

Improved Detection of Ice Accretion Status on Transmission Lines Using YOLOv5s

Yukai Lu *, Yongbiao Zhao, Yun Ju

School of Control and Computer Engineering, North China Electric Power University, Changping District, Beijing 102206, China

* Corresponding author: Yukai Lu (Email: 1360444162@qq.com)

Abstract: Ice accretion on transmission lines is one of the major hidden dangers to the safe operation of power systems. The current ice image monitoring system of power companies urgently needs to quickly and accurately detect the ice accretion status from massive real-time image data to assist in the smooth implementation of de-icing work. Based on the original dataset constructed with the help of Shanxi Power Company, this paper establishes a transmission line ice accretion image dataset through image screening, image enhancement, image sliding window segmentation, and image annotation. On the basis of the original YOLOv5s model, this paper proposes three improvements: 1. Designing a feature enhancement module based on atrous convolution to increase the receptive field and obtain more contextual information while preserving the texture details of the feature map. 2. This paper simplifies cross-scale connections and contextual information weighting operations in the feature pyramid to enhance the network's feature extraction capabilities. 3. Introducing Swin-Transformer network structure in the process of feature fusion, further enhancing the semantic information and global perception of small targets. It can be seen from the ablation experiment analysis that the mAP_{0.5:0.5} of the proposed algorithm has increased by 7.6% compared with the original YOLOv5s. At the same time, the performance comparison experiment validated that the detection accuracy of the proposed algorithm achieved better performance with little sacrifice in detection speed.

Keywords: Object Detection; Attention Mechanism; Multi-scale Feature Fusion; Transmission Line Ice Accretion.

1. Introduction

Transmission lines at high altitudes are prone to ice accretion during winter, especially in areas with low temperatures and high humidity. Ice accretion on transmission lines and electrical towers leads to a significant reduction in the insulation performance of electrical equipment. If the ice accumulation exceeds the designed anti-icing capability, it may cause accidents such as line overloads, broken wires, and tripping [1]. In recent years, the State Grid has established an ice accretion monitoring system for transmission lines [1-5]. Image monitoring generally includes long-distance line inspection conducted by workers, drone inspections, and the use of fixed industrial cameras installed at specific locations. Remote collection of on-site images is then performed through network transportation. Image monitoring eliminates the risks associated with manual inspections to some extent while providing rich data support for image analysis [6].

Currently, ice accretion monitoring research can be divided into three categories: mechanical methods, traditional image processing techniques, and deep learning-based image detection and segmentation. Ice accretion detection on transmission lines is one of the important research directions for the State Grid Company. After years of domestic and international research, it has been found that utilizing image processing technology for ice accretion detection on transmission lines is a more intuitive and reliable method compared to other approaches. However, this method also faces many practical challenges. First, long-distance aerial photography results in small proportions of target information within the images, which necessitates improvements in small-target detection capabilities. Second, aerial images have large resolutions, but most object detection algorithms have strict limitations on input resolution values. Third, there are significant differences in ice accretion equipment's shapes

and scale features. Fourth, to achieve a higher target recognition capability, object detection algorithms often have complex network structures. This leads to a large number of parameters and computations, as well as a need for more substantial storage space for weight files after training. This presents practical deployment challenges for embedded devices and other devices with limited storage and computational capabilities. Based on these considerations, the main tasks of this study are as follows:

First, we construct a dataset based on the on-site inspection images of transmission lines provided by Shanxi Power Company and a set of ice accretion images collected by the micro-meteorological laboratory. Through image screening, image enhancement, sliding window segmentation, and image annotation, we create a comprehensive dataset. Next, based on the lightweight YOLO-V5s model, we replace the SPPF module in the backbone network with the ASPP module, incorporating the concept of atrous convolution in the spatial pyramid pooling structure. This ensures that the receptive field is increased without causing a reduction in resolution. Subsequently, we incorporate a streamlined and efficient BiFPN [8] multi-scale feature fusion network, built upon FPN [7], enabling effective cross-scale connections and contextual information weighting processes that help prevent the loss of considerable semantic information for small targets. Finally, during the multi-scale feature fusion process, we incorporate the C3STR (Cross stage partial bottleneck with 3 convolutions and Swin Transformer) module with Swin Transformer [9] network characteristics to enhance the network's local perception capabilities and improve the detection accuracy for small-scale targets.

