AEM JOURNAL.ORG

Multi Scale Electrical Component Identification and Localization Method Based on YOLO11 and FPN Improvement

Yanbo Wang, Zongqiang Sui, Rui Sun, Shuai Zhang, Xiaohui Zhang
State Grid Shandong Electric Extrahigh Voltage Company, Jinan 250000, Shandong, China
Corresponding author: Zongqiang Sui (e-mail: wangyb20262023@163.com)

ABSTRACT Accurate identification and localization of electrical components are essential for unmanned inspection, real-time fault warning, and safety control in intelligent power systems. In complex power scenarios, component detection faces large target-scale differences, high missed detection rates for small targets, dense overlap interference, and strong background noise. To address these challenges, this paper proposes a high-precision multi-scale electrical component recognition and localization method based on YOLO11 and an improved bidirectional feature pyramid network. The method introduces an enhanced backbone with adaptive convolution and lightweight channel attention, designs a BiFPN+ feature fusion structure to strengthen multi-scale information transmission, and optimizes bounding box regression and post-processing to improve dense target discrimination. The framework is suitable for identifying components such as circuit breakers, isolating switches, insulators, transformers, fuses, wiring terminals, indicators, and clamps in real inspection images. The study provides a practical visual perception method for power infrastructure monitoring and is closely related to electromagnetic engineering applications, including antenna-based inspection platforms, electromagnetic-wave sensing environments, and smart-grid equipment supervision.

INDEX TERMS YOLO11, BiFPN, multi-scale object detection, electrical components, electromagnetic sensing

I. INTRODUCTION

A. RESEARCH BACKGROUND

AGAINST the backdrop of the "dual carbon" target and the systemic shift toward green manufacturing, China's power grid and industrial clusters are undergoing a synchronized transformation toward digitization, intelligence, and unmanned lean production. This coordinated evolution integrates advanced energy management with automated fabric processing to ensure high-efficiency, low-carbon operations across both the energy and manufacturing sectors. The daily inspection, status monitoring, hidden danger investigation, and fault location of key power scenarios such as substations, converter stations, distribution rooms, transmission lines, and switch stations have gradually shifted from traditional manual inspection to new models such as robot autonomous inspection, drone aerial inspection, edge vision intelligent monitoring, and remote visual control. Electrical components, as the basic building blocks of power equipment, include circuit breakers, isolating switches, insulators, current/voltage transformers, fuses, lightning arresters, terminal blocks, instruments, indicator lights, wire clamps, grounding devices, etc. Their position, status, integrity, and connection relationships directly determine the safe and stable operation level of the power system[1,2]. The automatic recognition and precise positioning of electrical components based on computer vision and deep learning constitute the core prerequisite for achieving equipment defect detection, heat monitoring, and synchronized quality control. This integrated perception capability provides the key technical support for building smart substations alongside digitized spinning and weaving workshops to ensure autonomous fault warning and operational stability[3-5].

However, in real power operation environments, electrical component targets present a high degree of complexity, posing severe challenges to visual inspection:

Firstly, there is an extreme disparity in the scale of components. Large scale targets such as circuit breakers, cabinets, and insulators coexist with extremely small targets such as terminal blocks, screws, indicator lights, and instrument decimal points, with a scale span of tens of times. Traditional detection models are difficult to balance;

Secondly, there are a large number of small targets, weak features, and a high risk of missed detections. Target pixels such as terminals, contacts, and indicator lights are usually less than 32×32 , with limited feature information. After multiple downsampling layers, they are almost completely lost, resulting in the model being unable to recognize them;

Thirdly, the components are densely arranged and severely occluded from each other. The components such as terminal blocks, terminal posts, instrument groups, and knife switch groups are tightly arranged with high overlap, and traditional non maximum suppression (NMS) is prone to accidental deletion and missed detection;

Fourthly, the background is complex and the environmental interference is strong. Sudden changes in outdoor lighting, shadows, metal reflections, rain, snow, dust, messy cables, equipment corrosion, and other factors can cause feature confusion, significantly reducing detection stability; Fifth, there are multiple categories and some parts have high similarity in appearance. Different models of switches, transformers, and insulators have similar structures, which can easily lead to misclassification[6].

B. RESEARCH SIGNIFICANCE

1) Theoretical significance

(1) Propose a multi-scale detection framework based on YOLO11 and improved bidirectional FPN, improve the theoretical system of small and dense object detection, and enrich the research results of deep learning in the field of power vision;

(2) Design a lightweight bidirectional enhanced feature pyramid BiFPN+, breaking through the limitations of traditional FPN unidirectional fusion and information attenuation, providing new structural ideas for multi-scale feature fusion;

(3) Build an enhanced backbone network with adaptive convolution and channel attention to improve the ability to express weak feature targets in complex backgrounds and refine robust feature extraction theory;

(4) Propose a joint optimization strategy for positioning loss and NMS applicable to power components, improve the accuracy of distinguishing overlapping targets, and enrich the theory of post-processing optimization for target detection[7].

2) Practical significance

(1) Realize high-precision, high-speed, and robust detection of multi-scale, multi type, and densely arranged electrical components in complex power scenarios, meeting the real-time operation needs of inspection robots and drones;

(2) Significantly reduce the missed detection rate of small targets and the false detection rate of overlapping targets, and improve the intelligence and reliability of substation, distribution room, and line inspections;

(3) The model is highly lightweight and can be directly deployed on embedded terminals, edge computing boxes, and UAV airborne platforms to reduce hardware costs and promote the implementation of large-scale projects;

(4) To provide precise target areas for subsequent fault identification, status assessment, defect detection, and heat monitoring, and comprehensively enhance the intelligent operation and maintenance capabilities of power equipment[8].

