International Journal of Power and Energy Systems, Vol. 46, No. 10, 2026
An Open Access Paper
Special Issue
Next Generation Energy Systems

REAL TIME PREDICTION OF WILDFIRE SPREAD IN TRANSMISSION LINES BASED ON YOLO-LNET

Guozhu Yang,* Wei Du,* Huamin Sun,* Shirui Sun,* Pengchao Sun,*

Abstract

Traditional methods have a limited monitoring range and delayed response in detecting wildfires on transmission lines, and existing image recognition methods are mostly limited to static fire source identification and lack dynamic prediction capabilities. A real-time prediction model based on a lightweight YOLOv5 detection module and a Bi-LSTM prediction module is introduced in this study. In terms of object detection, coordinate convolution and the Ghost module are introduced to improve the backbone network and neck structure of YOLOv5. Focus loss function and complete intersection-to-union ratio loss function are used to improve classification loss and achieve model lightweighting. In terms of spreading prediction, Bi-LSTM is employed to model the bidirectional dependence of temporal data, such as fire point location and fire area. The results indicate that the improved lightweight YOLOv5 model mAP@0.5 Reaches 74.3%, which is 4.1% higher than traditional YOLOv5s and better than mainstream models such as EfficientDet (68.5%) and MobileViT (70.6%); The computational complexity has been reduced by 48%, and the model size has been reduced to 6.5MB. The complete YOLO LNet prediction model not only has the lowest error in different scenarios, but its prediction accuracy (MAE: 5.23) and speed (43.8ms) are also superior to comparative models such as Seq2Seq and Transformer, demonstrating significant comprehensive advantages. This model provides a high-precision and light-weight solution for real-time detection and dynamic prediction of wildfires in transmission lines, and offers new technical support for disaster prevention and mitigation of transmission lines.

Key Words

Transmission line safety; Real-time detection of wildfires; Prediction of wildfire spread; Lightweight; YOLO-LNet

* State Grid Power Space Technology Co., Ltd, Beijing, 102211, China; e-mail: gzYang2025@outlook.com, weiu@pwedu.uu.me, huaminsun@pwedu.uu.me, shiruisun@pwedu.uu.me, pengchaosun@pwedu.uu.me
Corresponding author: Guozhu Yang

(DOI: 10.2316/J.2026.203-0658)

1. Introduction

The power transmission network is the lifeline of a country's energy, and its lines that pass through complex mountainous and forested areas are highly susceptible to the threat of wildfires. Mountain fires can cause insulator flashover, wire melting, and tower damage, leading to widespread power outages and causing significant economic losses and social impacts. In recent years, global warming has intensified the frequency and intensity of extreme wildfires, and the risk prevention and control of wildfires on transmission lines is severe [1] [2]. The traditional post disaster repair mode is costly, and real-time prediction of the spread of wildfires, achieving pre-disaster warning and active prevention and control, is an urgent need to guarantee the secure operation of the power grid [3] [4]. Traditional wildfire monitoring relies on manual inspections and fixed-point sensors, which have problems like restricted monitoring scope and delayed response. With the development of intelligent inspection technology, image recognition-based wildfire detection has gradually become mainstream [5] [6]. However, existing methods are mostly limited to static fire source identification and lack the ability to dynamically predict the spread of fire, making it difficult to support emergency decision-making [7] [8]. Therefore, building a wildfire spread model that combines real-time detection and dynamic prediction capabilities exerts a profound significance in augmenting the disaster prevention and mitigation level of transmission lines.

The You Only Look Once (YOLO) series model adopts a single-stage architecture, which combines the characteristics of fast inference speed and good multi-scale feature fusion (MSFF). It can efficiently extract target features and quickly complete target localization and recognition. As a result, it enjoys widespread utilization across various object detection (OD) tasks [9] [10]. Lee et al. developed a new YOLO structure with adjustable frame regulation to address the issue of high-end hardware required for real-time OD in YOLO. Multiple experiments showed that this architecture could preserve the high precision and user-friendliness of YOLO while cutting down on ser-

vice latency, thereby achieving real-time detection services [11]. Liu et al. introduced a novel image-adaptive YOLO framework to address the difficulty of OD in images of subpar quality captured in unfavorable weather circumstances. This framework introduces an image processing module possessing differentiability and predicts parameters through a small convolutional neural network (NN) for end-to-end joint training. The experiment showed that this method performed well in foggy and dim-light scenes [12]. To figure out the trouble of multi-sensor fusion in intelligent vehicles, Song et al. proposed an OD method grounded in the fusion of millimeter wave radar and vision. A mapping transformation NN was used to process radar data, and a multi-scale YOLO (MS-YOLO) detection network was constructed. The outcomes revealed that this approach improved detection accuracy while ensuring speed, with a large model mAP of 0.888 and a small model mAP of 0.841 at 65fps [13]. Li et al. introduced a progressive domain adaptive YOLO structure to deal with the issues of domain bias and insufficient annotated data in supervised OD models. By constructing an auxiliary domain and introducing an innovative domain-adaptive YOLO for cross-domain OD endeavors. Experiments showed that this method significantly improved OD performance in multiple domain offset scenarios [14].

Long Short-Term Memory (LSTM) is a better kind of recurrent NN that can effectively solve the problem of gradients disappearing or getting too big in long sequences by bringing in gating mechanisms and memory units. It has the ability to grasp long-span interrelationships and is therefore broadly applied in prediction tasks in various fields [15] [16]. To predict crop yield in advance, Kiran Kumar et al. proposed an advanced LSTM deep learning (DL) model through hyperparameter optimization configuration and fine-tuning. The results indicated that Bidirectional LSTM (Bi-LSTM) networks outperformed unidirectional LSTM and traditional machine learning models in terms of prediction accuracy [17]. He et al. proposed an enhanced Bi-LSTM model to address the issue of high nonlinearity, coupling, and complex dynamic characteristics affecting the prediction performance of traditional industrial soft sensing models. This model integrates distributed nonlinear extension and parallel input. The simulation results revealed that the precision of this model was better than that of other advanced models and had good industrial application prospects [18]. To tackle the problems of great computational difficulty and large forecasting errors in current approaches for river water quality modeling, Khullar et al. raised a Bi-LSTM model for forecasting water quality grounded in DL for the Yamuna River in India. The experiment showed that the predicted values of the model were highly consistent with the true values, which could reveal future trends. It outperformed other approaches in regard to prediction accuracy and error in indicators such as chemical oxygen demand and biochemical oxygen demand [19].

In summary, researchers have conducted extensive research on OD and prediction tasks in various fields and have made certain progress. However, there is still limited research on detecting wildfires on transmission lines, especially in predicting the spread of wildfires. The safety of transmission lines is crucial to people's normal lives. Therefore, a YOLO-LNet model grounded in a lightweight You Only Look Once Version 5 (YOLOv5) detection module and Bi-LSTM prediction module is proposed to achieve accurate wildfire detection and prediction, providing new technical support for disaster prevention and the reduction of transmission lines. The novelty of the research consists in the employment of Coordinate Convolution (CoordConv) and Ghost module to improve the backbone network (Backbone) and neck structure (Neck) parts of YOLOv5, respectively, and the introduction of focus loss function (LF) and complete intersection-to-union ratio (IoU) LF to enhance the classification loss of the model, thereby achieving the goal of model lightweighting.

2. Methods and Materials

2.1 Real-time Wildfire Detection Based on Lightweight YOLOv5

Mountain fire detection is the fundamental component of the entire prediction system, requiring fast and accurate identification of fire points to provide reliable input for subsequent spread prediction. YOLOv5, as a classic model for single-stage OD, performs well in general OD tasks and is broadly used in areas such as security monitoring, autonomous driving, industrial detection, and drone inspection, especially applicable to situations characterized by stringent real-time demands [20] [21]. YOLOv5 includes three elements: Backbone, Neck, and Prediction Head. Among them, Backbone serves the function of extracting fundamental features from input images, employing structures such as CSPDarknet, convolution, residual connections, and other operations to gradually compress image dimensions and extract semantic information. Neck employs Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) as the core to achieve MSFF, interact feature maps (FMs) at various levels, and raise the model's capability to identify targets with varying sizes. Head forecasts the target category and bounding box based on the fused features, and outputs the final detection result through regression and classification branches. However, in the scenario of detecting wildfires on transmission lines, traditional YOLOv5 has obvious shortcomings. On the one hand, transmission line inspection images often have complex backgrounds, while targets such as fire points and smoke have small scales and insignificant features. Backbone's ability to extract and distinguish such subtle features is limited. On the other hand, the proportion of background pixels in wildfire detection is much higher than that of fire point targets, and the problem of class imbalance is prominent. Traditional LFs are difficult to balance positive and negative sample learning, and bounding box localization is susceptible to complex environmental interference, resulting in difficulty in meeting the actual needs of power transmission line inspection in

