



Research article

Optimized YOLOv7-tiny model for smoke detection in power transmission lines

Chen Chen¹, Guowu Yuan^{1,3,*}, Hao Zhou^{1,3}, Yutang Ma² and Yi Ma²

¹ School of Information Science and Engineering, Yunnan University, Kunming 650504, Yunnan, China

² Electric Power Research Institute, Yunnan Power Grid Co., Ltd, Kunming 650217, Yunnan, China

³ Yunnan Key Laboratory of Intelligent Systems and Computing, Kunming 650504, Yunnan, China

* **Correspondence:** Email: gwyuan@ynu.edu.cn; Tel: +8687165033748.

Abstract: Fire incidents near power transmission lines pose significant safety hazards to the regular operation of the power system. Therefore, achieving fast and accurate smoke detection around power transmission lines is crucial. Due to the complexity and variability of smoke scenarios, existing smoke detection models suffer from low detection accuracy and slow detection speed. This paper proposes an improved model for smoke detection in high-voltage power transmission lines based on the improved YOLOv7-tiny. First, we construct a dataset for smoke detection in high-voltage power transmission lines. Due to the limited number of real samples, we employ a particle system to randomly generate smoke and composite it into randomly selected real scenes, effectively expanding the dataset with high quality. Next, we introduce multiple parameter-free attention modules into the YOLOv7-tiny model and replace regular convolutions in the Neck of the model with Spd-Conv (Space-to-depth Conv) to improve detection accuracy and speed. Finally, we utilize the synthesized smoke dataset as the source domain for model transfer learning. We pre-train the improved model and fine-tune it on a dataset consisting of real scenarios. Experimental results demonstrate that the proposed improved YOLOv7-tiny model achieves a 2.61% increase in mean Average Precision (mAP) for smoke detection on power transmission lines compared to the original model. The precision is improved by 2.26%, and the recall is improved by 7.25%. Compared to other object detection models, the smoke detection proposed in this paper achieves high detection accuracy and speed. Our model also improved detection accuracy on the already publicly available wildfire smoke dataset Figlib (Fire Ignition Library).

Keywords: transmission line; smoke detection; attention mechanism; deep learning

1. Introduction

Inspection of high-voltage power transmission lines in the power grid is one of the measures to ensure the safe operation of the transmission lines. However, due to the wide distribution of high-voltage power transmission lines in mountainous areas, manual inspections face challenges such as difficulty, time-consuming nature, high risks and low efficiency. With the development of image acquisition devices and transmission technologies, imaging devices installed on high-voltage power transmission lines, such as surveillance cameras, drones and satellite remote sensing, have been applied to line inspections [1]. Detecting key objects on the transmission corridors using videos and images captured by surveillance cameras, drones and other devices can improve the difficulties faced by manual inspections [2]. However, when detecting the collected images, the accuracy is limited by the observation skills of the inspectors. Additionally, the hidden danger of visual fatigue leads to the inability to observe sudden situations in real-time [3]. Therefore, applying computer vision technology to detect images in power transmission line inspections holds significant research significance.

In 2022, Yunnan Province had a power transmission volume of approximately 140 billion kilowatt-hours from the west to the east, making it a major province for external power transmission in China. However, Yunnan Province is located on a plateau, with mountainous areas accounting for about 94% of its total land area. Most high-voltage power transmission lines are located in these mountainous regions, making efficient inspection of these lines highly necessary. Several events affect the regular operation of transmission lines, such as hill fires, short circuits, equipment aging, bird touches and tree collapses. Among them, the incident with the most significant impact and the broadest range of losses are hill fires. The precursor of a hill fire is smoke, and transmission lines' safety needs to be dealt with after detecting smoke. So, automatic real-time detection of smoke around transmission lines is essential. Over the past 30 years, China has experienced over 13,000 wildfires on average per year, and due to the dry climate, Yunnan Province is particularly prone to wildfires. These wildfires seriously threaten the safety of high-voltage transmission lines that transmit power to other regions. Smoke, as a precursor to fires, can serve as an early warning sign. Suppose smoke can be detected accurately and promptly. In that case, the potential risks of wildfires can be mitigated, significantly reducing the impact of wildfires on the safe operation of the high-voltage power grid.

Researchers proposed fire smoke detection methods using computer vision methods in recent years. Khan et al. used VGG16 (Visual Geometry Group 16) as the backbone network for smoke detection [4]. They also employed an artificial smoke dataset to address the problem of haze recognition and improve the network model's robustness. Additionally, the authors performed pre-training on the weights before training their dataset to mitigate the interference caused by the natural environment. To tackle the issue of significant smoke shape variations in high outdoor wind speeds, Yin et al. proposed a cascade classification and AlexNet-based deep convolutional neural network for smoke detection [5]. This approach improved smoke detection in extreme environments and incorporated a BN (batch normalize) layer to normalize the scattered data in each layer, accelerating model training and alleviating overfitting. Li et al. made improvements to existing convolutional neural networks to reduce the detection difficulty and achieve real-time monitoring. They introduced a new convolutional neural network called SCCNN (Score Clustering Convolutional Neural Networks), which showed good performance for real-time smoke detection [6]. Jiang et al. adopted the lightweight object detection network Efficientdet-D2 [7]. By adding self-attention mechanisms to

the network, they solved the problem of false negatives in smoke detection results caused by insufficient consideration of scene information in actual smoke scenes. Furthermore, they addressed the issue of false alarms caused by similar smoke patterns by successfully fusing multi-level nodes for smoke multi-feature fusion. Li et al. developed a forest fire smoke recognition system based on satellite remote sensing technology by studying artificial neural networks and multi-threshold techniques [8]. Depending on the image size and smoke conditions, they used artificial neural networks (NN) and multi-threshold methods to detect smoke images separately or in combination. K. Muhammad proposed a low-cost fire detection CNN classification network model for monitoring videos [9]. Considering the nature of the target problem and fire data, the model was fine-tuned to obtain a suitable fire detection model for CCTV surveillance systems. Cai et al. refined the residual module with an efficient channel attention module in YOLOv5 [10]. They added DropBlock after each convolutional layer and ultimately proposed a robust and accurate smoke detection model. Zhou et al. proposed an unsupervised domain adaptation smoke detection algorithm based on multi-level feature fusion and collaborative alignment [11]. By coordinating the alignment of features at different scales, they reduced the differences between smoke data in the source and target domains. They also enhanced the representation capability of smoke features by embedding fusion modules at different depths in the Neck. Zhang et al. introduced an improved algorithm called Swin-YOLOv5 based on YOLOv5 for fire and smoke detection in fire accidents [12]. They incorporated the Swin Transformer as the feature extraction layer to enhance the model's capability.