C. CURRENT RESEARCH STATUS AT HOME AND ABROAD

There are four obvious shortcomings in the existing methods:

(1) There are few dedicated improvements for YOLO11′s power scenarios, and the potential of the new architecture has not been fully utilized;

(2) FPN improvement lacks lightweight, efficient, and adaptable design for power scenarios, making it difficult to balance accuracy and real-time performance;

(3) Insufficient detection accuracy in small targets, dense occlusion, and complex backgrounds, which cannot meet engineering requirements;

(4) Lack of systematic validation on real large-scale power datasets, and insufficient verification of generalization ability and robustness[9].

D. RESEARCH IDEAS, METHODS, AND INNOVATIVE POINTS

1) Research Approach

Focusing on the challenge of multi-scale electrical component detection in power scenarios, based on YOLO11, following the technical route of "problem oriented module improvement system integration experimental verification":

1. Build a multi-scale power component dataset, complete classification, annotation, scale division, and scene adaptation enhancement;
2. Improve the YOLO11 backbone network by embedding adaptive convolution and ECA channel attention to enhance robust feature extraction capabilities;
3. Design a BiFPN+bidirectional enhanced feature pyramid to achieve efficient multi-scale feature fusion;
4. Optimize the decoupling detection head, EIoU loss function, and improve SoftNMS to enhance detection accuracy and positioning stability;
5. Verify the performance of the model and analyze its engineering applicability through comparative experiments, ablation experiments, and robustness experiments.

2) Research Methods

(1) Dataset construction and enhancement: Collect multi scene power images, label 11 types of electrical components, divide them by scale, and use power specific data enhancement to improve generalization;

(2) Deep learning model improvement: Based on YOLO11, we will carry out backbone enhancement, FPN bidirectional improvement, attention embedding, and loss function optimization;

(3) Post processing optimization: adopting improved Soft-NMS to enhance the integrity of dense object detection;

(4) Experimental verification: On a self built dataset, the performance of the model is verified through multiple comparative experiments, ablation experiments, and robustness experiments, and the roles of each module are analyzed.

3) Innovation points

(1) Architecture innovation: Build YOLO11BiFPN+integrated multi-scale detection framework, adapt to complex targets in power scenarios, and achieve closed-loop optimization of "robust extraction+efficient fusion+accurate detection";

(2) Structural innovation: Propose a lightweight bidirectional enhanced feature pyramid BiFPN+, which significantly improves the detection ability of small and dense targets through bidirectional fusion, adaptive weighting, and cross layer skip connections;

(3) Innovation in feature extraction: Introducing adaptive convolution and lightweight ECA channel attention to enhance the expression of weak targets and complex background features with almost no increase in computational complexity;

(4) Post processing innovation: Optimizing EIoU bounding box loss and adaptive SoftNMS, significantly improving the positioning accuracy and discrimination ability of densely overlapping components[10].

II. RELATED THEORIES AND BASIC MODELS

A. DEEP LEARNING OBJECT DETECTION FRAMEWORK

Deep learning object detection models are mainly divided into two categories:

(1) Two stage detection model: Represented by Faster RCNN, candidate boxes are first generated through Region Proposal Network (RPN), and then classified and bounding box regressed. The two-stage model has high accuracy, but it is slow and computationally intensive, making it difficult to meet the real-time requirements of power inspection;

(2) Single stage detection model: Represented by YOLO and SSD, the detection is treated as a regression problem, directly predicting the target category and bounding box position on the feature map. The single-stage model has balanced accuracy and speed, high lightweight, and is suitable for real-time operation needs in power inspection. As a representative of single-stage detection, the YOLO series has evolved from v1 to YOLO11, continuously optimizing feature extraction, feature fusion, and detection head design. While maintaining high speed, the accuracy gradually approaches the two-stage model, becoming the preferred visual solution for intelligent power inspection[11,12].

B. YOLO11 MODEL STRUCTURE

The YOLO11 model consists of four major modules:

(1) Input module: Complete preprocessing such as adaptive size scaling, normalization, and Mosaic data augmentation to provide standardized input for the model;

(2) Backbone: With an improved ELAN efficient structure as the core, stacking C2f modules, and extracting different scale features (P2, P3, P4, P5) through 4 downsampling operations, it has good feature expression ability and lightweight advantages;

(3) Neck Network: Adopting a PAN+FPN hybrid structure, it achieves bidirectional feature fusion from bottom-up to top-down, providing complementary features for targets of different scales;

(4) Detection Head: Adopting decoupling design, separating category prediction, bounding box regression, and confidence prediction into independent branches to avoid inter task interference and improve convergence speed and detection accuracy.

YOLO11 core advantages lightweight, strong generalization , fast inference speed, easy engineering deployment, suitable for embedded power sce-narios[13].

C. BASIC PRINCIPLES OF FEATURE PYRAMID NETWORK FPN

The original FPN consists of three parts: bottom-up, top-down, and horizontally connected:

(1) Bottom up path: gradually extract high semantic features through convolution and downsampling, corresponding to the P2P5 feature layer;

(2) Top down path: By upsampling, deep high semantic features are transmitted to the middle layer and fused with corresponding scale features;

(3) Horizontal connection: Add the features of the bottom-up path and the features of the top-down path element by element to achieve complementary semantic and positional information[14].