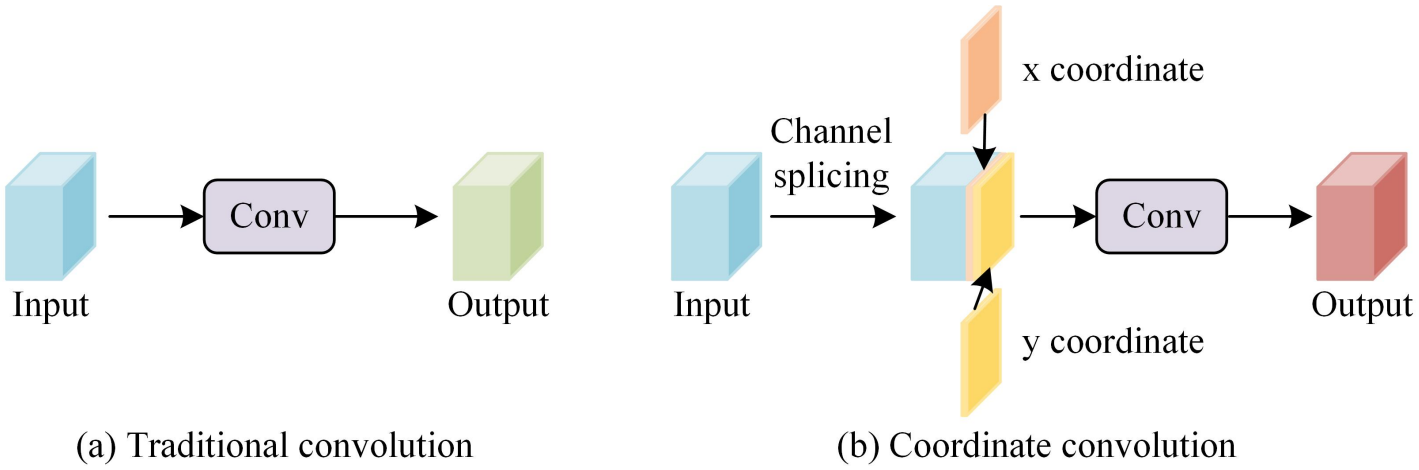


Figure 1. The Principles of Traditional Convolution and Coordinate Convolution

regard to fire point detection accuracy and real-time performance. Therefore, it is necessary to make lightweight improvements to the traditional YOLOv5 model. To enhance Backbone's feature extraction capability for transmission line wildfire targets, a lightweight module called CoordConv was introduced. Compared with ShuffleNet's channel shuffling mechanism, CoordConv can explicitly introduce coordinate information, which is more conducive to the accurate positioning of small targets such as fire points and smoke. The principles of traditional convolution and coordinate convolution are shown in Figure 1.

As shown in Figure 1, traditional convolution operations do not explicitly incorporate coordinate information when extracting features, resulting in insufficient spatial perception of small targets such as fire points, smoke, and irregularly shaped targets. CoordConv adds normalized coordinate channels to the input FM, allowing the convolution kernel (CK) to learn features that contain positional information. Let the original input FM be $X \in R^{H\times W\times C}$ (Height $H = 640$ pixels, width $W = 640$ pixels, number of channels $C = 3$, corresponding to RGB three channels), and the input after expanding the coordinate channel be $X' = [X, x_{coord}, y_{coord}] \in R^{H\times W\times(C+2)}$ The convolution operation process is presented in equation (1).

$$Y'_{i,j} = \sum_{m=0}^{k-1}\sum_{n=0}^{k-1}\sum_{c=0}^{C+1} X'_{i+m,j+n,c} \times W'_{m,n,c} \tag{1}$$

In equation (1), $k = 3 \times 3$ is the volume of the CK, W' is the CK parameters after adapting the extended channel, and Y' is the output FM. To satisfy the real-time needs of wildfire detection in transmission lines, research is conducted on lightweight optimization based on CoordConv combined with depthwise separable convolution (DSC). DSC divides conventional convolution into channel-wise convolution (CWC) and point-wise convolution (PWC). CWC only performs spatial convolution on single-channel features, and its computation is presented in equation (2).

$$Y_{dw,i,j,c} = \sum_{m=0}^{k-1}\sum_{n=0}^{k-1} X'_{i+m,j+n,c} \times W_{dw,m,n,c} \tag{2}$$

In equation (2), Y_{dw} is the output of CWC, and W_{dw} is the kernel of CWC. PWC performs channel fusion on the results of CWC, as presented in equation (3).

$$Y_{pw,i,j,c'} = \sum_{c=0}^{C+1} Y_{dw,i,j,c} \times W_{pw,c,c'} \tag{3}$$

In equation (3), Y_{pw} is the output of PWC, W_{pw} is the kernel of PWC, and c' is the amount of output channels. Substitution of the conventional convolutional layer in the Backbone's CSP module with a lightweight CoordConv architecture enables construction of a spatially-enhanced feature extractor. Through integration of explicit coordinate encoding with DSCs, the network achieves improved spatial localization accuracy for wildfire targets while maintaining computational efficiency via optimized convolution operations. This design meets the real-time detection requirements of transmission line inspection systems. The core function of Neck is multi-scale feature fusion, but the traditional YOLOv5 Neck structure has the problems of feature generation redundancy and high computational cost. The Ghost module generates "ghost" features through linear transformation, which can maintain rich feature representation at lower computational cost. Therefore, the research introduces the Ghost module and constructs rich feature maps through the method of "intrinsic feature generation+ghost feature transformation". Figure 2 presents a visual depiction of the Ghost module's architectural framework.

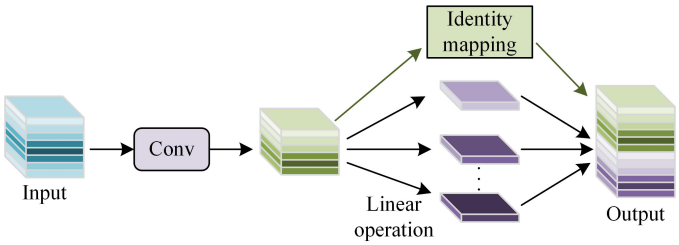


Figure 2. Structure Diagram of Ghost Module

From Figure 2, for the FM $X_{neck} \in R^{H\times W\times C_{in}}$ of

the input Neck, a small amount of conventional convolution is first applied to generate this FM $Y_{intrinsic} \in R^{H\times W\times C_{mid}}(C_{mid} << C_{in})$. Next, for each local FM, deep convolution is used to generate ghost FM Y_{ghost}, which are calculated as shown in equation (4).

$$\begin{cases} Y_{\text{intrinsic}} = \text{Xneck} * \sim \text{Wconv} \\ Y_{\text{ghost, k}} = \text{Yintrinsic,k} * \sim \text{Wdw,k}^{\text{nund}} \end{cases} \tag{4}$$

In equation (4), W_{conv} is the regular CK, $Y_{ghost,k}$ is the kth ghost FM, and $W_{dw,k}$ is the corresponding deep CK. Finally, the output FM $Y_{neck} = [Y_{intrinsic}; Y_{ghost}] \epsilon RH \times W \times Cout(Cout \approx Cin)$ is obtained through channel concatenation. The module embeds the Ghost module into Neck's FPN upsampling, PAN downsampling, and feature fusion steps, replacing the original conventional convolution. This can cut down the computational intricacy associated with the feature fusion process while preserving the multi-scale feature details required for wildfire detection, thereby improving the efficiency of feature fusion. In addition, there is a problem of class imbalance in the detection of wildfires on transmission lines, where the proportion of background pixels is much higher than that of fire point targets, leading to a tendency for the model to lean towards learning from background samples during training. Therefore, the study adopts a focus LF to solve this problem, which introduces a focus parameter to decrease the loss weight assigned to easily classifiable samples while enhancing the model's ability to concentrate on samples presenting significant classification difficulties. The expression for binary classification scenarios (fire points as positive class, background as negative class) is presented in equation (5).

$$FL(p_t) = -\alpha_t(1-p_t)^{\gamma} log(p_t) \tag{5}$$

In equation (5), p_t is the likelihood of being forecasted to belong to the positive category, the value of focusing parameter γ is 2 (the empirical optimal value, which can effectively suppress the loss of easily classified samples); The category balance coefficient α_t is set to 0.75, 0.7, 0.2, and 0.25 for the four categories of fire points, smoke, background, and mechanical equipment, respectively. By increasing the weights of minority categories such as fire points and smoke, the problem of category imbalance is addressed. Extending equation (5) to multiple categories results in a total classification loss as shown in equation (6).

$$L_{cls} = \sum_{i=1}^{N} FL(p_{ti}) \tag{6}$$

In equation (6), N is the category amount and p_{ti} is the forecasted probability of the category. In multi-classification tasks, an adaptive tuning strategy is adopted: in the initial stage of training, $\gamma = 1.5$ and α_t difference is 0.4. As the number of iterations increases ($\geq$ 150 times), $\gamma = 2$ and α_t difference is adjusted to 0.55 to ensure stable convergence of the model in the early stage and precise optimization in the later stage. Further research is being conducted to use a complete IoU LF to solve the tricky issue of inaccurate localization of detection target bounding boxes affected by complex environments. During the computation of the loss between the predicted box (PB) and the real box (RB), this function goes beyond simply considering the IoU, but also incorporates factors such as center distance and aspect ratio. Its calculation is presented in equation (7).

