

Article

TL-YOLO: Foreign-Object Detection on Power Transmission Line Based on Improved Yolov8

Yeqin Shao ^{1,*}, Ruwei Zhang ², Chang Lv ¹, Zexing Luo ¹ and Meiqin Che ¹¹ School of Transportation and Civil Engineering, Nantong University, Nantong 226019, China;

dragonquest@stmail.ntu.edu.cn (C.L.); 2233320026@stmail.ntu.edu.cn (Z.L.); meiqin.che@ntu.edu.cn (M.C.)

² School of Electrical Engineering, Nantong University, Nantong 226004, China; 2111310006@stmail.ntu.edu.cn

* Correspondence: hnsyk@ntu.edu.cn

Abstract: Foreign objects on power transmission lines carry a significant risk of triggering large-scale power interruptions which may have serious consequences for daily life if they are not detected and handled in time. To accurately detect foreign objects on power transmission lines, this paper proposes a TL-Yolo method based on the Yolov8 framework. Firstly, we design a full-dimensional dynamic convolution (ODConv) module as a backbone network to enhance the feature extraction capability, thus retaining richer semantic content and important visual features. Secondly, we present a feature fusion framework combining a weighted bidirectional feature pyramid network (BiFPN) and multiscale attention (MSA) module to mitigate the degradation effect of multiscale feature representation in the fusion process, and efficiently capture the high-level feature information and the core visual elements. Thirdly, we utilize a lightweight GSConv cross-stage partial network (GSCSP) to facilitate efficient cross-level feature fusion, significantly reducing the complexity and computation of the model. Finally, we employ the adaptive training sample selection (ATSS) strategy to balance the positive and negative samples, and dynamically adjust the selection process of the training samples according to the current state and performance of the model, thus effectively reducing the object misdetection and omission. The experimental results show that the average detection accuracy of the TL-Yolo method reaches 91.30%, which is 4.20% higher than that of the Yolov8 method. Meanwhile, the precision and recall metrics of our method are 4.64% and 3.53% higher than those of Yolov8. The visualization results also show the superior detection performance of the TL-Yolo algorithm in real scenes. Compared with the state-of-the-art methods, our method achieves higher accuracy and speed in the detection of foreign objects on power transmission lines.

Keywords: power transmission line; foreign-object detection; Yolov8; attention mechanism; feature fusion



Citation: Shao, Y.; Zhang, R.; Lv, C.; Luo, Z.; Che, M. TL-YOLO: Foreign-Object Detection on Power Transmission Line Based on Improved Yolov8. *Electronics* **2024**, *13*, 1543. <https://doi.org/10.3390/electronics13081543>

Academic Editor: Davide Astolfi

Received: 1 April 2024

Revised: 12 April 2024

Accepted: 16 April 2024

Published: 18 April 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

Transmission lines, as critical conduits for electric energy in power grid system, are negatively affected by foreign objects. The secure and reliable transmission of electric power requires the exclusion of foreign objects along these lines. With the accelerated construction of China's power grid and the increasingly complex and dense erection of transmission lines [1], traditional manual inspections are proving inadequate due to their operation risks and huge workload. In the real environment, the traditional manual inspection is limited by safety risks and efficiency problems. Indeed, devices like drones somehow guarantee operation safety and work efficiency, but on-board algorithms lack real-time detection capabilities and unbiased training data support, which greatly hinder their adaption to complex situations [2]. Therefore, more intelligent detection methods are required for efficiently combining the inspection and detection tasks, to meet the actual needs of the power grid construction [3].

1.1. Traditional Foreign Objects Detection Methods

To detect foreign objects on transmission lines, Li et al. [4] provided a comprehensive overview of Uncrewed Aerial Vehicle (UAV) inspection applications, highlighting their potential for revolutionizing inspection processes. The UAV inspection provides a more intelligent and efficient way than before. Furthermore, Li et al. [5] utilized a SWOT-PEST analysis to scrutinize the opportunities and challenges in globally promoting the standardized UAV inspection protocols in China. Jiang et al. [6] innovated an oblique photogrammetry system for outdoor data acquisition using UAVs, featuring a cost-effective dual-camera setup and a dual-frequency GNSS receiver. Zhong et al. [7] focused on the deployment of multi-wing UAVs in transmission line inspection, noting their agility and vast coverage capabilities, which markedly diminish inspection duration. Jin et al. [8] refined the OTSU method for image background segmentation, differentiating the foreground and background, and then employed the gradient operators and Hough transform algorithms to precisely identify the transmission line contours. Wang et al. [9] proposed an innovative method for transmission line extraction and foreign body detection. Based on line detection and the parallel characteristics of transmission lines, this method combines image processing and computer vision technology to realize the automatic detection of foreign objects in transmission lines. Hayal et al. [10] investigated and improved a relay-assisted free-space optical (FSO) optical wireless communication (OWC) system for unmanned aerial vehicles (UAVs) under the influence of pointing error (PE) and atmospheric turbulence (AT). It is proposed to incorporate UAVs as interference-assisted mobile relays in conventional FSO (CFSO) relay-assisted systems in order to improve the performance of pointing errors through AT.

1.2. Deep-Learning-Based Foreign Objects Detection Methods

With the rapid advancement of high-performance computation and artificial intelligence, the intelligent and automated detection of foreign objects on transmission lines has emerged as a significant area of research. Zhang et al. [11] proposed the RCNN4SPTL, a specialized deep-learning network for identifying stationary foreign objects on transmission lines. The RCNN4SPTL leverages a Region Proposal Network (RPN) to produce candidate frames that precisely align with the dimensions of such objects. Wang et al. [12] conducted a comprehensive analysis of the Single-Stage Detector (SSD) in detecting foreign objects. The SSD's performance was assessed against traditional detection methods like DPM and Faster R-CNN, demonstrating its effectiveness. Dou et al. [13] addressed the challenge of limited or unbalanced data in foreign object detection by proposing a data augmentation technique utilizing Generative Neural Networks (GANs) which can create high-quality synthetic samples to enhance sample diversity and recognition of rare categories. Li et al. [14] developed a regional convolutional neural network algorithm with an emphasis on region-of-interest mining for the automated detection of specific foreign objects, achieving higher average accuracy and F1 scores than conventional methods. Xia et al. [15] investigated the adaptation of the YOLOv3 network by fine-tuning the pre-trained model parameters on a large dataset. Qiu et al. [16] selected the YOLOv4-tiny network for its lightweight structure and rapid processing, enabling real-time detection in limited-resource settings. Shen et al. [17] introduced an innovative neural network architecture, termed TLFODNet (Transmission Line Foreign Object Detection Net) which has been specifically designed for the detection of foreign objects on transmission lines. This new structural approach has demonstrably enhanced the performance and expedited the speed of detection. Wu [18] further advanced the field by refining the YOLOX model. This approach augments the YOLOX detection framework by incorporating Atrous Spatial Pyramid Pooling (ASPP) to heighten sensitivity across varying scales of foreign objects, integrating a convolutional block attention module to bolster model precision, and employing GIoU loss to optimize performance. Yu et al. [19] developed a composite detection technique that merges multi-network feature fusion with a Random Forest (RF) algorithm. This method applies Otsu's binarization thresholding coupled with morphological processes to delineate the target area of foreign objects subse-

quently amalgamated using a tandem fusion strategy. Li et al. [20] proposed a lightweight model tailored for embedded devices in the detection of foreign objects on transmission lines. Mobilenetv2 is utilized as the foundational framework instead of Darknet-53, and leveraging depth-separable convolutions as opposed to standard 3×3 convolutions. This leads to a substantial reduction in network parameter size. Li et al. [21] introduced the DF-YOLO that was refined based on YOLOv7-Tiny by using the Focal-DIoU loss function. In addition, this algorithm further combines the Deformable Convolution Networks (DCN) module and the SimAM attention mechanism to enhance its performance. Wang et al. [22] unveiled an augmented detection model for foreign objects on transmission lines based on Yolov8. This model incorporates a global attention module into the backbone architecture to concentrate on obscured foreign objects. The SPPCSPC module supplants the SPPF module to amplify the model's capacity for multi-scale feature extraction. Zhang et al. [23] provided an improved YOLOv5 technique for detecting foreign objects in transmission lines. The method first reduces computation and memory consumption by introducing the RepConv structure, and then further improves the detection accuracy and speed of the model by embedding the C2F structure. Finally, the method further optimizes the neural network through the Meta-ACON activation function. Liu et al. [24] propose a deep learning-assisted object detection method, YOLO-CSM, combining two attentional mechanisms (Swin transformer and CBAM) and an additional detection layer. The proposed model can efficiently capture global information and key visual features, and improve its ability to identify small defects and distant objects in the field of view. Tang et al. [25] proposed an improved foreign object detection method (ST2Rep-YOLOX) based on Swin Transformer V2 and YOLOX. The feature extraction layer ST2CSP constructed in the original backbone network using Swin Transformer V2 is used to extract global and local features, and the Hybrid Spatial Pyramid Pool (HSPP) is designed to expand the receptive field and retain more feature information. Yu et al. [26] proposed a deep learning-based foreign object recognition algorithm for unmanned inspection of transmission lines. The algorithm is based on the YOLOv7 algorithm, combined with hyperparameter optimization based on Genetic Algorithm (GA) and Space-Profile Depth (SPD) convolution, to accomplish foreign object recognition on transmission line UAV images.

