International Journal of Innovative Computing, Information and Control
Volume **18**, Number **6**, December **2022**

ICIC International ©2022 ISSN 1349-4198
pp. **1755–1768**

DETECTION OF BIRD'S NEST ON TRANSMISSION LINES FROM AERIAL IMAGES BASED ON DEEP LEARNING MODEL

Jie Zhang[1], Qiye Qi[1], Huanlong Zhang[1,*], Qifan Du[1], Zhimin Guo[2] and Yangyang Tian[2]

[1]College of Electrical and Information Engineering
Zhengzhou University of Light Industry
No. 5, Dongfeng Road, Zhengzhou 450002, P. R. China
{ zhangjie1234; 2019031; 2019028 }@zzuli.edu.cn
*Corresponding author: 2016007@zzuli.edu.cn

[2]State Grid Henan Electric Power Research Institute
No. 85, Songshan Road, Zhengzhou 450052, P. R. China
{ guozhimin; tianyangyang }@ha.sgcc.com.cn

Received April 2022; revised August 2022

Abstract. *The bird's nest on transmission lines poses a threat to the safe operation of transmission equipment and even affects the stability of the entire power system. Recently, with the rapid development of 5G technology, unmanned aerial vehicle (UAV) technology, and artificial intelligence technology, intelligent patrol transmission lines based on UAVs have become an inevitable trend in the development of power inspection. To address the problems of low recognition accuracy and low Recall of bird's nest detection in complex backgrounds by traditional methods, an improved YOLOv3 automatic detection model of bird's nest based on attentional feature fusion (AFF-YOLOv3) is proposed in this paper. The model first adds an attentional feature fusion network to the YOLOv3 top-down sampling process, calculates semantic weights based on the deep-level feature map, and then uses the semantic weights as a guide for selecting low-level features to obtain more valuable low-level features. Finally, the selected low-level feature maps and the high-level feature maps are concatenated to obtain robust features with both location information and semantic information. The experimental results show that AFF-YOLOv3 achieves 87.58% average precision (AP) on the transmission line bird's nest dataset, and the model has stronger generalization ability and applicability compared with other detectors.*

Keywords: Deep learning, Bird's nest detection, AFF-YOLOv3, Attentional feature fusion, Intelligent inspection

1. **Introduction.** In recent years, with the continuous expansion of industrial production, the number of transmission lines has been increasing [1]. However, the activities of birds can pose a great threat to the safe operation of transmission lines [2]. The bird's nest (hereinafter referred to as nests for brevity) on transmission lines may lead to flashover of insulators and even cause short circuits of the lines [3].

With the growth of power demand, especially in the context of smart power grids, the safe operation of high-voltage transmission lines becomes more and more important. Since the nests on the transmission line can cause a series of harm, accurate and efficient detection of these nests becomes important for reducing the work intensity of patrol staff, improving efficiency, and ensuring the stable operation of the power grids.

DOI: 10.24507/ijicic.18.06.1755

Traditional image processing methods generally segment the nests by specific features (e.g., color, texture, gradient, and shape) and then implement object detection by matching algorithms. In the study of detecting the nests from images, Xu et al. [4] proposed to utilize the color, shape, and special texture of nests to detect the nests in UAV images, but the complex environment around the nests may have an impact on the detection effect. Wu et al. [5] proposed the use of a histogram of the orientation of streaks and histogram of the length of streaks to characterize nests and then used them for the detection of the nests in the overhead catenary system. However, these methods rely on artificially designed features, which are not suitable for bird's nest detection in complex backgrounds. The uncertainty of the correlation features of bird's nest in aerial images leads to low accuracy of traditional image processing methods, which are difficult to be applied in practice.

In recent years, with the rapid development of deep learning technology, convolutional neural networks (CNNs) have been widely used in the fields of object detection [6], semantic segmentation [7], automatic driving, and so on. Many excellent object detection models based on CNNs have been proposed, including two-stage network: R-CNN [8], Fast R-CNN [9], Faster R-CNN [10], Mask R-CNN [11] and one-stage network: SSD [12], YOLO series [13-17], CenterNet [18]. Naturally, the existing CNNs-based object detection models can be employed for bird's nest detection. In [19], a convolutional neural network-based transmission tower nests detection system was designed. Compared with the traditional image processing-based nests detection method, its detection accuracy has been significantly improved. In [2], the Faster R-CNN network was used to achieve automatic recognition of nests in the inspection images, and various data enhancement methods were used to solve the problem of insufficient samples. In [20], an improved Faster R-CNN network was proposed to improve the detection accuracy of nests on transmission lines by using the k-means clustering method and introducing the focal loss function. In [21], the accuracy and efficiency of the SSD algorithm are verified with VGG16 and EfficientNet-B7 for real-time detection of bird's nest on transmission lines. Due to the limitations of the original network, the bird's nest detection performance of the above methods needs further improvement. In [22], the RetinaNet model suitable for bird's nest detection was established by adjusting the appropriate network structure and parameters, and its validity was verified by comparison with Faster R-CNN, YOLO, and other methods. However, the applicability of the method is limited by the fact that transmission line inspection images often have complex backgrounds.