$$L_{CIoU} = 1 - CIoU + \lambda \times \upsilon \tag{7}$$

In equation (7), IoU is the IoU of the PB and the RB. λ is the balance coefficient used to control the impact of additional terms on the total loss. υ evaluates the degree of uniformity inherent in the aspect ratio between the PB and the RB. In equation (7), the calculations for $CIoU, IoU$, and υ are shown in equation (8).

$$\begin{cases} CIoU = IoU - \dfrac{\rho^2\left(b, b^{gt}\right)}{c^2} - \alpha \times \upsilon \\ IoU = \dfrac{|B \cap B^{gt}|}{|B \cup B^{gt}|} \\ \upsilon = \dfrac{4}{\pi^2}\left(\arctan\dfrac{w^{gt}}{h^{gt}} - \arctan\dfrac{w}{h}\right)^2 \end{cases} \tag{8}$$

In equation (8), B is the PB and B^{gt} is the RB. $\rho^2(b, b^{gt})$ is the square of the Euclidean distance from the center b of the PB to the center b^{gt} of the RB, and c is the min diagonal length of the bounding rectangle that surrounds the PB and the RB. w and h are the width and height of the PB, while w^{gt} and h^{gt} are the width and height of the RB.$\alpha = \frac{\upsilon}{(1-IoU)+\upsilon}$ is the weight coefficient. Integrating the improved classification loss with the regression loss, the total loss of the Head section is presented in equation (9).

$$L_{total} = L_{cls} + \beta \cdot L_{CIoU} + L_{obj} \tag{9}$$

In equation (9), L_{obj} is the original target existence loss of YOLOv5, used to check if the FM grid has the target.β is the regression loss weight coefficient, used to balance the contributions of classification loss and regression loss, ensuring that the model accurately locates the boundary boxes of fire points while learning to distinguish between wildfire categories. The overall enhanced lightweight YOLOv5 model architecture is illustrated in Figure 3.

2.2 Prediction of Wildfire Spread Based on Bi-LSTM and Complete YOLO-LNet Model

After completing real-time detection of wildfires on transmission lines, accurate prediction of the spread trend of wildfires is crucial for disaster prevention and mitigation decisions on transmission lines. The spread of wildfires is a dynamic process that changes over time, and the development of fire points is affected by many factors, such as terrain, meteorology, and vegetation, exhibiting temporal correlation characteristics. Bi-LSTM, with its bidirectional dependency modeling ability on time-series data, can effectively explore the historical and future correlation features of wildfire spread, providing technical support for

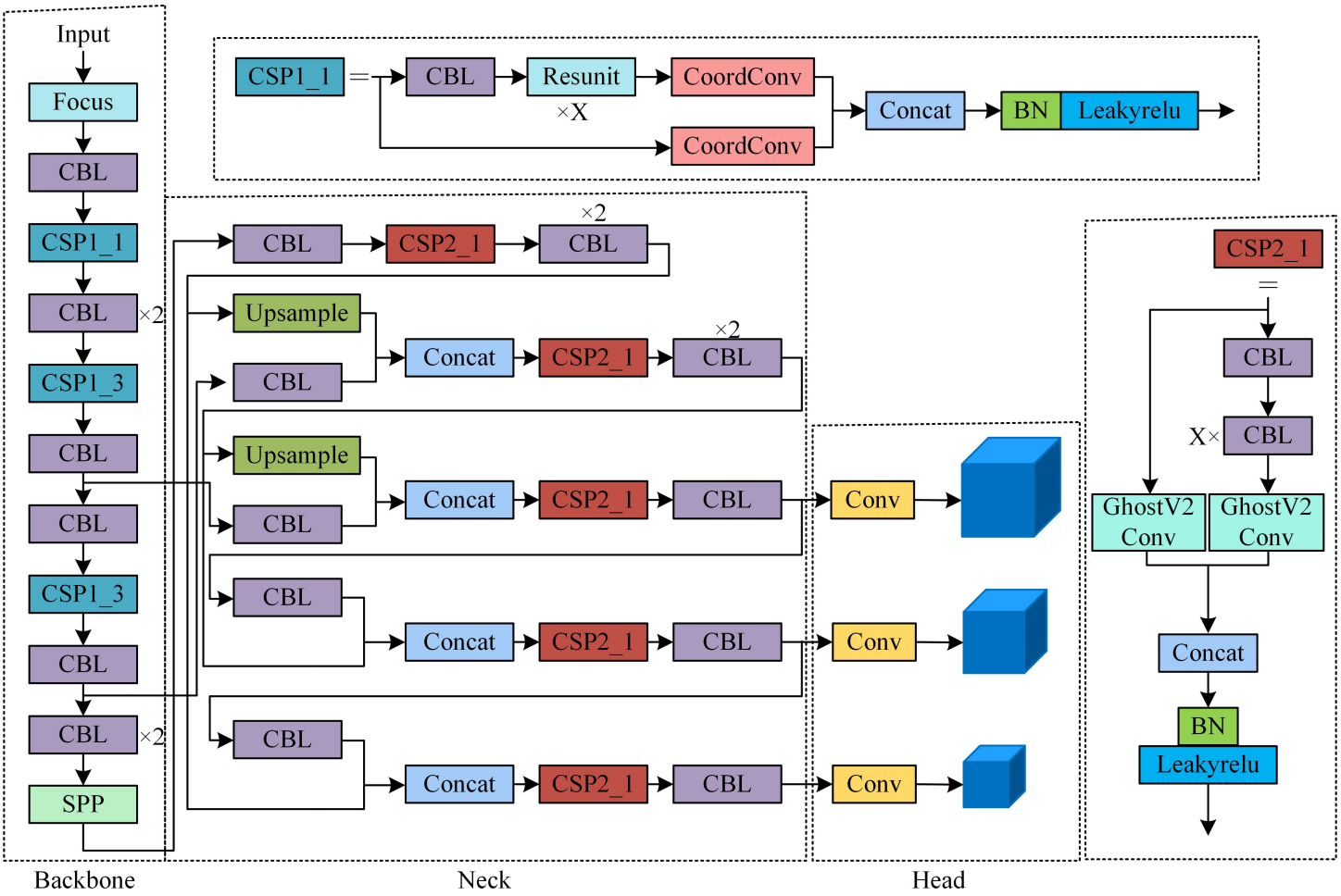


Figure 3. Improved Lightweight YOLOv5 Model Structure Diagram

wildfire spread prediction [22]. The construction process of Bi-LSTM input features is as follows. Firstly, feature extraction: Calculate the coordinates of the fire center, the burned area, and the fire intensity from the bounding box output by YOLOv5. The coordinates of the fire center are obtained from the upper left (x_t, y_t) and lower right corners of the bounding box, as shown in Equation (10).

$$\begin{cases} x_t = \dfrac{x_1 + x_2}{2} \\ y_t = \dfrac{y_1 + y_2}{2} \end{cases} \tag{10}$$

The fire area S_t is obtained by multiplying the bounding box area by the fire point pixel density (the proportion of flame pixels in the fire point area), ρ as shown in Equation (11).

$$S_t = (x_2 - x_1) \times (y_2 - y_1) \times \rho \tag{11}$$

The fire intensity I_t is the average brightness of the pixels in the fire area. If the fire area contains N pixels and the brightness value is $L_i (i = 1, 2, \cdots, N)$, then I_t it is as shown in equation (12).

$$I_t = \frac{1}{N} \sum_{i=1}^{N} L_i \tag{12}$$

Second, feature eature Encoding: Environmental factors such as terrain slope s (normalized to [0,1]), wind speed v (unit: m/s), and temperature t (unit: °) are combined with the extracted fire point features and (x_t, y_t, S_t, I_t) subjected to Z-score standardization. Environmental factors have a significant impact on the spread of fire: for every 10 ° increase in terrain slope, the speed of fire spread increases by about 15%; For every 5m/s increase in wind speed, the growth rate of burnt area increases by about 20%; The fire intensity in coniferous forest areas is 30% higher than that in broad-leaved forests. Therefore, vegetation types are subjected to single heat encoding during feature encoding to enhance the model's learning of key environmental factors. ③ Temporal Sequence Conversion: $t = 1, 2, \cdots, T$ The above features are sorted according to time steps to construct $T \times 7$ an input sequence of dimension $X = [x_1, x_2, \ldots, x_T]$, where each time step contains a feature vector $x_t = [x_t, y_t, S_t, I_t, s_t, v_t, t_t]$. The framework of Bi-LSTM as a prediction model is illustrated in Figure 4.

