electronics

MDPI

Article

DLCH-YOLO: An Object Detection Algorithm for Monitoring the Operation Status of Circuit Breakers in Power Scenarios

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

[1] School of Electrical Engineering and Automation, Xiamen University of Technology, Xiamen 361024, China; 2322131020@stu.xmut.edu.cn (R.S.); 2322131006@stu.xmut.edu.cn (L.C.); 2019000005@xmut.edu.cn (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@ec.sgcc.com.cn
* Correspondence: 2013110611@xmut.edu.cn

Abstract: In the scenario of power system monitoring, detecting the operating status of circuit breakers is often inaccurate due to variable object scales and background interference. This paper introduces DLCH-YOLO, an object detection algorithm aimed at identifying the operating status of circuit breakers. Firstly, we propose a novel C2f_DLKA module based on Deformable Large Kernel Attention. This module adapts to objects of varying scales within a large receptive field, thereby more effectively extracting multi-scale features. Secondly, we propose a Semantic Screening Feature Pyramid Network designed to fuse multi-scale features. By filtering low-level semantic information, it effectively suppresses background interference to enhance localization accuracy. Finally, the feature extraction network incorporates Generalized-Sparse Convolution, which combines depth-wise separable convolution and channel mixing operations, reducing computational load. The DLCH-YOLO algorithm achieved a 91.8% mAP on our self-built power equipment dataset, representing a 4.7% improvement over the baseline network Yolov8. With its superior detection accuracy and real-time performance, DLCH-YOLO outperforms mainstream detection algorithms. This algorithm provides an efficient and viable solution for circuit breaker status detection.

Keywords: deep learning; large kernel; attention mechanism; semantic screening

check for updates

Citation: Shu, R.; Chen, L.; Su, L.; Li, T.; Yin, F. DLCH-YOLO: An Object Detection Algorithm for Monitoring the Operation Status of Circuit Breakers in Power Scenarios. *Electronics* **2024**, *13*, 3949. https://doi.org/10.3390/electronics13193949

Academic Editor: Hyunjin Park

Received: 18 September 2024
Revised: 4 October 2024
Accepted: 5 October 2024
Published: 7 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

High-voltage circuit breakers are essential components of a power system, primarily used to protect and control ultra-high voltage transmission lines and substation equipment [1]. The operating status of high-voltage circuit breakers is crucial for maintaining the integrity of the power system. Therefore, real-time monitoring of these breakers is essential [2,3]. In smart power systems, computer vision technology is increasingly used to monitor the status of electrical equipment [4–6]. Typically, in smart power system monitoring, visual object detection methods are used to remotely detect and identify the switching status of circuit breakers in monitoring scenarios, enabling the real-time monitoring of their operating conditions.

The detection of the operating status of circuit breakers using visual object detection algorithms has generally progressed through two phases. In the early stages, research on detecting the operational status of circuit breakers mainly relied on traditional machine learning techniques, such as template matching, threshold segmentation, and morphological operations [7,8]. Although these methods achieved some success with high-resolution images, they still required manual intervention and were easily affected by environmental factors. With the development of deep learning, specifically the application of convolutional neural networks (CNNs) [9], deep learning-based object detection methods have provided a more automated and efficient solution for detecting the operating status of circuit breakers [10,11]. These methods integrate components such as fully connected layers, convolutional layers, pooling layers, and activation functions. The convolutional layers

extract multi-level and multi-dimensional visual features of objects in surveillance images; the pooling layers introduce translation invariance and reduce feature dimensionality through down-sampling; finally, the fully connected layers map the extracted features to specific status categories, enabling precise identification of the operating status of circuit breakers.

High voltage circuit breakers typically use oil-sealed or gas-insulated structures to ensure their proper functioning. When visually detecting the operating status of circuit breakers, the process begins with identifying labels such as "Open" or "Closed" on the instrument panel from the monitoring screen [12]. The operating status of the circuit breaker is then determined based on these labels. Accurately detecting the "Open" and "Closed" labels on the instrument in a complex, dynamic monitoring environment is crucial for identifying the operating status of high-voltage circuit breakers. In recent years, research has been conducted on instrument object detection. Zhang et al. [13] optimized the RCNN network structure and proposed an hourglass residual module network to effectively address issues of numeric deflection and ensure accurate detection and localization of instruments while accurately recognizing water meter readings. Waqar et al. [14] developed a two-stage detection and recognition model based on R-CNN, achieving the efficient detection of digital instruments under complex lighting conditions and improving model robustness. Peng et al. [15] combined corner detection loss functions to reconstruct the detection head, proposing a YOLO-based corner detection method that enables real-time meter reading recognition in natural scenes, addressing challenges in dynamic environments. Liu et al. [16] proposed a non-intrusive load identification method that monitors the charging process of electric bicycle batteries by integrating current signals. Meng et al. [17] proposed a stability monitoring method for railway power supply systems, which integrates the dynamic mobility of vehicles and their position distribution to achieve efficient stability analysis. Zou et al. [18] combined outer circle fitting with the YOLOv5 deep learning model and proposed an anchor design based on an arrester dial dataset, significantly improving reading accuracy in complex scenes. Zhang et al. [19] proposed an automatic recognition method combining object detection and ontology reasoning, using the Pellet inference engine to identify risks during construction processes in surveillance videos.

Research on meter detection methods in power scenarios differs from the detection scenes mentioned above. In monitoring scenarios of power equipment, the challenges in detecting meters and detecting the 'open' and 'closed' labels mainly focus on the following two aspects:

- Object size and rotation angle variations: In power systems, monitoring cameras are typically fixed in place, causing the size and rotation angle of the object to vary within the monitoring scene due to factors such as distance, viewing angle, and object type. These variations may result in the camera failing to accurately capture all the features of the object, leading to missed detections [20].
- Complex monitoring environments: In outdoor monitoring scenarios, the presence of numerous power equipment makes the image background complex, with factors such as uneven lighting and occlusions greatly impacting the accuracy of circuit breaker meter detection. These challenges make it difficult to extract precise details and features, leading to false detections [21].

For multi-scale objects, it is generally easier to detect large-scale objects due to their larger area and more abundant features, whereas small-scale objects are more challenging to detect because of their limited features [22]. With each downsampling operation, the resolution of the feature maps progressively decreases. At this point, CNN networks may inadvertently capture a significant amount of background information when extracting features of small objects. This problem arises from the fixed size of the receptive field. To better handle variations in object shape and scale within this constraint, we introduced deformable large kernel attention [23]. Additionally, the object's positional information

was improved during the feature fusion process, enhancing the ability to perceive objects of various sizes.

The power operation scene typically involves various structurally complex facilities, such as different electrical equipment, power towers, and power lines. These facilities often share visual appearances and texture features with circuit breakers and may obstruct each other. Due to these interferences, extracting features of circuit breakers from monitoring scenes becomes challenging. To overcome this challenge, we adopted a background subtraction method to reduce background interference and focus on the object. This paper utilizes an attention mechanism to enhance the focus of the model on local features. This mechanism generates a weight map to weigh the extraction of object features and suppress background noise interference, thereby allowing for a more reliable determination of the operation status of the circuit breaker.

Drawing from the above analysis, we propose DLCH-YOLO, an innovative algorithm for object detection that utilizes deformable large kernel attention along with semantic filtering. This algorithm accurately locates and identifies the meters with significant scale differences in complex power monitoring scenarios, thereby determining the operating status of the corresponding circuit breaker. This paper makes the following contributions:

1. This paper proposes a C2f_DLKA module based on deformable large kernel attention. This module leverages large convolution kernels to capture long-range dependencies in images, effectively associating distant features with the global features of the object, ensuring that object features of different scales within the same image receive sufficient attention. At the same time, by integrating deformable convolutions, the network adaptively adjusts based on the image content. This design better accommodates objects of varying scales. The module enhances the feature extraction capability of circuit breaker dials under different scales and angle variations. The proposed method is well-suited for object detection tasks that involve variations in scale and angle.
2. This paper proposes a Semantic Screening Feature Pyramid Network (SSFPN) based on HSFPN to achieve multi-scale feature fusion. This network enhances the positional information representation of circuit breaker dials by integrating lower-level feature channels. We introduced a Context Anchor Attention Module, which filters semantic information, highlights key features of the object regions, and suppresses background noise. This design significantly improves the network's accuracy in extracting object features in complex backgrounds. The method presented here can be used to tackle object detection problems in complex backgrounds.
3. This paper introduces generalized sparse convolution into the feature extraction network. This convolution combines depth-wise separable convolution with channel mixing operations, effectively reducing computational load through sparsification. Generalized sparse convolution reduces both the computational complexity of the model and the computational load, while also enhancing feature extraction efficiency, thus facilitating more effective object detection in resource-limited environments.
4. Images from different power system monitoring scenarios were collected and processed to construct a dataset containing circuit breaker meter panel indicators such as open, close, and their respective status. This dataset serves to train and evaluate the DLCH-YOLO algorithm.

