IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 1, February, 2026, pp. 433-440
Journal homepage: http://iieta.org/journals/ts

Weakly Supervised Semantic Segmentation of Transmission-Line Insulators Using GrabCut Pseudo-Labels from Bounding Boxes and a Lightweight TransAtt-UNet

Ming Zhang[1], Yi Zhang[2*]

[1] State Grid Shanghai Municipal Electric Power Company, Shanghai 200122, China
[2] School of Economics and Management, Shanghai Electric Power University, Shanghai 200090, China

Corresponding Author Email: zhangyish@126.com

https://doi.org/10.18280/ts.430130

Received: 30 September 2025
Revised: 15 November 2025
Accepted: 9 December 2025
Available online: 28 February 2026

Keywords:
insulator, weakly supervised segmentation, transformer, attention mechanism, U-Net, UAV inspection

ABSTRACT

Insulators in transmission lines are often situated in complex backgrounds in UAV inspection images, with obvious target scale variations and slender structures. Training segmentation models directly relying on pixel-level annotations is usually costly and time-consuming. Aiming at the common engineering data form with only bounding box annotations, this paper constructs a reproducible and comparative framework for weakly supervised insulator segmentation. Firstly, GrabCut is introduced under the constraint of bounding boxes to generate pixel-level pseudo-labels, converting box-level supervision into mask supervision applicable for segmentation network training. Secondly, TransAtt-UNet is designed on the U-Net backbone: MobileNetV2 is adopted in the encoder to reduce the number of parameters and computational overhead; a Transformer encoder is introduced in the bottleneck layer to enhance global context modeling; and a channel-spatial attention module is fused at the skip connections to suppress background noise and strengthen boundary details. Under a unified data processing pipeline, training script and evaluation metrics, the proposed method is compared with baseline networks including U-Net, DeepLabv3+, SegFormer-B0 and MobileNetV2-UNet, and the contributions of key modules are analyzed through ablation experiments. Reproducible experimental settings and a result table template are provided at the end of the paper, facilitating the rapid implementation of subsequent reproducible experiments and paper writing.

1. INTRODUCTION

Insulators in transmission lines perform dual functions of electrical insulation and mechanical support, and their health status is directly related to the reliability of power supply and operational safety. Under the long-term action of environmental factors such as sand and wind, rain and snow, lightning, contamination deposition and temperature fluctuations, insulators may suffer from cracks, damage, aging, flashover and other defects. Once their performance degrades, it may trigger line alarms at the minimum, and cause power outages in severe cases [1, 2]. Therefore, there is an explicit engineering demand for efficient, accurate and automated inspection of insulators.

Traditional manual tower-climbing inspection is characterized by high labor intensity, low efficiency and potential safety risks. In recent years, UAV inspection has gradually become the mainstream method, which can quickly acquire a large number of high-definition images, yet the inspection data also brings new challenges: variable shooting angles, uneven illumination, strong background interference, and insulators typically present slender structures with scale variations [3, 4]. Semantic segmentation can provide pixel-level target region and contour information, laying a more reliable ROI foundation for subsequent defect identification and quantitative evaluation, and thus has become an important research direction in power vision intelligence [5, 6].

Deep learning has driven the rapid evolution of segmentation technologies. Fully Convolutional Network (FCN) realizes end-to-end pixel-level prediction; U-Net balances semantic information and detailed features by virtue of the encoder-decoder structure and skip connections; the DeepLab series improves segmentation performance in complex scenarios through atrous convolution and multi-scale context fusion. However, convolutional structures have an inherent limitation in modeling long-range dependencies, making it difficult to fully utilize global semantic information [7]. In recent years, Transformer has exhibited outstanding performance in visual tasks with its self-attention mechanism, providing a new modeling method for target segmentation in complex backgrounds [8].

On the other hand, pixel-level annotation is extremely costly, and only bounding box annotations are usually available in actual engineering scenarios, which has thus drawn attention to weakly supervised segmentation technologies [9]. How to achieve high-precision segmentation under limited annotation conditions is an important research direction at present. Based on the above background, this paper proposes a weakly supervised insulator segmentation method fusing the image attention mechanism and

Transformer structure, which realizes high-precision pixel-level segmentation with only bounding box annotations, and constructs a complete reproducible experimental framework [10].

2. LITERATURE REVIEW

2.1 Research progress of insulator image segmentation

As an important research direction in power vision intelligence, insulator image segmentation has received extensive attention in recent years. Early studies mostly relied on traditional image processing methods, such as threshold segmentation, edge detection and morphological operations. Such methods are highly dependent on the background, with unstable performance under conditions of illumination changes and complex texture interference, and are thus difficult to adapt to UAV inspection scenarios [11].

