Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Research on the application of unmanned aerial vehicle multi-view image processing based on computational complexity analysis in the recognition and acceptance of infrastructure line defects

Haijun Su[1,✉], Huajun Xu[1], Lei Ma[1]

[1] *State Grid Gansu Electric Power Company Linxia Power Supply Company, Linxia, Gansu, 731100, China*

ABSTRACT

For national grid power line infrastructure construction construction, quality management and control can ensure improved safety standards, long-term reliability and cost savings through avoiding rework. In this paper, a high-definition image of a transmission line is collected from multiple viewpoints by a UAV, and a model for recognizing surface defects on infrastructure lines is proposed to reduce the computational complexity to improve the YOLOv8 algorithm. The model uses ResNet50 as the feature extraction backbone network and fuses convolution and attention mechanisms to enhance global and local feature extraction. A multi-scale feature aggregation diffusion module is added to the neck network of the model to enhance the detection of small targets on infrastructure lines. Finally, the classification loss function combined with the PIOU bounding box loss function is introduced to further enhance the recognition accuracy of infrastructure line surface defects. The experimental results show that the mAP of the infrastructure line surface defect recognition model is up to 0.935, which is 2.41% higher than that of the baseline model, and the performance is significantly better than that of some of the current mainstream defect recognition models. Therefore, from the computational complexity, combined with the target detection YOLOv8 algorithm can realize the accurate recognition of surface defects on infrastructure lines, and provide reliable data support for improving the timely repair of grid infrastructure lines.

Keywords: Computational complexity, UAV multi-view, improved YOLOv8 algorithm, defect recognition

1. Introduction

In recent years, UAV aerial photography has become a hot topic in the field of photography and visualization. By carrying high-definition cameras or sensors, drone aerial photography can capture

✉ Corresponding author.
E-mail address: SHKJYX2022@163.com (H. Su).
Received 25 March 2024; Revised 05 June 2024; Accepted 15 December 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-270

angles and horizons that were previously difficult to realize, bringing more creative inspiration to photographers and visual artists [1-4]. However, after the drone has finished shooting, image processing is an indispensable part, which can enhance the quality of the image and better show the beautiful scenery captured by the drone [5-7]. UAV aerial image processing refers to the use of UAVs for aerial data acquisition and processing of the acquired images to extract the feature information in the images and match them [8-10]. This technology is widely used in urban planning, environmental protection and other fields, and its application in the recognition and acceptance of infrastructure line defects has also shown an important role, especially the UAV multi-view image processing based on computational complexity analysis [11-14].

With the acceleration of urbanization and rapid economic development, various infrastructure construction projects have emerged. In order to ensure the quality and sustainability of these projects, the acceptance of infrastructure construction projects becomes particularly important [15-18]. The acceptance of infrastructure construction projects is of great significance to ensure the quality and sustainable development of the projects. In practice, specific acceptance standards should be formulated according to specific industry requirements and relevant laws and regulations, strictly enforced and timely revised and updated to ensure the quality and sustainable operation of construction projects [19-22]. And through the drone aerial photography can show the whole picture of the construction project in front of the acceptance personnel, whether the construction project meets the acceptance or not is also clear at a glance [23-24].

The article uses UAV to obtain multi-view images of infrastructure lines, extracts the features of the original pictures of infrastructure lines using HOG feature extraction method, and labels the defective parts in the original images, so as to construct a sample library of infrastructure line defects. Based on the YOLOv8 algorithm, we introduce the convolution and attention fusion module to obtain the global and local features of infrastructure line defects. Then, the multi-scale feature aggregation and diffusion module is combined to mine the defective features of infrastructure lines at small scales, and the bounding box regression PIoU loss function is introduced to enhance the recognition accuracy of defects at different scales. Applying the improved YOLOv8 model to the defect recognition of grid infrastructure lines can effectively improve the fault detection rate of electric power patrols and other faults, and reduce the safety hazards of grid line operation.

2. Infrastructure line defect identification data

Grid infrastructure lines are a crucial component of the power system, and their safe operation is directly related to the stability and reliability of power supply. With the expansion of the scale of the power system and the complexity of the operating environment, grid infrastructure lines are often faced with the threat of natural disasters and man-made damage, which may lead to line defects and even serious accidents. Therefore, rapid and accurate identification of grid infrastructure line defects is of great significance to ensure the safe operation of the power grid. Early detection of defects in grid infrastructure lines mainly relies on manual inspection, which is not only inefficient, but also limited by the experience of personnel and physical conditions, making it difficult to meet the demand for high-precision and large-scale detection. With the development of computer vision and machine learning technology, automatic defect recognition technology for grid infrastructure lines based on image processing has gradually emerged, and this technology analyzes the image or video data of grid infrastructure lines to automatically identify the defect types and locations, which greatly improves the efficiency and accuracy of defect detection for infrastructure lines.

2.1. Multi-view data collection by drones

2.1.1. Principle of spatio-temporal sequence image acquisition.The image characteristics of transmission infrastructure lines are complex, but once the transmission infrastructure lines are put

into operation, the location of most of the transmission equipment, such as towers and insulators, is relatively fixed. According to the requirements of the relevant documents, the transmission lines need to carry out 3~6 inspections per year. If each inspection is carried out by a UAV in the same geospatial location of the transmission equipment to shoot, fixed shooting distance, angle and other imaging parameters, can be obtained with the timing and spatial location information of the image library. According to the visible light imaging principle, after the relative position relationship between the visible light camera and the target to be photographed is fixed, the geometric features of the same target equipment on the image have very high similarity. If the environmental conditions such as solar radiation intensity and irradiation angle are similar, the radiation characteristics of the same target device on the image are also closer. According to the parameters of the mission equipment, the autonomous inspection of the curing path is set up, so that the historical inspection image has the ability to backtrack and analyze, which can effectively reduce the difficulty of the development of defect recognition algorithms and improve the defect recognition effect.

To realize that the UAV shoots at a fixed angle in a fixed position for many times, it is necessary for the UAV system to have high navigation and positioning precision, hovering control precision and flight control precision, so as to ensure that the images taken at different moments are highly overlapping. While ordinary UAVs usually adopt single-point navigation and positioning, RTK-type UAVs mostly adopt real-time dynamic differential technology, which can effectively improve navigation and positioning accuracy. When the UAV repeatedly hovers at the same set point and photographs the target equipment, changes in the real position or angle of the UAV have less impact on the reproduction of the target equipment. The performance of existing UAVs can basically meet the needs of inspection tasks at fixed distances and angles.

2.1.2. Data Acquisition Process for UAVs.The drone can clearly determine whether the important parts in the infrastructure lines are damaged, ensure the safety of transmission lines, and protect the residents' electricity consumption. Drone inspection improves the safety and reliability of line operation and is dozens of times more efficient than manual inspection. The drone has the hovering function, which can be manually or automatically manipulated, and can carry out observation and detection of the ground conductor and the tower in a close distance, which has the advantages of small volume, light weight, less constraints, short time consumption, less maintenance, etc., and is suitable for the inspection work of small components [25].

