Article

RCDAM-Net: A Foreign Object Detection Algorithm for Transmission Tower Lines Based on RevCol Network

Wenli Zhang [1], Yingna Li [1,2,*] and Ailian Liu [1,2]

[1] Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China; 20212104091@stu.kust.edu.cn (W.Z.); liu2024@stu.kust.edu.cn (A.L.)
[2] Computer Technology Application Key Lab of the Yunnan Province, Kunming 650500, China
* Correspondence: liyingna@kust.edu.cn

Abstract: As an important part of the power system, it is necessary to ensure the safe and stable operation of transmission lines. Due to long-term exposure to the outdoors, the lines face many insecurity factors, and foreign object intrusion is one of them. Traditional foreign object (bird's nest, kite, balloon, trash bag) detection algorithms suffer from low efficiency, poor accuracy, and small coverage, etc. To address the above problems, this paper introduces the RCDAM-Net. In order to prevent feature loss or useful feature compression, the RevCol (Reversible Column Networks) is used as the backbone network to ensure that the total information remains unchanged during feature decoupling. DySnakeConv (Dynamic Snake Convolution) is adopted and embedded into the C2f structure, which is named C2D and integrates low-level features and high-level features. Compared to the original BottleNeck structure of C2f, the DySnakeConv enhances the feature extraction ability for elongated and weak targets. In addition, MPDIoU (Maximum Performance Diagonal Intersection over Union) is used to improve the regression performance of model bounding boxes, solving the problem of predicted bounding boxes having the same aspect ratio as true bounding boxes, but with different values. Further, we adopt Decoupled Head for detection and add additional auxiliary training heads to improve the detection accuracy of the model. The experimental results show that the model achieves *mAP*50, Precision, and Recall of 97.98%, 98.15%, and 95.16% on the transmission tower line foreign object dataset, which is better to existing multi-target detection algorithms.

Keywords: transmission line; foreign object detection; RevCol algorithm; detection head; loss function

1. Introduction

The power grid is an important infrastructure related to people's livelihoods and national energy security. As the link of power transmission, the safe and stable operation of transmission lines in the power grid is a necessary guarantee for social production and people's life. In recent years, China's growing electricity demand has led to a significant increase in both the voltage levels and the number of transmission lines. This expansion, particularly in long-distance transmission, has introduced complexities in navigating diverse terrains [1,2] like plateaus, hills, basins, mountainous areas, etc. In crowded residential areas, commercial areas, and other places, power transmission lines are susceptible to the attachment of everyday items such as balloons, kites, trash bag, etc. In the natural environment, transmission lines are also susceptible to bird damage, such as bird nests in transmission lines. To ensure transmission line power delivery remains reliable and to mitigate safety risks [3], regular and thorough inspections of these lines are imperative [4–6]. The recognition method of computer vision can to some extent avoid manual operations, reduce the probability of misjudgment, and accelerate the screening process.

Traditional foreign object detection on transmission lines, as explored in the literature [7–10], primarily relies on manual feature extraction methods. These methods involve identifying foreign objects and physically removing them. Wang et al. [11] proposed a

detection method for broken strands and foreign object defects in transmission lines based on line structure perception, which uses gradient operators to extract line objects and identifies significant parallel wire groups in transmission line structures through parallelism calculations. Wire breakage and foreign object defects can be identified by calculating the width change and grayscale similarity of segmented wires. Zhao et al. [12] discusses the use of Hough's linear transform for transmission tower line extraction, combined with convolution operations in small areas of the transmission line. However, the detection efficiency of this method is impacted by complex backgrounds and noise, resulting in lower accuracy. Jiao et al. [13] adopts the frame difference method to label foreign objects and uses key-frame extraction and foreign object feature point tracking to achieve the purpose of foreign object detection in transmission lines. Liang et al. [14] used the Linear Segment Detection algorithm to extract power lines from images and designed a recognition algorithm to detect foreign objects based on foreign object features. Traditional foreign object detection methods can detect foreign objects to a certain extent; however, these methods exhibit low detection accuracy, focus on single-target detection, lack generalization ability, are not scalable, and suffer from false detection leaks.

With the development of deep learning and neural networks, the integration of image classification and detection with computer vision technology [15] has made further contributions to the development of the smart grid field [16–18]. Recent advancements in computer image classification and detection technologies have significantly reduced manual operations and the likelihood of misjudgments in identifying foreign objects on transmission tower lines. These technologies not only expedite the screening process but also ensure the safe and efficient functioning of power systems. Key developments in this area, as reported in the various literature, include the following. Wang et al. [19] compared and analyzed DPM (Deformable Part Model), Faster R-CNN, and SSD methods using actual datasets of transmission line foreign objects and verified the feasibility of deep learning-based recognition methods in the real-time detection of transmission line foreign objects. Gong et al. [20] used the TensorBoard module in the TensorFlow framework to design the deep convolutional neural network model structure and optimize the model parameters after grayscaling and denoising the foreign object dataset. Xiao et al. [21] used the K-means algorithm to cluster the size of foreign object images to set the size of anchor frames, and enhancements include an upsampling module, depth-separable convolution, and improved loss functions for better classification and detection. Shen et al. [22] optimized the candidate frame and designed an end-to-end joint training approach called TLFOD Net to improve the model training performance. Zhong et al. [23] improved three aspects of the YOLOv3 algorithm, including the width and height loss function of the prediction frame, the prediction category imbalance loss function, and the neural network structure, to improve the foreign object recognition effect. Zhang et al. [24] improved the feature pyramid pooling module based on YOLOv4 and optimized the loss function to improve the algorithm performance. Zhang et al. [25] used k-means clustering to generate anchor boxes and then improved the spatial pyramid pooling (SPPF) module and activation function, and this approach significantly enhanced the model performance. Tang et al. [26] used GhostNet module to replace the backbone feature extraction network of YOLOv4, improved the PANet module, and replaced the ordinary convolutional block with a depth-separable convolutional module, which improved the detection speed and reduced the model parameters, although the detection accuracy was degraded. Yu et al. [27] used Otsu to extract the target region of interest, DenseNet201 to extract the depth features of the target region, and ECOC-SVM for training and testing to improve the detection accuracy.