2. Related Work

2.1. Meter Detection

In smart power systems, visual meter detection aids in determining the operational status of equipment [5]. Recent advancements in deep learning have led to an increasing number of research efforts on meter detection using these methods. These methods are divided into two main types: object detection [24] and image segmentation [25]. Object detection methods perform one-step localization and detection of specific objects (such as pointers or scales) within an image, while image segmentation methods involve segmenting

regions such as pointers, scales, and digits, and further analyzing their position or angle to obtain specific readings.

Based on image segmentation technology, Ji et al. [26] developed a convolutional neural network (CNN) model based on Region-CNN (RCNN). This model classifies pointer meters while detecting the pointer and uses the angle method to read the meter. Subsequently, Liu et al. [5] used Faster R-CNN to detect the position of the meter and combined it with image processing techniques to determine the position of the pointer and obtain readings. On the other hand, object detection methods have also shown significant advantages in meter detection. Zuo et al. [27] used Mask-RCNN to classify various types of meters and performed digit recognition through the binary masks identified by Mask-RCNN. Additionally, Wang et al. [28] introduced a novel meter detection model built upon YOLOv3. They replaced the original K-means algorithm with the Canopy and K-means++ algorithms and substituted the path aggregation network with a feature pyramid network structure.

It is important to note that in smart power monitoring systems, both the precise identification of circuit breakers and the good adaptability of the detection algorithm to the hardware platform are crucial. However, most algorithms currently fail to meet both requirements simultaneously. To improve the practical detection performance of smart power monitoring systems, this paper proposes a novel object detection algorithm tailored to the operating conditions of circuit breaker status detection and the deployed hardware platform, achieving a balanced trade-off between detection accuracy and model complexity.

2.2. Multi-Scale Feature Fusion

Integrating features at multiple scales is crucial in the field of computer vision, especially for applications such as object detection and semantic segmentation. In real-world scenarios, there are often diverse objects and single-scale features may not meet these needs. By integrating features of different scales, the model more comprehensively captures the diversity of the objects, particularly showing significant improvements when handling small objects, occlusions, or deformations.

In early research, Lin et al. [29] proposed the Feature Pyramid Network (FPN) to overcome challenges associated with multi-scale variations in object detection. They achieved this by integrating features from different layers through bottom-up and top-down feature propagation, providing rich multi-scale features. Subsequently, Liu et al. [30] proposed PANet based on FPN to address detection issues for small and occluded objects. By adding a bottom-up path augmentation structure, shallow feature information is better preserved. Next, Tan et al. [31] proposed BiFPN, a more streamlined bidirectional feature fusion method that integrates multi-scale information more effectively by adaptively weighting the importance of features. Luo et al. [32] proposed CE-FPN, which utilizes high-level semantic features and attention mechanisms for selective feature fusion, improving feature integration and optimization. Additionally, Huang et al. [33] proposed FAPN, creating modules for feature selection and alignment aims to improve the accuracy of feature fusion in FPN while tackling potential misalignment challenges.

However, existing methods fall short of meeting the practical requirements of breaker status recognition tasks. Firstly, these methods are limited in terms of the types of detectable targets and fail to fully consider the unique characteristics of breaker status text labels. Secondly, the power system environment is highly variable, with numerous dashboards displaying status information, yet these methods have not adequately addressed how to effectively locate these dashboards. To address these issues, we use high-level features as weights to filter low-level features, enhancing the model's ability to detect text labels and dashboard locations. The choice of semantic filtering is not only due to its outstanding performance in text and specific object detection but also because it effectively selects key features relevant to the task, significantly reducing irrelevant or distracting information, especially in the complex background of power systems. By merging the filtered features with high-level features, we further enhance the model's spatial awareness, improving the

precision of multi-scale target detection. This approach overcomes the common issues of missed or incorrect detections in complex environments. This feature selection strategy not only improves the accuracy and robustness of breaker status recognition but also ensures the precise localization of targets, even when dashboards or status labels are partially obscured or overlapping.

2.3. Attention Mechanism

Attention mechanisms [34], inspired by human visual focus, dynamically adjust model attention to different features, thereby improving information utilization. This approach not only enhances overall model performance but also effectively addresses the limitations of traditional methods in handling long sequences and capturing long-range dependencies. The primary reason for the success of Transformers in image classification is their self-attention mechanism, which effectively models the long-distance dependencies in the input images.

Vaswani et al. [35] proposed the Transformer model, which successfully captured long-term dependencies through self-attention mechanisms, achieving outstanding performance. Subsequently, Liu et al. [36] proposed the Swin Transformer, which employs a window-based local attention mechanism. This mechanism facilitates cross-window information exchange through the movement of windows between layers, thereby addressing the issue of insufficient context modeling. Liang et al. [37] based on Swin Transformer, proposed the SwinIR method, comprising modules for shallow feature extraction, deep feature extraction, and high-quality image reconstruction. Chen et al. [38] proposed an efficient Transformer model that replaces the multi-head self-attention mechanism in ViT with rectangular window self-attention (Rwin-SA), forming the Cross-Aggregation Transformer Block (CATB) approach. Guo et al. [39] successfully modeled large kernels in CNNs by connecting cascaded depth-wise convolutions and expanded depth-wise convolutions, achieving similar functionality to self-attention mechanisms without significantly increasing computation and memory usage. To further optimize computational costs, Lau et al. [40] proposed a separable version of the LKA module, using cascaded $1 \times k$ and $k \times 1$ convolutions to replace the $k \times k$ convolutions in depth-wise convolutions, effectively mitigating the problem of quadratic parameter growth.

Self-attention mechanisms have gained increasing importance in computer vision due to their ability to capture long-range dependencies. However, their quadratic complexity leads to high computational costs when processing high-resolution images. To address this, we introduce the large-kernel attention mechanism, which combines the strengths of self-attention with those of convolutional networks, effectively utilizing local contextual information. While directly using large convolution kernels can expand the receptive field and capture more global information, it may also introduce additional background noise, potentially interfering with the detection of small objects and reducing accuracy. To overcome this issue, we propose the deformable large-kernel attention mechanism. This mechanism dynamically adjusts the shape and size of the receptive field based on the shape and position of the target, ensuring that the model focuses on critical areas and reduces interference from irrelevant backgrounds. Our method strikes a balance between capturing large-scale global information and accurately localizing small objects in multi-scale target detection, improving both detection accuracy and model robustness.

3. Methods

3.1. DLCH-YOLO

The operational state of circuit breakers is determined by recognizing the text labels ('open' or 'close') on the corresponding meters. In power system monitoring scenarios, the object size and rotation angle of the meters vary according to the monitoring distance and viewing angle of the camera. Furthermore, the complex outdoor scenario and lighting conditions make it challenging for general detection algorithms to accurately identify the circuit breaker meters, which increases the rate of miss detection and false detection. To

address these challenges, we propose the DLCH-YOLO method, developed based on the YOLOv8 model. As depicted in Figure 1, the DLCH-YOLO network is structured with three main elements: the backbone, neck, and head. Firstly, the backbone network consists of the improved C2f_DLKA module and Generalized Sparse Convolution [41], which are used to extract multi-scale features from the input images. The neck is constructed based on HSFPN [42], which enhances the recognition capability of the position information of the circuit breaker meter by fusing low-level channels. Additionally, low-level features are filtered with high-level features for multi-layer feature fusion. Layers convolution, pooling, and fully connected compose the head. The features in each layer are received from the neck network at different scales, and the final output is the position, category, and confidence of the objects at each scale.