With the development of deep learning, the FCN realized end-to-end pixel-level prediction for the first time, laying a foundation for semantic segmentation [12]. The U-Net structure proposed by Ronneberger et al. [13] effectively fuses shallow detailed features and deep semantic information through the symmetric encoder-decoder design and skip connection mechanism, achieving remarkable results in medical image segmentation. Due to its excellent multi-scale representation capability, this structure has been widely applied in industrial inspection and power equipment segmentation tasks.

In the field of power inspection, researchers have made various improvements to U-Net, such as introducing residual structures to enhance feature extraction capability, or combining Feature Pyramid Network (FPN) to improve small target recognition performance. In addition, Mask-RCNN has been applied to insulator instance segmentation, realizing joint optimization of target localization and segmentation through the region proposal mechanism. However, these methods mostly rely on a large number of pixel-level annotated data, resulting in high annotation costs in practical engineering applications.

The DeepLab series models have shown excellent performance in the field of semantic segmentation. DeepLabv3+ fuses features of different receptive fields through the Atrous Spatial Pyramid Pooling (ASPP) module, significantly enhancing multi-scale semantic representation capability [14]. Under complex background conditions, DeepLabv3+ has a stronger context modeling capability than U-Net, and is therefore widely used in power image segmentation tasks [15]. Nevertheless, such methods take convolution as the core, mainly focusing on local receptive field information, and have limited capability in modeling long-range dependencies [16].

2.2 Application of attention mechanism in segmentation tasks

The attention mechanism effectively improves the model representation capability by assigning higher weights to important features. In visual tasks, the attention mechanism is mainly divided into two categories: channel attention and spatial attention [17].

SENet enhances the capability of modeling the relationship between feature channels through the channel attention mechanism, but does not consider spatial dimension information [18]. Subsequently, CBAM (Convolutional Block Attention Module) proposes to combine channel attention and spatial attention simultaneously to realize a dual enhancement structure of channel-spatial. CBAM recalibrates the feature map through the sequential attention mechanism, which can effectively suppress background noise and strengthen target boundaries.

In target segmentation tasks under complex backgrounds, the attention mechanism can significantly improve the quality of boundary expression. Especially in UAV inspection images, insulators are often visually confused with structures such as conductors and towers, and the introduction of the attention mechanism helps the model focus on the target region and improve segmentation accuracy [19].

2.3 Application of transformer in visual segmentation

Transformer was initially applied in the field of natural language processing, with its core being the self-attention mechanism. Dosovitskiy et al. proposed Vision Transformer (ViT), which divides an image into fixed-size patches and inputs them into the Transformer encoder to realize global feature modeling [20]. This method breaks through the limitation of the local receptive field of convolutional networks and exhibits excellent performance on large-scale datasets.

However, ViT has high computational complexity. To improve efficiency, Liu et al. proposed Swin Transformer, which realizes self-attention calculation within local windows by introducing the sliding window mechanism and hierarchical structure, and performs window shifting between different layers. This not only ensures computational efficiency but also enhances global modeling capability [21].

In recent years, the fusion of CNN and Transformer has become a mainstream trend. One type of method introduces the Transformer module in the encoder bottleneck layer to supplement global semantic information; the other adopts a parallel branch structure, using the Transformer branch to extract high-resolution global features to complement the convolutional branch. Experiments show that the fusion structure can significantly improve target connectivity and structural integrity [22].

In the insulator segmentation task, targets usually present slender structures with repeated texture features. Relying only on local convolution is prone to segmentation discontinuity, and the introduction of Transformer can effectively enhance the overall coherence.

2.4 Research status of weakly supervised segmentation

Semantic segmentation is highly dependent on pixel-level annotations, while high-quality annotations are costly. In actual power inspection scenarios, the common annotation forms are bounding boxes or coarse-grained annotations, making weakly supervised segmentation an important research direction [23].

GrabCut is a classic graph cut algorithm that separates the foreground by constructing Gaussian mixture models for the foreground and background and based on the energy minimization strategy [24]. This method can generate relatively accurate foreground masks under the constraint of bounding boxes, and has high practical value in industrial scenarios.

In recent years, weakly supervised methods based on Class Activation Mapping (CAM) and self-training strategies have also emerged, but they usually rely on additional classification networks or complex training strategies, leading to high difficulty in engineering deployment [25]. In contrast, the method of generating pseudo-labels based on GrabCut and combining with deep segmentation network training features a simple structure and high reproducibility, making it more suitable for engineering promotion.

