electronics

MDPI

Article

TLI-YOLOv5: A Lightweight Object Detection Framework for Transmission Line Inspection by Unmanned Aerial Vehicle

Hanqiang Huang [1,*], Guiwen Lan [1,2,*], Jia Wei [1], Zhan Zhong [1], Zirui Xu [1], Dongbo Li [1] and Fengfan Zou [1]

[1] College of Geomatics and Geoinformation, Guilin University of Technology, Guilin 541006, China
[2] Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Guilin 541006, China
* Correspondence: 2120211860@glut.edu.cn (H.H.); 2009043@glut.edu.cn (G.L.)

Abstract: Unmanned aerial vehicles (UAVs) have become an important tool for transmission line inspection, and the inspection images taken by UAVs often contain complex backgrounds and many types of targets, which poses many challenges to object detection algorithms. In this paper, we propose a lightweight object detection framework, TLI-YOLOv5, for transmission line inspection tasks. Firstly, we incorporate the parameter-free attention module SimAM into the YOLOv5 network. This integration enhances the network's feature extraction capabilities, without introducing additional parameters. Secondly, we introduce the Wise-IoU (WIoU) loss function to evaluate the quality of anchor boxes and allocate various gradient gains to them, aiming to improve network performance and generalization capabilities. Furthermore, we employ transfer learning and cosine learning rate decay to further enhance the model's performance. The experimental evaluations performed on our UAV transmission line inspection dataset reveal that, in comparison to the original YOLOv5n, TLI-YOLOv5 increases precision by 0.40%, recall by 4.01%, F1 score by 1.69%, mean average precision at 50% IoU (mAP50) by 2.91%, and mean average precision from 50% to 95% IoU (mAP50-95) by 0.74%, while maintaining a recognition speed of 76.1 frames per second and model size of only 4.15 MB, exhibiting attributes such as small size, high speed, and ease of deployment. With these advantages, TLI-YOLOv5 proves more adept at meeting the requirements of modern, large-scale transmission line inspection operations, providing a reliable, efficient solution for such demanding tasks.

Keywords: object detection; SimAM attention module; Wise-IoU (WIoU) loss function; YOLOv5n; transmission line inspection

Citation: Huang, H.; Lan, G.; Wei, J.; Zhong, Z.; Xu, Z.; Li, D.; Zou, F. TLI-YOLOv5: A Lightweight Object Detection Framework for Transmission Line Inspection by Unmanned Aerial Vehicle. *Electronics* **2023**, *12*, 3340. https://doi.org/10.3390/electronics12153340

Academic Editors: Gábor Kertész, Sašo Tomažič, Sara Stančin, Peter Sarcevic and Akos Odry

Received: 18 July 2023
Revised: 1 August 2023
Accepted: 2 August 2023
Published: 4 August 2023

1. Introduction

With the mounting electricity demands of society, the inspection and maintenance of transmission lines have become increasingly crucial [1,2]. Traditional manual inspection methods are relatively slow, costly, and encompass inherent risks. The utilization of UAVs for electrical inspections offers many advantages, including safety, efficiency, flexibility, cost-effectiveness, and minimal constraints [3–7]. Consequently, UAVs have become widely adopted in power line inspections, becoming the principal tool for the electric utility sector [8–10].

UAV inspections require the collection of a large volume of image data, and manually inspecting this vast dataset is time-consuming. Furthermore, the quality of the inspection is subject to the subjective judgment and skill level of the personnel. The varying quality of image data could potentially lead to erroneous or missed detections [5]. In recent years, machine vision technology has significantly enhanced inspection efficiency [11]. Jenssen et al. [12] proposed a computer vision-based power line inspection method that uses UAV optical images as the data source, combined with deep learning algorithms for data analysis and detection. This approach allows for the automatic detection of safety risks such as missing pole top pads, incorrectly installed insulators, cracked poles,

and damage from woodpeckers. Han et al. [13] introduced a computer vision algorithm that can detect damaged insulator discs and foreign objects lodged between two discs. Ma et al. [14] proposed a method for the detection of transmission line insulators that combines UAV imagery with binocular vision perception technologies. This method can quickly and intelligently detect insulator damage and absences, while also using the global positioning system (GPS) and UAV flight parameters to calculate the spatial coordinates of the insulators.

These studies demonstrate that machine vision technology can significantly enhance the efficiency of inspection operations and is a crucial strategy in the advancement toward AI-driven power line inspections. However, these models require substantial computational resources, have large model sizes, and have slow processing speeds, creating significant challenges for practical deployment. Therefore, the industry is increasingly demanding lightweight and compact electrical inspection models with high efficiency.

Since 2012, computer vision technology based on deep convolutional neural networks has rapidly developed. Object detection has become a research hotspot in the field of computer vision. Object detection algorithms can be categorized into two main classes: Two-stage and one-stage. Two-stage algorithms involve candidate box generation in the first stage and accurate target localization in the second stage. Representative algorithms include R-CNN [15], Faster R-CNN [16], and Mask R-CNN [17]. In contrast, one-stage algorithms predict the target category and location simultaneously, making them suitable for real-time detection tasks. Notable examples include YOLO [18] and SSD [19]. Thanks to its speed and accuracy, YOLO quickly gained significant attention in various fields, including transportation [20–22], agriculture [23–26], epidemic prevention [27,28], geological monitoring [29], urban management [30], and medical diagnosis [31,32]. The power industry is also exploring the application of YOLO algorithms in power line inspection work.

For instance, Chen et al. [33] proposed an electrical component recognition algorithm framework based on SRCNN and YOLOv3. The SRCNN network is used to perform super-resolution reconstruction of blurry images to expand the dataset, while YOLOv3 is used to recognize electrical components. Chen et al. [34] used YOLOv3 to propose a solution for pole detection and counting based on UAV patrol videos, enabling rapid post-disaster assessment of fallen poles. Tu et al. [35] proposed a model for recognizing towers and insulators based on an improved YOLOv3 algorithm. The authors removed the 52×52 scale feature extraction layer and pruned the three-scale feature extraction layers down to two to enhance computational speed. Simultaneously, the K-means++ clustering method was employed to calculate anchor box dimensions, thereby improving detection accuracy. Zhan [36] created a dedicated dataset for electrical equipment and compared the detection performance of Faster R-CNN, Mask R-CNN, YOLO, and SSD. Bao et al. [37] improved the performance of the YOLOv5x model for detecting defects in insulators and vibration dampers using UAV remote sensing images by incorporating a coordinate attention (CA) module and replacing the original PANet feature fusion framework with a bidirectional feature pyramid network (BiFPN).

However, these existing studies mainly employed the early versions of the YOLO networks, which had inferior comprehensive performance. Even with improvements to these algorithms, their performance is hardly comparable to the new version of the YOLO network. Moreover, as the data sets used in these studies usually have lower image resolution, larger object sizes, and lower background complexity, with only a few types of objects, they cannot adequately meet the needs of actual complex inspection scenarios.

To resolve the limitations of previous works and better address the demands of real-world inspection, we have constructed a tailored, high-quality dataset with characteristics aligned with practical scenarios. On this basis, we propose TLI-YOLOv5, an advanced lightweight object detection framework specifically designed for transmission line inspection. The main contributions and innovations of this paper are summarized as follows:

1. We constructed a UAV transmission line inspection dataset. The dataset includes 1231 high-resolution images, with as many as 8 types of labeled targets, and a total

of 39,151 annotated boxes. The images were captured across various provinces and cities in China, resulting in a dataset with a rich variety of scenarios, a large volume, and high-quality images. This dataset provides a robust foundation for training high-quality models.

2. We introduced YOLOv5n to the field of transmission line inspection, a novel application that has not been explored before. YOLOv5n, the latest model released in the YOLOv5v6.0 version, exhibits faster speeds and a smaller size compared to its predecessors such as YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These characteristics make it particularly well-suited for large-scale, real-time transmission line inspection tasks.
3. We constructed the TLI-YOLOv5, a lightweight object detection framework for transmission line inspection, built upon the foundation of YOLOv5. Firstly, we incorporated the parameter-free attention module, SimAM, into the backbone of the YOLOv5n network, which enables a bolstered feature extraction ability without increasing the network parameters. Secondly, the loss function was improved to Wise-IoU (WIoU), enhancing the model's accuracy, robustness, and generalization ability. Furthermore, we employed transfer learning techniques to expedite the convergence rate during training and augment the learning performance of the model. We also adopted a cosine learning rate decay strategy to ensure a more stable training process, hasten the convergence speed, and thus improve the better generalization ability of the model.
4. We validated the proposed TLI-YOLOv5 on our transmission line inspection dataset. The experimental evaluations revealed that, in comparison to the original YOLOv5n model, the proposed TLI-YOLOv5 model exhibited measurable improvements. Specifically, precision improved by 0.40%, recall increased by 4.01%, the F1 score rose by 1.69%, the mean average precision at 50% IoU (mAP50) increased by 2.91%, and the mean average precision from 50% to 95% IoU (mAP50-95) also increased by 0.74%. Moreover, the model maintained a recognition speed of 76.1 frames per second (FPS) and a compact size of only 4.15 MB.

The remaining sections of this paper are structured as follows. In Section 2, we provide a detailed exposition on the construction of our UAV transmission line inspection dataset, along with a comprehensive elaboration on the architecture and principles of TLI-YOLOv5. This includes the principles and integration of YOLOv5n, the SimAM attention module, and the WIoU loss function. We also describe the training methods involving transfer learning and cosine learning rate decay. In Section 3, we present a comprehensive description of the rigorous experimental procedures undertaken. Section 4 is dedicated to a thorough discussion of the strengths and limitations of the proposed framework. Finally, in Section 5, we provide a concise summary that outlines the main findings and contributions of our study.

2. Materials and Methods

2.1. Dataset Construction

2.1.1. Transmission Line Inspection Dataset

The original imagery for the dataset used in this study was provided by a large-scale power grid company and comprises 13,744 real electrical inspection images. Adhering to the principles of maximizing scene diversity and image clarity, we selected 1231 images for annotation, and manually delineated 39,151 bounding box labels.

The image collection covers a wide range of regions across multiple provinces and cities in China and is characterized by high image quality and complex backgrounds, providing robust support for training high-quality models. The dataset encompasses a myriad of settings, including but not limited to urban landscapes, rural environs, mountainous terrains, and forested areas, as shown in Figure 1.

Figure 1. Showcase of the diverse scenarios covered in the dataset.

The captured images were taken at a diverse array of distances, angles, and weather conditions, as depicted in Figure 2. This diverse collection of samples is specifically curated to ensure a broad spectrum of scenarios and enhance the generalization capabilities of the model.

In alignment with the practical requirements of inspection work, we annotated eight types of targets, including towers, insulators, yoke plates, corona rings, line clamps, vibration dampers, tower signs, and bird nests, as shown in Figure 3. To make the model's output display more neat and easily interpretable, we abbreviated the names of some of the targets. The final annotation labels used for the eight types of targets are as follows: tower, insulator, plate, ring, clamp, damper, sign, and nest. Finally, we partitioned the dataset into three subsets: the training set, validation set, and test set, with a ratio of 8:1:1.

