Improved CNN Recognition Algorithm for Identifying Bird Hazards in Transmission Lines

Junzhou Li*, Yao Li, Wen Wang
State Grid Henan Electric Power Company, Hebi Power Supply Company, Hebi 458000, China

***Abstract*—With the expansion of the power grid, bird activities have become the main factor causing transmission line failures. How to accurately identify hazard birds has received widespread attention from all sectors of society. However, the current bird identification methods for transmission line hazards suffer from low accuracy due to the small size of bird targets. This study proposes an enhanced Convolutional Neural Network (CNN) with Support Vector Machines (SVM) to improve the accuracy of identifying hazardous birds on transmission lines. At the same time, a dataset of bird species affected by transmission lines is constructed, and data augmentation methods and denoising deep convolutional networks are used to process the data. Thus, a bird identification algorithm for transmission line hazards based on improved CNNs and SVM is constructed by combining the three. The study conducts a performance comparison analysis of the algorithm and finds that its average recognition speed and accuracy are 9.8 frames per second and 97.4%, respectively, significantly better than the compared algorithms. In addition, an analysis of the application effect of the algorithm is conducted, and it is found that the algorithm can accurately identify hazard birds. In some recognition results, the recognition results and confirmation probabilities for *Pica Pica, ciconia boyciana, egretta garzetta, and hirundo rusticas* are 98.73%, 97.68%, 96.54%, and 91.34%, respectively, all above 90%. The above findings indicate that the proposed identification algorithm has good performance and practical value, which helps to improve the accuracy of identifying hazard birds on transmission lines.**

***Keywords*—CNN; hazard birds; transmission line; distinguish; support vector machine**

I. Introduction

As the social economy rapidly develops and the power grid scale continuously expands, the safety and stability of overhead transmission lines, as an important infrastructure for power transmission, have become particularly important [1]. Birds build nests and excrete on transmission lines, causing damage to transmission equipment and short circuits, posing a significant threat to the stability of power lines [2]. Therefore, assisting transmission line inspection personnel in identifying birds that may pose a threat to the lines is important for ensuring the safety of transmission lines and preventing accidents [3]. However, the current bird identification methods for transmission line hazards suffer from low accuracy due to the small size of bird targets [4, 5]. Convolutional Neural Network (CNN) is a deep learning architecture that has strong feature extraction ability and good generalization ability, and is broadly utilized in fields such as image recognition and facial recognition [6]. However, CNN using a fixed architecture and parameters may not fully capture all the information in the data, which may limit its ability to express features [7, 8]. In addition, if the model structure is too complex or the training samples are insufficient, it may also lead to overfitting. Multi-convolutional feature fusion refers to the combination of feature maps from different convolutional layers in deep learning to improve the performance and feature representation ability of the model. It can effectively compensate for the limited feature representation ability of CNN. Support Vector Machine (SVM) is a binary classification model whose basic principle is to maximize the interval between sample points of different categories by finding an optimal hyperplane. SVM has the ability to avoid overfitting and handle high-dimensional data, which can effectively solve the problem of CNN overfitting. Therefore, this study utilizes backbone feature extraction networks (DarkNet-53), GoogleNet, Visual Geometry Group 19 Layer Network (VGG-19), and EfficientNet-B0 to extract features from images of hazard birds on transmission lines. Multiple convolution fusion methods are used to cascade fuse the extracted convolution features to construct an improved CNN. SVM is then used to classify and recognize the obtained features, and a transmission line hazard bird recognition model based on improved CNN and SVM is constructed. The innovation of this study lies in the convolutional feature fusion of CNN and the use of SVM to recognize and classify the fused images, aiming to raise the accuracy of bird recognition on transmission lines. It is expected that this method can contribute to enriching the theory of bird recognition of transmission line hazards.

*Corresponding Author.

II. Related Works

In recent years, with the rapid development of society and economy, the demand for electricity continues to increase, and transmission lines are regarded as an important infrastructure for power transmission. At present, the transmission line failure caused by the behavior of endangering the life of birds occurs frequently, and even leads to fire and other disasters. The transmission line failure caused by bird activity is particularly serious. The identification of birds endangering the safety of the transmission line is of great significance to ensure the safe and stable operation of the power system. Many experts have carried out relevant research on the identification of birds harmed by transmission lines. For example, to explore the problem of tripping caused by birds touching power lines, Rebolo-Ifran team adopted the method of literature review to make a summary of the harm to birds, but it is not practical [9]. To solve the problem of power interruption caused by electric shock of birds and thus damage the integrity of the power network, Biasotto team developed a framework to simulate the risk of electric shock of birds. After experimental verification, the framework identified 283 species facing the risk of electric shock, 38 of which were high risk, and birds of prey accounted for 76% [10]. Yuan et al. proposed an improved YOLOv5 technology to solve the problem of low bird identification accuracy in transmission lines, and the experimental verification showed that this technology improved the detection speed and accuracy of birds in transmission lines [11]. To solve the problem that it is difficult to identify birds endangering overhead transmission and distribution lines, Qiu's team proposed an automatic classification method of birds related to power line faults that combines deep convolutional features with error correction output code SVM. Experiments were conducted with this method and other methods, and the results showed that the average accuracy of this method was 94.39%, which was superior to the comparison method [12]. To solve the problem of woodpeckers' low accuracy in assessing composite insulator damage of UHV transmission lines, Zhang proposed a birding damage assessment method for composite insulators of UHV lines based on electric field simulation and deep learning. The results of simulation experiments showed that the average accuracy of the method was 0.79 [13].

Combining CNN and SVM is a common strategy, and by combining the advantages of both, the performance of image recognition tasks can be improved. Many experts have made some achievements in the field of combining CNN and SVM. For example, Ye proposed a method to improve CNN by SVM to solve the problem that image data cannot be processed with low capacity and depth in 3D Lidar visual position recognition technology, and the results showed that the method was effective [14]. To solve the problem of low accuracy of MRI image classification of brain cancer, Khairandish's team proposed a classification model based on CNN and SVM. Through comparative analysis and experiment with similar models, the accuracy of this model was 98.4959%, which was better than the comparison model [15]. To solve the problem of low accuracy of ECG image type recognition, Ozaltin and Yeniay proposed an image recognition method based on CNN-SVM. The validity experiment verified that the highest accuracy rate of this method was 99.21% [16]. To solve the problem of low classification efficiency of bread wheat varieties, Yasar proposed a classification model of bread wheat combining CNN and SVM. Through empirical experiments, the results showed that the highest classification accuracy of this model was 97.51% [17]. To solve the problem of low accuracy of skin image recognition, Anggriandi et al. proposed a skin image recognition method based on CNN and SVM. Through experimental verification, the classification accuracy and recall rate of this method were 93.55% and 93.74%, respectively [18].

In summary, there are few identification methods applied to birds endangered by transmission lines at present, and the existing CNN-SVM algorithm still has low accuracy in image recognition. However, there are still few methods to improve image feature extraction by combining backbone networks such as DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0. To solve the problem of low accuracy of current image recognition methods, a method combining multi-trunk feature extraction network and SVM was studied for transmission line bird recognition, and a model of transmission line bird recognition based on improved CNN and SVM was constructed. This model not only improved the accuracy of image recognition, but also broke through the limitations of previous qualitative studies, and had strong potential application value.

III. Methods and Materials

A. *Design of Feature Extraction Network Based on Improved CNN*

Bird hazards are a major cause of transmission line failures, ranking third after lightning strikes and external damage [19]. Identifying hazard birds and assisting transmission line inspectors in identifying birds that may pose a threat to the line has become an urgent problem to be addressed. CNN is a deep learning framework that has strong feature extraction capabilities and flexibility, and is broadly utilized in the computer vision [20]. However, due to the fixed architecture and parameters used by CNN, it is unable to fully capture all the information in the data, thereby limiting the expressive power of features [21]. In addition, if the model structure is too complex or the training samples are insufficient, it may also lead to overfitting [22]. An improved CNN-based feature extraction network is developed for identifying bird hazards in transmission lines. Subsequently, it is combined with SVM algorithm to construct a transmission line hazard bird recognition model that integrates improved CNN and SVM. Before building a bird identification model for transmission line hazards, it is necessary to construct an improved CNN. The basic structure of CNN is denoted in Fig. 1.