In general, the proposed smoke detection methods mentioned above have shown good detection performance. However, based on current research, there are still several challenges in smoke detection: 1) Detection accuracy of smoke detection: The outdoor monitoring scenes present complex backgrounds, and smoke exhibits significant scale variations, leading to a higher rate of false alarms and missed detections, resulting in lower detection accuracy. 2) Insufficient smoke training data in mountainous areas: Training deep learning models requires a large amount of data, and collecting thousands of smoke images in the mountainous scenes where high-voltage transmission lines are located is challenging. Some studies use artificially synthesized smoke datasets to compensate for this deficiency. However, in these synthetic datasets, some smoke instances may appear unrealistic, and the backgrounds may be too single, deviating from real-world scenarios. 3) Real-time performance of smoke detection algorithms: For certain applications, smoke detection needs to have real-time capabilities. However, current research has shown that some algorithms have high computational complexity and cannot meet the requirements for real-time performance. 4) Lightweight smoke detection algorithms: The size of smoke detection algorithms can affect their deployments in large-scale applications. Larger algorithms consume more computational resources and storage space, which can lead to performance degradation and increased deployment costs.

To further improve the smoke detection accuracy of the high-voltage transmission lines in Yunnan's power grid, this paper constructs a real smoke dataset based on actual transmission line images in Yunnan's power grid. In addition, synthetic smoke vector graphics are incorporated into non-smoke power grid images, creating a synthetic smoke image dataset with complex backgrounds typical of mountainous high-voltage transmission lines. The lightweight object detection model YOLOv7 has demonstrated promising results in speed and accuracy [13]. Taking into account the specific requirements of smoke detection in Yunnan's mountainous areas along high-voltage transmission lines through comparative experiments among different models on the dataset, this

paper proposes an improved smoke detection model based on YOLOv7-tiny.

This paper has the following main contributions:

- 1) This paper constructs a smoke detection dataset for mountainous high-voltage transmission lines. The dataset will be publicly released after desensitization.
- 2) This paper proposes a robust smoke detection model capable of accurately and rapidly detecting multi-scale smoke in complex mountainous backgrounds.

2. Smoke detection dataset construction

2.1. Image dataset and sample labeling

The dataset used in this paper was captured from Yunnan Power Grid Co., Ltd of China Southern Power Grid. The images were captured using surveillance cameras mounted on transmission towers or unmanned aerial vehicles (UAVs) used for transmission line inspections. The obtained real smoke dataset consists of a total of 7990 images, and all sample images were labeled using the LabelImg labeling tool. An example of smoke detection labeling on an image is shown in Figure 1.



Figure 1. Sample labeling of smoke detection.

2.2. Construction of synthetic smoke image dataset

Due to the complex backgrounds, and small and thin smoke in transmission line images, smoke can easily hide within the complex environment, resulting in a challenging training task for algorithms. To ensure the effectiveness of the model training while improving efficiency, we created a synthetic smoke image dataset by integrating virtual smoke into the complex background of transmission lines.

In virtual reality systems, particle systems are commonly used to simulate natural phenomena such as fire, smoke, water flow, clouds, fog and snow, aiming to generate more realistic and dynamic effects [14]. The irregular motion of particle systems fundamentally enhances the realism of the simulation. Simulating smoke using a particle system involves adjusting parameters influencing particle behavior, such as gravity, resistance, obstruction and wind. Particle systems can obtain

smoke with different thicknesses, colors, shapes and motion trajectories. Table 1 shows the parameters and effect of the particle system.

Table 1. The parameters and effect of the particle system.

parameter	effect
Emitter Type	Control the emitter's type, such as point, line, surface, sphere, box, etc.
Emitter Size	Control the emitter's size in the X, Y and Z directions.
Particle Type	Control the particles' type, which can be points, lines, faces, spheres, cubes, droplets, smoke, etc.
Particle Size	Control the size of each particle.
Particle Life	Control the duration of each particle's existence. Longer lifetimes allow particles to remain visible for a longer time.
Particle Birth Rate	Control the number of particles generated per second.
Physics	Physical parameters, such as particle velocity, gravity, and air resistance are controlled to control the trajectory of the particle.
Turbulence	Control the irregular movement of particles to simulate the natural movement effect of smoke.
Render Settings	Control rendering parameters, such as smoke color and transparency.

Since images generated by particle systems are dynamic, we saved them as simulated smoke videos. We extracted individual frames from the smoke videos to save as static images. In the end, we obtained 127 pure smoke images, each exhibiting a different form due to the randomness of the particle system. Subsequently, random selections of pure smoke images and background images from the pure smoke dataset and the background dataset of the power grid were linearly blended to construct a virtual smoke dataset with the power grid as the background.



Figure 2. Virtual smoke generated by particle system.

Figure 2 illustrates four samples of pure smoke images generated by a particle system. Each sample is an 800×800 RGBA image consisting of four channels: three RGB color channels (\mathcal{S}) and one alpha channel (α). α is a separate channel in a RGBA image that controls the transparency of each pixel in the image. The synthetic smoke image B_{image} , obtained by linearly combining the pure smoke image (\mathcal{S}) with the background image of the power grid (B), can be represented as follows:

$$B_{image} = \alpha(\theta)S(\gamma * x, \gamma * y) + B \quad (1)$$

where, x and y represent the placement coordinates of the pure smoke image, γ represents the scale

factor for resizing the pure smoke image, $x \in [0, B.width - S.width]$, $y \in [0, B.height - S.height]$, and $\gamma \in [0.0, 1.0]$. When γ is 1.0, the smoke image is the original size. $\alpha(\theta)$ represents the transparency of the smoke image, and based on smoke characteristics, it is generally chosen within the range of $[0.7, 1]$, indicating high transparency. When θ is 1, the smoke image is completely opaque.

The smoke image S is subjected to size transformation by multiplying both the horizontal and vertical coordinates by γ . The resulting image is then multiplied by the transparency α and randomly placed on the background image B , resulting in the synthesized virtual smoke image dataset. Figure 3 compares the synthesized smoke images with the real smoke images.



(a) Synthesized smoke images

(b) Real smoke image

Figure 3. Comparison of the synthesized smoke image with the real smoke image.

Since the smoke generated by the particle system is random and highly similar to real smoke, the synthesized virtual smoke image dataset exhibits high similarity to real smoke. Additionally, variations in smoke scale and transparency and the random selection of background images from the power line scenes contribute to the dataset's diversity. Therefore, the synthesized virtual smoke image dataset resembles real smoke and offers various variations. We generated 6364 images by synthesizing smoke images, significantly expanding our training and testing data. In this paper, the real smoke and synthesized smoke datasets were randomly divided into training, validation and testing sets in a ratio of 2:1:1. The specific distribution is shown in Table 2.

Table 2. Dataset division.

Dataset	Total number	Training set	Validation set	Testing set
Real smoke samples	7990	3995	1997	1998
Synthetic smoke samples	6364	3182	1591	1591

3. Proposed methods

Due to the installation of numerous surveillance cameras on the power grid, a single high-voltage transmission line can have thousands of cameras. The data transmission from these cameras

relies on the 4G mobile communication network, which has limited bandwidth. Moreover, if all the images were transmitted to the data center for processing, the data traffic from the surveillance cameras would be overwhelming. Therefore, this application requires front-end detection of anomalies such as smoke, and only the images with abnormalities are returned. However, the computational hardware resources integrated into the front-end cameras are limited, so lightweight models must be considered.