2. Related Techniques and Theories

YOLO v5 belongs to the family of lightweight models,

featuring fast execution satisfying real-time accurate detection requirements while maintaining low memory occupancy. This ensures the model can be ported to mobile devices, offering more portable and flexible applications. YOLO v5 includes four models: YOLO v5s, YOLO v5m, YOLO v5l, and YOLO v5x, with progressively increasing network width and depth. In this study, YOLO v5s is chosen as the base model. The YOLO v5s network architecture is comprised of Input, Backbone, Neck, and Output components.

3. Improved Ice Accretion Detection on Transmission Lines Using YOLOv5s

3.1. Atrous Convolution-based Feature Enhancement Module

YOLOv5s utilizes multi-scale extraction and BiFPN, enriching feature map information. However, with varying target sizes and complex backgrounds, the custom transmission line ice accretion dataset may result in missed detections and false positives. To improve performance, this study designs a feature enhancement module based on atrous convolution. Increasing the receptive field with atrous convolution enables greater extraction of target contextual information, reducing missed detections and improving false-positive performance. The proposed module is shown in Figure 1

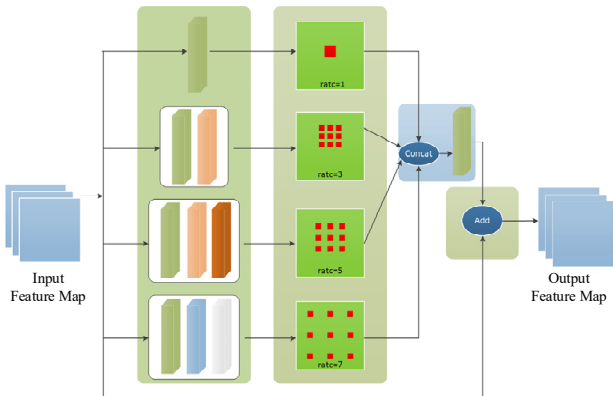


Fig 1. Atrous convolution feature enhancement module

This study develops a four-stage feature enhancement module for input maps using a multi-branch structure, reducing channels, enhancing non-linearity, and modifying dimensions. By replacing larger kernels with smaller convolutions, it simplifies the model and accelerates training. These stages, along with the residual connections, consolidate the model's discriminative power and multi-scale contextual information capture. The second part aims to maintain feature map resolution while capturing more extensive contextual information using dilated 3x3 convolution kernels. Different dilation rates provide varying receptive fields for detecting varied object sizes. In the third part, the module fuses multi-branch structures along the channel dimension, leveraging the semantic information from different scales, which enhances localization and classification accuracy. A 1x1 convolution reduces channel dimensions, and a residual connection is added, keeping the channel number unchanged. This facilitates improved ice accretion detection, especially regarding localization and classification accuracy.

3.2. Weighted Bidirectional Feature Pyramid Network (BiFPN) Model

This study employs a simplified cross-scale connection and contextual information weighting operation in the feature pyramid, as shown in Figure 2. The dashed outline represents the bidirectional feature pyramid model used in this study, mainly consisting of four parts: top-down process, cross-scale connection process, contextual information weighting, and bottom-up process.

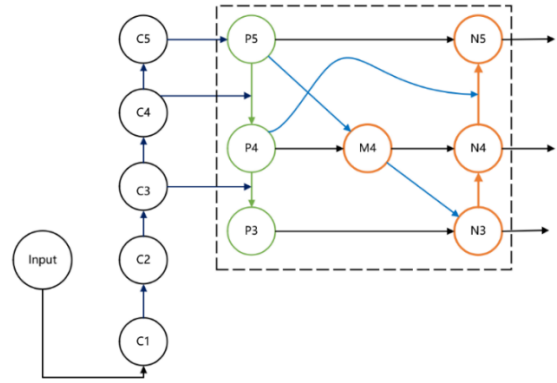


Fig 2. Weighted Bidirectional Feature Pyramid Network Model

The model involves the following steps: first, input a pixel value of, and obtain different-sized feature maps C1, C2, C3, C4, C5 through multiple convolutional downsampling operations, retaining the last three layers for model construction. Next, perform top-down processing by reducing the dimensionality of feature map C5 using 1x1 convolutions, and obtain feature map P5. Then, upsample P5 and fuse it with C4, extracting cross-stage local network features to get P4. Upsample P4, and fuse it with C3 to obtain P3. Use convolutions to achieve cross-scale connectivity between feature maps P4 and P5, selecting M4 as their feature map. Assign different weight information to input streams N3, N4, and N5 using fast normalization fusion, obtaining various feature map representations. Finally, combine the interaction information between top-down and bottom-up process feature maps, outputting the result feature map.