Nests in transmission line inspection images are of great diversity and randomness. Due to the fact that the above-mentioned methods fail to process the characteristic information about nests in the network, they are incapable of distinguishing nests from the complex background. As a result, these methods can hardly complete transmission line nests detection quickly and accurately.

The one-stage network YOLOv3 employs the idea of feature pyramid network (FPN) [23] to detect objects at different scales and combines high-level semantic features and low-level location features through a top-down sampling process, which improves the localization and classification abilities of objects and achieves excellent detection performance. However, YOLOv3 directly fuses unprocessed low-level features with high-level features, introducing irrelevant feature information.

To overcome the above problems, an improved YOLOv3 object detection model based on attention feature fusion network is proposed. In summary, the main contributions of this paper are as follows.

1) Design the attentional feature fusion network, calculate the semantic weight based on the deep-level feature map, and use this semantic weight as a guide for selecting low-level features to obtain more valuable low-level features.

2) To improve the feature fusion efficiency of the YOLOv3 detection model, an attentional feature fusion network is introduced in the top-down sampling process of YOLOv3 to help the model obtain positional and semantic information about the target and improve the detection accuracy.

3) The experimental results show that the AFF-YOLOv3 model can effectively detect nests in various complex environments, and has strong applicability and generalization capability.

2. Our Approach.

The details of our work are proposed in this section. In Section 2.1, the basic theoretical knowledge of YOLOv3 is introduced in detail. In Section 2.2, the detailed design process of the attention feature fusion network is presented. In Section 2.3, the overall framework of AFF-YOLOv3 is described.

2.1. YOLOv3 network.

The network structure of YOLOv3 can be divided into a feature extraction network and a multi-scale prediction network (as shown in Figure 1). It draws on the idea of FPN to extract semantic features layer by layer on the input image, extract the features of the previous layer through a top-down upsampling operation, and fuse them with the features of the current layer to enhance the semantic information in the features, and then predicts from three scales to achieve the detection of objects of different sizes.

In Figure 1, the feature extraction network of YOLOv3 is Darknet-53 similar to ResNet [24], and the residual blocks added to the network effectively alleviate the network degradation problem of deep convolutional neural networks, which enables the network to be built deeper. The Darknet-53 network mainly consists of the DBL module and several RES_n modules. DBL module consists of Conv layer, Batch Normalization (BN) operation, and Leaky ReLU. RES_n indicates that each residual block contains n RES Units. In YOLOv3, the multi-scale prediction network is used to obtain the location and category of objects. Given the input image size of 416×416 for the YOLOv3 network, the image is divided into $S \times S$ grids from three scales (13×13, 26×26, 52×52), and each grid is responsible for predicting the objects within it.

Compared with the previous YOLO series detectors, YOLOv3 keeps a better balance between detection accuracy and detection speed, but it still has several challenges when directly applied to the detection of nests on transmission lines (e.g., insufficient extraction of feature, recall of the model is low and high missed detection rate). To solve the above issues, an improved YOLOv3 automatic bird's nest detection model based on attentional feature fusion network is proposed to obtain robust features of nest objects through attention feature fusion networks, reduce the rate of missed detection and improve the detection accuracy.

2.2. Attention feature fusion network.

The attention feature fusion network can select more effective feature information, which not only more accurately computes channel attention, but also provides guidance information for the low-level feature in a simple way. To efficiently compute the channel attention, the spatial dimension of the input feature map needs to be squeezed, many methods commonly adopt averaging pooling to extract feature information [25,26]. In [27], max-pooling is proposed to extract distinctive object features to infer finer channel-wise attention, and experimental results confirm that using max-pooling and average-pooling greatly improves the representation capability of