Of all the algorithms above, the convolution neural networks exhibit inadequate feature extraction capabilities for irregular and small foreign objects on transmission lines, leading to inaccurate detection and false positives for small targets. Moreover, existing methodologies inadequately address the efficiency of foreign object detection under complex environments, resulting in subpar generalization capabilities. To address the aforementioned issues, this paper presents TL-Yolo to detect foreign objects on transmission lines by refining Yolov8. The contributions of this work are as follows:

1. Adopting Omni-Dimensional Dynamic Convolution (ODConv) as the feature extracting module to capture richer semantic information.
2. Designing an improved weighted bidirectional feature pyramid module with a Multi-Scale Attention (MSA) mechanism to fuse multiscale features and thereby mitigate representation degradation in fusion operation.
3. Using a lightweight cross-platform partial network, the GSConv Cross Stage Partial Network (GSCSP), to facilitate effective cross-layer feature fusion while markedly diminishing model complexity.
4. Employing an Adaptive Training Sample Selection (ATSS) strategy to equilibrate the distribution of positive and negative samples, thereby accelerating algorithm convergence.

2. Materials and Methods

2.1. Yolov8 Algorithm

Yolov8 is a prominent algorithm in the object detection realm. Its backbone network, Darknet-53, is composed of repeated convolutional modules and residual blocks that are sequentially stacked four times. The use of C2f configuration in the residual block improves the gradient flow information of the model through more cross-layer connectivity and

additional split operations. This C2f design further augments the model’s gradient flow through increased cross-layer connectivity and introduces a stack of 3, 6, 6, and three convolutional residual blocks within the backbone to bolster feature extraction capabilities. In the model’s neck part, the Yolov8 follows the structure of the PANet [27], executing a cross-layer fusion via up-sampling and down-sampling to ensure comprehensive integration of contextual feature information. In the model’s head part, it adopts the decoupled-head structure, effectively bifurcating the classification and detection heads to independently extract categorical and locational features. Subsequently, each branch employs a singular 1×1 convolutional layer to fulfill the tasks of classification and localization. The structure of the Yolov8 network is shown in Figure 1.

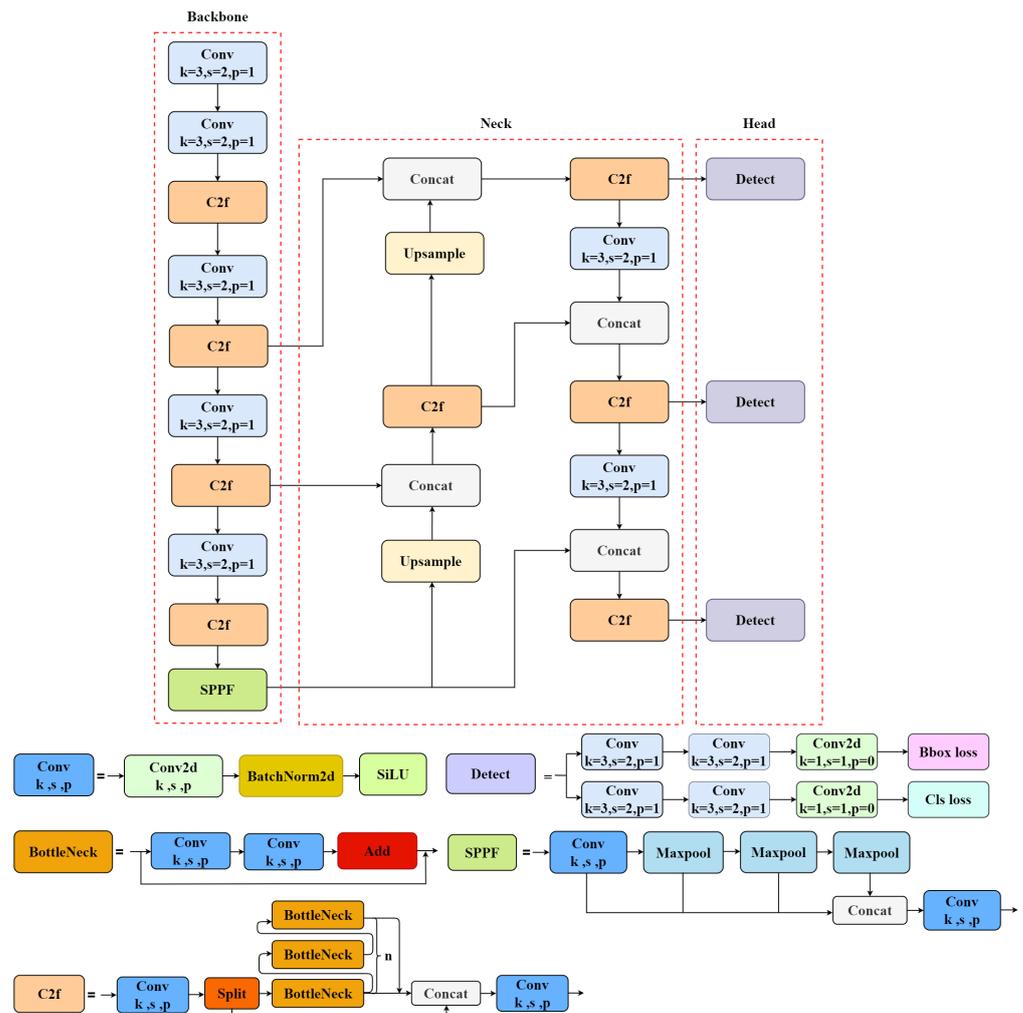


Figure 1. Network structure diagram of Yolov8 algorithm.

2.2. TL-YOLO Algorithms

The improved network framework, TL-Yolo, is depicted in Figure 2. Initially, the Omni-Dimensional Dynamic Convolution is used, and its seamless integration into the C2f module is delineated [28]. The new C2f_ODConv structure is introduced to enhance the capabilities for complex feature extraction. Subsequently, to capture pairwise associations at the pixel level, an effective Multiscale Attention module should be incorporated to further aggregate the output features of two parallel branch output features [29]. In the third phase, cross-layer cascading is utilized to amalgamate semantic information across feature maps of disparate scales thoroughly. A lightweight, efficient cross-stage feature fusion layer, predicated on GSConv and GSCSP is engineered to minimize the model’s computational and parametric demands [30]. Furthermore, an optimized weighted bidirectional feature

pyramid, the BiFPN, is employed during the feature fusion phase to intensify the depth of the pyramid [31]. Finally, the Adaptive Training Sample Selection strategy is implemented to judiciously segregate positive and negative samples [32].

Experimental results show that the average detection accuracy of the TL-Yolo network has risen by 91.30%, which is 4.20% higher than that of Yolov8. The accuracy and recall are also improved by 4.64% and 3.53%, respectively. Compared with other state-of-the-art methods, our TL-Yolo network proffered superior precision in detecting foreign objects on transmission lines.

