applied sciences

MDPI

Article

EMB-YOLO: A Lightweight Object Detection Algorithm for Isolation Switch State Detection

Haojie Chen [1], Lumei Su [1,2,*], Riben Shu [1], Tianyou Li [1] and Fan Yin [3]

[1] School of Electrical Engineering and Automation, Xiamen University of Technology, Xiamen 361024, China; 2322131004@stu.xmut.edu.cn (H.C.); 2322131020@stu.xmut.edu.cn (R.S.); ltyxm@163.net (T.L.)
[2] Xiamen Key Laboratory of Frontier Electric Power Equipment and Intelligent Control, Xiamen 361024, China
[3] The State Grid East China Branch, Shanghai 200001, China; yin_f@126.com
* Correspondence: sulumei@163.com

Abstract: In power inspection, it is crucial to accurately and regularly monitor the status of isolation switches to ensure the stable operation of power systems. However, current methods for detecting the open and closed states of isolation switches based on image recognition still suffer from low accuracy and high edge deployment costs. In this paper, we propose a lightweight object detection model, EMB-YOLO, to address this challenge. Firstly, we propose an efficient mobile inverted bottleneck convolution (EMBC) module for the backbone network. This module is designed with a lightweight structure, aimed at reducing the computational complexity and parameter count, thereby optimizing the model's computational efficiency. Furthermore, an ELA attention mechanism is used in the EMBC module to enhance the extraction of horizontal and vertical isolation switch features in complex environments. Finally, we proposed an efficient-RepGDFPN fusion network. This network integrates feature maps from different levels to detect isolation switches at multiple scales in monitoring scenarios. An isolation switch dataset was self-built to evaluate the performance of the proposed EMB-YOLO. The experimental results demonstrated that the proposed method achieved superior detection performance on our self-built dataset, with a mean average precision (mAP) of 87.2%, while maintaining a computational cost of only 6.5×10^9 FLOPs and a parameter size of just 2.8×10^6 bytes.

Keywords: isolation switch detection; lightweight object detection algorithm; attention mechanism; efficient mobile inverted bottleneck convolution

check for updates

Citation: Chen, H.; Su, L.; Shu, R.; Li, T.; Yin, F. EMB-YOLO: A Lightweight Object Detection Algorithm for Isolation Switch State Detection. *Appl. Sci.* **2024**, *14*, 9779. https://doi.org/10.3390/app14219779

Academic Editor: Stéfano Frizzo Stefenon

Received: 17 September 2024
Revised: 12 October 2024
Accepted: 23 October 2024
Published: 25 October 2024

Copyright: © 2024 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/).

1. Introduction

Isolation switches are critical components in power systems. Accurate and reliable determination of their operational status is essential for power grid security. With the rapid development of smart grids and the increasing prevalence of unmanned substations, there is a growing need for real-time monitoring and automatic identification of the operational status of isolation switches [1]. Manual inspections traditionally consume both time and effort, while also facing challenges in maintaining accuracy and timeliness, particularly during unfavorable weather conditions. With the rise of automation and intelligent inspection technologies, computer vision and pattern recognition techniques are increasingly being applied in modern power systems to assist in monitoring the operational status of isolation switches [2–4].

In the early stages, most research on visual isolation switch detection methods primarily focused on traditional machine learning methods. These methods first design feature descriptors for isolation switches, such as histogram of oriented gradients (HOG) and local binary patterns (LBP), to extract image feature information. Traditional machine learning classifiers, such as support vector machines (SVM) and random forests, are designed to determine the operational status of isolation switches [5]. Compared to deep-learning-based methods, traditional machine learning methods do not necessitate large image datasets

of isolation switches for model training and are comparatively easier to deploy on inspection robots. However, in complex power monitoring scenarios, the isolation switch state recognition results of traditional machine learning methods are susceptible to factors such as lighting variations, occlusions, and diverse viewpoints. The dual-layer cellular neural network (CNN) with a constant template proposed by P. Arena et al. [6] is well-suited for generating self-organizing patterns and can simulate complex phenomena, offering a potential solution for detection in complex and dynamic power scenarios.

In recent years, the rise of deep learning has led to significant progress in the intelligent identification of power system equipment. Object detection algorithms based on deep learning, such as YOLO [7], SSD [8], and Faster R-CNN [9], have been widely used to identify the status of power equipment. These algorithms have demonstrated exceptional performance in detecting objects with nonlinear and high-dimensional features. The accuracy and robustness of detection results have significantly improved. Currently, research on deep-learning-based methods for power equipment detection primarily focuses on the identification and fault detection of other types of power equipment, while little research focuses on isolation switch state detection.

Numerous recent studies on deep-learning-based power equipment recognition have concentrated on identifying electrical devices of various sizes in monitoring scenarios, particularly in environments such as substations. The diverse morphological differences in the equipment within these images pose a significant challenge for visual detection models in adapting to objects of different scales, often resulting in suboptimal recognition outcomes. Wu et al. [10] proposed an enhanced YOLOv5-based visual fault detection algorithm for substation equipment, which adjusts the combination of multi-level features in the network through a floating adaptive weighted fusion strategy, enabling adaptive learning for objects of different scales. Similarly, Bi et al. [11] introduced a YOLOX++ detector, based on YOLOX, which utilizes a multi-scale cross-stage partial network (MS-CSPNet) to fuse multi-scale feature information and expand the receptive field for objects through channel combinations, thereby optimizing the localization of small objects. The primary focus of these studies was to enhance the multi-scale fusion network in the neck of object detection models. By integrating feature maps of different sizes, these models enhance their ability to extract features at various scales, ultimately improving detection accuracy.

In the field of deep-learning-based visual detection for power equipment, significant efforts are being made to improve detection performance across a range of electrical devices. Ou et al. [12] put forth a object detection model based on an improved faster R-CNN, tailored for the automatic detection of five types of power equipment in substations. This method involves modifying the feature extraction network and adjusting the aspect ratios of anchor boxes to elevate the accuracy when identifying elongated equipment. To tackle the challenges presented by the diverse shapes and sizes of distribution equipment, Hu et al. [13] introduced a multi-device detection method for distribution line inspection based on YOLOx-s. This innovative approach enables the simultaneous recognition of multiple types of power equipment, thereby bolstering the intelligence of autonomous UAV inspections of distribution lines. These studies predominantly sought to address the visual detection challenges of electrical equipment with significant appearance variations. By proposing object detection algorithms, they aimed to refine the recognition accuracy of different types of electrical devices.

In the domain of deep-learning-based visual detection for power equipment, there has been a focus on creating lightweight detection models to improve real-time performance. Yu et al. [14] introduced the RepVGG-YOLOv5 model, aimed at transmission line fault detection with drones, which enhances real-time detection without compromising accuracy by optimizing the weight distribution, adjusting single-branch inference, and improving the normalization performance. Similarly, Su et al. [15] restructured the detection head and neck to develop a lightweight insulator defect detection method. This approach minimizes the number of parameters, while preserving robust feature extraction and perception abilities, thereby enhancing the accuracy of insulator defect detection. These studies

typically achieve a light weight by modifying the backbone network and redesigning the detection head, thereby reducing model parameters, while preserving the detection accuracy. This research provides valuable insights for developing real-time object detection models in power system scenarios.

Overall, existing research has primarily focused on improving the accuracy, multi-scale detection, and real-time performance of power equipment detection models. These improvements can be summarized as follows: (1) Multi-scale fusion methods have been introduced to enhance a model's ability to recognize objects of varying sizes; (2) Convolutional modules and attention mechanisms have been designed to strengthen feature extraction, thereby improving model accuracy; (3) Lighter network architectures have been proposed to accelerate model inference, thereby enhancing real-time performance.

It is important to note that isolation switch status detection differs from other electrical equipment detection tasks, due to its unique characteristics. First, compared to other electrical devices, isolation switches have a relatively elongated shape with a large length-to-width ratio. Second, the structure of isolation switches is variable; they exhibit significant visual differences between open and closed states, and may also appear in a separated state. Third, in monitoring scenarios, the scale of isolation switches can vary within the same image, due to differences in the distance from the monitoring equipment. Finally, power monitoring systems often generate large volumes of image data. To ensure faster system response and reduce data transmission delays, lightweight models must be deployed at the edge to process images in real time.

To address the aforementioned challenges, this paper first proposes an efficient mobile inverted bottleneck convolution (EMBC) module within the backbone network. This module leverages an ELA attention mechanism [16] to enhance the feature extraction capability of the detection model for complex backgrounds. The ELA attention mechanism focuses on the spatial features of object positions in both the horizontal and vertical directions. By independently processing the feature vectors of each direction, it activates attention weights, guiding the model to focus on the location information of the isolation switch in different states. Second, an efficient-RepGDFPN feature fusion module is employed, which uses lightweight convolutional modules and a fast normalized fusion method to enhance the model's multi-scale feature fusion capability. Finally, depthwise separable convolutions are used to replace the traditional convolutions, achieving a lightweight model for isolation switch detection and improving the real-time inference performance.