In Figure 4, Bi-LSTM comprises forward and backward LSTM. The forward LSTM learns the forward dependency of temporal data, while the backward LSTM learns its backward dependency. Finally, the bidirectional information output features are fused to capture the temporal cor-

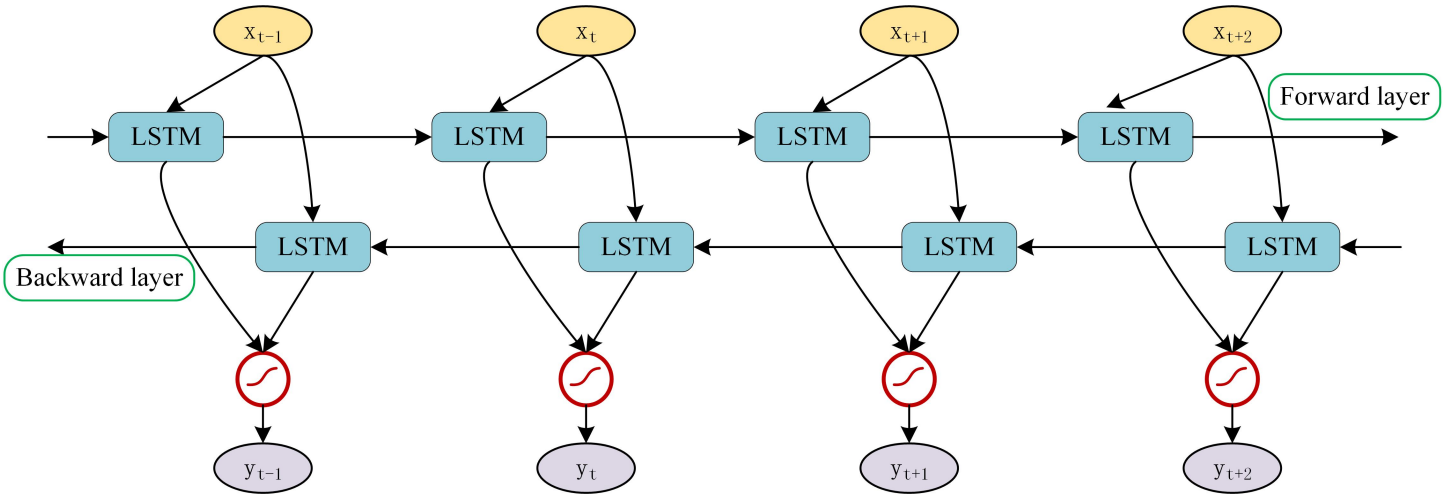


Figure 4. Structure Diagram of Bi-LSTM

relations of multiple factors. The LSTM core unit achieves selective memory and update of multi-factor temporal features through input gate i_t, forget gate f_t, output gate o_t, and cell state c_t. The input gate of LSTM controls the input of multi-factor features, and its calculation is presented in equation (13).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_i) \tag{13}$$

In equation (13), σ is the Sigmoid activation function, W is the weight matrix (The dimension is $\text{hidden}_{\text{size}} \times \text{input}_{\text{size}}$, $\text{input}_{\text{sine}} = 7$), h_{t-1} is the hidden state (HS) of the previous time step (The dimension is $\text{batch}_{\text{sixe}} \times \text{hidden}_{\text{size}}$, $\text{batch}_{\text{size}} = 32$), and b is the bias term. The forgetting gate determines the retention of historical multi-factorial association information in the cell state, as calculated by equation (14).

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f) \tag{14}$$

The candidate cell states are generated by integrating the current multi-factor features, and their calculation is presented in equation (15).

$$\widetilde{c}_t = \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \tag{15}$$

In equation (15), tanh is the hyperbolic tangent activation function. Then, by integrating historical and current multi-factor information, the cell state is updated, and the calculation is presented in equation (16).

$$c_t = f_t * c_{t-1} + i_t * \widetilde{c}_t \tag{16}$$

The output gate screens the multi-factor correlation features used for output in the cell state, and its calculation is presented in equation (17).

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o) \tag{17}$$

The final output contains hidden features with multiple temporal correlations, which are calculated as shown in equation (18).

$$h_t = o_t * tanh(c_t) \tag{18}$$

Bi LSTM concatenates forward HS $h_t{}'$ and backward HS ht'' to obtain $h_t^{bi} = ht'; ht''$, thereby capturing the bidirectional temporal correlation of wildfire temporal data. The HS h_t^{bi} output by Bi LSTM is input to the fully connected layer for prediction output. To predict the future spread of wildfires in k steps, a prediction LF needs to be constructed. If the forecasted value is $\hat{y}_{T+1}, \ldots, \hat{y}_{T+k}$ and the true value is $y_{T+1}, \ldots, y_{T+k}$, then the LF is presented in equation (19).

$$L_{pred} = \frac{1}{k}i = 1kyT + i - yT + i22 \tag{19}$$

The study aims to train a Bi-LSTM network to fit the changing patterns of wildfire time-series data. By utilizing the detected historical fire point time-series features, the study predicts the location and extent of wildfires in the future, providing a basis for predicting the spread trend of wildfires in transmission lines and making protective decisions. This enables a progressive application from detecting wildfires to predicting the development of wildfires. The research will connect the lightweight YOLOv5 detection module with the Bi-LSTM prediction module to form a complete YOLO-LNet model. The complete process of using YOLO-LNet to predict the spread of wildfires on transmission lines is presented in Figure 5.

In Figure 5, the images captured by unmanned aerial vehicles or fixed cameras are input into the lightweight YOLOv5 for real-time detection of fire points and smoke positions, and corresponding feature sequences are generated. Meanwhile, the corresponding data of multiple influencing factors, such as terrain and weather, are encoded and input into the feature interaction layer along with the feature sequence. The feature interaction layer between the detection and prediction modules enables bidirectional transmission of detection features and intermediate features generated during the prediction process. Detecting features can assist in the initialization of the prediction module, and the predicted intermediate features can be fed back to the detection module to optimize the next frame detection result, thus forming a closed-loop structure. The final output includes real-time detection results

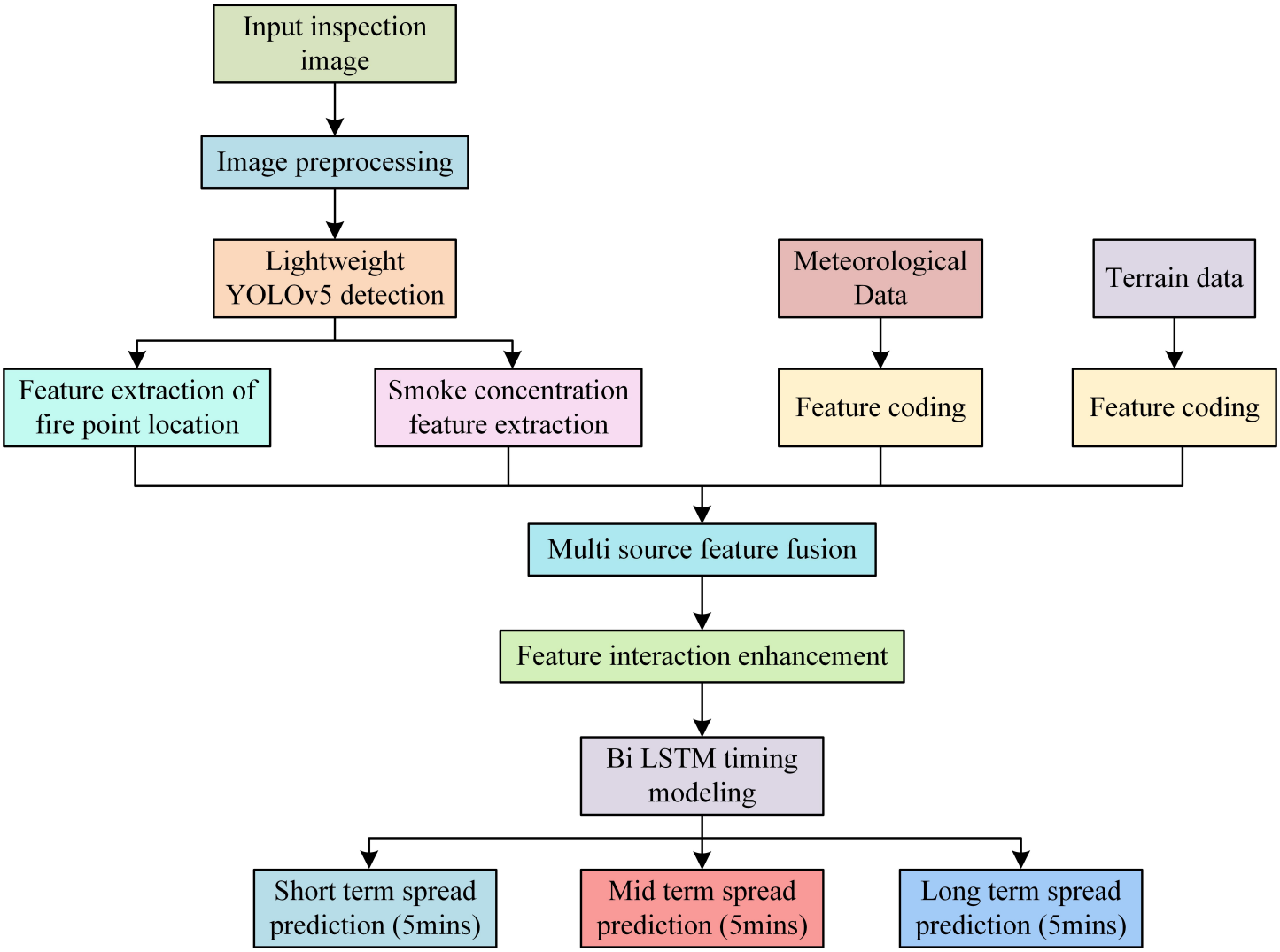


Figure 5. The Complete Process of YOLO-LNet for Predicting the Spread of Wildfires on Transmission Lines

of wildfires and short-term predictions for the future, providing a decision-making basis for disaster prevention and mitigation of transmission lines.

