOLO model attains mAP@0.5 and mAP@0.5–0.95 values of 88.45% (95.37%) and 64.23% (77.25%) on the EPOVR-v1.0 (Alibaba Tianchi) dataset, respectively, surpassing all other YOLO models and confirming its superior detection accuracy. With a parameter count of 2.45 M, the DFA-YOLO model slightly exceeds those of YOLOv9t and YOLOv10n but remains lower than the other seven models. Additionally, the DFA-YOLO model exhibits optimal computational efficiency and inference speed, with GFLOPs and FPS values of 5.77 (5.77) and 97.86 (98.47) on the EPOVR-v1.0 (Alibaba Tianchi) dataset, respectively, further validating its advantages in real-time performance.

**Table 7.** Comprehensive quantitative comparison results of DFA-YOLO model and other official YOLO models on the Alibaba Tianchi dataset.\* indicates that the Aux-Heads are used for ensemble prediction in the inference stage, and <sup>†</sup> denotes that the resolution of input images is changed to  $1280 \times 1280$ .

Model	mAP@0.5 (%)	mAP@0.5–0.95 (%)	Parameters (M)	GFLOPS	FPS
YOLOv5n	$90.92 \pm 0.12$	$66.44 \pm 0.13$	2.52	7.08	$89.23 \pm 2.2$
YOLOv6n	$90.31 \pm 0.11$	$65.83 \pm 0.12$	4.18	12.12	$84.16 \pm 1.9$
YOLOv7t	$92.13 \pm 0.06$	$69.51 \pm 0.15$	6.03	13.13	$80.24 \pm 1.6$
YOLOv8n	$91.38 \pm 0.05$	$69.18 \pm 0.14$	3.04	8.07	$87.12 \pm 1.4$
YOLOv9t	$90.48 \pm 0.11$	$65.73 \pm 0.12$	2.1	7.68	$88.47 \pm 1.2$
YOLOv10n	$89.22 \pm 0.10$	$64.18 \pm 0.12$	2.28	6.53	$90.24 \pm 2.0$
YOLOv11n	$90.38 \pm 0.12$	$65.16 \pm 0.14$	2.58	6.31	$91.28 \pm 1.5$
YOLOv12n	$94.01 \pm 0.08$	$74.96 \pm 0.11$	2.51	5.83	$95.21 \pm 1.5$
YOLOv13n	$93.89 \pm 0.07$	$74.78 \pm 0.10$	2.45	6.22	$92.68 \pm 1.8$
DFA-YOLO *	$94.78 \pm 0.04$	$73.86 \pm 0.05$	2.58	6.38	$91.14 \pm 1.2$
DFA-YOLO <sup>†</sup>	$95.89 \pm 0.05$	$78.37 \pm 0.06$	2.45	23.10	$24.21 \pm 1.3$
DFA-YOLO (Ours)	$95.37 \pm 0.03$	$77.25 \pm 0.04$	2.45	5.77	$98.47 \pm 1.1$

In addition, all five metrics of DFA-YOLO\* are inferior to those of DFA-YOLO, indicating that involving the Aux-Heads in the inference stage not only increases the number of parameters and computational cost but also reduces detection accuracy. In fact, the Aux-Heads perform inference based on shallow features; therefore, their predictions are less accurate than those of the main detection heads. Nonetheless, the detection accuracy of DFA-YOLO<sup>†</sup> is slightly higher than that of DFA-YOLO, which demonstrates that the DFA-YOLO is still effective in  $1280 \times 1280$  resolution. However, the computational cost and inference speed of DFA-YOLO<sup>†</sup> are significantly inferior than that of DFA-YOLO. Therefore, the overall performance of DFA-YOLO achieves the optimal balance when the size of input image is  $640 \times 640$ .

In conclusion, the DFA-YOLO model outperforms other official YOLO models in both detection precision and real-time inference, demonstrating its practical value in EPOVR and other application scenarios.

#### 4.10. Subjective Evaluation

To provide a more intuitive assessment of the DFA-YOLO model's performance in power violation detection, this section presents a visual comparison between the improved model and the baseline model (YOLOv12n) using several representative images from power operation sites.