2.1.2. Dataset Annotation

For annotation, we employed the annotation tool labelImg, with the annotation process depicted in Figure 4. Due to the diversity of the categories and the wide viewing angles captured by the drones, a single image could contain dozens of bounding boxes. Across the 1231 images, a total of 39,151 bounding boxes were annotated, averaging approximately 31.8 bounding boxes per image.

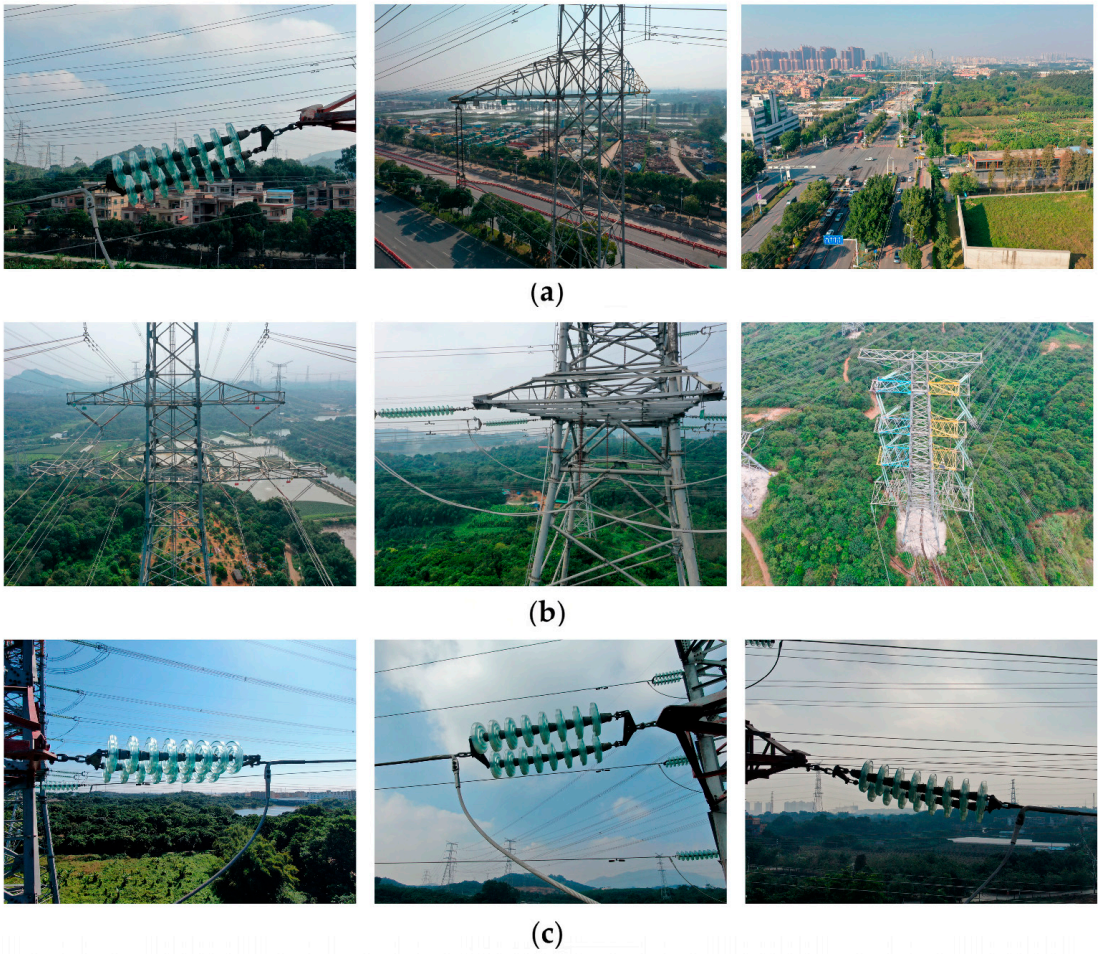


Figure 2. Display of the dataset captured at various distances, angles, and weather conditions: (**a**) Imagery captured at different distances; (**b**) imagery captured from different angles; (**c**) imagery captured under different weather/lighting conditions.

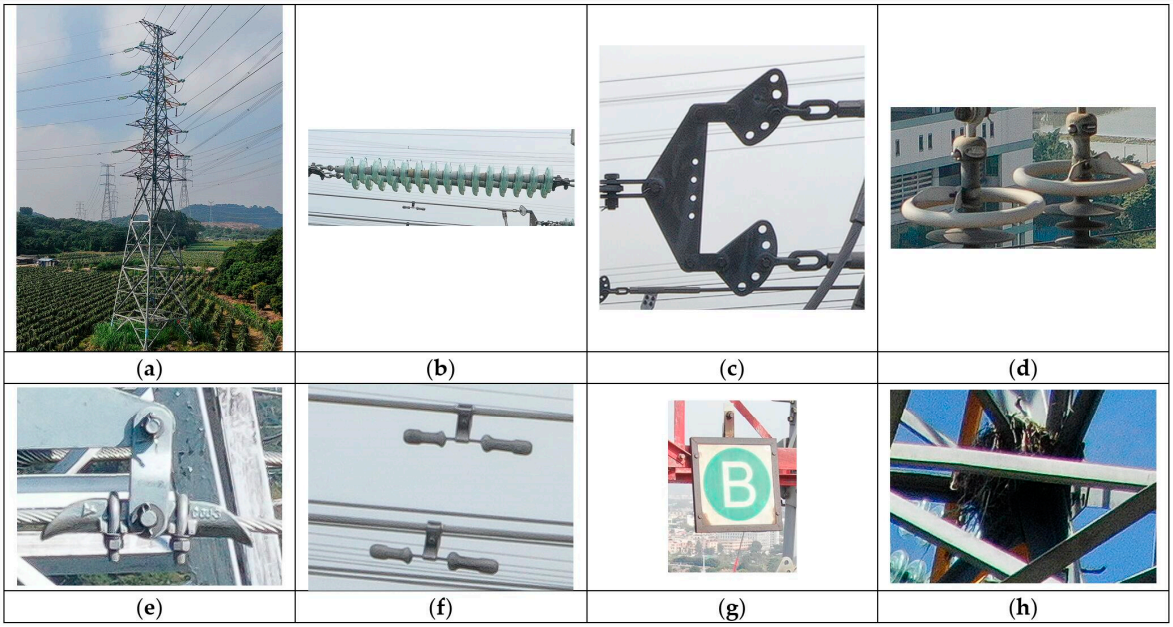


Figure 3. Examples of the 8 types of targets: (**a**) tower; (**b**) insulator; (**c**) yoke plate; (**d**) corona ring; (**e**) line clamp; (**f**) vibration damper; (**g**) tower sign; (**h**) bird nest.

Figure 4. Demonstration of the annotation method.

A statistical overview of the dataset is presented in Figure 5. Figure 5a depicts the distribution of bounding box annotations across the eight object categories in the dataset, providing an overview of the classes and their respective quantities. Figure 5b illustrates the size of each annotated bounding box in the dataset, revealing that a substantial proportion of the objects in the dataset are of small size. This presents a considerable challenge for object detection algorithms. Figure 5c provides a plot of the normalized center coordinates of each bounding box, remapped to a range of 0–1, displaying the spatial distribution of the objects. Figure 5d presents the width-to-height ratio of each bounding box, also normalized to the range of 0–1. The notably darker points in the lower left corner further emphasize the high proportion of smaller targets in the dataset.

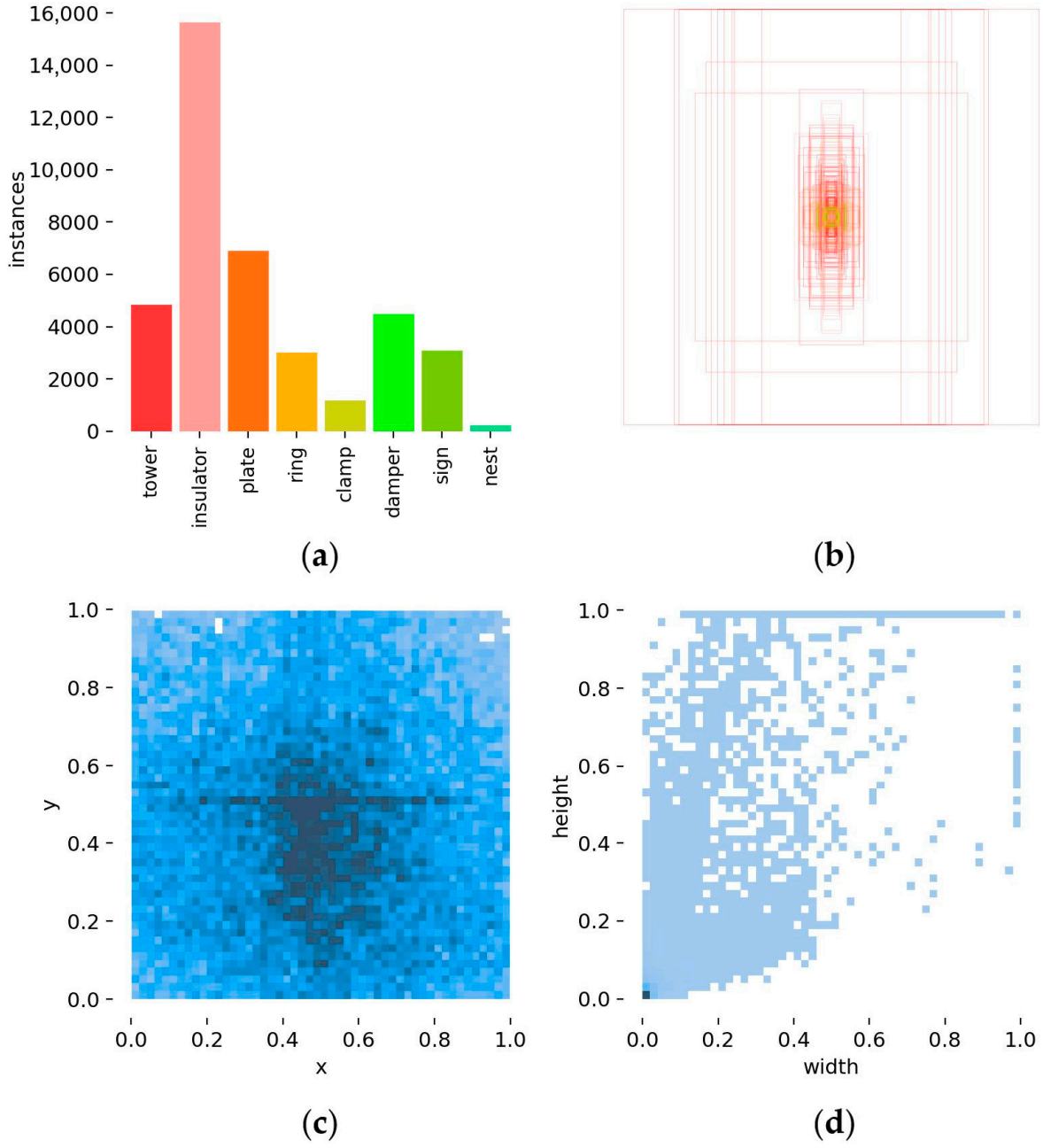


Figure 5. Statistical overview of the dataset: (**a**) Number of instances per class; (**b**) sizes of bounding boxes; (**c**) normalized bounding boxes centers; (**d**) normalized width and height of bounding boxes.