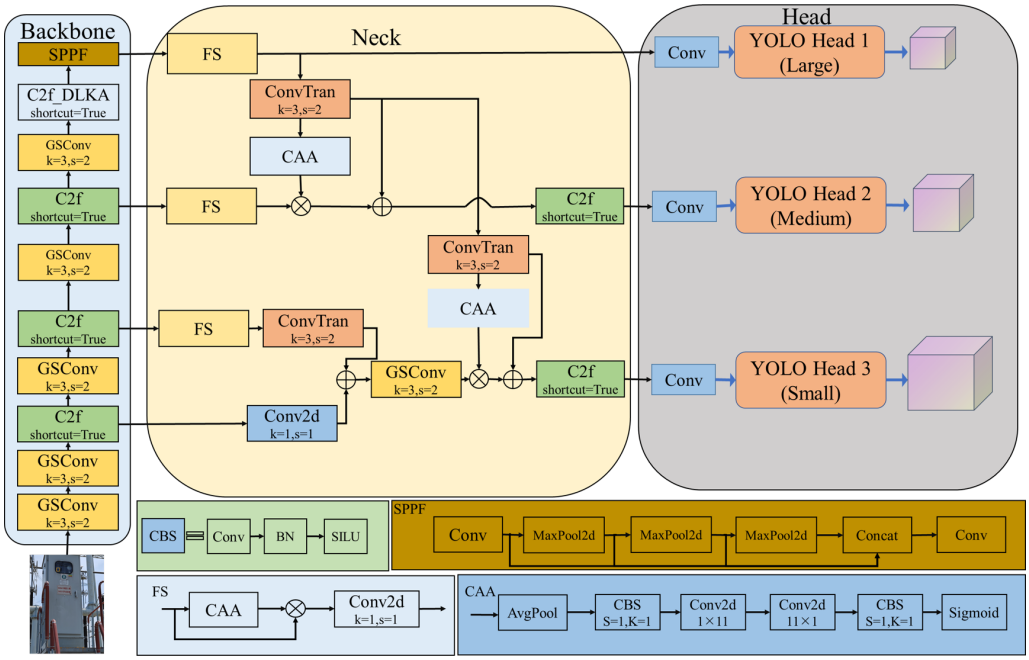


Figure 1. The architecture of DLCH-YOLO network.

3.2. Feature Extraction Network

A feature extraction network transforms input images into complex feature expressions that encapsulate details about different objects, which is a key component of object detection. In YOLOv8, the feature extraction network mainly consists of Conv and C2f modules. The C2f module extracts features by stacking multiple DarknetBottle blocks based on residual connections. However, this structure faces challenges when the scale of the detected object varies in complex backgrounds, such as the meters of the circuit breaker in intelligent surveillance. Current residual connection methods primarily use simple CBS blocks (Conv2d + BatchNorm2d + SiLU), which limit the ability to effectively expand the receptive field, thereby constraining adaptation to objects of different scales. Therefore, a broader receptive field is needed to capture rich semantic and contextual information. Additionally, feature extraction networks are major consumers of computational resources, involving substantial computational and parameter storage requirements, which directly affect the overall performance of the detection model. In applications like power monitoring, where detection speed and accuracy are crucial, it is essential to optimize the feature extraction network for efficiency.

In this paper, a novel Bottleneck architecture is proposed in the C2f module, integrating deformable large-kernel attention [23] to achieve a wide receptive field similar to that of self-attention mechanisms. By using this approach, finer-grained features can be extracted without a significant increase in parameters and computational costs. Through

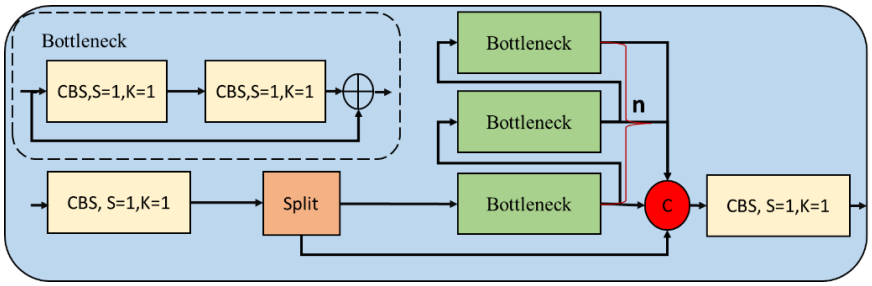
these improvements, we have enhanced feature extraction and detection accuracy while maintaining efficiency. Figure 2 illustrates the improved C2f_DLKA module.

The deformable large-kernel attention mechanism is implemented by incorporating deformable convolutions into the large-kernel convolution. We construct large convolutional kernels by using depth-wise convolution, depth-wise multilayer convolution, and 1 × 1 convolution so that the receptive field is large while the parameter complexity is relatively low. The kernel sizes for depth-wise separable convolution and depth-wise convolution with input dimension H × W and channel number C are determined and constructed using the following equation:

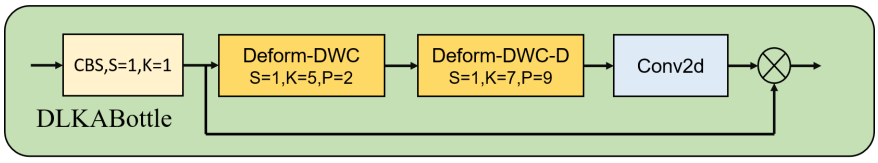
$$DWC = (2d - 1) \times (2d - 1) \tag{1}$$

$$DWC - S = \left[\frac{K}{d}\right] \times \left[\frac{K}{d}\right] \tag{2}$$

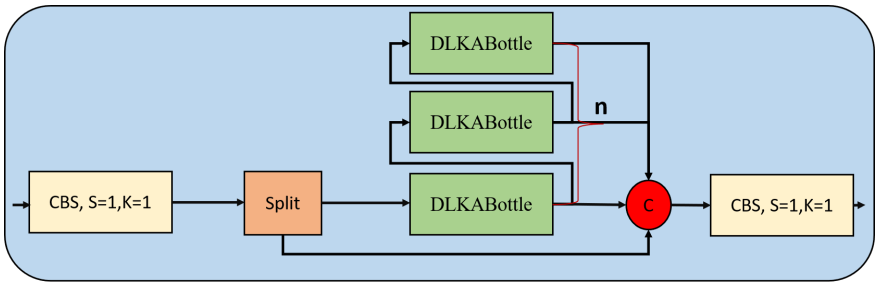
where d denotes the dilation rate and K represents the kernel size.



(a) C2f



(b) DLKABottle



(c) C2f_DLKA

Figure 2. C2f module in the feature extraction network. (**a**) shows the base module; (**b**) shows the improved module that introduces large kernel convolution; (**c**) combines the improved DLKABottle module with C2f.

Furthermore, the large convolutional kernels are designed to be deformable by using an additional convolutional layer to learn offset fields [43], which makes the convolutional kernel dynamically adjust to fit the input features. Given an input feature map $X \in R^{H \times W \times C}$ and an output feature map $Y \in R^{H \times W \times D}$, the deformable convolution is described as follows:

$$Y[h, w, d] = \sum_{c=1}^{C} \sum_{i=1}^{S_h} \sum_{j=1}^{S_w} W_{ic,jd} * X[h + i - 1, w + j - 1, c] \tag{3}$$

The Equations h, w, and d, respectively, represent the pixel coordinates and channel coordinates on the output feature map; S_h and S_w refer to the range of offsets for the kernel along the height and width dimensions; $\dot{W_{ic,id}}$ denotes the weights of the convolutional kernel, respectively. Deformable convolution, in contrast to traditional convolution, adjusts the position of each convolutional kernel by incorporating learnable offsets, allowing for a better match to the local features of the input data. By incorporating deformable convolution into large kernel attention, the model better adapts to multi-scale objects, enhancing the accuracy of object boundary localization. The improved Bottleneck structure, the DLKABottle module, be expressed as follows:

$$Output = Conv2d\big(DeformConv7 \times 7\big(DeformConv5 \times 5(F')\big)\big) + F' \tag{4}$$

where F represents the input features, $F' = CBS(F)$.

To reduce the calculation of the feature extraction network, Grouped Spatial Convolution (GSConv) [41] is used in our object detection algorithm. The GSConv module is designed to combine standard convolution, depth-wise separable convolution, and channel mixing operations to reduce the burden with lower computational costs. As shown in Figure 3, when the input feature map enters the GSConv module, it is first downsampled using standard convolution, reducing the number of channels to c2/2, and generating one feature map. Next, another feature map with c2/2 channels is generated using depth-wise separable convolution (DWConv) [44]. Finally, these two feature maps are concatenated, and channel mixing is used to achieve information fusion. Channel mixing facilitates the uniform exchange of local feature information across various channels, enabling this information to be fully integrated into the output of the depth-wise separable convolution. The GSConv module effectively reduces the negative impact of non-interacting channel information in depth-wise separable convolution and fully utilizes its advantage of low parameter count.

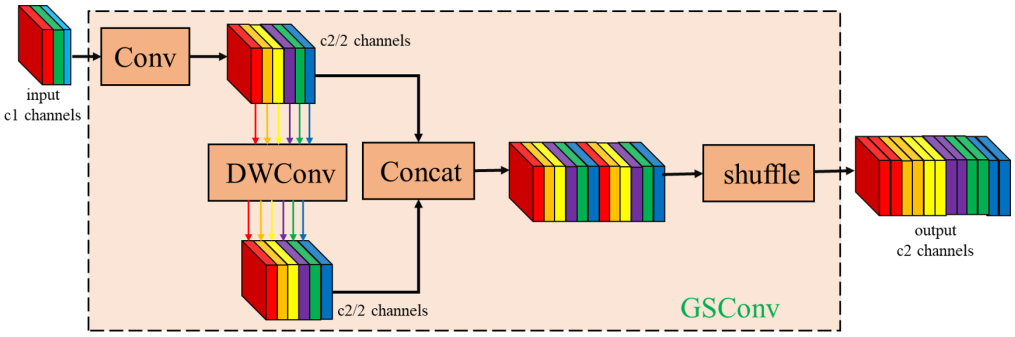


Figure 3. Generalized Sparse Convolution.