3. Results

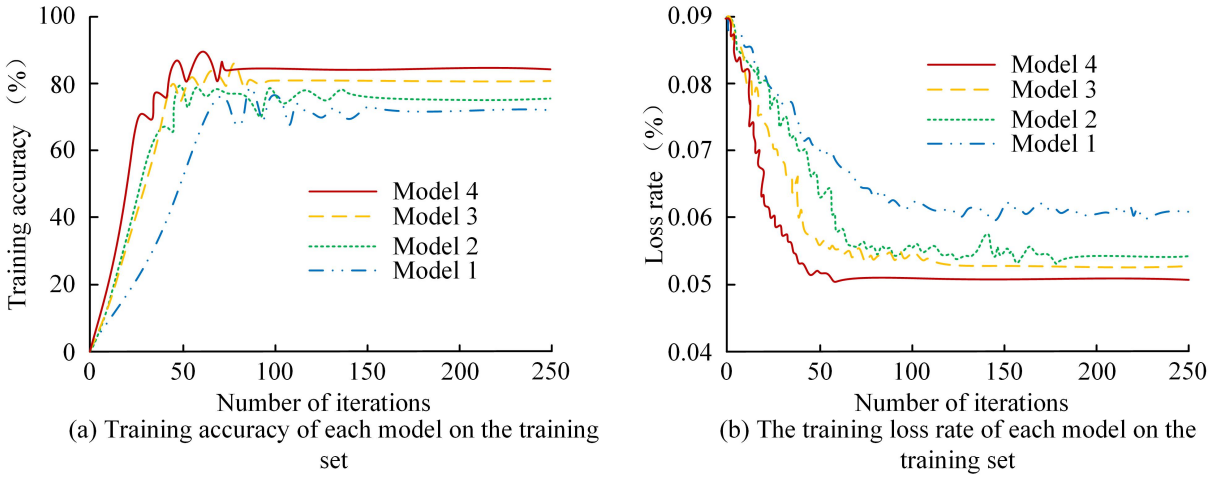
3.1 Performance Analysis of Transmission Line Wildfire Detection

To guarantee the validity of the introduced lightweight YOLOv5 model for detecting wildfires in transmission lines, ablation and comparison experiments were done on the dataset. The study collected a substantial quantity of forest fire images from public datasets and drone inspections, covering different terrains, vegetation, lighting, weather, and fire scale conditions. The sample images were expanded to 5000 through the following data augmentation methods: ① Rotation: Random rotation angle range of -15° ∼ 15 °, step size of 5 °; ② Crop: Randomly crop at a ratio of 0.7 to 1.0 times the original image, and resize to 640 × 640 after cropping; ③ Contrast adjustment: The brightness and contrast are randomly adjusted within the range of 0.5 to 1.5 times; ④ Additional enhancements: Add simulated smoke (transparency 0.2 ∼ 0.5), random Gaussian noise (variance 0.001 ∼ 0.005), and lighting changes (simulating the difference in day night lighting intensity, brightness 0.3 ∼ 1.2 times) to enhance the model's adaptability to complex scenes. Considering the time series characteristics of wildfire spread, the study adopts the rolling window method to partition the dataset: in chronological order, the first 70% of the time period samples are selected as the training set, the middle 20% as the validation set, and the last 10% as the test set, to avoid the problem of time series leakage caused by random partitioning and ensure that the prediction model learns the true temporal dependencies. The software and hardware environment required for the experiment is in Table 1.

In the experiment, the max iteration count for the model was set to 300, the batch size was 32, the original learning rate was 0.01, and the min learning rate was 0.0001. The study used the Adam optimizer to accelerate the convergence rate of the model. The study first verified the improvement effect of the lightweight YOLOv5 model

Table 1
Experiment Setup

Name	Set up
CPU	Intel(R) Xeon(R) CPU E5-5135 v3@ 3.65GHz
GPU	4×NVIDIA GeForce RTX 2080 Ti
Operating system	Ubuntu16.04-based Linux OS
DL framework	Pytorch 1.8.1
GPU acceleration library	CUDA 10.2
Torchvision Version	0.9.1
Python Version	3.8.10
Other Dependencies	OpenCV 4.5.3, NumPy 1.21.2, Scikit-learn 0.24.2



(a) Training accuracy of each model on the training set

(b) The training loss rate of each model on the training set

Figure 6. Training Accuracy and Loss Rate of Each Model on the Dataset

through ablation experiments. For the convenience of displaying the results, four model variants were set up in the study. Model 1 is the traditional YOLOv5s, Model 2 is YOLOv5+CoordConv, Model 3 is YOLOv5+Ghost, and Model 4 is the proposed lightweight YOLOv5. The training accuracy and loss rate of each model on the dataset are presented in Figure 6.

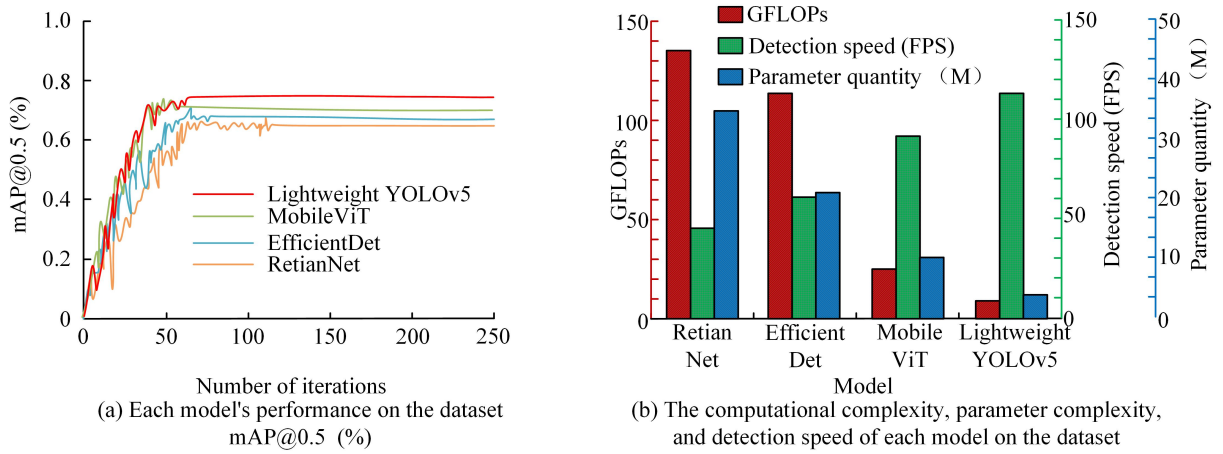
In Figure 6 (a), the detection accuracy of Model 1 on the training set was relatively low, and the initial increase was slow. It gradually stabilized at 71.8% after 150 iterations. The improvement of models 2 and 3 both resulted in performance enhancement, converging to 75.2% and 80.3% after 90 and 130 iterations, respectively. However, the initial convergence speed of Model 3 is slightly faster than Model 2, mainly because the Ghost module generates "ghost" features through linear transformation, effectively reducing the redundancy of the feature map and accelerating the feature learning process in the early stage of the network; However, CoordConv improves spatial perception ability by embedding coordinate information, and its benefits are more reflected in the improvement of positioning accuracy, with a relatively small direct impact on convergence speed. After only 70 iterations, the detection accuracy of detection model 4 proposed by the research institute converged to 84.7%, an improvement of 12.9% compared to model 1. As shown in Figure 6 (b), the training loss rate of Model 1 is the highest and fluctuates greatly, reflecting its insufficient fitting ability and unstable training in complex wildfire scenarios. The loss rates of models 2 and 3 gradually stabilized at 0.055 and 0.053 after 180 and 110 iterations, respectively. Compared with other models, the training loss rate of Model 4 decreases the fastest in the early stage and converges to 0.051 after only 50 iterations. This further confirms that the improved Backbone and Neck structures enhance the ability to extract and distinguish features of wildfire targets, making the model learning process more efficient and stable. Table 2 offers a comprehensive display of the performance contrast among various models when evaluated on the dataset.