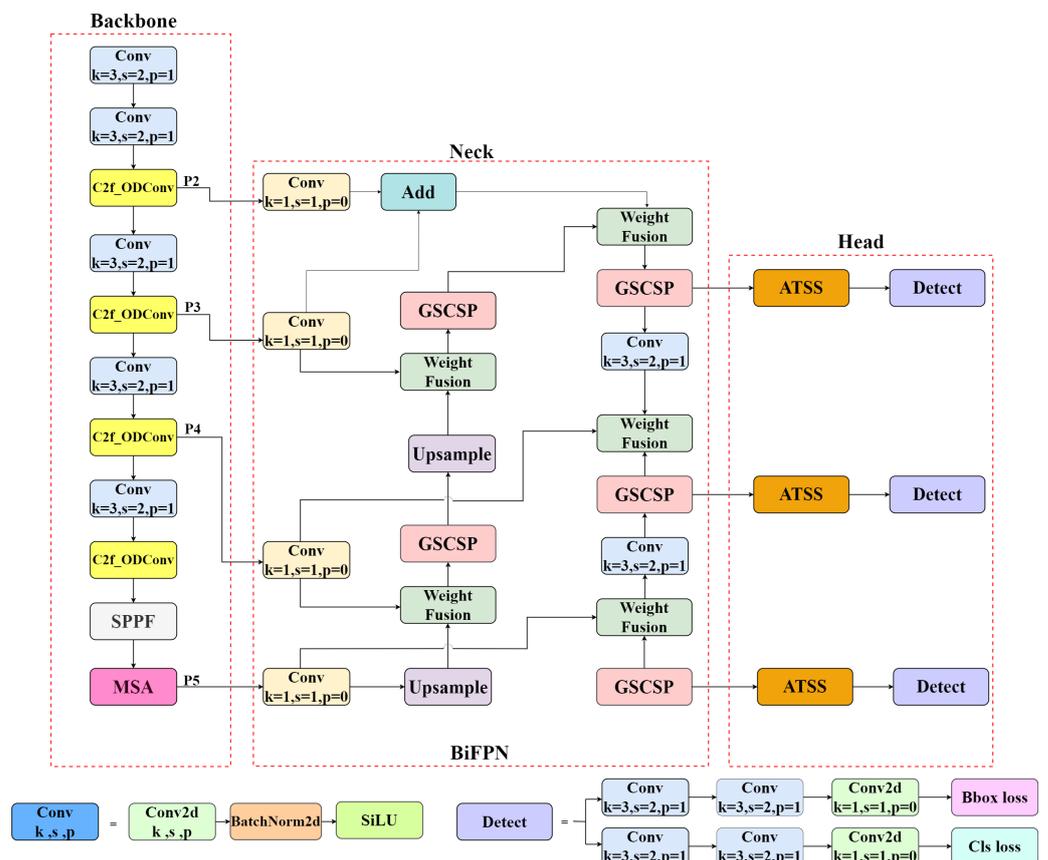


Figure 2. Network structure diagram of our TL-Yolo algorithm. Our algorithm employs C2f_ODConv as the Backbone network to facilitate feature extraction. It uses an improved weighted Bi-Directional Feature Pyramid as the Neck network to complete the fusion of semantic information between different feature layers, and balances the positive and negative samples using an Adaptive Training Sample Selection strategy.

2.2.1. C2f_ODConv

The conventional C2f module is computed using conventional static convolution, which uses the same convolution kernel for different input information, causing information loss to a large extent. Omni-Dimensional Dynamic Convolution adopts a multidimensional attention and parallel strategy, which pays full attention to the input channel dimension, the output channel dimension, and the spatial dimension of the feature map, as well as the convolution kernel dimension so that the output feature maps are fully integrated with the context information, as shown in Figure 3. Specifically, the attention module includes four branches: channel attention (branch 1), filter attention (branch 2), spatial attention (branch 3), and kernel attention (branch 4), which correspond to the input channel dimension of the feature map, the output channel dimension of the feature map, the spatial dimension of the feature map, and the convolution kernel dimension [33]. In branch 1, we obtain the attention along the input channel dimension of the feature map, then the input feature map is multiplied by channel attention to focus on important input feature channels.

In branch 2, we obtain the attention along the output channel dimension of the feature map. In branch 3 and branch 4, we obtain two attentions along the spatial dimension of the feature map and the convolution kernel dimension independently and multiply the two attention outputs before summarizing them along the first dimension. Finally, the summarized result is multiplied by the output of branch 2 and then further multiplied by the output of branch 1 to obtain the final output.

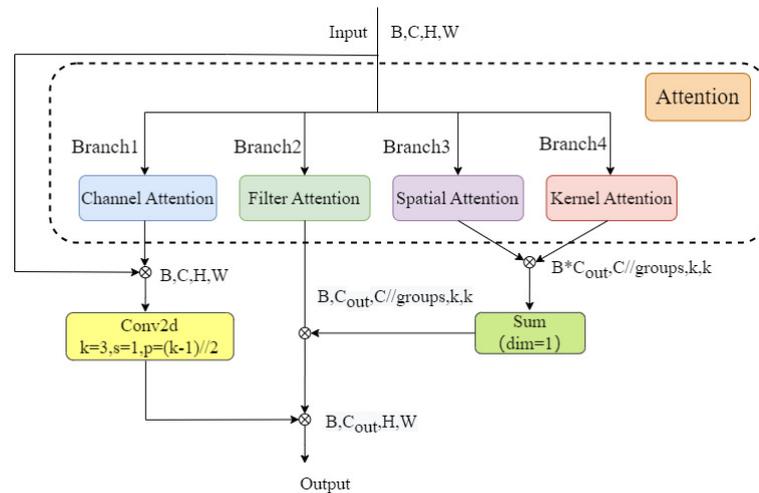


Figure 3. Network structure of ODConv. The ODConv adopts a multidimensional attention and parallelism strategy, which pays full attention to the number of input channels, the number of output channels, the dimension and the size of the convolution kernel. This design ensures that the resultant feature maps are thoroughly amalgamated with contextual data enriching the feature representation.

ODConv is used to replace the traditional convolution in the C2f structure. A new bottleneck structure, Bottleneck_ODConv, is proposed and incorporated into C2f to obtain the C2f_ODConv module which can significantly improve the feature extraction capability, as shown in Figure 4.

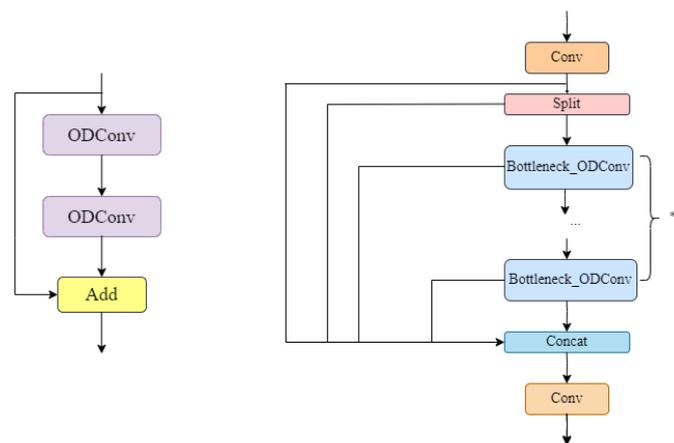


Figure 4. The Bottleneck_ODConv (left) and C2f_ODConv (right) structure.

2.2.2. Multi-Scale Attention (MSA)

The Multiscale attention(MSA) mechanism presents an efficient approach for preserving channel-specific information and emphasizing pertinent spatial details. By partitioning the channels into several sub-feature groups and rescaling certain dimensions, the MSA ensures that critical semantic content is maintained within each group, thereby mitigating the potential information loss typically associated with convolutional processes. Furthermore, the MSA facilitates local information exchange across parallel sub-networks and employs a cross-dimensional spatial learning strategy to amalgamate the output features from these

branches. This technique effectively discerns pixel-level pairwise associations, which are illustrated in Figure 5.