Current object detection algorithms often compromise detection accuracy by compressing or discarding features during extraction. To address this, our paper introduces a novel algorithm specifically tailored for detecting foreign objects on transmission tower lines. This algorithm diverges from the norm by improving upon the conventional target detection model. Key innovations include: 1. Utilizing RevCol as the primary backbone network to ensure that the total information remains unchanged during the feature decoupling

process. 2. Integrating an enhanced C2f module behind the backbone network to enable the model to capture both low-level detail information and high-level semantic information and replacing the standard C2f in the BottleNeck with DySnakeConv to better capture relevant features. 3. Employing a decoupled detection head, which separates classification and regression tasks, and adding an auxiliary training head. This head is actively involved during the training phase and removed during final inference, thereby refining the model's detection accuracy. 4. The MPDIoU method is incorporated to overcome challenges in bounding box regression, especially when predicted and actual boxes share aspect ratios but differ in dimension. To validate the effectiveness of our proposed model, we conducted comparative analyses with other mainstream target detection models. The results affirm the superiority of our approach in accurately detecting foreign objects on transmission tower lines.

2. Model Construction

In this paper, we design a convolutional neural network-based foreign object detection model for transmission towers and lines, RCDAM-Net. First, in order to solve the problem of how the feature extraction of traditional target detection algorithms will compress or discard useful features, this paper introduces RevCol as the backbone network to extract more comprehensive feature information. Second, in order to strengthen the feature extraction and fusion capability of the model, an improved C2D structure is designed. A decoupled detection head is used in the detection head part to assist the training head for detection. Finally, to further enhance the model's ability to localize the target, the MPDIoU loss function is used as the bounding box regression function. The network architecture of RCDAM-Net is depicted in Figure 1.
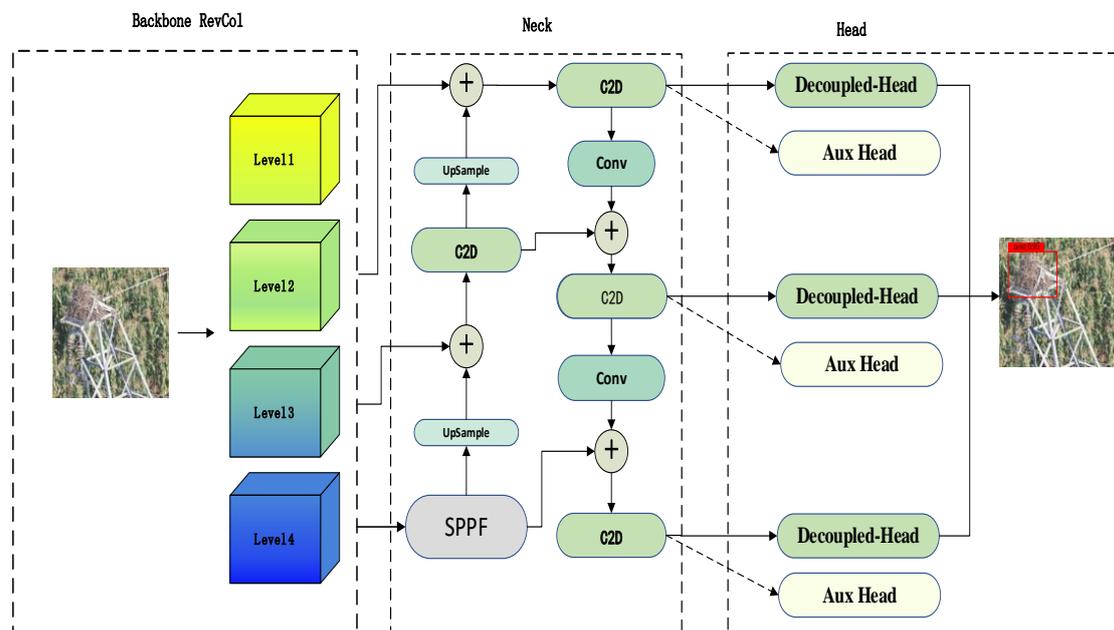

Figure 1. RCDAM-Net Network Framework.

2.1. Backbone Network-RevCol

Mainstream visual models predominantly concentrate on representational and perceptual capabilities, particularly in the context of supervised training. These models tend to compress or discard features during layer transfer, retaining only the most pertinent information aligned with supervisory input. However, this approach often leads to an inadequacy of features [28]. Given that these models undergo supervised pre-training, there is an implicit expectation of their efficacy in downstream tasks. It is essential to recognize that during pre-training, the specifics of downstream tasks are typically unknown,

highlighting the risk of prematurely discarding potentially valuable information. Therefore, to cultivate a more generalized representation, it is advisable to retain a broader spectrum of original information rather than compressing or eliminating it. This approach could facilitate richer knowledge acquisition in pre-training, thereby enhancing performance in various downstream applications. The targets studied in this paper (bird's nests, balloons, kites, trash bags) have the characteristics of small targets and subtle features. They have a small spatial distribution in the image, which may make it difficult for the model to extract enough features to correctly identify these small targets. Taking this feature into account, we use the RevCol network to allow the model to retain a wider range of original information when extracting features. Because during pre-training, the model is unknown to subsequent downstream tasks, useful features may be discarded during the feature extraction stage. The RevCol network can enhance the acquisition of feature information in the pre-training stage, thereby improving the performance of downstream tasks.