The calculation formula for the fast normalization fusion method is shown below.

$$O = \sum_i \frac{\omega_i \times X_i}{\varepsilon + \sum_j \omega_j} \quad (1)$$

In this formulation, O denotes the output feature map, ω_i represents the weight coefficient of the current layer feature map, X_i corresponds to the feature maps that need to be fused, and $\varepsilon < 0.001$.

In Figure 2, the fusion situations of M4 and N4 are shown in the formulas below, respectively.

$$M_4 = Conv\left(\frac{\omega_1 \times P_4^{in} + \omega_2 \times Resize(P_5^{in})}{\omega_1 + \omega_2 + \varepsilon}\right) \quad (2)$$

$$N_4 = Conv\left(\frac{\omega'_1 \times P_4^{in} + \omega'_2 \times P_4^{td} + \omega'_3 \times Resize(P_3^{out})}{\omega_1 + \omega_2 + \omega_3 + \varepsilon}\right) \quad (3)$$

In this setup, M4 represents the intermediate features of the 4th layer in the top-down fusion process, and N4 denotes the intermediate features of the 4th layer in the bottom-up fusion process. P_i^{in} and P_i^{out} represent the input and output of the i -th layer feature map, respectively; P_i^{td} refers to the output of

the intermediate node of the i -th layer feature map in the top-down fusion process; Resize indicates the operation of maintaining consistent feature map dimensions through sampling; $\varepsilon = 0.0001$.

3.3. C3STR Model

As the network deepens and undergoes multiple convolution operations, most of the target features that small objects in remote sensing images should possess are lost in the high-level feature maps. Therefore, in the feature fusion part, the concept of Swin Transformer [9] is employed. The C3 convolution block incorporates it, employing the C3STR structure as a supplementary module while introducing select discrete parameters from the Transformer. The window self-attention module serves to improve the semantic information and feature representation for small targets. The improved convolution structure is shown in Figure 3

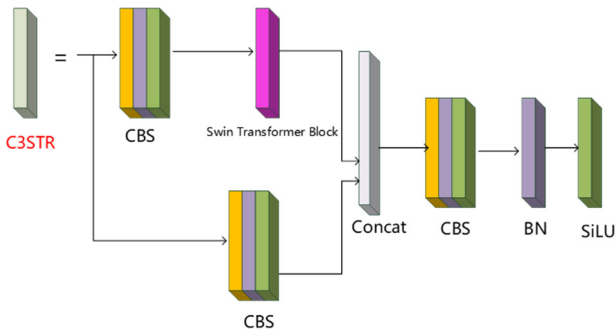


Fig 3. C3STR Model Structure

Comprising paired Window Multi-head Self-Attention (W-MSA) modules, Shifted Window Multi-head Self-Attention (SW-MSA) modules, and Multi-layer Perceptrons (MLP), the Swin Transformer Block (STB) is designed with residual connections in its modules. With a local window size of 7 and a multi-layer perceptron hidden layer embedding dimension set to 4, the STB forms an optimized structure. The calculation process for the multi-head self-attention mechanism is as follows:

$$Attention(Q, K, V) = SoftMax\left(\frac{QK^T}{\sqrt{d} + B}\right)V \quad (4)$$

Table 1. Results of the ablation experiment of the improved model

| Serial Number | Algorithm | Recall | Precision | mAP0.5:0.5 |
|---------------|---------------------------|--------|-----------|------------|
| T0 | YOLOV5s | 63.25% | 50.38% | 57.51% |
| T1 | YOLOV5s+ASPP | 74.42% | 52.79% | 59.34% |
| T2 | YOLOV5s+Bi-FPN | 75.31% | 51.04% | 57.94% |
| T3 | YOLOV5s+C3STR | 78.65% | 52.28% | 59.76% |
| T4 | YOLOV5s+ASPP+Bi-FPN | 76.06% | 51.89% | 61.45% |
| T5 | YOLOV5s+ASPP+C3STR | 76.42% | 53.60% | 63.66% |
| T6 | YOLOV5s+Bi-FPN+C3STR | 79.67% | 54.65% | 64.90% |
| T7 | YOLOV5s+ASPP+Bi-FPN+C3STR | 81.27% | 54.53% | 64.69% |

By analyzing Table 1, we can draw the following conclusions:

(1) From the experimental results of groups T0 to T3, the

(Q, K, V) represent (Query, Key, and Value) matrices, respectively; d denotes the number of input feature map channels, and B is the relative position bias. By introducing B , a significant improvement in performance can be achieved. Compared to the Multi-head Self-Attention (MSA) module in traditional Transformers, the C3STR module in this improved network architecture controls the calculation area in each window by dividing local windows, enabling cross-window information exchange while reducing computational complexity and network cost. This approach allows the model to efficiently manage small targets and complex scenes in remote sensing images without overburdening the computation.