2.2. A Brief Introduction to YOLOv5

2.2.1. The Network Architecture of YOLOv5

In June 2020, Glenn Jocher unveiled YOLOv5 [38], which has since undergone continuous optimization and refinement and is still actively being updated. The network architecture of YOLOv5 is illustrated in Figure 6.

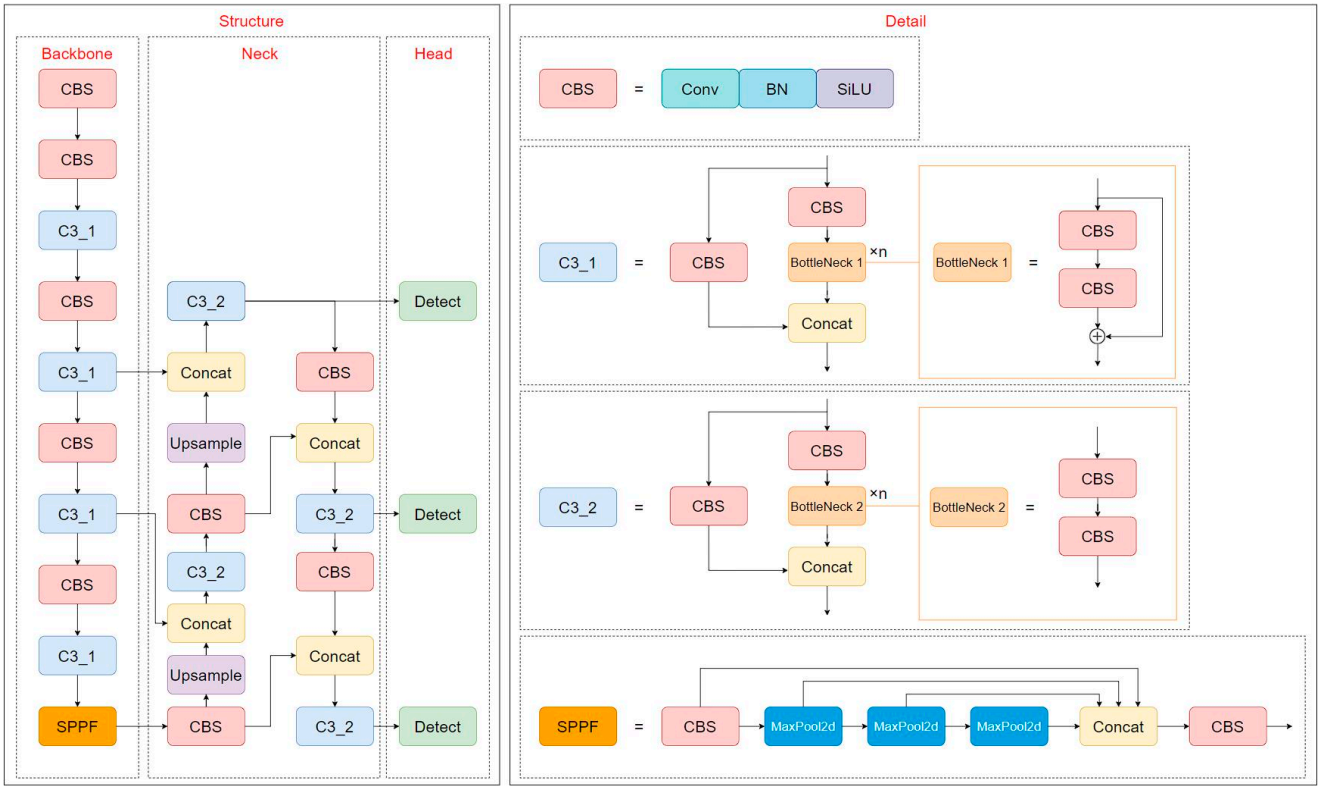


Figure 6. Network architecture of YOLOv5.

The YOLOv5 model is structured into three primary sections: the backbone, the neck, and the head.

The backbone forms the core of the network and is responsible for the initial stages of feature extraction. It is built upon the new CSP-Darknet53 structure, an evolved version of the Darknet architecture from earlier models. The backbone includes CBS, C3, and SPPF modules. The convolution, batch normalization, and SiLU (CBS) module is a fundamental building block in the YOLOv5 architecture. It begins with a convolution operation to extract specific features from the input, followed by batch normalization to standardize these features and increase the stability of the network. Finally, the SiLU (sigmoid linear unit) activation function introduces non-linearity, allowing the model to learn and represent more complex patterns. The C3 module, standing for "CSP Bottleneck with 3 convolutions", further enhances the feature extraction process by splitting the input feature map into two parts, applying a sequence of convolution layers, and then merging the results. The SPPF module, a variant of the spatial pyramid pooling (SPP) module, extracts features at different scales, crucial for handling objects of various sizes in the input images.

The neck of the YOLOv5 model connects the backbone and the head. It utilizes the new CSP-PAN structure to further process the features extracted by the backbone. The neck is responsible for consolidating the high-level semantic feature maps from the backbone and combining them in a way that preserves both the high-resolution details and the broader context.

The head of the YOLOv5 model is responsible for generating the final output, i.e., the object detection predictions. It uses the YOLOv3 head structure, which consists of multiple convolutional layers and output layers. The output layers predict the class probabilities and bounding box coordinates for each grid cell in the feature map. These predictions are then processed to produce the final object detection results.

YOLOv5 introduces several modifications compared to its predecessors, such as the replacement of the Focus structure with a 6×6 Conv2d structure for enhanced efficiency and the substitution of the SPP structure with SPPF to accelerate processing speed.

To enhance the model's generalization capabilities and minimize overfitting, YOLOv5 employs a diverse range of data augmentation techniques. These include mosaic augmentation, copy-paste augmentation, random affine transformations, MixUp augmentation, Albumentations, HSV augmentation, and random horizontal flip. Each of these methods plays a crucial role in improving the model's adaptability to various object scales, translations, and image conditions, thereby boosting its overall performance.

YOLOv5 also utilizes a series of sophisticated training strategies to optimize the model's performance. These include multiscale training, AutoAnchor, exponential moving average (EMA), mixed precision training, and hyper-parameter evolution. Collectively, these strategies optimize the learning process, stabilize training, reduce generalization error, and ensure optimal performance.

2.2.2. Overview of YOLOv5n

The YOLOv5 model is available in different variants: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. These variants are designed to offer a range of trade-offs between computational efficiency and detection performance, allowing users to choose the variant that best suits their specific needs.

YOLOv5n, introduced as a lightweight model in version 6.0 of YOLOv5 [39], is the smallest and fastest model among the YOLOv5 variants. It is designed for applications where computational resources are limited, making it ideal for deployment on devices with low computational power. The compactness of YOLOv5n is achieved by setting lower values for two key hyper-parameters: depth multiple and width multiple.

The depth multiple is a hyper-parameter that controls the depth of the network, which corresponds to the number of layers in the model. A lower-depth multiple results in fewer layers, thereby reducing the computational complexity of the model and making it faster to run. However, this also means that the model may extract fewer features from the input data, which could potentially impact the detection performance.

The width multiple is another hyper-parameter that controls the width of the network, corresponding to the number of channels in each layer. A lower width multiple results in fewer channels, further reducing the computational complexity of the model. However, similar to the depth multiple, a lower width multiple may limit the amount of information that the model can capture in each layer, which could also potentially impact the detection performance.

By adjusting these hyper-parameters, it is possible to modify the network's depth and width without altering the underlying architecture. This results in five variants—YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x—each with varying levels of depth and width. A comparative overview of the depth and width hyper-parameters across these five models is provided in Table 1.

Table 1. The depth and width parameters of the five YOLOv5 variants.

	YOLOv5n	**YOLOv5s**	**YOLOv5m**	**YOLOv5l**	**YOLOv5x**
Depth multiple	0.33	0.33	0.67	1.00	1.33
Width multiple	0.25	0.50	0.75	1.00	1.25

YOLOv5n is particularly advantageous for transmission line inspections due to its computational efficiency and speed, enabling real-time detection crucial for timely issue resolution. Its lower computational demands make it ideal for deployment on resource-constrained devices such as drones, commonly used in these inspections. Furthermore, its efficiency facilitates long-term, high-frequency monitoring, leading to significant energy and resource savings. In emergency scenarios, YOLOv5n's ability to rapidly process data is critical for prompt issue identification and response. While other YOLOv5 variants may offer higher detection performance, their increased computational requirements make them less suitable for power line inspections, particularly on devices with limited computational

power. Therefore, YOLOv5n's balance of efficiency and performance makes it the preferred choice for these tasks.

2.3. The Proposed TLI-YOLOv5

UAV transmission line inspection images are characterized by high image resolution, high background complexity, small detection targets, and a variety of target types, posing significant challenges for real-time multi-class detection. To address these challenges, we propose TLI-YOLOv5, a lightweight object detection framework specifically designed for transmission line inspection. The overall framework of our proposed TLI-YOLOv5 is illustrated in Figure 7. It is mainly comprised of the following components.