Taking a transmission line as an example, the process of UAV multi-view acquisition of its line image is shown in Figure 1, which mainly contains personnel preparation, site investigation, pre-takeoff preparation, takeoff operation, inspection flight, return landing, equipment recovery and inspection data processing. After the inspection of the infrastructure line by the drone is completed, the acquired data should be organized in a timely manner, and the photographs, videos and records should be transmitted to the relevant units to determine whether to carry out follow-up work. According to the final inspection record operators initially screen out suspected defects and transmit them to the equipment operation and maintenance unit for confirmation.

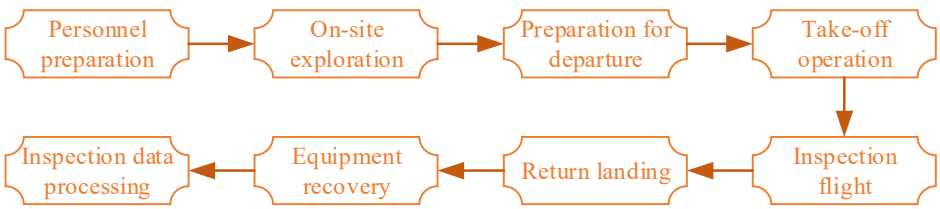


Fig. 1. UAV data acquisition process

2.2. *Infrastructure Line Defect Data Processing*

2.2.1. HOG feature extraction for defective data.After the UAV has acquired the multi-view sample data of the infrastructure line, the local features in the image need to be further extracted in order to better assist in realizing the image annotation, and provide data support for the defect recognition of the infrastructure line.

Based on this, this paper proposes a feature extraction method based on HOG, which is a local region description operator that describes the image by calculating the statistics of the histogram of the gradient direction on the local region of the image [26]. Its specific computational steps are as follows:
1) Normalization of the image.HOG uses the Gamma correction method to normalize the color space of the input image by performing the normalization operation to reduce the impact on the image due to changes in light intensity and shadows. Then:

$$I(x,y) = I(x,y)^{gamma}, gamma \in (0,1) \tag{1}$$

2) Image gradient calculation. The gradient of each pixel point in the processed image is calculated, mainly including the magnitude and direction of the gradient, for the pixel points in the image, the magnitude of the gradient and the direction of the gradient can be expressed as respectively:

$$G_x(x,y) = H(x+1,y) - H(x-1,y) \tag{2}$$

$$G_y(x,y) = H(x,y+1) - H(x,y-1) \tag{3}$$

$$G(x,y) = \sqrt{G_x(x,y)^2 - G_y(x,y)^2} \tag{4}$$

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_y(x,y)}{G_x(x,y)}\right) \tag{5}$$

where $H(x,y)$ is the pixel value at (x,y) in the image, $G_x(x,y)$ is the horizontal gradient value at (x,y) in the image, and $G_y(x,y)$ is the vertical gradient value at (x,y) in the image.
3) Calculate cellular unit feature descriptors. The image to be detected is divided into connected areas of the same size and named as cell unit Cells, and the feature histogram of the cell unit Cells is calculated by constructing the gradient direction histogram for each cell unit Cells and counting the gradient direction features of each cell in the cell unit. Depending on the features and size of the image, there are generally differences in the selection of the size of the cell cells, and different cell cell sizes will have different effects on the feature extraction of the image.
4) Constructing Feature Description Blocks and Normalizing Histograms.HOG uses Block blocks of the same size to combine individual Cells into large, spatially connected regions. Since Block blocks are generated by performing sliding windows on Cells, there is generally overlap between blocks. In addition, due to local illumination and changes in foreground-background contrast, the gradient intensities within the blocks may vary over a very large range, and thus a normalization operation of the gradient intensities within the blocks is also required to compress the illumination, shadows, and edge features in the blocks. The block after normalization is generally referred to as HOG feature descriptor.
5) Generate HOG features. The feature vectors of all the Blocks are concatenated after the feature extraction is completed, and finally the HOG feature vectors of the image are composed.

2.2.2. Labeling of Infrastructure Line Sample Pictures.Figure 2 shows the labeling process of the sample pictures of infrastructure lines. After nearly 1 month of centralized labeling work, this paper has collected about 60,000 pictures. Among them, the number of defective images that have been labeled is about 5,000, the number of images that clearly contain defects is about 10,000, and the possibility of defects contained in these images is about 50%, and the remaining about 50,000 images

are cycle patrol images, which contain a low amount of defects. The main task of the image labeling work is to identify and record the areas of parts and defects in the transmission line pictures, which can be divided into manual labeling and automatic machine labeling. The labeling strategies adopted are as follows:

1) For 10,000 image data parts with high defect content rate, manual labeling is mainly adopted.

2) For 50,000 images with a low defect content rate, manual labeling of some pictures is used first, and then automatic machine labeling is used. After the labeling is completed, the labels and corresponding images are stored, and the open data application interface forms a sample library of transmission line parts and defect images.