RevCol [29] structure adopts a multi-input approach, which consists of multiple reversible multilevel fusion modules. The RevCol introduced the idea of disentangled feature learning [30] into model design and proposed using reversible column as a unit to transfer information to ensure feature decoupling, while ensuring that information is not lost in network transmission. The network structure includes multiple columns, which can increase the sensory field. And reversible connections [31] are added between the columns, so that the low-level texture details and high semantic information can be separated gradually by accessing the inputs to the columns repeatedly. The specific structure is shown in Figure 2. The merit of this approach lies in its dual capacity to maintain high accuracy during pre-training while preserving essential low-level information. This balance is crucial for achieving superior results in subsequent detection tasks.
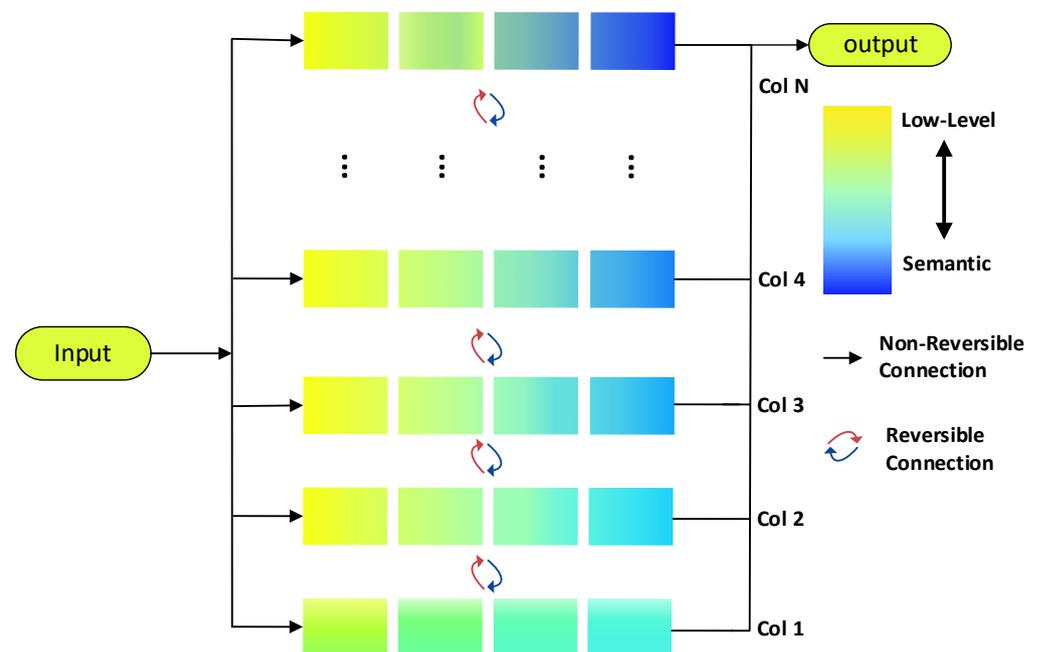

Figure 2. Reversible Column Networks Framework.

This method begins by dividing the input image into several non-overlapping blocks. Each block is then processed through a distinct sub-network (column), each employing a unique ConvNeXt structure [32] with different weights. Within these sub-networks, the image blocks undergo a four-layer propagation process. Initially, a fusion unit harmonizes the dimensions of inputs across different layers. Subsequently, these unified inputs pass through a series of ConvNeXt blocks, eventually combining with the inputs of reversible operations to produce the final output. Notably, each column yields four levels of feature

maps. The feature maps closer to the input emphasize high-level semantics, whereas those at further levels focus on low-level semantics.

Each hierarchy has two inputs, one input is the previous hierarchy in the same column, and the other input is the next hierarchy in the previous column, as shown in Figure 3.
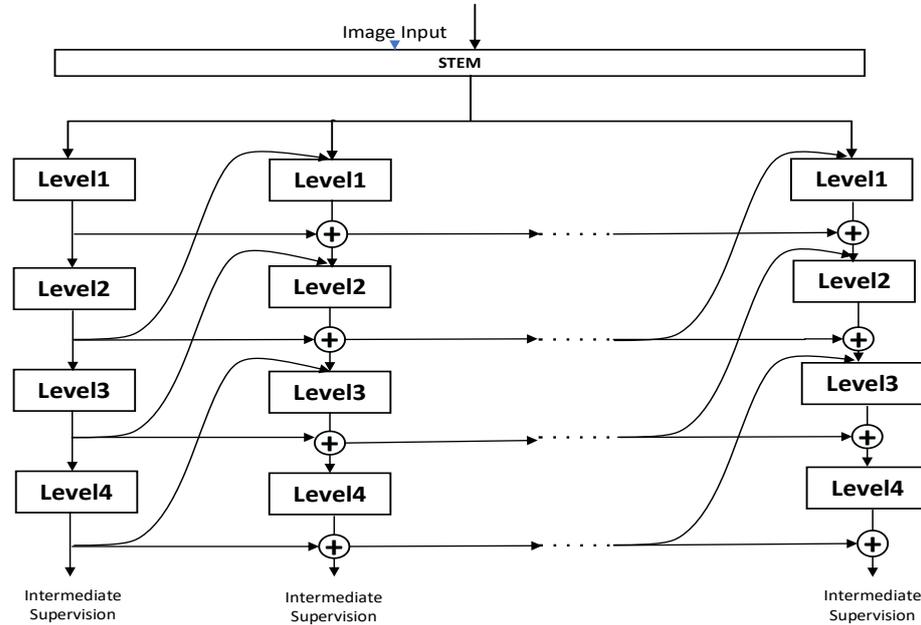

Figure 3. Reversible Column Networks Detailed Structure.

The two inputs represent the high-level semantic information and the low-level texture information, and the equation is expressed as follows:

$$Forward: x_t = F_t(x_{t-1}, x_{t-m+1}) + \gamma x_{t-m} \tag{1}$$

$$Inverse: x_{t-m} = \gamma^{-1}[x_t - F_t(x_{t-1}, x_{t-m+1})] \tag{2}$$

where *Forward* is the forward propagation, *Inverse* is the inverse operation, x_t is the t-th level feature, $F(\bullet)$ is the activation function, γ is the reversible operation, and γ^{-1} is its inverse function.