From Table 2, the mAP@0.5 of Model 1, which did not undergo any improvements, was 70.2%, with GFLOPs of 15.8, a parameter count of 7.02M, and a model size of 14.4MB. And the improved model 2 by CoordConv mAP@0.5 Improved by 2.3% compared to Model 1; Model 3 improved by Ghost module has a more significant contribution to reducing computational complexity (20.3%) and parameter count (14.8%). The fully improved Model 4 had an mAP@0.5 of 74.3%, GFLOPs of 8.2, a parameter count of 3.12M, and a model size of 6.5MB. Compared to Model 1, its mAP@0.5 increased by 4.1%, while GFLOPs, parameter count, and model size decreased by 48.1%, 55.5%, and 54.8%, respectively. This demonstrated that the proposed model ensured wildfire detection accuracy and also fit real-time and lightweight deployment requirements. Compare several basic wildfire detection models selected for research with the proposed model, including RetinaNet, Efficient

Table 2
Comparative Analysis of the Performance Exhibited by Diverse Models

Model	CoordConv	Ghost	mAP@0.5 (%)	Computational Complexity (GFLOPs)	GFLOPs decrease (%)	Parameter Count (M)	Parameter reduction (%)	Model Size (MB)
Model1	×	×	70.2	15.8	/	7.02	/	14.4
Model2	√	×	72.5	14.5	8.2	6.85	2.4	13.9
Model3	×	√	73.1	12.6	20.3	5.98	14.8	12.1
Model4	√	√	74.3	8.2	48.1	3.12	55.5	6.5

Note: GFLOPs: Giga Floating point Operations, used to measure computational complexity.



(a) Each model's performance on the dataset mAP@0.5 (%)

(b) The computational complexity, parameter complexity, and detection speed of each model on the dataset

Figure 7. Performance of Each Model on the Dataset

Detection (EfficientDet), and Mobile Vision Transformer (MobileViT). The performance of each model is presented in Figure 7.

According to Figure 7 (a), the accuracy of RetianNet gradually stabilized at 66.7% after 120 iterations. The accuracy of EfficientDet and MobileViT converged to 68.5% and 70.6% after 75 and 50 iterations, respectively. The lightweight YOLOv5 had the highest accuracy, converging to 74.2% after 50 iterations. In Figure 7 (b), compared with the other three comparison models, the lightweight YOLOv5 had the lowest computational complexity on the dataset, the fastest detection speed, and the smallest quantity of model parameters, which were 9.8, 112.3 FPS, and 3.12M, respectively. To further demonstrate the superiority of the proposed model's detection performance, the study also compared it with current mainstream wildfire detection models, including Faster R-CNN based on a two-stage detection framework: its Backbone was replaced with ResNet-50, and the FireNet idea was borrowed to enhance smoke texture feature extraction in the region proposal network RPN, referred to as Faster R-CNN FineNet, and Swin Transformer Tiny detector with pure Transformer architecture. The comparison results are shown in Table 3.

According to Table 3, Faster R-CNN FineNet achieved high detection accuracy (72.8%) with its two-stage detection architecture for fine processing of candidate regions. However, its computational complexity (95.6 GFLOPs) and parameter count (36.8M) far exceeded the proposed model, resulting in a detection speed of only 22.5 FPS, which is difficult to meet the demanding requirements of real-time inspection. The Swin Transformer Tiny model utilizes the global attention mechanism of Transformer and achieves accuracy similar to Faster R-CNN (71.9%), while significantly reducing computational overhead (25.4 GFLOPs), demonstrating the advantage of Transformer architecture in balancing accuracy and efficiency. However, compared with the lightweight YOLOv5 proposed by the research institute, its computational complexity is still 3.1 times that of the proposed model, the number of parameters is 5.3 times, and the detection speed (48.1 FPS) is much lower than that of the proposed model (112.3 FPS). In summary, the lightweight YOLOv5 model proposed in the study achieved the best detection accuracy (74.3%), while leading the two mainstream advanced architectures in computational efficiency, model lightweighting, and inference speed with significant advantages, fully demonstrating its stronger practicality and competitiveness in real-time detection tasks of transmission line wildfires.

3.2 Analysis of the Prediction Results of Mountain Fire Spread on Transmission Lines

After verifying the detection capability of lightweight YOLOv5, the study further validated the effectiveness of YOLO-LNet in predicting wildfire spread through comparative experiments. YOLO-LNet is an overall model based on the lightweight YOLOv5 detection module and Bi-LSTM prediction module. Therefore, the selected prediction comparison models all took the detection output

Table 3
Comparison of Detection Performance of Various Models on the Dataset

Model	mAP@0.5 (%)	Computational Complexity (GFLOPs)	Parameter Count (M)	Detection speed (FPS)	Model Size (MB)
Faster R-CNN-FireNet	72.8	95.6	36.8	22.5	147.2
Swin-Transformer-Tiny	71.9	25.4	16.5	48.1	66.0
Lightweight YOLOv5 (Ours)	74.3	8.2	3.12	112.3	6.5

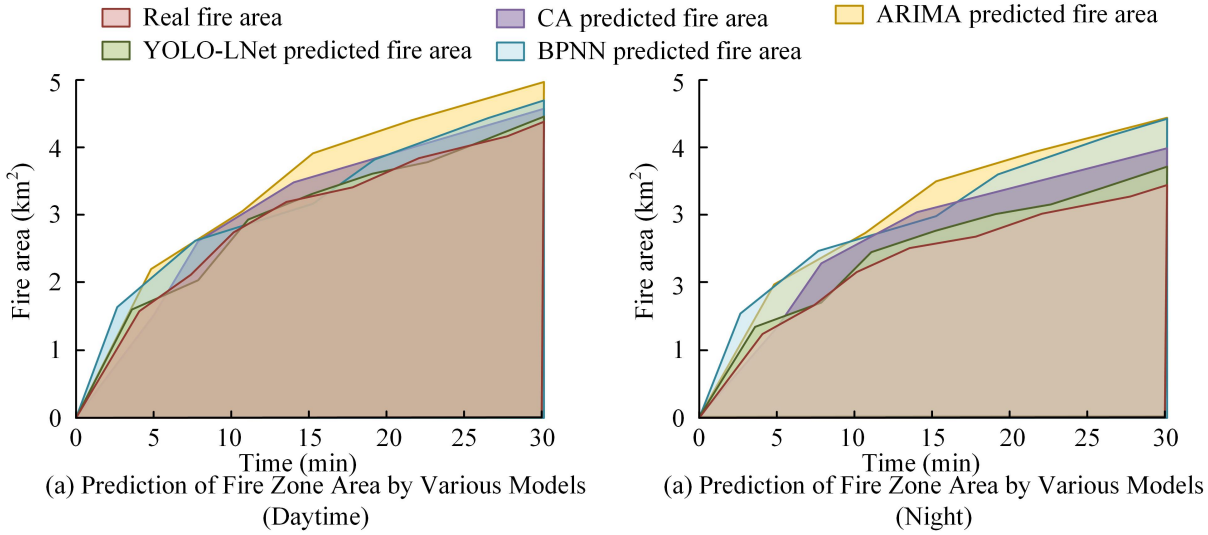


Figure 8. Comparison Between the Predicted and the Actual Fire Area

of lightweight YOLOv5 as input. The comparative models selected for the study were Cellular Automata (CA), Backpropagation Neural Network (BPNN), and Autoregressive Integrated Moving Average (ARIMA) models. As time increased, the predicted fire area by each model and the actual fire area are shown in Figure 8.

From Figure 8 (a), under sufficient daylight conditions, YOLO-LNet's predicted fire area had the closest growth trend and size to the actual fire area. At 30 minutes, its predicted area was about 4.5 km2, slightly larger than the actual area of 4.4 km2. To quantitatively evaluate the stability of the prediction, five independent repeated experiments were conducted on YOLO LNet, and the standard deviation of its prediction results was calculated. At all predicted time points, the standard deviation of the predicted area is less than 0.15 km ^{2}, and the coefficient of variation (CV) is less than 3.5%. This indicates that the prediction results of the model have good consistency and reliability, with a small fluctuation range. Secondly, CA had better prediction performance, but its prediction accuracy was slightly lower than YOLO-LNet. ARIMA had the worst prediction performance, as its predicted area was much larger than the actual fire area. As shown in Figure 8 (b), due to the low visibility and more complex environment at night, the prediction accuracy of each model decreased. However, compared with other models, YOLO-LNet had the highest trend fit and the closest predicted area. Despite the overall increase in prediction uncertainty for nighttime scenes, the standard deviation of YOLO LNet remained within 0.25 km ^{2} in 5 repeated experiments, indicating that the proposed model has stronger robustness and prediction stability for complex situations such as nighttime smoke interference and insufficient lighting. The comparison of the mean absolute error (MAE) of each model under various wind speeds and terrain slopes is presented in Figure 9.