Compared to existing methods, the innovation of EMB-YOLO in the field of power monitoring mainly lies in its lightweight design and optimization for the specific object of isolator switches. Through the combination of the EMBC module and the efficient-RepGDFPN fusion network, we not only improved the computational efficiency of the model, but also addressed the instability issues of traditional methods in complex environments. These designs ensure that EMB-YOLO can provide a higher detection accuracy and real-time performance in practical applications within power monitoring scenarios.

The paper proposes a lightweight object detection algorithm, EMB-YOLO, for identifying the states of isolation switches. This algorithm ensures accurate detection of isolation switch states, even in complex background environments, while significantly reducing the computational overhead. It can be applied to intelligent power monitoring systems, enhancing the detection performance of isolation switch states, while maintaining the detection speed. The contributions of this paper are as follows:

- To achieve timely and accurate detection with complex backgrounds, we modified the bottleneck in the original C2f module of YOLOv8 based on the characteristics of isolation switches. A more efficient EMBC-Module was designed, enhancing the model's feature extraction capabilities to accurately capture isolation switch features from complex substation environments.
- A novel feature fusion module, efficient-RepGDFPN, was developed. By utilizing the lightweight convolutional module GSconv and a fast normalized fusion method, this

module efficiently extracts features of isolation switches at different scales. The capability of the proposed model for extracting multi-scale features is ultimately enhanced.
- Considering the characteristics of horizontally and vertically extensible isolation switches, we integrated the ELA attention mechanism into the network. By employing strip pooling in the spatial dimension, the model effectively extracts feature vectors in both horizontal and vertical directions, capturing long-range dependencies. This approach is highly beneficial for handling global features in images, such as object shapes and positions.
- We collected and processed surveillance images from various power operation sites to create a dataset for detecting the open and closed states of isolation switches. This dataset includes objects such as single-arm horizontal telescopic isolation switches, single-arm vertical telescopic isolation switches, and double-arm vertical telescopic isolation switches, in both their open and closed states.

2. Related Work

2.1. Vision-Based Object Detection for Fault and Anomaly Identification in Power Equipment

In the field of power system automation, the fault and anomaly detection of power equipment is a typical problem in monitoring the operational status of such equipment. With the widespread application of deep learning technologies, an increasing number of studies have employed automated analysis of image data from power equipment. By utilizing visual object detection methods, these studies swiftly identified faults and anomalies, enabling early warnings and timely interventions. This approach helps ensure the stable operation of power systems.

Zheng et al. [17] employed an improved YOLOv4 model for detecting circuit breaker and insulation switch faults. Their approach achieved intelligent diagnosis of thermal faults by rapidly identifying and extracting device temperature features. Peng et al. [18] proposed a defect recognition method called EDF-YOLOv5 based on YOLOv5s, which enhanced the accuracy of transmission line defect detection. Liu et al. [19] introduced the CSPD-YOLO model for detecting insulation switch faults in complex background aerial images. Their method improved the detection accuracy by utilizing feature pyramid networks and an enhanced loss function. Zhao et al. [20] developed a limited slip network (LSNet) for detecting minor defects in transmission line infrastructure. Li et al. [21] proposed a lightweight power equipment detection network (PEDNet) based on YOLOv4-tiny, addressing challenges such as complex infrared backgrounds, low contrast in infrared images, and object rotation, resulting in improvements in both detection speed and accuracy.

In summary, while significant progress has been made in deep learning models for detecting power equipment faults, most research has focused on devices such as insulators and transmission lines, with relatively less attention given to the status recognition of isolation switches. Therefore, this paper proposes algorithmic optimizations specifically for the characteristics of isolation switches. The proposed approach achieves an effective balance between accuracy and model efficiency in isolation switch status detection, enhancing the practicality of the model for real-world applications.

2.2. Attention Mechanism

Attention mechanisms enable neural networks to focus exclusively on crucial features of an image, while disregarding less significant features. This capability has demonstrated substantial potential in enhancing the performance of deep convolutional neural networks. Consequently, attention mechanisms have been widely adopted across various object detection algorithms within the field of computer vision [22].

Huang et al. [23] proposed the cross-cross attention model, which introduced a cross-crossover attention mechanism to more efficiently capture global image dependencies, ultimately demonstrating advanced model performance across multiple benchmark datasets. Hu et al. [24] introduced an effective channel attention mechanism known as squeeze-and-excitation (SE) attention, which enhances model detection accuracy by leveraging 2D

global pooling and fully connected structures. Wang et al. [25] designed an efficient channel attention (ECA) mechanism to better capture global information and reduce computational complexity. This mechanism improves the channel relationships in feature maps, thereby enhancing the performance of deep learning models. Woo et al. [26] proposed a convolutional block attention module (CBAM), which integrates channel attention and spatial attention modules. By using average pooling and maximum pooling to aggregate features, the CBAM effectively enhances the feature representation capabilities of convolutional neural networks.

The introduction of attention mechanisms has been proven to significantly enhance both the accuracy and efficiency of detection tasks. This mechanism enables detection models to focus more precisely on critical regions within an image. In the context of detecting the open and closed states of isolation switches, incorporating attention mechanisms can strengthen a model's ability to capture switch features and improve the detection performance. By leveraging attention mechanisms, a model can better interpret and analyze visual data, thereby achieving efficient and accurate detection of isolation switch states.

2.3. Lightweight Object Detection Algorithm

Lightweight object detection algorithms enhance model efficiency by reducing complexity, optimizing computational processes, and compressing parameters, while maintaining accuracy. To improve the detection efficiency for power equipment, the application of real-time lightweight algorithms in power systems is essential. Furthermore, efficiently running complex deep learning models on resource-constrained embedded systems and mobile devices has become a key focus of research.

Han et al. [27] designed GhostNet, which enhances the model performance by extracting feature maps with reduced computational resources. Howard A G et al. [28] developed MobileNets, which employs depthwise separable convolution to create lightweight models. MobileNetV2 [29] introduces an inverted residual structure, while MobileNetV3 [30] leverages AutoML techniques to improve the model performance, while reducing computational complexity. Zhang [31] proposed ShuffleNet, which optimizes the information flow between different channel groups using channel shuffle operations. ShuffleNetV2 [32] further refined the model design for improved compatibility with object hardware. Tan et al. [33] developed EfficientNet through a neural architecture search, optimizing the width and depth ratios of the network to enhance efficiency. Ding et al. [34] introduced RepVGG, which unifies different convolutional layer structures (such as standard and depthwise separable convolutions) during training, thereby achieving a higher inference efficiency by simplifying the network structure during inference.

In recent years, with the introduction of the ViT [35] model, it has been demonstrated that transformers also hold great potential in the field of computer vision. Some researchers have focused on how to make ViT models more lightweight. LeViT [36] proposed a multi-stage transformer architecture that uses an attention mechanism for downsampling, significantly improving the computational efficiency. MobileViT [37], by introducing local convolution operations and a simplified transformer structure, retains the ViT's advantage in capturing global information, while maintaining the efficiency of a lightweight network. EdgeViTs [38] developed the transformer block and introduced a cost-effective bottleneck, achieving a better accuracy–latency balance. Additionally, DETRs [39] have garnered widespread attention in academia for eliminating various handcrafted components. Recently, lightweight object detection models based on DETRs have also emerged. Lite-DETR [40], specifically designed for efficiency, introduces an efficient encoder block, enhancing the model efficiency through optimizations to the architecture and computational requirements. RT-DETR [41], developed by Baidu, leverages ViT to efficiently process multi-scale features, achieving real-time performance, while maintaining a high accuracy.

These models have achieved impressive performance with minimal FLOPs, ensuring high-performance object detection even under resource-constrained conditions, which is essential for practical applications in power equipment monitoring. To improve the real-

time inference capability of the model, this study employs more lightweight convolutional modules to replace the standard convolutions, thereby maintaining a low parameter count and computational complexity, while still delivering effective detection results.

3. Methods

3.1. EMB-YOLO Lightweight Attention Detection Network

In practical isolation switch status detection scenarios, images captured by monitoring systems or inspection robots are typically processed to identify the coordinates of the isolation switch and determine whether it is in an open or closed state. To improve the detection accuracy and better adapt to multi-scale scenarios, we propose a network named EMB-YOLO based on YOLOv8, as illustrated in Figure 1. First, we adopt several of our proposed C2f_EMBC modules, along with depthwise separable convolutions. This combination efficiently extracts features from different channels and promotes information exchange between them. Through a series of convolutional operations, multi-scale features are extracted from the input image, enhancing the feature extraction capability for isolation switches. In addition, compared to standard convolution, depthwise separable convolution breaks the standard convolution operation into two smaller, more efficient operations, significantly reducing the number of parameters. Next, the efficient-RepGFPN [42] fusion network, combined with GSConv convolutions [43] and a weighted feature fusion method, is used to integrate multi-scale features extracted by the backbone, improving the detection capability across different isolation switch scales. Finally, the multiple detection heads consist of convolutional layers, pooling layers, and fully connected layers. Each detection head receives feature maps of different scales from the fusion network and ultimately outputs the bounding box coordinates, class labels, and confidence scores for the corresponding scales.