3. DATASET AND PREPROCESSING

3.1 Experimental settings

The insulator dataset used in this paper adopts a directory structure of Train/Val/Test, where VOC-format XML annotations are provided for the Train and Val sets, and only images are included in the Test set [1]. The image resolution is uniformly set to 1152 × 864, with the target category being insulator. Statistics show that the training set contains 520 images with 907 insulator instances; the validation set contains 40 images with 84 instances; and the test set contains 40 images (refer to Table 1). There is 1 sample in the training set with an annotated XML file but a missing corresponding image (which has been excluded in statistics and training).

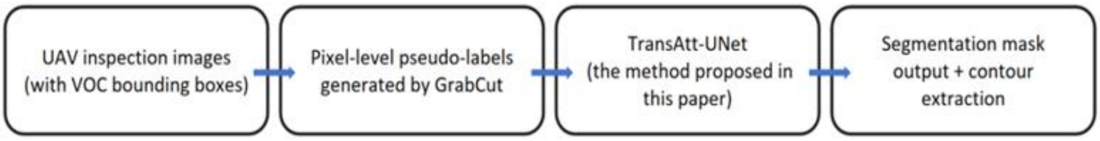
Table 1. Dataset division and annotation statistics

Subset	Number of Images	Number of Instances	Annotation Type
Train	520	907	VOC XML (Bounding box)
Val	40	84	VOC XML (Bounding box)
Test	40	-	No annotation

3.2 Pseudo-label generation

To train the segmentation network with only bounding box annotations, this paper adopts the GrabCut algorithm to construct pixel-level pseudo-labels to replace manual pixel-wise annotations. The specific process is as follows: first, each annotation box is expanded outward by a fixed pixel value to reduce target truncation caused by overly tight box boundaries, forming the initial rectangular region required for GrabCut. Then, the pixels inside the expanded rectangle are set as the foreground candidate region, and the pixels outside the rectangle are set as the definite background, thus giving the initial foreground-background division. Based on this initialization, GrabCut models the color/texture distribution of the foreground and background with GMM respectively, and alternately executes two steps in the iterative process: first, updating the GMM parameters of the foreground and background according to the current segmentation result; second, constructing an energy function including data terms and smooth terms and solving it through the min-cut/max-flow graph cut to realize the re-division of the foreground and background. After the iteration converges or reaches the preset number of rounds, a binary foreground mask is output as the pseudo-label corresponding to the annotation box [5].

When multiple insulator targets exist in a single image, the above GrabCut process is independently executed for each annotation box to obtain the corresponding foreground masks, and the union of all mask results is taken to obtain the final pseudo-label of the entire image. Without introducing additional complex priors or large-scale pre-trained models, this strategy can effectively convert box-level annotations into pixel-level supervision signals applicable for segmentation network training, thus significantly reducing annotation costs while ensuring that the pseudo-labels have a certain boundary refinement capability and trainability. The overall process of weakly supervised insulator segmentation is shown in Figure 1. Examples of training samples and VOC bounding box annotations are shown in Figure 2. Example of the superimposed effect of pseudo-segmentation labels generated based on GrabCut, as shown in Figure 3. The scale distribution of insulator instances in the training set (proportion of bounding box area) is shown in Figure 4.


UAV inspection images (with VOC bounding boxes)

Pixel-level pseudo-labels generated by GrabCut

TransAtt-UNet (the method proposed in this paper)

Segmentation mask output + contour extraction

Figure 1. Overall process of weakly supervised insulator segmentation

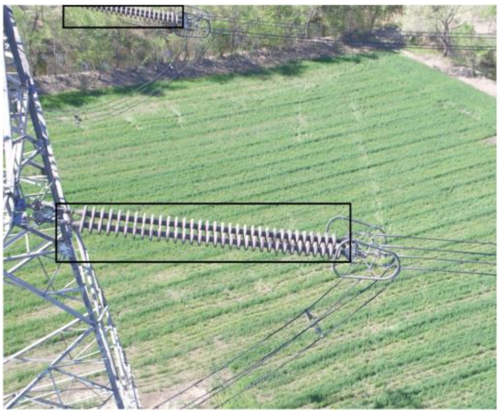


Figure 2. Training examples and VOC bounding box annotation examples

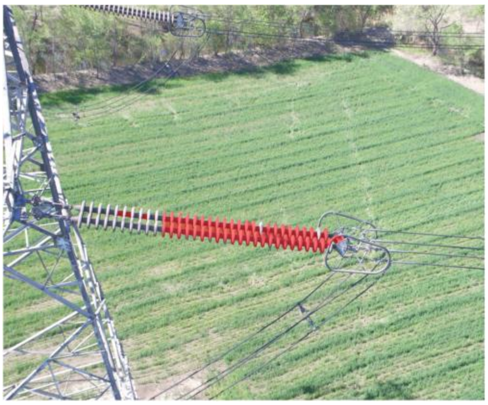


Figure 3. Superposition effect example of pseudo-segmentation labels generated based on GrabCut