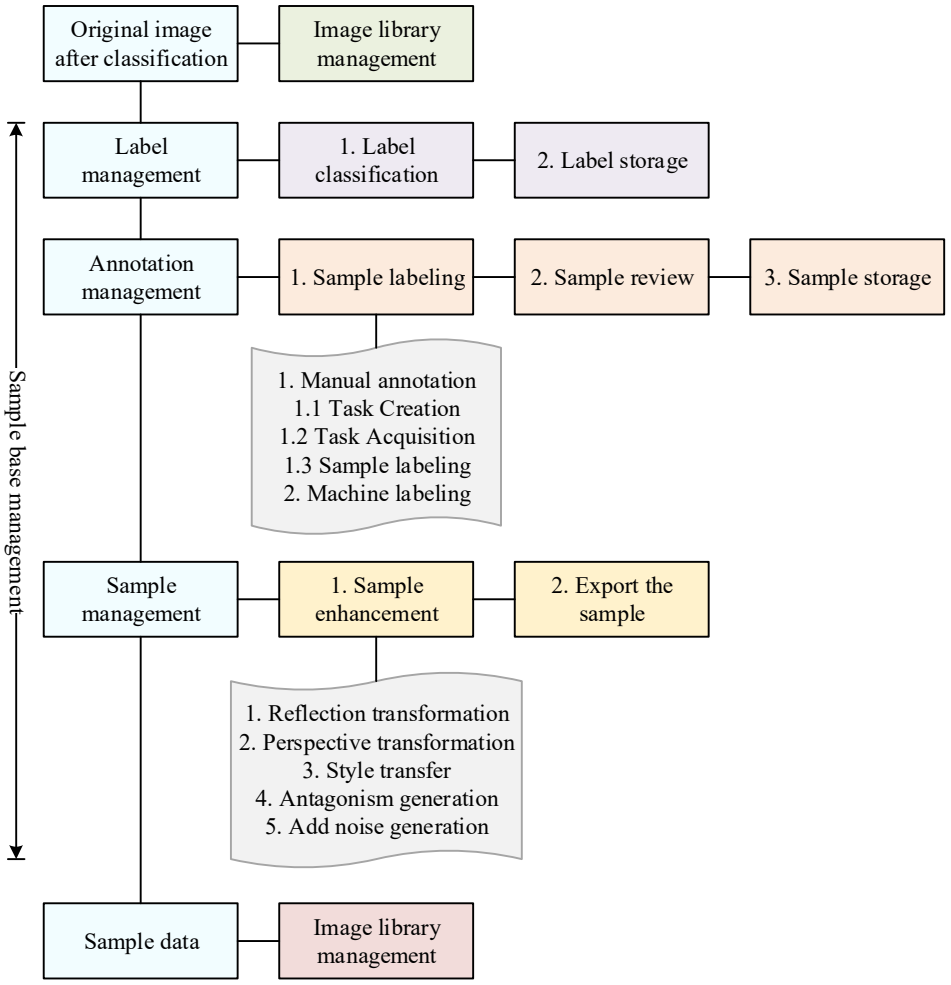


Fig. 2. The project of the infrastructure line sample image annotation process

2.2.3. Sample Library of Infrastructure Line Defects.After completing the feature extraction and image annotation of the UAV inspection sample data, a total of 60,000 high-resolution images taken by UAV infrastructure line inspection are acquired in this paper, which are named as UAV dataset. Among them, 50,000 are used as training samples and 10,000 are used as test samples. A large number of positive samples of infrastructure lines are intercepted from the images with a sample size of about 128*128 pixels.

Randomly intercept ROI rectangle images on the UAV inspection sample images without infrastructure lines as negative samples for training, the ROI rectangle area size is 256*256 pixels. The samples were uniformly scaled to the same size, and the scaled size size was 128*128 pixels. From the training samples, 30,000 positive samples of infrastructure lines and 10,000 negative samples are intercepted as a collection of training learning samples for infrastructure line defect recognition.

3. Improved YOLOv8 model

Due to the rapid development of China's electric power construction, the distance of transmission lines is getting longer and longer, and the number of long-distance overhead infrastructure lines is also gradually increasing, which makes it difficult to inspect and maintain them. In order to realize more accurate inspection and maintenance work, the drone inspection technology for long-distance overhead infrastructure lines came into being. With the drone multi-perspective infrastructure line to carry out all aspects of data collection, and then introduce deep learning technology to realize the effective identification of infrastructure line defects and accurate detection, and further enhance the stable operation of the transmission infrastructure line to provide reference.

3.1. Infrastructure Line Defect Identification Model

3.1.1. Overall Framework of Defect Identification Models.Taking the infrastructure line image samples collected from UAVs with multiple viewpoints as the data source, this paper constructs an efficient and accurate infrastructure line defect recognition model from the improvement of YOLOv8. The improved YOLOv8 model proposed in this paper mainly contains three parts, i.e., Backbone, Neck and Head, and its specific framework is shown in Fig. 3. First, the input image is preprocessed as necessary. Then, the processed image is sent to Backbone for feature extraction. Second, feature maps of the same size and semantic level extracted from the Backbone network are fused and processed to obtain feature representations of multiple scales, which usually include features of large, medium and small sizes. Finally, the fused features are sent to the detection head to generate target detection results.

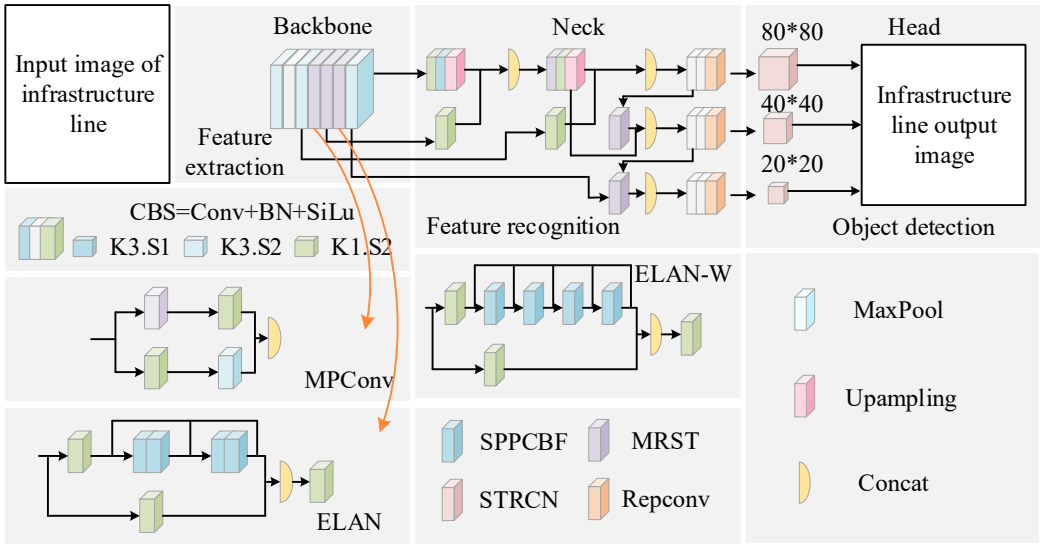


Fig. 3. The framework of the Improve YOLOv8 model

The backbone network consists of multiple convolutional and attentional mechanisms involving the Efficient Layer Aggregation Network Module, the Maximum Pooling Convolution Module, and the Convolutional and Attention Based Fusion Module. The Efficient Layer Aggregation Network Module consists of multiple convolutional modules linked across layers, enabling the network to learn more features by limiting the range of the gradient or adjusting the direction of the gradient. The Maximum Pooling Convolutional Module is used for downsampling and consists of convolutional modules and a maximum pooling layer. The convolutional attention based spatial pyramid module adds multiple maximal pooling layers in series to the convolution, which combines with the convolutional attention to form a cross-stage localized network structure that helps the network capture feature information at different scales. The neck section uses path aggregation networks for multi-scale feature fusion.

Introducing a multiscale feature aggregation diffusion module after the spatial pyramid module of convolutional attention better integrates the features output from the backbone network to achieve more effective multiscale feature fusion. On the basis of the multiscale feature aggregation diffusion module, a window transformer-based downsampling module is designed to downsample the neck features. The output side uses the IDetect detection head to categorize large, medium, and small images based on the fused feature values, and finally outputs the bounding box of the target, its confidence level, and its corresponding category.