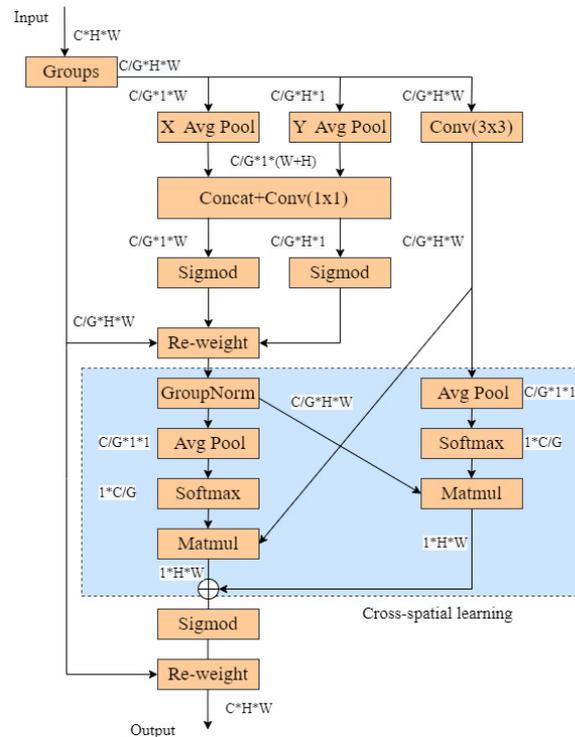


Figure 5. MSA structure. The MSA allows spatial semantic features to be uniformly distributed in each feature group by reshaping some of the channels into batch dimensions and then grouping the channel dimensions into multiple sub-features, as well as further aggregating the output features of the two parallel branches through cross-dimensional interactions to capture pixel-level pairwise relationships.

The MSA divides the input feature map into G sub-features along the channel dimension to facilitate diverse information learning. It employs a tri-branch structure for capturing spatial information of the grouped features: two branches operate with 1×1 convolutions, while the third branch utilizes 3×3 convolutions. Within the 1×1 branches, global average pooling extracts image features across height and width followed by a non-linear Sigmoid activation that molds the dual output vectors produced through identical convolutions. These vectors are then synergistically combined via element-wise multiplication, fostering an interaction of channel-wise feature information. The 3×3 branch performs convolution to garner global channel insights, thereby broadening the feature space. A novel cross-space information fusion procedure is conducted to aggregate the outputs from both branches, generating two spatial attention maps through global average pooling, Softmax normalization, and matrix dot product computations. Here, the new cross-spatial information fusion technology can extract multi-scale features into the attention mechanism to help detect small or difficult-to-identify objects, thus improving the model’s detection accuracy. Ultimately, the output feature maps within each group are computed as an aggregation of the two generated spatial attention weight values, encapsulating all the contextual details.

2.2.3. Improved Bi-Directional Feature Pyramid Network (BiFPN)

When conventional feature fusion approaches are employed, feature maps of different scales are weighted equally, resulting in uniform weights being computed for the output feature maps. This strategy significantly contributes to information loss [34].

The Yolov8 facilitates the PANet approach to fuse features, where information is transferred through separate top-sampling and bottom-sampling operations, respectively. Although this technique does leverage the intrinsic information present in the feature

maps, it necessitates more effective information than what is available through a singular node design, consequently increasing model complexity. Instead, the TL-Yolo structure introduces the Bi-Directional Feature Pyramid Network(BiFPN) for feature fusion, which establishes cross-scale channel connectivity through a weighted bidirectional structure that operates in top-down and bottom-up directions. This bidirectional structure assigns varying weights to input features according to their relative importance and is recursively adopted to bolster feature fusion. This recursive application preserves shallow positional details while mitigating the loss of deep semantic information. A comparative analysis between the PANet (left) and BiFPN (right) structures is given in Figure 6.

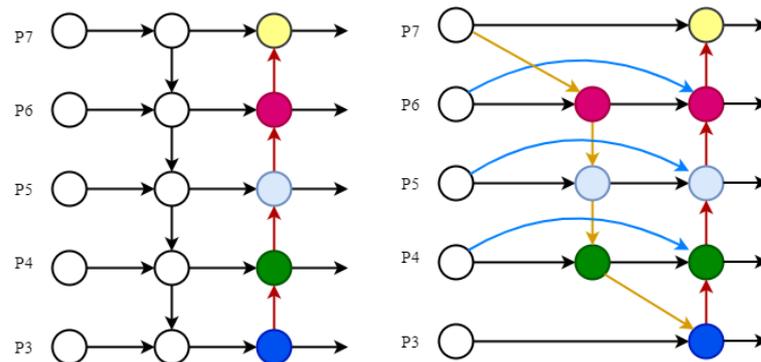


Figure 6. Structure comparison of PANet (left) and BiFPN (right).

As shown in Figure 7, three feature layers–P3, P4, and P5 are ultimately generated and then fed into the designed BiFPN. Figure 7 delineates the refined BiFPN, which is optimized by: (1) eliminating nodes that have a singular input path, since such nodes contribute minimally to feature fusion, and their removal has a negligible impact on network performance while concurrently streamlining the architecture. (2) introducing feature map P2 into the fusion procedure which facilitates up-sampling P2 using standard convolution techniques. The P2 feature map is introduced into BiFPN for feature fusion which enhances the expressive ability of the feature pyramid and enables the model to understand the multi-scale feature information of the image in a more comprehensive way. With 1×1 convolution, the model can achieve feature dimension reduction. Also, the 1×1 convolution has fewer parameters and computation. To a certain extent, it improves the computational efficiency and reduces the training time and inference time of the model.

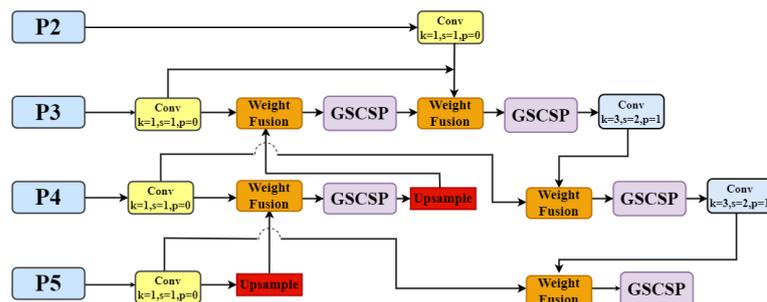


Figure 7. Improved weighted bidirectional feature pyramid BiFPN. The component GSCSP is designed for introducing feature map P2 into fusion procedure and adopting Weight Fusion method to achieve efficient cross-layer cascade information interaction.

The feature fusion is performed using the Weight Fusion method [35], as shown in Figure 8. Specifically, assuming that the input sample size is (bs, C, H, W), the spatially adaptive weights of (bs, 3, H, W) are obtained through convolutional cascade as well as splicing operations, and the weighted summation of the weights can yield an output that fully fuses the contextual information of each channel.

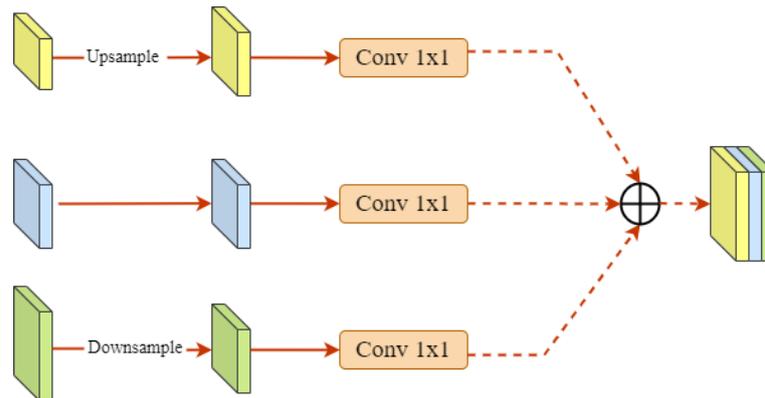


Figure 8. The Weight Fusion module. The sampling operations are performed on feature maps at different levels and the concat operations are performed after convolution by 1×1 . Here the concat operations enable to capture and integrate image information at different scales.

A Group Shuffle Convolution (GSCConv) based Cross Stage Partial Network (GSCSP) is designed at each node of the feature fusion. Here, the GSCConv is a hybrid convolution method combining Standard Convolution(SC) and Depth-wise Separable Convolution (DSC) hybrid convolution method [36].

The input undergoes downsampling through a standard convolution, followed by the application of a depth-separable convolution. The outputs, preceding and succeeding the depth-separable convolution are conjoined, culminating in a blending operation. This operation effectively amalgamates the information processed by both the standard and depth-separable convolutions. The aim is to align the convolutional output as closely as possible with that obtained from standard convolution while concurrently curtailing computational expenses. This approach serves to mitigate the potential adverse effects that are incurred by the intrinsic limitations of depth-separable convolutions on model training.