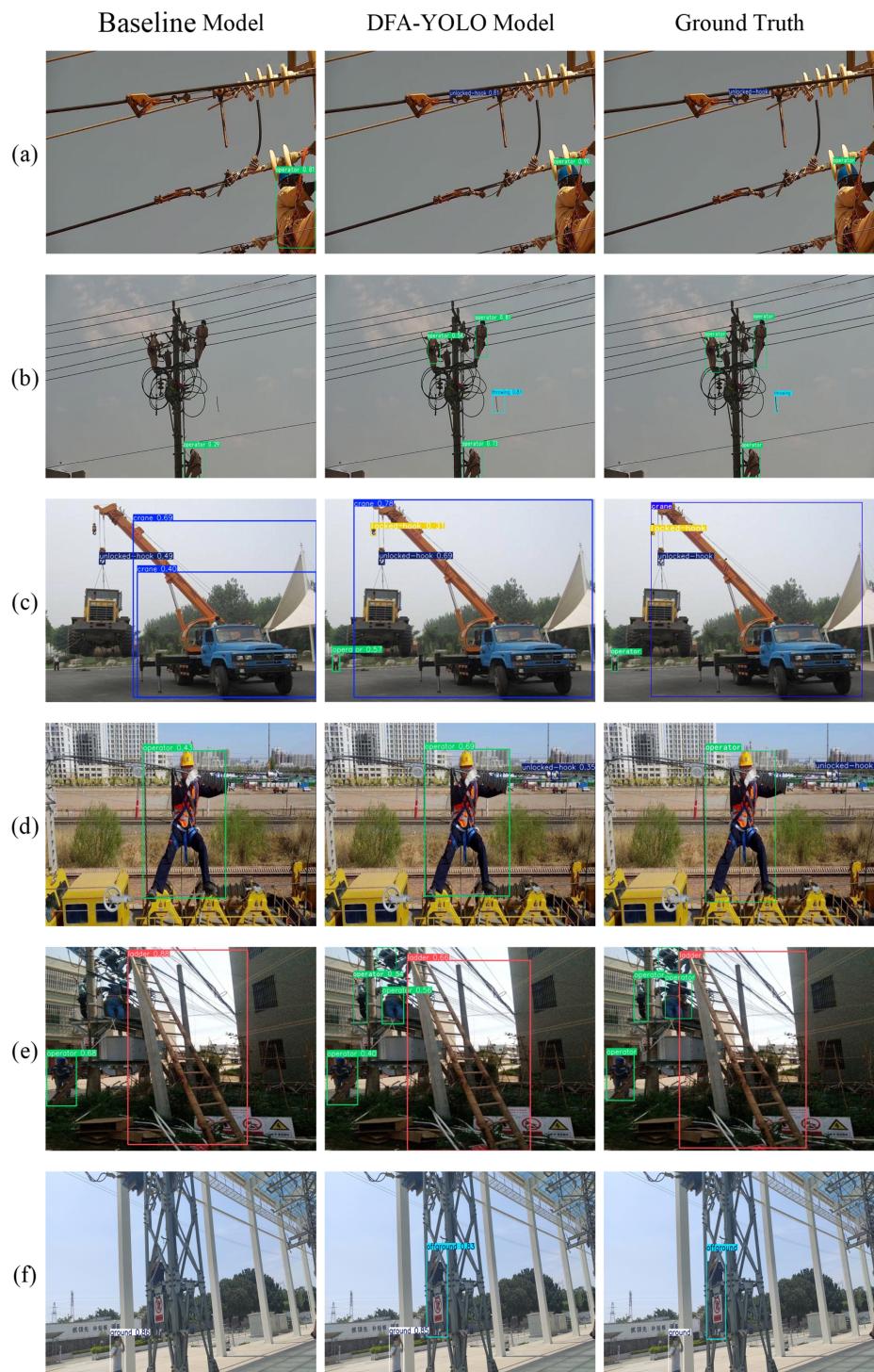
To provide a more intuitive evaluation of the DFA-YOLO model in EPOVR, our model is subjectively compared with the baseline model using several representative testing images from the EPOVR-v1.0 dataset. As indicated in Figure 8a–c, the baseline model cannot detect small-sized high-altitude throwing objects, distant operators, locked-hooks, and unlocked-hooks and can only partially detect larger-sized cranes. In contrast, the DFA-YOLO model can successfully detect the aforementioned objects, which indicates that the DWMConv module markedly enhances the DFA-YOLO model's capability to detect multi-scale objects. Moreover, it also validates the effectiveness of Aux-Heads in improving the detection of small-sized objects. As shown in Figure 8d,e, the baseline model fails to detect unlocked-hooks and partially covered operators that are similar in color to the background, while the DFA-YOLO model accurately detects these objects. This shows that the FDA module effectively improves the discriminative capability of features, which permits the DFA-YOLO model to achieve robust detection performance even for challenging objects severely affected by background interference. As shown in Figure 8f, the baseline model fails to detect the partially occluded off-ground person, whereas the DFA-YOLO model successfully identifies them.

Furthermore, the detection result in Figure 8b demonstrates that the DFA-YOLO model can still accurately detect objects when the light is slightly dim. The result in Figure 8f also indicates that the DFA-YOLO model remains capable of detecting objects when the level of occlusion is not severe.