In July 2022, the original team behind YOLOv4 [15] proposed the YOLOv7 model. As the latest iteration of the YOLO series of object detection algorithms, YOLOv7 combines cutting-edge academic research achievements with the practical needs of engineering. YOLOv7-tiny is the most lightweight model in the YOLOv7 series and employs data augmentation methods similar to YOLOv4 and YOLOv5 [16], such as mosaic and random scaling [17]. In the backbone network, a concatenation of convolutional neural networks is used to extract multi-scale features. The Feature Pyramid Network (FPN) structure [18] enhances multi-scale semantic and spatial information. Batch normalization layers are directly connected to the convolutional layers, allowing the batch normalization mean and variance to be integrated into the biases and weights of the convolutional layers during the inference stage. At the end of the network, the prediction head consists of three prediction layers of different scales, each responsible for detecting objects of large, medium and small sizes.

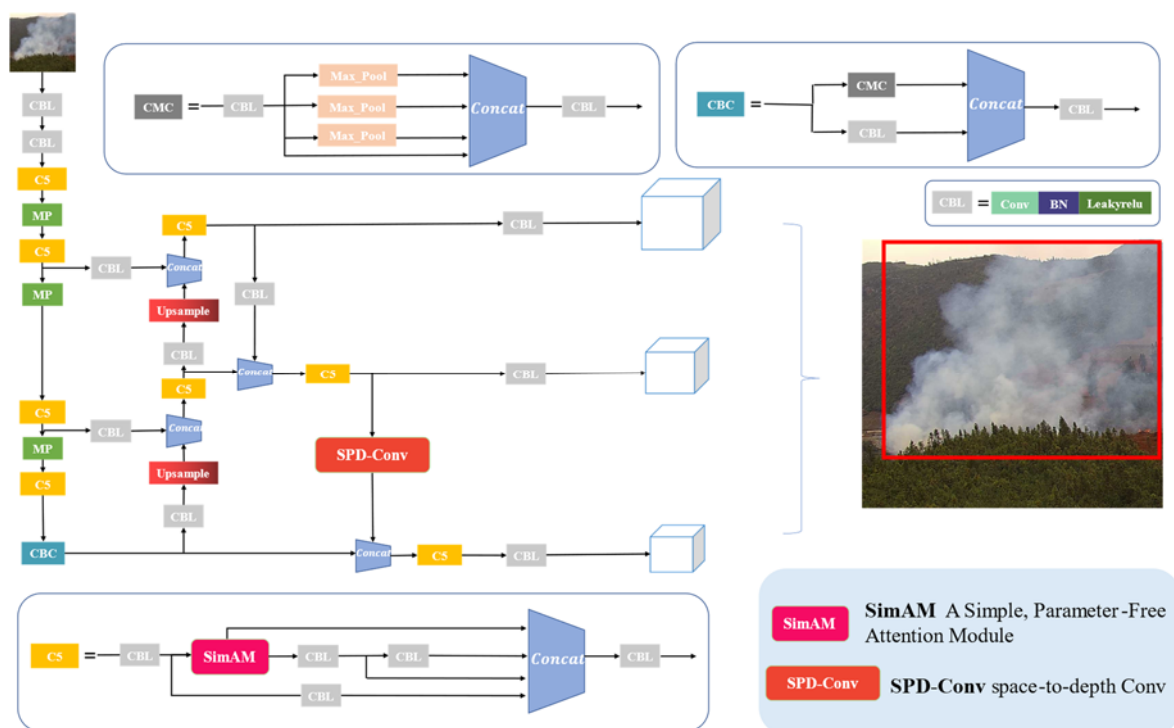


Figure 4. Our improved YOLOv7-tiny model.

To address the smoke detection problem at the front-end of high-voltage transmission line surveillance cameras, we propose a smoke detection method based on the lightweight YOLOv7-tiny model, which ensures implementation on hardware with limited computational capabilities. First, we add multiple parameter-free attention mechanisms [19] to the backbone of the lightweight object

detection model YOLOv7-tiny to enhance the model's perception of features. Then, we replace regular convolutions in the detection head with Spd-Conv (Space-to-depth Conv) [20] to improve the model's accuracy in detecting blurry images. Next, we employ a transfer learning strategy for the improved model. We use a synthetic virtual smoke dataset with the mountains and forests background as the source domain for model pretraining. Finally, we trained the model using a real dataset. The framework of our smoke detection model is illustrated in Figure 4.

According to application requirements, this chapter introduces the improvements for YOLOv7-tiny. Section 3.1 introduces the parameter-free attention mechanism module SimAM; Section 3.2 presents SPD-Conv for low-resolution images and small objects; Section 3.3 introduces the transfer supervised learning used in this paper.

3.1. SimAM

Inserting attention modules into convolutional neural networks has effectively enhanced the network's rich representation capabilities [21]. Considering the diverse and intricate morphology of smoke images, their similarity to fog and haze, and the complex backgrounds in transmission line channels, this paper introduces multiple parameter-free attention modules in the C5 module part of the original YOLOv7-tiny model. The modules enhance the model's perception of target features, improve their saliency and maintain detection speed. The SimAM parameter-free attention mechanisms were chosen because they enhance the model's attention capability without introducing additional parameters. The attention mechanism can focus on important features without increasing computational complexity or sacrificing efficiency. With YOLOv7-tiny, as a lightweight model, time-for-space problems may occur if using a parametric attention mechanism. The non-parametric attention mechanism is undoubtedly the best choice when prioritizing the time factor. By incorporating SimAM, the YOLOv7-tiny model can better understand and emphasize relevant smoke details without compromising the detection speed.

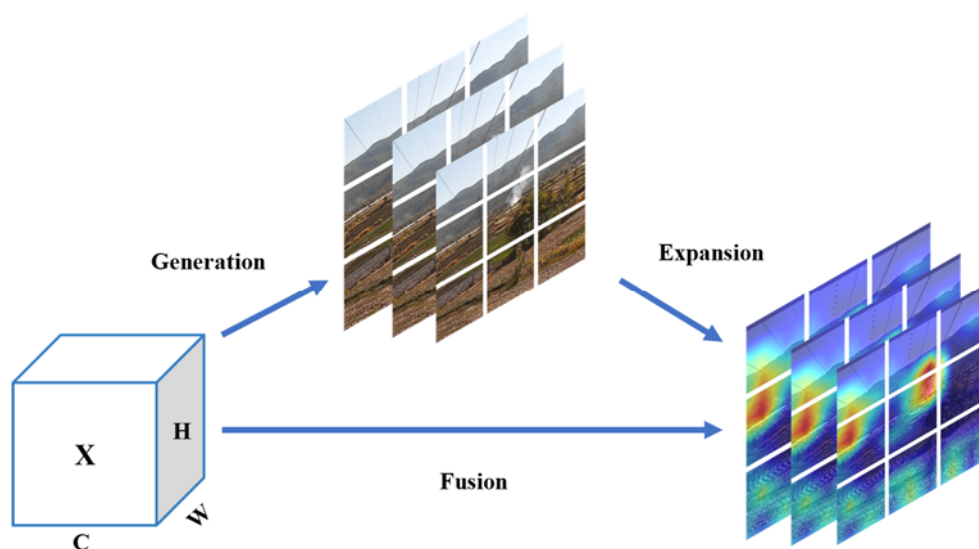
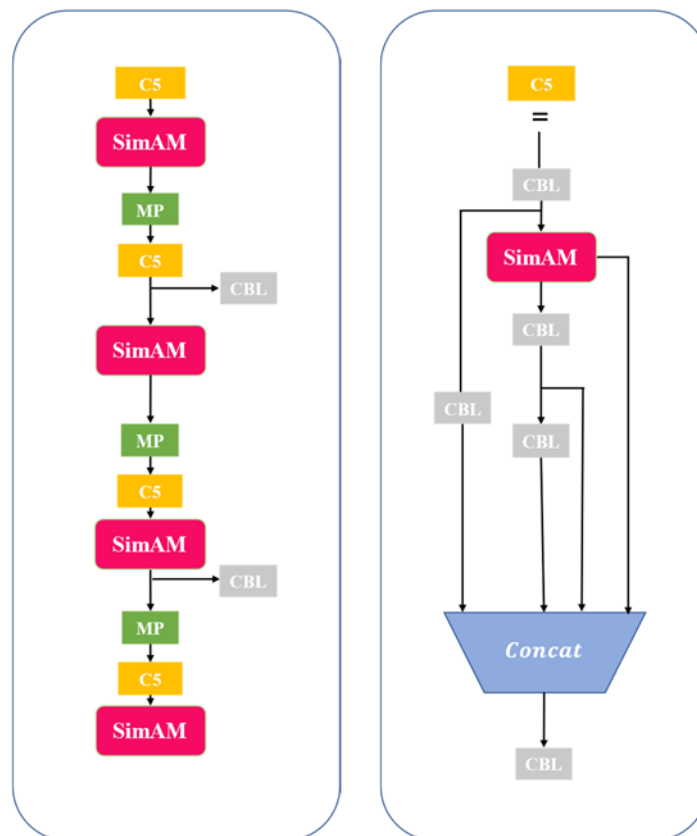