The cross-stage partial network, GSCSP, utilizes a singular aggregation approach, as illustrated in Figure 9. With the progressive deepening of network layers, the feature map’s channel attributes are optimized. At the same time, spatial information becomes increasingly sparse, diminishing redundancy. Consequently, further scaling alterations become unnecessary. In other words, implementing the GSCSP module at this juncture effectively reduces both computational demand and network complexity, yet maintains adequate model accuracy.

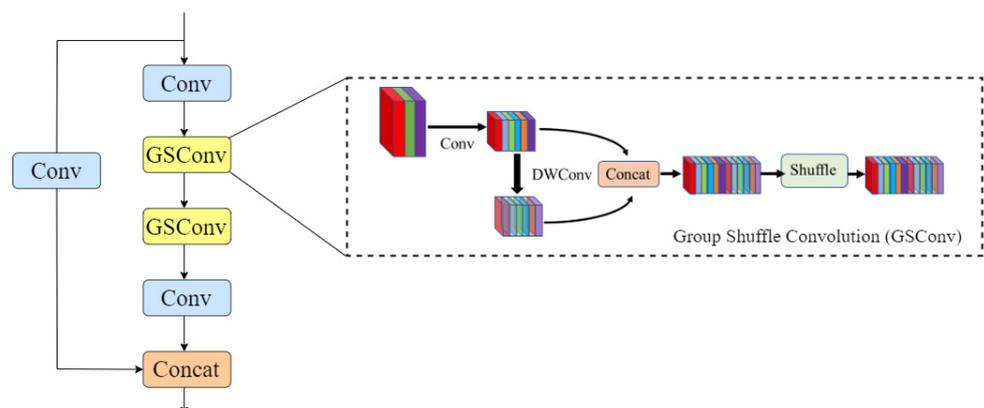


Figure 9. Network structure of GSCSP.

2.2.4. Adaptive Training Sample Selection (ATSS)

In addition to the architectural design of the network structure, the differentiation and equilibrium between positive and negative samples are critical considerations that significantly impact the network performance [37]. Most existing detection networks use

uniform criteria for positive and negative sampling. This results in unsatisfactory predictions whether using a bounding box or anchor regression. Therefore, our work introduced an Adaptive Training Sample Selection (ATSS) strategy to automatically categorize the positive and negative training samples guided by the statistical characteristics (variance and mean) of the object. This strategy aims at improving the detection performance of the network without adding any overhead.

The pseudo-code is given in Algorithm 1. Initially, for each level of the FPN layer around each ground-truth g , the k anchors closest to the center of the specific real frame are selected to create a candidate pool of positive samples c_g . In detail, the Intersection over Union (IoU) between this candidate positive samples and the ground-truth g are calculated, from which the mean m_g and variance v_g of the set of IoUs are derived. At last, the m_g plus v_g are used for thresholding in sample classification: candidates c_g exceeding this threshold with center points of anchor box located inside the real box are recognized as positive and added to the positive sample pool. This process is iteratively applied to each ground-truth g and to curate the final positive samples.

Algorithm 1 Adaptive Training Sample Selection (ATSS)

Input:

- \mathcal{G} is a set of ground-truth boxes on the image
- \mathcal{L} is the number of feature pyramid levels
- \mathcal{A}_i is a set of anchor boxes from the i -th pyramid levels
- \mathcal{A} is a complete set of anchor boxes
- k is a hyperparameter with a default value of 9

Output:

- \mathcal{P} is a set of positive samples
 - \mathcal{N} is a set of negative samples
- 1: **for** each ground-truth $g \in \mathcal{G}$ **do**
 - 2: build an empty set for candidate positive samples of the ground-truth g : $\mathcal{C}_g \leftarrow \emptyset$;
 - 3: **for** each level $i \in [1, 2, \dots, \mathcal{L}]$ **do**
 - 4: $\mathcal{S}_i \leftarrow$ select k anchors from \mathcal{A}_i whose center are closest to the center of ground-truth based on L2 distance;
 - 5: $\mathcal{C}_g = \mathcal{C}_g \cup \mathcal{S}_i$;
 - 6: **end for**
 - 7: compute IoU between \mathcal{C}_g and g : $\mathcal{D}_g = \text{IoU}(\mathcal{C}_g, g)$;
 - 8: compute mean of \mathcal{D}_g : $m_g = \text{Mean}(\mathcal{D}_g)$;
 - 9: compute standard deviation of \mathcal{D}_g : $v_g = \text{Std}(\mathcal{D}_g)$;
 - 10: compute IoU threshold for ground-truth g : $t_g = m_g + v_g$;
 - 11: **for** each candidate $c \in \mathcal{C}_g$ **do**
 - 12: **if** $\text{IoU}(c, g) \geq t_g$ and center of c in g **then**
 - 13: $\mathcal{P} = \mathcal{P} \cup c$;
 - 14: **end if**
 - 15: **end for**
 - 16: **end for**
 - 17: $\mathcal{N} = \mathcal{A} - \mathcal{P}$;
 - 18: **return** \mathcal{P}, \mathcal{N} ;
-

3. Results and Discussion

3.1. Dataset

The foreign objects dataset of transmission lines, including kites, bird nests, trash, and balloons (3156 images). The dataset size is limited, which could potentially result in underfitting, overfitting, or failure in convergence, therefore, we employ data augmentation techniques to increase its volume and diversity. In detail, geometric transformations, brightness adjustments, and noise introduction [38] were utilized to obtain an augmentative dataset of 8526 images with complex backgrounds under adverse weather conditions and illumination levels. To enhance the model's generalization capabilities, the dataset was

partitioned in a ratio of 8:1:1 to create training, validation, and test sets. Some augmented examples are shown in Figure 10 below. The classes and the number of foreign objects in the dataset are shown in Table 1.

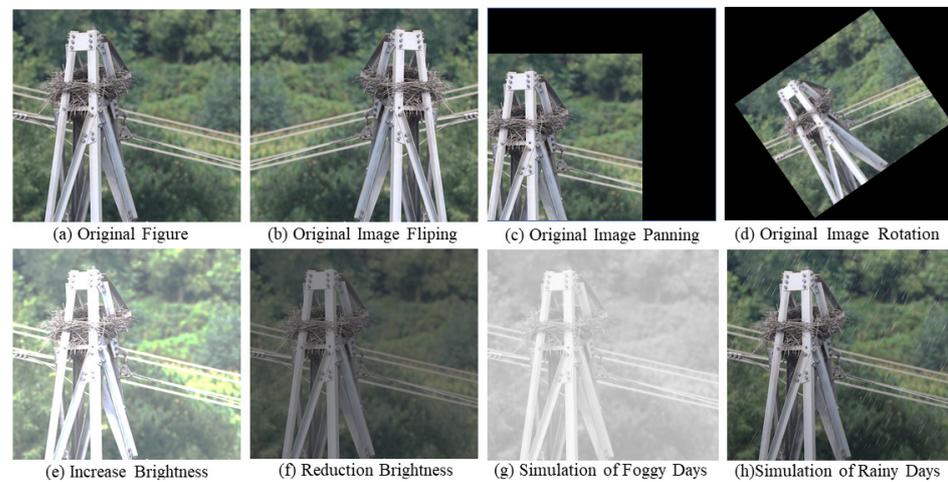


Figure 10. Illustration of data augmentation. The detection scenarios in different environments is simulated using data augmentation.

Table 1. The distribution of objects in the dataset.

Classes	Number
bird nest	2313
balloon	2146
kite	1970
trash	2097

3.2. Parameter Settings

The experiments of this paper were conducted on the Ubuntu 18.04 system powered by the platform with an Intel(R) Core(TM) i7-8565U CPU and an NVIDIA GTX-2080 TI GPU, complemented by 32 GB of RAM. The deep learning framework utilized was Pytorch, running on Python 3.8, and accelerated by the CUDA 11.2 and cuDNN 8.1.1 libraries for enhanced computational efficiency.

The optimal configuration of our model parameters is initiated before training, e.g., the learning rate is initialized empirically at 0.01 and then subsequently modulated via a cosine learning rate scheduler. The epoch number of training is fixed at 150. Considering the experimental environment and hardware configuration, the batch size was set to 32 images. Critical parameters are enumerated in Table 2.