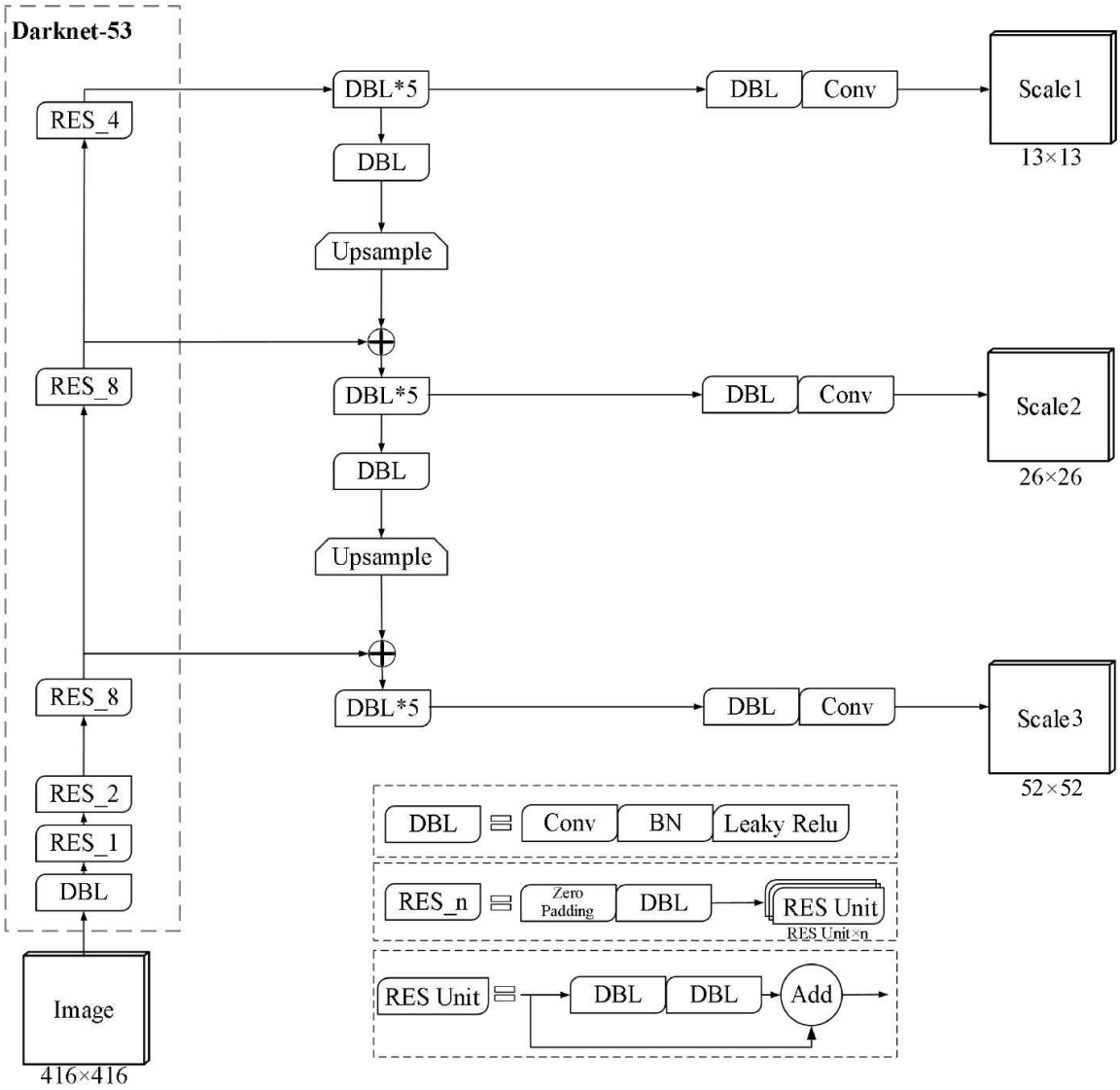


FIGURE 1. YOLOv3 network structure

the network, rather than using them alone. Inspired by this, max-pooling and average-pooling are adopted to extract important information about high-level features, and this process does not add too much to the computational parameters.

As shown in Figure 2, the attention feature fusion network is divided into two branches: upper and lower. The upper branch computes semantic weights using high-level semantic input feature H. The semantic weight matrix $M(H)$ is calculated as follows. First, average-pooling and max-pooling operations are performed on the input feature map H to aggregate feature information and generate average-pooled features and max-pooled features, respectively. Then, these two pooled features are forwarded to a shared multi-layer perceptron (MLP) to produce a channel attention map. The semantic weight matrix $M(H)$ is obtained by element-wise summation and the Sigmoid function. The semantic weight matrix $M(H)$ is calculated as follows:

$$M(H) = \sigma(MLP(Avgpool(H)) + MLP(Maxpool(H))) \tag{1}$$

where σ denotes the sigmoid function. To make full use of the information aggregated in the pooling operation, the sigmoid function σ is needed to capture channel-wise dependencies. The sigmoid function is capable of learning a nonlinear interaction between

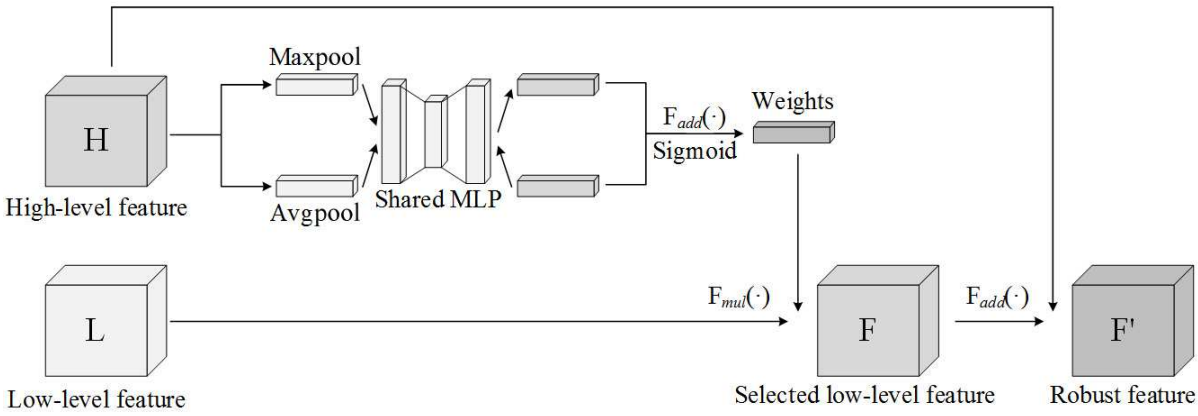


FIGURE 2. The main architecture of attention feature fusion network. The network takes as input high-level semantic features and low-level detail features, respectively.

channels and ensuring multiple channels instead of one thermal activation. It is capable of adaptively learning the weights of each channel and fixing these weights to between 0 and 1. *Avgpool* and *Maxpool* represent average-pooling and max-pooling operations, respectively.