3.3. Feature Fusion Network

Feature fusion networks significantly improve detector performance in object detection tasks [45]. These networks are typically used to combine feature maps from various scales and levels of the backbone, thereby capturing rich semantic and positional information. In the field of power system monitoring, the task of detecting the operating status of circuit breakers faces the challenge of multi-scale problems. This challenge arises from variations in monitoring perspectives, causing the size and rotation angle of the object to vary depending on the distance, viewpoint, and type of object, making it challenging to accurately determine the operating status of the circuit breakers.

To address the multi-scale challenges in detecting the operating status of circuit breakers, we developed the Semantic Screening Feature Pyramid Network (SSFPN), as illustrated in Figure 4. First, feature maps of different scales undergo feature screening through the Context Anchor Attention (CAA) module [46]. This module captures local area features by applying a 1 × 1 convolution after average pooling. Next, two depth-wise separable convolutions are used as approximations of standard large-kernel depth-

wise convolution, where $DWConv_{1\times kb}$ and $DWConv_{kb\times 1}$ represent depth-wise separable convolutions with a width of 1 × kb and a height of kb × 1, respectively, to capture long-range contextual information. Finally, the Sigmoid function is used to calculate channel weights, which enhances feature maps with contextual details. These feature maps are then merged with the original feature maps to capture contextual details across different scales. Through this process, the CAA module enhances the ability of the model to capture long-range dependencies, thereby improving the performance of instrument object detection.

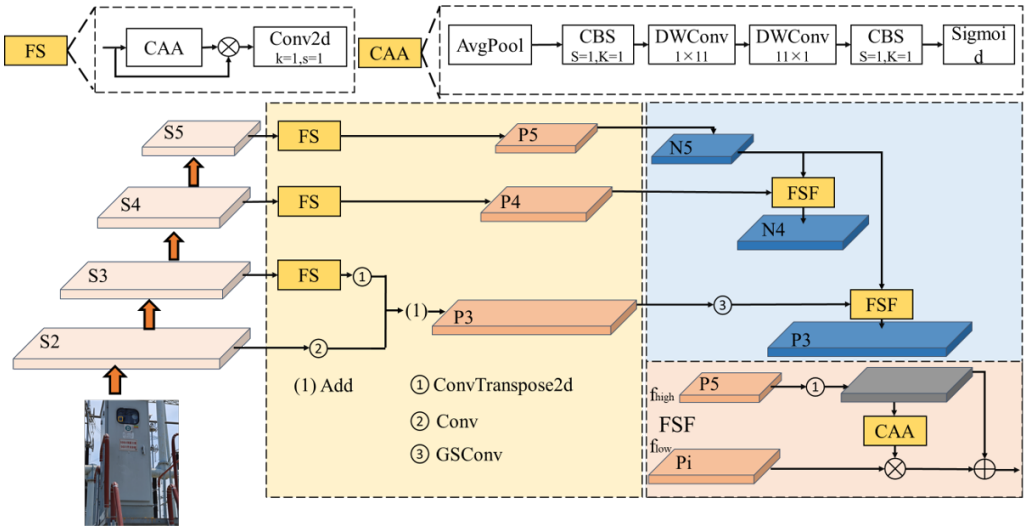


Figure 4. The Framework of Semantic Screening Feature Pyramid Network.

Subsequently, while retaining the feature fusion mechanism, we use only the P3, P4, and P5 channels to generate the output feature maps, and we upward fuse the P2 channel features, which contain positional information, into the P3 channel. This design ensures a thorough fusion of feature information, particularly preserving crucial positional information. Simultaneously, decreasing channel counts simplifies the network architecture, lowering computational complexity and resource consumption.

Finally, through the designed Feature Selection Fusion Module (FSF), High-level features act as weights, filtering and integrating essential information from the low-level features. Specifically, given an input high-level feature $f_{high} \in R^{C\times H\times W}$ and a low-level feature $f_{low} \in R^{C\times H\times W}$, the high-level feature is upsampled with a transposed convolution that uses a 3 × 3 kernel and a stride of 2, adjusting it to the feature size $f_{\widehat{high}} \in R^{C\times 2H\times 2W}$. Then, bilinear interpolation is applied to align the dimensions of high- and low-level features. The CAA module then transforms the high-level features into attention weights, filtering the low-level features. Once the dimensions are unified, the refined low-level features are integrated with the high-level features to enhance the overall feature representation. The specific process of feature selection and fusion is described by Formula (5).

$$f_{out} = f_{low} * CAA\Big(BL\Big(T_Conv(f_{high})\Big)\Big) + BL(T_Conv(f_{high})) \quad (5)$$

where T_Conv represents a transposed convolution.

3.4. Detection Head

The Head output represents the final stage of the detection process. It takes the compressed and fused feature maps from the Neck for classification and prediction. The Head includes three detection branches, each designed to handle feature maps of different scales from the Neck, enabling precise detection of objects of varying sizes. It segments these maps into grids of 80 × 80, 40 × 40, and 20 × 20, predicting bounding box coordinates,

confidence scores, and class probabilities for each grid. This method provides the location, presence, and category of objects, facilitating efficient real-time detection.

During model training, a comprehensive loss function assesses the performance of the detection predictions. This function includes three components: confidence loss L_{obj}, classification loss L_{cls}, and bounding box regression loss L_{box}. This design ensures a comprehensive assessment of bounding box and class predictions, supporting end-to-end model optimization. The confidence loss L_{obj} measures the reliability of bounding boxes, while the classification loss L_{cls} gauges the discrepancy between true and predicted classes in each grid. Both are calculated using the BCEWithLogitsLoss function, as shown in Formula (6):

$$Loss = -\frac{1}{n}\sum_{i}^{n}[y_i \cdot log(\sigma(x_i)) + (1 - y_i) \cdot log(1 - \sigma(x_i))] \tag{6}$$

$$\sigma(a) = \frac{1}{1 + exp(-a)} \tag{7}$$

where n denotes the number of samples. y_i represents the true class of the sample. x_i denotes the predicted class probability value. σ represents the sigmoid function.

The bounding box loss function L_{box} evaluates the positional difference between the ground truth box and the predicted box, using the CIOU function. This loss function not only considers the diagonal distance between the boxes but also takes into account the aspect ratio of the bounding boxes, providing a more accurate position estimate. The calculation is as follows:

$$Loss_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha \tag{8}$$

$$v = \frac{4}{\pi^2}(\arctan\frac{w^{gt}}{h^{gt}} - \arctan\frac{w}{h})^2 \tag{9}$$

$$\alpha = \frac{v}{(1 - IOU) + v} \tag{10}$$

where b and b^{gt}, respectively, represent the center coordinates of the two bounding boxes; c represents the diagonal length of the smallest bounding box enclosing both bounding boxes, and ρ^2 denotes the squared Euclidean distance between these two points. α represents the weighting parameter, and v is used to measure the similarity of the aspect ratios; $\frac{w^{gt}}{h^{gt}}$ and $\frac{w}{h}$ denote the width and height ratios of the ground truth and predicted boxes, respectively.

4. Experiments Results

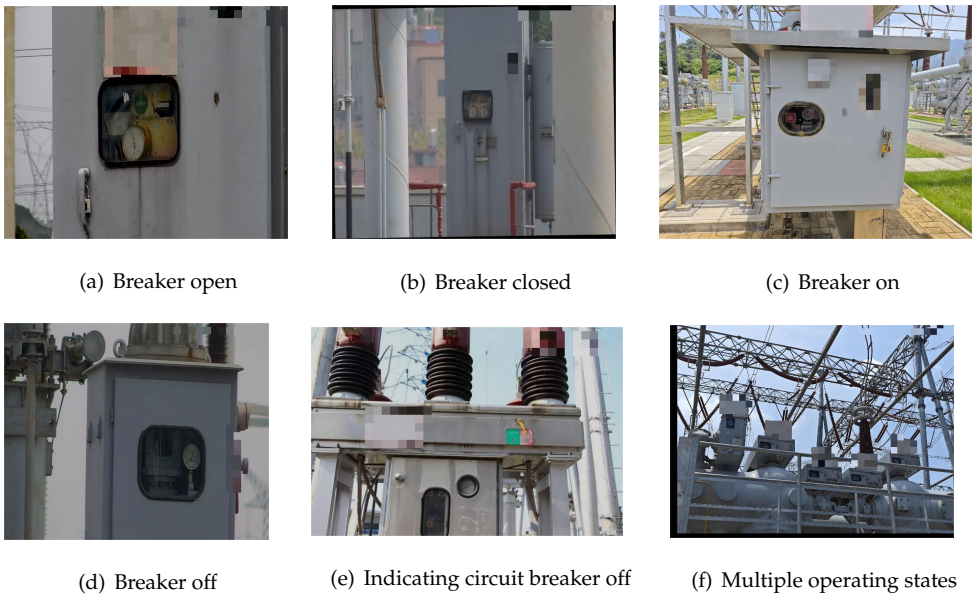
4.1. Experimental Dataset

Since there is no publicly available dataset suitable for detection at present of the operational status of circuit breaker panels, this paper constructs a custom dataset for power equipment to validate the proposed method's detection performance.