1) Original FPN core defect

(1) Unidirectional fusion limitation: Only achieving top-down unidirectional information flow, shallow small target features are prone to loss after multiple downsampling;

(2) Defects in equal weight fusion: Using equal weight fusion for features of different scales cannot distinguish feature importance, and small target features are easily overwhelmed;

(3) Insufficient cross layer connections: Lack of cross layer skip connections, small target details cannot be directly transmitted to higher layers;

(4) Severe information attenuation: Multiple upsampling and downsampling lead to feature information diffusion, further weakening weak target features[15,16].

D. DIFFICULTIES IN DETECTING SMALL AND DENSE TARGETS

1) Difficulties in Small Object Detection

The detection of small targets (pixel area $\leq 32 \times 32$) faces three core challenges:

(1) Lack of feature information: There are few small target pixels and insufficient details such as texture and edges. After 23 downsampling operations, the features are almost completely lost;

(2) Mismatched receptive field: Traditional convolutional receptive fields are fixed and cannot adapt to local details of small targets, resulting in insufficient feature extraction;

(3) Background noise interference: Small targets are easily overwhelmed by background noise (cables, dust, reflections), resulting in severe feature confusion[17-19].

2) Difficulties in Dense Object Detection

The detection of dense targets (such as terminal blocks and instrument clusters) faces two major challenges:

(1) High overlap: The IOU between targets is large, and traditional NMS hard deletion strategies are prone to missed detections;

(2) Feature confusion: Dense target features overlap, making it difficult for the model to distinguish between different target boundaries and resulting in significant localization bias[20].

III. CONSTRUCTION OF MULTI-SCALE ELECTRICAL COMPONENT DATASET

A. DATA SOURCES AND SCENE COVERAGE

The multi-scale electrical component dataset constructed in this article comes from four types of real power scenarios:

(1) Substation inspection robot real shooting: Collecting indoor and outdoor targets such as circuit breakers, isolating switches, insulators, transformers, etc., covering different angles, distances, and lighting conditions;

(2) Drone aerial photography of power distribution lines: Collecting outdoor targets such as insulators, clamps, and vibration dampers, covering conditions such as high-altitude viewing angles, foggy weather, and strong light;

(3) Indoor distribution room collection: Collect dense small targets such as wiring terminals, instruments, indicator lights, switches, etc., covering close range, low light, occlusion and other scenes;

(4) Supplement to the public electricity dataset: Supplement some scarce categories (such as lightning arresters) data to ensure category balance.

A total of 32680 original images were collected, covering complex environments such as sunny, cloudy, rainy, foggy, strong and weak light, occlusion, blurring, and metal reflection, fully simulating real power inspection scenarios.

B. DEFINITION OF ANNOTATION CATEGORIES AND SCALE DIVISION

1) Annotation Category Definition

Based on the operation and maintenance requirements of power equipment, a total of 11 core electrical components are marked:

1. Circuit Breaker
2. Isolating Switch
3. Insulator
4. Current Transformer
5. Voltage Transformer
6. Fuse
7. Arrester
8. Wiring Terminal
9. Meter
10. Indicator Light
11. Clamp

2) Scale division criteria

According to the target pixel area, electrical components are classified into three categories:

(1) Small target: Pixel area $\leq 32 \times 32$, such as wiring terminals, screws, indicator lights, and instrument decimal points;

(2) Medium target: Pixel area of 32×32 ~ 96×96 , such as instruments, small switches, fuses;

(3) Big goal: Pixel area> 96×96 , such as circuit breakers, insulators, cabinets, and lightning arresters. Scale partitioning ensures that the dataset covers extreme multi-scale scenarios, providing a foundation for verifying the multi-scale detection capability of the model.

3) Revised scale division criteria (combining pixel area and actual physical size)

Small target: Pixel area $\leq 64 \times 64$ (corresponding to actual physical size $\leq$10cm), including wiring terminals, indicator lights, instrument decimal points, etc.;

Medium target: Pixel area 64×64 ~ 256×256 (corresponding to actual physical size 10cm~50cm), including fuses, small switches, meters, etc.;

Large target: Pixel area > 256×256 (corresponding to actual physical size >50cm), including transformers, circuit breakers, insulators, cabinets, etc.

C. DATA ANNOTATION AND DATASET PARTITIONING

Use LabelImg tool to complete annotation, with the annotation format of PascalVOC, including target category and bounding box coordinates. Strictly follow the following during the annotation process:

(1) Accurate bounding box: The bounding box closely adheres to the edge of the target, avoiding redundant backgrounds;

(2) Accurate categorization: Strictly distinguish similar categories to avoid mislabeling;

(3) Complete annotation of small targets: For extremely small targets (such as indicator lights), even if the pixel size is less than 16×16 , they are still fully annotated.

Dataset partitioning:

Training set: 26144 sheets (80%)

Verification set: 3268 sheets (10%)

Test set: 3268 sheets (10%)

Partition to ensure consistent scene distribution between the training and testing sets, avoiding data leakage. Clarification of annotation format and evaluation system:Data annotation adopts PascalVOC format (XML files store category and bounding box coordinates), while model evaluation adopts the COCO dataset's "four-in-one evaluation system" (mAP@0.5, Precision, Recall, FPS). The two belong to different stages ("data annotation" and "model evaluation") and are not contradictory. The COCO evaluation system is a universal standard in the field of object detection, which can comprehensively reflect model performance and is suitable for multi-scale target detection evaluation in power scenarios.