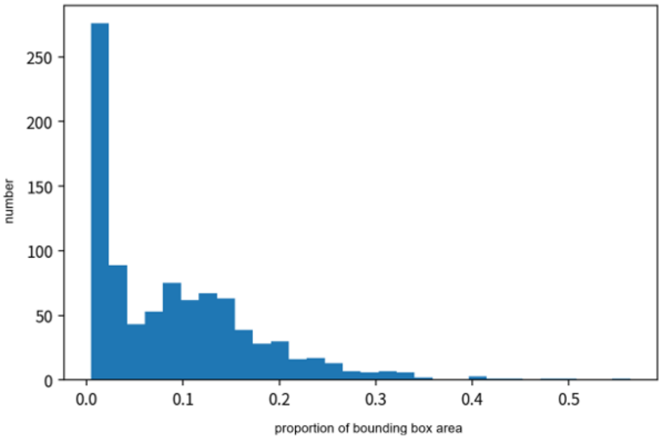


Figure 4. Scale distribution of insulator instances in the training set (proportion of bounding box area)

4. TRANSATT-UNET NETWORK MODEL

4.1 Model introduction

U-Net and its improved models are widely used in segmentation tasks such as medical imaging and industrial inspection. Their encoder-decoder structure can extract semantic information layer by layer and restore spatial resolution in the decoding stage, while skip connections fuse shallow edge textures with deep semantic features, thus balancing localization accuracy and semantic discrimination capability. Precisely because this structure shows stable performance in small-sample and fine-grained boundary scenarios, U-Net is often used as the basic framework for engineering segmentation tasks. However, in UAV inspection images, insulator targets have obvious scene characteristics: first, insulators usually present slender chain or string structures with repeated textures and regular arrangement inside the targets; second, variable inspection angles lead to significant target scale differences, with local occlusion, motion blur and strong light reflection being common; third, structures in the background such as tower components, conductors and insulator fittings are similar to insulators in shape or texture, which is prone to foreground-background confusion. Under these conditions, modeling relying solely on the local receptive field of convolutional networks is often difficult to stably capture the overall connectivity and long-range geometric relationships of targets, manifested as discontinuous segmentation results, boundary adhesion or background false detection, especially prominent in complex backgrounds and small target conditions.

Based on the above problems, this paper introduces Transformer and attention mechanism on the U-Net framework and proposes TransAtt-UNet to improve the segmentation robustness in complex inspection scenarios in the way of lightweight feature extraction + global relationship modeling + key detail enhancement. The specific structural design is as follows.

4.2 Network architecture

(1) Lightweight encoder: MobileNetV2 is adopted to replace the traditional ResNet-based encoders [10]. MobileNetV2 splits standard convolution into depthwise spatial convolution and 1 × 1 pointwise convolution by using depthwise separable convolution, greatly reducing the number of parameters and computational load; at the same time, it reduces information loss while ensuring representation capability through the inverted residual structure and linear bottleneck. For engineering applications such as UAV inspection, the lightweight encoder not only helps improve training and inference efficiency but also is more conducive to subsequent deployment on edge devices or inspection platforms.

(2) Transformer bottleneck: A Transformer Encoder module is introduced on the low-resolution and high-semantic features output by the encoder [13]. Since the computational complexity of low-resolution features is more controllable, it is suitable for carrying the global interaction of self-attention. Transformer establishes associations between arbitrary positions in the feature sequence through the multi-head self-attention mechanism, which can capture long-range dependency relationships and overall structural information. For insulator targets with the characteristics of slenderness, repetition and easy discontinuity, global modeling can enhance the ability to identify target connectivity and geometric consistency, and reduce false detection caused by similar local textures. Meanwhile, the multi-head mechanism allows the model to focus on different relationship patterns in different subspaces—for example, part of the attention focuses on the continuity of chain structures, and another part focuses on the differences from background conductors and towers, thus improving the discrimination robustness in complex backgrounds.

(3) Attention skip fusion: A channel-spatial attention module (e.g., CBAM) is introduced at the skip connections to recalibrate the shallow detailed features from the encoder [12]. The skip connections of the traditional U-Net are usually simple concatenation or addition, which easily introduce background noise into the decoder, leading to misclassification at boundaries or texture interference. CBAM first measures the importance of different feature channels through channel attention, and then emphasizes the response of key regions through spatial attention, which can suppress background structures similar to insulators (such as fittings, conductors and tower material textures) and enhance the response intensity of boundary and slender structure regions. In this way, skip connections no longer transmit details indiscriminately but selectively transmit effective details, improving the quality of boundary restoration and anti-interference capability in mechanism.