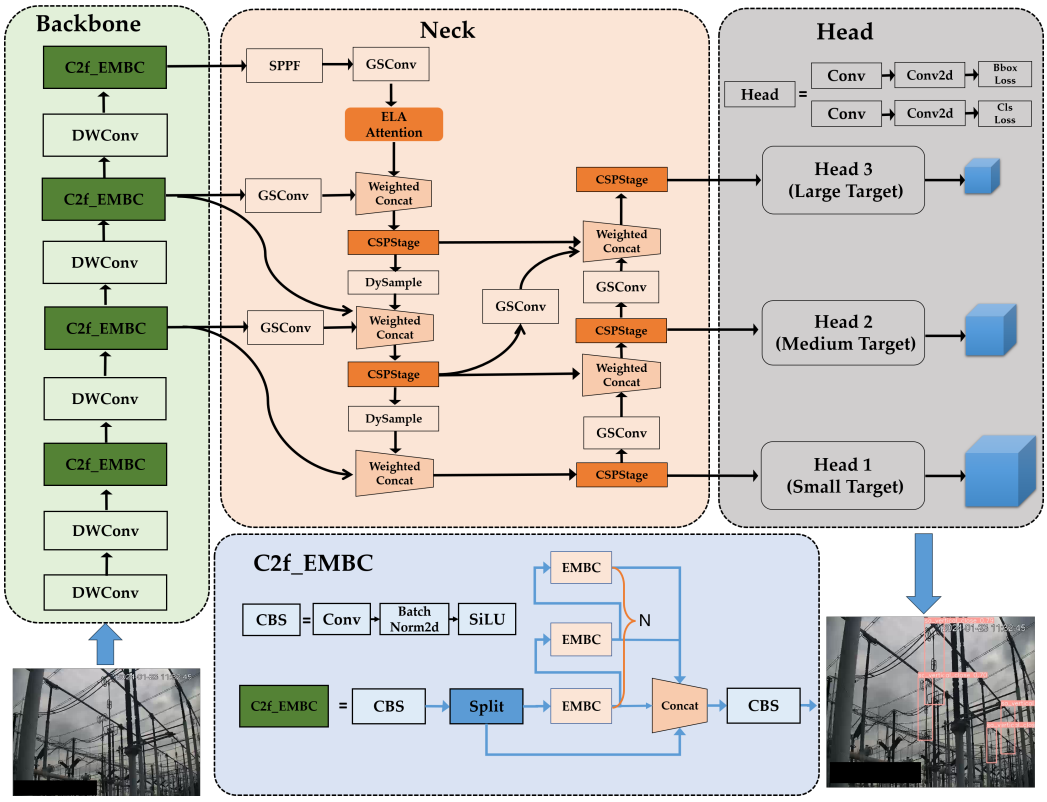


Figure 1. The architecture of the EMB-Net, a novel lightweight attention detection network. Here, EMBC stands for efficient mobile inverted bottleneck convolution module, CSPStage is the cross stage partial module, and DySample refers to the dynamic upsampler.

3.2. Efficient Mobile Inverted Bottleneck Convolution Module

To enhance the detection performance for isolation switches, specific filters can be employed to emphasize their structural features. Additionally, to reduce the parameter count and computational load of the backbone network, efficient and lightweight convolutions can be used to construct these filters. Consequently, this paper improves the bottleneck module in C2f by incorporating the MBConv module from EfficientNet [33], resulting in the design of an efficient and lightweight network module, EMBC, as illustrated in Figure 2.

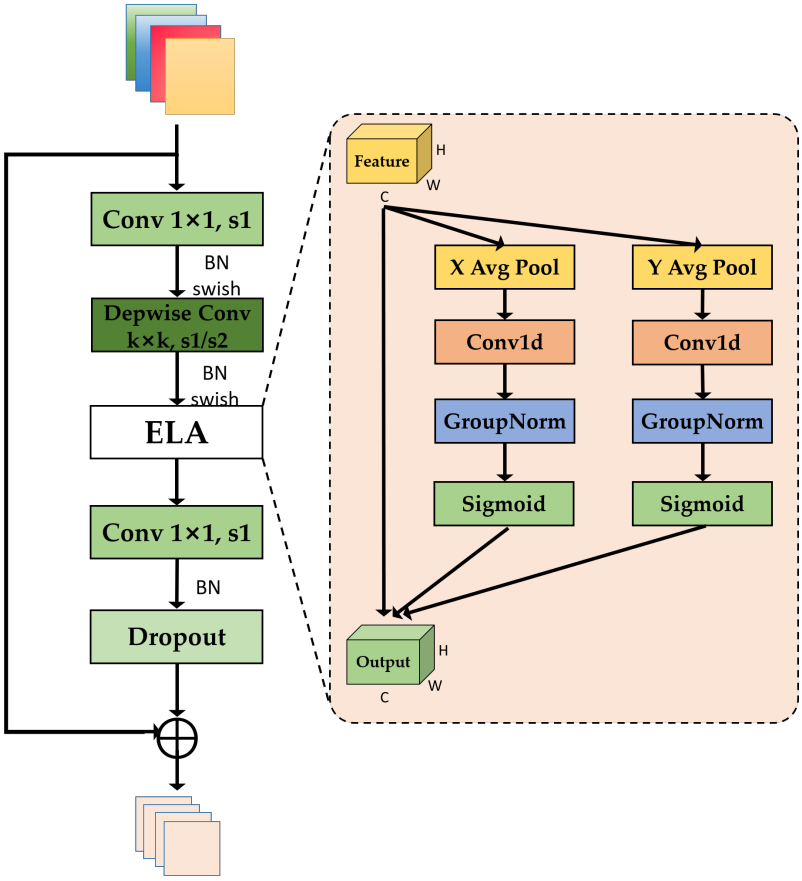


Figure 2. The structure of the EMBC (efficient mobile inverted bottleneck convolution) module.

First, to enable the model to learn different abstract features from the input data, we applied pointwise convolution to increase the network's depth. This helps the model acquire more fine-grained features and makes it easier to train. Meanwhile, to ensure that the network remains lightweight, while capturing richer and more complex features, we utilized depthwise convolutions for feature extraction. As shown in Equations (1) and (2), δ represents the swish activation function.

$$X_{2,c_2} = \delta\left(BN\left(PWConv(X_{1,c_1})\right)\right) \tag{1}$$

$$X_{3,c_3} = \delta(BN(DWConv(X_{2,c_2}))) \tag{2}$$

Secondly, to enhance the spatial feature representation in different directions during isolation switch detection, we introduced the lightweight and efficient ELA attention mechanism. The ELA mechanism extracts horizontal and vertical feature vectors through strip pooling in the spatial dimension, as shown in Equations (3) and (4). Initially, each channel undergoes average pooling in both directions, maintaining an elongated kernel shape to capture long-range dependencies, while avoiding interference from irrelevant regions in label prediction. Subsequently, 1D convolutions are applied to strengthen the positional information in both the horizontal and vertical directions. Group normalization (GN) is

then employed to process the enhanced positional information, representing positional attention in these directions, as demonstrated in Equations (5) and (6). Here, F_h represents the 1×1 convolution, and σ denotes the sigmoid activation function. This approach generates informative object position features in each direction. ELA independently handles these directional feature vectors for attention prediction, and then integrates them via multiplication to produce the final output Y, as shown in Equation (7). This ensures precise location information for the regions of interest, while maintaining the model's lightweight nature. Such processing is crucial for detecting the open and closed states of telescopic isolation switches.

$$z_c(h) = \frac{1}{H} \sum 0 \le i < H x_c(h, i) \tag{3}$$

$$z_c(w) = \frac{1}{W} \sum 0 \le j < W x_c(j, w) \tag{4}$$

$$y^h = \sigma\big(G_n\big(F_h(z_h)\big)\big) \tag{5}$$

$$y^w = \sigma(G_n(F_w(z_w))) \tag{6}$$

$$Y = x_c \times y^h \times y^w, \tag{7}$$

Lastly, a residual module is used to integrate the input features with those processed by the ELA attention mechanism, in order to enhance the feature extraction capability of the EMBC module. The residual connection allows the fusion of input features with the ELA enhanced features, effectively leveraging multi-scale feature information. With this approach, the network convergence is accelerated, so that the model is capable of effectively learning features associated with isolation switches in a short period of time. Moreover, residual connections mitigate the gradient vanishing problem, enabling higher-level features to be extracted from networks. Consequently, by incorporating both ELA and residual modules, the EMBC module is better equipped to extract features from both horizontal and vertical types of isolation switches, leading to a more accurate detection and localization.