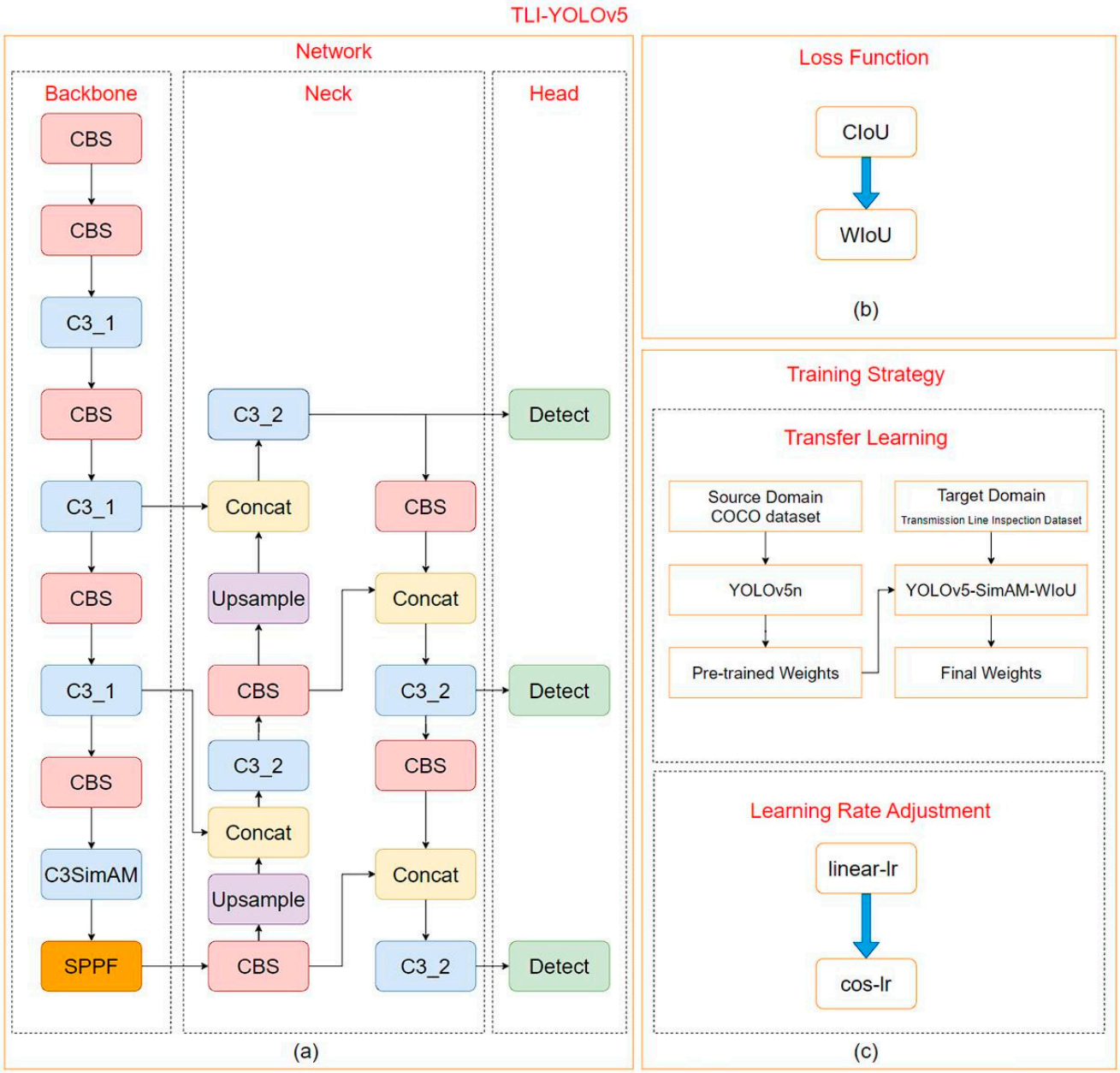


Figure 7. TLI-YOLOv5 framework: (**a**) Improved YOLOv5 network: the last C3 module of the backbone is modified to C3SimAM; (**b**) enhanced loss function: CIoU is refined to WIoU; (**c**) training strategy: implements both transfer learning and cosine learning rate decay strategies.

Firstly, we introduce the SimAM attention module [40] into the YOLOv5n network to enhance its feature extraction capabilities. Specifically, we incorporate the SimAM attention

module into the C3 module to build the C3SimAM module, which replaces the last C3 module in the original YOLOv5 backbone. Additionally, we improve the model's performance by enhancing the loss function using WIoU [41]. To optimize the training process, we employ two effective strategies: transfer learning [42–44] and cosine learning rate decay [45]. These strategies contribute to overall improved detection performance and enable the model to effectively handle the intricacies involved in transmission line inspection.

2.3.1. SimAM Attention Module

In the field of computer vision, a neuron refers to a node in a neural network that receives input from the previous layer and passes it on to the next. In a neural network, each neuron corresponds to a specific feature, such as edges, textures, colors, etc. In visual neuroscience, the most informative neurons are often those that exhibit discharge patterns that are distinct from surrounding neurons. Inspired by this, Yang et al. proposed the SimAM attention module in 2021 [40]. The authors considered neurons that exhibit significant spatial suppression effects to be more important and assigned them higher priority. To identify these neurons, the authors proposed a simple implementation method that measures the linear separability between a target neuron and other neurons. To this end, they defined an energy function to find neurons that exhibit significant spatial suppression effects and assigned them higher priority. Unlike existing channel-domain and spatial-domain attention modules, this module can infer three-dimensional attention weights in the feature map without increasing the parameters in the original network.

The transmission line inspection images captured by UAVs contain complex backgrounds and multiple types of objects. To accurately detect objects of interest, the model needs to learn discriminative features that can distinguish between objects and background areas. However, the original YOLOv5 backbone may not be able to sufficiently extract such discriminative features. To address this issue, we incorporated the SimAM attention module into the YOLOv5 backbone.

For transmission line inspection images, each neuron in the feature maps of the YOLOv5 backbone may respond to different basic visual patterns. By measuring the linear separability between a target neuron and its surrounding neurons, SimAM can identify the most distinctive neurons within each channel, which likely correspond to informative visual patterns critical for distinguishing objects in complex environments. The 3D attention weights generated by SimAM then allow for selectively highlighting these informative neurons while suppressing irrelevant ones. This enhances YOLOv5's capability to extract discriminative features from transmission line inspection images.

We conducted a series of experiments to determine the optimal way to integrate the SimAM attention module into the YOLOv5 network. Initially, we attempted to add the SimAM attention module as an additional layer at various positions within the YOLOv5 backbone, but these modifications did not result in significant performance improvements. So, we tried to combine the SimAM attention module with the C3 module, the i.e., C3SimAM module. The structure of the C3SimAM module is depicted in Figure 8. We then replaced some C3 modules within the backbone with the C3SimAM module. Optimal performance was achieved when only the last C3 module in the backbone was replaced.

Our experiments validate the effectiveness of this improvement, significantly enhancing the network's feature extraction capability. Detailed experimental results and analyses are presented in Section 3.3.4.

The SimAM attention module mainly consists of two steps:

1. Calculate the minimum energy e_t^* with Equation (1).

$$e_t^* = \frac{4\left(\hat{\sigma}^2 + \lambda\right)}{(t - \hat{\mu})^2 + 2\hat{\sigma}^2 + 2\lambda} \tag{1}$$

where $\hat{\mu} = \frac{1}{M}\sum_{i=1}^{M} x_i$, $\hat{\sigma}^2 = \frac{1}{M}\sum_{i=1}^{M} (x_i - \hat{\mu})^2$, t and x are the target neuron and other neurons within a single channel of the input features, respectively. i is the index number

in the spatial dimension, M is the number of neurons in the channel, and λ is a weight constant.

From this, it can be seen that the smaller e_t^*, the larger the difference between t and x, indicating that the importance of the target neuron in the channel is higher, and its weight $1/e_t^*$ is also larger.

2. Enhance the input features are enhanced with Equation (2).

$$\widetilde{X} = sigmoid\left(\frac{1}{E}\right) \odot X \tag{2}$$

In this equation, X denotes the input features, $\widetilde{X}$ denotes the output features, E denotes the combination of e_t^* across all channels and spatial dimensions.

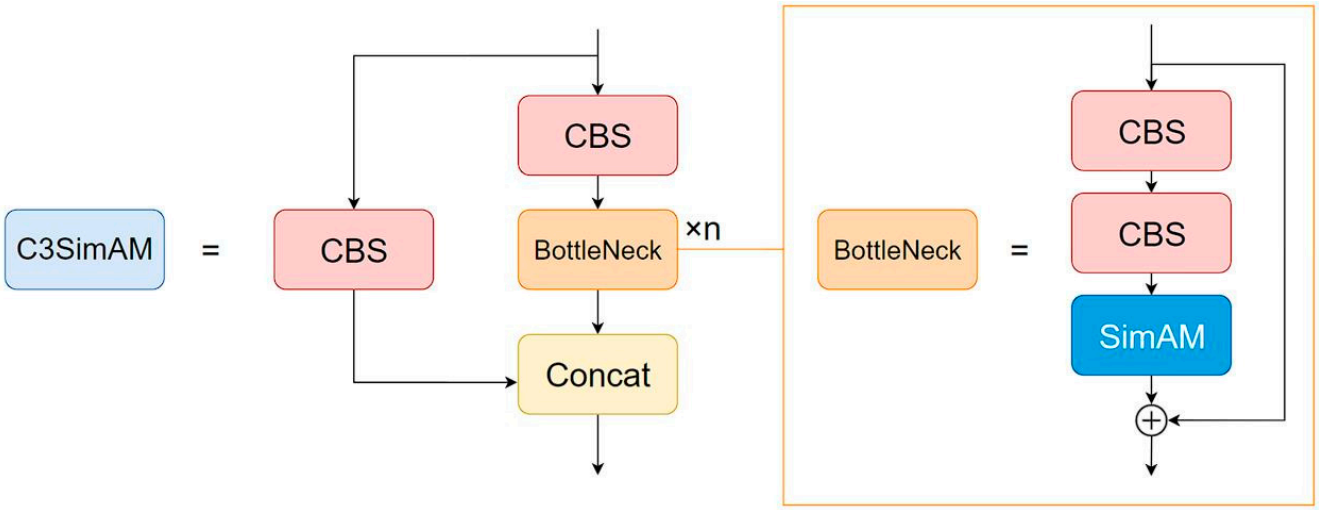


Figure 8. Schematic diagram of the C3SimAM module structure.

2.3.2. Selection of Loss function

The default bounding box regression loss function in YOLOv5 is complete IoU (CIoU) [46]. The CIoU loss function takes into account the differences between the ground truth and predicted bounding boxes in terms of overlap area, center distance, and aspect ratio. Compared to the traditional IoU loss function, the CIoU loss function exhibits superior performance and stability, thus being widely used in various object detection algorithms. However, there still exists room for optimization in the CIoU loss function.

In 2023, Tong et al. [41] proposed a loss function with a dynamic focusing mechanism called Wise-IoU (WIoU). The WIoU loss function evaluates the quality of anchor boxes by calculating their outlier degree and accordingly allocates different gradient gains. Specifically, for anchor boxes with a small outlier degree, corresponding to high-quality anchor boxes, a smaller gradient gain is assigned to prevent overfitting and enhance the model's generalization ability. For anchor boxes with a large outlier degree, corresponding to low-quality anchor boxes, a smaller gradient gain is assigned to mitigate the harmful gradients induced by low-quality examples, thereby improving the model's accuracy and robustness. For anchor boxes with a moderate outlier degree, corresponding to ordinary-quality anchor boxes, a larger gradient gain is assigned to concentrate the bounding box regression task on these anchor boxes. Furthermore, the classification standard for anchor box quality is dynamic, allowing the WIoU loss function to continually determine the most appropriate gradient gain allocation strategy under the current circumstances.

The WIoU loss can be calculated by Equation (3).

$$\mathcal{L}_{WIoU} = r \cdot exp\left(\frac{\rho^2\left(b, b^{gt}\right)}{\left(c^2\right)^*}\right) \cdot (1 - IoU) \tag{3}$$

where $r = \frac{\beta}{\delta\alpha^{\beta-\delta}}$, $\beta = \frac{(1-IoU)^*}{1-IoU} \in [0, +\infty)$, IoU stands for the Intersection over Union between the ground truth and predicted bounding boxes, r is the non-monotonic focusing coefficient, β is the outlier degree, and α and δ are hyper-parameters, the superscript * denotes the operation of detaching variables from the computation graph to prevent gradients that could impede convergence, $\rho\left(b, b^{gt}\right)$ represents the Euclidean distance between the centers of the ground truth and predicted bounding boxes, c represents the diagonal distance of the smallest rectangular area that can contain both the ground truth and predicted bounding boxes.

We performed a series of experiments to determine the optimal selection of α and δ hyper-parameters for the WIoU loss function. The experimental results indicated that the model performed optimally when α and δ were set to 1.9 and 3, respectively. Detailed experimentations related to this are presented in Section 3.3.1.

Furthermore, we conducted a series of experiments that demonstrated the positive impact of the WIoU loss function within our proposed framework. The experimental results show that the model trained using the WIoU loss function exhibits better overall performance than the model trained using the CIoU loss function. Relevant experimental results and analyses can be found in Section 3.3.4.