Figure 5. SimAM structure.

The principle of SimAM is a novel parameter-free attention module that draws inspiration from neural science theories and is based on the theory of spatial suppression [22]. It constructs an energy function to explore the importance of neurons and derives attention weights based on it. Rather than only having different attention styles for different channels or different positions, SimAM employs 3D attention, treating each position differently. Figure 5 illustrates its structure.



(a) SimAM(before the MP) (b) SimAM(in the C5)

Figure 6. Two ways to introduce SimAm.

Figure 6 presents two ways to introduce the SimAM attention module. Figure 6(a) illustrates the addition of the attention mechanism in the feature extraction part of the model, and Figure 6(b) shows the introduction of the attention mechanism in the C5 module of the model. Table 3 in Section 4.3.2 of this paper indicates that adding the attention mechanism in the C5 module is more effective for smoke detection. Because the C5 module spans the entire model, we believe it can enhance the model's detection performance by integrating SimAM in C5. SimAM is effective for the feature extraction part and the whole detection process.

3.2. SPD-Conv

Traditional convolutional neural networks (CNNs) often experience performance degradation when dealing with low-resolution images or small objects. This is generally attributed to using

convolutional strides or pooling layers in CNNs, which result in the loss of fine-grained information and inefficient learning of feature representations. To address this issue, we introduce a new CNN module called SPD-Conv into the neck of the YOLOv7-tiny model. By replacing the convolutional layers with SPD-Conv building blocks, the high resolution of the feature map can be maintained to avoid information loss while maintaining the same parameter size level, thus significantly improving the detection accuracy of the YOLOv7-tiny model. It satisfies the lightweight goal we pursue.

The SPD-Conv CNN module replaces the pooling and strided convolution layers commonly found in traditional CNNs with a space-to-depth (SPD) layer and a non-strided convolution layer. The SPD layer performs downsampling on the feature maps throughout the network while preserving all information in the channel dimension. Following each SPD layer, a non-strided convolution layer is added to reduce the number of channels using learnable parameters introduced in the additional convolution layer. The SPD-Conv module performs well in tasks involving low-resolution images and detecting tiny objects, significantly reducing information loss. Figure 7 illustrates the structure of the SPD-Conv module.

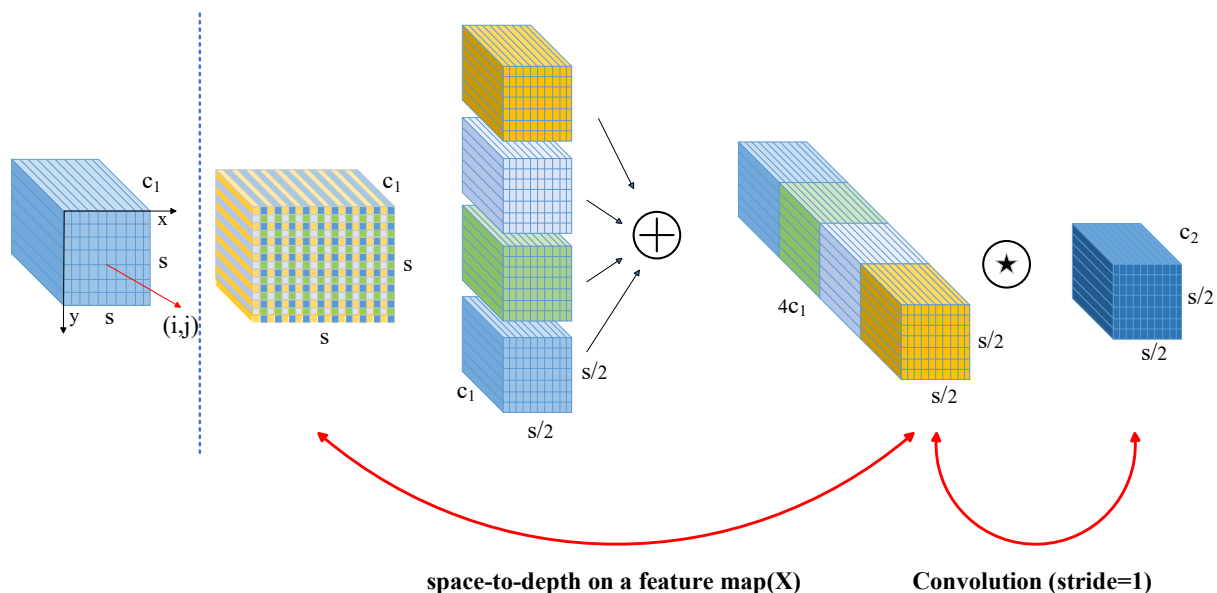


Figure 7. SPD-Conv structure.

In the neck of the original YOLOv7-tiny model, there are two convolutional layers with a stride of 2. To determine the number of convolutions to be replaced by SPD-Conv, we conducted an ablation experiment, and the results can be found in Table 3 in the Experimental section. Figure 8 shows the specific locations of the convolutions with a stride of 2 in the original model. The experimental results in Table 3 demonstrate that replacing the second convolution with SPD-Conv instead of replacing all convolutions achieves the best performance improvement for the model.

We believe this is because, when the images have good resolution, there is a large amount of redundant pixel information. Even with information loss, the model can still learn features effectively. However, when detecting smaller feature maps, more target feature information is lost, making the remaining information more valuable. Therefore, it is crucial to preserve fine-grained information in such cases.