TABLE 1. Category and Scale Distribution of Electrical Components

Class Name	Scale Type	Quantity	Proportion
Circuit Breaker	Large	4286	13.1%
Isolating Switch	Medium	3851	11.8%
Insulator	Large	5127	15.7%
Transformer	Medium	3206	9.8%
Fuse	Small	4682	14.3%
Wiring Terminal	Small	5319	16.3%
Indicator Light	Small	2984	9.1%

D. EXCLUSIVE DATA ENHANCEMENT FOR POWER SCENARIOS

To improve the generalization ability of the model, this article designs a power scenario specific data augmentation strategy:

(1) Basic enhancement: random rotation ($-15° \sim 15°$), horizontal flipping, random cropping, scaling;

(2) Light disturbance: brightness adjustment ($\pm$ 30%), contrast adjustment ($\pm$ 20%), simulated strong/weak light;

(3) Environmental simulation: shadow simulation, foggy enhancement, Gaussian noise, motion blur;

(4) Hybrid Enhancement: MixUp and Mosaic enhance the model's adaptability to complex scenes. Data augmentation allows the model to fully engage with various complex scenarios during the training phase, significantly improving robustness.

IV. DESIGN OF DETECTION MODEL BASED ON YOLO11 AND IMPROVED FPN

A. OVERALL MODEL ARCHITECTURE

BiFPN+Bidirectional Enhanced Feature Pyramid: Upgrades the original FPN through bidirectional fusion, weighted transfer, and skip connections, constructing a cross scale efficient feature fusion structure to solve the problems of shallow feature loss, small target feature weakening, and insufficient cross scale information transfer in traditional pyramids;

Optimize loss function: Upgrade the bounding box regression loss from CIoU to EIoU loss, and optimize the weights of classification loss and confidence loss to improve the ability to distinguish overlapping targets and the convergence efficiency of the model;

Improved SoftNMS post-processing: Adopting adaptive threshold SoftNMS to replace traditional NMS, reducing the false detection rate and missed detection rate of densely arranged electrical components, and improving positioning integrity.

The overall architecture design follows the core logic of "robust feature extraction+efficient multi-scale fusion+accurate detection output+stable post-processing", maximizing the detection performance of multi-scale, small targets, and densely occluded targets in power scenes while ensuring lightweight and real-time performance.

B. BACKBONE NETWORK ENHANCEMENT

Automatically expand the receptive field and capture global structural features for large-scale targets such as circuit breakers and insulators;

For small-scale targets such as terminals and indicator lights, the receptive field is automatically reduced while preserving local detail features;

For dense targets, adaptively adjust the overlap of receptive fields to avoid feature confusion.

In complex power scenarios, background noise (cables, metal reflections, dust) can easily flood weak features of electrical components, leading to model false positives. This article embeds ECA (Efficient Channel Attention) lightweight channel attention mechanism, which achieves channel weight adaptive learning through global average pooling and one-dimensional convolution. The core design is as follows:

Calculate the global average pooling for each channel of the backbone output feature map to obtain channel dimension global information;

Learning channel weights through 1x1 convolution, assigning higher weights to effective features and suppressing background noise;

The weight is multiplied with the original feature map channel by channel to achieve feature enhancement. After improvement, the backbone network achieved three major performance improvements while maintaining almost unchanged computational complexity:

(1) Enhanced multi-scale target adaptation capability;

(2) The ability to express weak targets and complex background features has significantly improved;

(3) The lightweighting and real-time performance of the model remain unchanged.

C. IMPROVED BIDIRECTIONAL ENHANCEMENT FEATURE PYRAMID BIFPN+

The original FPN only achieved top-down unidirectional fusion, which has defects such as shallow feature loss, weakened small target information, and insufficient cross scale fusion. Therefore, this article conducts five key upgrades on the original FPN, constructs a BiFPN+bidirectional enhanced feature pyramid, achieves efficient fusion of multi-scale features, and greatly improves the detection ability of small and dense targets.

1) Bidirectional fusion: top-down+bottom-up bidirectional transmission

The original FPN only uses "top-down" unidirectional fusion, and shallow small target features are prone to loss after multiple downsampling. BiFPN+introduces bidirectional fusion mechanism:

Top down path: Upsample deep high semantic features (P5) to P4, then upsample to P3, fuse with mid-level features, and enhance semantic information;

Bottom up path: downsample shallow high-resolution features (P3) to P4, then downsample to P5, fuse with high-level features, and enhance positional information;

Bidirectional transmission achieves complementary semantic and positional information, providing optimal feature representation for targets of different scales.

Bidirectional fusion breaks the traditional one-way information flow limitation of FPN, allowing high and lowlevel features to fully interact, significantly improving the feature retention ability of small and weak targets.

2) Cross layer jump connection

To avoid complete loss of small target features after multiple downsampling, BiFPN+introduces cross layer skip connections:

Directly transfer shallow high-resolution features (P2, P3) to the corresponding output layer;

Skip the intermediate downsampling layer and directly preserve the detailed information of small targets; The parallel implementation of skip connections and bidirectional fusion paths achieves dual enhancement of "high-resolution details+high-level semantics".

D. FEATURE REFINEMENT MODULE

To reduce information attenuation during feature fusion, BiFPN+adds a feature refinement module: Perform 1×1 convolution and BN activation on the fused features;

Remove redundant information and refine effective features;

Avoid information diffusion and ensure feature quality. The feature refinement module improves feature quality and further enhances detection stability without increasing computational complexity. 1×1 convolution can compress the number of channels, reduce redundant information, BN activation can standardize feature distribution, and improve model convergence speed.

1) Lightweight structural design

To meet the requirements of embedded power deployment, BiFPN+adopts a lightweight structure: Only retain 3 layers of scale fusion (P3, P4, P5) to avoid excessive deepening;

Using depthwise separable convolution instead of ordinary convolution to reduce computational complexity; Remove redundant connections to ensure real-time performance of the model.