The lower branch takes the low-level detail feature L as input and multiplies them with the semantic weight matrix $M(H)$ extracted from the upper branch to obtain the filtered low-level feature map F. Finally, the high-level semantic feature is fused with the filtered low-level detail feature. The intent is to combine high-level semantic features with more effective detail features. In short, the output F' of the two-branch attention network can be summarized as follows:

$$F = M(H) \times L \tag{2}$$

$$F' = F + H \tag{3}$$

To effectively improve the feature extraction capability of the detection network, the attention feature fusion network enhances the saliency of the object in the low-level features by guiding the low-level detail features through the semantic weight matrix generated by the upper branch. Therefore, the fusion of information extracted from high and low-level features by an attention feature fusion network using different levels of features as input helps to enhance the expression of features and improve the detection accuracy of the model.

2.3. **The location of attention feature fusion network in AFF-YOLOv3.** High-level features contain richer semantic information that is beneficial for object classification, while low-level features contain more detailed information that is essential for the location of object bounding boxes. However, the low-level features may be adulterated with information that is irrelevant to the object, and not all low-level features are beneficial for detection. To solve this issue, the attention feature fusion network is introduced in YOLOv3.

Figure 3 shows the location of the attention feature fusion network in AFF-YOLOv3 (red dashed box). The left half of Figure 3 is the bottom-up feature extraction process, and C1, C2, and C3 represent the output feature maps of each stage of Darknet-53, respectively.

The right half of Figure 3 shows the top-down upsampling process. The high-level semantic feature map C1 from Darknet-53 is processed by the DBL*5 module to obtain two outputs, one of which is fed to the multiscale prediction network for prediction ($13 \times$