#### 4.11. Analysis of Failure Results

Although the overall performance of the DFA-YOLO model is satisfactory, failure results are still inevitable in certain cases. Some typical failure results and their reasons can be summarized as follows:

1. Failure results induced by the severe occlusion of object. As shown in Figure 9a, an off-ground person is not detected because the person is heavily occluded by an electric pole.
2. Failure results induced by terrible illumination. As shown in Figure 9b, due to the blockage of light by leaves, the lighting conditions are extremely poor, resulting in the failed detection of an off-ground person.



**Figure 8.** Subjective comparison between DFA-YOLO model and baseline model. (a–f) are selected from the EPOVR-v1.0 dataset and Alibaba Tianchi dataset, respectively.



**Figure 9.** Quantitative Illustration of failure results. (a,b) are selected from the Alibaba Tianchi dataset.

## 5. Conclusions

The DFA-YOLO model, of which the baseline is YOLOv12n, is proposed in this paper, and its contributions are as follows. Firstly, a DWMConv module is proposed, which dynamically fuses lightweight multi-scale convolutional features using learnable weights. The fused features exhibit enhanced representational capability for multi-scale objects, significantly improving the YOLO model's capability to detect the objects with varying sizes. Secondly, a FDA module that integrates attentions across height, width, and channel dimensions is proposed, which effectively enhances feature discriminability, enabling the YOLO model to achieve superior detection performance for challenging objects that are heavily influenced by background interference. Thirdly, a set of Aux-Heads is incorporated into the backbone network, positioned far from the primary detection heads. These Aux-Heads strengthen the training effect of labels on the low-level feature extraction module, thereby effectively enhancing the detection capability of the YOLO model for small-sized objects. Ablation experiments conducted on the EPOVR-v1.0 dataset verify the validity of the DWMConv module, FDA module, Aux-Heads, and their combinations. Furthermore, comprehensive quantitative comparisons with nine official YOLO models, ranging from YOLOv5 to YOLOv13, demonstrate that the DFA-YOLO model achieves the best performance in both overall detection precision and real-time inference.

To date, the DFA-YOLO model has been successfully deployed on the PC platform. The next step is to deploy it on edge devices for commercial applications. Considering factors such as code compatibility, computational capability, and price, we are currently evaluating suitable edge devices. In addition, we plan to incorporate one-dimensional convolutions along the temporal dimension into the backbone network of the DFA-YOLO model to further enhance detection accuracy and stability.

**Author Contributions:** Conceptualization, X.Q.; formal analysis, X.Q.; funding acquisition, W.W.; methodology, X.Q. and X.D.; project administration, P.X.; resources, W.W.; software, X.D., P.W. and J.G.; supervision, W.W.; validation, W.W. and H.C.; writing—original draft, X.D.; writing—review and editing, X.Q. and P.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Natural Science Foundation of Henan (Grant No. 252300421063), the Key Research Project of Henan Province Universities (Grant No. 24ZX005), and the National Natural Science Foundation of China (Grant No. 62076223).

**Institutional Review Board Statement:** The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Ethics Committee of Zhengzhou University of Light Industry (41580459-6) on 18 September 2025.

**Informed Consent Statement:** Informed consent for participation was obtained from all subjects involved in the study.

**Data Availability Statement:** The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding authors.

**Conflicts of Interest:** Author Jungang Guo was employed by Zhengzhou Fengjia Technology Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

YOLO	You Only Look Once
DWMConv	Dynamic weighted multi-scale convolution
FDA	Full-dimensional attention
Aux-Heads	Auxiliary detection heads
EPOVR	Electric power operation violation recognition
A2C2f	Area attention-enhanced cross-stage fusion
SPPF	Spatial pyramid pooling-fast
DSConv	Depthwise separable convolution
SGD	Stochastic gradient descent
NMS	Non-maximum suppression
Params	Parameter count
FLOPs	Floating-point operations
FPSs	Frames per second
mAP	mean Average Precision

## References

1. Meng, L.; He, D.; Ban, G.; Xi, G.; Li, A.; Zhu, X. Active Hard Sample Learning for Violation Action Recognition in Power Grid Operation. *Information* **2025**, *16*, 67.
2. Vukicevic, A.M.; Petrovic, M.; Milosevic, P.; Peulic, A.; Jovanovic, K.; Novakovic, A. A systematic review of computer vision-based personal protective equipment compliance in industry practice: Advancements, challenges and future directions. *Artif. Intell. Rev.* **2024**, *57*, 319.
3. Ji, X.; Gong, F.; Yuan, X.; Wang, N. A high-performance framework for personal protective equipment detection on the offshore drilling platform. *Complex Intell. Syst.* **2023**, *9*, 5637–5652.
4. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
5. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1137–1149.
6. Qian, X.; Wu, B.; Cheng, G.; Yao, X.; Wang, W.; Han, J. Building a Bridge of Bounding Box Regression Between Oriented and Horizontal Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5605209.
7. Qian, X.; Jian, Q.; Wang, W.; Yao, X.; Cheng, G. Incorporating Multiscale Context and Task-consistent Focal Loss into Oriented Object Detection. *IEEE Trans. Geosci. Remote Sens.* **2025**, *11*, 5628411.
8. Qian, X.; Zhang, B.; He, Z.; Wang, W.; Yao, X.; Cheng, G. IPS-YOLO: Iterative Pseudo-fully Supervised Training of YOLO for Weakly Supervised Object Detection in Remote Sensing Images. *IEEE Trans. Geosci. Remote Sens.* **2025**, *14*, 5630414.
9. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
10. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
11. Farhadi, A.; Redmon, J. Yolov3: An incremental improvement. In *Computer Vision and Pattern Recognition*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 1804, pp. 1–6.
12. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
13. Ultralytics. YOLOv5 (v7.0). 2023. Available online: <https://github.com/ultralytics/yolov5> (accessed on 22 November 2022).
14. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.

15. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
16. Jocher, G. YOLOv8. 2025. Available online: <https://github.com/ultralytics/ultralytics> (accessed on 10 January 2023).
17. Wang, C.Y.; Yeh, I.H.; Liao, H.Y.M. Yolov9: Learning what you want to learn using programmable gradient information. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2024; pp. 1–21.
18. Wang, A.; Chen, H.; Liu, L.; Chen, K.; Lin, Z.; Han, J.; Ding, G. Yolov10: Real-time end-to-end object detection. *arXiv* **2024**, arXiv:2405.14458.
19. Khanam, R.; Hussain, M. YOLOv11: An overview of the key architectural enhancements. *arXiv* **2024**, arXiv:2410.17725.
20. Lei, M.; Li, S.; Wu, Y.; Hu, H.; Zhou, Y.; Zheng, X.; Gao, Y. YOLOv13: Real-Time Object Detection with Hypergraph-Enhanced Adaptive Visual Perception. *arXiv* **2025**, arXiv:2506.17733.
21. Wu, C.; Cai, C.; Xiao, F.; Wang, J.; Guo, Y.; Ma, L. YOLO-LSM: A Lightweight UAV Target Detection Algorithm Based on Shallow and Multiscale Information Learning. *Information* **2025**, 16, 393.
22. Wang, Y.; Cao, M.; Yang, Q.; Zhang, Y.; Wang, Z. YOLO-SSFA: A Lightweight Real-Time Infrared Detection Method for Small Targets. *Information* **2025**, 16, 618.
23. Scapinello Aquino, L.; Rodrigues Agottani, L.F.; Seman, L.O.; Cocco Mariani, V.; Coelho, L.D.S.; González, G.V. Fault Detection in Power Distribution Systems Using Sensor Data and Hybrid YOLO with Adaptive Context Refinement. *Appl. Sci.* **2015**, 15, 9186.
24. Ji, Y.; Ma, T.; Shen, H.; Feng, H.; Zhang, Z.; Li, D.; He, Y. Transmission Line Defect Detection Algorithm Based on Improved YOLOv12. *Electronics* **2025**, 14, 2432.
25. Qian, X.; Li, Y.; Ding, X.; Luo, L.; Guo, J.; Wang, W.; Xing, P. A Real-Time DAO-YOLO Model for Electric Power Operation Violation Recognition. *Appl. Sci.* **2025**, 15, 4492.
26. Tian, Y.; Ye, Q.; Doermann, D. Yolov12: Attention-centric real-time object detectors. *arXiv* **2025**, arXiv:2502.12524.
27. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
28. Howard, A.G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Adam, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
29. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
30. Ramachandran, P.; Zoph, B.; Le, Q.V. Searching for activation functions. *arXiv* **2017**, arXiv:1710.05941.
31. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by back-propagating errors. *Nature* **1986**, 323, 533–536.
32. Tianchi Platform, Guangdong Power Information Technology Co., Ltd. Guangdong Power Grid Smart Field Operation Challenge, Track 3: High-Altitude Operation and Safety Belt Wearing Dataset. Dataset. 2021. Available online: <https://tianchi.aliyun.com/specials/promotion/gzgrid> (accessed on 23 July 2024).
33. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Zitnick, C.L. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*; Springer International Publishing: Cham, Switzerland, 2014; pp. 740–755.
34. Wang, Q.; Wu, B.; Zhu, P.; Li, P.; Zuo, W.; Hu, Q. ECA-Net: Efficient channel attention for deep convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 11534–11542.
35. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
36. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
37. Akyon, F.C.; Altinuc, S.O.; Temizel, A. Slicing aided hyper inference and fine-tuning for small object detection. In Proceedings of the 2022 IEEE International Conference on Image Processing, Bordeaux, France, 16–19 October 2022; pp. 966–970.

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.