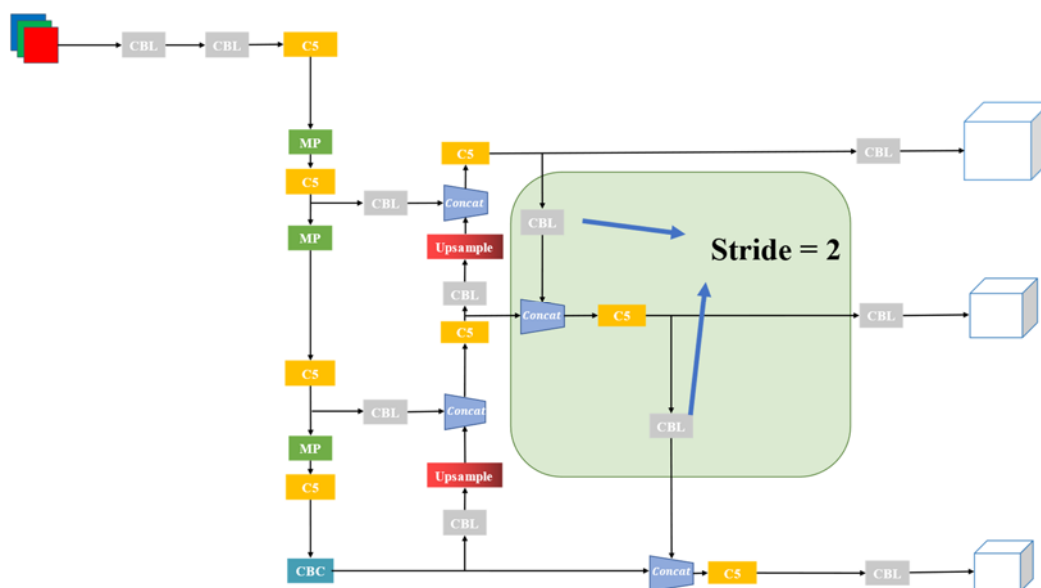


Figure 8. Convolution with a stride of 2.

3.3. Transfer supervised learning

Transfer learning refers to the machine learning method of applying the knowledge and experience learned from one task to another related task. It can transfer the general features learned from a synthetic smoke dataset, such as edges and textures, to real smoke detection by sharing the underlying feature extractor. This reduces training time and computational resource consumption in smoke detection tasks while improving the performance and generalization ability of the smoke detection model. Due to the availability of labeling information while generating synthetic datasets with added smoke, we chose to employ transfer supervised learning. Figure 9 illustrates the workflow of transfer learning. The synthetic dataset is the source domain for model transfer learning, with real power grid images as the background. This approach reduces the gap between the source and target domains, thereby directly improving the detection performance of real smoke. Through comparative experiments with real datasets, we observed significant enhancements in smoke targets' detection performance and robustness using this synthetic dataset.

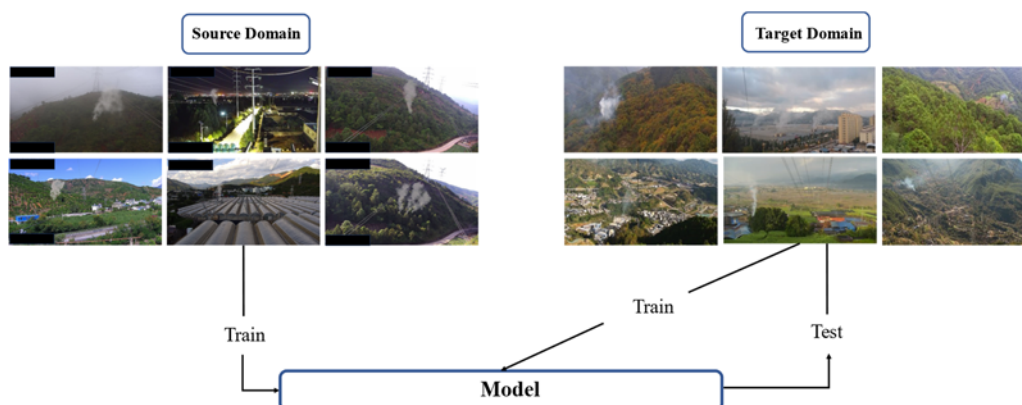


Figure 9. Transfer supervised learning process.

4. Experimental analysis and discussion

4.1. Experimental setup

Our hardware configuration for training and testing algorithms consists of an AMD Ryzen 5 3600x CPU, 32 GB memory and NVIDIA GeForce RTX 2080Ti GPU. The software environment includes the Windows 10 operating system, PyTorch 1.9.0 deep learning framework, CUDA version 11.3 and Python version 3.6. The pre-training weights are the official YOLOv7-tiny weights. The training was conducted for 300 epochs, with a batch size of 16, adapted to the available GPU memory. The initial learning rate was set to 0.01, and the SGD momentum was set to 0.937.

4.2. Model evaluation metrics

The evaluation metrics used in the experiments include precision (P), recall (R), mean average precision (mAP) and $F1$ score. Precision (P) represents the ratio of correctly predicted positive samples to the total number of predicted positive samples. Recall (R) represents the ratio of correctly predicted positive samples to the total number of actual positive samples. The $F1$ score is the harmonic mean of precision and recall, providing a balanced evaluation metric that considers both precision and recall to avoid overemphasizing either one. The formulas for calculating detection precision (P), recall (R), mean average precision (mAP) and $F1$ score are as follows:

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 p(r) dr \quad (4)$$

$$mAP = \frac{\sum_{i=1}^k AP_i}{k} \quad (5)$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (6)$$

where TP represents the cases where positive samples are correctly predicted as positive, FP represents the cases where negative samples are incorrectly predicted as positive, and FN represents the cases where positive samples are incorrectly predicted as negative, i.e., the cases of false negatives or missed detections.

4.3. Experimental results and analysis

4.3.1. Experiments on selecting attention mechanisms

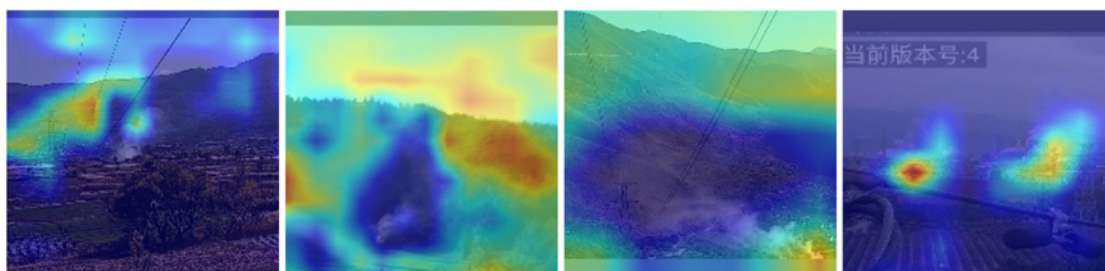
To validate the effectiveness of the attention modules used, we attempted to add the SE [23], ECA [24], CBAM [25] and SimAM attention modules. We compared the performance of the YOLOv7-tiny model after adding different attention mechanisms. Table 2 shows the comparison results after incorporating various attention mechanisms.

Table 3. Comparative experiment on attention mechanism.