3.1.2. Convolution and Attention Fusion Module.The convolution and attention fusion module consists of a tandem combination of a single-line channel feature extraction module and a parallel spatial feature extraction module, which enhances the sensitive information from the channel and spatial dimensions, respectively, to realize the enhanced extraction of important features in the process of multi-scale feature fusion [27]. The convolution and attention fusion module (C2F-CAF) first performs 2D convolution operation on the input feature maps, then performs batch normalization operation to accelerate the model convergence and alleviate the problems of gradient vanishing and gradient explosion, and finally introduces a nonlinear transformation using Hardswish activation function to complete the activation operation on the batch of normalized feature maps, and ultimately obtains the feature maps.Hardswish as the improved version of Swish and ReLU activation function, which can significantly reduce the computational cost while guaranteeing the accuracy. Its expression is:

$$Hardswish(x)=\begin{cases}0 & x\le -3\\ x & x\ge 3\\ \dfrac{x(x+3)}{6} & -3<x<3\end{cases} \tag{6}$$

First F_1 is fed into the channel attention module, and the feature map undergoes a global average pooling operation in the channel dimension to obtain the aggregated feature F_{CA}. Cross-channel information interaction is then achieved using a 2D convolution, with the convolution kernel size adaptively adjusted by a function; Finally, the channel feature weight vector W_{CH} is obtained through the Sigmoid activation function, and the weighted channel feature Figure F_{CH} can be obtained by multiplying W_{CH} by the original input feature F_1. The spatial attention mechanism is directly connected in series after the channel attention mechanism, and the adaptive global maximum pooling and global average pooling in the spatial dimension are performed respectively using the channel weight feature map F_{CH} as input. Then the pooling results F_{SM} and F_{SA} are completed spliced on the channel dimension to increase the number of channels, and the spatial feature weight vector W_{SP} is obtained after 2D convolution operation and Sigmoid function activation, which is multiplied with the channel weight feature map F_{CH} to obtain the hybrid weight feature map F_2. Considering the computational cost and the characteristics of the feature information of the small objects, a 3×3 convolution kernel is used. The above process can be expressed as:

$$F_1(x)=Hardswish(BatchNorm(Conv2d^{3\times 3}(x,W_{conv},b_{conv}))) \tag{7}$$

$$W_{CH}(F_1)=\sigma(Conv2d(Avgpool(F_1))) \tag{8}$$

$$W_{SP}(F_{CH})=\sigma(Conv2d^{3\times 3}(Concat(Maxpool\ (F_{CH})Avgpool(F_{CH}))) \tag{9}$$

$$W_{SP}(F_{CH})=\sigma(Conv2d^{3\times 3}([F_{SM},F_{SA}])) \tag{10}$$

3.1.3. Multi-scale feature aggregation diffusion module.In order to solve the situation of low defect recognition rate as well as high computational complexity caused by different scale features of

infrastructure line image samples collected by UAVs from multiple viewpoints, this paper proposes a multi-scale feature aggregation and diffusion pyramid network. The network enhances the global interaction ability of different scale features through the strategy of multi-scale aggregation and diffusion, shortens the information transfer path, and is more capable of overcoming the defect recognition limitations caused by different scale features on the basis of optimizing the computational complexity.

In the feature aggregation stage, a multi-scale feature aggregation module (MSFAM) is designed to integrate feature maps from different levels to realize the direct fusion of fine-grained and semantic information and enhance the comprehensive expression ability of features. Specifically, MSFAM first splices feature maps P_2, P_3, P_4 from different layers in the channel dimension to generate feature maps containing multiscale information as:

$$F = Concat(P_2, P_3, P_4) \tag{11}$$

where Concat denotes the channel splicing operation and F denotes the feature map obtained by splicing P_2, P_3 and P_4 over the channel.

Subsequently, the convolution operation is performed using dilation convolutions with dilation rates of 1, 3 and 5. In the feature fusion stage, by combining information from these different scales, the model is able to form a more expressive feature representation. This process can be expressed in a mathematical formula can be expressed as:

$$F_1 = DWConv_{3\times3}^{5}(F) \tag{12}$$

$$F_2 = DWConv_{3\times3}^{3}(F) \tag{13}$$

$$F_3 = DWConv_{3\times3}^{1}(F) \tag{14}$$

where $DWConv_{k\times k}^{d}$ represents a deep convolution with convolution kernel size k and dilation rate d. F_1, F_2 and F_3 represent the feature maps obtained after the convolution operation with different dilation rates, respectively.

In addition, considering that convolution is usually good at capturing spatial information but tends to have limited capture of edge details, an additional edge detection branch based on the Sobel operator is added to extract the edge information of the image, i.e.:

$$F_4 = Sobel(F) \tag{15}$$

where F_4 represents the feature map generated after convolution based on the Sobel operator.

Finally, by fusing the original feature map F with the feature map F_1, F_2, F_3 after multi-scale convolution and the feature map F_4 containing a large amount of edge information, MSFAM is able to generate a more comprehensive feature representation F_5, which helps the model to better understand the image content. The feature aggregation stage can be expressed as Eq:

$$F_5 = Conv_{1\times1}(F + F_1 + F_2 + F_3 + F_4) + F \tag{16}$$

where $Conv_{1\times1}$ denotes the convolution operation with a convolution kernel size of 1, F_5 denotes the output feature map of the MSFAM module, and the symbol + denotes the element-by-element addition operation.

In the diffusion stage, the aggregated and integrated features are assigned to each feature layer of different scales, so that each scale can directly benefit from the rich information after fusion, which in turn enhances the model's ability to detect multi-scale targets. Specifically, the propagation path of feature information is shortened through the aggregation of multi-scale features at the intermediate layer followed by the diffusion of features from the intermediate layer to the neighboring layers on both sides. This effectively prevents the possible attenuation or loss of information in the long

transmission path, effectively reduces the computational complexity of the model, and further improves the computational efficiency of the model for the identification of defects in infrastructure lines.

3.2. Defect Recognition Loss Function Design

3.2.1. Classification Loss Function Design.In this paper, the Focal Loss function and the balanced cross-entropy function are used to improve the classification loss function of YOLOv8. In the infrastructure line defect recognition problem, the infrastructure line to be localized is called foreground, and the other parts are called background, and there is an imbalance between the complexity of foreground and background in the samples of UAV multi-view aerial images of infrastructure lines. In order to solve this problem, this paper adds new modulation parameters to improve the cross-entropy function in the loss function of YOLOv8.The original loss function of YOLOv8 consists of 4 parts, i.e.:

$$L_{YOLOv8} = l_{xy} + l_{wh} + l_c + l_{class} \tag{17}$$

Where L_{YOLOv8} is the original loss function of YOLOv8, l_{xy} is the loss of the center point calculated by the cross-entropy function, l_{wh} is the loss of the size of the box calculated by the mean square error function, l_c is the loss of confidence calculated by the cross-entropy function, and l_{class} is the loss of the category calculated by the cross-entropy function. In this paper, we replace the cross-entropy function in the confidence loss with the FocalLoss function.The FocalLoss function is as follows:

$$L_1 = \alpha(1-p_t)^{\gamma} \lg p_t \tag{18}$$

Where L_1 is the FocalLoss value, α,γ are hyperparameters, α is used to solve the insulator string sample complexity imbalance. γ is used to solve the foreground and background imbalance in the image, and p_t is the sigmoid function, defined as follows:

$$p_t = \frac{1}{1+e^{-x}} \tag{19}$$

Then the back propagation process of FocalLoss is as follows:

$$\begin{aligned} \frac{dL_1}{dx} &= \frac{dL_1}{dp_t}\frac{dp_t}{dx} \\ &= \alpha(1-p_t)^{\gamma}(\gamma(\lg p_t)p_t - (1-p_t)) \\ &= \alpha(\frac{e^{-x}}{1+e^{-x}})^{\gamma}(\gamma(\lg\frac{1}{1+e^{-x}})\frac{1}{1+e^{-x}} - \frac{e^{-x}}{1+e^{-x}}) \end{aligned} \tag{20}$$

For the category loss function of YOLOv8, it is improved by using the balanced cross-entropy function, which is as follows:

$$L_2 = -\beta_s \log(q_t) \tag{21}$$

Among them:

$$q_t = \begin{cases} q, & y=1 \\ 1-q, & \text{otherwise} \end{cases} \tag{22}$$

$$\beta_s = \begin{cases} \beta, & y=1 \\ 1-\beta, & \text{otherwise} \end{cases} \tag{23}$$

where L_2 is the balanced cross-entropy value, γ is the true label of the infrastructure line samples, q is the predicted value of the infrastructure line samples, and β is a modulation variable to address the imbalance in the categories of the infrastructure line samples.