As shown in Figure 9 (a), the MAE of each model was relatively low when the wind speed was 5-8 m/s. The MAE of YOLO-LNet was significantly lower than that of the other three models, being only 4.15. When the wind speed rose to 8-12 m/s, the MAE of ARIMA, BPNN, and CA sharply increased to 11.12, 8.97, and 10.15, respectively, while the MAE of YOLO-LNet was only 5.23, indicating that YOLO-LNet was minimally affected by wind speed. As shown in Figure 9 (b), regardless of whether the slope was greater than or less than 25°, YOLO-LNet performed significantly better than the other comparative models. When the slope was less than 25° and greater than 25°, the MAE values of YOLO-LNet were 4.22 and 5.38, respectively, which verified the strong robustness of the model to tricky surroundings. At the end of the study, the training time and prediction time of each model were compared, and the results are shown in Figure 10.

As shown in Figure 10 (a), ARIMA had the shortest training time, lasting only 18.9 minutes, due to its reliance solely on temporal statistics. The training time of YOLO-LNet was slightly longer than that of the CA model and shorter than that of the complex BPNN model, at 31.5

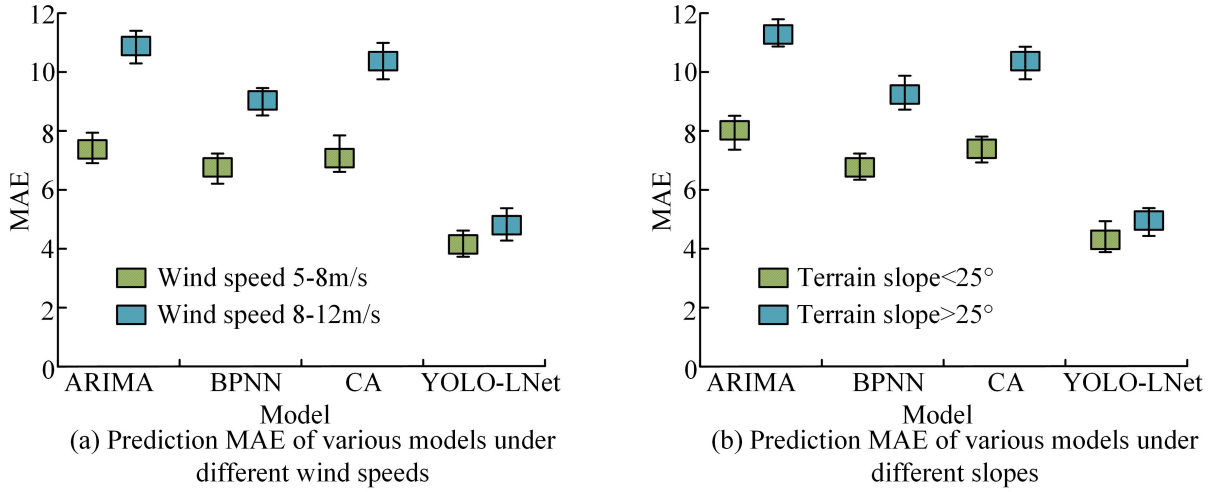


Figure 9. MAE of Each Model Under Different Wind Speeds and Terrain Slopes

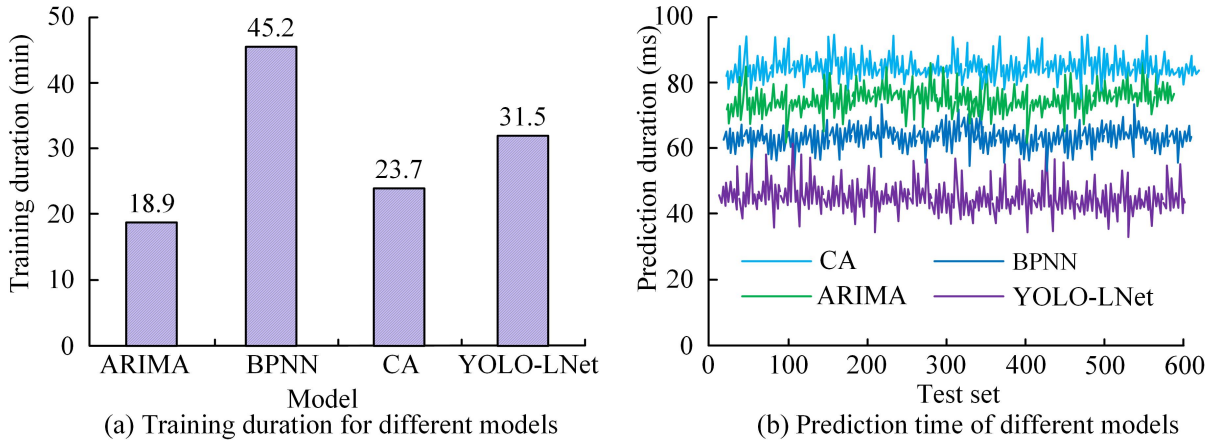


Figure 10. Training Duration and Prediction Duration of Each Model

minutes. Although the training time of YOLO-LNet was relatively long, its prediction accuracy was high, indicating that the model balanced accuracy and efficiency well. As shown in Figure 10 (b), the prediction time of the CA model was as long as 82.6 ms, which could not meet the real-time prediction requirements. Thanks to the end-to-end temporal fusion mechanism, YOLO-LNet's prediction time was only 43.8 ms. The duration encompasses the entire end-to-end process, from image input and lightweight YOLOv5 detection to feature sequence construction and Bi-LSTM prediction. To further verify real-time performance, the study conducted tests on the embedded platform Jetson Xavier NX, achieving an average inference time of 89.5ms. This still meets the real-time requirements of most inspection scenarios, demonstrating its potential for edge deployment. At the end of the study, the proposed YOLO LNet model was compared with two advanced deep learning temporal models to verify its superiority, including the sequence-to-sequence model with attention mechanism (Seq2Seq Attn) and the Transformer temporal prediction model. All comparison models use the detection output of lightweight YOLOv5 as a unified input feature to ensure fairness. The performance of each model is shown in Table 4.

Table 4
Quantitative Comparison of Prediction Performance Among Deep Learning Models

Model	MAE (km^2)	Training Time (min)	Inference Time (ms)	Model Size (MB)
Seq2Seq-Attn	7.45	38.2	68.5	8.7
Transformer	5.98	75.6	125.7	19.3
YOLO-LNet (Ours)	5.23	31.5	43.8	6.5

According to Table 4, the Transformer model demonstrates strong sequence modeling capabilities, with a prediction accuracy (MAE: 5.98) closest to YOLO LNet. However, its model complexity and computational cost are enormous, resulting in training time (75.6 minutes) and single prediction time (125.7ms) being 2.4 times and 2.9 times longer than the proposed model, respectively. The model size is also larger. The Seq2Seq Attn model outperforms Transformer in training and prediction speed, but its ability to capture complex spatiotemporal dynamics of wildfire spread is limited, resulting in significantly higher

prediction error (MAE: 7.45) compared to the other two models. Through comprehensive comparison, the YOLO LNet model proposed in the study has achieved the best level in three core indicators: prediction error, inference efficiency, and model lightweighting, achieving the best balance between accuracy and speed, fully demonstrating its superiority as an edge deployment solution.

4. Discussion and Conclusion

To achieve accurate detection and dynamic prediction of wildfires on transmission lines, a YOLO-LNet real-time prediction model was proposed. This model integrated lightweight YOLOv5 and Bi-LSTM, enhanced spatial positioning through CoordConv, reduced computational redundancy through the Ghost module, and optimized the detection module by combining focus loss and CIoU loss. The study simultaneously utilized the Bi-LSTM bidirectional temporal modeling capability to capture the spread patterns of wildfires. The outcomes revealed that the lightweight YOLOv5 achieved an mAP@0.5 of 74.3%, which marked a 4.1% increase relative to the traditional YOLOv5s. Its computational complexity was reduced by 48%, and the model size was compressed to 6.5 MB. The reason was that CoordConv enhanced the feature extraction capability for small targets through coordinate information embedding, while the Ghost module reduced redundant computations using an intrinsic feature + ghost feature mechanism. Together, they achieved a balance between accuracy and efficiency. YOLO-LNet exhibited an average absolute error of only 5.23 when the wind speed was 8-12 m/s. During nighttime predictions, the deviation between the forecasted fire area and the true value was less than 0.1 km^2, with a prediction duration of 43.8 ms. This was because Bi-LSTM effectively fused multi-dimensional data, such as fire point locations and environmental factors, through forward and backward temporal modeling, capturing the temporal dependencies of fire spread. When juxtaposed with models like CA, BPNN, Seq2Seq, and Transformer, the present model exhibited stronger robustness and superior real-time performance under complex terrain and meteorological conditions, providing an effective technical solution for forest fire prevention and control in transmission lines.