2.3.3. Transfer Learning

Transfer learning is a method that enhances the detection capability and training speed of a model in a target domain by leveraging features learned from a pre-trained network in a similar domain. Transfer learning can reduce the dependency on the target domain dataset, accelerate training speed, and prevent overfitting, among other benefits.

In this study, we utilize the COCO dataset as the source domain and input it into YOLOv5n to obtain pre-trained weights. These pre-trained weights, along with the transmission line inspection dataset, are then input into the YOLOv5n model integrated with the Simam attention module and WIoU loss function to obtain the final weights.

The COCO dataset comprises over 330,000 images with more than 2 million annotated boxes, covering 80 common object categories in daily life. The images in the COCO dataset are diverse in terms of angles and backgrounds. Using the weight file trained on this dataset as the source domain for transfer learning can positively contribute to the training of the transmission line inspection model.

The effectiveness of this transfer learning approach is verified in our experiments, which can be found in Section 3.3.2.

2.3.4. Learning Rate Adjustment

Different learning rate adjustment strategies have been employed in various versions of YOLOv5. Up until version v6.0, the default adjustment method incorporated a combination of warmup and cosine learning rate decay (cos-lr). Starting with version v6.1 and in subsequent iterations, the default strategy was changed to warmup and linear learning rate decay (linear-lr).

Each of these strategies has distinct advantages and disadvantages that must be considered within the context of the specific task. The cosine learning rate decay strategy applies a cosine function to decay the learning rate, leading to smoother convergence and better fine-tuning of the model parameters towards the end of the training. Conversely, the linear learning rate decay strategy applies a constant rate of decay, providing more predictable training dynamics and potentially benefiting cases where the initial learning rate has been finely tuned.

In our experiments, we have critically assessed both strategies within our detection task. Our results, detailed in Section 3.3.3, led us to opt for the cosine learning rate decay strategy for our final model due to its superior performance in our task.

2.4. Research Methodology Overview

Figure 9 provides a comprehensive and concise representation of the research methodology and workflow employed in our study. Specifically, the process began with the annotation of eight types of targets on real transmission line inspection images, resulting in the construction of a transmission line inspection dataset. Subsequently, the TLI-YOLOv5 framework was developed by integrating the parameter-free attention module, SimAM, the WIoU loss function, and advanced training strategies into the YOLOv5n network. This framework resulted in the realization of a lightweight object detection method tailored for transmission line inspection.

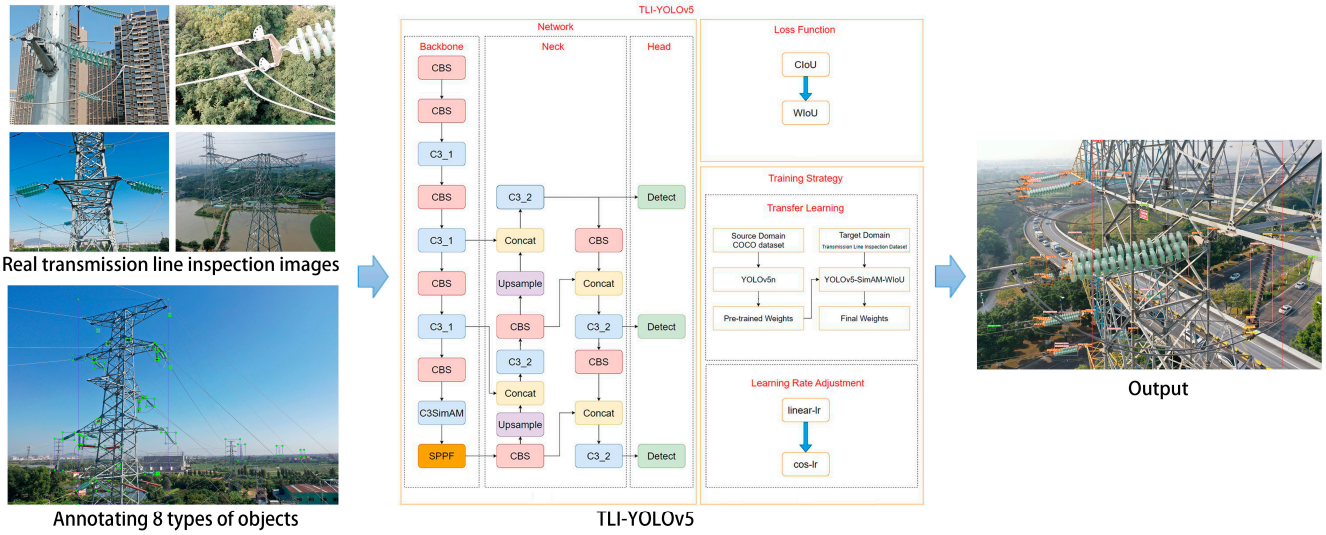


Figure 9. Flowchart of our study method.

3. Results

3.1. Experiment Configuration

To validate the effectiveness of the TLI-YOLOv5 framework, we carried out a series of experiments. The detailed configuration of the experimental environment is outlined in Table 2.

Table 2. System configuration and software environment.

Component	Specification
CPU	Intel(R) Xeon(R) Gold 6240 CPU @ 2.60 GHz
RAM	128 GB
GPU	NVIDIA Quadro RTX5000 16 GB
Operating system	Windows10 pro
CUDA version	10.2
cuDNN version	7.6.5
Deep learning framework	PyTorch 1.8.2
Python version	3.8.8

When inputting images into the YOLOv5 network, the images are adjusted to a predetermined size, with the default size being 640 pixels. Considering that the images from the transmission line inspection dataset used in this study all have a resolution exceeding 4000 × 3000 pixels, substantial image compression occurs when these images are input into the network. This compression can result in significant information loss, leading

to suboptimal model performance. To balance model performance and computational resources, we have chosen an input size of 1280 pixels, a batch size of 4, and 300 epochs for training.

3.2. Evaluation Metrics

In this study, the assessment of the model's performance was conducted using commonly utilized metrics in object detection tasks, which include precision (P), recall (R), F1 score (F1), mean average precision at 50% IoU (mAP50), and mean average precision from 50% to 95% IoU (mAP50-95). Furthermore, we adopt frames per second (FPS) as the metric for evaluating the speed of the model.

Precision, also known as positive predictive value, measures the proportion of true positives (TP) out of the sum of true positives and false positives (FP), which are instances incorrectly identified as positive. The computation formula for precision is shown in Equation (4). Precision serves to evaluate how many of the detected objects are correct detections.

$$P = \frac{TP}{(TP + FP)} \tag{4}$$

Recall, sometimes referred to as sensitivity or true positive rate, quantifies the proportion of true positives (TP) out of the sum of true positives and false negatives (FN), which are positive instances incorrectly identified as negative. The computation formula for recall is presented in Equation (5). Recall is utilized to evaluate how well the model can identify all positive instances.

$$R = \frac{TP}{(TP + FN)} \tag{5}$$

The F1 score is the harmonic mean of precision and recall. It aims to balance precision and recall by providing a single metric that encapsulates the model's performance in terms of both precision and recall.

The mean average precision at 50% IoU (mAP50) is computed by averaging the precision-recall curve's area under the curve across all classes at an IoU threshold of 50%.

The mean average precision from 50% to 95% IoU (mAP50-95) is calculated by averaging the mAP values over IoU thresholds from 0.50 to 0.95 (inclusive) with a step size of 0.05.

3.3. Experimental Results and Analysis

3.3.1. Experiment on WIoU Hyper-Parameter Selection

In paper [41], the authors enumerated four sets of values for the hyper-parameters α and δ, which are $\alpha = 2.5$, $\delta = 2$; $\alpha = 1.9$, $\delta = 3$; $\alpha = 1.6$, $\delta = 4$; and $\alpha = 1.4$, $\delta = 5$. The outlier degree β and gradient gain r curves corresponding to these four sets of values are depicted in Figure 10.

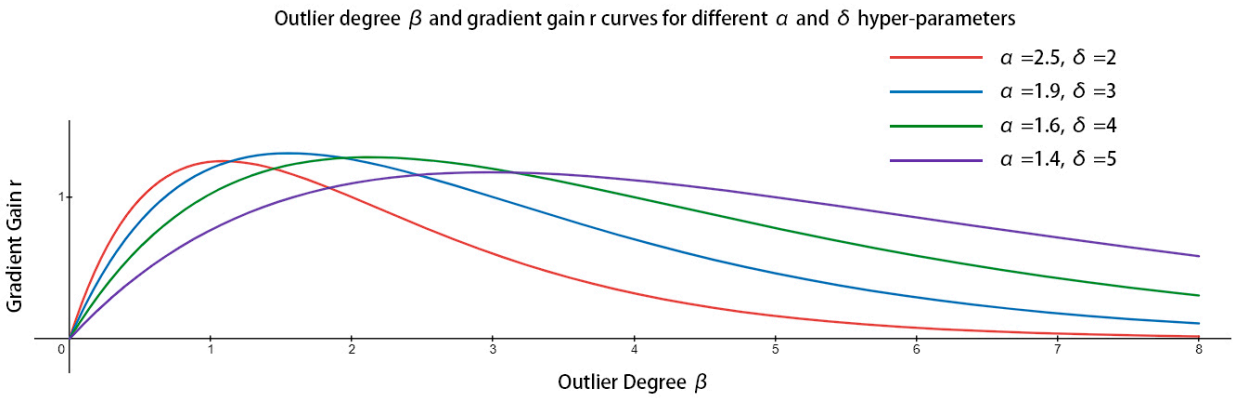


Figure 10. Outlier degree β and gradient gain r curves for different α and δ hyper-parameters.

In the context of the TLI-YOLOv5 framework, we have tested the model performance for these four sets of hyper-parameters, as shown in Table 3. From these experiments, we have determined the optimal values of α and δ to be 1.9 and 3, respectively.

Table 3. Model performance for different α and δ hyper-parameters.

Values of α and δ	P	R	F1	mAP50	mAP50-95
$\alpha = 2.5$ $\delta = 2$	0.736	0.503	0.59	0.545	0.271
$\alpha = 1.9$ $\delta = 3$	0.752	0.519	0.60	0.566	0.271
$\alpha = 1.6$ $\delta = 4$	0.770	0.497	0.60	0.554	0.265
$\alpha = 1.4$ $\delta = 5$	0.713	0.497	0.58	0.541	0.266

3.3.2. Experiment on the Effectiveness of Transfer Learning

To validate the effectiveness of the transfer learning strategy, we conducted two sets of experiments within the TLI-YOLOv5 framework. As shown in Table 4, significant improvements were observed across all metrics when the transfer learning strategy using the COCO dataset as the source domain was implemented.

Table 4. Impact of transfer learning.

Training Strategy	P	R	F1	mAP50	mAP50-95
Without transfer learning	0.660	0.432	0.52	0.477	0.211
With transfer learning	0.752	0.519	0.60	0.566	0.271

The positive impact of using the COCO dataset as the source domain can be attributed to several factors. Firstly, the COCO dataset contains a variety of images with diverse and complex backgrounds, mirroring the conditions often encountered in transmission line inspection tasks. Secondly, the COCO dataset covers 80 common object categories in daily life with over 2 million annotated boxes. Many object features and edge characteristics in the COCO dataset share similarities with features found in transmission line components.