The improved YOLOv8 loss function still consists of four parts, l_{xy} for calculating the centroid prediction loss, l_{wh} for calculating the box size prediction loss, FocalLoss improved l_c and Equilibrium Cross Entropy improved l_{class}. After the improvement, the contribution of the samples that are in the minority in the dataset to the loss function will be enhanced.

3.2.2. Bounding box regression loss function design.In target detection, the bounding box regression loss function is used to measure the difference between the model's predicted bounding box and the true bounding box and optimize the model by minimizing the difference. In YOLOv8 algorithm, CIOU is used as the bounding box regression loss function of the model [28]. CIOU introduces the Euclidean distance penalty term of the centroid of the predicted box and the real box and the bounding box aspect ratio penalty term on the basis of IOU.The CIOU bounding box regression loss function is defined as follows:

$$IOU = \frac{B_{gt} \cap B_{prd}}{B_{gt} \cup B_{prd}} \tag{24}$$

$$CIOU = IOU - \frac{\rho^2(b_{gt}, b_{prd})}{C^2} - \alpha V \tag{25}$$

$$V = \frac{4}{\pi^2}\left(\arctan\frac{w_{gt}}{h_{gt}} - \arctan\frac{w_{prd}}{h_{prd}}\right)^2 \tag{26}$$

$$\alpha = \frac{V}{1 - IOU + V} \tag{27}$$

where B_{gt} and B_{prd} represent the true and predicted boxes, respectively, b_{gt} and b_{prd} are the coordinates of the centers of the true and predicted boxes, ρ^2 is the square of the distance between the two centers, and C is the length of the diagonal of the smallest outer bounding box that contains the predicted and true boxes. α and V are parameters that measure the aspect ratio of the bounding box, w_{gt} and h_{gt} represent the width and height of the real box, and w_{prd} and h_{prd} represent the width and height of the predicted box.

CIOU uses $R = \frac{\rho^2(b_{gt}, b_{prd})}{C^2}$ in the penalty term. As C increases, R becomes smaller and vice versa. Therefore, when the real and predicted frames do not overlap, L_{IoU} stays the same and only the size of the predicted frame needs to be increased to decrease L_{CloU}, which shows that it is not reasonable to use C as the denominator in the penalty term.

To address this problem, an alternative to CIOU for PIOU is introduced, which uses the target frame size as the penalty factor in the denominator and combines it with an adaptive anchor frame quality function to enhance the ability to pay attention to medium-quality anchor frames and guide the anchor frames to regress efficiently along more direct paths, so as to provide the model with faster convergence speed and higher accuracy.

The PIOU bounding box regression loss function is defined as follows:

$$P = \left(\frac{dw_1}{w_{gt}} + \frac{dw_2}{w_{gt}} + \frac{dh_1}{h_{gt}} + \frac{dh_2}{h_{gt}}\right) / 4 \tag{28}$$

$$f(x) = 1 - e^{-x^2} \tag{29}$$

$$PIoU = IoU - f(P), -1 \le PIoU \le 1 \tag{30}$$

$$L_{PIoU} = L_{IoU} + f(P), 0 \le L_{PIoU} \le 2 \tag{31}$$

where P is a penalty factor adapted to the target size, dw_1, dw_2, dh_1 and dh_2 are the absolute values of the distances between the corresponding edges of the prediction frame and the target frame, respectively, w_{gt} and h_{gt} are the width and height of the target frame, and $f(x)$ is the adaptive anchor frame quality function.

The denominator of the penalty factor P depends only on the size of the target box and is independent of the anchor box and the smallest outer box of the target box, and the size of the anchor box does not affect P. Unless the anchor box completely overlaps with the target box, P never degrades to 0. For very poor anchor boxes, such as those that do not overlap at all or those that are very inaccurately positioned, which are difficult to be adjusted to the correct position during the optimization process, their gradients should be small in order to avoid having an have an excessive impact. For very good anchor boxes, such as those close to the true box, the error is very small and the gradient should be small to avoid over-adjustment, which causes the predicted box to deviate from the correct position.

4. Infrastructure line defect identification validation

For transmission infrastructure line inspection, there are a variety of ways to choose from, mainly divided into two categories: manual inspection and machine inspection. With the development of artificial intelligence and drone technology, drone power inspection has become a viable option. Drone power inspection is safer and more efficient than manual inspection, and it can carry visible light cameras, infrared cameras, temperature and humidity sensors and other equipment to photograph the power equipment in the transmission line, and identify the inspection images through artificial or AI to determine whether there is a fault. Introducing the application of deep learning technology, the use of computer vision and image processing technology can reduce the problem of artificial misjudgment, combined with embedded devices, to realize intelligent inspection to improve the accuracy and efficiency of inspection.

4.1. Experimental Setting and Assessment Indicator Setting

4.1.1. Simulation experiment environment setup.The operating system used in this experiment is Windows 10 64-bit flagship, the GPU is NVIDIA RTX 3090Ti, the development platform is PyCharm, the programming language is Python, the deep learning framework is PyTorch, and the GPU acceleration library is CUDA.Table 1 shows the specific parameter settings.

Table 1. Experiment parameter settings

Parameter	Value	Parameter	Value
Image size	128*128 pix	Momentum	0.952
Bulk size	20	Attenuation factor	0.00001
Learning rate	0.001	Training number	300

4.1.2. Model performance evaluation indicators.The average precision is used to evaluate the accuracy of the model's pose estimation over a range of detected instances. The formula for the OKS-based AP is:

$$AP = \frac{\sum_{i=1}^{n} \delta(OKS_i > T)}{n} \tag{32}$$

Where n represents the number of instances in a series of detected images, and T represents the manually set threshold of OKS, so the meaning of average precision is the ratio of the number of instances in which the similarity of keypoints among p instances is greater than a threshold value to the total number of instances.