2.2. Neck-Network

In the neck-network part, the C2f structure is selected as the main module to fuse the low-level feature information with the high-level features in the backbone network part. The C2f module performs a series of convolution operations on the inputs, and then fuses the information by splitting and splicing to obtain the output. The C2f module enhances the ability of feature expression through the dense residual structure and changes the number of channels through the splitting and splicing operations based on the scaling coefficients to reduce the computational complexity and the model capacity. The C2f module is a pivotal component of our neural network architecture, comprising two integral parts: the Context module and the Focus module. The Context module, primarily a series of convolutional layers with residual connections, excels in extracting high-level semantic features. These features are then relayed to the Focus module via lateral connections. In the Focus module, contextual information is utilized effectively. This involves a concatenation operation and a 1×1 convolutional layer, strategically fusing feature information from diverse layers to enhance the expressiveness and perceptual capabilities of the network. Specifically, for detection objects with extreme shapes, we employ Dynamic Snake [33] within the C2f structure. This replacement for the BottleNeck significantly improves the extraction of slender and weak local structural features as well as complex and changeable global morphologies, which is particularly effective for objects like kites and garbage that have

slender and variable shapes. Dynamic Snake Convolution learns deformation based on input feature maps and adaptively focuses on slender and tortuous local features based on information about tubular structure morphology. The Snake Model is used, which is a closed curve that represents the outline of an object. It has the ability to adaptively adjust its shape. In Dynamic Snake Convolution, convolution operations are introduced into the Snake Model to enhance the perception of image features. By combining with convolution operations, the Snake Model can dynamically adjust its shape based on local information in the image to more accurately fit the object contour. This is very beneficial when working with complex, irregular or changing object shapes. The kites and trash bags detected in this paper meet these characteristics, so dynamic snake convolution is used to replace the BottleNeck structure in C2f to enhance the model detection effect. The architecture's improved structure is illustrated in Figure 4.
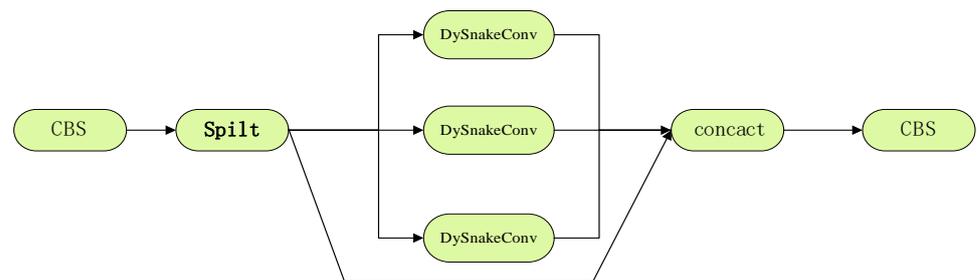

Figure 4. Modified C2f Structure.

2.3. Detection Head Networks

In contemporary mainstream object detection models, the detection head typically comprises three sensory field branches, each functioning as a coupled head that concurrently undertakes classification and regression tasks. However, this coupling presents challenges due to the distinct nature of these tasks. Localization requires boundary-aware features for precise bounding box regression, while classification, a more coarse-grained segmentation task, demands a richer semantic context. This disparity often results in spatial misalignment, adversely affecting the model's convergence speed. Moreover, while a fully-connected head offers higher spatial sensitivity, crucial for differentiating between complete and partial objects, a convolutional head excels in robustly regressing the entire object. To mitigate these issues, we adopt a decoupled head for detection, enabling more efficient and accurate task-specific processing.

The decoupled head subdivides into three branches, namely, classification, regression, and confidence. The use of a decoupled head can make the model network converge faster. The structure of the improved detection head is illustrated in Figure 5.
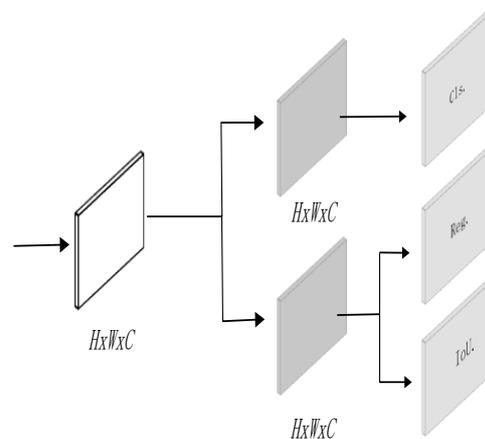

Figure 5. Decoupled Head Structure.

Adding an auxiliary training head to the detection head improves the accuracy by increasing the training cost without affecting the inference time because the auxiliary head will only appear during the training process. The structure of the structure after adding the auxiliary training head is illustrated in Figure 6.
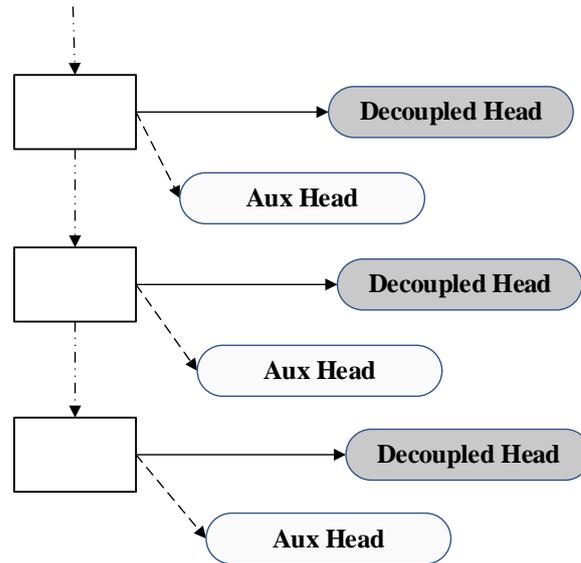

Figure 6. Detection Head Structure.