Consequently, the application of transfer learning from the COCO dataset to the transmission line inspection dataset proves to be an effective strategy, enhancing both the performance and robustness of the TLI-YOLOv5 model.

3.3.3. Selection of Learning Rate Adjustment Strategy

The performance comparison of models trained with two different learning rate decay strategies, linear learning rate decay and cosine learning rate decay, is presented in Table 5. Notably, the model trained with linear learning rate decay exhibits superior precision. However, the model trained with cosine learning rate decay outperforms the linear learning rate decay in recall, F1 score, mAP50, and mAP50-95 metrics. Overall, the model trained with the cosine learning rate decay strategy shows better comprehensive performance.

Table 5. Performance comparison of models trained with linear learning rate decay (linear-lr) and cosine learning rate decay (cos-lr).

Training Strategy	P	R	F1	mAP50	mAP50-95
With linear-lr	0.760	0.500	0.59	0.551	0.269
With cos-lr	0.752	0.519	0.60	0.566	0.271

To further investigate the impact of cosine learning rate decay on our model, we incorporated it as an independent component in the ablation experiments of TLI-YOLOv5, as detailed in Section 3.3.4.

3.3.4. Ablation Experiments of TLI-YOLOv5

To gain insights into the contribution of each component—the SimAM attention module, the WIoU loss function, and the cosine learning rate decay strategy—we conducted a total of eight experiments, including different combinations of these components. The results of these experiments are summarized in Table 6.

Table 6. Ablation results of each component of TLI-YOLOv5.

SimAM	WIoU	cos-lr	P	R	F1	mAP50	mAP50-95	FPS	GFLOPs
			0.749	0.499	0.59	0.550	0.269	76.9	4.2
√			0.748	0.501	0.59	0.554	0.270	76.1	4.2
	√		0.707	0.520	0.59	0.551	0.270	76.6	4.2
		√	0.724	0.502	0.58	0.556	0.268	76.9	4.2
√	√		0.760	0.500	0.59	0.551	0.269	75.2	4.2
√		√	0.723	0.516	0.59	0.561	0.274	75.2	4.2
	√	√	0.766	0.509	0.61	0.560	0.270	76.9	4.2
√	√	√	0.752	0.519	0.60	0.566	0.271	76.1	4.2

The baseline experiment employs the original YOLOv5n model without any of these enhancements.

The experimental results indicate that when the SimAM attention module and the WIoU loss function are independently implemented, they both result in marginal improvements to the overall performance. However, the standalone application of the cosine learning rate decay strategy exhibits a slight decrement in performance. Yet, it should be noted that these effects are quite subtle.

Upon combining any two enhancements—either SimAM and WIoU, SimAM and cos-lr, or WIoU and cos-lr—the performance measures notably surpass those when any single component is applied independently. This suggests that the amalgamation of these techniques leads to a synergistic boost in model performance.

Remarkably, the simultaneous incorporation of all three modifications yields peak performance across all combinations, illustrating a potential synergistic effect. Specifically, improvements are seen as increases of 0.40% in precision, 4.01% in recall, 1.69% in the F1 score, 2.91% in mAP50, and 0.74% in mAP50-95, respectively.

Importantly, the implementation of these three enhancements does not alter the GFLOPs value of the model, and the model's weight file retains the same 4.15 MB size as the original YOLOv5n. The difference in FPS is negligible. This indicates that the TLI-YOLOv5 framework successfully maintains the lightweight characteristics of YOLOv5n while simultaneously improving detection performance.

3.3.5. Comparison with Mainstream Networks

A comparative analysis was conducted between our proposed TLI-YOLOv5 and several mainstream networks using our transmission line inspection dataset.

Specifically, we included YOLOv3 in the comparison since it was used in the studies [33–35] for transmission line inspection. Additionally, we selected YOLOv3-tiny, a lightweight version of YOLOv3, aimed at achieving faster inference speed at the cost of some accuracy. Furthermore, we incorporated the latest lightweight algorithms from the YOLO series, namely YOLOv6n and YOLOv8n (YOLOv7 currently does not have a lightweight variant available), into our comparison. YOLOv6n is renowned for its exceptional performance with respect to both accuracy and efficiency. On the other hand, YOLOv8n is regarded as the state-of-the-art lightweight variant among the YOLO series.

The results are summarized in Table 7.

Table 7. Comparison with mainstream networks.

Model	P	R	mAP50	mAP50-95	FPS	GFLOPs
YOLOv3	0.795	0.566	0.625	0.318	15.8	154.7
YOLOv3-tiny	0.588	0.365	0.385	0.159	72.5	12.9
YOLOv6n	0.704	0.421	0.453	0.249	74.6	11.4
YOLOv8n	0.753	0.432	0.475	0.255	76.9	8.9
TLI-YOLOv5	0.752	0.519	0.566	0.271	76.1	4.2

The results clearly demonstrate the advantages and characteristics of each network in terms of accuracy, speed, and model size. Notably, our proposed TLI-YOLOv5 model achieves competitive performance.

Overall, these comparative analyses highlight the effectiveness of TLI-YOLOv5 as a promising object detection framework for transmission line inspection, offering a balanced combination of accuracy and efficiency for real-world applications.

4. Discussion

Table 8 displays the detection performance of TLI-YOLOv5 for eight different objects. The confusion matrix of TLI-YOLOv5's detection results is illustrated in Figure 11. The *x*-axis represents the actual categories, while the *y*-axis indicates the predicted categories. Each cell in the matrix represents the proportion of instances where the model predicted a given actual category as the predicted category. All proportion values in the matrix have been normalized to the range of 0 to 1.

Table 8. Performance of TLI-YOLOv5 on different object recognition.

Object	P	R	mAP50	mAP50-95
All	0.752	0.519	0.566	0.271
Tower	0.755	0.594	0.699	0.340
Insulator	0.753	0.469	0.542	0.263
Plate	0.668	0.504	0.542	0.252
Ring	0.732	0.283	0.348	0.183
Clamp	0.734	0.534	0.533	0.316
Damper	0.870	0.565	0.629	0.310
Sign	0.680	0.341	0.377	0.175
Nest	0.822	0.864	0.860	0.328

Among all eight categories, the model performs best in detecting nest, followed by vibration damper and tower. The detection results for insulator, yoke plate, and line clamp are at an average level. In contrast, the worst-performing categories are corona ring and tower sign, whose performance is significantly lower than the other six objects.

Delving into the images of these two problematic categories, we can identify the contributing factors to these discrepancies.

For the corona ring, it holds the position of the smallest object among all eight categories. Its small size inherently hampers detection effectiveness. Specifically, smaller objects occupy fewer pixels in an image, making it more challenging for the model to learn and extract meaningful features.

As for the tower sign, in some images captured at mid to long distances, tower signs still have high recognizability due to their distinctive features, despite their small size. Thus, we annotated them in the dataset. However, similar to other small objects, the model faces difficulty accurately detecting them. As illustrated in Figure 12, among the seven tower signs that we annotated in the original image, only one was successfully detected by the model.