Generally according to the different settings of T the meaning represented by the AP is different, in order to ensure the objectivity of the evaluation, this paper will T set the value to take a value every 0.05 from 0.5-0.95 for AP calculation, where $T = 0.5 + 0.05k, k = 0,1,2,3...9$ and then average the results, and ultimately get the mean average accuracy mAP, the calculation formula can be expressed as:

$$mAP = \frac{\sum_{k=0}^{9} AP_{0.5+0.05k}}{10} \tag{33}$$

The number of billion floating point operations per second (GFLOPs) is commonly used to evaluate the computational performance of computing devices. In deep learning tasks, it can be used to evaluate the computational complexity of the model. Taking the neural network convolution module as an example, the complexity calculation formula of this module can be expressed as:

$$FLOPs = 2HW(C_{in}K^2)C_{out} \tag{34}$$

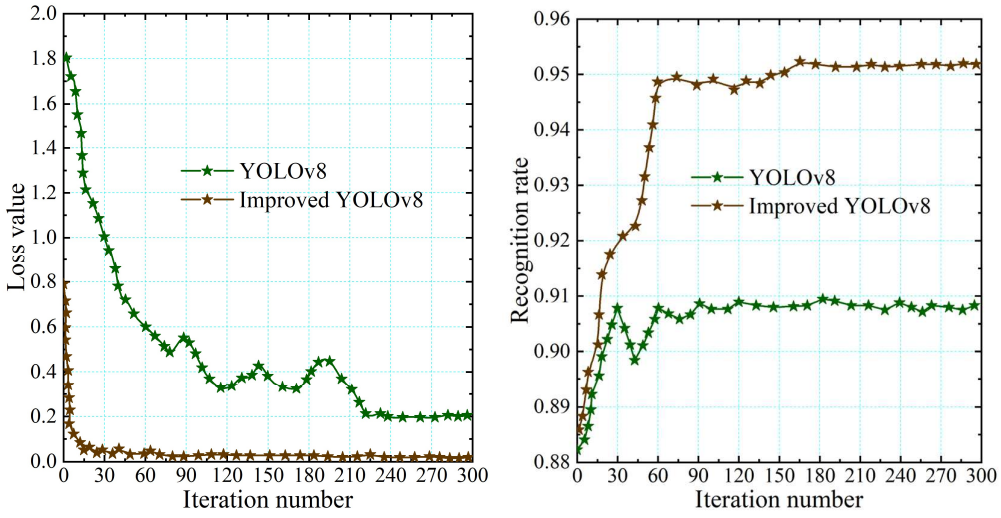
$$1GFLOPs = 1 \times 10^9 FLOPs \tag{35}$$

Where C_{in} refers to the number of input tensor channels, C_{out} refers to the number of output tensor channels, and K refers to the size of the convolutional kernel. In deep learning models, the smaller the value of GFLOPs represents the lower the computational complexity of the model, i.e., theoretically on the same performance of the computing device, the faster the inference speed of the model.

4.2. Infrastructure Line Defect Identification Model Performance

4.2.1. Effect of different network training.The UAV dataset established in the previous section is divided into training set and test set according to the ratio of 9:1, and the recognition rate and loss function are used to compare the infrastructure line defect recognition effect of the improved YOLOv8 model and the original YOLOv8 model. Fig. 4 shows the recognition effect of infrastructure line defects trained by different networks, where Fig. 4(a)~(b) shows the comparison results of loss function and recognition rate, respectively.

Based on the recognition effect of different network training the following conclusions can be drawn:
1) The loss function training in both networks shows a decreasing trend, and the first half of the decline rate is relatively fast, but the loss function of the unimproved YOLOv8 model tends to stabilize after 225 iterations, and the loss function decreases slowly and fluctuates non-stop before 225 iterations, while the improved YOLOv8 model tends to stabilize after 50 iterations, which indicates that the improved YOLOv8 model converges This indicates that the convergence of the improved YOLOv8 model is faster. The unimproved YOLOv8 model converges to about 0.2, while the improved YOLOv8 model converges to about 0.01, which is much smaller than that of the unimproved YOLOv8 model, indicating that the training effect of the improved YOLOv8 model is better.
2) The recognition effect of the unimproved YOLOv8 model is around 0.909, while the improved YOLOv8 model tends to stabilize after 165 iterations, and its recognition rate for the defects of the infrastructure line is around 0.955, which is due to the fact that this paper combines the attention mechanism and the multi-scale features with each other to improve the model's ability of mining the defects of the infrastructure line for the small targets, and retains the image's feature vectors, so the recognition accuracy is high.



(a) Loss function (b) Recognition rate

Fig. 4. Identification of different network training

4.2.2. Defect Recognition Comparison Experiment.In order to study the performance of the infrastructure line defect recognition model designed in this paper, the number of model parameters (Params), training accuracy (TA), inference accuracy (IA), inference time (IT) and frame rate (FR) are additionally selected as evaluation metrics in addition to the computational complexity of the model (GFLOPs) given in the previous paper. While keeping the experimental conditions unchanged, we selected the YOLOv9 algorithm as well as other YOLO series as representatives of advanced single-stage detectors, and added the classical two-stage detection model Faster RCNN as well as another efficient single-stage detection model, SSD, to comprehensively compare the performance differences of the different models when deployed on the device.The backbone of the Faster RCNN network uses Resnet and the backbone network of SSD uses Mobilenet.Table 2 shows the performance comparison results of different detection and recognition models.

As can be seen from the table, the improved YOLOv8 model proposed in this paper shows significant advantages compared with the v6 to v8 models of the same scale in the YOLO series. In terms of inference accuracy, the model in this paper achieves 3.84, 7.46 and 3.17 percentage points of improvement compared to YOLOv6, YOLOv7 and YOLOv8, respectively. Meanwhile, the inference time is reduced by 16.91%, 21.09% and 11.46% compared to the above three models, which effectively shortens the inference time. Although the YOLOv9 algorithm improves 0.77 percentage points in inference accuracy compared with the model in this paper, this advantage comes at the expense of inference duration, which increases by 1.33 times. In contrast, this paper's model achieves a better balance between inference accuracy and inference duration. The model in this paper is only 0.85×109 higher than YOLOv5 in terms of model computational complexity, which is lower than other same-scale models, and 0.25×106 and 0.17×106 higher than v5 and v9 same-scale models in the YOLO series in terms of the number of model parameters, which effectively controls the model scale and facilitates deployment in resource-constrained environments. In terms of frame rate performance, the model in this paper reaches 36.71 f/s, which is 20.09%, 26.85%, 13.09%, and 131.9% higher than the v6, v7, v8, and v9 same-scale models. It can be seen that the model in this paper ensures high accuracy and maintains efficient inference speed while considering the computational complexity of the model.

In the comparison with the two-stage model Faster-RCNN, the computational complexity and the number of parameters of this paper's model are much lower than that of Faster-RCNN, and at the same time, it reduces the inference duration by 93.83% and improves the inference accuracy by 6.64 percentage points over Faster-RCNN, which can not be deployed on a device due to its excessive number of parameters and model complexity for inference. Compared with another popular single-

stage detection model SSD, the model in this paper reduces the inference speed by 17.56% and improves the inference accuracy by 14.12 percentage points compared to SSD. The experimental results show that the improved YOLOv8 model proposed in this paper significantly outperforms other mainstream target detection models in defect identification of infrastructure lines, with better inference accuracy, shorter inference time, and relatively low computational complexity of the model, which provides reliable support for efficiently analyzing defect localization of infrastructure lines.