2.4. MPDIoU Loss Function

Most state-of-the-art target detection models rely on a bounding box regression (BBR) module to determine the location of a target. Based on this model, a well-designed loss function is very important for the success of BBR. However, most of the existing BBR loss functions have the same value for different predictions, which reduces the convergence speed and accuracy of bounding box regression. In this paper, we use the MPDIoU [34] optimization model performance by directly minimizing the distance between the upper-left and lower-right corner points between the predicted bounding box and the real bounding box, rather than through *IoU* (Intersection over Union), the details are shown in Figure 7. This helps to more directly measure the positional differences between bounding boxes. By optimizing the loss function, MPDIoU can make the predicted bounding box more accurately align with the actual target, thereby improving the accuracy of detection. By considering the width and height of the bounding box, MPDIoU can obtain more detailed position information, thereby better guiding the model for regression. MPDIoU adopts a maximization strategy in loss calculation, aiming to maximize accuracy. This method may make the model pay more attention to the accuracy of the bounding box during the learning process. The MPDIoU formula is as follows:

$$d_1^2 = \left(x_1^B - x_1^A\right)^2 + \left(y_1^B - y_1^A\right)^2 \tag{3}$$

$$d_2^2 = \left(x_2^B - x_2^A\right)^2 + \left(y_2^B - y_2^A\right)^2 \tag{4}$$

$$\text{MPDIoU} = \frac{A \cap B}{A \cup B} - \frac{d_1^2}{w^2 + h^2} - \frac{d_2^2}{w^2 + h^2} \tag{5}$$

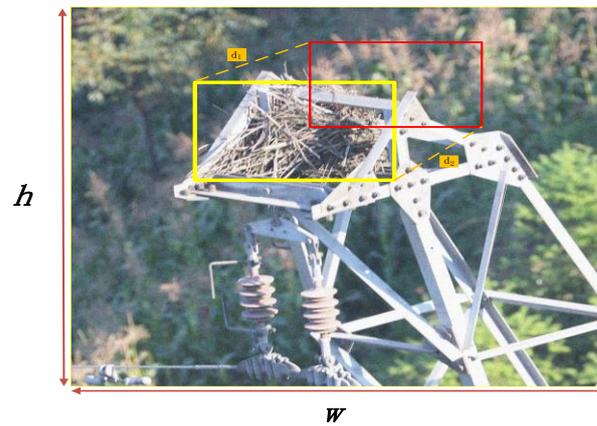$$L_{\text{MPDIoU}} = 1 - \text{MPDIoU} \tag{6}$$

Figure 7. Example of MPDIoU Real-picture.

3. Experimental Results and Analysis

3.1. Experimental Environment

The operating system used for the experiment was Linux 18.04, the CPU was Intel(R) Xeon(R) Gold 6326 CPU@2.90GHz, the RAM was 24 GB, the graphics card was NVIDIA GeForce RTX 3090 24 G memory with CUDA11.4, and the deep learning framework corresponding to CUDA11.4 was selected. Framework was PyTorch 1.12.1, and the programming language environment was Python 3.9.16.

3.2. Evaluation Metrics

In this study, we utilize Precision, Recall, Mean Average Precision (MAP) 50, and *mAP*50:95 as key metrics to evaluate our model's detection accuracy. Precision is defined as the proportion of true positive samples among those identified as positive by the model. Recall, also known as the true positive rate, measures the ratio of correctly predicted positive samples to the total actual positive samples. Both *mAP*50 and *mAP*50:95 serve as crucial metrics for assessing the model's proficiency in both localization and classification of detection objects. The *mAP* is calculated as the average of the Average Precision (*AP*) for all categories at a given Intersection over Union (*IoU*) threshold. Specifically, *mAP*50 represents the mean accuracy across categories at an *IoU* threshold of 0.5, while *mAP*50:95 indicates the mean accuracy at a more stringent *IoU* threshold of 0.95. The formulae are shown below:

$$P = \frac{TP}{TP + FP} \tag{7}$$

$$R = \frac{TP}{TP + FN} \tag{8}$$

$$AP = \int P \, dR \tag{9}$$

$$mAP = \frac{1}{n} \sum_{j=1}^{n} AP(j) \tag{10}$$

$$AP50 = \frac{1}{n} P_i^{IOU=0.5}(R_i^{IOU=0.5}) \tag{11}$$

$$AP50:95 = \frac{1}{10}(AP50 + AP55 + \cdots\cdots + AP95) \tag{12}$$

where *n* is the detection target type, *TP* (True positives): the number of correctly classified positive samples, i.e., the number of samples that are actually positive and classified as positive by the classifier. *FP* (False positives): the number of incorrectly classified positive samples, i.e., the number of samples that are actually negative but classified as positive by the classifier. *FN* (False negatives): the number of samples incorrectly classified as negative, i.e., the number of samples that are actually positive but classified as negative by

the classifier. TN (True negatives): the number of samples correctly classified as negative, i.e., the number of samples that are actually negative and classified as negative by the classifier. The PR curve is obtained by taking the Recall value as the horizontal axis and the Precision value as the vertical axis. AP is the integral of the PR curve, i.e., the area enclosed by the curve; in the experiments of this paper. The size of the input image is 640×640, in which the initial learning rate is 0.01. The optimizer is SGD, and the SGD momentum is 0.937. And the learning rate adjustment strategy is cosine annealing strategy, the number of iterations is 250 rounds and batch size is 16.

3.3. Data Processing

The dataset used in the paper is a dataset of foreign objects in a power station within the power grid, containing four types: bird nest, kite, balloon, trash bag. This dataset contains a total of 1111 images of four categories of foreign objects in the transmission tower line, including 103 balloons, 284 kites, 331 trash bags, and 642 bird nests. Due to the insufficient number of samples of the dataset of this study, the imbalance in the number of each category of foreign objects exists. Through the data enhancement methods such as mirroring, cropping, scaling, panning and rotating, the spatial geometric transformations were used to change the spatial position of the pixels in the images without changing the content of the images, increasing the number of samples and avoiding the occurrence of overfitting [35]. Field-captured images of transmission lines often suffer from external interferences, leading to issues such as unclear visuals, motion-induced distortions, and weather-related blurriness. To address these real-world complexities, this study implements image preprocessing techniques, including Gaussian noise addition, random luminance adjustment, and motion blur simulation. These methods enrich the training dataset, enabling the model to learn and recognize foreign object features more effectively in varied environmental conditions. This approach simulates actual field scenarios, thereby enhancing the model's validation accuracy. The preprocessing resulted in a final dataset of 5084 images, categorized by different types of foreign objects as follows Table 1:

Table 1. Number of foreign object categories.