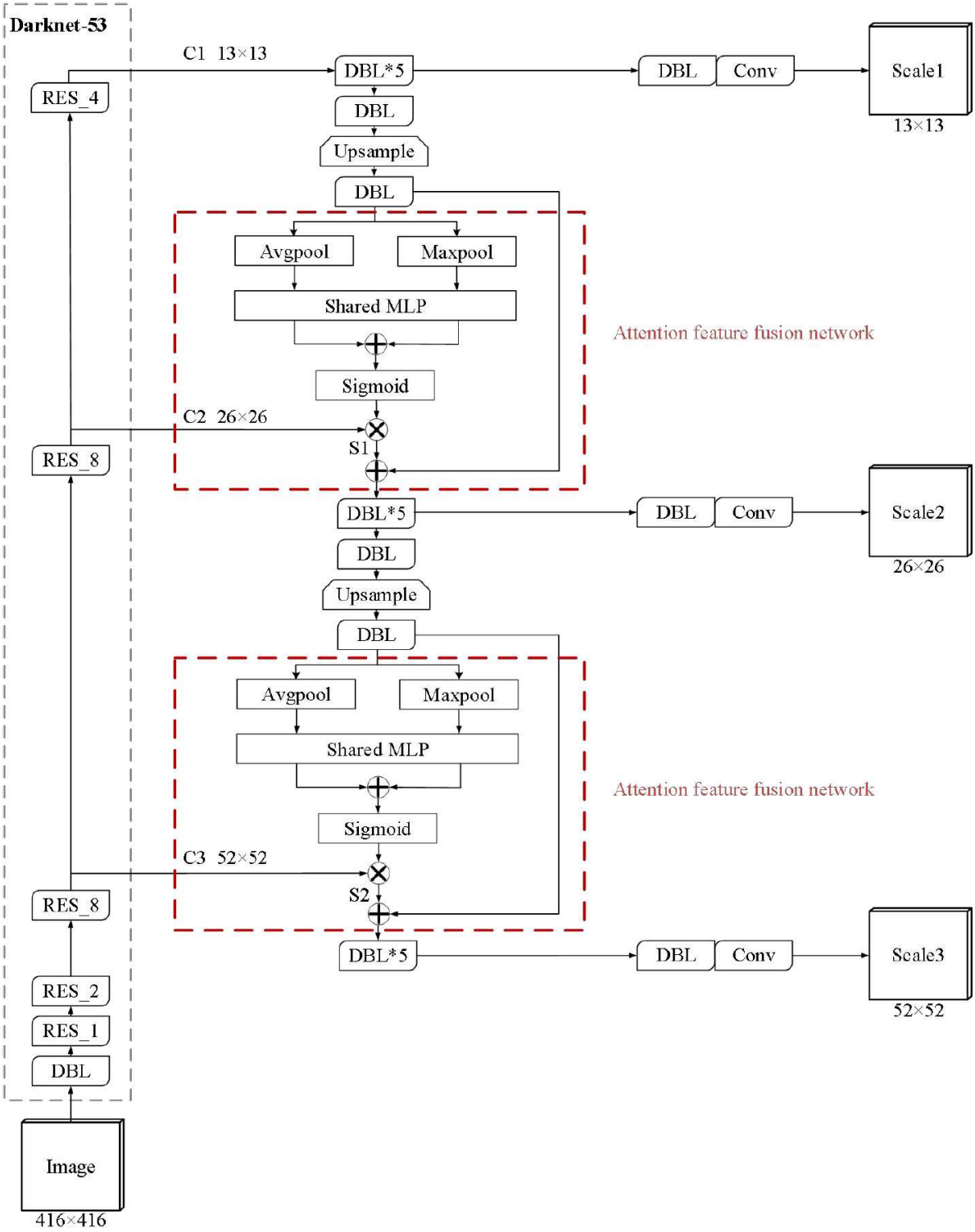


FIGURE 3. AFF-YOLOv3 network structure

13), and the other output is processed by two DBL modules and upsampling as the input to the attention feature fusion network (where the second DBL module is used for increasing channel dimensions to contain richer feature information). The attention feature fusion network calculates the semantic weight matrix of the input feature map and multiplies it with the low-level detail feature map C2 to filter out the redundant information in the low-level features. Finally, the high-level semantic feature map C1

is fused with the selected low-level detail feature map S1 to obtain a robust feature map that contains both semantic and location information. Another attentional feature fusion network is introduced between the Darknet-53 output feature maps C2 and C3. Similarly, the DBL*5 module processes the robust feature map of the output from the previous step to obtain two outputs, one of which is fed to the multiscale prediction network for prediction (26×26), and the other output is processed by two DBL modules and upsampling as the input to the attention feature fusion network. Finally, the output of the attentional feature fusion network is processed by the DBL*5 module and fed to the multiscale prediction network for prediction (52×52) to complete the whole target detection process.

The fusion of information extracted from high and low-level features by an attention feature fusion network using different levels of features as input helps to enhance the expression of features and improve the detection accuracy of the YOLOv3 network.

3. **Experiment.** The hardware platforms used for all experiments in this paper are as follows: Intel platinum 8163 processor, NVIDIA TITAN RTX GPU, and 256 GB of memory. During the training of AFF-YOLOv3, Adam was used to perform gradient updates on the network. In addition, the Precision, Recall, Average Precision (AP), F1 score, and running time were used as evaluation metrics for detection models.

3.1. **Establishing the bird's nest dataset.** Since there is no publicly available transmission line nests dataset, this paper uses images of transmission line nests in Henan Province taken by UAVs. However, the detection effect obtained by training the model using only the collected nest images is far from the expected and cannot meet the demand of actual power inspection, so we use random brightness, adding Gaussian noise, and other data enhancement methods to expand the initial image to obtain 3328 augmented images. The purpose of data enhancement is to compensate for the lack of nest images and reduce the occurrence of over-fitting of the network so that the trained model can get better detection results. Finally, a transmission line faults dataset named 'Bird's Nests' is established, which includes 3328 aerial images, and the details of the "Bird's Nest" dataset are shown in Table 1.

Table 1. The details of the "Bird's Nest" dataset

Image number	3328
Number of single nest pictures	351
Number of multi nest pictures	2977
Number of small object nest pictures	510
Number of nest distortion pictures	998
Number of background clutter pictures	2995

As shown in Figure 4, we acquired images containing various challenging nests objects such as distortion, small objects, and background clutter to better evaluate the robustness of different detection models.

3.2. **Results and analyses.** To evaluate the effectiveness of the proposed AFF-YOLOv3, experiments are conducted on the bird's nest dataset. Both quantitative and qualitative comparisons are used to evaluate the performances of the improved YOLOv3. The experimental results of the original YOLOv3 and AFF-YOLOv3 are shown in Table 2.

As shown in Table 2, the Precision values of the two network models are YOLOv3 (92.01%) and AFF-YOLOv3 (93.16%), and the Precision of the AFF-YOLOv3 model is 1.15% higher than that of YOLOv3. The Recall values of the two network models are