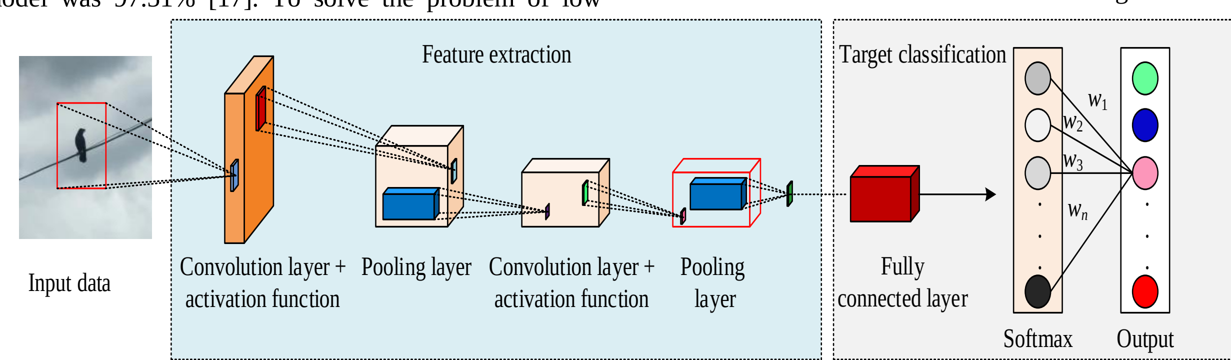


Fig. 1. The basic structure of CNN.

From Fig. 1, the basic structure of CNN is mainly composed of convolutional layers, activation layers, pooling layers, and fully connected layers. Among them, w is the weight, and the calculation expression for $Softmax\ function$ is shown in (1).

$$softmax\ function(x_i) = exp(x_i) / \sum_{j=1}^{n} exp(x_i) \tag{1}$$

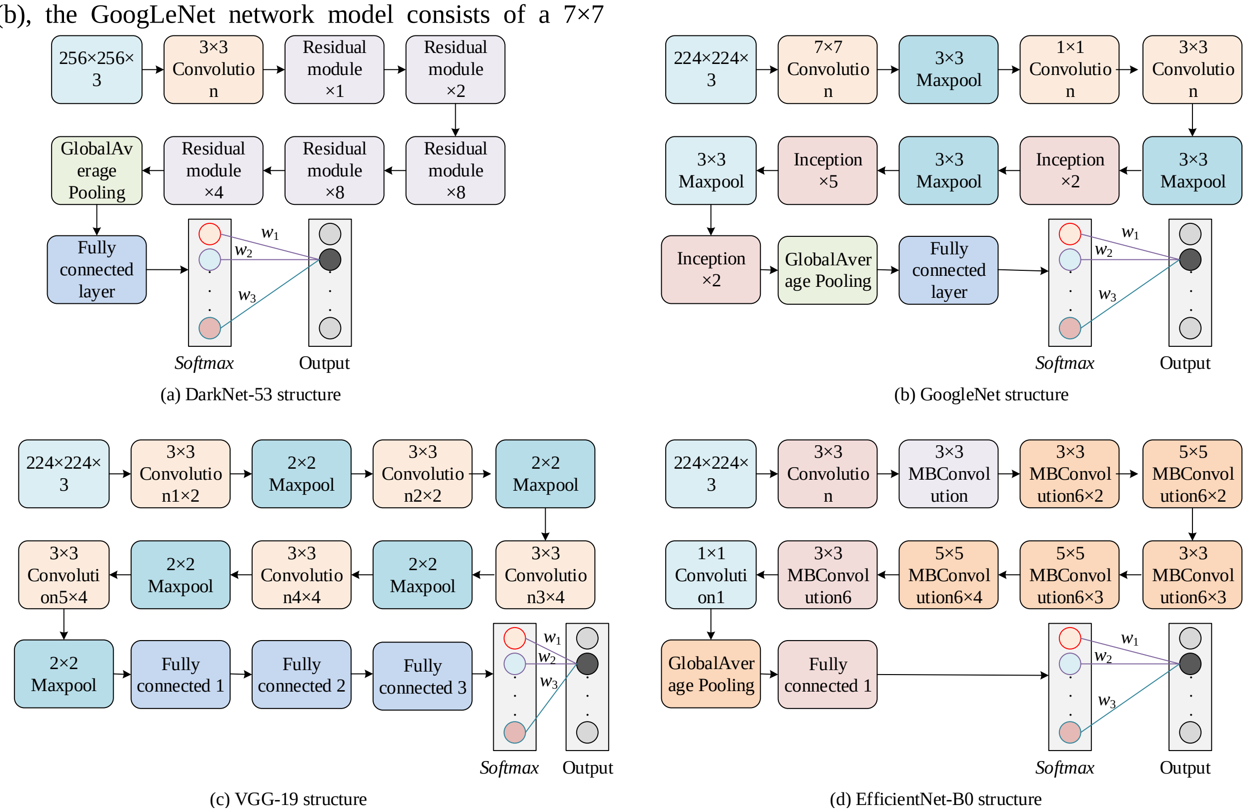
In (1), x_i means the i th element of the input vector, $softmax(x_i)$ represents the output, and n represents the dimension of the vector. However, due to the fixed architecture and parameters used by CNN, it is unable to fully capture all the information in the data, thereby limiting the expressive power of features. To solve this problem, research combined with relevant literature analysis find that there are many types of networks derived from CNN. Therefore, based on the characteristics of these networks, an improved CNN network is constructed. Firstly, DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0 are used to extract features from images of bird hazards caused by transmission lines. The basic structures of each network are shown in Fig. 2.

From Fig. 2(a), the feature extraction part of DarkNet-53 network consists of a 3×3 convolutional layer and five residual blocks. Each residual block is downsampled using a 1×1 convolution, achieving a total of five dimensionality reductions to control the dimensionality of the feature channel. From Fig. 2(b), the GoogLeNet network model consists of a 7×7 convolutional layer and an $Softmax$ output layer. To capture features at different scales, the network adopts the Inception module. From Fig. 2(c), VGG-19 consists of 19 parameterized layers. The network structure contains 16 convolutional layers and 3 fully connected layers. According to Fig. 2(d), the network architecture consists of 16 Mobile Inverted Bottleneck Convolution (MBConv) units and a 1×1 convolutional layer. Secondly, to better understand the impact of feature maps on the final classification decision, the study introduces gradient weighted class activation mapping for visual analysis of the model. Gradient weighted class activation mapping generates a class activation map by calculating weights, revealing the influence of each pixel in the feature map on the classification probability gradient. The weight calculation can be represented by (2).

$$\omega_z^m = \frac{1}{k} \sum_i \sum_j \frac{\partial y^m}{\partial A_{ij}^z} \tag{2}$$

In (2), z and k respectively represent the sequence number and number of current feature maps, y^m refers to the score of category m, and A_{ij}^z denotes the pixel value of the feature map. The generation of the class activation mapping $L_{Grad-CAM}^m$ can be represented by (3).

$$L_{Grad-CAM}^m = ReLU\left(\sum_z \omega_z^m A^z\right) \tag{3}$$



(a) DarkNet-53 structure

(b) GoogleNet structure

(c) VGG-19 structure

(d) EfficientNet-B0 structure

Fig. 2. Basic network structure.

Next, the study uses cascaded fusion to perform convolutional feature fusion on the above four network structures. In this fusion network, the part that extracts convolutional features from bird images is the trained four networks. The Dropout layer of the last fully connected layer in the trained VGG-19-C, along with the global average pooling layers of DarkNet-53-C, GoogleNet-C, and EfficientNet-B0-C, is used for feature extraction. So, the convolutional features extracted by DarkNet-53-C, GoogleNet-C, EfficientNet-B0-C, and VGG-19-C networks can be set as F_D, F_G, F_V, and F_E, respectively, and each convolutional feature dimension is 1024, 1024, 1280, and 4096 dimensions, respectively. The convolutional fusion feature F obtained by cascading fusion can be represented by (4).

$$F = Concatenate(F_D, F_G, F_V, F_E) \tag{4}$$

By concatenating features from different levels, the output of each layer is sequentially passed on to the next layer as input, gradually extracting and integrating richer information. Finally, transfer learning is used to fine tune the test and training sets into 30% and 70% structures. Finally, based on the above content, a feature extraction network based on improved CNN is constructed, and the network process is indicated in Fig. 3.