Table 2. Training parameters.

Batch_Size	Epochs	Learning Rate	Momentum
32	150	0.01	0.937

In the experiments, a suite of evaluation metrics was employed to evaluate the efficacy of model training. This suite encompasses conventional indicators of model performance including mean Average Precision (mAP), precision, and recall. Additionally, it incorporates metrics indicative of model complexity and computational efficiency, namely Giga Floating Point Operations (GFLOPs), which corresponds to the computational resources and performance requirements required by the model, the total quantity of model parameters (Params) which reflects the parameters to be learned in the model, and the Frames Per

Second(FPS) evaluate the inference speed and efficiency of deep network models. These specific formulas are delineated below:

$$Recall = \frac{TP}{TP + FN} \quad (1)$$

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$AP_i = \int_0^i P(R)dR(i = 0, 1, 2, \dots) \quad (3)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N} \quad (4)$$

where True Positive (TP) represents the count of objects correctly identified as positive samples. False Positive (FP) denotes the number of objects incorrectly detected as positive samples when they are, in fact, negative. False Negative (FN) correspond to the objects that are wrongly classified as negative despite being positive samples. Average Precision (AP) reflects the prediction accuracy for each class, whereas mean Average Precision (mAP) averages the APs across all the classes, thereby providing a measure of the model's overall accuracy.

3.3. Ablation Experiments

To evaluate the contribution of each component in our method, a series of ablation experiments are conducted as follows: (1) Integration of the C2f_ODConv module in the backbone network to bolster the detection capability for small objects. (2) Implementation of the MSA module enables the rapid concentration on the key area over targets and filtering of redundant information thereby improving the efficiency and accuracy of this task. (3) Usage of the improved BiFPN structure to augment the capability of feature fusion and thereby improve the detection accuracy. (4) Application of the ATSS label-matching strategy to more equitably divide the positive and negative samples. The results of the ablation experiments are shown in Table 3.

Table 3. Ablation experiment.

Experiment Number	ATSS	C2f_ODConv	MSA	Improved BiFPN	Params (M)	GFLOPs	mAP@0.5 (%)
1	-	-	-	-	11.15	28.40	87.10
2	✓	-	-	-	11.17	28.56	88.23
3	✓	✓	-	-	11.68	28.66	89.10
4	✓	✓	✓	-	11.82	28.72	90.44
5	✓	✓	✓	✓	12.22	29.13	91.30

The Yolov8 network is chosen as the baseline model for this experiment. The second experiment is conducted with the ATSS label-matching strategy in the baseline model. The third experiment is conducted with the C2f_ODConv module while retaining the ATSS label-matching strategy. The fourth experiment incorporates the MSA attention mechanism. The fifth experiment additionally employs the improved BiFPN structure. The ablation study shows that the baseline model achieves an mAP of 87.10%. The second experiment yields an mAP of 88.23%, marking a 1.13% enhancement over the Yolov8 network. The third, fourth, and fifth experiments demonstrate incremental mAP improvements of 2.00%, 3.34%, and 4.20%, respectively, over the baseline model. With all the improvements, the mAP for the TL-Yolo algorithm reached 91.30%, signifying a 4.20% improvement compared with the baseline model.

3.4. Comparison with the State-of-the-Art Methods

To validate the performance of our proposed TL-Yolo, we compare our TL-Yolo with the state-of-the-art methods. The experimental results, from Table 4, show that our TL-Yolo outperforms the Faster R-CNN [39], SPP-Net [40], Yolov5 [41], Yolov7 [42], and Yolov8 [43] networks, which shows its improved detection accuracy and superior efficacy.

Table 4. Comparison with the state-of-the-art methods.

Model	Recall (%)	Precision (%)	mAP@0.5 (%)	mAP@0.5:0.95 (%)	Params (M)	GFLOPs	FPS
Faster R-CNN	67.52	53.68	83.26	56.10	128.70	67.26	23.68
SPP-Net	70.23	66.24	85.93	56.47	84.36	70.39	31.24
Yolov5	81.36	75.28	88.64	58.68	24.66	27.41	98.41
Yolov7	83.26	80.47	88.92	59.36	23.68	29.68	110.36
Yolov8	87.10	83.92	89.25	61.32	11.15	28.40	187.63
TL-Yolo	90.63	88.56	91.30	63.20	12.22	29.13	173.28

Table 3 shows that Faster R-CNN and SPP-Net, as two-stage target detection algorithms, are not suitable for real-time target detection due to their high model complexity, large number of parameters and slow reasoning speed. Compared with Yolov5 and Yolov7 algorithms, the detection accuracy of the TL-Yolo algorithm is improved to some extent in mAP@0.5 and mAP@0.5:0.95 evaluation indicators and the detection precision and recall rate are also higher. Meanwhile, the complexity and number of parameters of the TL-Yolo algorithm are optimized, which improves the overall detection efficiency of the algorithm. In addition, we can see that our TL-Yolo outperforms the methods under comparison on almost all metrics, except Yolov8. Although Yolov8 is better than TL-Yolo on computational efficiency metrics, the TL-Yolo achieves better detection accuracy than Yolov8. The reason is that the Omni-Dimensional Dynamic Convolution significantly boosts the backbone network's feature extraction capabilities. Simultaneously, the proposed MSA structure automatically focuses on important features. The weighted BiFPN with a lightweight feature fusion layer not only augments the model's precision and adaptability but also substantially reduces computational demands. Besides, the implementation of adaptive training sample selection harmonizes the ratio of positive to negative samples through strategic label matching, thus enhancing the model's performance. It proves the effectiveness of our TL-Yolo.

3.5. Visualization

We also qualitatively compare the detection results of our proposed TL-Yolo and the state-of-the-art methods. From Figure 11, we can see that, under the real scenes with complex backgrounds, the Faster R-CNN misses the kite in the first column and inaccurately locates the trash though they can effectively locate other kinds of foreign objects. The SPP-Net also inaccurately locates the trash; however, its detection confidence scores for objects are relatively low. The Yolov5 wrongly detects a kite on the tower in the second column and repeatedly detects the trash and the balloon. Although Yolov5, Yolov7, and Yolov8 wrongly detect a kite on the tower in the second column and repeatedly detect the balloon, they achieve higher detection confidence. Our TL-Yolo accurately detects foreign objects and addresses the missed detection, false detection, and repeated detection mentioned above. It is because our TL-Yolo boosts the detection performance by improving effective feature extraction, multi-scale feature integration, and cross-layer feature fusion based on Yolov8.

In foggy environments, the images captured by the devices may be blurred and have low contrast due to the low visibility, which makes it difficult to detect foreign objects on time. When the surface of foreign objects is wet on rainy days, its color, texture, and other characteristics tend to change which affects the performance of the foreign object detection algorithm and decreases the detection confidence in some cases. Under low-light conditions, the overall brightness of the image is reduced, which can lead to a decrease in the contrast between foreign objects and the background, making it difficult to accurately detect the

foreign objects. If the light is too strong, the image may be overexposed, which will also affect foreign object detection. We test the TL-Yolo algorithm in different environments. Compared with normal environmental conditions (Figure 11), the TL-Yolo algorithm can achieve accurate foreign object detection in foggy and rainy days, dark, and bright scenes, which proves that the TL-Yolo algorithm has good generalization performance, as shown in Figure 12.