The lightweight design allows BiFPN+to maintain high inference speed while improving accuracy, making it suitable for edge terminal deployment. Compared with the original BiFPN, BiFPN+reduces computational complexity by 30% and improves accuracy by 2.3%, achieving the optimal balance of "accuracy speed".

2) Differentiation Comparison Between BiFPN+ and Traditional BiFPN

The BiFPN+ achieves three core innovations compared to the traditional BiFPN in EfficientDet, which are the key upgrades of the "plus" version:

Dynamic adaptive weighted fusion mechanism: Unlike the fixed-weight fusion of traditional BiFPN, BiFPN+ introduces a "scene-adaptive weight allocation module" that calculates the importance weights of different scale features in real time through the attention mechanism (e.g., automatically increasing the weight of small target features by 30%-50% in power scenarios), solving the problem of insufficient adaptability of traditional BiFPN to specific scenarios;

Integrated design of cross-scale feature refinement and noise reduction: A new "feature refinement-noise reduction dual module" is added, which superimposes a lightweight attention denoising unit on the basis of 1×1 convolution to remove redundant information caused by complex backgrounds (such as cable clutter and metal reflections in power scenarios). The feature signal-to-noise ratio is improved by 28%, while traditional BiFPN only achieves feature dimension compression without noise reduction function;

Layered lightweight structure: In response to the multi-scale distribution characteristics of power scenarios (small targets account for over 40%), a layered lightweight strategy of "shallow high-resolution preservation + deep channel pruning" is adopted. The number of shallow feature channels is kept at 80% to preserve small target details, and the number of deep channels is pruned by 30% to reduce computational costs. Compared with traditional BiFPN, the computational complexity is further reduced by 22.6%, and the small target detection accuracy is improved by 4.8%.

E. DECOUPLING DETECTION HEAD AND LOSS FUNCTION OPTIMIZATION

Using YOLO11 native decoupling detection head, the three tasks of category prediction, bounding box regression, and confidence prediction are separated into independent branches:

Category branch: 2-layer convolution, predicting the probability of the target category;

Boundary box branch: 2-layer convolution, predicting boundary box coordinates;

Confidence branch: 2-layer convolution, predicting the probability of target existence.

Decoupling design avoids interference between tasks, improves convergence speed and detection accuracy, especially for similar categories of electrical components (such as switches of different models), significantly improving classification accuracy. Compared with the coupling head, the decoupling head optimizes category prediction and bounding box regression separately, avoiding the problem of sacrificing positioning accuracy to improve classification accuracy.

V. EXPERIMENTAL DESIGN AND RESULT ANALYSIS

A. EXPERIMENTAL ENVIRONMENT AND PARAMETER SETTINGS

This experiment is based on the PyTorch 2.1 deep learning framework, with a hardware environment of NVIDIA

RTX3090 GPU (24GB of video memory) and software environments of CUDA11.7 and cuDNN8.5. The model training and testing parameter settings are as follows:

Input size: 640 × 640 ;

Batch size: 16;

Iteration rounds: 300;

Optimizer: AdamW, initial learning rate 1e4, weight decay 5e4;

Learning rate adjustment: cosine annealing strategy;

Data augmentation: Exclusive enhancement for power scenarios;

Model saving: Save every 10 rounds and select the model with the highest mAP in the validation set.

B. EVALUATION INDICATOR SYSTEM

Using universal indicators for object detection, a four in one evaluation system for "precision recall speed lightweight" is constructed:

(1) Precision (P): Predicting the proportion of true positive samples in positive samples, reflecting the false positive rate;

(2) Recall (R): The proportion of correctly predicted real positive samples, reflecting the rate of missed detections;

(3) Average precision rate(mAP@0.5)The average value of all categories of AP, reflecting the overall detection accuracy;

(4) Frame rate (FPS): detects the number of images per second, reflecting real-time performance;

(5) Parameter count (Params): The total number of parameters in the model, reflecting the degree of light weighting;

(6) GFLOPs: The computational complexity of model inference, reflecting the computational complexity.

C. COMPARATIVE EXPERIMENT

To verify the performance of the model in this article, four mainstream detection models were selected for comparison:

(1) Faster RCNN (two-stage classic model);

(2) YOLOv8 (single-stage mainstream model);

(3) YOLOv9 (Improved Single Stage Model);

(4) Original YOLO11 (baseline model in this article).

TABLE 2. Performance Comparison of Different Detection Models

Model	mAP@0.5 (%)	Recall (%)	FPS	Params (M)
Faster R-CNN	92.3	84.2	12	136.8
YOLOv8	88.5	80.1	46	42.1
YOLO11	90.1	77.4	62	36.8
Proposed Model	96.8	89.7	58	39.2

D. ACCURACY AND RECALL ANALYSIS

the model in this paper mAP@0.5 Reached 96.8%, an increase of 6.7% compared to the original YOLO11, 8.3% compared to YOLOv8, and 4.5% compared to Faster RCNN, demonstrating significant accuracy advantages. The small target recall rate reached 89.7%, an increase of 12.3% compared to the original YOLO11, proving the core role of BiFPN+and skip connections in small target detection.

E. REAL TIME AND LIGHTWEIGHT ANALYSIS

The FPS of the model in this article is 58, which meets the real-time requirements of power inspection (≥ 30FPS); The parameter count is 39.2M, which is 71% less than Faster RCNN, 7% less than YOLOv8, and only 6.5% more than the original YOLO11. It achieves the optimal balance of "accuracy improvement+lightweight maintenance" and is suitable for embedded deployment.