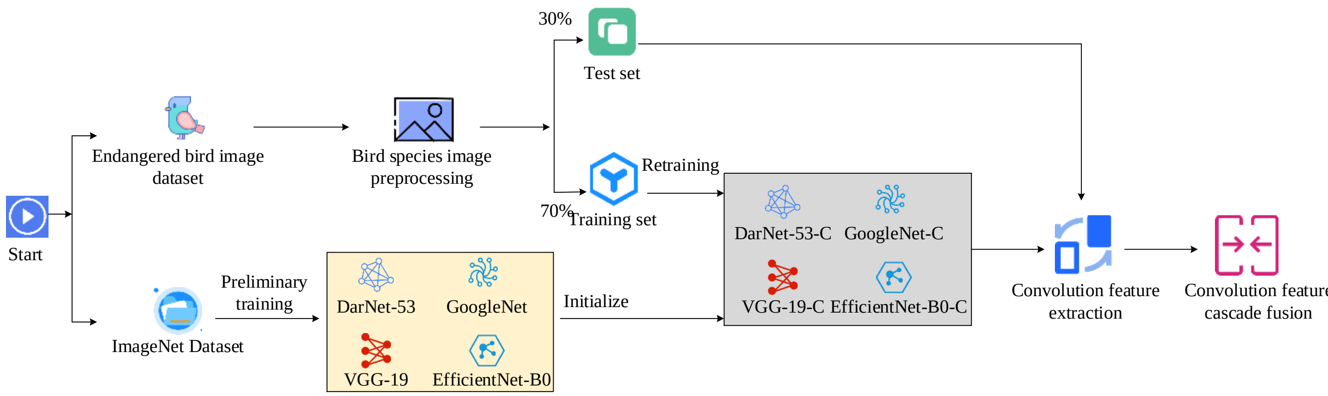


Fig. 3. Feature extraction network based on improved CNN.

The feature extraction of the network can be obtained from Fig. 3. Firstly, four networks are pre-trained using hazard bird data in the ImageNet dataset, and the four networks are initialized to obtain the trained four CNNs. Secondly, the preprocessed dataset of bird images of power line hazards will be preprocessed, and the preprocessed data will be divided into test and training sets with 30% and 70% ratios, respectively. Then, the four networks are retrained using the training set. Finally, the processed 30% test set and the convolutional features extracted by the four trained networks are cascaded and fused to obtain the extracted hazard bird features.

B. *Construction of a Hazard Bird Classification and Recognition Algorithm Integrating Improved CNN and SVM*

The prerequisite for extracting and recognizing bird characteristics is to understand the types and image data of birds that are hazard to transmission lines, and to construct a dataset of birds that are hazard to transmission lines. Therefore, based on the collection of bird records that have caused faults in transmission lines in the past, the study summarized a total of 80 hazard bird species and types of faults involved, among which 20 high-risk bird species are shown in Table I.

Note: Fault type: Bird droppings (Dung); Bird's nest (Nest); Bird body contact (Catch); Bird peck (Peck).

From Table I, there are a total of 20 bird species that pose a threat to transmission lines, including ciconia nigra, egretta garzetta, and pond herons. Based on the above bird species, the study utilizes web crawling technology to collect a massive amount of bird image data from the internet, covering images of birds in various environmental conditions, target sizes and quantities, and other different contexts. Due to the uneven sample size of the collected bird images, the model may have poor recognition performance for these birds. To solve this problem, the specific steps of data augmentation processing on images are to first randomly scale and rotate the images. Secondly, fogging is performed on the image, which can be represented by (5).

$$new\_pixel = old\_pixel \cdot td + U \cdot (1 - td) \tag{5}$$

TABLE I. TRANSMISSION LINE HAZARDS TO BIRDS AND FAULT TYPES

Name	Fault type	Name	Fault type
Ciconia nigra	Dung, nest, catch	*Pica pica*	Dung, nest, pick, peck
Ciconia boyciana	Dung, nest, catch	*Pycnonotus sinensis*	Dung
Egretta garzetta	Dung, nest	*Oriolus chinensis*	Dung
Ardeola bacchus	Dung, nest	*Hirundo rustica*	Dung
Falco tinnunculus	Dung, nest	*Anser cygnoides*	Catch, dung
Sturnus nigricollis	Dung, nest	*Asio otus*	Catch, dung
Spodiopsar sericeus	Dung, nest	*Spilopelia chinensis*	Dung, nest, catch
Acridotheres cristatellus	Dung, nest	*Cuculus canorus*	Dung
Cyanopica cyanus	Dung, nest, catch	*Otis tarda*	Catch
Corvus macrorhynchos	Dung, nest, catch	*Upupa epops*	Dung

In (5), U represents the brightness of fog, new_pixel and old_pixel represent the brightness of new and original pixels, respectively, and td is variable. The calculation expression for td is shown in (6).

$$td = exp(-\beta \cdot td) \tag{6}$$

In (6), d and β represent the distance from the pixel to the center of the mist and the concentration of the mist, respectively. Next, a linear transformation is performed, and its calculation expression is denoted in (7).

$$r_n(l,h) = \alpha g(l,h) + (1-\alpha) g_0 + \varsigma \tag{7}$$

In (7), g_0 and ς represent the zero pixel image with the same $g(l,h)$ and the added pixel value, $g(l,h)$ and $r_n(l,h)$ represent the original image and the converted image, respectively, and α represents the original image multiple. Finally, the image is denoised using a Denoising Convolutional Neural Network (DnCNN) and labeled with the image annotation software LabelImg before saving. The construction of a dataset on bird species affected by transmission lines is completed. After completion, an improved CNN-based feature extraction network is used to extract features, and the birds affected by transmission lines are classified and recognized. To avoid the problem of overfitting in CNN, SVM is used for classification and recognition. SVM segments samples of different categories by finding a hyperplane in the feature space, maximizing the distance from the nearest sample point on the hyperplane to the hyperplane. The calculation expression for this hyperplane is shown in (8).

$$\varpi \mu + b = 0 \tag{8}$$

In (8), ϖ is the weight, b is the intercept, and μ is the eigenvector. However, in some cases, SVM cannot find the hyperplane. To solve this problem, SVM uses kernel functions to map data to a linearly separable high-dimensional feature space, and the hyperplane of this high-dimensional space can be represented by (9).

$$f(\mu) = \varpi^T \phi(\mu) \tag{9}$$

In (9), $\phi(\mu)$ is the feature vector after μ mapping. Thus, the classification problem can be transformed into a quadratic programming problem, which can be represented by (10).

$$\begin{cases} \min\limits_{\varpi,b,\zeta} \ \frac{1}{2}\|\varpi\|^2 + H\sum\limits_{\gamma=1}^{M}\zeta_\gamma \\ s.t\ p_\gamma[\varpi^T\phi(\mu_\gamma + b] \geq 1-\zeta_\gamma, \zeta_\gamma \geq 0, \gamma = 1,2,\ldots,M \end{cases} \tag{10}$$

In (10), H is the penalty coefficient, ζ_γ is the relaxation variable, and p_γ is the category label of the γ th sample point. Introducing Lagrange multipliers to simplify it can be represented by (11).

$$L(\varpi,b,\zeta,\tau,\psi) = \frac{1}{2}\|\varpi\|^2 + H\sum_{\gamma=1}^{M}\zeta_\gamma + \sum_{\gamma=1}^{M}\tau_\gamma\{1-\zeta_\gamma - p_\gamma[\varpi^T\phi(\mu_\gamma)+b]\} - \sum_{\gamma}^{M}\psi_\gamma\zeta_\gamma \tag{11}$$

In (11), τ and ψ are Lagrange multipliers, respectively. By taking the derivative of each variable using the Lagrange function and making it zero, a set of candidate values can be obtained, and then the optimal value can be verified. So, the quadratic programming problem can be transformed into a Lagrangian dual problem, which can be represented by (12).

$$\begin{cases} \max\limits_{\tau} \ \sum\limits_{\gamma=1}^{M}\tau_i - \frac{1}{2}\sum\limits_{\gamma=1}^{M}\sum\limits_{j=1}^{M}\tau_\gamma\tau_j p_\gamma p_j \phi(\mu_\gamma)^T\phi(\mu_j) \\ s.t\ \sum\limits_{\gamma=1}^{M}\tau_\gamma p_\gamma = 0, 0 \leq \tau_\gamma \leq H, i = 1,2,\ldots,M \end{cases} \tag{12}$$

Thus, by solving it, the expression of the support vector decision function can be obtained as shown in (13).

$$f(\mu) = \sum_{\gamma=1}^{M}\tau_\gamma p_\gamma \kappa(\tau,\tau_\gamma) + b \tag{13}$$

In (13), $\kappa(\tau,\tau_\gamma)$ is the kernel function. Therefore, the study combines SVM with a feature extraction network based on improved CNN to construct a classification algorithm based on improved CNN and SVM, as shown in Fig. 4.