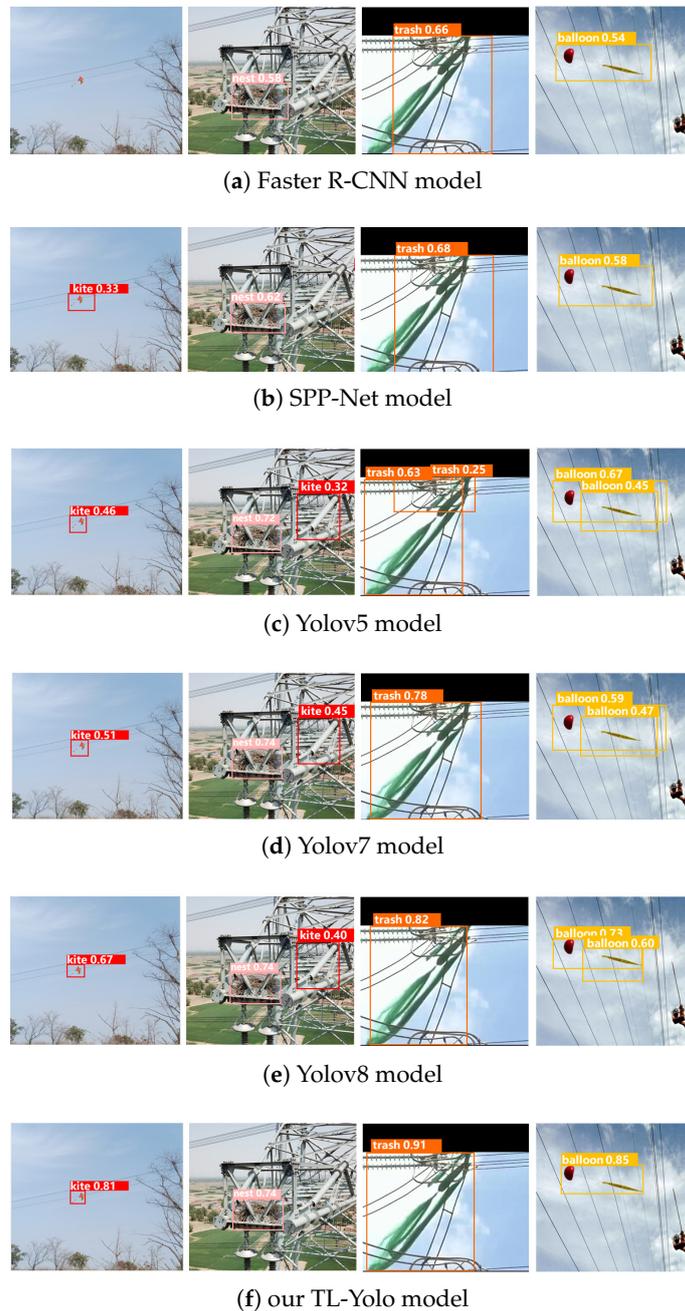


Figure 11. Visualization comparison of detection results with the state-of-the-art methods. Compared with the mainstream two-stage detection algorithms Faster RCNN, SPP-Net, and the single-stage Yolo series of detection algorithms, our TL-Yolo successfully solves the missed detection, false detection, and repeated detection.



Figure 12. Visualisation of detection results under varying environmental conditions using the TL-YOLO algorithm. The TL-YOLO algorithm has consistent detection performance in a variety of working environments with weather conditions and lighting conditions.

4. Conclusions

This paper presents a foreign-object detection method TL-Yolo on power transmission lines based on improved Yolov8. Our TL-Yolo employs Omni-Dimensional Dynamic Convolution (ODConv) as the feature extraction module, an improved weighted bidirectional feature pyramid module with Multi-Scale Attention (MSA) mechanism to fuse multiscale features, a lightweight cross-platform GSCov Cross Stage Partial Network (GSCSP) to facilitate effective cross-layer feature fusion, and an Adaptive Training Sample Selection (ATSS) strategy to balance the positive and negative samples. The experimental results show that the average detection accuracy of the TL-Yolo algorithm is as high as 91.30%, which is 4.20% higher than the Yolov8 algorithm. Meanwhile, our method's accuracy and recall rates also increased by 4.64% and 3.53% respectively. In the comparison experiment, the mAP value of the TL-Yolo algorithm exceeds that of other state-of-the-art target detection algorithms, which fully shows its superior performance in the task of detecting foreign objects on transmission lines.

Traditional methods for inspecting transmission lines often require manual inspection and detection, leading to the issues of missed and false detections. TL-YOLO, on the other hand, can accurately identify and locate target objects, thereby improving the precision and reliability of inspections. Furthermore, the TL-YOLO algorithm enables automated real-time detection, significantly reducing the need for manual intervention. This facilitates the timely identification of line faults and enhances the power supply stability. The low complexity and less computation of the TL-YOLO algorithm make it suitable for outdoor portable detection equipment such as drones; therefore, the TL-YOLO can be used in a wide range of open-air inspection tasks.

It is important to note that this study is conducted in a supervised manner, which necessitates a large amount of labeled data for training. Obtaining labeled data in practice is a time-consuming and labor-intensive process. Therefore, in future research related to foreign object detection on transmission lines, it may be beneficial to integrate with semi-supervised learning methods. In the foreign object detection task, there is often a large number of unlabeled data and a small number of labeled data. Traditional supervised learning methods may have difficulties in dealing with such problems, while semi-supervised learning can make full use of the information of unlabeled data besides labeled data to identify foreign objects more accurately. It can reduce the labor and time costs. This makes semi-supervised learning more practical and feasible in the field of foreign object detection. This has certain practical significance for improving the situation of limited data resources.

Author Contributions: Conceptualization, Y.S.; Methodology, Y.S., R.Z. and C.L.; Software, R.Z. and C.L.; Validation, R.Z. and C.L.; Formal Analysis, Y.S.; Investigation, C.L. and Y.S.; Resources, Y.S. and M.C.; Data Curation, Z.L. and R.Z.; Writing—Original Draft Preparation, R.Z. and C.L.; Writing—Review and Editing, R.Z., M.C., Y.S. and C.L.; Visualization, R.Z., Z.L. and Y.S.; Supervision, M.C., C.L. and Y.S.; Project Administration, C.L. and Y.S.; Funding Acquisition, M.C. and Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Natural Science Foundation of China (Grant No. 61671255 and No. 62201295), Fundamental Science Research Program of Nantong (Grant No. JC12022101), and “Qinglan Project” of Jiangsu Universities.

Data Availability Statement: Data are contained within the article.

Acknowledgments: This paper and research would not have been possible without the support of the Transportation Safety and Intelligence Analysis Team from Nan Tong University.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tang, X.; Shen, W.; Zhu, M.; Bao, W. The foreign object algorithm for transmission lines based on the improved YOLOv4. *J. Anhui Univ. Nat. Sci.* **2021**, *45*, 58–63.
2. Zhu, J.; Guo, Y.; Yue, F.; Yuan, H.; Yang, A.; Wang, X.; Rong, M. A deep learning method to detect foreign objects for inspecting power transmission lines. *IEEE Access* **2020**, *8*, 94065–94075. [[CrossRef](#)]
3. Zhang, H.; Zhou, H.; Li, S.; Li, P. Improved YOLOv3 foreign body detection method in transmission line. *Laser J.* **2022**, *43*, 82–87.
4. Li, J.; Duan, Y.; Wang, C.; Wang, X.; Guo, P.; Zhang, Y. Application of the Unmanned Aerial Vehicle in the Transmission Line Inspection. *Power Syst. Clean Energy* **2017**, *33*, 62–65+70.
5. Li, X.; Li, Z.; Wang, H.; Li, W. Unmanned aerial vehicle for transmission line inspection: Status, standardization, and perspectives. *Front. Energy Res.* **2021**, *9*, 713634–713647. [[CrossRef](#)]
6. Jiang, S.; Jiang, W.; Huang, W.; Yang, L. UAV-based oblique photogrammetry for outdoor data acquisition and offsite visual inspection of transmission line. *Remote Sens.* **2017**, *9*, 278. [[CrossRef](#)]
7. Yuwu, W.; Jialong, L.; Chun, C. Research on autonomous patrol inspection technology of power transmission channel based on multi-rotor UAV. *Power Syst. Clean Energy* **2020**, *36*, 84–89.
8. Lijun, J.; Chunyu, Y.; Shujia, Y. Recognition of extra matters on transmission lines based on aerial images. *J. Tongji Univ. Nat. Sci.* **2013**, *41*, 277–281.
9. Wang, W.; Zhang, J.; Han, J.; Liu, L.; Zhu, M. Broken strand and foreign body fault detection method for power transmission line based on unmanned aerial vehicle image. *J. Comput.* **2015**, *35*, 2404–2408.
10. Hayal, M.R.; Elsayed, E.E.; Kakati, D.; Singh, M.; Elfikky, A.; Boghdady, A.I.; Grover, A.; Mehta, S.; Mohsan, S.A.H.; Nurhidayat, I. Modeling and investigation on the performance enhancement of hovering UAV-based FSO relay optical wireless communication systems under pointing errors and atmospheric turbulence effects. *Opt. Quantum Electron.* **2023**, *55*, 625–648. [[CrossRef](#)]
11. Zhang, W.; Liu, X.; Yuan, J.; Xu, L.; Sun, H.; Zhou, J.; Liu, X. RCNN-based foreign object detection for securing power transmission lines (RCNN4SPTL). *Procedia Comput. Sci.* **2019**, *147*, 331–337. [[CrossRef](#)]
12. Wang, B.; Wu, R.; Zheng, Z.; Zhang, W.; Guo, J. Study on the method of transmission line foreign body detection based on deep learning. In Proceedings of the 2017 IEEE Conference on Energy Internet and Energy System Integration, (EI2) 2017, Beijing, China, 26–28 November 2017; pp. 1–5.
13. Dou, Y.; Yu, X.; Li, J. Feature GANs: A Model for Data Enhancement and Sample Balance of Foreign Object Detection in High Voltage Transmission Lines. In Proceedings of the International Conference on Computer Analysis of Images and Patterns, CAIP 2019, Salerno, Italy, 3–5 September 2019; pp. 568–580.