Object Category	Object Number	Object Ratio
bird nest	2289	44.68%
kite	1168	22.79%
balloon	607	11.84%
trash bag	1020	19.91%

In this paper, the dataset is divided into training set, validation set, and test set according to the ratio of 8:1:1. The training set and validation set are used for model training, and the test set is used for testing the accuracy of the model. The dataset annotation tool used in this paper is LabelImg, and after the annotation is completed, corresponding txt file will be generated for each image after labeling.

3.4. Experimental Result and Comparative Analysis

3.4.1. Comparative Experiments

Figure 8 shows the training results of the proposed model $mAP50$, from which it can be seen that the training model gradually converges after 220 epochs, and the map value is 97.87% at the 300th epoch. And Figure 9 shows the PR curve, from which it can be seen that among the recognized objects. The recognition accuracy is slightly lower in comparison with the garbage bags, kites, and balloons, due to the existence of the phenomenon of occlusion in some of the bird's nests.
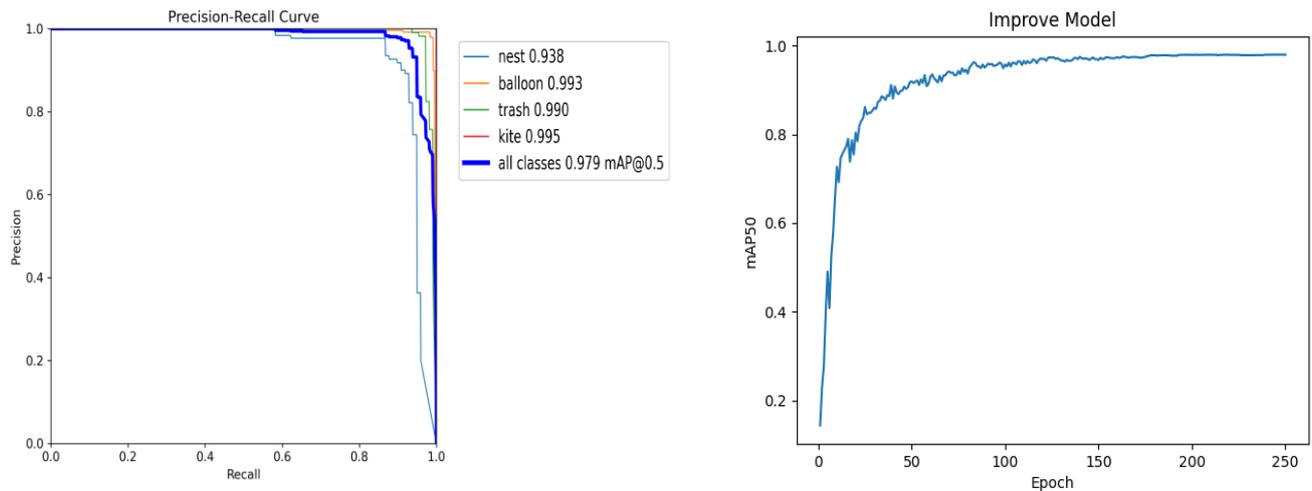
Figure 8. Experimental results of RCDAM-Net model.

Figure 9. Example image of RCDAM-Net model detection.

This paper conducts eight comparative experiments, benchmarking our proposed method against classical models: YOLOV5, YOLOV6, YOLOV7, YOLOV8, FCOS, SSD, RetinaNet, and Faster R-CNN. The comparative analysis, detailed in Table 2 below, demonstrates that our method outperforms these models in *mAP*50, *mAP*50-95, Recall, and Precision metrics. Our model's distinctive advantage lies in the adoption of the RevCol framework, which overcomes the traditional feature extraction network's tendency to lose some features. We have enhanced the RevCol detection capability by integrating a feature fusion part using the C2f module. Additionally, for targets with extreme aspect ratios, we replaced the BottleNeck in the C2f structure with the DySnakeConv module. The detection head incorporates a decoupled detection head and an auxiliary training head, contributing to the model's superior performance. Experimental results validate that richer feature information significantly improves the detection of foreign objects on transmission tower lines.

Table 2. Comparison of results of common models.

Models	*mAP*50	*mAP*50-95	Recall	Precision
YOLOv5	0.9545	0.6836	0.8978	0.9636
YOLOV6	0.9107	0.6397	0.8804	0.9110
YOLOV7	0.9471	0.6457	0.8896	0.9287
YOLOV8	0.9686	0.6923	0.9310	0.9580
FCOS	0.9576	0.6533	0.8897	0.9297
SSD	0.7570	0.5673	0.6571	0.7632
Faster R-CNN	0.9528	0.6523	0.8923	0.9312
RetinaNet	0.9253	0.6399	0.8838	0.9245
Ours model	0.9798	0.7211	0.9516	0.9815

3.4.2. Ablation Experiments

To rigorously assess the impact of our proposed module on model performance, we conducted a series of ablation experiments. In order to verify the effectiveness of each module we used, we selected the YOLOV8 model with the best effect in the comparative experiment as a benchmark and compared each improved module with the unimproved YOLOV8. The experimental results are shown in Table 3, where "-" represents the original structure of YOLOV8, and "+" represents the module used in this paper.

Table 3. Single improvement comparison result.