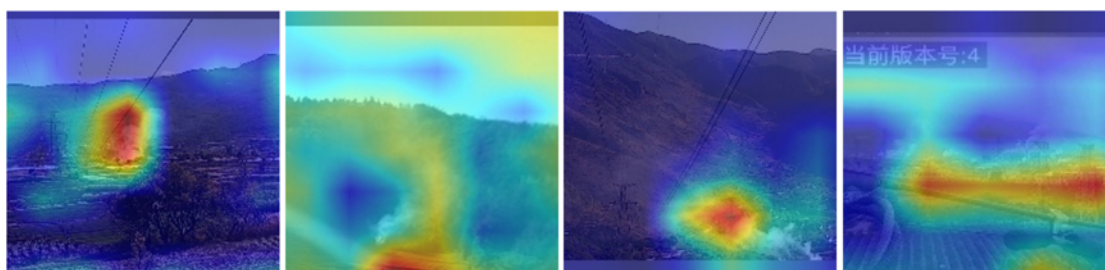
Model	Precision/%	mAP@0.5/%	Recall/%	FPS
YOLOv7-tiny	80.94	76.86	70.92	78.13
+ SE	76.8	72.8	70.8	84.74
+ CBAM	82.1	76.1	74.4	66.66
+ ECA	76.2	73.7	71.9	83.3
+ SimAM	81.4	75.2	73.2	86.95



(a) Detected original image



(b) Attention visualization of YOLOv7-tiny model



(c) Attention visualization of our improved YOLOv7-tiny model

Figure 10. Comparison of attention visualization between the original model and our improved model.

Upon summarizing the results in Table 3, it is observed that adding attention modules has varying degrees of impact on the detection performance of the model. However, considering all evaluation metrics, the SimAM attention mechanism is the most suitable attention module for balancing detection accuracy and speed on our dataset.

Figure 10 shows the attention visualization results of the SimAM module compared to the original model on the smoke test dataset. Figure 10(a) shows the original input image, Figure 10(b) displays the attention visualization results of the original model on the detected objects, and Figure 10(c) demonstrates the improved YOLOv7-tiny algorithm with attention visualization results. A darker color indicates greater attention from a model. Figure 10 shows that the improved YOLOv7-tiny model in this paper enhances the saliency of smoke objects, thereby improving the detection accuracy of smoke in power transmission lines.

4.3.2. Ablation experiments

We made a series of ablation experiments to analyze the importance and necessity of each improvement. The results of the ablation experiments demonstrate that the SimAM attention mechanism and SPD-Conv play crucial roles in smoke detection in power transmission lines, and the transfer learning based on synthetic datasets significantly improves the accuracy and robustness of the smoke detection algorithm.

The results of our ablation experiments are presented in Table 4. Based on the results, we can observe that adding the attention mechanism in two different ways in the original model leads to improvements in the F1 score. Specifically, incorporating the attention mechanism in the C5 module of the original model enhances precision and recall. Using SPD-Conv alone in the model does not result in any improvement in accuracy. However, when combined with the attention mechanism, the F1 score of the model increases from 75.60 to 77.57 and 75.75. This highlights the importance of combining SPD-Conv with the attention mechanism for accuracy improvement. Furthermore, after performing transfer learning, the F1 score of the model improves by 5.01, indicating that our improvements are effective and result in the best detection performance of the model.

Table 4. Results of ablation experiment.

SimAM ¹	SimAM ²	SPD-Convs ³	SPD-Conv	Transfer Learning	Precision/%	mAP@0.5/%	Recall/%	F1
					80.94	76.86	70.92	75.60
√					81.4	75.2	73.2	77.08
	√				80.9	76.7	79.9	80.40
		√			77.86	74.78	69.05	73.19
			√		79.73	76.18	71.4	75.34
√		√			76	72.76	72.8	74.37
√			√		82	78.66	73.6	77.57
	√	√			77.2	73.4	71.7	74.35
	√		√		80.8	75.5	71.3	75.75
√		√		√	81	75.7	71.9	76.18
	√	√		√	78.33	74.49	73.8	76.00
	√		√	√	81.66	78.18	75.66	78.55
√			√	√	83.2	79.47	78.17	80.61

Note: ¹Using SimAM in the C5 module. ²Using SimAM before the MP module. ³Replace all convolutions with stride 2 using SPD-Conv.

4.3.3. Model training comparison

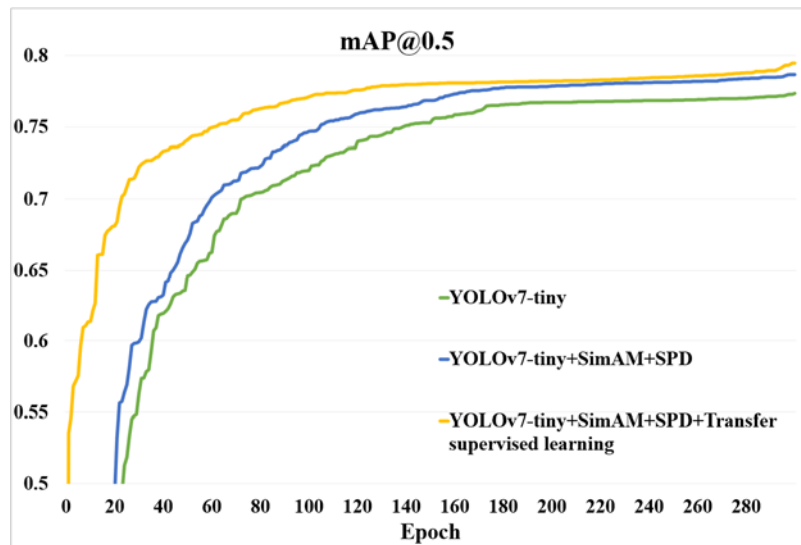
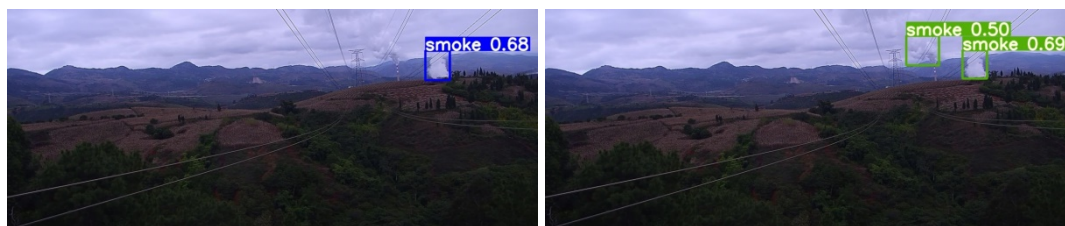


Figure 11. Comparison curve for $mAP@0.5$.

The comparison of mAP curves between the improved YOLOv7-tiny model and the original YOLOv7-tiny model is shown in Figure 11. From Figure 11, we can observe that, even before performing transfer learning, the improved YOLOv7-tiny model already achieves a higher $mAP@0.5$ compared to the original YOLOv7-tiny model. After transfer learning, the detection performance of the model is significantly enhanced. The $mAP@0.5$ value of the model stabilizes around 0.7947 after 300 iterations, while the original YOLOv7-tiny model stabilizes around 0.7686. Therefore, compared to the original model, the improved YOLOv7-tiny method achieves a 2.61% increase in mAP for smoke detection in power transmission lines. Here, $mAP@0.5$ represents the mAP when the IoU threshold is set to 0.5.