1. Data Collection: The dataset consists of 1230 images from substation sites, categorized into two types. The first type includes high-resolution images (5184 × 3888 pixels) taken by field inspection personnel, typically from horizontal or inclined angles. The second type consists of power grid monitoring images, which have either 1920 × 1080 or 1280 × 720 pixel resolutions. These images are typically captured from a top-down view.
2. Data Cleaning: First, we removed images that lacked detectable objects, were severely occluded, or excessively blurred, as they hindered model training. To boost the robustness and adaptability of the training process, and to improve performance across various conditions—including noise and environmental changes—while avoiding overfitting, we applied data augmentation techniques. These included blurring, rotation, and adding Gaussian noise to the cleaned images to expand the dataset. After augmentation, the circuit breaker operation dataset totaled 4920 images. Finally, due to the varying pixel sizes resulting from different image acquisition methods, we

normalized image sizes before model training. YOLOv8 generally uses 640 × 640 pixel images as input. However, the images gathered from power operation scenarios generally have sizes of 5184 × 3888, 1280 × 720, or 1920 × 1080 pixels. In this experiment, we resized these images to 1280 × 800 pixels to expedite the model training process. The computational complexity of deep learning models is closely related to the input image size. The larger the image, the greater the computation required for convolution operations and feature extraction at each layer of the network. When larger images (e.g., 5184 × 3888 pixels) are downscaled to smaller dimensions (e.g., 1280 × 800 pixels), the total number of pixels fed into the network significantly decreases, which directly reduces computational load and noticeably accelerates training speed.

3. Data Annotation: For the detection tasks related to the operating status of circuit breakers, we annotated six types of objects in the power equipment dataset using LabelImg 1.8.6 software. The annotated objects include open Figure 5a, closed Figure 5b, on Figure 5c, off Figure 5d, indicate_off Figure 5e and indicate_on Figure 5f.



(a) Breaker open (b) Breaker closed (c) Breaker on

(d) Breaker off (e) Indicating circuit breaker off (f) Multiple operating states

Figure 5. Power Equipment Dataset. Due to confidentiality requirements, the above images have been desensitized.

4.2. Experimental Details

The experiments were conducted on a system configured with an Intel® Core™ i9-9900K CPU and a GeForce RTX 2080 Ti 11G GPU, running on a Windows 10 operating system. The deep learning framework used was Pytorch 1.8.0, with CUDA version 10.0.

To validate the effectiveness of the proposed method in power operation scenarios, the model was trained using a dataset of 4,920 images. The YOLOv8 architecture was used as the baseline model, and network modifications and optimizations were applied. No pre-trained weights were used during training. The training process was set to 300 epochs, with a batch size of 4. The initial learning rate was 0.01, with a learning rate momentum of 0.9, and the weight decay coefficient was set to 0.0005.

4.3. Evaluation Metrics

To evaluate the effectiveness of the proposed detection algorithm, in this experiment, we evaluate the model using common deep learning metrics such as Precision (P), Recall (R), mAP, and FPS. Additionally, GFLOPs and the total number of parameters are employed to measure the computational complexity and model size.

Precision and Recall are commonly used metrics in detection and classification tasks, and their calculation formulas are as follows:

$$Precison = \frac{TP}{FP + TP} \tag{11}$$

$$Recall = \frac{TP}{FN + TP} \tag{12}$$

where FP refers to negative samples that were incorrectly classified as positive, TP indicates the correctly identified positive samples, and FN denotes the positive samples that were overlooked.

AP and mAP serve as key indicators for assessing model performance. AP calculates the Average Precision for an individual object category, whereas mAP measures the Mean Average Precision across all categories. The calculation formulas are as follows:

$$AP = \int_0^1 P(R)d(R) \tag{13}$$

$$mAP = \frac{1}{M}\sum_{j=1}^{M} AP_j \tag{14}$$

M indicates the complete set of object categories.

The following formula outlines how $FLOPs$ and Parameters are calculated for detection algorithms with L convolutional layers:

$$FLOPs \approx \sum_{l=1}^{L} FLOPs^l = 2 \times \sum_{l=1}^{L} \frac{(K^l)^2 \cdot H^l \cdot W^l \cdot C_{in}^l \cdot C_{out}^l}{(S^l)^2} \tag{15}$$

$$Parameter = \sum_{l=1}^{L} Parameter^l \approx \sum_{l=1}^{L} (K^l)^2 \cdot C_{in}^l \cdot C_{out}^l \tag{16}$$

where K^l indicates the kernel size of the l-th convolutional layer. The height and width of the input feature map are represented by H^l and W^l, respectively. The variables c_{in}^l and c_{out}^l denote the number of input and output feature channels for the l-th convolutional layer, while S^l signifies the stride of the convolutional kernel.

FPS measures the performance of a model by indicating how many images are processed per second. The formula for calculating FPS is as follows:

$$FPS = \frac{T_{total}}{N_{images}} \tag{17}$$

In this formula, T_{total} stands for the model's total inference time, and N_{images} represents the number of images processed during this period.

4.4. Ablation Experiment

To assess the impact of the enhancements in the DLCH-YOLO algorithm for detecting circuit breaker operation statuses, we performed ablation experiments using the power equipment dataset. These experiments evaluate how each individual improvement and their combinations affect the performance of the model. We used mAP@50, mAP@50-95, GFlops, and Parameters as evaluation metrics to analyze the experimental results. The table below shows the average values of each metric under different conditions.

According to the ablation experiment results presented in Table 1, for Model 1, we introduced GSConv into the backbone network of the baseline model YOLOv8n. This not only improved detection accuracy but also reduced the number of parameters and computational load, resulting in a 1.1% increase in mAP and a reduction of 0.3G in the

number of parameters. For Model 2, the improved C2f_DLKA module, with its deformable large kernel attention, demonstrated strong adaptability to multi-scale features, achieving a 3% increase in mAP, although GFlops increased by 0.6. For Model 3, the introduction of the improved Semantic Screening Feature Pyramid Network (SSFPN) in the neck improved detection accuracy by 3.8%. Although the number of parameters increased by 1.4G, the computational load decreased by 0.8M. For Model 4, when both the GSConv and C2f_DLKA modules were introduced, mAP increased by 2.8%, and the number of parameters was reduced by 0.4G compared to Model 2, which only used the C2f_DLKA module. For Model 5, with the introduction of both the GSConv and SSFPN modules, mAP reached 91.2%, which is a 4.1% improvement over the baseline model, and the number of parameters was reduced by 0.3G compared to Model 3, which only used the SSFPN module. For Model 6, when both the C2f_DLKA module and SSFPN module were introduced, mAP increased by 3.6%. The final model proposed in this paper incorporates improvements from GSConv, C2f_DLKA, and SSFPN modules, achieving the highest mAP of 91.8%, a 4.7% increase over the baseline model, with only a 1.5% increase in parameters. The results show that the algorithm achieves more precise and comprehensive object identification, with fewer false positives and missed detections.

Table 1. Ablation experiments for the DLCH-YOLO algorithm using the Power Equipment dataset.

Model	GSConv	C2f_DLKA	SSFPN	mAP@50 (%)	mAP@50-95 (%)	GFlops (G)	Parament (M)
YOLOv8n	-	-	-	87.1	56.9	8.1	3.0
1	✓			88.2	59.5	7.7	2.8
2		✓		90.1	60.4	8.7	3.7
3			✓	90.9	59.9	9.5	2.2
4	✓	✓		89.9	60.0	8.3	3.5
5	✓		✓	91.2	61.6	9.2	2.0
6		✓	✓	90.7	59.9	10.1	2.8
DLCH-YOLO	✓	✓	✓	91.8	61.6	9.6	2.6

C2f_DLKA expands the receptive field through large convolutional kernels, enhancing the understanding of contextual information for the model, which helps to better distinguish between targets and backgrounds in complex scenarios. SSFPN, through its feature filtering mechanism, enables precise target localization in challenging backgrounds, significantly reducing the false positives and missed detections caused by background interference. GSConv, combining standard and Depthwise Separable Convolutions, maintains performance similar to dense convolutions while improving the model's inference speed. Both the large convolutional kernels in C2f_DLKA and SSFPN are implemented using an equivalent structure, expanding the receptive field significantly with minimal additional computation and parameters. By introducing GSConv, the detection performance for the model is further enhanced with only a slight increase in parameters. In high-risk power operation scenarios, detection accuracy is often more critical than speed.