3.3. Efficient-RepGDFPN Feature Fusion Network

In the detection of isolation switch states, variations in the sizes of isolation switches pose a challenge, particularly for distant switches, where significant detail loss occurs due to their smaller dimensions. To enhance the accuracy of multi-scale isolation switch detection, while maintaining a lightweight model, we improved the efficient-RepGFPN feature fusion network by incorporating the GSConv and DySample modules [44], resulting in a more lightweight version, efficient-RepGDFPN. Our improvements were based on two key considerations: First, the original efficient-RepGFPN utilizes standard convolutions for feature extraction, which does not meet the requirements for a lightweight model design. Second, the original network performs feature fusion by concatenating feature maps, a method that directly combines different input feature maps, without any distinction. However, in object detection networks, input feature maps have varying resolutions and contribute differently to the final fused output. Therefore, simply concatenating them may not be the optimal approach.

To address this, we proposed a more efficient feature fusion network Efficient-RepGDFPN, designed to serve as the bottleneck stage of the detection model. This network meets the requirements for a lightweight object detection network in the context of isolation switch status recognition, by providing an enhanced feature fusion mechanism. The overall structure of this network is illustrated in Figure 3. Compared to the original efficient-RepGFPN feature fusion network, the key improvements of efficient-RepGDFPN are as follows:

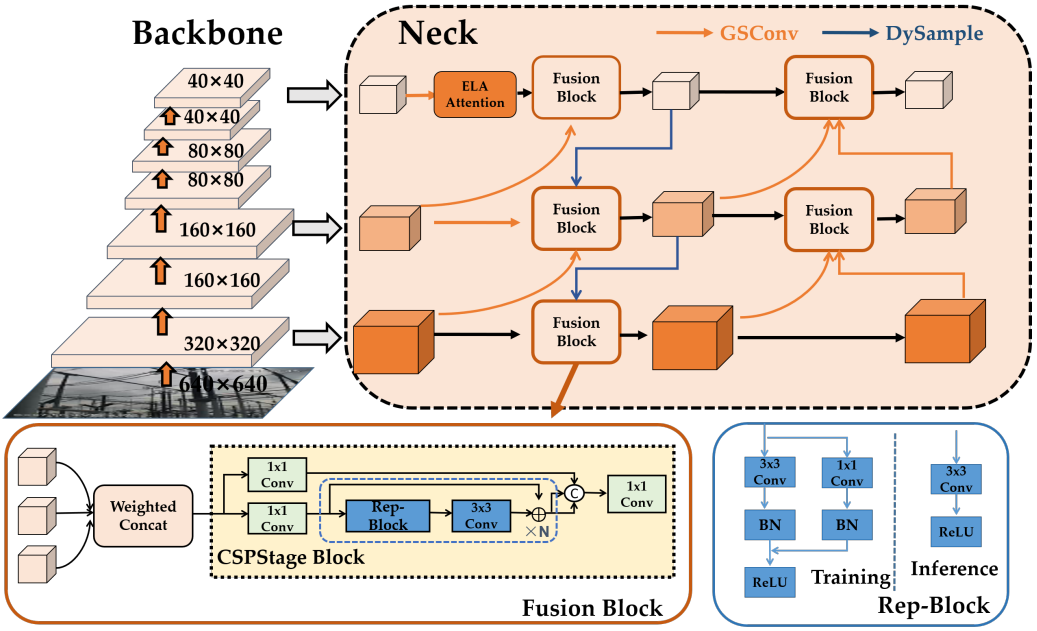


Figure 3. Structure of the efficient-RepGDFPN feature fusion network.

Firstly, we introduce the lightweight convolution module GSConv, which integrates standard convolution, depthwise convolution, and channel shuffle operations. This module achieves a feature extraction performance comparable to standard convolution at a significantly lower computational cost, effectively reducing the model's burden, while maintaining a high performance.

As illustrated in Figure 4, when input feature maps are fed into the GSConv convolution module, the process is as follows: First, a standard convolution performs channel transformation, outputting feature maps with $c_2/2$ channels. Next, depthwise convolution is applied, also producing feature maps with $c_2/2$ channels. Finally, these two feature maps, which have the same number of channels, are concatenated and passed through a channel shuffle operation to achieve information fusion between the feature channels.

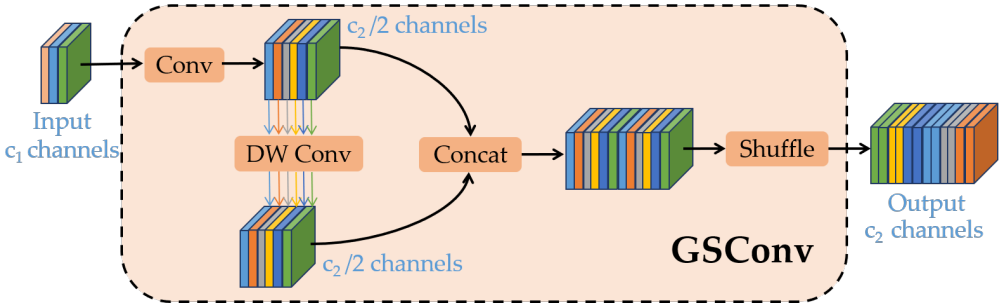


Figure 4. The GSConv convolutional structure.

Secondly, we introduce a weighted feature fusion method, fast normalized fusion, that enables the network to more efficiently learn the importance of different input features by performing distinct fusion based on their significance. As shown in Equation (8), fast normalized fusion scales each normalization weight to the range [0,1], while offering faster training speeds.

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i \tag{8}$$

Thirdly, in the upsampling section, a highly efficient and ultra-light dynamic upsampler, DySample, is utilized. Unlike conventional kernel-based dynamic upsampling methods, DySample is designed from a point-sampling perspective. It consumes fewer

computational resources and demonstrates superior performance in terms of latency, training memory, training time, GFLOPs, and parameter count compared to other upsampling methods. As illustrated in Figure 5, given a feature map $\mathcal{X}$ with dimensions $C \times H_1 \times W_1$ and a sampling set $\mathcal{S}$ with dimensions $2 \times H_2 \times W_2$ (where the first dimension of 2 represents the x and y coordinates), the "grid sample" function uses the positions in $\mathcal{S}$ to perform bilinear interpolation on the assumed $\mathcal{X}$, generating an upsampled feature map $\mathcal{X}'$ with dimensions $C \times H_2 \times W_2$.

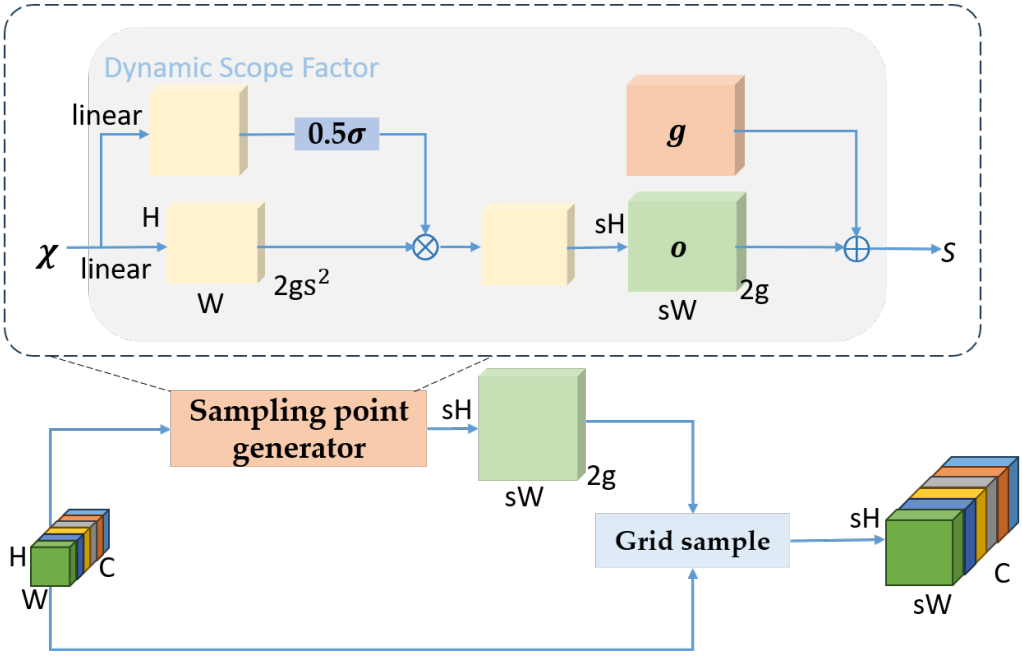


Figure 5. The DySample upsampling structure.

Lastly, The efficient-RepGDFPN feature fusion network retains the basic structure of the fusion block from the original efficient-RepGFPN. This module introduces structural reparameterization techniques and adopts the connection form of efficient layer aggregation networks (ELAN) [45] to effectively enhance the model accuracy, with a relatively low computational cost. The fusion block compresses the input feature map into two parts using two 1×1 convolutional layers, which are then output to two separate branches. This design significantly reduces the computational complexity of the module. One of the branches is a skip connection branch that directly concatenates with the output of the other branch. The other branch is the feature extraction branch, mainly composed of a residual combination of N 3×3 convolutions and structural reparameterization modules (Rep-Blocks). This branch facilitates sufficient interaction of the input features across different stages through skip connections, enhancing the reuse of features. This approach improves the feature extraction capability of the isolation switch state detection model, while boosting the model's inference efficiency.