From Fig. 4, the specific classification of the algorithm is to first encode. In this process, an encoding matrix S with a value of $\{+1,0,-1\}$ needs to be constructed, and the behavior s of the encoding matrix represents the number of categories and category labels. The column is $s(s-1)/2$, and its vector represents the binary classifier. For the j th column of the matrix, if the values of $S[\lambda_1, j]$ and $S[\lambda_2, j]$ are +1 and -1, respectively, and the other elements are 0, then the binary classifier for this column is used to distinguish between λ_1 and λ_2, where λ represents the category. Then, in the training process, the feature vectors and labels of different bird species of different categories are used as input values, and SVM is used to perform two class classification training on the species, thereby obtaining $s(s-1)/2$ trained SVMs. After predicting all test samples through a classifier, an output vector $J(x) = [\eta_1(x), \eta_2(x), \eta_4(x), \ldots \eta_{s(s-1)}(x)]$ is generated, with η being the output value. The value of each element is defined as -1 or +1. Finally, decoding is performed. During decoding, the Hamming distance decoding method is used to determine the Hamming distance dr from $J(x)$ to each row of S, and the category corresponding to the shortest χ_δ is selected as the predicted output. The calculation expression for Hamming distance decoding is shown in (14).

$$\chi_\delta = \sum_{\gamma=1}^{s(s-1)/2}\frac{|S(\lambda,\gamma) - J_\gamma(x)|}{2}, \lambda \in \{1,2,\ldots,s\} \tag{14}$$

In (14), $J_{\gamma}(x)$ is the value of the output vector $J(x)$ relative to the test sample x, and $S(\lambda,\gamma)$ is the value of the γ th element in row λ of S. Finally, based on the above content, a bird recognition algorithm for transmission line hazards is constructed using improved CNN and SVM. The algorithm flow is shown in Fig. 5.

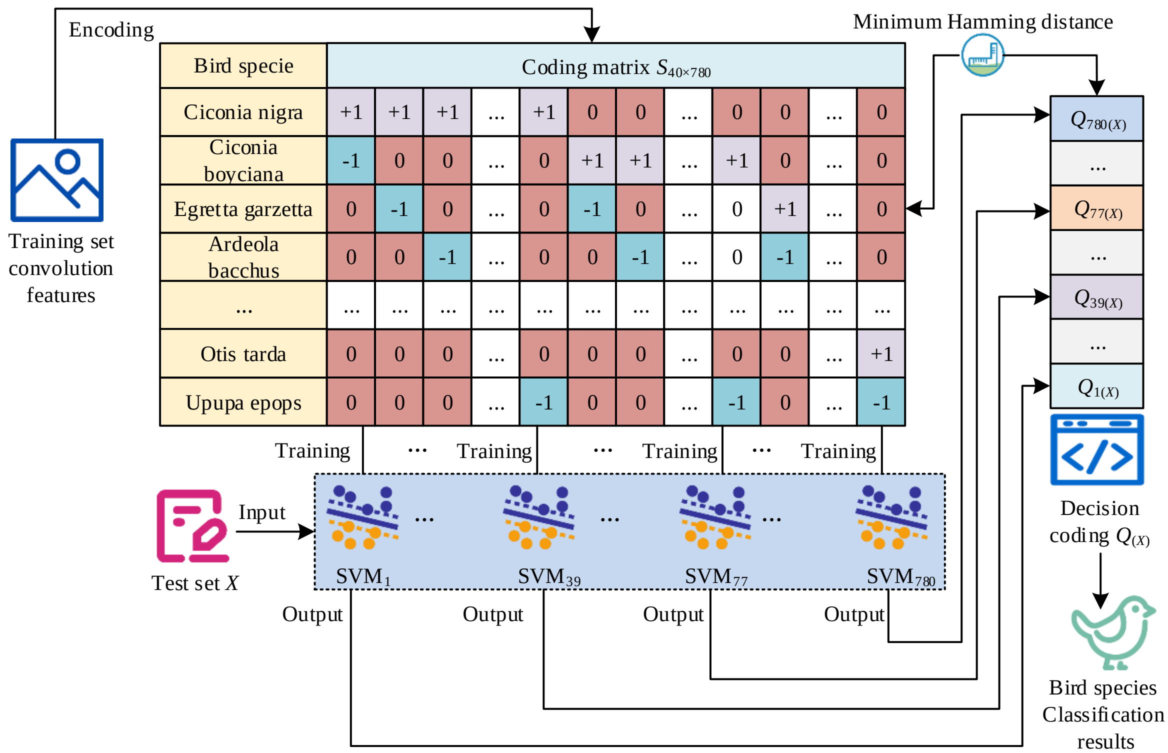


Fig. 4. Classification model based on improved CNN and SVM.

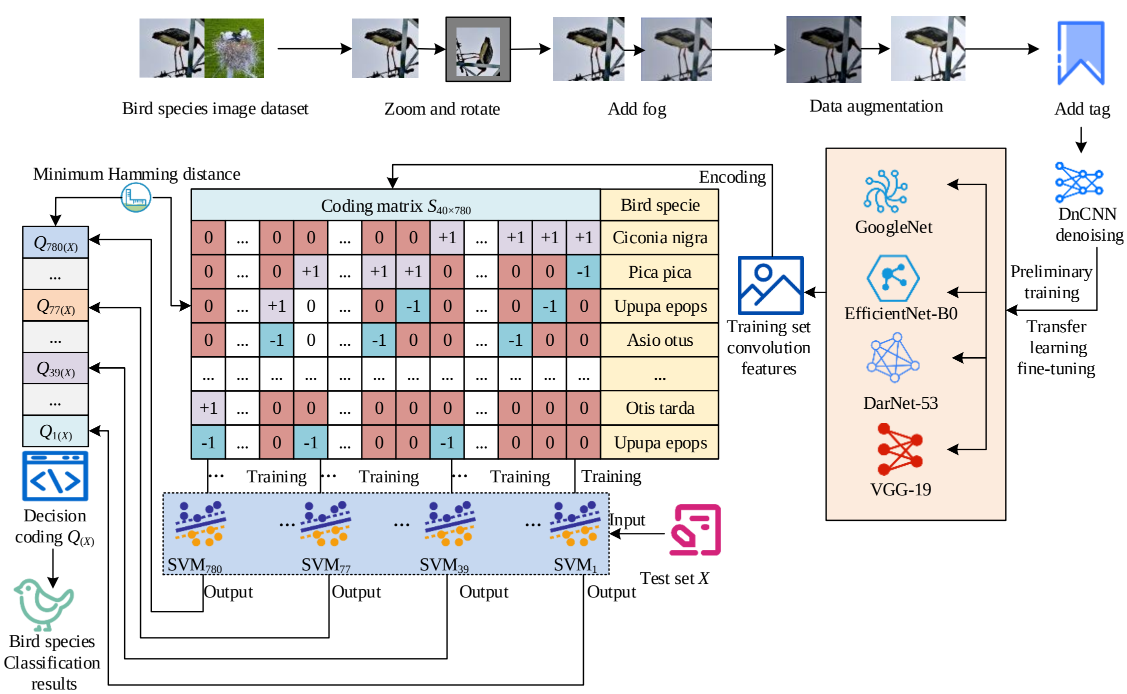


Fig. 5. Algorithm flow of transmission line hazard bird identification based on improved CNN and SVM.