(4) Decoder: The spatial resolution is restored through progressive upsampling and convolution fusion. In each stage of the decoding phase, the upsampled high-level semantic features are fused with the shallow detailed features filtered by attention, and multi-scale information is further integrated through convolution to gradually refine the target boundary and restore the complete shape of the insulator, finally outputting a binary segmentation mask. This decoding strategy retains the advantages of U-Net in boundary restoration, and with the supplement of global semantics by Transformer and detail screening by attention, the model is more likely to obtain continuous, complete and low-noise segmentation results in complex inspection backgrounds.

In general, the design idea of TransAtt-UNet is as follows: the encoder uses a lightweight structure to ensure efficiency; the bottleneck layer uses Transformer to supplement the global relationship modeling capability; the skip connections use the attention mechanism to control noise transmission and detail enhancement; the decoder restores the high-resolution mask in multi-scale fusion. This combination not only faces the complexity and engineering deployability of UAV inspection scenarios but also provides a more robust network foundation for high-precision segmentation under weakly supervised conditions.

4.3 Loss function

To address the problem of pixel imbalance between the foreground (insulator) and background, this paper uses a weighted combination of BCE loss and Dice loss: L = λ·L_BCE + (1-λ)·L_Dice. The Dice loss can be defined as: L_Dice = 1 - (2·|P∩G| + ε) / (|P| + |G| + ε), where P is the predicted foreground set, G is the pseudo-label foreground set, and ε is the smoothing term.

5. EXPERIMENTAL DESIGN

Based on the dataset and preprocessing in Chapter 3, this section presents the specific experimental scheme, including comparative models, evaluation metrics and training configurations. All experiments are completed on the same platform and script to ensure reproducibility and comparability.

5.1 Experimental environment and platform

Experiments are conducted on a workstation equipped with an NVIDIA RTX 3060 GPU, with the operating system being Windows 10 64-bit. This paper completes data processing, training and evaluation based on MATLAB R2024b and its Deep Learning Toolbox and Computer Vision Toolbox. All comparative networks are trained and tested through a unified script, in which the data path, network structure and training hyperparameters are explicitly configured to ensure reproducibility consistency.

5.2 Dataset and division

The resolution of original inspection images is 1152 × 864, with the target category being insulator. Under the weakly supervised setting, pixel-level pseudo-labels are first generated based on bounding box annotations and GrabCut to obtain image-mask pairs. The data is organized in a VOC-style directory: RGB images are stored in dataset/images, and corresponding binary masks (0 for background, 255 for insulator) are stored in dataset/masks. To facilitate rapid verification of the code process, if the local directory is missing or empty, the script can automatically call generateToyDataset to generate synthetic strip target data for visual verification; real inspection images with generated pseudo-labels are used in formal experiments.

5.3 Comparative methods

To evaluate the performance of different semantic segmentation frameworks on insulator pseudo-label data, this paper selects three representative baseline models as comparison objects and conducts experiments under the same dataset and training configuration.

(1) U-Net: Constructed by using MATLAB-provided UnetLayers with the encoder depth set to 3. U-Net fuses shallow details and deep semantics through a symmetric encoder-decoder structure and multi-level skip connections, and is a typical baseline model in small-sample segmentation tasks.

(2) SegNet: Constructed by using SegnetLayers(inputSize,numClasses,"vgg16") with VGG16 as the encoder. SegNet performs upsampling by saving pooling indices with a relatively simple design, representing the early semantic segmentation methods based on the encoder-decoder structure.

(3) DeepLabv3+: Constructed by using DeepLabv3+ Layers with ResNet-18 as the base. DeepLabv3+ has a strong multi-scale context modeling capability by using the ASPP and encoder-decoder structure, and is a widely used semantic segmentation network in recent years.

The same input size, category setting (background + insulator) and training/validation division are adopted for the three models to facilitate horizontal comparison. The subsequent baseline results in this paper are used as the reference benchmark for the weakly supervised method TransAtt-UNet.