4. Experimental Results

4.1. Experimental Environment and Parameter Settings

The experimental environment used in this study includes the Windows 10 operating system, an NVIDIA GeForce RTX 3060 GPU, CUDA 12.0, the PyTorch 1.10.2 deep learning framework, and Python 3.6.0.

4.2. Datasets and Evaluation Indicators

With assistance from Shanxi Electric Power Company, an original image dataset was created using handheld cameras, drones, and simulated ice accretion images. After image screening, augmentation, sliding window segmentation, and annotation, the final dataset contains 6100 images. The dataset includes normal conductors, normal insulators, ice-conductors, and ice-insulators, with a train-test-validation split of 8:1:1. This study uses AP, mAP, and detection rate as evaluation metrics.

4.3. Experimental Results and Analysis

4.3.1. Results and Analysis of Ablation Experiments on Improved Models

The ablation experiment results of the method proposed in this study on the transmission line ice accretion dataset are documented.

recall, classification accuracy, and mAP of the improved algorithm in groups T1, T2, and T3 are higher than those of the original YOLOv5s model. This indicates that the ASPP, Bi-FPN, and C3STR modules proposed in this study all

contribute to improving the target detection performance. The improvement in localization capability is particularly significant, while the enhancement in classification capability is relatively smaller.

(2) By observing the experimental results of groups (T1, T2, T4), the combination of ASPP and Bi-FPN shows better recall performance than using them separately. Comparing the results of groups (T2, T3, T6), the combination effects are also superior to the separate ones in terms of classification accuracy. In terms of mAP, the stacking of different modules outperforms individual modules.

(3) Comparing the experimental results of group T7 and groups (T4, T5, T6), although the performance of the proposed method on precision and mAP is slightly lower than that of group T6, the recall of group T7 is 1.6% higher. This indicates that the improved algorithm effectively reduces the possibility of missed detections in ice accretion target detection.

(4) Observing the experimental results of groups T0 and T7, the improved algorithm has higher recall, accuracy, and mAP than the original YOLOv5s model, with increases of 18.02%, 4.15%, and 7.18%, respectively. This demonstrates the effectiveness and feasibility of the proposed method.

Figure 4 shows the change in accuracy during the training process of the YOLOv5s and improved-YOLOv5s models. As can be seen from Figure 4, as the number of epochs increases, the classification accuracy of the improved-YOLOv5s model gradually increases and performs better than the YOLOv5s model. This further validates the contribution of the proposed modifications in the improved-YOLOv5s model, resulting in a more accurate object detection for ice accretion on transmission lines.

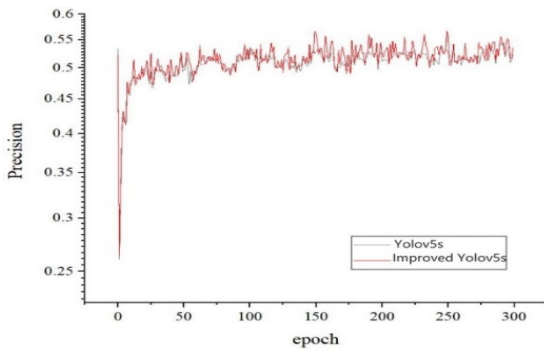


Fig 4. Comparison of Model Accuracy Change Process

Figure 5 shows the comparison of the AP for various detection categories between YOLOv5s and improved-YOLOv5s models, reflecting the specific distribution of mAP values. The following analysis can be drawn from Figure 5:

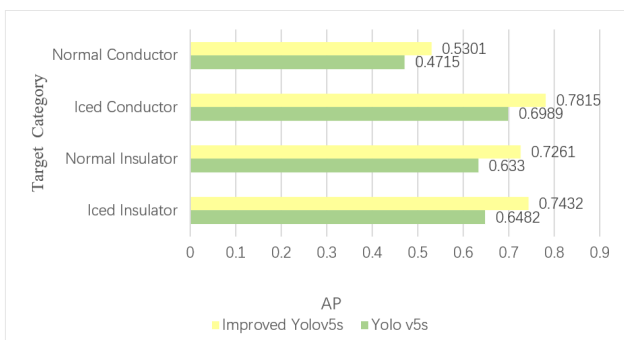


Fig 5. Comparison of Average Precision (AP) for Different Categories Detection

(1) In different detection categories, the AP values of the improved-YOLOv5s model are higher than those of the YOLOv5s model, indicating that the proposed method in this study generally improves the recognition capabilities.