4.5. Comparison Experiments

To comprehensively evaluate our improvements, this study comparison experiments will be conducted from various perspectives. The first step involves comparing the effects of different attention modules and their application at various positions, with results presented in Table 2. This will reveal the differences in how each attention module enhances model performance. Next, we will compare the improved backbone networks, with relevant results shown in Table 3, to analyze the impact of different backbone networks on detection performance. Subsequently, an analysis of mainstream feature fusion methods will be conducted, with results provided in Table 4 to understand the advantages and disadvantages of various feature fusion strategies. Finally, to assess the overall effectiveness of the proposed methods, we will evaluate them against several existing approaches. The comparison results are shown in Table 5. These comparison

experiments aim to thoroughly assess the advantages and practical application effectiveness of the proposed improvements.

Table 2 presents the results of the comparison experiments for the attention mechanism. The experiments indicate that on the electrical equipment dataset, the model with the "DLKA improving the C2f module" outperforms other configurations in terms of mAP@0.5 and mAP@0.5:0.95. Specifically, compared to the model without the DLKA module, this improvement increases mAP by 3% and 3.5%, respectively; compared to the model with the DLKA module inserted at the end of the backbone (network layer 10), it improves mAP by 8.3% and 5.7%, respectively. Although the "DLKA module improving the C2f module" results in a minor rise in computational demands and the number of parameters, and a minor reduction in FPS, it significantly enhances detection accuracy, reflecting a high cost-effectiveness ratio.

To evaluate the DLKA mechanism in detection tasks, we compared its performance with that of four mainstream attention modules (CA [47], SE [48], CBAM [49], EMA [50]). All four modules were used to improve the C2f module in the same manner, with the experimental results shown in Table 2. The results demonstrate that the DLKA module excels in improving model detection performance. Specifically, compared to CA, SE, CBAM, and EMA modules, the DLKA module achieves mAP@0.5 and mAP@0.5:0.95 that are higher by 5.4%, 9.6%, 1.6%, 3.4% and 5.8%, 7.6%, 3.5%, 4.7%, respectively. Additionally, the DLKA mechanism increases computational load and parameter count by only 0.4G while maintaining strong performance in FPS. This indicates that the DLKA mechanism not only improves detection accuracy in electrical equipment tasks but also maintains a good balance in computational efficiency.

Table 2. Results of the attention mechanism comparison experiment.

Method	P (%)	R (%)	mAP@50 (%)	mAP@50-95 (%)	GFlops (G)	Parament (M)
YOLOv8n	97.7	80.9	87.1	56.9	8.1	3.00
+DLKA attention (C2f)	98.4	80.1	90.1	60.4	8.6	3.70
+DLKA attention (10)	98.1	80.7	81.8	54.7	9.4	4.60
+DAttention (C2f)	97.8	79.8	84.8	54.8	8.1	3.07
+CA (C2f)	95.8	79.9	84.7	54.6	8.1	3.00
+SE (C2f)	95.8	79.0	80.5	52.8	8.1	3.00
+CBAM (C2f)	94.2	78.0	88.5	56.9	8.1	3.02
+EMA (C2f)	93.1	79.2	86.7	55.7	8.1	3.00

The backbone network, which combines GSConv and the DLKA module, effectively reduces redundant computations in feature maps while enhancing dashboard localization and detection through contextual information. Table 3 presents the performance comparison of different backbone networks, including GSConv+DLKA, YOLOv8n, FasterNet [51], MobileNetV4 [52], and HGNetV2 [53], on the custom dataset. The results show that although our improved backbone network slightly increases the number of parameters, it demonstrates excellent detection accuracy, being only 0.4% lower than MobileNetV4, while reducing the parameter count by 14.2G, showcasing superior performance.

Table 3. Results of the backbone network comparison experiment.

Method	P (%)	R (%)	mAP@50 (%)	mAP@50-95 (%)	GFlops (G)	Parament (M)
Base	97.7	80.9	87.1	56.9	8.1	3.00
FasterNet	97.0	80.4	81.3	53.7	10.7	4.17
MobileNetV4	96.1	79.6	90.3	59.1	22.5	5.70
HgNetV2	94.0	78.1	85.7	52.7	6.9	2.35
GSConv + DLKA	96.6	80.2	89.9	60.0	8.3	3.50

Additionally, we compared SSFPN with other feature fusion networks such as PAN [30], AFPN [54], BiFPN [31], Slimneck [41], and HSFPN [42], as shown in Table 4. Under similar parameter and computation conditions, SSFPN achieved the highest detection accuracy in the circuit breaker operational status detection task. Through an efficient network structure design, SSFPN not only reduced computational complexity and resource consumption but also better preserved the complex edges and detailed information of defect objects, further enhancing detection performance.

Table 4. Results of the feature fusion evaluation.

Method	P (%)	R (%)	mAP@50 (%)	mAP@50-95 (%)	GFlops (G)	Parament (M)
Base + PAN	97.7	80.9	87.1	56.9	8.1	3.00
Base + AFAN	87.9	75.7	85.6	53.5	8.4	2.59
Base + BiFAN	96.7	80.2	81.6	54.4	7.1	1.99
Base + Slimneck	96.5	79.6	80.8	55.3	7.3	2.79
Base + HSFAN	97.4	80.8	88.0	59.2	10.2	2.43
SSFPN	98.3	81.1	90.9	59.9	9.5	2.20

The results of the ablation study validate the effectiveness of the various network components in DLCH-YOLO. Furthermore, we conducted a comprehensive comparison of DLCH-YOLO with other algorithms, including classical models like Faster-RCNN [10] and Cascade-RCNN [55], as well as widely adopted models such as YOLOv5, YOLOv8, YOLOv10, and RT-DETR-L [53]. Compared to classical algorithms (Faster-RCNN, Cascade-RCNN), DLCH-YOLO not only significantly reduced parameters and computational complexity but also achieved superior performance in terms of mAP@0.5, mAP@0.5:0.95, and FPS. When compared with lightweight models such as YOLOv5n, YOLOv8n, and YOLOv10, DLCH-YOLO showed notable improvements in mAP@0.5 and mAP@0.5:0.95, despite an increase in parameters by 5.4G, 1.5G, and 3.1G, respectively. Specifically, compared to the model with the lowest mAP, YOLOv5n, DLCH-YOLO achieved an 8.6% increase in mAP. When compared to YOLOv8s, YOLOv10s, and RT-DETR, DLCH-YOLO maintained a lower parameter count and computational load while still outperforming them in mAP@0.5:0.95, mAP@0.5 and FPS.

Table 5. Experimental comparison between DLCH-YOLO and SOTA models.

Method	mAP@50 (%)	mAP@50-95 (%)	GFlops (G)	Parament (M)	FPS
Faster-RCNN	71.6	39.2	174	41.37	16.0
Cascade-RCNN	70.5	43.1	201	69.23	12.9
YOLOv5n	83.2	55.0	4.2	1.90	47.2
YOLOv8n	87.1	56.9	8.1	3.01	146.0
YOLOv8s	89.8	60.5	28.4	22.50	127.4
YOLOv10n	88.8	55.3	6.5	2.26	115.3
YOLOv10s	90.0	60.3	21.6	7.20	108.8
RT-DETR-L	79.0	51.5	105	29.70	17.6
DLCH-YOLO	86.7	55.7	9.6	2.60	72.8

Notably, while DLCH-YOLO utilizes large kernel convolutions that add complexity to the model, it achieves a performance of 72.8 FPS, which is substantially higher than that of two-stage algorithms and surpasses the real-time detection standard of over 60 FPS. This demonstrates that DLCH-YOLO improves detection accuracy while maintaining high speed and efficiency. In the field of power system monitoring, algorithms need to process images and make decisions in a very short time. Too high a delay may cause the system to fail to respond in time, thus affecting the decision effect. The actual performance of a model during deployment largely depends on the hardware platform. The computational

power and memory constraints of different devices directly affect the inference speed and load-handling capabilities of the model.