FIGURE 4. The samples of bird's nest with diverse scenes

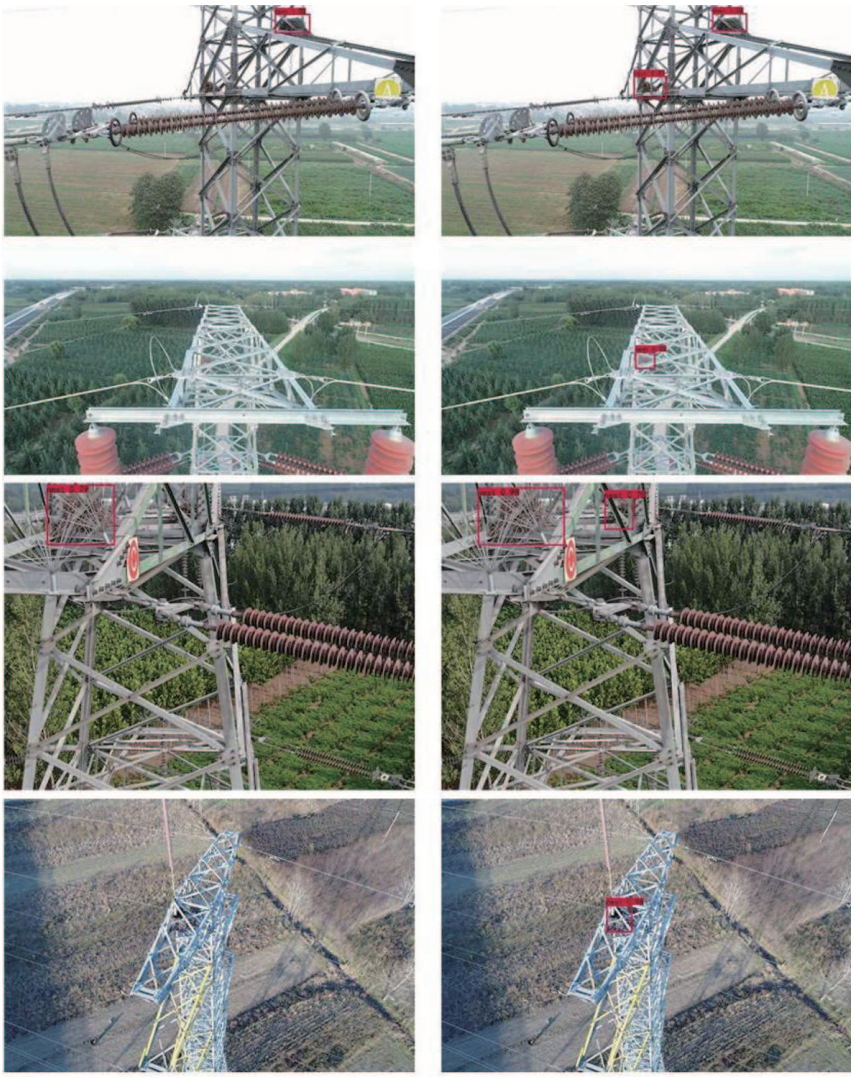
TABLE 2. The experimental effects of original YOLOv3 and AFF-YOLOv3

Detector	Precision	Recall	AP	F1 score	Running time (ms/piece)
YOLOv3	92.01%	78.12%	82.93%	85%	40.2
AFF-YOLOv3	93.16%	82.68%	87.58%	88%	41.9

YOLOv3 (78.12%) and AFF-YOLOv3 (82.68%), and the Recall of the AFF-YOLOv3 model is 4.56% higher than that of YOLOv3. The F1 score values of the two network models are YOLOv3 (85%) and AFF-YOLOv3 (88%), and the F1 score of the AFF-YOLOv3 model is 3% higher than that of YOLOv3. The average precision (AP) values of the two network models are YOLOv3 (82.93%) and AFF-YOLOv3 (87.58%), and the AP of the AFF-YOLOv3 model is 4.65% higher than that of YOLOv3. Based on the values of Precision, Recall, AP, and F1 score, it is proved that the proposed AFF-YOLOv3 is superior to the YOLOv3. The running times of the networks were YOLOv3 (40.2 ms) and AFF-YOLOv3 model (41.9 ms), and it can be concluded that the proposed AFF-YOLO model can still maintain a fast detection speed. Consequently, as it achieves a good trade-off between Precision, Recall, AP, F1 score, and running time, the proposed AFF-YOLOv3 model may be more advantageous than YOLOv3. AFF-YOLOv3 considers the difference in semantic information contained between high-level features and low-level features, and the robust feature map containing both semantic and location information can be obtained by an attention feature fusion network, which achieves a better feature representation without adding too many computational parameters while obtaining improved detection accuracy.

The detection performance of AFF-YOLOv3 on the bird's nest dataset is qualitatively evaluated in Figure 5(a). It can be concluded that the original YOLOv3 is prone to missed detections, which is due to the fact that feature information about the nest is difficult to be mined by the detection network in complex backgrounds, which in turn reduces the generalization ability of the network. Compared with YOLOv3, Figure 5(b) shows that AFF-YOLOv3 performs well. It is because the attention feature fusion network combines filtered low-level detail features with high-level semantic features to better distinguish between the nest and the background.

The experimental effects of different models for nests detection are shown in Table 3. For a fair comparison, experiments on the Faster RCNN, SSD, Retinanet [28], YOLOv4-tiny and AFF-YOLOv3 are conducted by using the same evaluation method on the Bird's Nest dataset. Compared to other detection models, AFF-YOLOv3 achieves a better balance between detection accuracy and detection speed. It is effectively proved that integrating



(a) (b)

FIGURE 5. Qualitative comparison between YOLOv3 and AFF-YOLOv3 on bird's nest dataset: (a) YOLOv3; (b) AFF-YOLOv3

TABLE 3. The experimental effects of different models

Detector	Backbone	Precision (%)	Recall (%)	AP (%)
Faster-RCNN	Resnet-50	80.88	83.48	80.68
SSD	VGG	90.59	73.04	78.92
Retinanet	Resnet-50	90.85	83.30	85.75
YOLOv4-tiny	CSPDarknet-53-tiny	83.53	62.50	73.39
YOLOv3	Darknet-53	92.01	78.12	82.93
AFF-YOLOv3	Darknet-53	93.16	82.68	87.58

the attention feature network into YOLOv3 is very effective for the detection of bird's nest.