YOLOV8	+RevCol	+C2D	+Detection Head	+MPDIoU	*mAP*50	Group
+	-	-	-	-	0.9686	A
+	+	-	-	-	0.9712	B1
+	-	+	-	-	0.9702	C1
+	-	-	+	-	0.9692	D1
+	-	-	-	+	0.9689	E1

Group A: YOLOV8 was used for detection, and the model *mAP*50 was 0.9686;

Group B1: Using the RevCol network to replace the backbone network of YOLOV8, the *mAP*50 value is 0.9712;

Group C1: Replacing the C2f structure of YOLOV8 with the C2D structure, the *mAP*50 value is 0.9702;

Group D1: Replacing the detection head of YOLOV8 with the detection head designed in this article, the *mAP*50 value is 0.9692;

Group E1: Replacing the bounding box loss function of YOLOV8 with the MPDIoU loss function, the *mAP*50 value is 0.9689.

In order to verify the overall detection effect of the module we designed, after we gradually added the improved structure to the design model, the *mAP*50 value of the model increased, which proved the effectiveness of the model we designed for the transmission tower line foreign matter dataset. In these experiments, we divided the proposed algorithm structure into five distinct groups for systematic comparative analysis. The experimental results are shown in Table 4, where "-" means that the module does not exist, and "+" means that the module is included.

Table 4. Improve results step by step.

YOLOV8	+RevCol	+C2D	+Detection Head	+MPDIoU	*mAP*50	Group
+	-	-	-	-	0.9686	A
+	+	-	-	-	0.9712	B2
+	+	+	-	-	0.9755	C2
+	+	+	+	-	0.9785	D2
+	+	+	+	+	0.9798	Ours model

Group A: YOLOV8 was used for detection, and the model *mAP*50 was 0.9686;

Group B2: Using the RevCol network to replace the backbone network of YOLOV8, the *mAP*50 value is 0.9712;

Group C2: Replacing the C2f structure of YOLOV8 with the C2D structure based on the structure of Group B2, the *mAP*50 value is 0.9755;

Group D2: Based on the structure of Group C2, the detection head was replaced with a decoupled detection head, and an auxiliary training head was added for training. The *mAP*50 value was 0.9785;

Ours model: Based on group D2 experiments, the bounding box loss function is replaced with MPDIoU loss function, and the *mAP*50 value is 0.9798.

3.4.3. Detection Result

Figure 9 shows the detection results of the proposed model RCDAM-Net. The detection pictures represent the types of foreign objects that are easy to be attached to common transmission tower lines, and this result is able to check the practical effectiveness of the model.

4. Conclusions

In this paper, we propose a novel method for foreign object detection on transmission towers and lines, called RCAM-NET. This method addresses the limitations of main-stream models, particularly in terms of feature extraction that focuses on information compression or loss. We adopt RevCol as the core network, ensuring that feature information is not lost during the process of target feature decoupling. To effectively detect targets with extreme aspect ratios, we incorporate DySnakeConv, enhancing the extraction of features from thin and weak targets. We embed DySnakeConv into a C2f structure and rename it C2D, which allows for the fusion of both detailed and semantic features. To address the challenge of optimizing bounding boxes with the same aspect ratio but different dimensions, we utilize MPDIoU, improving the regression performance of the model's bounding boxes. Considering the overall detection performance of the model, we introduce a decoupled head for detection, separating the classification task from the regression task. We also add an auxiliary training head to enhance detection accuracy. The RCAM-NET method achieves results on a dedicated dataset for foreign objects on transmission towers and lines, with a Map50 value of 97.98%, recall rate of 72.11%, and precision rate of 95.16%. It demonstrates superior robustness compared to other models. Future research will focus on improving the detection of bird nests with occlusion, balancing datasets with complex backgrounds, and developing lightweight versions of the model to meet mobile deployment needs.

Author Contributions: W.Z., Y.L. and A.L. contributed to the study conception and design. Material preparation, data collection, and analysis were performed by W.Z., Y.L. and A.L. The first draft of the manuscript was written by W.Z. and all authors commented on previous versions of the manuscript. All authors have read and agreed to the published version of the manuscript.

References

1. Jalil, B.; Leone, G.R.; Martinelli, M.; Moroni, D.; Berton, A. Fault Detection in Power Equipment via an Unmanned Aerial System Using Multi Modal Data. *Sensors* **2019**, *19*, 3014. [CrossRef] [PubMed]
2. Menendez, O.; Cheein, F.A.A.; Perez, M.; Kouro, S. Robotics in Power Systems: Enabling a More Reliable and Safe Grid. *IEEE Ind. Electron. Mag.* **2017**, *11*, 22–34. [CrossRef]
3. Mann, B.J.; Morrison, I.F. Digital calculation of impedance for transmission line protection. *IEEE Trans. Power Appar. Syst.* **1971**, 270–279. [CrossRef]
4. Wale, P.B. Maintenance of transmission line by using robot. In Proceedings of the 2016 International Conference on Automatic Control and Dynamic Optimization Techniques (ICACDOT), Pune, India, 9–10 September 2016; pp. 538–542.
5. Xie, X.; Liu, Z.; Xu, C.; Zhang, Y. A multiple sensors platform method for power line inspection based on a large unmanned helicopter. *Sensors* **2017**, *17*, 1222. [CrossRef] [PubMed]
6. Alhassan, A.B.; Zhang, X.; Shen, H.; Xu, H. Power transmission line inspection robots: A review, trends and challenges for future research. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105862. [CrossRef]

7. Rublee, E.; Rabaud, V.; Konolige, K.; Bradski, G. ORB: An efficient alternative to SIFT or SURF. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; IEEE: Piscataway, NJ, USA, 2011; pp. 2564–2571.

8. Kumar, N.S.; Shobha, G.; Balaji, S. Key frame extraction algorithm for video abstraction applications in underwater videos. In Proceedings of the 2015 IEEE Underwater Technology (UT), Chennai, India, 23–25 February 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 1–5.