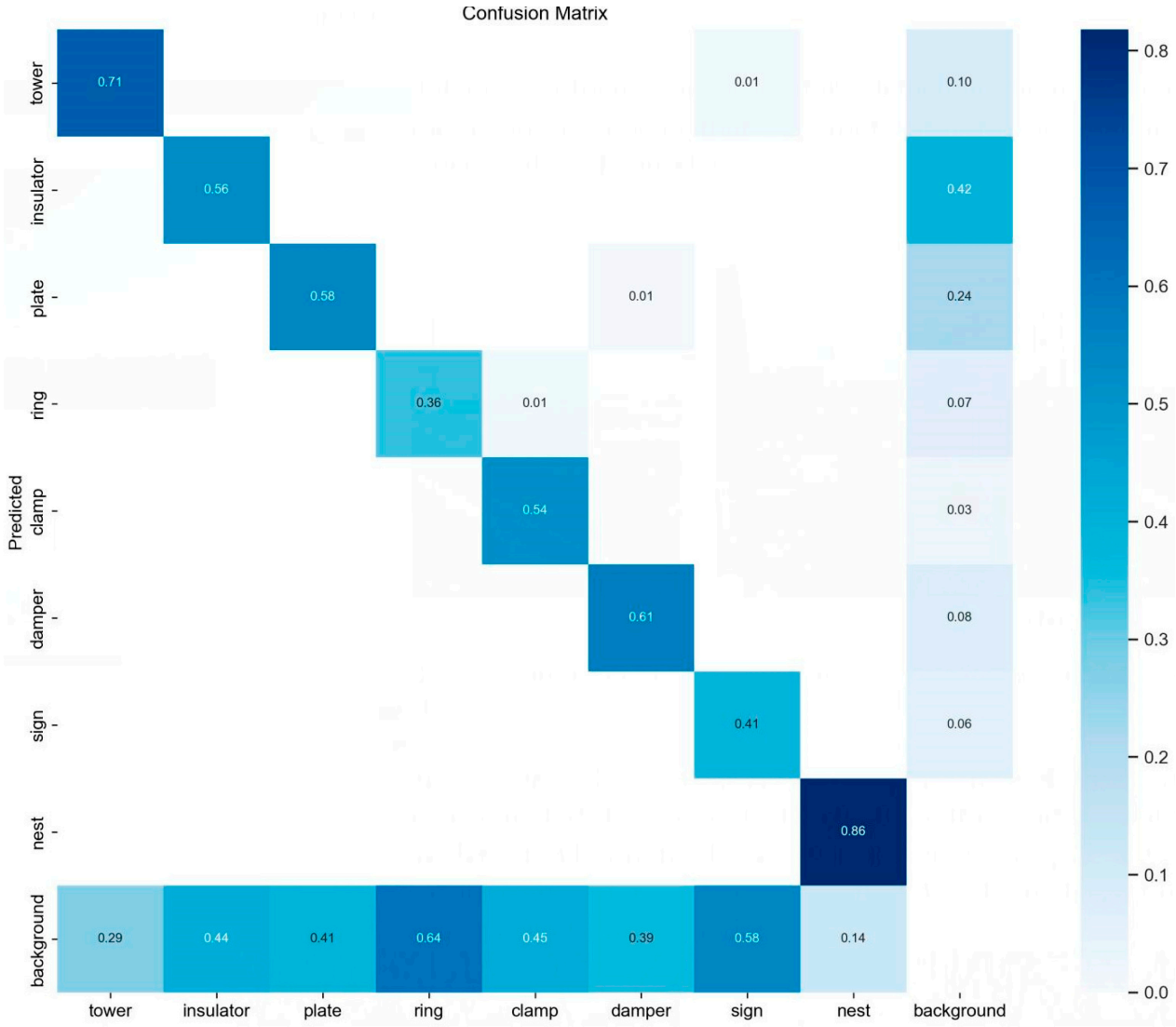


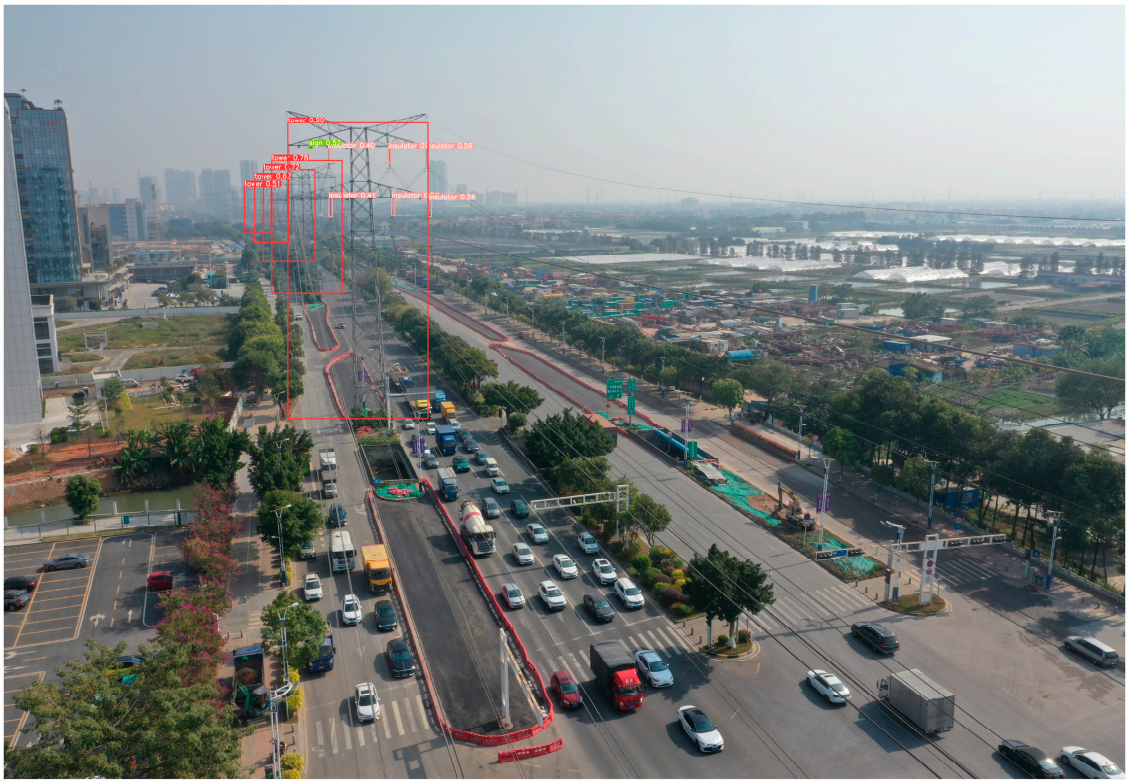
Figure 11. Confusion matrix.

Figure 13 provides a clear visual comparison of the performance differences between the original YOLOv5n and TLI-YOLOv5. It is evident that the confidence threshold of the detection bounding boxes by TLI-YOLOv5 is generally higher than that of the original YOLOv5n. TLI-YOLOv5 successfully detects the largest object in the image, the tower, which YOLOv5n fails to detect.

Figure 14 presents the detailed sections of Figure 13. As demonstrated in Figure 14a, TLI-YOLOv5 successfully detected the obscured object hidden behind the tower body, which the original YOLOv5n was unable to detect. Furthermore, as illustrated in Figure 14b, TLI-YOLOv5 generates more accurate detection bounding boxes than the original YOLOv5n for densely arranged objects.

As a lightweight object detection framework designed for real-time detection and video detection, TLI-YOLOv5, similar to other lightweight models, compromises a degree of precision for a higher detection speed. In the context of UAV transmission line inspection, the UAV's visuals constitute a continuous video stream. The speed of detection not only determines the fluidity of the UAV's operation but also significantly affects the efficiency of the inspection work. Through testing, TLI-YOLOv5's performance metrics satisfy the demands of such tasks, striking a well-managed balance between precision and speed.

(a)



(b)

Figure 12. Detection result of tower signs in a sample image: (**a**) original annotations; (**b**) detection results.

(**a**) (**b**)

Figure 13. Comparison of detection results between YOLOv5n and TLI-YOLOv5: (**a**) YOLOv5n; (**b**) TLI-YOLOv5.

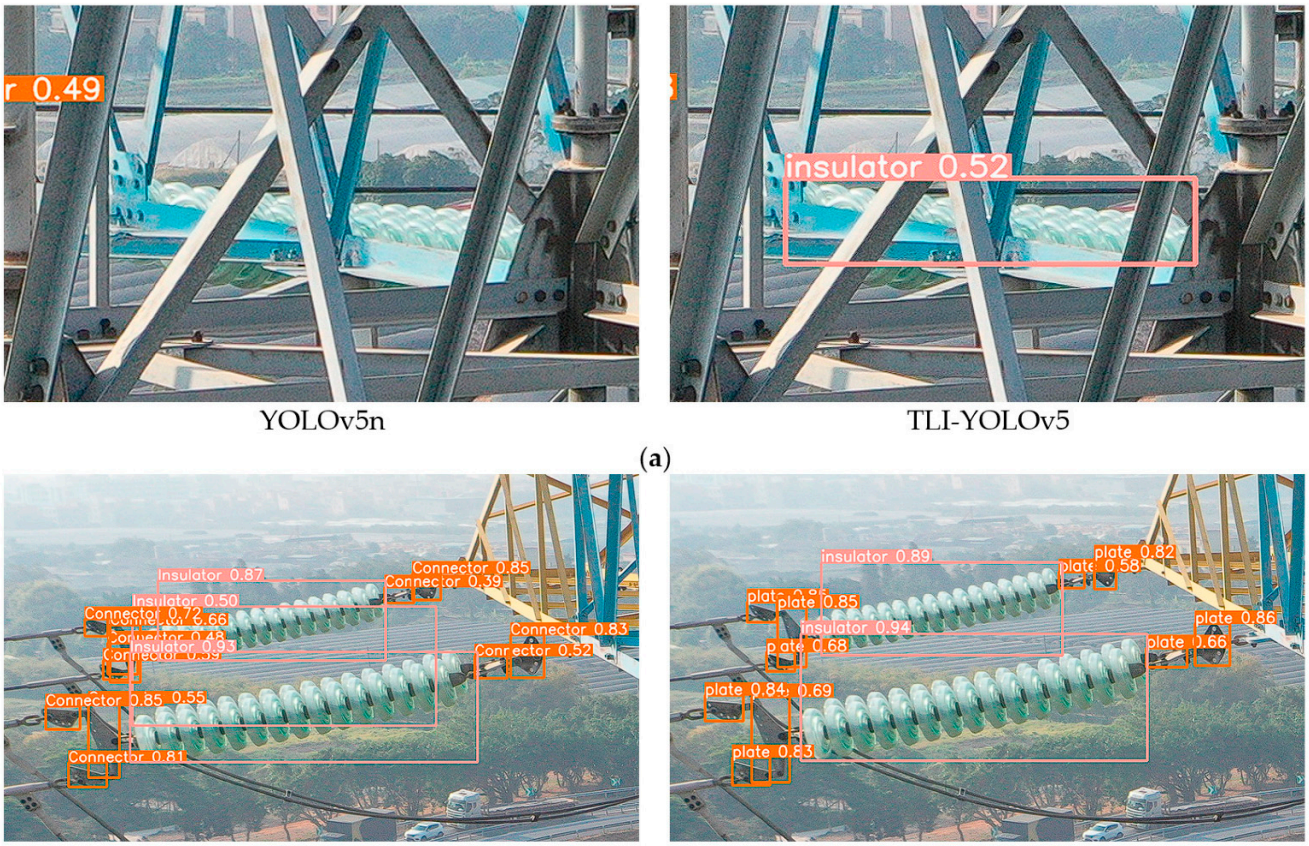


Figure 14. Comparison of detection results: (**a**) obscured object; (**b**) densely arranged objects.

To further analyze the effectiveness of our proposed TLI-YOLOv5, we compare it with some other existing object detection methods designed for UAV-based transmission line inspection.

Previous works [33–35] adopted YOLOv3 as the detection algorithm. However, YOLOv3 has been widely recognized to be surpassed by YOLOv5 in comprehensive performance, which is also validated in our experiments in Section 3.3.5. Previous works [36,37] employed earlier versions of YOLOv5, whose capabilities fall far behind the 7.0 version

used in our study. Moreover, they did not utilize the most lightweight YOLOv5n model, thus their detection speed is much lower compared to our approach.

Most importantly, our dataset contains eight types of transmission line objects and is annotated based on real-world UAV inspection images. Therefore, our proposed method has a wider range of application scenarios and is more suitable for the needs of modern unmanned aerial vehicle transmission line inspection. It demonstrates noticeable superiority over existing approaches.

Despite its strengths, our work has certain limitations. First, transmission line inspections involve a wide array of components. Due to constraints of time and manpower, our research only annotated and trained eight types of components. In future work, we plan to incorporate more components into our detection targets. Additionally, we aim to augment our dataset with a diverse range of medium- and large-sized objects, enabling the model to learn and capture more intricate feature details. Second, with the continuous introduction of more sophisticated algorithms and improvement strategies in the field of object detection, it is worth exploring their application to UAV transmission line inspection tasks. For example, recent works such as [47,48] on network pruning could help reduce model complexity. Techniques such as weight quantization [49,50] may be adopted for significant memory reduction. We hope our work can provide a baseline for future explorations to further advance UAV-based transmission line inspection capabilities.

5. Conclusions

This study proposes TLI-YOLOv5, a lightweight object detection framework specially developed for transmission line inspection. By integrating the parameter-free attention module SimAM, the WIoU loss function, and incorporating advanced training strategies, TLI-YOLOv5 demonstrates enhanced performance in various metrics compared to the original YOLOv5n, while maintaining high recognition speed and compactness. This work contributes a practical and efficient solution for large-scale, real-time transmission line inspection tasks.

However, there remain some limitations in our current research. First, the number of object categories is still limited. In future work, more types of components could be incorporated as detection targets to expand the model's capabilities. Second, While TLI-YOLOv5 achieves a balance between accuracy and efficiency, using more advanced networks could potentially further boost detection performance. We aim to explore state-of-the-art architectures to continue pushing the frontiers. Third, creating larger datasets covering more environments would also be beneficial. Overall, this study offers a robust baseline model, and future research will focus on expanding in terms of category diversity, maximizing accuracy, and improving generalizability.

Although object detection has made rapid progress, there remain challenges in integration and implementation within real-world transmission line inspection applications. Factors such as dataset limitations, model generalization, efficiency constraints, and transition costs can hinder adoption. Our future work will concentrate on tackling these pitfalls to facilitate the broader deployment of object detection algorithms in transmission line inspection systems. We believe continued research and technical development efforts will progressively bridge the gap between state-of-the-art techniques and practical usage in the field of transmission line inspection.

Author Contributions: Conceptualization, H.H. and G.L.; methodology, H.H. and J.W.; software, H.H. and Z.Z.; validation, H.H., G.L., J.W., Z.Z., Z.X., D.L. and F.Z.; formal analysis, H.H., G.L. and J.W.; investigation, H.H., J.W. and D.L.; resources, G.L.; data curation H.H., Z.Z., Z.X. and F.Z.; writing—original draft preparation, H.H.; writing—review and editing, H.H. and G.L.; visualization, Z.Z., Z.X. and D.L.; supervision, G.L.; project administration, G.L.; funding acquisition, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by a grant from the National Natural Science Foundation of China (Grant number 41861050).