Similarly, the comparison of the detection effect of AFF-YOLOv3 proposed in this paper with the existing work is shown in Table 4. It is found that the Precision of the AFF-YOLOv3 is 16.96% and 5.95% higher than that of [29] and [30], 3.16% higher than

TABLE 4. Comparison of AFF-YOLOv3 with existing work

Methods	**Precision (%)**
Dual-scale YOLOv3 by [29]	76.20
AFCNN by [30]	87.21
RFCN by [31]	90.00
AFF-YOLOv3	93.16

(a) The experimental effects with the background of trees

(b) The experimental effects with the power tower as the background

(c) The experimental effects of the small object

FIGURE 6. Experimental results with different scenes conducted by the AFF-YOLOv3 model

that of [31]. Therefore, the AFF-YOLOv3 proposed in this paper is more suitable for the detection of bird's nest in practical applications.

To further verify the effectiveness and robustness of the proposed AFF-YOLOv3 model, several typical aerial images are selected to exhibit the visualization performance, and the positions of the nests are located by red rectangular boxes.

Specifically, the experimental effects with the background of trees are shown in Figure 6(a). Although the color of the background is similar to that of the nests and the tree background is complex, the AFF-YOLOv3 model can still detect all the nests correctly. Figure 6(b) shows the experimental effects with the power tower as the background. Although the nests are obscured or deformed, all nests are detected by the AFF-YOLOv3 model. The experimental effects of the small object are shown in Figure 6(c). Due to the

influence of image acquisition equipment, shooting distance, and other factors, the proportion of nests to the whole image may be very small, but AFF-YOLOv3 still demonstrates excellent detection results.

The above analysis shows that the proposed AFF-YOLOv3 model is advantageous for nest detection in aerial images with a complex background, distortion, and small objects. In future research, the AFF-YOLOv3 model is used for automatic inspection of nests in UAV-based high-voltage transmission lines, and the AFF-YOLOv3 model is extended to detect faults in other high-voltage transmission line components (e.g., insulator faults, missing anti-vibration hammers, and missing metal fittings).

4. **Conclusions.** This study proposes an improved YOLOv3 detection model for bird's nest detection in aerial images. Firstly, a novel dataset named 'Bird's Nest' is established, consisting of 3328 aerial images. Secondly, an improved YOLOv3 model based on the attention network model is proposed to achieve bird's nest detection in aerial images with various environments in this paper. The attentional feature fusion network calculates the attention weight based on the deep-level feature map and uses this attention weight as a guide for selecting low-level features to obtain more valuable low-level features. Then the selected low-level feature maps and the high-level feature maps are concatenated to obtain robust features with both location information and semantic information. Finally, the proposed AFF-YOLOv3 model, YOLOv3 model, and other object detection models are carefully trained and tested on the established dataset. The experimental results fully validate that the proposed AFF-YOLOv3 model is effective in detecting bird's nest targets in various complex environments and has extremely important application value in the power industry.

In this work, AFF-YOLOv3 considers the difference in semantic information contained between high-level features and low-level features. However, the attention feature fusion network ignores the differences in spatial location relationships between features at different levels, which limits the further improvement of detection accuracy. In the next work, we will further explore the information differences between different levels of features to obtain a better feature representation and enhance the detection performance.

Acknowledgment. This work is supported by the grants from the National Natural Science Foundation of China (Nos. 62102373, 61873246, 62006213), the Science and Technology Research Project of Henan Province (No. 212102310053) and Henan University Science and Technology Innovation Talents Program (No. 21HASTIT028).

REFERENCES

[1] Y. Shu and W. Chen, Research and application of UHV power transmission in China, *High Voltage*, vol.3, no.1, pp.1-13, 2018.

[2] F. Li, J. Xin, T. Chen, L. Xin, Z. Wei, Y. Li, Y. Zhang, H. Jin, Y. Tu, X. Zhou et al., An automatic detection method of bird's nest on transmission line tower based on Faster_RCNN, *IEEE Access*, vol.8, pp.164214-164221, 2020.

[3] H. Yi, Y. Xiong, G. Zhou and H.-W. He, Analysis on bird-caused damages of overhead transmission lines and countermeasures, *Power System Technology*, vol.32, no.20, pp.95-100, 2008.

[4] J. Xu, J. Han, Z. Tong and Y. Wang, Method for detecting bird's nest on tower based on UAV image, *Computer Engineering and Applications*, vol.53, no.6, pp.231-235, 2017.

[5] X. Wu, P. Yuan, Q. Peng, C.-W. Ngo and J.-Y. He, Detection of bird nests in overhead catenary system images for high-speed rail, *Pattern Recognition*, vol.51, pp.242-254, 2016.

[6] K. L. Masita, A. N. Hasan and T. Shongwe, Deep learning in object detection: A review, *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp.1-11, 2020.