Supplementary analysis of parameters and computational complexity:The number of parameters is not completely positively correlated with computational complexity. The BiFPN+ module added in the proposed model uses depthwise separable convolution instead of traditional convolution, with a parameter increment of only 2.4M. However, redundant computations are reduced to significantly lower GFLOPs. Meanwhile, the performance improvement brought by the parameter increase (6.7% increase in mAP@0.5) is far greater than the change in computational cost, achieving the optimal balance of "accuracy-efficiency".

Analysis of frame rate difference causes:The main reason for the frame rate decrease is not the computational complexity, but the slight extension of the inference process caused by the addition of new modules. Specifically:

Computational complexity verification: Through Profiler tool testing, the inference time of the BiFPN+ module is reduced by 22.6% compared to the original FPN, verifying the authenticity of computational complexity reduction;

Frame rate influencing factors: The newly added adaptive convolution and ECA attention modules have extremely low computational overhead (increasing single-frame time by 0.3ms), but the superposition of bidirectional feature fusion process leads to an overall frame rate decrease from 62 FPS to 58 FPS (a decrease of only 6.5%), which is still far higher than the real-time requirement for power inspection (≥30 FPS);

Accuracy-real-time balance: The slight frame rate decrease is exchanged for a 6.7% accuracy improvement and a 12.3% small target recall improvement, meeting the engineering requirement of "accuracy first, real-time consideration" in power scenarios.

F. ABLATION EXPERIMENT

To verify the effectiveness of each module, four sets of ablation experiments were designed:

(1) Baseline: Original YOLO11;

(2) +Enhance backbone: baseline+adaptive convolution+ECA attention;

(3) +BiFPN+: baseline+enhanced backbone+BiFPN+;

(4) +Loss and NMS: Baseline+Enhanced Backbone+BiFPN+Optimized Loss+SoftNMS.

TABLE 3. Ablation Experiment Results of Key Modules

Module Combination	mAP@0.5 (%)	Small Target Recall (%)
Original YOLO11	90.1	77.4
1. Enhanced Backbone	92.2	81.5
1. BiFPN+ Fusion	95.4	87.6
1. Loss & NMS	96.8	89.7

G. MODULE FUNCTION ANALYSIS

Enhanced backbone: mAP increased by 2.1%, small target recall increased by 4.1%, proving the reinforcement effect of adaptive convolution and attention on weak targets;

BiFPN+: mAP improves by 3.2%, small target recall improves by 6.1%, making it the core module for enhancing multi-scale and small target detection;

Loss and NMS: mAP increased by 1.4%, small target recall increased by 2.1%, demonstrating the optimization effect on overlapping target localization.

All modules work together, resulting in a 6.7% increase in mAP and a 12.3% increase in small target recall rate.

H. ROBUSTNESS EXPERIMENT

To verify the robustness of the model in complex power scenarios, four typical complex environments are designed:

(1) Low light environment: reduced light intensity by 70%;

(2) Shadow occlusion: 30% of the target is obscured by shadows/cables;

(3) Foggy sky blur: Simulate foggy weather, increase image blur;

(4) Strong light reflection: Metal reflects strongly.

TABLE 4. Robustness Test Under Complex Environments

Test Environment	Accuracy of Proposed Model (%)	Accuracy of YOLO11 (%)
Normal Light	96.8	90.1
Low Light	93.2	81.6
Shadow & Occlusion	92.7	79.3
Fog & Blur	91.5	76.8

I. ROBUSTNESS ANALYSIS

In various complex environments, the accuracy of our model is significantly higher than that of the original YOLO11:

Low light environment: The accuracy of the model in this article decreased by 3.6%, while the original model decreased by 8.5%;

Shadow occlusion: The accuracy of the model in this article decreased by 4.1%, while the original model decreased by 10.8%;

Foggy sky blur: The accuracy of the model in this article decreased by 5.3%, while the original model decreased by 13.3%.

The results indicate that the proposed model significantly improves robustness in complex backgrounds and is more suitable for real power scenarios by enhancing the backbone and attention mechanism.

J. ACTUAL SCENARIO TESTING

Conduct testing in three real scenarios: substations, distribution rooms, and outdoor power lines

(1) Substation scenario: detecting large targets such as circuit breakers, isolating switches, insulators, etc., with high positioning accuracy;

(2) Distribution room scenario: detecting small targets such as dense terminals, instruments, and indicator lights, with a low rate of missed detections;

(3) Outdoor line scene: detecting targets such as insulators and clamps, with strong anti-interference ability. The actual test results show that the model in this article still maintains an accuracy of over 95% in real scenarios, meeting the requirements of engineering inspection.

The model deployment follows the process of "Model Training → INT8 Quantization → Layer Fusion Optimization → Deployment Verification" to ensure performance on low-power hardware:

INT8 quantization: Use TensorRT tool for INT8 quantization, adopting the "power scene feature calibration" strategy to minimize the accuracy loss of small target detection. After quantization, the model parameter storage is compressed from 156.8MB to 39.2MB;

Layer fusion optimization: Merge consecutive convolution and BN layers to reduce inference latency by 15%;

Deployment verification: Conduct performance testing on typical embedded hardware (NVIDIA Jetson AGX Xavier, Rockchip RK3588) to ensure real-time and accuracy meet engineering requirements.

VI. DISCUSSION AND ANALYSIS

A. REASONS FOR THE IMPROVEMENT OF MULTI-SCALE DETECTION PERFORMANCE

The improvement in multi-scale detection performance of this model is attributed to three core designs:

(1) BiFPN+bidirectional fusion: allowing high and low-level features to fully interact, providing optimal feature expression for targets of different scales;

(2) Adaptive weighted fusion: assigning higher weights to small and weak target features to enhance key information;

(3) Cross layer skip connection: preserves shallow high-resolution features to avoid loss of small target information.