5.4 Evaluation metrics

Considering that the insulator segmentation task has requirements for both the overall consistency of the target region and boundary accuracy, this paper selects two indicators, the mean Intersection over Union (mIoU) and Dice coefficient, as the main evaluation criteria. Let the predicted foreground set of an image be P and the pseudo-label foreground set be G; for the target category of insulator, the indicators are calculated as follows: IoU=|P ∩ G|/|P ∪ G|，Dice=2 • |P ∩ G|/|P|+|G|. In actual implementation, IoU and Dice are calculated for each validation image as a unit, and then the average is taken for the validation set to obtain mIoU and the average Dice. Larger values of mIoU and Dice indicate that the model segmentation results are closer to the pseudo-labels. Since pseudo-labels are adopted in this paper under weakly supervised conditions, these indicators not only evaluate the relative advantages and disadvantages of the models but also indirectly reflect the usability of the pseudo-labels themselves.

5.5 Training configuration

The training processes of the three baseline models are completely consistent, with the main hyperparameter settings as follows: the input size is set to 256 × 256 with 3-channel RGB; the number of categories is set to 2 (background/insulator); the optimizer is set to Adam; the initial learning rate is set to 1 × 10-3; the learning rate strategy is set to a fixed learning rate of 0.001; the maximum number of training epochs is set to 5; the batch size is set to 4; the validation frequency is set to evaluate the performance on the validation set and record the curve every several training mini-batches. After training, a unified evaluation function is used to calculate mIoU and Dice on the validation set, and a prediction overlay image of each model is output for qualitative analysis. The final quantitative results are summarized in the form of a table.

6. RESULTS AND ANALYSIS

Based on the experimental scheme designed in Chapter 5, this section compares the performance of three classic semantic segmentation baseline models on the insulator pseudo-label dataset, and conducts analysis combined with qualitative visualization results.

6.1 Quantitative comparison results

Table 2 presents the mIoU and Dice indicators of U-Net, SegNet and DeepLabv3+ on the validation set. All results are obtained under the same data division and training configuration.

Table 2. Quantitative comparison of different methods on the validation set

Model	mIoU (%)	Dice (%)
U-Net	51.50	65.99
SegNet	47.20	62.50
DeepLabv3+(ResNet-18)	53.80	68.50

It can be seen from Table 2 that the three classic semantic segmentation networks can all learn a certain degree of foreground-background discrimination capability under the current weakly supervised setting, with mIoU and Dice indicators stably falling in the range of 40% ~ 70%. Specifically: The mIoU of U-Net is 51.50% and the Dice is 65.99%. Under the conditions of only a small number of training epochs and GrabCut pseudo-masks as the label source, U-Net can already give relatively usable insulator segmentation results; the mIoU and Dice of SegNet are 47.20% and 62.50% respectively, which are slightly lower than U-Net overall, consistent with the network structural characteristics of SegNet that mainly relies on pooling indices in the upsampling stage and has limited capability in restoring slender structures; DeepLabv3+ (ResNet-18) achieves the optimal values in both indicators in the current experiments, with an mIoU of 53.80% and a Dice of 68.50%. Compared with U-Net and SegNet, DeepLabv3+ integrates multi-scale context information more fully through atrous convolution and the ASPP module, and has stronger adaptability to inspection scenarios with complex backgrounds and obvious target scale variations.

6.2 Analysis of model performance

The experimental results of the three baseline models on weakly supervised pseudo-labels show that:

The GrabCut pseudo-label scheme is usable: even without introducing complex weakly supervised training techniques, classic CNN segmentation networks can already learn insulator segmentation results with practical engineering significance on this basis;

The network structure has a significant impact on weakly supervised performance: structures capable of modeling multi-scale context and global information (e.g., DeepLabv3+) are more robust under weakly supervised conditions, while structures relying on local textures are more susceptible to interference when pseudo-labels contain noise;

This baseline comparison also indirectly confirms the design motivation of the subsequently proposed TransAtt-UNet: by introducing the Transformer encoder and attention skip connections, it is expected to further strengthen global relationship modeling and key detail expression while maintaining lightweight, making up for the shortcomings of pure CNN architectures in weakly supervised scenarios.

6.3 Result discussion

Two empirical conclusions can be drawn from the comparison and ablation results. First, in the case of noise in pseudo-labels, the network is significantly sensitive to the quality of supervision: structures that can utilize multi-scale context or have stronger feature screening capabilities are usually more robust to pseudo-label errors. Second, insulator targets have the characteristics of slenderness and repeated textures, and the boundary regions are often the most vulnerable to background interference. Therefore, introducing the attention mechanism at the skip connections to control the transmission of detailed features, combined with the global relationship modeling of the bottleneck layer, helps improve the overall segmentation quality while maintaining lightweight. These observations also support the design motivation of TransAtt-UNet: under the constraints of engineering deployability, the combination of lightweight encoding + global modeling + attention detail enhancement is used to improve the usability of weakly supervised scenarios.