From Fig. 5, the specific process of the recognition algorithm can be seen. Firstly, the image is preprocessed using image scaling, rotation, fogging, and DnCNN to increase image pixels and remove image noise. Secondly, the convolutional feature fusion method is used to cascade and fuse the four CNN models, DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0, to raise the robustness and feature extraction ability of the models. Then, using transfer learning theory, the model is fine-tuned through the training set to achieve the optimal state. Finally, SVM is used to find the optimal Hamming distance, which is used as the solution for the classification result, thus achieving the effect of classification recognition.

IV. Results

A. Algorithm Performance Analysis

To validate the superiority of the raised algorithm, a performance comparison analysis experiment was conducted with other algorithms. The experiment was simulated using Matable software, and the algorithm was trained before the experiment. After training, specific parameter settings were obtained: the output channels of DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0 were set to 40, and the final output layer of the original network layer was replaced by a 40 class output layer. After optimization using momentum stochastic gradient descent algorithm, the batch processing size was obtained to be 64, the initial learning rate was 1×10^{-4}, the momentum value was 0.9, and the regularization factor in the network was set to 1×10^{-4}. DnCNN adopts $ReLU$ activation function and optimizes with momentum stochastic gradient descent algorithm to obtain an initial learning rate of 0.1, momentum value of 0, and batch processing size of 64. The source of the dataset consists of two parts, one of which is live images captured by surveillance cameras installed in and near transmission line towers in Anhui Province, China. The other part is the target images of birds with similar scenes collected from the network by crawler technology. In the end, the dataset contained a total of 6,876 images, of which 80% was used as a training set and 20% as a test set. In the pre-processing stage, the image was randomly scaled, rotated and fogged to increase the data diversity, and the image quality was improved by DnCNN. The annotation work was manually completed by LabelImg software, and the annotation results were saved as corresponding annotation files for subsequent model training. Since the data mainly come from field shooting and network crawling in specific areas, there may be scene bias, regional bias and category bias. These biases can lead to limitations in the model's ability to generalize, especially when dealing with images from other regions or different scenes. In addition, the class imbalance problem may affect the model's ability to recognize a few classes, thereby reducing the overall performance. To solve this problem, the research adopted data enhancement methods such as random scaling, rotation and fog processing to increase data diversity. A few classes of samples were also over-sampled to increase their proportion in the data set to alleviate the problem of data imbalance. The experimental comparison algorithms included Faster-RCNN, EF-YOLOv5, YOLOv7-BiFormer, and the experimental comparison indicators included accuracy, precision, and recognition speed. The specific experimental environment is indicated in Table II.

TABLE II. EXPERIMENTAL ENVIRONMENTAL CONFIGURATION

Parameter names	Parameter
Processor	Intel Core i9-13900K
Main frequency	5.8 GHz
Internal memory	32 GB
Hard disk capacity	1 TB
Operating system	Windows 10 64
Matlab version	Matlab 2021a
Data analysis software	Spss24.0

To verify the benchmark performance and hardware requirements of the algorithm proposed in the research, a benchmark test was conducted between the algorithm and other algorithms with the indexes of each image recognition time, model size, accuracy improvement, scalability, and hardware requirements. The test results are shown in Table III.

TABLE III. TEST RESULTS

Algorithm	Image recognition (ms/pcs)	Model size (MB)	Processor	Improves accuracy (%)	Scalability	Hardware requirement
Research	19.8	41	Intel Core i9-13900K	+43.6	High	Medium
Faster-RCNN	74.6	56	Intel Core i9-13900K	+27.3	Medium	High
EF-YOLOv5	40.3	47	Intel Core i9-13900K	+32.1	Low	High
YOLOv7-BiFormer	51.7	44	Intel Core i9-13900K	+29.5	Medium	High

From Table III, the recognition time of the proposed algorithm, Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer for each image was 19.8 ms, 74.6 ms, 40.3 ms, and 51.7 ms, respectively, among which the proposed algorithm had the shortest recognition time for each image. This showed that the proposed algorithm had obvious advantages in processing speed and was suitable for application scenarios requiring fast response. In terms of model size, the proposed algorithm was 41 MB, which was lower than the 56 MB of Faster-RCNN, 47 MB of EF-YOLOv5 and 44 MB of YOLOv7-BiFormer. Smaller model sizes helped reduce storage requirements and potentially increased deployment flexibility. In addition, the accuracy of the proposed algorithm was improved to 43.6%, which was significantly higher than other algorithms. This showed that the algorithm could provide high recognition accuracy while maintaining high recognition speed. In terms of scalability, the proposed algorithm was rated as high, which indicates that the proposed algorithm can adapt to different application requirements and environments, and has good adaptability. The medium hardware requirement indicated that the algorithm required neither the lowest nor the highest hardware, thus striking a balance between performance and cost. The above results showed that the proposed algorithm had the best performance in recognition speed and accuracy, high scalability, and low hardware requirements, and could achieve real-time recognition well. To verify the contribution of each component of the model proposed in the study to the performance improvement, ablation experiments were conducted on it, and the experimental results are shown in Table IV.

TABLE IV. ABLATION RESULTS

Experiment No.	Model architecture	Feature extraction network	Feature fusion method	Classifier	Accuracy (%)
1	CNN-only	DarkNet-53	None	CNN	82.5
2	CNN-only	GoogleNet	None	CNN	80.0
3	CNN-only	VGG-19	None	CNN	78.5
4	CNN-only	EfficientNet-B0	None	CNN	81.0
5	CNN-only	Multiple network fusion	None	CNN	83.0
6	CNN+SVM	DarkNet-53	None	SVM	84.0
7	CNN+SVM	GoogleNet	None	SVM	82.0
8	CNN+SVM	VGG-19	None	SVM	81.5
9	CNN+SVM	EfficientNet-B0	None	SVM	86.5
10	CNN+SVM	Multiple network fusion	Multiple convolution fusion	SVM	96.6

From the ablation experiment results in Table IV, a single CNN model had different performances in bird recognition tasks, among which DarkNet-53 had the best performance, with an accuracy rate of 82.5%. After the introduction of SVM classifier, the model performance was generally improved, especially the EfficientNet-B0+SVM combination, the accuracy rate increased to 86.5%, indicating that SVM has significant advantages in the feature classification stage. After further use of multi-network fusion and multi-convolutional feature fusion methods, the model accuracy was significantly improved to 96.6%, indicating that feature fusion technology contributes significantly to the performance improvement. Therefore, the CNN+SVM model architecture combined with multi-network fusion and feature fusion had the best performance in the identification of transmission line endangered birds, which provides an effective way to improve the identification accuracy.

In the above environment, firstly, 1000 bird images were selected and four algorithms were used to classify and recognize 8 high-risk birds on transmission lines. The classification and recognition results were represented by a confusion matrix. The classification accuracy results of each algorithm are indicated in Fig. 6.