Data Availability Statement: The data are not publicly available due to the confidentiality of the research projects.

Acknowledgments: The authors extend their profound gratitude to the anonymous reviewers for their insightful critiques and constructive suggestions on the manuscript.

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

References

1. Luo, Y.; Yu, X.; Yang, D.; Zhou, B. A survey of intelligent transmission line inspection based on unmanned aerial vehicle. *Artif. Intell. Rev.* **2023**, *56*, 173–201. [CrossRef]
2. Alhassan, A.B.; Zhang, X.; Shen, H.; Xu, H. Power transmission line inspection robots: A review, trends and challenges for future research. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105862. [CrossRef]
3. Larrauri, J.I.; Sorrosal, G.; González, M. Automatic system for overhead power line inspection using an Unmanned Aerial Vehicle—RELIFO project. In Proceedings of the 2013 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 28–31 May 2013; pp. 244–252.
4. Luque-Vega, L.F.; Castillo-Toledo, B.; Loukianov, A.; Gonzalez-Jimenez, L.E. Power line inspection via an unmanned aerial system based on the quadrotor helicopter. In Proceedings of the MELECON 2014—2014 17th IEEE Mediterranean Electrotechnical Conference, Beirut, Lebanon, 13–16 April 2014; pp. 393–397.
5. Matikainen, L.; Lehtomäki, M.; Ahokas, E.; Hyyppä, J.; Karjalainen, M.; Jaakkola, A.; Kukko, A.; Heinonen, T. Remote sensing methods for power line corridor surveys. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 10–31. [CrossRef]
6. Wang, C.-N.; Yang, F.-C.; Vo, N.T.; Nguyen, V.T.T. Wireless communications for data security: Efficiency assessment of cybersecurity industry—A promising application for UAVs. *Drones* **2022**, *6*, 363. [CrossRef]
7. Wang, C.-N.; Yang, F.-C.; Vo, N.T.; Nguyen, V.T.T. Enhancing Lithium-Ion Battery Manufacturing Efficiency: A Comparative Analysis Using DEA Malmquist and Epsilon-Based Measures. *Batteries* **2023**, *9*, 317. [CrossRef]
8. Bian, J.; Hui, X.; Zhao, X.; Tan, M. A novel monocular-based navigation approach for UAV autonomous transmission-line inspection. In Proceedings of the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 1–7.
9. Li, X.; Li, Z.; Wang, H.; Li, W. Unmanned aerial vehicle for transmission line inspection: Status, standardization, and perspectives. *Front. Energy Res.* **2021**, *9*, 713634. [CrossRef]
10. Hui, X.; Bian, J.; Zhao, X.; Tan, M. Vision-based autonomous navigation approach for unmanned aerial vehicle transmission-line inspection. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1729881417752821. [CrossRef]
11. Mirallès, F.; Pouliot, N.; Montambault, S. State-of-the-art review of computer vision for the management of power transmission lines. In Proceedings of the 2014 3rd International Conference on Applied Robotics for the Power Industry, Foz do Iguacu, Brazil, 14–16 October 2014; pp. 1–6.
12. Jenssen, R.; Roverso, D. Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *Int. J. Electr. Power Energy Syst.* **2018**, *99*, 107–120.
13. Han, Y.; Liu, Z.; Lee, D.; Liu, W.; Chen, J.; Han, Z. Computer vision–based automatic rod-insulator defect detection in high-speed railway catenary system. *Int. J. Adv. Robot. Syst.* **2018**, *15*, 1729881418773943. [CrossRef]
14. Ma, Y.; Li, Q.; Chu, L.; Zhou, Y.; Xu, C. Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. *Remote Sens.* **2021**, *13*, 230. [CrossRef]
15. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
16. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 91–99. [CrossRef]
17. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
18. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
19. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.-Y.; Berg, A.C. Ssd: Single shot multibox detector. In Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, Part I 14, 2016. pp. 21–37.
20. Dewi, C.; Chen, R.-C.; Liu, Y.-T.; Jiang, X.; Hartomo, K.D. Yolo V4 for advanced traffic sign recognition with synthetic training data generated by various GAN. *IEEE Access* **2021**, *9*, 97228–97242. [CrossRef]
21. Lan, W.; Dang, J.; Wang, Y.; Wang, S. Pedestrian detection based on YOLO network model. In Proceedings of the 2018 IEEE International Conference on Mechatronics and Automation (ICMA), Changchun, China, 5–8 August 2018; pp. 1547–1551.
22. Laroca, R.; Severo, E.; Zanlorensi, L.A.; Oliveira, L.S.; Gonçalves, G.R.; Schwartz, W.R.; Menotti, D. A robust real-time automatic license plate recognition based on the YOLO detector. In Proceedings of the 2018 International Joint Conference on Neural Networks (IJCNN), Shah Alam, Malaysia, 7 August 2018; pp. 1–10.

23. Liu, G.; Nouaze, J.C.; Touko Mbouembe, P.L.; Kim, J.H. YOLO-tomato: A robust algorithm for tomato detection based on YOLOv3. *Sensors* **2020**, *20*, 2145. [CrossRef] [PubMed]
24. Liu, J.; Wang, X. Tomato diseases and pests detection based on improved Yolo V3 convolutional neural network. *Front. Plant Sci.* **2020**, *11*, 898. [CrossRef] [PubMed]
25. Wu, D.; Lv, S.; Jiang, M.; Song, H. Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments. *Comput. Electron. Agric.* **2020**, *178*, 105742. [CrossRef]
26. Tian, Y.; Yang, G.; Wang, Z.; Wang, H.; Li, E.; Liang, Z. Apple detection during different growth stages in orchards using the improved YOLO-V3 model. *Comput. Electron. Agric.* **2019**, *157*, 417–426. [CrossRef]
27. Loey, M.; Manogaran, G.; Taha, M.H.N.; Khalifa, N.E.M. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustain. Cities Soc.* **2021**, *65*, 102600. [CrossRef]
28. Punn, N.S.; Sonbhadra, S.K.; Agarwal, S.; Rai, G. Monitoring COVID-19 social distancing with person detection and tracking via fine-tuned YOLO v3 and Deepsort techniques. *arXiv* **2020**, arXiv:2005.01385.
29. Cheng, L.; Li, J.; Duan, P.; Wang, M. A small attentional YOLO model for landslide detection from satellite remote sensing im ages. *Landslides* **2021**, *18*, 2751–2765. [CrossRef]
30. Du, Y.; Pan, N.; Xu, Z.; Deng, F.; Shen, Y.; Kang, H. Pavement distress detection and classification based on YOLO network. *Int. J. Pavement Eng.* **2021**, *22*, 1659–1672. [CrossRef]
31. Aly, G.H.; Marey, M.; El-Sayed, S.A.; Tolba, M.F. YOLO based breast masses detection and classification in full-field digital mammograms. *Comput. Methods Programs Biomed.* **2021**, *200*, 105823. [CrossRef] [PubMed]
32. Ünver, H.M.; Ayan, E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm. *Diagnostics* **2019**, *9*, 72. [CrossRef] [PubMed]
33. Chen, H.; He, Z.; Shi, B.; Zhong, T. Research on recognition method of electrical components based on YOLO V3. *IEEE Access* **2019**, *7*, 157818–157829. [CrossRef]
34. Chen, B.; Miao, X. Distribution line pole detection and counting based on YOLO using UAV inspection line video. *J. Electr. Eng. Technol.* **2020**, *15*, 441–448. [CrossRef]
35. Renwei, T.; Zhongjie, Z.; Yongqiang, B.; Ming, G.; Zhifeng, G. Key parts of transmission line detection using improved YOLO v3. *Int. Arab J. Inf. Technol.* **2021**, *18*, 747–754. [CrossRef]
36. Zhan, W. Electric Equipment Inspection on High Voltage Transmission Line Via Mobile Net-SSD. *CONVERTER* **2021**, *2021*, 527–540.
37. Bao, W.; Du, X.; Wang, N.; Yuan, M.; Yang, X. A Defect Detection Method Based on BC-YOLO for Transmission Line Components in UAV Remote Sensing Images. *Remote Sens.* **2022**, *14*, 5176. [CrossRef]
38. Jocher, G.; Chaurasia, A.; Stoken, A.; Borovec, J.; Kwon, Y.; Michael, K.; Fang, J.; Yifu, Z.; Wong, C.; Montes, D. Ultralytics/yolov5: v7. 0-yolov5 Sota Realtime Instance Segmentation. Zenodo. 2022. Available online: https://zenodo.org/record/7347926 (accessed on 4 July 2023).
39. Jocher, G.; Stoken, A.; Chaurasia, A.; Borovec, J.; Kwon, Y.; Michael, K.; Changyu, L.; Fang, J.; Skalski, P.; Hogan, A. ultralytics/yolov5: v6. 0-Yolov5n'nano'models, Roboflow Integration, TensorFlow Export, OpenCV DNN support. Zenodo. 2021. Available online: https://zenodo.org/record/5563715 (accessed on 4 July 2023).
40. Yang, L.; Zhang, R.-Y.; Li, L.; Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Virtual, 18–24 July 2021; pp. 11863–11874.
41. Tong, Z.; Chen, Y.; Xu, Z.; Yu, R. Wise-IoU: Bounding Box Regression Loss with Dynamic Focusing Mechanism. *arXiv* **2023**, arXiv:2301.10051.
42. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [CrossRef]
43. Zheng, J.; Lu, C.; Hao, C.; Chen, D.; Guo, D. Improving the generalization ability of deep neural networks for cross-domain visual recognition. *IEEE Trans. Cogn. Dev. Syst.* **2020**, *13*, 607–620. [CrossRef]
44. Hao, C.; Chen, D. Software/hardware co-design for multi-modal multi-task learning in autonomous systems. In Proceedings of the 2021 IEEE 3rd International Conference on Artificial Intelligence Circuits and Systems (AICAS), Washington DC, USA, 6–9 June 2021; pp. 1–5.
45. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
46. Zheng, Z.; Wang, P.; Liu, W.; Li, J.; Ye, R.; Ren, D. Distance-IoU loss: Faster and better learning for bounding box regression. In Proceedings of the AAAI Conference on Artificial Intelligence, Virtual, 22 February–1 March 2020; pp. 12993–13000.
47. Hu, W.; Che, Z.; Liu, N.; Li, M.; Tang, J.; Zhang, C.; Wang, J. Channel Pruning via Class-Aware Trace Ratio Optimization. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**. [CrossRef] [PubMed]
48. Tang, Z.; Luo, L.; Xie, B.; Zhu, Y.; Zhao, R.; Bi, L.; Lu, C. Automatic sparse connectivity learning for neural networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**. [CrossRef] [PubMed]

49. Huang, Q. Weight-quantized squeezenet for resource-constrained robot vacuums for indoor obstacle classification. *AI* **2022**, *3*, 180–193. [CrossRef]
50. Ding, C.; Wang, S.; Liu, N.; Xu, K.; Wang, Y.; Liang, Y. REQ-YOLO: A resource-aware, efficient quantization framework for object detection on FPGAs. In Proceedings of the 2019 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays, Seaside, CA, USA, 24–26 February 2019; pp. 33–42.