[7] K. Kawagoe, S. Murakami, T. Kamiya and T. Aoki, Automatic segmentation of finger bone regions from CR images using improved U-Net and MSGVF Snakes, *ICIC Express Letters, Part B: Applications*, vol.13, no.2, pp.155-160, 2022.
[8] R. Girshick, J. Donahue, T. Darrell and J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.580-587, 2014.
[9] R. Girshick, Fast R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.1440-1448, 2015.
[10] S. Ren, K. He, R. Girshick and J. Sun, Faster R-CNN: Towards real-time object detection with region proposal networks, *Advances in Neural Information Processing Systems*, vol.28, pp.91-99, 2015.
[11] K. He, G. Gkioxari, P. Dollár and R. Girshick, Mask R-CNN, *Proc. of the IEEE International Conference on Computer Vision*, pp.2961-2969, 2017.
[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu and A. C. Berg, SSD: Single shot multibox detector, *European Conference on Computer Vision*, pp.21-37, 2016.
[13] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You Only Look Once: Unified, real-time object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.779-788, 2016.
[14] J. Redmon and A. Farhadi, YOLO9000: Better, faster, stronger, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7263-7271, 2017.
[15] J. Redmon and A. Farhadi, YOLOv3: An incremental improvement, *arXiv Preprint*, arXiv: 1804.027 67, 2018.
[16] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, YOLOv4: Optimal speed and accuracy of object detection, *arXiv Preprint*, arXiv: 2004.10934, 2020.
[17] *YOLOv5 Code Repository*, https://gitee.com/atari/ultralytics_yolov5, Accessed on 20 January 2021.
[18] X. Zhou, D. Wang and P. Kräahenbühl, Objects as points, *arXiv Preprint*, arXiv: 1904.07850, 2019.
[19] M. Chen and C. Xu, Bird's nest detection method on electricity transmission line tower based on deeply convolutional neural networks, *2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, vol.1, pp.2309-2312, 2020.
[20] J. Li, D. Yan, K. Luan et al., Deep learning-based bird's nest detection on transmission lines using UAV imagery, *Applied Sciences*, vol.10, no.18, 2020.
[21] D. Satheeswari, L. Shanmugam and N. M. J. Swaroopan, Recognition of bird's nest in high voltage power line using SSD, *2022 1st International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*, 2022.
[22] Z. Hui, Z. Jian, C. Yuran et al., Intelligent bird's nest hazard detection of transmission line based on retinanet model, *Journal of Physics: Conference Series*, DOI: 10.1088/1742-6596/2005/1/012235, 2021.
[23] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, Feature pyramid networks for object detection, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2117-2125, 2017.
[24] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.770-778, 2016.
[25] J. Hu, L. Shen and G. Sun, Squeeze-and-excitation networks, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.7132-7141, 2018.
[26] B. Zhou, A. Khosla, A. Lapedriza et al., Learning deep features for discriminative localization, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp.2921-2929, 2016.
[27] S. Woo, J. Park, J.-Y. Lee and I. S. Kweon, CBAM: Convolutional block attention module, *Proc. of the European Conference on Computer Vision (ECCV)*, pp.3-19, 2018.
[28] T. Y. Lin, P. Goyal, R. Girshick, K. He and P. Dollár, Focal loss for dense object detection, *Proc. of the IEEE International Conference on Computer Vision*, pp.2980-2988, 2017.
[29] J. Ding, L. Huang, D. Zhu et al., Learning to detect high altitude tower nest using dual-scale YOLOv3 network, *Journal of Xi'an University of Technology*, vol.37, no.2, pp.253-260, 2021.
[30] W. Dong, L. Wu, Q. Wang et al., An automatic detection method of bird's nest on electric tower based on attention full convolutional neural networks, *2021 4th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp.304-308, 2021.
[31] J. Zou, R. Xu and Y. Diao, Application of deep learning in intelligent defect identification of transmission line project acceptance, *Jiangxi Electric Power*, vol.44, no.2, pp.5-9, 2020.

Author Biography

Jie Zhang received the bachelor's degree in Automation from Henan University of Science and Technology, China, 2010; the master of engineering in Control Theory and Control Engineering from Harbin Engineering University; the Ph.D. degree in Control Science and Engineering from Harbin Institute of Technology, China, 2018.

Dr. Zhang is currently a full-time teacher at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, China. His research interests include object detection and recognition, image processing, and pattern recognition. He has published over 40 papers in journals and conferences.

Qiye Qi obtained the bachelor's degree in Electrical Engineering and Its Automation from Henan University of Engineering, China, in 2020. He is currently working towards the master's degree in Zhengzhou University of Light Industry, China. His main research interests include object detection, artificial intelligence and computer vision.

Huanlong Zhang received the Ph.D. degree from the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China, in 2015. He is currently an Associate Professor with the College of Electrical and Information Engineering, Zhengzhou University of Light Industry, Henan, Zhengzhou, China. His research has been funded by the National Natural Science Foundation of China (NSFC), the Key Science and Technology Henan Province, etc. He has published more than 40 technical articles in refereed journals and conference proceedings. His research interests include pattern recognition, machine learning, image processing, computer vision, and intelligent human-machine systems.

Qifan Du obtained the bachelor's degree in Building Electrical and Intelligent from Zhengzhou University of Light Industry, China, in 2019. He is currently working towards the master's degree in Zhengzhou University of Light Industry. His main research interests include artificial intelligence, deep learning, and object detection.

Zhimin Guo received a B.E. degree from Shanghai University of Electric Power in 2000. He is currently serving as a Full Professor Senior Engineer in the State Grid Henan Electric Power Research Institute. His current research interests include power artificial intelligence and industrial control information security. Now he has conducted and is conducting a number of standard-setting and state-grid science and technology project research as project principal or main researcher.

Yangyang Tian received a B.E. degree from Henan Normal University in 2015 and a M.Sc. degree from Donghua University in 2018. She is currently an engineer at State Grid Henan Electric Power Research Institute. She is mainly engaged in the research of machine learning.