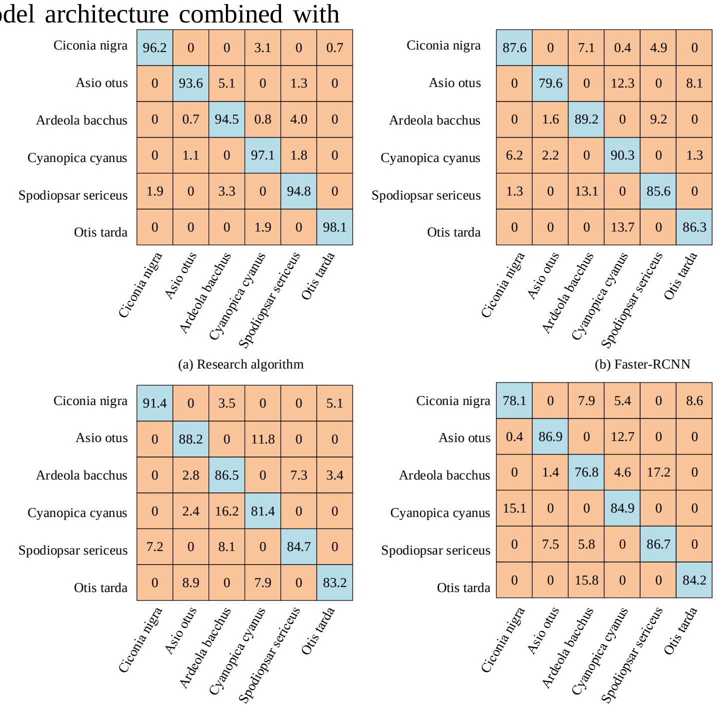
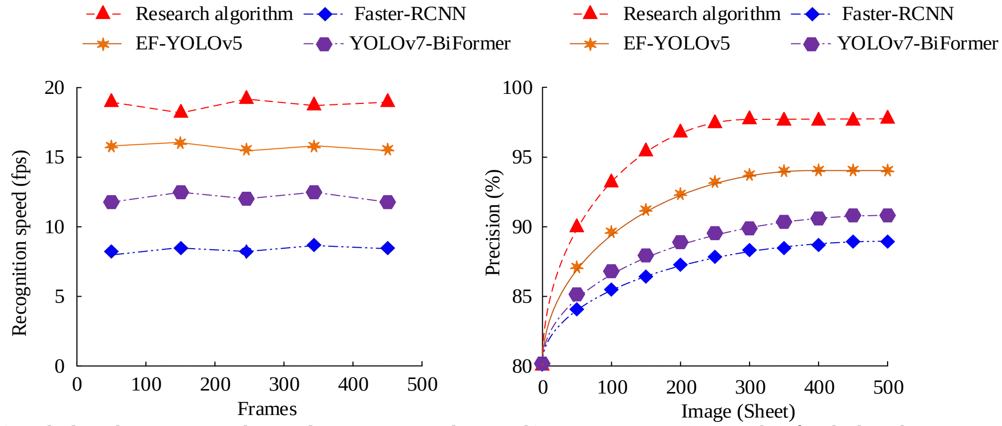


Fig. 6. Comparison results of classification accuracy of each algorithm.

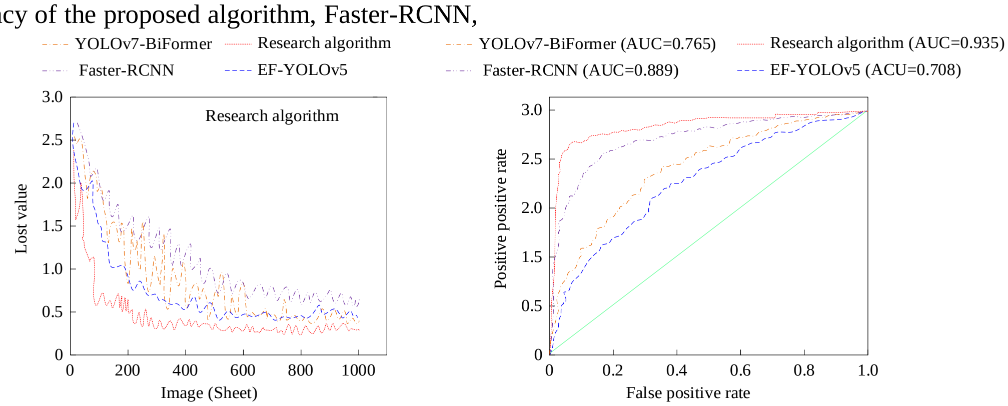
From Fig. 6(a), the proposed algorithm had recognition accuracies of 96.2%, 93.6%, 94.5%, 97.1%, 94.8%, and 98.1% for ciconia nigra, asio otus, ardeola bacchus, cyanopica cyanus, spodiopsar sericeus, and otis tarda, respectively, all of which were above 90%. The average recognition accuracy of Fig. 6(b), 6(c), and 6(d) was all below 90%, significantly lower than the average recognition accuracy of Fig. 6(b). The above results indicate that, from the perspective of recognition accuracy, the proposed recognition algorithm is significantly better than the comparative algorithm. The recognition speed and accuracy results of each algorithm are shown in Fig. 7.



(a) Each algorithm recognizes the speed comparison results (b) Precision comparison results of each algorithm

Fig. 7. Results of recognition speed and recognition accuracy of each algorithm.

From Fig. 7(a), the average recognition speed of the proposed algorithm was 19.8 frames per second (FPS), the average recognition speed of Faster RCNN was 8.2 FPS, the average recognition speed of EF-YOLOv5 was 16.6 FPS, and the average recognition speed of YOLOv7-BiFormer was 13.1 FPS. Among them, the algorithm proposed in the study had the fastest average recognition speed. From Fig. 7(b), the recognition accuracy of the proposed algorithm, Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer were 97.4%, 86.4%, 94.3%, and 93.7%, respectively. Among them, the algorithm proposed in the study had the highest recognition accuracy. The above results indicate that, in terms of recognition speed and accuracy, the proposed algorithm outperforms the compared algorithms in terms of performance. The loss values and ROC curve results of each algorithm are shown in Fig. 8.



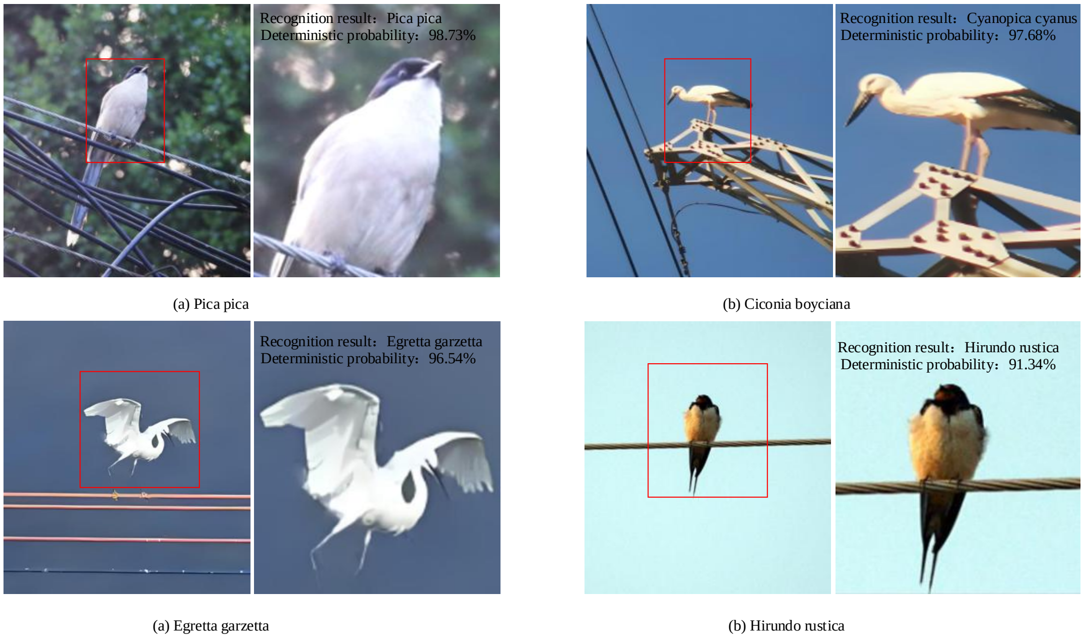
(a) The loss value of each algorithm is compared (b) ROC curve results of each algorithm

Fig. 8. Loss values and ROC curve results for each algorithm.

From Fig. 8(a), the proposed algorithm converged first with an average loss value of 0.37, Faster-RCNN had an average loss value of 0.9, EF-YOLOv5 had an average loss value of 0.52, and YOLOv7-BiFormer had an average loss value of 0.67. According to Fig. 8(b), the AUC values of the proposed algorithm, Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer were 0.935, 0.889, 0.708, and 0.765, respectively, with the proposed algorithm having the highest AUC value. The lower the loss value in the range of 0.1 to 1, the better the model's generalization ability. The higher the AUC value of the ROC curve in the range of 0.5 to 1, the stronger the model's discriminative ability. Based on the loss value and ROC curve dimensions, the proposed algorithm outperformed the compared algorithms in terms of performance. In summary, from the perspectives of accuracy, precision, recognition speed, loss value, and ROC curve dimensions, the proposed algorithm outperforms the compared algorithms in terms of performance and is effective.

B. *Analysis of Algorithm Application Effectiveness*

After verifying the performance superiority of the algorithm, an application effect analysis experiment was conducted on the proposed algorithm. The study randomly captured images of hazard birds on transmission lines in a certain area for identification, and some of the identification results are shown in Fig. 9.