4.6. Visualization Results

Figure 6 presents the DLCH-YOLO recognition results for six circuit breaker states from the custom electrical equipment dataset. Each detection box provides both the position and category of the object, along with a confidence score that reflects the probability of the model correctly identifying the category. The following Algorithm 1 presents the detection results.

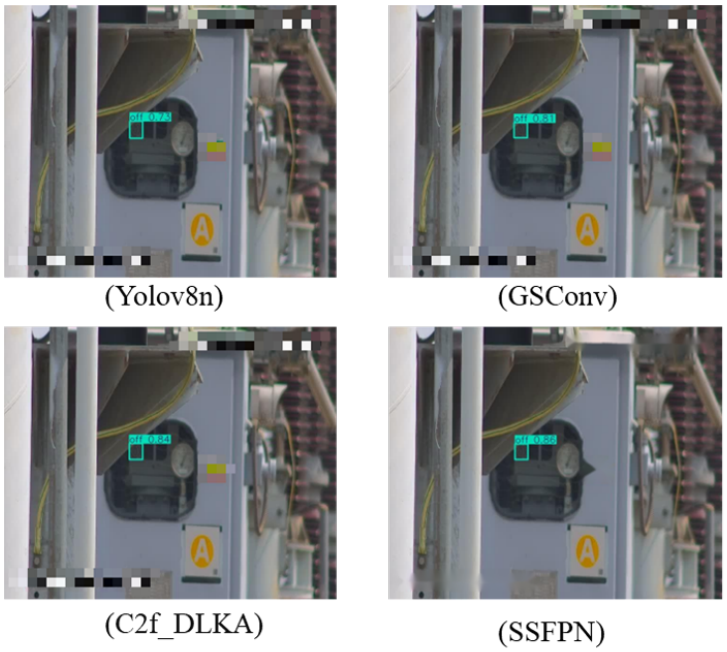


Figure 6. Ablation visualization. All images have been desensitized.

Algorithm 1 Algorithm of Inference

Input: X^n test dataset; *Class*: all categories; α: Minimum confidence score (e.g., 0.5)
Output: *Bounding_boxes*; *Confidence*; *Class_labels*

for x in X^n **do**
  for *each_project* in *Class* **do**
    $decoded \leftarrow Feature_extraction(X)$
    $Bounding_boxes \leftarrow sigmoid(decoded)$
    $Confidence \leftarrow softmax(Bounding_boxes)$
    if $Confidence < \alpha$ **then**
      continue
    end if
  end for
end for
return Outputs

The data in the figure show that DLCH-YOLO accurately detects various circuit breaker states, such as open and close operations, with no missed detections or false positives. Notably, even with partial occlusion of the targets, DLCH-YOLO maintains high recognition accuracy and provides high confidence levels. These results not only validate the superior performance of DLCH-YOLO in electrical equipment status monitoring but also demonstrate its robustness in handling complex and partially occluded targets. These advantages suggest that DLCH-YOLO has significant potential and value in practical applications.

5. Conclusions

In power monitoring scenarios, the operating status of circuit breaker detection faces numerous challenges, such as varying object sizes, complex backgrounds, and diverse monitoring angles. The DLCH-YOLO algorithm proposed in this paper demonstrates exceptional performance and significant application potential in addressing these challenges. First, the C2f_DLKA module, based on deformable large kernel attention, enhances feature extraction by capturing long-range dependencies in images using large convolutional kernels. Second, the introduction of the SSFPN feature fusion network improves object feature extraction accuracy in complex backgrounds through semantic information filtering. Finally, the use of Grouped Spatial Convolution (GSConv) effectively reduces computational load. Experimental results show that DLCH-YOLO achieves a mAP of 91.8% on the custom power equipment dataset, a 4.7% improvement over YOLOv8n, and outperforms in detecting various object categories. The visualized results, as shown in Figure 7, clearly demonstrate the improved model's outstanding performance in detection accuracy, confidence scores, and handling of occluded objects. This algorithm not only offers high precision and adaptability but also meets the parameter constraints required for deployment on edge devices in deep learning applications. With a detection speed of 72.8 FPS, DLCH-YOLO offers an efficient and practical solution for real-time detection of circuit breaker statuses in resource-constrained industrial settings.

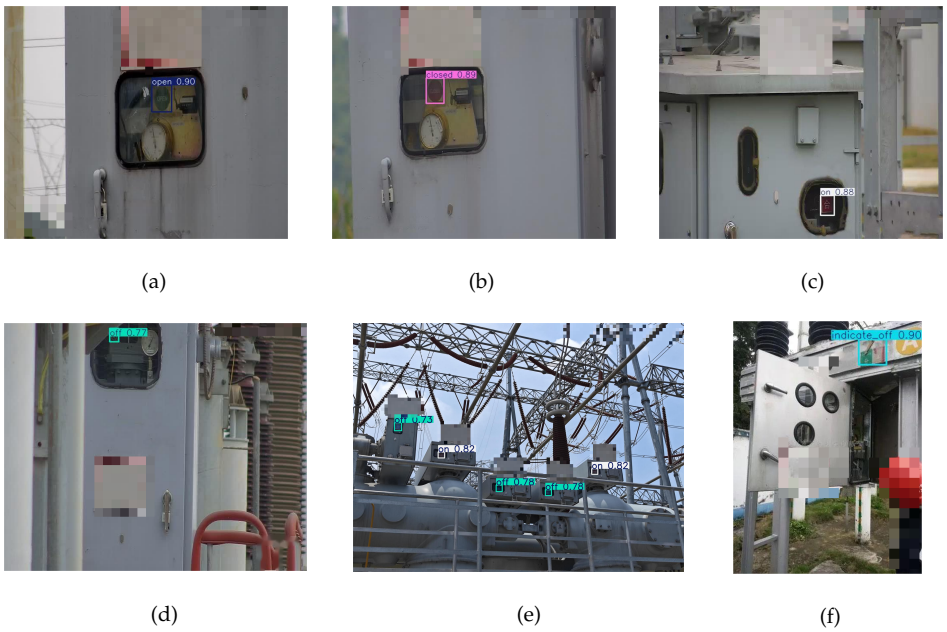


Figure 7. The detection results of the improved method on the power equipment dataset for identifying the working status of circuit breakers. (**a**–**e**) represent detection results from different monitoring perspectives, while (**f**) represents a complex background.

Moving forward, our work will focus on the following issues to further improve the performance and applicability of the model:

- Exploration of Diffusion Models: We will study diffusion models to generate more natural and realistic training samples for circuit breaker dashboards. This approach will help improve the detection accuracy and generalization ability of the model across all categories, enabling it to more effectively tackle complex and variable industrial environments. Additionally, we will explore how to further enrich the training dataset through data augmentation and synthetic samples to enhance the model's robustness.
- Model Lightweighting and Real-Time Detection: To ensure the feasibility of the algorithm in practical applications, we will focus on optimizing the lightweight design of the model, employing techniques such as model pruning and quantization to

reduce computational complexity. At the same time, we will investigate real-time detection technologies based on deep learning to meet the demands for efficient real-time detection in resource-constrained environments. This includes exploring methods for edge computing and distributed processing to enhance the inference speed and responsiveness.

Through sustained exploration in these areas, we aim to further enhance the performance of the model and increase its value in industrial applications.

Author Contributions: Conceptualization, R.S. and L.S.; methodology, R.S.; software, R.S. and L.C.; validation, R.S.; formal analysis, R.S.; investigation, R.S. and T.L.; resources, R.S. and L.C.; data curation, R.S. and L.C.; writing—original draft preparation, R.S.; writing—review and editing, R.S.; visualization, R.S.; supervision, L.S., F.Y. and T.L.; project administration, R.S.; funding acquisition, L.S., F.Y. and T.L. 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: Data are not available for privacy reasons.

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: Fan Yin was employed by 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.