Table 2. Comparison results with other detection models

Model	GFLOPs/109	Params/106	TA/%	IA/%	IT/ms	FR/(f/s)
YOLOv5	4.52	1.98	80.95	77.14	26.65	37.69
YOLOv6	11.48	4.72	87.04	83.45	32.83	30.57
YOLOv7	13.75	6.25	82.93	79.83	34.57	28.94
YOLOv8	8.81	3.17	88.46	84.12	30.81	32.46
YOLOv9	7.63	2.06	89.51	88.06	63.46	15.83
Faster-RCNN	370.26	138.74	81.15	80.65	442.12	2.34
SDD	1.89	6.26	73.06	73.17	33.09	30.15
Ours	5.37	2.23	90.48	87.29	27.28	36.71

In order to further verify the effectiveness of this paper's model in identifying defects in infrastructure lines, mAP is used as an evaluation index to compare the defect identification accuracy of infrastructure lines under different models. Table 3 shows the comparison results of the detection indexes of different models, where AP1 and AP2 are the average detection accuracies of normal and defective infrastructure lines, respectively.

From the data in the table, it can be seen that compared with Faster-RCNN, SDD and YOLOv8 models, YOLOv8 is better in overall detection, and is slightly lower than YOLOv5 in the detection of defects in infrastructure lines, probably due to the fact that YOLOv5 has been updated from its launch to now with many versions so that its functions are relatively mature while YOLOv8 is still in the research stage. In this paper, the mAP value of the improved YOLOv8 algorithm reaches 0.935, which is 2.41% higher than that of the original YOLOv8, and the average accuracies of recognizing the normal and defective infrastructure lines are 0.945 and 0.929, which are 1.39% and 3.92% higher than that of the original 0.913 network, respectively.The average detection accuracy of the model of YOLOv8 algorithm is 1.39% and 3.92% higher than that of the original 0.913 network, after the fusion of the convolutional attention module and the multi-scale feature aggregation diffusion module, the average detection accuracy of the model is improved, especially for the infrastructure line defect dataset. From the experimental results, it can be seen that the introduction of the convolutional attention module and multi-scale feature aggregation and diffusion module in the YOLOv8 backbone network realizes the multi-scale feature fusion of local and global information, enhances the semantic expression, and better realizes the defect identification and detection of transmission lines.

Table 3. Comparison of detection indexes of different models

Model	mAP@0.5	AP1	AP2
YOLOv5	0.893	0.913	0.898
YOLOv6	0.905	0.917	0.896
YOLOv7	0.909	0.924	0.895
YOLOv8	0.913	0.932	0.894
YOLOv9	0.926	0.938	0.911
Faster-RCNN	0.897	0.906	0.887
SDD	0.891	0.909	0.882
Ours	0.935	0.945	0.929

4.3. Comparative validation analysis of different improvement strategies

4.3.1. Comparison of different loss functions.In order to verify the optimization effect of different loss functions on the model, the experiments compare the performance of the improved YOLOv8 model with IoU, CIoU, DIoU, EIoU, SIoU, and PIoU loss functions respectively. On the basis of mAP, frame rate, and model size, precision (P), recall (R), and F1 are additionally selected as evaluation indexes, and the experimental results are shown in Table 4.

As can be seen from the table, compared to using IoU as the loss function, the P, R, mAP, F1 values and detection speed of the improved model after using PIoU as the loss function are increased by 2.18%, 1.27%, 1.12%, 0.09% and 5.45%, respectively, and the model size is unchanged. When CIoU, DIoU, EIoU, and SIoU are used as loss functions, respectively, the mAP and F1 values of the improved model are significantly reduced, which cannot satisfy the high accuracy requirements of the defect recognition model for infrastructure lines. The results show that PIoU as a loss function can focus on anchor frames with different qualities and improve the detection accuracy and speed of the infrastructure line defect recognition model.

Table 4. Performance comparison of different loss function

Loss	P/%	R/%	mAP/%	F1/%	FR(f/s)	Size/MB
IoU	91.15	90.31	95.68	91.21	133.38	6.73
CIoU	92.21	90.12	95.34	91.15	134.72	6.73
DIoU	92.21	89.45	95.52	90.74	122.83	6.73
EIoU	91.06	90.53	95.46	90.63	134.51	6.73
SIoU	92.38	89.62	95.17	90.96	130.15	6.73
PIoU	93.14	91.46	96.75	92.07	140.65	6.73

4.3.2. Comparison of Attention Mechanisms.This experiment compares the impact on the defect recognition and detection performance of infrastructure lines after introducing different attention modules in front of each component of the multi-scale feature aggregation diffusion module of the improved YOLOv8 model. Its specific experimental results are shown in Table 5. Where SE, CAM, C2F-CAF are channel attention, hybrid attention and convolutional attention, respectively.

From the data in the table, it can be seen that the mAP of the network is not improved after the fusion of the SE module, and although the precision rate is improved by 1.19 percentage points, the recall rate decreases by 0.85 percentage points, which indicates that for the detection task such as defect recognition and detection of infrastructure lines, which has a complex background and a large number of targets, the effect of weighting the feature maps only in the channel dimension is not ideal. The mAP of the network after integrating the CAM module is slightly higher than that of SE.

Considering that the weighting on a single dimension does not enhance the effect well, the CAM performs the serial connection of the attention module on both the channel and spatial dimensions, and thus there is a 1.33% improvement in the mAP. The C2F-CAF module proposed in this paper, on the other hand, performs better compared to the CAM module, with a 0.64% improvement in mAP compared to the network incorporating CAM. This implies that the C2F-CAF module not only better captures the details of the image in terms of channel and spatial dimensions, but also fully utilizes the original feature maps to preserve the rich semantics, thus focusing on the defective features of the infrastructure lines more comprehensively and accurately.

Table 5. Contrast of the attention mechanism (%)

Model	P	R	F1	mAP
YOLOv8	79.15	90.27	89.38	90.12
YOLOv8+SE	80.34	89.42	90.25	90.12
YOLOv8+CAM	80.93	90.53	91.07	91.45
YOLOv8+C2F-CAF	82.76	91.84	91.42	92.09

4.3.3. Analysis of model ablation experiments.In order to prove that each improvement point improves the performance of the model and show the experimental situation more clearly, the ablation experiment was carried out by setting up the models that added different modules one by one as a comparison. The results of the ablation experiment are shown in Table 6. As can be seen from the table, no improvement module was added to the baseline model of experiment A, and the mAP@0.5 value and parameter amount were 90.58% and 11.15MB, respectively. Taking the results of experiment A as the evaluation criteria, the convolutional attention, multi-scale feature aggregation diffusion and PIoU loss function were added to experiments B, C and D, respectively, and the mAP@0.5 values were increased by 1.26%, 2.17% and 1.18%, respectively, but the introduction of multi-scale feature aggregation diffusion module would slightly increase the number of model parameters. On the whole, the improved YOLOv8 model after adding each module has a great improvement in the defect identification ability of infrastructure lines, while the change range of parameters is small. This shows that the improved YOLOv8 model designed in this paper is feasible to be applied to the defect identification of infrastructure lines, has faster computational efficiency, and can effectively reduce the influence of complex background of infrastructure line images, so as to improve the defect recognition ability of the model.