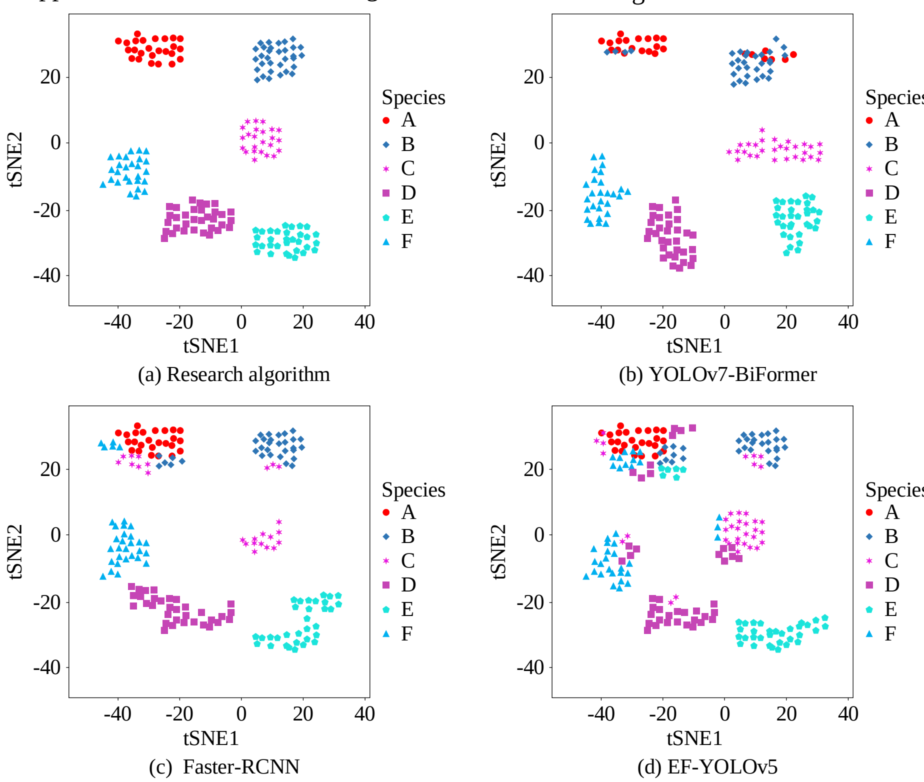


(a) Pica pica (b) Ciconia boyciana

(a) Egretta garzetta (b) Hirundo rustica

Fig. 9. Identification results of bird parts of transmission lines.

From Fig. 9, the recognition algorithm proposed in the study had recognition results and confirmation probabilities of 98.73%, 97.68%, 96.54%, and 91.34% for magpies, ciconia boyciana, egretta garzetta, and hirundo rustica, respectively, all of which were above 90%. This result indicates that the proposed recognition algorithm can effectively identify birds that pose a threat to transmission lines and has practical value. To further verify the application effect of the recognition algorithm proposed in the research, the classification and recognition of randomly captured birds were studied. The t distribution random neighborhood embedding technique was used to select six bird species for visual analysis and comparative experiments. The comparative algorithms included Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer algorithms. The visual recognition results of each algorithm are denoted in Fig. 10.



(a) Research algorithm (b) YOLOv7-BiFormer

(c) Faster-RCNN (d) EF-YOLOv5

Fig. 10. Visualization of recognition results for each algorithm.

From Fig. 10(a), the algorithm proposed in the study had the best classification performance, with the highest clustering degree for each category. From Fig. 10(b), YOLOv7-BiHomer had good classification performance, with two categories not clearly distinguished. From Fig. 10(c), Faster-RCNN had poor classification performance, with four categories confused and not significantly distinguished. From Fig. 10(d), EF-YOLOv5 had the worst classification performance, with six bird species confused. The above results indicate that the proposed algorithm has the best visualization effect and good application performance.

V. Discussion

This study conducted comparative experimental analysis on the performance of bird recognition algorithms for transmission line hazards based on improved CNN and SVM, and conducted application effect analysis experiments on the algorithm. The findings denoted that the algorithm had significant advantages in accuracy, precision, and recognition speed. In the accuracy comparison experiment, the proposed algorithm achieved recognition accuracies of 96.2%, 93.6%, 94.5%, 97.1%, 94.8%, and 98.1% for ciconia nigra, asio otus, ardeola bacchus, cyanopica cyanus, spodiopsar sericeus, and otis tarda, respectively, all of which were above 90%, significantly better than the comparison algorithms. This result indicates that the introduction of DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0 raises the algorithm's ability to extract deep features and optimizes the accuracy of algorithm recognition. This result is similar to the improved CNN algorithm proposed by Elangovan et al. [23]. In the precision comparison experiment, the recognition accuracy of the proposed algorithm, Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer were 97.4%, 86.4%, 94.3%, and 93.7%, respectively, with the proposed algorithm having the highest recognition accuracy. This result indicates that the introduction of SVM algorithm raises the classification accuracy of the algorithm. The Chaudhari team reached consistent conclusions in their research on combining SVM and CNN [24]. In the recognition speed comparison experiment, the average recognition speeds of the proposed algorithm, Faster-RCNN, EF-YOLOv5, and YOLOv7-BiFormer were 19.8 FPS, 8.2 FPS, 16.6 FPS, and 13.1 FPS, respectively. Among them, the algorithm proposed in the study had the fastest average recognition speed. This outcome indicates that the convolution feature fusion of DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0, as well as the introduction of SVM, improved the computational efficiency of the algorithm. At the same time, in the comparative experiments of loss value and ROC, the average loss value and AUC value of the proposed algorithm were 0.37 and 0.935, respectively, which were better than the comparative algorithms. This result further validates the superiority of the algorithm proposed in the study. The Okomba team reached similar conclusions in SVM-CNN related research [25]. Secondly, in the application effect analysis experiment, the algorithm proposed in the study had good application effects in identifying hazard birds and visualizing the results. In the experiment of identifying hazard birds, the results showed that the proposed identification algorithm could effectively identify hazard birds. This result indicated that the convolutional feature fusion of DnCNN, DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0, as well as the introduction of SVM, improved the accuracy and precision of the algorithm for bird recognition on transmission lines. In the visual classification comparative analysis experiment, the algorithm proposed in the study had the best classification performance and the highest degree of category aggregation. This conclusion is similar to the one obtained by the Gao team in their relevant research in 2022 [26]. To conduct a thorough analysis of the performance of the model in practical applications, a detailed discussion was conducted on the failure cases of the model. Research has found that common classification errors included difficulty in detecting small targets, misclassification caused by occlusion, and the negative impact of lighting changes on detection performance. For example, small birds are easily misclassified or missed due to occupying fewer pixels in the image; partially occluded bird targets often lead to inaccurate recognition by the model. In addition, changes in lighting conditions (such as shadows and highlights) can mask key features of birds, further reducing the detection accuracy of the model. To address these issues, techniques such as data augmentation, feature enhancement, and multimodal data fusion can be used to improve the adaptability of the model to complex environments. However, factors such as dynamic scenes, seasonal changes, and environmental noise in the real world still pose challenges to the generalization ability of the model. Therefore, future work needs to further optimize the model architecture, regularly update the dataset, and combine online learning strategies to improve the robustness and accuracy of the model in practical deployment. In addition, there are limitations in the research data set and potential biases in the data collection process, such as data mainly from Anhui province, which may lead to poor adaptability of the model to other regions. In addition, the large number of certain bird samples in the dataset may result in the model being better at identifying these birds and less able to identify others. This bias can have an unfair impact on grid maintenance and bird protection, for example, false positives can lead to unnecessary waste of resources, while missed positives can pose a threat to grid security. Future work will reduce this bias by introducing data from more regions and optimizing the balance of the dataset, and its ethical implications will be discussed in detail in the paper. The following ethical issues and practical application challenges arise in the study of transmission line hazard bird identification model based on multi-backbone network and SVM. First, the geographic limitations of the data set and sample imbalances can lead to model bias, which in turn affects the fairness of grid maintenance and the comprehensiveness of bird conservation. Second, models in real-world deployments can be disturbed by dynamic scenarios, seasonal changes, and ambient noise, leading to false positives and triggering unwanted interventions. Finally, to address these challenges, future work will optimize the diversity and balance of datasets, reduce false positives by combining multi-sensor data and real-time verification mechanisms, and promote harmonious symbiosis between the grid and birds through ecological conservation measures and public education.