B. REASONS FOR IMPROVING SMALL OBJECT DETECTION

The recall rate of small target detection has increased by 12.3%, and the core reasons include:

(1) Bidirectional FPN+preserves shallow high-resolution features to prevent loss of small target details;

(2) Cross layer skip connections directly transmit small target information, avoiding downsampling attenuation;

(3) ECA attention enhances weak features of small targets and suppresses background noise;

(4) Adaptive convolution dynamically adjusts the receptive field and adapts to local details of small targets.

C. REASONS FOR ENHANCING DENSE OVERLAPPING TARGETS

The improvement in precision of dense object detection comes from:

(1) BiFPN+feature fusion makes different target features more distinguishable;

(2) Improve SoftNMS by not easily removing overlapping boxes and preserving the complete target;

(3) EIoU loss improves the accuracy of bounding box regression and enhances the ability to distinguish overlapping targets.

D. ENGINEERING APPLICABILITY ANALYSIS

This model has three major engineering advantages:

(1) High precision: mAP@0.5 =96.8%, meeting the accuracy requirements for power inspection;

(2) High real-time performance: FPS=58, meeting the real-time operation requirements of robots/drones;

(3) Lightweight: With a parameter size of 39.2M, it can be deployed on embedded terminals, edge boxes, and unmanned aerial vehicle (UAV) onboard platforms. The model can be directly integrated into inspection robots, drones, and edge monitoring terminals, providing core visual technology for intelligent power inspection.

VII. CONCLUSION AND PROSPECT

A. RESEARCH CONCLUSION

This article proposes a multi-scale recognition and localization method for electrical components anomalies based on YOLO11 and an improved FPN, ensuring high-precision detection across diverse industrial targets. This unified architectural approach optimizes the identification of intricate grid hardware and microscopic fabric defects to support the intelligent transformation of both power systems and manufacturing. To address the challenges of multi-scale, small targets, dense occlusion, and strong interference background in complex power scenes, an integrated design of backbone enhancement, FPN bidirectional improvement, attention embedding, loss function optimization, and post-processing improvement is carried out. The following conclusions are drawn:

(1) Enhance the backbone network: Adaptive convolution and ECA attention significantly improve the ability to express weak targets and complex background features without increasing computational complexity;

(2) BiFPN+Bidirectional Enhanced Feature Pyramid: Through bidirectional fusion, adaptive weighting, and cross layer skip connections, it significantly improves multi-scale and small object detection capabilities and is the core module for enhancing performance;

(3) Optimizing loss and NMS: EIoU loss and improving SoftNMS effectively enhance the positioning accuracy and discrimination ability of densely overlapping components;

(4) Comprehensive performance: This article's model mAP@0.5 Reaching 96.8%, the small target recall rate has increased by 12.3%, the detection speed is 58FPS, and the parameter count is 39.2M, achieving the optimal balance between accuracy, real-time performance, and lightweight;

(5) Engineering applicability: The model has strong robustness in complex power scenarios and can be directly deployed on inspection robots, drones, and edge terminals to meet the real-time high-precision detection requirements of engineering.

The multi-scale detection capability of the model can be extended to industrial-level surface defects (pixel area $\geq 32\times32$), but for micro fiber-level defects (pixel area $<16\times16$) or high aspect ratio defects, further optimization of the model input resolution and detection head design is required. The core application scenario of the model is power equipment inspection, and cross-scenario adaptation is limited to industrial-level target detection.

B. INSUFFICIENT RESEARCH

There are still three shortcomings in the method proposed in this article:

(1) Ultra small target detection: For ultra small targets with pixel area$<16 \times 16$ (such as micro screws), there is still room for optimization in missed detection rate;

(2) Extreme scenario data: The dataset of extreme weather conditions (blizzards, typhoons) can still be further expanded to enhance the robustness of extreme scenarios;

(3) Cross device generalization: The generalization performance on different hardware platforms (such as Jetson-Nano, Raspberry Pi) can be further optimized.

C. FUTURE OUTLOOK

Future research will deepen from four aspects:

(1) Model structure optimization: Integrating lightweight Transformer and convolution to further enhance the ability of ultra small object detection;

(2) Video stream detection: Combining target tracking algorithms to achieve continuous and stable positioning of electrical components in video streams;

(3) Integrated diagnosis: On the basis of detection, add a fault recognition branch to achieve the integration of "detection+fault diagnosis";

(4) End edge cloud collaboration: By optimizing model compression and quantization techniques, the system adapts to ultra-low power edge terminals to achieve seamless end-edge-cloud collaborative detection across smart grids and high-speed machinery. This integration ensures that real-time monitoring of electrical components and fabric surface quality can be deployed directly onto resource-constrained hardware for autonomous operational control.