4. Results

4.1. Experimental Setting

The hardware environment for the experiments in this study included the Windows 10 operating system, a 2.10GHz Intel Xeon Silver 4110 CPU, and an NVIDIA GeForce RTX 2080 Ti GPU. The programming environment was based on Python 3.8.15, utilizing CUDA 11.3.1 and cuDNN 8.2.1 for deep learning acceleration, with PyTorch 1.13.0 serving as the deep learning framework. The neural network model was optimized using stochastic gradient descent (SGD) with a learning rate set to 0.01. The input image size was 640×640 pixels, and each batch contained 64 samples. To ensure thorough training, the model underwent 300 epochs of iterative training.

4.2. Experimental Data

Since there is currently no publicly available dataset for detecting the open and closed states of isolator switches, we needed to construct a dataset for algorithm training and evaluation. During the dataset construction process, we carried out a significant amount of work, including image collection, image augmentation, and annotation. In this paper, we obtained a set of isolator switch open and closed state images from actual substation operation and maintenance environments. The collected images were cleaned and annotated to create a dataset suitable for this study. The specific implementation was as follows:

Firstly, we collected a large number of high-voltage isolator switch images from real substations, totaling 1654 images. The images were captured using high-resolution cameras installed within the substations, with resolutions of 704 × 576 and 1280 × 720. Typically, the images were taken from a low-angle view. These images feature various types of isolator switches and cover a range of angles and distances, both near and far, as well as different sizes. The dataset includes images under various lighting conditions, weather environments, and times of day, including morning, noon, and evening. The weather conditions represented include cloudy, foggy, and sunny days, as shown in Figure 6.

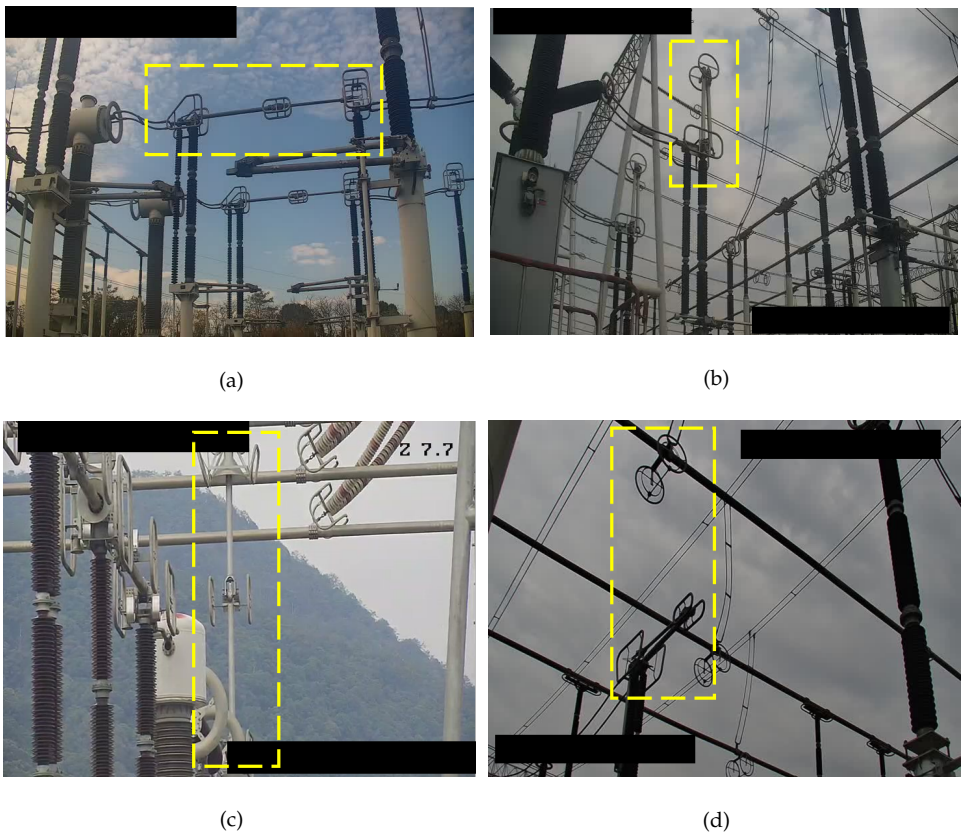


Figure 6. The self-constructed isolator switch dataset includes various angles and weather conditions. (**a**) Represents the closed state of a horizontal telescopic isolator switch, (**b**) represents the open state of a horizontal telescopic isolator switch, (**c**) represents the open state of a vertical telescopic isolator switch, and (**d**) represents the closed state of a vertical telescopic isolator switch. We have the object to be detected with a yellow box.

Secondly, we performed data cleaning and augmentation on the collected data, as some images were blurry, incomplete, or duplicated. Before annotation, we cleaned the dataset by removing invalid or unclear images. To enhance the robustness and generalization of the training algorithm and improve its performance under noise, environmental changes, and other anomalies, we applied data augmentation techniques during training. These

techniques included the following: Scaling and flipping: By scaling and flipping the images, we generated new training samples. Lighting adjustments: We modified the hue and saturation of the images to simulate various lighting conditions, helping the model adapt to real-world changes in lighting. Blur effects: We added blurred images to improve the model's robustness in noisy environments. Mosaic augmentation: We combined four training images into one, simulating different scene compositions and object interactions, thereby enhancing the model's ability to recognize complex scenarios.

Finally, we used the LabelImg 1.8.6 software to annotate six objects in the power equipment dataset. These annotations covered three common types of isolator switches: single-arm horizontal telescopic isolator switches, single-arm vertical telescopic isolator switches, and double-arm vertical telescopic isolator switches. Each type of switch image was further divided into two states: open and closed. We annotated the images using rectangular bounding boxes, manually marking the position of the isolator switches and assigning corresponding labels based on their open or closed state. After augmenting the dataset, we obtained a total of 4962 images, with 3473 images selected as the training set, 992 as the validation set, and 497 as the test set. Figure 6 shows a sample from the dataset. Specifically, the dataset was designed to detect the open and closed states of isolator switches.

4.3. Evaluation Indicators

4.3.1. Evaluation Metrics for Accuracy

The accuracy metrics used in this paper included precision, recall, and mean average precision (mAP). Precision refers to the ratio of correctly detected objects (true positives, TP) to the total number of detected objects, i.e., the ratio of TP to the sum of true positives (TP) and false positives (FP). The formula for calculating precision is provided in Equation (9).

$$Precision = \frac{k}{N} = \frac{TP}{TP + FP} \tag{9}$$

Recall refers to the ratio of correctly detected objects (true positives, TP) to the total number of actual objects, i.e., the ratio of TP to the sum of true positives (TP) and false negatives (FN). The formula for calculating recall is provided in Equation (10).

$$Recall = \frac{k}{M} = \frac{TP}{TP + FN} \tag{10}$$

Mean average precision (mAP) is a comprehensive metric commonly used in multi-class object detection tasks. To calculate mAP, the average precision (AP) for each object category must first be determined. Then, the mAP is obtained by averaging the AP values across all categories. The specific formula for calculating AP is provided in Equation (11).

$$AP = \sum_{i=1}^{n-1} \left(Recall_{i+1} - Recall_i\right) Precision_{interp}\left(Recall_{i+1}\right) \tag{11}$$

In this context, $Recall_i$ represents the recall value corresponding to the first interpolated point on the precision–recall curve in ascending order, and n denotes the number of recall values considered. After calculating the AP values for each object category, the mean average precision (mAP) is obtained by summing these AP values and taking the average. The formula for calculating mAP is shown in Equation (12).

$$mAP = \frac{\sum_{i=1}^{K} AP_i}{K} \tag{12}$$

In this study, mAP@0.5 and mAP@.5:0.95 were used as accuracy evaluation metrics. Here, mAP@0.5 is the mAP value calculated at an intersection over union (IoU) threshold of 0.5, while mAP@.5:0.95 is a more comprehensive evaluation metric. It calculates mAP at 11

different IoU thresholds, ranging from 0.5 to 0.95, in increments of 0.05, and averages these mAP values to provide an overall assessment.