Table 6. Ablation experiment results

Test	C2F-CAF	MSFAM	PIOU	mAP@0.5/%	Params/MB
A	×	×	×	90.58	11.15
B	√	×	×	91.84	11.15
C	×	√	×	92.75	12.14
D	×	×	√	91.76	11.15
E	√	√	×	92.68	11.26
F	√	×	√	93.42	12.14
G	×	√	√	93.75	12.14
H	√	√	√	94.29	12.25

5. Conclusion

In this article, image data of grid infrastructure lines are acquired by UAV, and the YOLOv8 model is

improved to realize the accurate recognition of infrastructure line defects based on the consideration of computational complexity. The inference accuracy of this paper's model is improved by 3.84, 7.46 and 3.17 percentage points compared with YOLOv6, YOLOv7 and YOLOv8, respectively. The model in this paper is only 0.85×109 higher than YOLOv5 in terms of model computational complexity, which is significantly lower than other models of the same scale, and only 0.25×106 and 0.17×106 higher than v5 and v9 models in the YOLO series in terms of model parameter counts.The improved YOLOv8 model combines convolutional attention with multiscale feature-aggregated diffusion, which can better obtain the infrastructure line defects in the complex background of the images taken by the UAV, which better guarantees the safe and stable operation of the grid infrastructure lines.

References

[1] Suroso, I., & Irmawan, E. (2018). Analysis Of Aerial Photography With Drone Type Fixed Wing In Kotabaru, Lampung. Journal of Applied Geospatial Information, 2(1).

[2] Wong, Y. W. A., Ernesto, P., & Elias, J. (2022). Comparative study of aerial photography/(UAV)-drone vs 16th century cityscape art. IDA: International Design and Art Journal, 4(1), 57-75.

[3] Li, R., & Wu, M. (2024). Revealing Urban Color Patterns via Drone Aerial Photography—A Case Study in Urban Hangzhou, China. Buildings, 14(2), 546.

[4] Sugden, J. (2020). Drone Photography: Art and Techniques. The Crowood Press.

[5] Petrou, M. M., & Kamata, S. I. (2021). Image processing: dealing with texture. John Wiley & Sons.

[6] Kuruvilla, J., Sukumaran, D., Sankar, A., & Joy, S. P. (2016, March). A review on image processing and image segmentation. In 2016 international conference on data mining and advanced computing (SAPIENCE) (pp. 198-203). IEEE.

[7] Syeda, I. H., Alam, M. M., Illahi, U., & Su'ud, M. M. (2021). Advance control strategies using image processing, UAV and AI in agriculture: a review. World Journal of Engineering, 18(4), 579-589.

[8] Lei, B., Wang, N., Xu, P., & Song, G. (2018). New crack detection method for bridge inspection using UAV incorporating image processing. Journal of Aerospace Engineering, 31(5), 04018058.

[9] Dewangan, V., Saxena, A., Thakur, R., & Tripathi, S. (2023). Application of image processing techniques for uav detection using deep learning and distance-wise analysis. Drones, 7(3), 174.

[10] Ansari, E., Akhtar, M. N., Abdullah, M. N., Othman, W. A. F. W., Bakar, E. A., Hawary, A. F., & Alhady, S. S. N. (2021). Image processing of UAV imagery for river feature recognition of Kerian River, Malaysia. Sustainability, 13(17), 9568.

[11] Mohd Noor, N., Abdullah, A., & Hashim, M. (2018, July). Remote sensing UAV/drones and its applications for urban areas: A review. In IOP conference series: Earth and environmental science (Vol. 169, p. 012003). IOP Publishing.

[12] Kujawski, A., & Nürnberg, M. (2023). Analysis of the potential use of unmanned aerial vehicles and image processing methods to support road and parking space management in urban transport. Sustainability, 15(4), 3285.

[13] Santos, V. G., Pereira, D. S., Alsina, P., Fernandes, D. H., Nascimento, L. B., Leite, D. L., ... & Souza, E. S. (2019). Multi-uav system architecture for environmental protection area monitoring. Proc. Anais do Simpsio Brasileiro de Automao Inteligente, 1-6

[14] Green, D. R., Hagon, J. J., Gómez, C., & Gregory, B. J. (2019). Using low-cost UAVs for environmental monitoring, mapping, and modelling: Examples from the coastal zone. In Coastal management (pp. 465-501). Academic Press.

[15] Lyapuntsova, E., Belozerova, I., Drozdova, I., & Korol, O. (2019). Safety in construction in the field of investment in urban infrastructure. In E3S Web of Conferences (Vol. 97, p. 06034). EDP Sciences.

[16] Zhang, X., Wu, Y., Skitmore, M., & Jiang, S. (2015). Sustainable infrastructure projects in balancing urban–rural development: Towards the goal of efficiency and equity. Journal of Cleaner Production, 107, 445-454.

[17] Sepasgozaar, S. M., Shirowzhan, S., & Wang, C. C. (2017). A scanner technology acceptance model for construction projects. Procedia engineering, 180, 1237-1246.

[18] Lee, S., & Yu, J. (2017). Discriminant model of BIM acceptance readiness in a construction organization. KSCE Journal of Civil Engineering, 21, 555-564.

[19] Olaniyan, R., Halkias, D., & Neubert, M. (2020). The Technology Acceptance Among Construction Project Managers in Nigeria: A Narrative Literature. Olaniyan, Rasaq," Barriers to Technology Adoption Among Construction Project Managers in Nigeria"(2019). Walden Dissertations and Doctoral Studies, 7832.

[20] Goulias, D. G., & Karimi, S. (2017). Risk Assessment for the Acceptance of Construction Materials. Risk, 13(12), 54-58.

[21] Zhang, J. (2022). Supervision method of indoor construction engineering quality acceptance based on cloud computing. Journal of Intelligent Systems, 31(1), 795-805.

[22] Kreiger, M., Kreiger, E., Mansour, S., Monkman, S., Delavar, M. A., Sideris, P., ... & Jones, S. (2024). Additive construction in practice–Realities of acceptance criteria. Cement and Concrete Research, 186, 107652.

[23] Liu, H., Sun, Y., Cao, J., Chen, S., Pan, N., Dai, Y., & Pan, D. (2022). Study on UAV parallel planning system for transmission line project acceptance under the background of industry 5.0. IEEE Transactions on Industrial Informatics, 18(8), 5537-5546.

[24] Senvar, O., & Ünver, S. (2022). An Overview to Unmanned Aerial Vehicles from Perspectives of Quality and Safety Management in Aviation. International Journal of Engineering Research and Development, 14(2), 917-940.

[25] Xiaomin Zhang,Mingjie Yang,Dong Wang,Bo Chen & Kai Song. (2024). A transmission line tower tilt detection method based on BeiDou positioning signals and UAV inspection images. Journal of Radiation Research and Applied Sciences(4),101192-101192.

[26] Yuhai He,Jiye Huang & Yiming Pan. (2024). A novel low-resource consumption and high-speed hardware implementation of HOG feature extraction on FPGA for human detection. Integration102208-.

[27] Lóránd Vatamány & Siamak Mehrkanoon. (2025). Graph Dual-stream Convolutional Attention Fusion for precipitation nowcasting. Engineering Applications of Artificial Intelligence109788-109788.

[28] Siliang Ma & Yong Xu. (2025). FPDIoU Loss: A loss function for efficient bounding box regression of rotated object detection. Image and Vision Computing105381-105381.