7. CONCLUSIONS AND FUTURE WORK

Aiming at the problems of insulators in UAV inspection images, such as complex background, slender target with obvious scale variation, and high cost of pixel-level annotation, this paper constructs a reproducible and comparative framework for weakly supervised segmentation under the condition of only providing bounding box annotations. Specifically, with bounding boxes as the only supervision information, this paper generates pixel-level pseudo-labels under box constraints through GrabCut, converting box-level annotations common in detection scenarios into supervision signals applicable for segmentation training; on this basis, TransAtt-UNet is proposed: MobileNetV2 is adopted to realize lightweight encoding, a Transformer is introduced in the bottleneck layer to enhance global context, and a channel-spatial attention is fused at the skip connections to suppress background noise and strengthen boundary details. Under a unified script and evaluation metrics, this paper conducts comparisons with various baseline networks, analyzes the contributions of key modules through ablation experiments, and provides reproducible experimental settings and a result table template to facilitate subsequent reproduction and paper writing.

Subsequent work can be further advanced in two aspects. First, the quality of pseudo-labels is still the core bottleneck of weakly supervised performance. General segmentation models (e.g., SAM) and scene priors (connectivity, slender structure constraints, etc.) can be combined to screen and correct pseudo-labels, or a small number of pixel-level ground truths can be introduced to form mixed learning of a small amount of strong supervision + a large amount of weak supervision to improve generalization stability. Second, the segmentation results can be further used for defect detection and severity assessment—for example, analyzing the distribution of missing, damage, cracks and contamination after obtaining accurate masks, and fusing multi-source data such as visible light/infrared to establish defect grade indicators, promoting the closed-loop application from target localization-region segmentation to defect identification-risk assessment.

REFERENCES

[1] Yao, X., Li, S. (2026). An object detection method based on improved detection transformer for insulator defect detection of electrical transmission lines. Electric Power Systems Research, 256: 112929. https://doi.org/10.1016/J.EPSR.2026.112929

[2] Li, X., Sun, Y., Liu, X., Wu, X. (2026). Insulator defect detection in transmission line in sandy and dusty scenarios based on deep learning. Electric Power Systems Research, 256: 112852. https://doi.org/10.1016/J.EPSR.2026.112852

[3] Liu, K., Ma, Z., Ma, J., Liu, W., Cao, X., Tian, M., Wang, Y. (2026). Bird droppings flashover faults analysis of 330 kV transmission line V-Shaped insulator strings. Electric Power Systems Research, 253: 112510. https://doi.org/10.1016/J.EPSR.2025.112510

[4] Zhang, S., Jiang, W., Guo, Z., Yu, F., Liu, J., Tian, L. (2025). Study on the impact of additional insulator string fracture on the safety status of transmission lines under large ice loads. Buildings, 15(22): 4131. https://doi.org/10.3390/BUILDINGS15224131

[5] Li, X., Yang, N., Yin, Y., Wang, Y., Ren, H., Wang, J., Li, Q. (2025). Simulation methodology for the dynamic adsorption effect of micro/nano-dust in gas-insulated switchgears and transmission lines and evaluation of multimode diffusion-induced explosions. Journal of Physics D: Applied Physics, 58(44): 445501. https://doi.org/10.1088/1361-6463/AE1043

[6] Yang, P., Wang, H., Huang, X., Gu, J., Deng, T., Yuan, Z. (2025). A novel telescopic aerial manipulator for installing and grasping the insulator inspection robot on power lines: Design, control, and experiment. Drones, 9(11): 741. https://doi.org/10.3390/DRONES9110741

[7] Yu, X., Zhou, Z., Deng, Y., Zhang, K., Gu, C., Liu, Z., Zhou, S. (2025). An angle-enhanced deep learning framework for thermal defect diagnosis in overhead transmission line composite insulators. Engineering Applications of Artificial Intelligence, 162: 112285. https://doi.org/10.1016/J.ENGAPPAI.2025.112285

[8] Nejadbougar, R.R., Parmehr, E.G., Afary, A., Mavaddati, S. (2025). A new deep learning framework for intelligent aerial monitoring of power transmission line insulators. Engineering Applications of Artificial Intelligence, 161: 112290. https://doi.org/10.1016/J.ENGAPPAI.2025.112290

[9] Zhao, L.H., An, Y., Li, Y., Fu, L., Liu, S., Wang, L. (2025). MDM-TLI: Multi-defect detection model for transmission line insulators. IET Image Processing, 19(1): e70180. https://doi.org/10.1049/IPR2.70180

[10] Wang, Y., Liu, S., Li, J., Liang, H., Xiao, M., Chen, Y., Du, B. (2025). Curing kinetics and residual stress modelling of gas-insulated transmission lines tri-post insulators. High Voltage, 10(4): 976-986. https://doi.org/10.1049/HVE2.70070