4.3.2. Evaluation Metrics for Lightweightness

In the design of lightweight object detection algorithms, it is essential, not only to evaluate the model's performance, but also to pay special attention to its complexity. Therefore, this study introduced two commonly used evaluation metrics: floating point operations (FLOPs) and the total number of parameters (parameters) in the convolutional layers. These metrics were used to quantify the model's computational load and parameter count, respectively. The formula for calculating FLOPs is provided in Equation (13).

$$FLOPs \approx \sum_{l=1}^{L} FLOPs^l = 2 \times \sum_{l=1}^{L} \frac{(K^l)^2 \cdot H^l \cdot W^l \cdot C_{in}^l \cdot C_{out}^l}{\left(S^l\right)^2} \tag{13}$$

If the detection model includes L convolutional layers, the total number of parameters (parameters) can be calculated using Equation (14).

$$Parameter = \sum_{l=1}^{L} Parameter^l \approx \sum_{l=1}^{L} (K^l)^2 \cdot C_{in}^l \cdot C_{out}^l \tag{14}$$

In addition, this paper introduced FPS (frames per second) as a metric to measure the actual performance of the model on embedded devices. It represents the number of images a model can detect per second, with the calculation formula shown in Equation (15).

$$FPS = \frac{T_{total}}{N_{images}} \tag{15}$$

4.4. Comparison Experiment

To validate the performance of the proposed EMB-YOLO model in the task of isolation switch detection, we conducted a series of comparative experiments with various object detection models. Specifically, EMB-YOLO was compared with transformer-based object detection algorithms, such as RT-DETR [41]. Additionally, it was evaluated against single-stage object detection algorithms including YOLOv5, YOLO8, YOLOX-tiny [46], and YOLOv10 [47]. Furthermore, we performed comparative experiments using the two-stage object detection algorithm faster-RCNN [9]. The results of these comparative experiments on our custom isolation switch status dataset were as follows:

As shown in the Tables 1 and 2, the proposed EMB-YOLO model outperformed the baseline model YOLOv8n of similar scale, with mAP@0.5 and mAP@.5:0.95 improving by 3.6% and 2.2%, respectively, while also reducing the computational complexity and parameter count. Compared to several well-known algorithms, such as RT-DETR-L (80.8), YOLOv5n (82.1), Faster-RCNN (84.2), and YOLOv10 (72.3), our EMB-YOLO achieved a significant improvement in mAP@0.5, increasing by 6.4%, 5.1%, 3%, and 14.9%, respectively. Additionally, EMB-YOLO demonstrated reductions in both parameter count and computational cost when compared with RT-DETR-L, YOLOv5s, YOLOv8, YOLOX-tiny, and faster-RCNN. Our EMB-YOLO achieved the highest F1 score, indicating that it strikes a good balance between precision and recall, resulting in a more stable performance. In terms of real-time capabilities, it was comparable to YOLOv5s and YOLOv8s, meeting the requirements for real-time detection. This performance advantage can be attributed to the use of the EMBC module in EMB-YOLO, which enhances positional information in both horizontal and vertical directions, captures long-range dependencies more effectively, and strengthens the model's ability to detect isolation switch features, while reducing the number of parameters and computational complexity. This highlights the model's outstanding lightweight performance.

Table 1. Performance comparison experiments based on the isolation switch state detection dataset.

Model	P (%)	R (%)	F1	mAP@0.5 (%)	mAP@0.5:0.95 (%)
RT-DETR-L	84.9	68.1	0.755	80.8	50.7
YOLOv5n	70.1	82.8	0.759	82.1	47.8
YOLOv5s	72.9	85.9	0.788	82.4	47.7
YOLOv8n	87.2	77.3	0.819	83.6	49.3
YOLOv8s	86.5	72.8	0.791	82.7	51.6
YOLOX-tiny	89.2	60.2	0.718	88.2	49.1
Faster-RCNN	86.1	81.0	0.834	84.2	55.4
YOLOv10n	87.6	70.7	0.782	72.3	43.6
EMB-YOLO(Our)	82.6	87.4	0.849	87.2	51.5

Table 2. Comparison of lightweightness metrics.

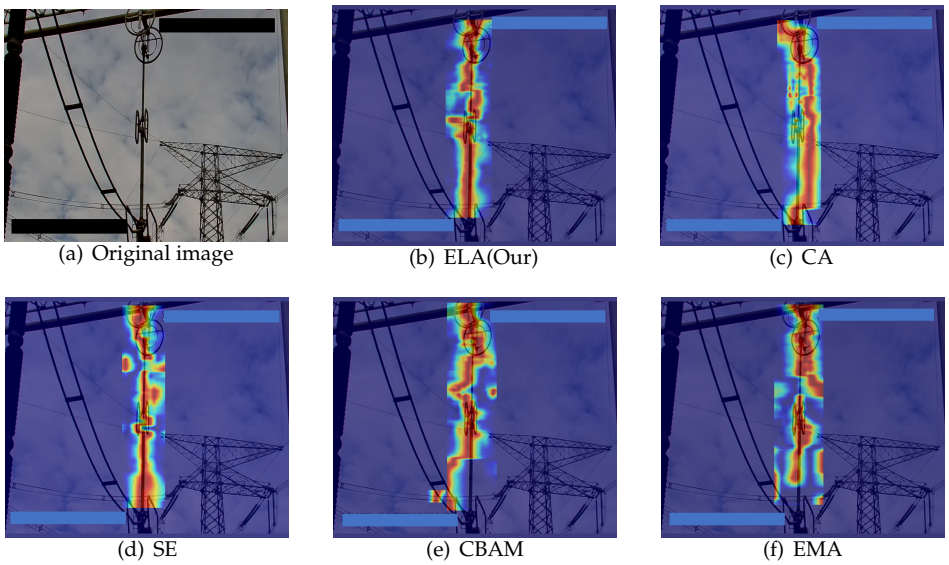
Model	FLOPs ($\times 10^9$)	Paraments (Byte$\times 10^6$)	Model Size (MB)	FPS (img/s)
RT-DETR-L	100.6	28.4	59.1	18.5
YOLOv5n	4.2	1.767	3.9	52.6
YOLOv5s	15.8	7.026	14.4	47.8
YOLOv8n	8.1	3.007	6.2	62.5
YOLOv8s	28.4	11.127	22.5	52.9
YOLOX-tiny	7.5	5.034	38.8	56.3
Faster-RCNN	166	41.374	322	25.5
YOLOv10n	6.7	2.696	5.8	50.7
EMB-YOLO(Our)	6.5	2.861	6.1	48.4

To validate the effectiveness of the ELA attention module used in the EMB-YOLO detection model, we conducted experiments by replacing the ELA attention mechanism within the C2f module. Various comparative experiments were performed to demonstrate the superiority of the ELA attention module in EMB-YOLO. The results, as shown in Table 3, indicated that the model incorporating the ELA attention module in the C2f module achieved a better performance for both mAP@0.5 and mAP@0.5:0.95 on the self-constructed isolation switch dataset. Specifically, compared to the CA attention, SE attention, CBAM attention, and EMA attention modules, the mAP@0.5 improved by 0.3%, 5.2%, 3.7%, and 0.4%, respectively. Additionally, in terms of model complexity, the computational load decreased compared to the EMA and CA attention modules, and the FPS reached 38.4, outperforming the CA attention, SE attention, and EMA attention modules.

We further visualized the training weights of the detection models by implementing five different attention modules (CA, SE, ECA, CBAM, ELA) on the custom isolation switch state detection dataset. The visualization results are presented in Figure 7. It can be observed that, compared to other attention mechanisms, the ELA attention mechanism employed in this study allowed the model to focus more intensively on the elongated features of the isolation switch. This better concentrated and sensitive focus significantly aided the detection model in accurately identifying and localizing the isolation switch, thereby enhancing the model's precision in detecting isolation switch objects.

Table 3. Attention comparison experiments based on self-constructed isolation switch state detection dataset.

Attention Mechanism	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FLOPs ($\times 10^9$)	Paraments (Byte$\times 10^6$)	FPS (t/n)
CA Attention	86.9	55.0	9.2	3.332	35
SE Attention	82.0	49.6	6.5	2.481	37.6
CBAM Attention	83.5	50.8	6.5	2.859	39.4
EMA Attention	86.8	52.7	7.9	2.495	32.4
ELA Attention	87.2	51.5	6.5	2.861	38.4



(a) Original image (b) ELA(Our) (c) CA

(d) SE (e) CBAM (f) EMA

Figure 7. Comparison of different attention mechanisms on the custom isolation switch state detection dataset using heatmaps.

4.5. Ablation Experiments

To validate the effect of the proposed optimization techniques on model performance, ablation experiments were conducted on the self-constructed dataset. As shown in Table 4, ablation studies were performed on the EMB-YOLO series models by replacing the backbone and neck networks of YOLOv8n with those of EMB-YOLO, and experiments were conducted on the self-built isolation switch dataset. The results demonstrated improvements in mAP@0.5 and mAP@0.5:0.95 compared to the baseline model after introducing these techniques. Replacing the standard backbone network with the EMB-backbone improved the accuracy, especially in multi-scale object detection (mAP@0.5:0.95), increasing from 49.3% to 51.8%. This indicates that the EMB-backbone is more effective at feature extraction, reducing the computational overhead, while enhancing accuracy. When only the neck network was introduced, mAP@0.5 improved by 1.6%. By introducing both the backbone and neck networks, mAP@0.5 and mAP@0.5:0.95 increased by 3.6% and 2.2%, respectively. This shows that the combination of the EMB-backbone and EMB-neck more effectively extracts multi-scale features, while reducing the computational complexity.