4.3.4. Smoke detection results

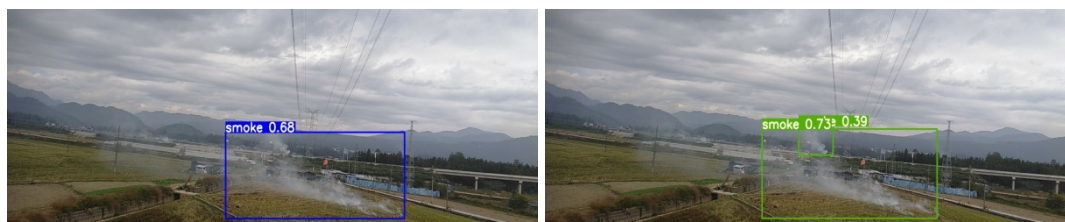
To visually compare the detection performance of the algorithm before and after improvement, we provide detection results in Figure 12. The left figure shows the results of the original YOLOv7-tiny model, while the right figure shows the results of the improved YOLOv7-tiny model proposed in this paper. In Figure 12(a), the original YOLOv7-tiny model exhibits instances of missed detection, while the improved YOLOv7-tiny model accurately detects the smoke targets. In Figure 12(b), the original YOLOv7-tiny model produces false positives under the influence of lighting conditions. However, the improved YOLOv7-tiny model avoids false detections and provides more accurate bounding boxes. In Figure 12(c), the original YOLOv7-tiny model experiences missed detection when other objects partially occlude smoke objects. On the other hand, the improved smoke detection model in this paper accurately detects all smoke targets. These visual comparisons demonstrate the effectiveness of the proposed improvements in enhancing the detection accuracy and robustness of the YOLOv7-tiny model for smoke detection.



(a) Detection result 1



(b) Detection result 2



(c) Detection result 3

Figure 12. Comparison of model detection results.

4.3.5. Comparison experiment with other detection models

To objectively evaluate the advantages of our improved model, we compared the improved YOLOv7-tiny model with other detection models, including Fast R-CNN, Faster R-CNN, YOLOx-nano, YOLOv5s, YOLOv5n, nanodet-plus and the original YOLOv7-tiny model. All object detection models were trained using the same dataset, and the training process utilized identical parameters. The results obtained from the comparison models are presented in Table 5.

Table 5. Comparison of different model detection experiments.

Model	Precision/%	mAP@0.5/%	Recall/%	F1	FPS
Fast R-CNN	71.91	68.16	64.31	67.18	8.91
Faster R-CNN	71.23	67.85	64.06	67.04	9.03
Yolox-nano	78.51	73.61	68.17	72.98	62.01
Yolov5s	78.9	75.5	71.3	74.91	58.13
Yolov5n	78.8	74.7	72.3	75.41	68.96
Nanodet-plus	72.53	69.8	65.62	68.90	83.54
Yolov7-tiny	80.94	76.86	70.92	75.60	78.13
Ours	83.2	79.47	78.17	80.61	76.63

The four models compared in this study are lightweight and have small model sizes, making them suitable for real-time object detection in resource-constrained environments. However, their performance decreases when faced with complex tasks such as smoke detection. The detection results of the comparison models in Table 5 show that the improved YOLOv7-tiny model in this paper achieves higher mAP@0.5 and F1 scores compared to the other models. Although its FPS is significantly lower than nanodet-plus and is disadvantaged compared to the original YOLOv7-tiny model, its detection speed is noticeably better than the other three models. The speed of the one-stage model is significantly faster than that of the single-stage model, and the two-stage model is difficult to meet real-time requirements. Overall, the experimental results demonstrate the trade-off between detection speed and accuracy achieved by our improved model.

4.3.6. Experiments on Figlib dataset

In addition to using our smoke dataset for transmission lines, we also conducted experiments on the Figlib dataset provided in [26]. The Figlib dataset addresses the need for a large-scale, publicly available labeled dataset for wildfire smoke detection. It consists of a sequence of wildfire images captured by a fixed-angle camera on remote mountaintops in Southern California through the HPWREN (High-Performance Wireless Research and Education Network). The dataset contains 24,800 high-resolution images with sizes of either 1536×2048 or 2048×3072 pixels. After removing images without smoke, the dataset consists of 2521 wildfire smoke images.

During the experiment, the smoke image data was randomly divided, with 50% of the images used as the training set, 25% as the validation set, and the remaining 25% as the test set.

We compared the results of our improved model with the experiments conducted in [26] and [27] on the Figlib dataset. The comparison results are shown in Table 6. Since [27] did not provide specific data for Precision and Recall, only the F1 value is displayed in the table. It can be observed that, compared to [26] and [27], our method achieves slightly better accuracy and recall in smoke detection.

Table 6. Experiments on the Figlib dataset.

Method	Precision/%	Recall/%	F1
Reference [26]	90.85	76.11	82.83
Reference [27]	-	-	89
Ours	93.61	90.45	92

5. Conclusions

We have established a smoke detection dataset for high-voltage transmission lines in mountainous areas, using monitoring images provided by Yunnan Power Grid. Due to the limited number of samples in the dataset, we employed a particle system to generate smoke. We synthesized the generated smoke into randomly selected real background images of mountainous high-voltage transmission lines. This data augmentation method, different from conventional geometric operations such as flipping, rotation, scale transformation and cropping, produces more realistic synthesized images with richer features, making them more suitable for training deep learning models. The image dataset contains some textual information, which will be anonymized before being publicly released.

Next, based on the actual needs of Yunnan Power Grid for safe operation, we propose an

improved YOLOv7-tiny-based smoke detection model for transmission lines, aiming to achieve fast and highly accurate detection of smoke, thereby balancing detection accuracy and speed. This model plays a preventive role in detecting fires on transmission lines. Through comparative experiments, the constructed network model achieves an accuracy rate of 83.2%, a recall rate of 78.17% and a mAP of 79.47%.

However, due to the limited number of flame samples in the real transmission line dataset, this paper only focused on smoke detection. It did not address flame detection, which makes monitoring transmission line safety not comprehensive enough. Additionally, since the monitoring cameras cover a large area, the smoke is too tiny or located too far in the images. YOLOv7-tiny has a smaller network structure and fewer parameters than YOLOv7 or other larger models, so it has relatively low detection accuracy. This tiny and fuzzy smoke may be missed. Future research will focus on improving the feature extraction part to enhance the accuracy of detecting wildfires near transmission lines from these two aspects.

Availability of data

The dataset used in the paper was obtained from Yunnan Limited Company of China Southern Power Grid, which is not publicly available due to privacy restrictions.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

This research was funded by the Key R&D Projects of Yunnan Province (Grant No. 202202AD080004), the Graduate Supervisor Team of Yunnan Province, the China Natural Science Foundation of China (Grant No. 62061049, 12263008), the Yunnan Provincial Department of Science and Technology-Yunnan University Joint Special Project for Double-Class Construction (Grant No. 202201BF070001-005) and the Application and Foundation Project of the Yunnan Province (Grant No. 202001BB050032).