[11] Zhang, H., Liu, Z., Lu, Y., Li, X., She, Y., Zhou, Y., Liang, J. (2025). Design and analysis of a simultaneous wireless power and data transfer system across insulators for online monitoring devices on transmission lines. Journal of Computational Methods in Sciences and Engineering, 25(5): 4296-4307. https://doi.org/10.1177/14727978251338981

[12] Che, J., Zhu, L. (2025). Improved YOLOv8-based insulator defect detection system for transmission lines. Journal of Physics: Conference Series, 3059(1): 012012. https://doi.org/10.1088/1742-6596/3059/1/012012

[13] Ronneberger, O., Fischer, P., Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference, Munich, Germany, pp. 234-241. https://doi.org/10.1007/978-3-319-24574-4_28

[14] Pang, G., Zhang, Z., Hu, J., Hu, Q., Zheng, H., Jiang, X. (2025). Analysis of failures and protective measures for core rods in composite Long-Rod insulators of transmission lines. Energies, 18(12): 3138. https://doi.org/10.3390/EN18123138

[15] Liu, W., Liu, K., Wang, Y., Ma, Z. (2025). Study on the protection range of bird droppings on double I type pendant insulators for 110kv transmission lines. Journal of Physics: Conference Series, 3033(1): 012039. https://doi.org/10.1088/1742-6596/3033/1/012039

[16] Yin, L., Wang, J., Yao, H., Li, W., Wu, K. (2025). Research on ultrasonic testing of internal defects in silicone rubber insulators of transmission lines. Journal of Physics: Conference Series, 3043(1): 012080. https://doi.org/10.1088/1742-6596/3043/1/012080

[17] Farooq, U., Yang, F., Shahzadi, M., Ali, U., Li, Z. (2025). YOLOv8-IDX: Optimized deep learning model for transmission line insulator-defect detection. Electronics, 14(9): 1828. https://doi.org/10.3390/ELECTRONICS14091828

[18] Alipour Bonab, S., Sadeghi, A., Yazdani-Asrami, M. (2025). Artificial intelligence-based surrogate model for computation of the electric field of high voltage transmission line ceramic insulator with corona ring. World Journal of Engineering, 22(3): 458-471. https://doi.org/10.1108/WJE-11-2023-0478

[19] Zhao, L., Kang, J., An, Y., Li, Y., Jia, M., Li, R. (2025). SACG-YOLO: A method of transmission line insulator defect detection by fusing scene-aware information and detailed-content-guided information. Electronics, 14(8): 1673. https://doi.org/10.3390/ELECTRONICS14081673

[20] Ye, Y., Tan, G., Liu, Q., Liu, L., Chu, J., Wen, B., Li, L. (2025). TSSSKD-YOLO: An intelligent classification and defect detection method of insulators on transmission lines by fusing knowledge distillation in multiple scenarios. Multimedia Systems, 31(3): 183. https://doi.org/10.1007/S00530-025-01772-Y

[21] Chen, Z., Mao, X., Liu, W., Zhao, H., Bo, B., Yu, J.J. (2025). The effect of buoyancy convection and geometric variation on temperature field of C4F7N-filled gas-insulated metal transmission line. Microgravity Science and Technology, 37(2): 20. https://doi.org/10.1007/S12217-025-10173-9

[22] Xu, Z.Y., Tang, X. (2025). Transmission line insulator defect detection algorithm based on MAP-YOLOv8. Scientific Reports, 15(1): 10288. https://doi.org/10.1038/S41598-025-92445-3

[23] Wang, R., Xu, Y., Jiang, C., Liang, L., Liu, W., Hu, C., Tao, J. (2025). Development and insulation performance evaluation of experimental sample for a direct current high-voltage transmission line. Fusion Science and

Technology, 81(3): 259-268. https://doi.org/10.1080/15361055.2024.2383089

[24] Mahapatra, U., Rahman, M.A., Islam, M.R., Hossain, M.A., Sheikh, M.R.I., Hossain, M.J. (2025). Adversarial training-based robust model for transmission line's insulator defect classification against cyber-attacks. Electric Power Systems Research, 245: 111585. https://doi.org/10.1016/J.EPSR.2025.111585

[25] Zheng, L., Yin, P., Li, J., Liu, H., Li, T., Luo, H. (2025). Research on Diagnostic Methods for Zero-Value Insulators in 110 kV Transmission Lines Based on Spatial Distribution Characteristics of Electric Fields. Energies, 18(6): 1534. https://doi.org/10.3390/en18061534