To evaluate the impact of different modules in the fusion network, the experiment progressively added key components from the neck network into the EMB-YOLO model, such as the efficient-RepGFPN fusion block, lightweight convolution GSConv, weighted feature fusion (Weighted Concat), DySample upsampling module, and more. The effect of each component on the performance of the fusion network was analyzed. After adding each module, the model was retrained and its performance on the validation set was recorded. The experimental results are shown in Table 5.

Table 4. EMB-YOLO model ablation experiments based on a self-constructed isolation switch state detection dataset.

Model	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FLOPs ($\times 10^9$)	Paraments (Byte$\times 10^6$)
YOLOv8n	83.6	49.3	8.1	3.007
+EMB-Backbone	85.7	51.8	6.5	2.859
+EMB-Neck	85.2	48.9	8.1	3.069
+EMB-Backbone+EMB-Neck	87.2	51.5	6.5	2.861

The results indicate that with the unmodified fusion network, the mAP@0.5 and mAP@0.5:0.95 were only 81.5% and 48.9%, respectively. The inclusion of GSConv sig-

nificantly improved the performance of the fusion network, especially for complex environments and multi-scale object detection, where mAP@0.5 and mAP@0.5:0.95 increased by 5.4% and 4.9%, respectively, while reducing the computational load and model size. After adding DySample, both mAP@0.5 and mAP@0.5:0.95 saw notable improvements, indicating that DySample improved the sampling accuracy and enhanced the model's performance in multi-scale scenarios. Although the computational load slightly increased, the performance gains were substantial. The introduction of the weighted feature fusion (Weighted Concat) provided a limited mAP improvement. Our analysis suggests that the fast normalized fusion method used in this experiment trained relatively slowly, and its advantages in weight computation were not fully leveraged in this application. However, when combined with other techniques, it performed better overall.

Table 5. Fusion network ablation experiments based on a self-constructed isolation switch state detection dataset.

Neck	mAP@0.5 (%)	mAP@0.5:0.95 (%)	FLOPs ($\times 10^9$)	Model Size (MB)
Efficient-RepGFPN	81.5	48.9	6.7	6.3
+Weighted Concat	82.0	47.8	6.7	6.3
+GSConv	86.9	53.8	6.5	6.1
+DySample	85.0	51.5	6.8	6.5
+GSConv+DySample	85.9	52.6	6.5	6.1
+GSConv +Weighted Concat +DySample	87.2	51.5	6.5	6.1

5. Visualization Results

Figure 8 displays the detection results of EMB-YOLO on four randomly selected images from the isolation switch status detection dataset. Representative images are shown, including those with varying backgrounds and lighting conditions. Each image highlights the detected switch locations and confidence scores, with the boundary colors representing different detection classes. It can be observed that EMB-YOLO accurately identified targets such as single-arm vertical isolation switches, single-arm horizontal isolation switches, and double-arm vertical isolation switches in both open and closed states. Additionally, EMB-YOLO successfully detected multi-scale targets, including partially occluded isolation switches, with high confidence, further demonstrating the algorithm's exceptional performance in isolation switch status detection.

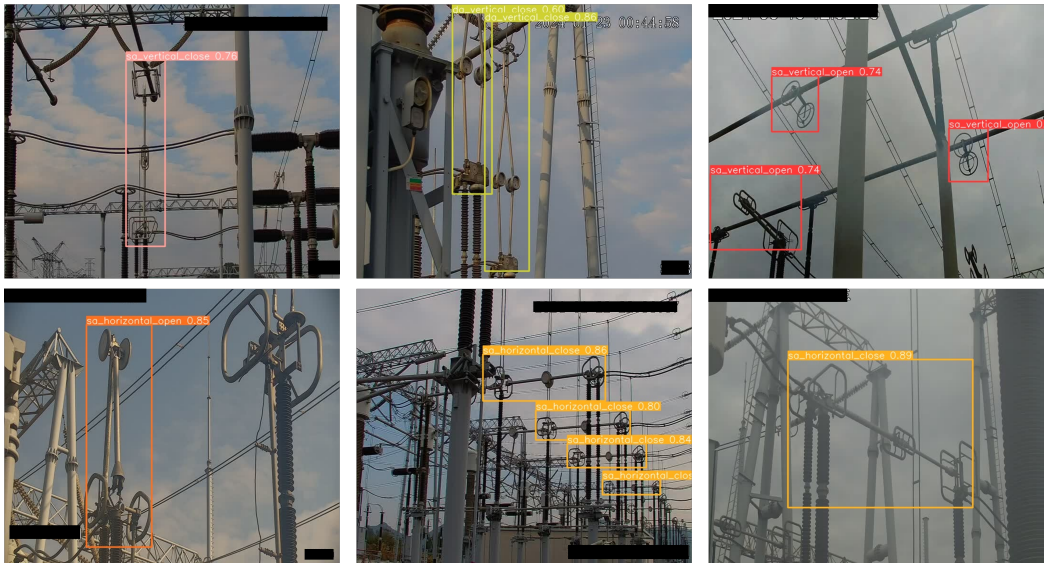


Figure 8. Detection results on the custom isolation switch state detection dataset.

6. Conclusions

In the recognizing the open and closed states of isolation switches, traditional visual detection methods often struggle with a low accuracy, poor adaptability, and insufficient robustness in complex environments. To address these challenges, we proposed a novel algorithm, EMB-YOLO, specifically designed for isolation switch state recognition. Given the characteristics of the isolation switch dataset, we designed an efficient mobile inverted bottleneck convolution module to extract features in complex environments, while reducing the computational costs. This module was specifically designed for isolator switch detection in complex environments. Its novelty lies in combining depthwise separable convolution with pointwise convolution and integrating the efficient local attention (ELA) mechanism. This enhances the model's ability to extract key switch features along the vertical and horizontal axes, which is crucial for power equipment monitoring. The efficient-RepGDFPN fusion network was also utilized to integrate multi-scale features, effectively addressing the scale diversity problem in isolation switch images.

Our model can also be deployed on inspection robots or edge cameras, making it more convenient than traditional visual inspections. For power inspections, this reduces inspection time and labor costs. Especially in large-scale power equipment inspections, it enables automated processing, allowing inspection personnel to monitor the status of isolator switches at any time from a central control center. This reduces the workload and time required for manual inspections.

The EMB-YOLO algorithm mainly relies on the shape features of isolator switches and is relatively insensitive to the influence of texture and color. Therefore, it can even maintain good performance under poor lighting conditions. However, the model still encounters challenges in certain situations. Firstly, when the isolator switch is partially obscured by wires, structural components, or nearby equipment, the detection performance decreases significantly, especially when more than 70% of the switch is obscured. In such cases, the model may fail to detect the switch or misidentify its state. This is because the model relies on distinct vertical and horizontal features of the isolator switch, and severe occlusion greatly affects the detection accuracy. To address this issue, the training dataset could be expanded with more images containing occlusions, and more powerful attention mechanisms or context learning techniques could be introduced to enhance the model's focus on local features, thereby reducing the detection errors in occluded situations.

Secondly, the issue of viewing angles is also a challenge. When the isolator switch is observed from a side angle, the model's reliance on horizontal and vertical shape features can lead to changes in the switch's appearance, thereby affecting the detection accuracy. To improve the model's performance under different viewing angles, diverse angle data samples could be added, and multi-view learning methods could be incorporated to enhance the model's detection ability across various perspectives.

Additionally, small object detection is also a challenge, especially when the isolator switch occupies a very small portion of the image or is captured at a long distance. In such cases, the model often struggles to capture key features, leading to detection failure. This is because small objects have fewer pixels, providing insufficient information for the model. To address this, the efficient-RepGDFPN fusion network could be optimized to enhance the detection capability for small objects, or the image resolution during training could be increased. Additionally, using super-resolution techniques could help to improve the detection accuracy for small objects.

Finally, in images with complex backgrounds, such as those containing many wires, transformers, or other electrical equipment, the model occasionally experiences missed detections or false positives. This is because substations contain many elongated or tubular objects similar in shape to isolator switches, which can cause confusion. To address this, the ELA attention mechanism could be further strengthened or combined with other deep learning techniques to help the model better distinguish isolator switches from other similar objects in complex backgrounds, thereby reducing the occurrence of false positives.

Additionally, we plan to adopt pruning and quantization methods in our future work, which will help reduce the model's computational requirements and improve its detection speed. We will also develop and adjust loss functions to adapt to different types of switchgear. This approach will enhance the model's generalization, making it more reliable in diverse power inspection environments, allowing for more accurate and faster real-time monitoring.