Conflict of interest

The authors declare there is no conflict of interest.

References

1. Z. B. Zhao, Z. G. Jiang, Y. X. Li, Y. C. Qi, Y. J. Zhai, W. Q. Zhao, et al., Overview of visual defect detection of transmission line components, *J. Image Graphics*, **26** (2021), 2545–2560. <https://doi.org/10.11834/jig.200689>
2. Y. Sui, P. F. Ning, P. J. Niu, C. Y. Wang, D. Zhao, W. L. Zhang, et al, Review on mounted UAV for transmission line inspection, *Power Syst. Technol.*, **45** (2021), 3636–3648. <http://doi.org/10.13335/j.1000-3673.pst.2020.1178>

3. Z. Y. Liu, X. R. Miu, J. Chen, H. Jiang, Review of visible image intelligent processing for transmission line inspection, *Power Syst. Technol.*, **44** (2020), 1057–1069. <http://doi.org/10.13335/j.1000-3673.pst.2019.0349>
4. S. Khan, K. Muhammad, S. Mumtaz, S. W. Baik, V. H. C. Albuquerque, Energy-efficient deep CNN for smoke detection in foggy IoT environment, *IEEE Internet Things J.*, **6** (2019), 9237–9245. <http://doi.org/10.1109/JIOT.2019.2896120>
5. H. Yin, Y. R. Wei, An improved algorithm based on convolutional neural network for smoke detection, in *2019 IEEE International Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNS)*, IEEE, (2019), 207–211. <http://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00063>
6. C. H. Li, B. Yang, H. Ding, H. L. Shi, X. P. Jiang, J. Sun, Real-time video-based smoke detection with high accuracy and efficiency, *Fire Saf. J.*, **117** (2020), 103184. <http://doi.org/10.1016/j.firesaf.2020.103184>
7. M. H. Jiang, Y. X. Zhao, F. Yu, C. L. Zhou, T. Peng, A self-attention network for smoke detection, *Fire Saf. J.*, **129** (2022), 103547. <http://doi.org/10.1016/j.firesaf.2022.103547>
8. Z. Q. Li, A. Khananian, R. H. Fraser, J. Cihlar, Automatic detection of fire smoke using artificial neural networks and threshold approaches applied to AVHRR imagery, *IEEE Trans. Geosci. Remote Sens.*, **39** (2001), 1859–1870. <http://doi.org/10.1109/36.951076>
9. K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, S. W. Baik, Convolutional neural networks based fire detection in surveillance videos, *IEEE Access*, **6** (2018), 18174–18183. <http://10.1109/ACCESS.2018.2812835>
10. W. B. Cai, C. Y. Wang, H. Huang, T. Z. Wang, A real-time smoke detection model based on YOLO-smoke algorithm, in *2020 Cross Strait Radio Science & Wireless Technology Conference (CSRSWTC)*, IEEE, (2020), 1–3. <http://10.1109/CSRSWTC50769.2020.9372453>
11. F. R. Zhou, G. Wen, Y. Ma, Y. F. Wang, Y. T. Ma, G. F. Wang, et al., Multilevel feature cooperative alignment and fusion for unsupervised domain adaptation smoke detection, *Front. Phys.*, **11** (2023), 81. <https://doi.org/10.3389/fphy.2023.1136021>
12. S. G. Zhang, F. Zhang, Y. Ding, Y. Li, Swin-YOLOv5: Research and application of fire and smoke detection algorithm based on YOLOv5, *Comput. Intell. Neurosci.*, **2022** (2022). <https://doi.org/10.1155/2022/6081680>
13. C. Y. Wang, A. Bochkovskiy, H. Y. M. Liao, YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2023), 7464–7475. <https://doi.org/10.48550/arXiv.2207.02696>
14. Y. C. Zhou, L. H. Fang, X. Y. Zheng, X. L. Chen, Virtual battlefield smoke effect simulation based on particle system, *Comput. Simul.*, **32** (2015), 417–420. <https://doi.org/10.3969/j.issn.1006-9348.2015.07.093>
15. A. Bochkovskiy, C. Y. Wang, H. Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, preprint, arXiv:2004.10934. <https://doi.org/10.48550/arXiv.2004.10934>
16. G. Jocher, A. Stoken, J. Borovec, L. Changyu, A. Hogan, L. Diaconu, et al., ultralytics/yolov5: v3.0, Zenodo, 2020. Available from: <https://ui.adsabs.harvard.edu/abs/2020zndo...3983579J/abstract>.

17. Y. Liu, X. Wang, Ship detection based on improved YOLOv7-tiny, in *2022 IEEE 8th International Conference on Computer and Communications (ICCC)*, IEEE, (2022), 2166–2170. <https://doi.org/10.1109/ICCC56324.2022.10065775>
18. T. Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, S. Belongie, Feature pyramid networks for object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 2117–2125. <https://doi.org/10.1109/CVPR.2017.106>
19. L. Yang, R. Y. Zhang, L. Li, X. Xie, Simam: A simple, parameter-free attention module for convolutional neural networks, in *International Conference on Machine Learning*, (2021), 11863–11874.
20. R. Sunkara, T. Luo, No more strided convolutions or pooling: A new CNN building block for low-resolution images and small objects, in *Machine Learning and Knowledge Discovery in Databases*, Springer Nature, Cham, Switzerland, (2023), 443–459. https://doi.org/10.1007/978-3-031-26409-2_27
21. Q. Tian, R. Hu, Z. Li, Y. Cai, Z. Yu, Insulator detection based on se-YOLOv5s, *Chin. J. Intell. Sci. Technol.*, **3** (2021), 312–321. <https://doi.org/10.11959/j.issn.2096-6652.202132>
22. B. S. Webb, N. T. Dhruv, S. G. Solomon, C. Tailby, P. Lennie, Early and late mechanisms of surround suppression in striate cortex of macaque, *J. Neurosci.*, **25** (2005), 11666–11675. <https://doi.org/10.1523/JNEUROSCI.3414-05.2005>
23. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 7132–7141. <https://doi.org/10.1109/TPAMI.2019.2913372>
24. Q. L. Wang, B. G. Wu, P. F. Zhu, P. H. Li, W. M. Zuo, Q. H. Hu, ECA-Net: Efficient channel attention for deep convolutional neural networks, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2020), 11531–11539. <https://doi.org/10.1109/CVPR42600.2020.01155>
25. S. Woo, J. Park, J. Y. Lee, I. S. Kweon, CBAM: convolutional block attention module, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 3–19. <https://doi.org/10.48550/arXiv.1807.06521>
26. A. Dewangan, Y. Pande, H. W. Braun, F. Vernon, I. Perez, I. Altintas, et al., Figlib & smokeynet: Dataset and deep learning model for real-time wildland fire smoke detection, *Remote Sens.*, **14** (2022), 1007. <https://doi.org/10.3390/rs14041007>
27. K. Govil, M. L. Welch, J. T. Ball, C. R. Pennypacker, Preliminary results from a wildfire detection system using deep learning on remote camera images, *Remote Sens.*, **12** (2020), 166. <https://doi.org/10.3390/rs12010166>



AIMS Press

©2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)