VI. Conclusion

To address the problem of low recognition accuracy in bird identification methods for transmission line hazards due to the small size of bird targets, this study introduced CNNs to solve the problem of CNN being unable to fully capture all information in the data due to its fixed architecture and parameters, which limits its ability to express features. After fine-tuning and migration learning, four CNN models, DnCNN, DarkNet-53, GoogleNet, VGG-19, and EfficientNet-B0, were cascaded for feature fusion and improvement to construct an improved CNN feature extraction network. To accurately identify hazard birds, SVM was introduced for classification. A dataset of bird species affected by transmission lines was constructed, and data augmentation methods and DnCNN were introduced for noise reduction processing of bird image data. The above classification algorithms were applied to this dataset and a bird classification and recognition algorithm for transmission line hazards was constructed based on improved CNN and SVM. Comparative performance analysis experiments were conducted on the algorithm, and the results showed that the algorithm performed significantly better than the compared algorithms in terms of accuracy, precision, recognition speed, loss value, and ROC curve dimensions. Subsequently, the algorithm was subjected to application effect analysis experiments, and the results showed that the algorithm not only accurately identified hazard birds, but also had better classification performance than the comparative algorithms in visualization effect analysis. The above findings denote that the proposed algorithm has strong robustness. The limitation of this study is that the recognition method adopts a multi-CNN structure for fusion, which may have redundant parameters and a large number of parameters. Therefore, further research is needed to reduce the dimensionality of the fused features to shorten the algorithm recognition time and improve recognition speed.

References

[1] Luo Y, Yu X, Yang D, Zhou B, “A survey of intelligent transmission line inspection based on unmanned aerial vehicle,” Artif Intell Rev, vol. 56, no. 1, pp. 173-201, April 2023.

[2] Gauld J G, Silva J P, Atkinson P W, Record P, Acácio M, Arkumarev V, “Hotspots in the grid: Avian sensitivity and vulnerability to collision risk from energy infrastructure interactions in Europe and North Africa,” J Appl Ecol, vol. 59, no. 6, pp. 1496-1512, Apr 2022.

[3] Qiu Z, Zhu X, Liao C, Shi D, Kuang Y, Li Y, “Detection of bird species related to transmission line faults based on lightweight convolutional neural network,” IET Gener Transm Dis, vol. 16, no. 5, pp. 869-881, Oct 2022.

[4] Zhang J, Qi Q Y, Zhang H L, Du Q, Guo Z, Tian Y Y, “Detection of bird’s nest on transmission lines from aerial images based on deep learning model,” Int J Innov Comput I, vol. 18, no. 6, pp. 1755-1768, Dec 2020.

[5] Shakiba F M, Azizi S M, Zhou M, Abusorrah, A, “Application of machine learning methods in fault detection and classification of power transmission lines: a survey,” Artif Intell Rev, vol. 56, no. 7, pp. 5799-5836, Nov 2023.

[6] Bhosle K, Musande V, “Evaluation of deep learning CNN model for recognition of devanagari digit,” Artificial Intelligence and Applications. Vol. 1, no. 2, pp. 114-118, Feb 2023.

[7] Ding Y, Zhang Z, Zhao X, Hong D, Cai W, Yu C, “Multi-feature fusion: Graph neural network and CNN combining for hyperspectral image classification,” Neurocomputing, vol. 501, no. 3, pp. 246-257, Aug 2022.

[8] Das S, Mishra M, Majumder S, “Identification of Glaucoma from Retinal Fundus Images using Deep Learning Model, MobileNet,” ECTI Transactions on Computer and Information Technology (ECTI-CIT), vol. 18, no. 3, pp. 371-380, Jul 2024.

[9] Rebolo-Ifrán N, Plaza P, Pérez-García J M, Gamarra-Toledo V, Santander F, Lambertucci S A, “Power lines and birds: An overlooked threat in South America,” Perspect Ecol Conser, vol. 21, no. 1, pp. 71-84, Jan 2023.

[10] Biasotto L D, Moreira F, Bencke G A, D’Amico M, Kindel A, Ascensão F, “Risk of bird electrocution in power lines: a framework for prioritizing species and areas for conservation and impact mitigation,” Anim Conserv, vol. 25, no. 2, pp. 285-296, Apr 2022.

[11] Yuan J, Zheng X, Peng L, Qu K, Luo H, Wei L, “Identification method of typical defects in transmission lines based on YOLOv5 object detection algorithm,” Energy Rep, vol. 9, no. 6, pp. 323-332, Sep 2023.

[12] Qiu Z, Zhou Z, Wan Z. “Automatic classification of bird species related to power line faults using deep convolution features and ECOC-SVM model”. IET GTD, vol. 18, no. 19, pp. 3138-3149, Sep 2024.

[13] Zhang Y, Sun H, Li H, Qi, D, Yan, Y, Chen, Z. “Bird pecking damage risk assessment of UHV transmission line composite insulators based on deep learning.” IET GTD, vol. 17, no. 12, pp. 2788-2798, May 2023.

[14] Ye M, Tanaka K, “Improved Visual Robot Place Recognition of Scan-Context Descriptors by Combining with CNN and SVM,” J Robot Mechatron, vol. 35, no. 6, pp. 1622-1628, Dec 2023.

[15] Khairandish M O, Sharma M, Jain V, Chatterjee, J. M., Jhanjhi, N. Z. “A hybrid CNN-SVM threshold segmentation approach for tumor detection and classification of MRI brain images.” IRBM, vol. 43, no. 4, pp. 2788-2798, Aug 2022.

[16] Ozaltin O, Yeniay O. “A novel proposed CNN–SVM architecture for ECG scalograms classification.” SOFT COMPUT, vol. 27, no. 8, pp. 4639-4658, Dec 2023.

[17] Yasar A. “Analysis of selected deep features with CNN-SVM-based for bread wheat seed classification”. EUR FOOD RES TECHNOL, vol. 250, no. 6, pp. 1551-1561, Mar 2024.

[18] Anggriandi D, Utami E, Ariatmanto D. “Comparative analysis of CNN and CNN-SVM methods for classification types of human skin disease.” Sinkron, vol. 7, no. 4, pp. 2168-2178, Oct 2023.

[19] Zhang Y, Sun H, Li H, Qi D, Yan Y, Chen Z, “Bird pecking damage risk assessment of UHV transmission line composite insulators based on deep learning,” IET Gener Transm Dis, vol. 17, no. 12, pp. 2788-2798, May 2023.

[20] Das S K, Roy P, Mishra A K, “DFU_SPNet: A stacked parallel convolution layers based CNN to improve Diabetic Foot Ulcer classification,” ICT Express, vol. 8, no. 2, pp. 271-275, Nov 2022.

[21] Dong Z, Zhao D, Cui L, “An intelligent bearing fault diagnosis framework: one-dimensional improved self-attention-enhanced CNN and empirical wavelet transform,” Nonlinear Dynamics, vol. 112, no. 8, pp. 6439-6459, Mar 2024.

[22] Waheed S R, Rahim M S M, Suaib N M, Salim A A, “CNN deep learning-based image to vector depiction,” Multimed Tools Appl, vol. 82, no. 13, pp. 20283-20302, Jan 2023.

[23] Elangovan P, Vijayalakshmi D, Nath M K, “Covid-19net: An effective and robust approach for covid-19 detection using ensemble of convnet-24 and customized pre-trained models,” Circ Syst Signal Pr, vol. 43, no. 4, pp. 2385-2408, Dec 2024.

[24] Chaudhari D J, Malathi K, “Detection and prediction of rice leaf disease using a hybrid CNN-SVM model.” Opt Memory Neural, vol. 32, no. 7, pp. 39-57, Apr 2023.

[25] Okomba N S, Adedayo S A, Aviara C V, Esan A O, Omodunbi B, “Development of Glaucoma Detection System using CNN and SVM,” Arid Zone Journal of Engineering, Technology and Environment, vol. 20, no. 1, pp. 193-214, Mar 2024.

[26] Gao Q, Yang Y, Kang Q, Tian Z, Song Y. “EEG-based emotion recognition with feature fusion networks.” Int J Mach Learn Cyb, vol. 13, no. 2, pp. 421-429, Feb 2022.