Author Contributions: Conceptualization, H.C. and L.S.; methodology, H.C.; software, H.C.; validation, H.C. and R.S; formal analysis, H.C. and R.S; investigation, H.C. and T.L.; resources, H.C.; data curation, H.C. and R.S.; writing—original draft preparation, H.C.; writing—review and editing, H.C.; visualization, H.C.; supervision, L.S., T.L. and F.Y.; project administration, H.C. and F.Y; funding acquisition, L.S., T.L. and F.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science and Technology Project of East China Branch of State Grid under Grant No. SGHD0000AZJS2310287.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

Acknowledgments: We sincerely thank all contributors of the open-source datasets used in this study. We also appreciate the support and funding from the State Grid East China Branch for our work.

Conflicts of Interest: Author Fan Yin was employed by the company The State Grid East China Branch. 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. The authors declare that this study received funding from The State Grid East China Branch. The funder had the following involvement with the study: data collection, analysis.

References

1. Bozhong, W.; Wenqi, M.; Yizhou, J.; Mingwei, H.; Hui, Z. Review on breaking-closing position monitoring method for intelligent disconnecting switches. In Proceedings of the 2018 International Conference on Advanced Technologies in Energy, Environmental and Electrical Engineering (AT3E 2018), Qingdao, China, 26–28 October 2018; p. 012026
2. Liu, Z.; Wang, H. Automatic detection of transformer components in inspection images based on improved faster R-CNN. *Energies* **2018**, *11*, 3496. [CrossRef]
3. Chen, H.; Zhao, X.; Tan, M.; Sun, S. Computer vision-based detection and state recognition for disconnecting switch in substation automation. *Int. J. Robot. Autom.* **2017**, *32*, 1–12. [CrossRef]
4. Lu, X.; Quan, W.; Gao, S.; Zhang, G.; Feng, K.; Lin, G.; Chen, J.X. A segmentation-based multitask learning approach for isolating switch state recognition in high-speed railway traction substation. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15922–15939. [CrossRef]
5. Nassu, B.T.; Marchesi, B.; Wagner, R.; Gomes, V.B.; Zarnicinski, V.; Lippmann, L. A computer vision system for monitoring disconnect switches in distribution substations. *IEEE Trans. Power Deliv.* **2021**, *37*, 833–841. [CrossRef]
6. Arena, P.; Baglio, S.; Fortuna, L.; Manganaro, G. Self-organization in a two-layer CNN. *IEEE Trans. Circuits Syst. I Fundam. Theory Appl.* **1998**, *45*, 157–162. [CrossRef]
7. Redmon, J. 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.
8. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37.
9. 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. [CrossRef]
10. Wu, Y.; Xiao, F.; Liu, F.; Sun, Y.; Deng, X.; Lin, L.; Zhu, C. A Visual Fault Detection Algorithm of Substation Equipment Based on Improved YOLOv5. *Appl. Sci.* **2023**, *13*, 11785. [CrossRef]
11. Bi, Z.; Jing, L.; Sun, C.; Shan, M. YOLOX++ for transmission line abnormal target detection. *IEEE Access* **2023**, *11*, 38157–38167. [CrossRef]
12. Ou, J.; Wang, J.; Xue, J.; Wang, J.; Zhou, X.; She, L.; Fan, Y. Infrared image target detection of substation electrical equipment using an improved faster R-CNN. *IEEE Trans. Power Deliv.* **2022**, *38*, 387–396. [CrossRef]

13. Hu, L.; Lu, Y.; Wang, S.; Zhang, Y. Multi-Equipment Detection Method for Distribution Lines Based on Improved YOLOx-s. *Comput. Mater. Contin.* **2023**, *77*, 3. [CrossRef]
14. Yu, Z.; Lei, Y.; Shen, F.; Zhou, S.; Yuan, Y. Research on identification and detection of transmission line insulator defects based on a lightweight YOLOv5 network. *Remote Sens.* **2023**, *15*, 4552. [CrossRef]
15. Su, J.; Yuan, Y.; Przystupa, K.; Kochan, O. Insulator defect detection algorithm based on improved YOLOv8 for electric power. *Signal Image Video Process.* **2024**, 1–13. [CrossRef]
16. Xu, W.; Wan, Y. ELA: Efficient Local Attention for Deep Convolutional Neural Networks. *arXiv* **2024**, arXiv:2403.01123.
17. Zheng, H.; Ping, Y.; Cui, Y.; Li, J. Intelligent diagnosis method of power equipment faults based on single-stage infrared image target detection. *IEEJ Trans. Electr. Electron. Eng.* **2023**, *13*, 1706–1716. [CrossRef]
18. Peng, H.; Liang, M.; Yuan, C.; Ma, Y. EDF-YOLOv5: An Improved Algorithm for Power Transmission Line Defect Detection Based on YOLOv5. *Electronics* **2023**, *13*, 148. [CrossRef]
19. Liu, C.; Wu, Y.; Liu, J.; Sun, Z.; Xu, H. Insulator faults detection in aerial images from high-voltage transmission lines based on deep learning model. *Appl. Sci.* **2021**, *11*, 4647. [CrossRef]
20. Zhao, J.; Zhang, K.; Wang, Z.; Liu, F.; Sun, G.; Chou, J.; Xu, M.; Zhang, X.; Liu, X.; Li, Z. Limited sliding network: Fine-grained target detection on electrical infrastructure for power transmission line surveillance. *Int. J. Circuit Theory Appl.* **2021**, *49*, 1212–1224. [CrossRef]
21. Li, J.; Xu, Y.; Nie, K.; Cao, B.; Zuo, S.; Zhu, J. PEDNet: A lightweight detection network of power equipment in infrared image based on YOLOv4-Tiny. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–12. [CrossRef]
22. Xuanhao, Q.I.; Min, Z.H.I. Review of Attention Mechanisms in Image Processing. *J. Front. Comput. Sci. & Technol.* **2024**, *18*, 345.
23. Huang, Z.; Wang, X.; Huang, L.; Huang, C.; Wei, Y.; Liu, W. CCNet: Criss-cross attention for semantic segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 603–612.
24. 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.
25. 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, 13–19 June 2020; pp. 11534–11542.
26. 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.
27. Han, K.; Wang, Y.; Tian, Q.; Guo, J.; Xu, C.; Xu, C. Ghostnet: More features from cheap operations. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1580–1589.
28. Howard, A.G. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv* **2017**, arXiv:1704.04861.
29. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
30. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R. Vasudevan, V. and Le, Q.V. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1314–1324.
31. Zhang, X.; Zhou, X.; Lin, M.; Sun, J. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6848–6856.
32. Ma, N.; Zhang, X.; Zheng, H.T.; Sun, J. Shufflenet v2: Practical guidelines for efficient cnn architecture design. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 116–131.
33. Tan, M. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv* **2019**, arXiv:1905.11946.
34. Ding, X.; Zhang, X.; Ma, N.; Han, J.; Ding, G.; Sun, J. Repvgg: Making vgg-style convnets great again. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13733–13742.
35. Dosovitskiy, A. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929.
36. Graham, B.; El-Nouby, A.; Touvron, H.; Stock, P.; Joulin, A.; Jégou, H.; Douze, M. Levit: A vision transformer in convnet's clothing for faster inference. In Proceedings of the IEEE/CVF International Conference on Computer Visionand and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12259–12269.
37. Mehta, S.; Rastegari, M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. *arXiv* **2021**, arXiv:2110.02178.
38. Pan, J.; Bulat, A.; Tan, F.; Zhu, X.; Dudziak, L.; Li, H.; Tzimiropoulos, G.; Martinez, B. Edgevits: Competing light-weight cnns on mobile devices with vision transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Tel Aviv, Israel, 23–24 October 2022; pp. 294–311.
39. Carion, N.; Massa, F.; Synnaeve, G.; Usunier, N.; Kirillov, A.; Zagoruyko, S. End-to-end object detection with transformers. In Proceedings of the European Conference on Computer Vision (ECCV), Online, 23–28 August 2020; pp. 213–229.
40. Li, F.; Zeng, A.; Liu, S.; Zhang, H.; Li, H.; Zhang, L.; Ni, L.M. Lite detr: An interleaved multi-scale encoder for efficient detr. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Paris, France, 1–6 October 2023; pp. 18558–18567.

41. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
42. Xu, X.; Jiang, Y.; Chen, W.; Huang, Y.; Zhang, Y.; Sun, X. Damo-yolo: A report on real-time object detection design. *arXiv* **2022**, arXiv:2211.15444.
43. Li, H.; Li, J.; Wei, H.; Liu, Z.; Zhan, Z.; Ren, Q. Slim-neck by GSConv: A lightweight-design for real-time detector architectures. *J. Real-Time Image Process.* **2024**, *21*, 62. [CrossRef]
44. Liu, W.; Lu, H.; Fu, H.; Cao, Z. Learning to upsample by learning to sample. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 1–6 October 2023; pp. 6027–6037.
45. Wang, C.Y.; Liao, H.Y.M.; Yeh, I.H. Designing network design strategies through gradient path analysis. *arXiv* **2022**, arXiv:2211.04800.
46. Ge, Z. Yolox: Exceeding yolo series in 2021. *arXiv* **2021**, arXiv:2107.08430.
47. 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.

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.