(2) Although the mAP of the improved-YOLOv5s model is only 64.69%, the detection performance for some individual subclasses can still reach around 75%.

4.3.2. Model Performance Comparison Experimental Results and Analysis

To further verify the robustness and generalization capabilities of the model, an additional 50 ice-covered transmission line images outside the test set are selected for evaluating the network model's detection performance. Table 2 presents the performance comparison of different models under the same conditions, including both accuracy and speed metrics.

Table 2. Model Performance Comparison

| Algorithm | Detection Speed/fps | mAp/% |
|---------------------------|---------------------|-------|
| SSD | 46 | 48.26 |
| Faster R-CNN | 7 | 50.61 |
| YOLOv3 | 30 | 51.67 |
| YOLOv4 | 28 | 63.42 |
| YOLOv5s | 48 | 61.36 |
| Algorithm In This Article | 34 | 64.81 |

The detection speed of the model proposed in this study is 14 fps lower than the original YOLOv5s model due to the increased number of parameters introduced by the attention module. However, the overall performance is still superior to YOLOv4, YOLOv3, and Faster R-CNN. At the same time, the mAP value has increased by 3.45% compared to the original YOLOv5s algorithm, achieving the highest mAP metric value among all comparison models. This demonstrates that the proposed model successfully maintains a balance between detection speed and accuracy, providing a more effective solution for ice accretion detection tasks.

5. Summary and Discussion

In response to the urgent need of electric power companies to identify ice accretion conditions in transmission line images, this study proposes an improved algorithm based on YOLOv5s. Through model ablation experiments and model comparison experiments, this study's algorithm achieves significant improvements in detection accuracy compared to YOLOv5s.

However, the detection speed of the proposed algorithm is somewhat lower than that of YOLOv5s. Also, merely obtaining the location and category information of ice-covered transmission lines in image recognition and status evaluation does not accurately reflect the ice accretion status. In the future, we will research transmission line ice accretion identification methods based on YOLO-optimized edge detection. By extracting the ice accretion edges and calculating ice thickness, we aim to more accurately reflect

the ice accretion status of transmission lines.

References

- [1] Yang Lin, Hao Yanpeng, Li Weiguo, et al. A mechanical calculation model for on-line icing-monitoring system of overhead transmission lines [J]. Proceedings of the CSEE, 2010, 30(19):100-105.
- [2] Zhang Yan, Chi Zhixiang, Zhang Guoyan. Relationship between Daban mountain transmission line ice cover phenomenon and meteorologic condition [J]. Qinghai Electric Power, 2005, 24(1):33-36.
- [3] Cheng Shengfeng. Study and Application Status of On-line Icing-monitoring Technology for Overhead Transmission Lines [C] //Proceedings of the 2018 Annual Conference of Jiangxi Society of Electrical Engineering, Nanchang: China Electrical Engineering Society, 2019.
- [4] Li Hao, Fu Chuang, Liu Xu, et al. Overhead transmission line ice monitoring system of China southern power grid and its operation analysis [J]. Shaanxi Electric Power, 2013, 41(4): 20-23.
- [5] Hua Shenbing, Feng Shuanglei, Jin Shuanglong, et al. Evaluation of transmission line icing prediction products based on fraction skill score [J]. High Voltage Engineering, 2020, 46(4):1450-1457.
- [6] Sun Fengjie, Yang Zhenxun, Li Yuanyuan, et al. Methods of transmission line target recognition [J]. Journal of Image and Graphics, 2012, 17(3): 349-356.
- [7] Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.
- [8] Tan Mingxing, Pang Ruoming, Le Quoc. EfficientDet: Scalable and efficient object detection [EB/OL]. (2019-11-20) [2022-02-12]. <https://arxiv.org/abs/1911.09070>.
- [9] Liu Ze, Lin Yutong, Cao Yue, et al. Swin Transformer: hierarchical vision transformer using shifted windows [C] // 2021 IEEE/CVF International Conference on Computer Vision, 2021: 9992-10002.