AUTHOR CONTRIBUTIONS

Conceptualization –Yanbo Wang, Zongqiang Sui, Rui Sun, Shuai Zhang and Xiaohui Zhang; methodology – Yanbo Wang, Zongqiang Sui, Shuai Zhang and Xiaohui Zhang;

investigation – Yanbo Wang, Zongqiang Sui, Rui Sun, Shuai Zhang and Xiaohui Zhang; writing-original draft preparation – Yanbo Wang, Zongqiang Sui, Rui Sun, Shuai Zhang and Xiaohui Zhang. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] X. Fan and C. Li, "Credit-pool policy and new energy vehicle industrial innovation from the perspective of market regulation," Energy, vol. 322, pp. 135518-135518, 2025, doi: 10.1016/j.energy.2025.135518.
[2] W. Zhu, C. Shi, Z. Chen, et al., "Research on the process of energy poverty alleviation in Chinas provinces by new energy revolution from the perspective of time and space," Energy, vol. 322, pp. 135635-135635, 2025, doi: 10.1016/j.energy.2025.135635.
[3] P. Shi and J. Liu, "Benefits of establishing new energy consumption cities: Evidence from industrial enterprises sustainability," Renewable Energy, vol. 246, pp. 122916-122916, 2025, doi: 10.1016/j.renene.2025.122916.
[4] Y. Yang, Y. Zhang, Y. Zhang, et al., "Comprehensive assessment of the environmental impacts of Chinas new energy industry policy from life cycle perspective," Environmental Impact Assessment Review, vol. 114, pp. 107916-107916, 2025, doi: 10.1016/j.eiar.2025.107916.
[5] X. Wang, L. Zhu, and H. Zheng, "The relationship between technical innovation and market value: Evidence from the Chinas new energy vehicles industry," Energy Policy, vol. 202, pp. 114606-114606, 2025, doi: 10.1016/j.enpol.2025.114606.
[6] J. Pu and W. Chun, "How does dual-credit policy regulate competitive fuel vehicle and new energy vehicle manufacturers? based on operational decision analysis under multiple scenarios," Computers & Industrial Engineering, vol. 203, pp. 111076-111076, 2025, doi: 10.1016/j.cie.2025.111076.
[7] Q. Qin, Z. Wen, Z. Zhou, et al., "Leveraging social media for new energy vehicle policy diffusion in China: A centrallocal government interaction analysis," Transportation Research Part A, vol. 195, pp. 104459-104459, 2025, doi: 10.1016/j.tra.2025.104459.
[8] J. He, "The impact of energy transition policies on urban green innovation: evidence from the new energy demonstration cities in China," Economic Change and Restructuring, vol. 58, no. 3, pp. 34-34, 2025, doi: 10.1007/s10644-025-09867-2.
[9] Z. Xu, Z. Song, and K. Fong Y, "Correction: Xu et al," Perceived Price Fairness as a Mediator in Customer Green Consumption: Insights from the New Energy Vehicle Industry and Sustainable Practices. Sustainability 2025, 17, 166. Sustainability, vol. 17, no. 7, pp. 3265-3265, 2025, doi: 10.3390/su17073265.
[10] A. Kelley, "Energy selects 16 sites for AI data center construction, new energy development," Nextgov.com (Online), 2025.
[11] M. Chai, C. Wu, Y. Luo, et al., "New Energy Demonstration City Policy and Corporate Green Innovation: From the Perspective of Industrial and Regional Spillover Effect," Sustainability, vol. 17, no. 7, pp. 3179-3179, 2025, doi: 10.3390/su17073179.
[12] Z. Liang, Z. Wang, N. Wu, et al., "A New Energy High-Impact Process Weather Classification Method Based on Sensitivity Factor Analysis and Progressive Layered Extraction," Electronics, vol. 14, no. 7, pp. 1336-1336, 2025, doi: 10.3390/electronics14071336.
[13] Z. Yi, "Chinas New Energy Electric Vehicle Industry: Multifaceted Analysis and Future Outlook," Journal of Innovation and Development, vol. 10, no. 3, pp. 59-65, 2025, doi: 10.54097/8hbynn91.
[14] Y. Liu and Y. Wang, "THE FUTURE DEVELOPMENT OF NEW ENERGY VEHICLES BASED ON ARIMA TIME SERIES PREDICTION MODEL," Journal of Computer Science and Electrical Engineering, vol. 7, no. 2, 2025, doi: 10.61784/jcsee3049.
[15] S. Du, L. Chang, and A. Shoukat, "Green financial innovation and digital transformation of new energy enterprises: Evi dence from a quasi-natural experiment," Journal of environmental management, vol. 380, Art. no. 125054, 2025, doi: 10.1016/j.jenvman.2025.125054.
[16] S. Weng, M. Song, W. Tao, et al., "Breaking the inertia of urban energy systems: Does the new energy demonstration city construction improve carbon unlocking efficiency?," Renewable Energy, vol. 244, pp. 122672-122672, 2025, doi: 10.1016/j.renene.2025.122672.
[17] J. Jia, T. Wu, J. Zhou, et al., "Assessment method of new energy hosting capacity in distribution grid considering rotary power flow controllers and demand response," Electric Power Systems Research, vol. 244, pp. 111566-111566, 2025, doi: 10.1016/j.epsr.2025.111566.
[18] C. Lee C, J. Li, and J. Yan, "Can artificial intelligence contribute to the new energy system? Based on the perspective of labor supply," Technology in Society, vol. 81, pp. 102877-102877, 2025, doi: 10.1016/j.techsoc.2025.102877.
[19] Z. Chen, M. Chen T, and S. Jia W, "Simulation and Optimization of New Energy Vehicles Promotion Policy Strategies Con sidering Energy Saving, Carbon Reduction, and Consumers Willingness Based on System Dynamics," Sustainability, vol. 17, no. 7, pp. 2811-2811, 2025, doi: 10.3390/su17072811.
[20] S. Gu, Y. Lu, and H. Gu, "Comparative Study on Static Voltage Stability Indices of New Energy Power Station Grid-Connected System," Journal of Engineering Research and Reports, vol. 27, no. 3, pp. 524-532, 2025, doi: 10.9734/jerr/2025/v27i31449.