References

1. Razi-Kazemi, A.A.; Niayesh, K. Condition monitoring of high voltage circuit breakers: Past to future. *IEEE Trans. Power Deliv.* **2020**, *36*, 740–750. [CrossRef]
2. Andruşcă, M.; Adam, M.; Irimia, F.D.; Baraboi, A. Aspects about the monitoring and diagnosis of high voltage circuit breakers. In Proceedings of the 2012 International Conference and Exposition on Electrical and Power Engineering, Iasi, Romania, 25–27 October 2012; IEEE: Piscataway, NJ, USA; pp. 154–158.
3. Li, S.; Li, J. Condition monitoring and diagnosis of power equipment: Review and prospective. *High Volt.* **2017**, *2*, 82–91. [CrossRef]
4. Kotsiopoulos, T.; Sarigiannidis, P.; Ioannidis, D.; Tzovaras, D. Machine learning and deep learning in smart manufacturing: The smart grid paradigm. *Comput. Sci. Rev.* **2021**, *40*, 100341. [CrossRef]
5. Liu, Y.; Liu, J.; Ke, Y. A detection and recognition system of pointer meters in substations based on computer vision. *Measurement* **2020**, *152*, 107333. [CrossRef]
6. 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]
7. Parihar, A.S.; Singh, K. A study on Retinex based method for image enhancement. In Proceedings of the 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 19–20 January 2018; IEEE: Piscataway, NJ, USA; pp. 619–624.
8. Pham, T.Q. Asymmetric recursive Gaussian filtering for space-variant artificial bokeh. In Proceedings of the 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 10–13 December 2018; IEEE: Piscataway, NJ, USA; pp. 1–8.
9. Sharif Razavian, A.; Azizpour, H.; Sullivan, J.; Carlsson, S. CNN features off-the-shelf: An astounding baseline for recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Columbus, OH, USA, 23–28 June 2014; pp. 806–813.
10. 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]
11. Khalyasmaa, A.I.; Senyuk, M.D.; Eroshenko, S.A. High-voltage circuit breakers technical state patterns recognition based on machine learning methods. *IEEE Trans. Power Deliv.* **2019**, *34*, 1747–1756. [CrossRef]
12. Ullah, I.; Khan, R.U.; Yang, F.; Wuttisittikulkij, L. Deep learning image-based defect detection in high voltage electrical equipment. *Energies* **2020**, *13*, 392. [CrossRef]
13. Zhang, J.; Liu, W.; Xu, S.; Zhang, X. Key point localization and recurrent neural network based water meter reading recognition. *Displays* **2022**, *74*, 102222. [CrossRef]

14. Waqar, M.; Waris, M.A.; Rashid, E.; Nida, N.; Nawaz, S.; Yousaf, M.H. Meter digit recognition via Faster R-CNN. In Proceedings of the 2019 International Conference on Robotics and Automation in Industry (ICRAI), Rawalpindi, Pakistan, 21–22 October 2019; IEEE: Piscataway, NJ, USA; pp. 1–5.
15. Peng, J.; Zhou, W.; Han, Y.; Li, M.; Liu, W. Deep learning-based autonomous real-time digital meter reading recognition method for natural scenes. *Measurement* **2023**, *222*, 113615. [CrossRef]
16. Liu, J.; Wang, C.; Xu, L.; Wang, M.; Hu, D.; Jin, W.; Li, Y. A Study of Electric Bicycle Lithium Battery Charging Monitoring Using CNN and BiLSTM Networks Model with NILM Method. *Electronics* **2024**, *13*, 3316. [CrossRef]
17. Meng, X.; Hu, G.; Liu, Z.; Wang, H.; Zhang, G.; Lin, H.; Sadabadi, M.S. Neural Network-Based Impedance Identification and Stability Analysis for Double-Sided Feeding Railway Systems. *IEEE Trans. Transp. Electrif.* **2024**, *1*, 1. [CrossRef]
18. Zou, L.; Wang, K.; Wang, X.; Zhang, J.; Li, R.; Wu, Z. Automatic recognition reading method of pointer meter based on YOLOv5-mr model. *Sensors* **2023**, *23*, 6644. [CrossRef] [PubMed]
19. Zhang, M.; Zhu, M.; Zhao, X. Recognition of high-risk scenarios in building construction based on image semantics. *J. Comput. Civ. Eng.* **2020**, *34*, 04020019. [CrossRef]
20. Arulprakash, E.; Aruldoss, M. A study on generic object detection with emphasis on future research directions. *J. King Saud Univ. Comput. Inf. Sci.* **2022**, *34*, 7347–7365. [CrossRef]
21. Zhang, Q.; Zhang, H.; Lu, X. Adaptive feature fusion for small object detection. *Appl. Sci.* **2022**, *12*, 11854. [CrossRef]
22. Chen, G.; Wang, H.; Chen, K.; Li, Z.; Song, Z.; Liu, Y.; Chen, W.; Knoll, A. A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. *IEEE Trans. Syst. Man Cybern. Syst.* **2020**, *52*, 936–953. [CrossRef]
23. Azad, R.; Niggemeier, L.; Hüttemann, M.; Kazerouni, A.; Aghdam, E.K.; Velichko, Y.; Bagci, U.; Merhof, D. Beyond self-attention: Deformable large kernel attention for medical image segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Seattle, WA, USA, 17–21 June 2024; pp. 1287–1297.
24. Amit, Y.; Felzenszwalb, P.; Girshick, R. Object detection. In *Computer Vision: A Reference Guide*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 875–883.
25. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [CrossRef]
26. Ji, D.; Zhang, W.; Zhao, Q.; Yang, W. Correction and pointer reading recognition of circular pointer meter. *Meas. Sci. Technol.* **2022**, *34*, 025406. [CrossRef]
27. Zuo, L.; He, P.; Zhang, C.; Zhang, Z. A robust approach to reading recognition of pointer meters based on improved mask-RCNN. *Neurocomputing* **2020**, *388*, 90–101. [CrossRef]
28. Wang, Y.; Dong, G. Target Detection of Pointer Instrument based on Deep Learning. *Highlights Sci. Eng. Technol.* **2022**, *24*, 182–190. [CrossRef]
29. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
30. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
31. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
32. Luo, Y.; Cao, X.; Zhang, J.; Guo, J.; Shen, H.; Wang, T.; Feng, Q. CE-FPN: Enhancing channel information for object detection. *Multimed. Tools Appl.* **2022**, *81*, 30685–30704. [CrossRef]
33. Huang, S.; Lu, Z.; Cheng, R.; He, C. Fapn: Feature-aligned pyramid network for dense image prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 19–25 June 2021; pp. 864–873.
34. 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, Glasgow, Scotland, 23–28 August 2020; Springer: Berlin/Heidelberg, Germany, 2020; pp. 213–229.
35. Vaswani, A. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5999–6099.
36. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 19–25 June 2021; pp. 10012–10022.
37. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. Swinir: Image restoration using swin transformer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Online, 19–25 June 2021; pp. 1833–1844.
38. Chen, Z.; Zhang, Y.; Gu, J.; Zhang, Y.; Kong, L.; Yuan, X. Cross aggregation transformer for image restoration. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 25478–25490.
39. Guo, M.H.; Lu, C.Z.; Liu, Z.N.; Cheng, M.M.; Hu, S.M. Visual attention network. *Comput. Vis. Media* **2023**, *9*, 733–752. [CrossRef]
40. Lau, K.W.; Po, L.M.; Rehman, Y.A.U. Large separable kernel attention: Rethinking the large kernel attention design in cnn. *Expert Syst. Appl.* **2024**, *236*, 121352. [CrossRef]
41. 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]

42. Chen, Y.; Zhang, C.; Chen, B.; Huang, Y.; Sun, Y.; Wang, C.; Fu, X.; Dai, Y.; Qin, F.; Peng, Y.; et al. Accurate leukocyte detection based on deformable-DETR and multi-level feature fusion for aiding diagnosis of blood diseases. *Comput. Biol. Med.* **2024**, *170*, 107917. [CrossRef]
43. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 764–773.
44. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
45. Long, H.; Chung, Y.; Liu, Z.; Bu, S. Object detection in aerial images using feature fusion deep networks. *IEEE Access* **2019**, *7*, 30980–30990. [CrossRef]
46. Cai, X.; Lai, Q.; Wang, Y.; Wang, W.; Sun, Z.; Yao, Y. Poly kernel inception network for remote sensing detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 27706–27716.
47. Hou, Q.; Zhou, D.; Feng, J. Coordinate attention for efficient mobile network design. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 19–25 June 2021; pp. 13713–13722.
48. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 18–22 June 2018; pp. 7132–7141.
49. 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.
50. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
51. Chen, J.; Kao, S.h.; He, H.; Zhuo, W.; Wen, S.; Lee, C.H.; Chan, S.H.G. Run, don't walk: Chasing higher FLOPS for faster neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 12021–12031.
52. Qin, D.; Leichner, C.; Delakis, M.; Fornoni, M.; Luo, S.; Yang, F.; Wang, W.; Banbury, C.; Ye, C.; Akin, B.; et al. MobileNetV4-Universal Models for the Mobile Ecosystem. *arXiv* **2024**, arXiv:2404.10518.
53. 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 Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
54. Yang, G.; Lei, J.; Zhu, Z.; Cheng, S.; Feng, Z.; Liang, R. AFPN: Asymptotic feature pyramid network for object detection. In Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Maui, HI, USA, 1–4 October 2023; IEEE: Piscataway, NJ, USA; pp. 2184–2189.
55. Cai, Z.; Vasconcelos, N. Cascade r-cnn: Delving into high quality object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake, UT, USA, 18–22 June 2018; pp. 6154–6162.

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.