14. Li, J.; Yan, D.; Luan, K.; Li, Z.; Liang, H. Deep learning-based bird's nest detection on transmission lines using UAV imagery. *Appl. Sci.* **2020**, *10*, 6147. [[CrossRef](#)]
15. Xia, P.; Yin, J.; He, J.; Gu, L.; Yang, K. Neural detection of foreign objects for transmission lines in power systems. *J. Phys. Conf. Ser.* **2019**, *1267*, 012043–012050. [[CrossRef](#)]
16. Qiu, Z.; Zhu, X.; Liao, C.; Shi, D.; Kuang, Y.; Li, Y.; Zhang, Y. Detection of bird species related to transmission line faults based on lightweight convolutional neural network. *IET Gener. Transm. Distrib.* **2022**, *16*, 869–881. [[CrossRef](#)]
17. Shen, M.; Pei, J.; Fu, X.; Zhang, J.; Gong, F.; Liu, X. A New Transmission Line Foreign Object Detection Network Structure: TLFODNet. *Comput. Mod.* **2019**, *2*, 118–122.
18. Wu, M.; Guo, L.; Chen, R.; Du, W.; Wang, J.; Liu, M.; Kong, X.; Tang, J. Improved YOLOX foreign object detection algorithm for transmission lines. *Wirel. Commun. Mob. Com.* **2022**, *2022*, 5835693. [[CrossRef](#)]
19. Yu, Y.; Qiu, Z.; Liao, H.; Wei, Z.; Zhu, X.; Zhou, Z. A method based on multi-network feature fusion and random forest for foreign objects detection on transmission lines. *Appl. Sci.* **2022**, *12*, 4982. [[CrossRef](#)]
20. Li, H.; Liu, L.; Du, J.; Jiang, F.; Guo, F.; Hu, Q.; Fan, L. An improved YOLOv3 for foreign objects detection of transmission lines. *IEEE Access* **2022**, *10*, 45620–45628. [[CrossRef](#)]
21. Li, M.; Ding, L. DF-YOLO: Highly accurate transmission line foreign object detection algorithm. *IEEE Access* **2023**, *11*, 108398–108406. [[CrossRef](#)]
22. Wang, Z.; Yuan, G.; Zhou, H.; Ma, Y.; Ma, Y. Foreign-Object Detection in High-Voltage Transmission Line Based on Improved Yolov8m. *Appl. Sci.* **2023**, *13*, 12775. [[CrossRef](#)]
23. Zhang, H.; Zhou, X.; Shi, Y.; Guo, X.; Liu, H. Object Detection Algorithm of Transmission Lines Based on Improved YOLOv5 Framework. *J. Sens.* **2024**, *2024*, 1–14. [[CrossRef](#)]
24. Liu, C.; Ma, L.; Sui, X.; Guo, N.; Yang, F.; Yang, X.; Huang, Y.; Wang, X. YOLO-CSM-Based Component Defect and Foreign Object Detection in Overhead Transmission Lines. *Electronics* **2023**, *13*, 123. [[CrossRef](#)]
25. Tang, C.; Dong, H.; Huang, Y.; Han, T.; Fang, M.; Fu, J. Foreign object detection for transmission lines based on Swin Transformer V2 and YOLOX. *Vis. Comput.* **2023**, *23*, 1–19. [[CrossRef](#)]
26. Yu, C.; Liu, Y.; Zhang, W.; Zhang, X.; Zhang, Y.; Jiang, X. Foreign objects identification of transmission line based on improved YOLOv7. *IEEE Access* **2023**, *11*, 51997–52008. [[CrossRef](#)]
27. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the 2018 Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8759–8768.
28. Wang, R.; Liang, F.; Wang, B.; Mou, X. ODCA-YOLO: An Omni-Dynamic Convolution Coordinate Attention-Based YOLO for Wood Defect Detection. *Forests* **2023**, *14*, 1885. [[CrossRef](#)]
29. 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 International Conference on Acoustics, Speech and Signal Processing, ICASSP 2023, Rhodes Island, Greece, 4–10 June 2023; pp. 1–5.
30. Guan, Q.; Yu, K.; Wang, H.; Li, Q.; Gao, S. Lightweight Protective Clothing Detection Algorithm Based on Ghost Convolution and GSConv Convolution. In Proceedings of the 2023 International Conference on the Cognitive Computing and Complex Data, ICCD 2023, Huaian, China, 21–22 October 2023; pp. 73–77.
31. Yan, B.; Li, J.; Yang, Z.; Zhang, X.; Hao, X. AIE-YOLO: Auxiliary information enhanced YOLO for small object detection. *Sensors* **2022**, *22*, 8221. [[CrossRef](#)]
32. Zhang, S.; Chi, C.; Yao, Y.; Lei, Z.; Li, S.Z. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, 13–19 June 2020; pp. 9759–9768.
33. Qian, J.; Lin, J.; Bai, D.; Xu, R.; Lin, H. Omni-dimensional dynamic convolution meets bottleneck transformer: A novel improved high accuracy forest fire smoke detection model. *Forests* **2023**, *23*, 833. [[CrossRef](#)]
34. He, L.; Wei, H.; Wang, Q. A New Target Detection Method of Ferrography Wear Particle Images Based on ECAM-YOLOv5-BiFPN Network. *Sensors* **2023**, *23*, 6477. [[CrossRef](#)] [[PubMed](#)]
35. Ma, X.; Ma, Y. Relation and context augmentation network for facial expression recognition. *Image Vision Comput.* **2022**, *127*, 104556–104565. [[CrossRef](#)]
36. Wang, T.; Li, Y.; Zhai, Y.; Wang, W.; Huang, R. A Sewer Pipeline Defect Detection Method Based on Improved YOLOv5. *Processes* **2023**, *11*, 2508. [[CrossRef](#)]
37. Wang, K.; Liu, M. YOLO-Anti: YOLO-based counterattack model for unseen congested object detection. *Pattern Recognit.* **2022**, *131*, 108814–108830. [[CrossRef](#)]
38. Sun, X.; Jia, X.; Liang, Y.; Wang, M.; Chi, X. A defect detection method for a boiler inner wall based on an improved YOLO-v5 network and data augmentation technologies. *IEEE Access* **2022**, *10*, 93845–93853. [[CrossRef](#)]
39. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9. [[CrossRef](#)]
40. He, K.; Zhang, X.; Ren, S.; Sun, J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 1904–1916. [[CrossRef](#)] [[PubMed](#)]
41. Lu, L.; Chen, Z.; Wang, R.; Liu, L.; Chi, H. Yolo-inspection: Defect detection method for power transmission lines based on enhanced YOLOv5s. *J. Real-Time Image Process.* **2023**, *20*, 104–116. [[CrossRef](#)]

42. Wang, C.-Y.; Bochkovskiy, A.; Liao, H.-Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, 17–24 June 2023; pp. 7464–7475.
43. Hou, C.; Li, Z.; Shen, X.; Li, G. Real-time defect detection method based on YOLO-GSS at the edge end of a transmission line. *IET Image Process.* **2024**, *18*, 1315–1327. [[CrossRef](#)]

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.