However, there are still certain limitations to the research. First of all, the performance verification of the model is mainly carried out under conventional meteorological conditions, and its detection and prediction generalization ability in extreme weather (such as dense fog, rainstorm) needs further special testing and enhancement. Secondly, the research mainly focuses on the innovation and verification at the algorithm level. The focus of future work is to deploy the model to the actual embedded devices (such as patrol drones, edge computing nodes), and comprehensively evaluate its power consumption, memory occupation, and long-term operation stability in practical applications to verify the feasibility of its engineering application. In addition, exploring more refined dynamic environmental factor fusion mechanisms, combined with hardware acceleration technologies such as TensorRT, will be the key to further improving model robustness and extreme real-time processing capabilities. Finally, from the perspective of expanding application scenarios, intelligent inspection of transmission lines is a multitasking scenario that covers fire warning, equipment inspection, and channel environment monitoring. Therefore, in the future, it is possible to explore the integration of the core capabilities of this model with tasks such as identifying defects in line equipment and monitoring vegetation encroachment in channels, in order to construct a unified multi-task learning model and further enhance its comprehensive application value and deployment efficiency in real operation and maintenance scenarios. In summary, the YOLO LNet model proposed by the research institute performs well in real-time detection and propagation prediction of wildfires on transmission lines, and has good application prospects. Subsequent research will focus on improving its robustness and deploying it in complex real-world environments.

Funding

This research was funded by the 2025 Science and Technology Project of State Grid Power Space Technology Co., Ltd, titled "Research on Key Technologies for Monitoring and Early Warning of Typical Meteorological Disasters of Transmission Lines in Microtopography and Microclimate Areas" (No. 529500250007).

References

[1] S. Dong, Y. Ding, L. Ma, W. He, and X. Lei, "Design of mountain fire prevention monitoring system for transmission lines based on machine vision algorithms," *International Journal of Emerging Electric Power Systems*, vol. 24, no. 4, pp. 529–539, 2023.

[2] Y. Limei, H. Junjie, G. Zhengxu, W. Jun, and Z. Liwen, "Analysis of the distribution characteristics of mountain fires based on the disaster data of hubei transmission lines," *Southern Energy Construction*, vol. 11, no. 1, pp. 196–204, 2024.

[3] S. Li, Y. Dong, W. Zhang, *et al.*, "Self-adaptive optimal allocation strategy of emergency resources for power distribution network failures," *International Journal of Power and Energy Systems*, vol. 43, no. 10, 2023.

[4] J. Li, "Electrical equipment condition monitoring and predictive maintenance strategy based on optimised ant colony algorithm," *International Journal of Power and Energy Systems*, vol. 44, no. 10, 2024.

[5] M. Grari, M. Yandouzi, I. Idrissi, M. Boukabous, O. Moussaoui, M. Azizi, and M. Moussaoui, "Using iot and ml for forest fire detection, monitoring, and prediction: a literature review," *Journal of Theoretical and Applied Information Technology*, vol. 100, no. 19, pp. 5445–5461, 2022.

[6] Z. Wu, R. Xue, and H. Li, "Real-time video fire detection via modified yolov5 network model," *Fire Technology*, vol. 58, no. 4, pp. 2377–2403, 2022.

[7] R. Vikram and D. Sinha, "A multimodal framework for forest fire detection and monitoring," *Multimedia Tools and Applications*, vol. 82, no. 7, pp. 9819–9842, 2023.

[8] P. V. A. B. Venâncio, R. J. Campos, T. M. Rezende, A. C. Lisboa, and A. V. Barbosa, "A hybrid method for fire detection based on spatial and temporal patterns," *Neural Computing and Applications*, vol. 35, no. 13, pp. 9349–9361, 2023.

[9] C. H. Kang and S. Y. Kim, "Real-time object detection and segmentation technology: an analysis of the yolo algorithm," *JMST Advances*, vol. 5, no. 2, pp. 69–76, 2023.

[10] J. Purohit and R. Dave, "Leveraging deep learning techniques to obtain efficacious segmentation results," *Archives of Advanced Engineering Science*, vol. 1, no. 1, pp. 11–26, 2023.

[11] J. Lee and K. I. Hwang, "Yolo with adaptive frame control for real-time object detection applications," *Multimedia Tools and Applications*, vol. 81, no. 25, pp. 36375–36396, 2022.

[12] W. Liu, G. Ren, R. Yu, S. Guo, J. Zhu, and L. Zhang, "Image-adaptive yolo for object detection in adverse weather conditions," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 1792–1800, 2022.

[13] Y. Song, Z. Xie, X. Wang, and Y. Zou, "Ms-yolo: Object detection based on yolov5 optimized fusion millimeter-wave radar and machine vision," *IEEE Sensors Journal*, vol. 22, no. 15, pp. 15435–15447, 2022.

[14] G. Li, Z. Ji, X. Qu, R. Zhou, and D. Cao, "Cross-domain object detection for autonomous driving: A step-wise domain adaptive yolo approach," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 603–615, 2022.

[15] N. K. Karnam, S. R. Dubey, A. C. Turlapaty, and B. Gokaraju, "Emghandnet: A hybrid cnn and bi-lstm architecture for hand activity classification using surface emg signals," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 1, pp. 325–340, 2022.

[16] X. Hu, T. Liu, X. Hao, and C. Lin, "Attention-based conv-lstm and bi-lstm networks for large-scale traffic speed prediction," *The Journal of Supercomputing*, vol. 78, no. 10, pp. 12686–12709, 2022.

[17] V. Kiran Kumar, K. V. Ramesh, and V. Rakesh, "Optimizing lstm and bi-lstm models for crop yield prediction and comparison of their performance with traditional machine learning techniques," *Applied Intelligence*, vol. 53, no. 23, pp. 28291–28309, 2023.

[18] Y. L. He, P. F. Wang, and Q. X. Zhu, "Improved bi-lstm with distributed nonlinear extensions and parallel inputs for soft sensing," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 3748–3755, 2023.

[19] S. Khullar and N. Singh, "Water quality assessment of a river using deep learning bi-lstm methodology: forecasting and validation," *Environmental Science and Pollution Research*, vol. 29, no. 9, pp. 12875–12889, 2022.

[20] S. Gothane, "A practice for object detection using yolo algorithm," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 7, no. 2, pp. 268–272, 2021.

[21] Y. Chen, H. Xu, X. Zhang, P. Gao, Z. Xu, and X. Huang, "An object detection method for bayberry trees based on an improved yolo algorithm," *International Journal of Digital Earth*, vol. 16, no. 1, pp. 781–805, 2023.

[22] A. S. A. Ali, S. Ebrahimi, M. M. Ashiq, M. S. Alasta, and B. Azari, "Cnn-bi-lstm neural network for simulating groundwater level," *Environmental Engineering*, vol. 8, no. 1, pp. 1–7, 2022.

Biographies

Guozhu Yang received his Bachelor of Engineering degree in Surveying and Mapping Engineering from Shandong University of Science and Technology in 2013 and his Master of Engineering degree in Surveying and Mapping Engineering from China University of Mining and Technology (Beijing) in 2015. Since 2022, he has been Deputy Director of the Aerospace Technology Office in the Space Technology Application Center at State Grid Electric Power Space Technology Co., Ltd. His research interests include the application of aerospace technology in power grids.

Wei Du received his Bachelor's degree in Mechatronic Engineering from Beijing University of Posts and Telecommunications. Since then, he has been with State Grid Power Space Technology Co., Ltd. His research interests include power grid operation and maintenance, surveying and mapping technology, and digital technology.

Huamin Sun received her Bachelor of Science degree in Atmospheric Sciences from Ocean University of China in 2018 and her Ph.D. degree in Meteorology from the University of Chinese Academy of Sciences in 2023. Since 2023, she has been a Principal Specialist in Meteorological Application Analysis at State Grid Electric Power Space Technology Co., Ltd. Her research interests include power grid disaster prediction and early warning technology.

Shirui Sun received his Bachelor of Engineering degree in Surveying and Mapping Engineering from Hebei GEO University in 2017 and his Master of Engineering degree in Surveying and Mapping Engineering from China University of Mining and Technology (Beijing) in 2020. Since 2020, he has been a Principal Specialist in Digital Power Grid Technology R&D at State Grid Electric Power Space Technology Co., Ltd. His research interests include UAV remote sensing applications.

Pengchao Sun received his Bachelor of Engineering degree in Surveying and Mapping Engineering from Chang'an University in 2017 and his Ph.D. degree in Geodesy and Surveying Engineering from the University of Chinese Academy of Sciences in 2024. Since 2024, he has been an Engineer at State Grid Electric Power Space Technology Co., Ltd. His research interests include satellite remote sensing technology and climate change.