9. Changan, L.; Jin, S.; Hua, W.; Guotian, Y.; Chunyang, L. Research on keyframes extraction pretreatment of power-tower in flying robot inspection video. *J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.)* **2015**, *43*, 477–480. [CrossRef]

10. Lijun, J.; Chunyu, Y.; Shujia, Y.; Wenhao, Z. Recognition of Extra Matterson Transmission Lines Basedon Aerial Images. *J. Tongji Univ. (Nat.)* **2013**, *41*, 277–281.

11. Wanguo, W.; Jingjing, Z.; Jun, H.; Liang, L.; Mingwu, Z. Broken strand and foreign body fault detection method for power transmission line based on unmanned aerial vehicle image. *J. Comput. Appl.* **2015**, *35*, 2404–2408.

12. Yongsheng, Z.; Haiqing, X.; Ligang, W.; Ruizhi, Y.; Chong, L. Application of Hough's Linear Transform-based Foreign Object Recognition on Transmission Lines. *Digit. Technol. Appl.* **2017**, 127–129. [CrossRef]

13. Shengxi, J.; Haiyang, W. Research on foreign object recognition of transmission line based on ORB algorithm. *Sci. Technol. Eng.* **2016**, *16*, 236–240.

14. Xinfu, L.; Richeng, L.; Shixuan, D.; Jing, Z.; Guanfei, Y. Research on foreign object recognition method of power line based on digital image processing. *Electr. Eng.* **2022**, *23*, 73–78.

15. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]

16. Jinglin, H.; Xiangang, P.; Shengchao, J.; Haoliang, Y. Transmission Line Fault Classification Based on Deep Learning and Imbalanced Sample Set. *Smart Power* **2021**, *49*, 114–119.

17. Xinlan, J.; Wenbo, J. Machine Vision Detection Method for Foreign Object Intrusion in High-Speed Rail Contact Net. *Comput. Eng. Appl.* **2019**, *55*, 250–257.

18. Zhenmin, Z.; Liangkai, X. Detection of birds' nest in catenary based on relative position invariance. *J. Railw. Sci. Eng.* **2018**, *15*, 1043–1049. [CrossRef]

19. 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), Beijing, China, 26–28 November 2017.

20. Gangjun, G.; Shuai, Z.; Qiuxin, W.; Zhimin, C.; Ren, L.; Chang, S.; Alhassan, A.B. TensorFlow-based foreign object recognition for high-voltage transmission lines. *Electr. Power Autom. Equip.* **2019**, *39*, 204–209. [CrossRef]

21. Zengxiang, X.; Qifeng, X. Recognition of Foreign Objects Intrusion in Substation Based on Improved Convolutional Neural Network. *Sci. Technol. Eng.* **2022**, *22*, 1465–1471.

22. Maodong, S.; Pei, J.; Xinyang, F.; Junling, Z.; Fankui, G.; Xia, L.; Alhassan. A New Transmission Line Foreign Object Detection Network Structure: TLFOD Net. *Jisuanji Yu Xiandaihua* **2019**, 118–122.

23. Yingchun, Z.; Siyu, S.; Shuai, L.; Zhiyong, L.; Yongliang, X.; Ching, H.W.; Alhassan. Recognition of Bird's Nest on Transmission Tower in Aerial Image of High-volage Power Line by YOLOv3 Algorithm. *J. Guangdong Univ. Technol.* **2020**, *37*, 42–48.

24. Qiuyan, Z.; Zhu, T.; Xiao, S.; Yang, Z.; Zeng, H.; Chi, Z.; Li, G. Foreign object detection of high voltage transmission line based on improved YOLOv4 algorithm. *Appl. Sci. Technol.* **2023**, *50*, 59–65.

25. Hongmin, Z.; Hao, Z.; Shunyuan, L.; Pingping, L. Improved YOLOv3 foreign body detection method in transmission line. *Laser J.* **2022**, *43*, 82–87. [CrossRef]

26. Zheng, T.; Huilin, Z.; Lixin, M.; Jinzhi, L.; Hao, W. Identification of Foreign Objects on Transmission Lines Using Lightweight Network Algorithm. *Electron. Sci. Technol.* **2023**, *36*, 71–77. [CrossRef]

27. Yanzhen, Y.; Zhibin, Q.; Yinbiao, Z.; Xuan, Z.; Qing, W. Foreign Body Detection for Transmission Lines Based on Convolutional Neural Network and ECOC-SVM. *Smart Power* **2022**, *50*, 87–92.

28. Zamir, A.R.; Sax, A.; Shen, W.; Guibas, L.J.; Malik, J.; Savarese, S. Taskonomy: Disentangling task transfer learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3712–3722.

29. Cai, Y.; Zhou, Y.; Han, Q.; Sun, J.; Kong, X.; Li, J.; Zhang, X. Reversible Column Networks. *arXiv* **2023**, arXiv:2212.11696.

30. Hinton, G. How to represent part-whole hierarchies in a neural network. *Neural Comput.* **2023**, *35*, 413–452. [CrossRef]

31. Chang, B.; Meng, L.; Haber, E.; Ruthotto, L.; Begert, D.; Holtham, E. Reversible Architectures for Arbitrarily Deep Residual Neural Networks. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.

32. Liu, Z.; Mao, H.; Wu, C.; Feichtenhofer, C.; Darrell, T.; Xie, S. A convnet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 11976–11986.

33. Qi, Y.; He, Y.; Qi, X.; Zhang, Y.; Yang, G. Dynamic Snake Convolution based on Topological Geometric Constraints for Tubular Structure Segmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Vancouver, BC, Canada, 17–24 June 2023; pp. 6070–6079.

34. Xing, B.; Wang, W.; Qian, J.; Pan, C.; Le, Q. A Lightweight Model for Real-Time Monitoring of Ships. *Electronics* **2023**, *12*, 3804. [CrossRef]
35. Zeng, G.; Yu, W.; Wang, R.; Lin, A. Research on mosaic image data enhancement for overlapping ship targets. *arXiv* **2021**, arXiv:2105.05090.