

# Smart UAV Infrared Image Defect Detection using YOLOv3 and Gaussian Mixture Models

Rong MENG\*, Zhilong ZHAO, Mengtian HE, Zhiyuan WANG, Yanbo DUAN

**Abstract:** To improve detection performance, a model with strong universality and adaptability to different environments was established. This study proposed a migration and optimization method for drone infrared image defect detection and recognition models based on deep learning and transfer learning. This method proposed a temperature prediction method for infrared images based on fully convolutional neural networks to reduce storage and transmission costs. In addition, the study introduced a pixel level infrared image defect segmentation method based on Gaussian mixture model to solve the background interference problem caused by large defect area angles. Experiments confirmed that the algorithm performed well in depth detection, achieving an accuracy of 90% and processing time of only 11 ms. Compared to other methods, this algorithm not only had a low false detection rate of 2.6%, but also reduced the running speed by 63 ms and 27 ms compared to the wavelet texture feature statistical analysis and the spectral residual visual saliency methods. In addition, through filtering and morphological processing, this algorithm provided a more complete defect area, almost eliminating noise, which was helpful for further defect area analysis. This has important practical significance for further improving the application of drone infrared image defect detection.

**Keywords:** detection; gaussian mixture model; infrared images; transfer learning; yoloV3

## 1 INTRODUCTION

Infrared image analysis is the analysis of images obtained by infrared spectroscopy technology, by detecting the temperature difference and heat distribution of objects, making abnormal hot spots appear quickly, helping to identify potential defects of structures or equipment, especially used in electrical, mechanical and construction industries [1]. This technology is closely linked to defect detection, which can effectively assist fault location and analysis, thus enhancing the robustness of production systems. With the widespread application of low altitude technology, the advantages of flight inspection methods have become increasingly prominent [2]. The Unmanned Aerial Vehicle (UAV) inspection system can transmit the captured infrared images back to the ground monitoring station. This can not only use ground systems to automatically judge based on infrared image features, but also provide professional personnel with manual judgment, to timely detect and eliminate potential faults in power systems, building equipment, et al. [3]. According to the principle of deep learning, Wang K. and other scholars designed an infrared imaging resolution method specifically to deal with power device faults. The method is verified by a series of experiments [4]. The experimental results finally confirm the validity of this resolution method, and demonstrate its excellent performance in identifying various defects of electric power devices. In response to the damage of spacecraft in hypervelocity collisions, Yang X. and other researchers proposed an innovative algorithm for spacecraft defect detection based on infrared imaging [5]. The algorithm incorporates multi-objective optimization techniques to balance detail preservation and noise reduction. In order to improve the accuracy of damage region segmentation, the researchers used a decomposition based multi-objective evolutionary algorithm to optimize. Experimental results show that the algorithm is effective in detecting spacecraft damage accurately. However, despite these significant advances, techniques for defect detection with infrared images still face several limitations. Infrared imaging is affected by environmental factors such as temperature, humidity, and

light conditions, which can affect image quality and algorithm accuracy. For some complex or small defects, existing infrared imaging techniques may be difficult to detect accurately, especially in the case of high background noise. Although deep learning offers high accuracy, it often requires a lot of computational resources, which may limit its application in resource-limited environments. Infrared non-destructive testing captures and converts infrared radiation into two-dimensional images, and intuitively reflects the temperature distribution of the object through Infrared Image (II) without touching or damaging it. However, currently, II files generally use reserved space to store the image temperature matrix to achieve temperature identification. This method can achieve accurate temperature inversion of each pixel in II [6]. However, the method of reserving space for storing temperature data results in the II file being too large, resulting in high storage and transmission costs, which affects the efficiency of using the II file. And a large number of temperature markers result in excessive redundancy of II temperature storage information. When it is necessary to collect a large number of II files on site, it can lead to excessive consumption of network transmission and file storage resources [7]. Therefore, this paper proposes a migration and optimization strategy for UAV infrared image defect detection based on deep learning and transfer learning. The key innovations include: 1) An FCN-based temperature prediction method to reduce storage costs of images; 2) A YoloV3 model optimized for multi-scale infrared defects; 3) An accelerated Gaussian mixture model for accurate defect segmentation. In order to improve the accuracy of defect recognition in UAV infrared images, the YoloV3 network was improved and a multi-scale adaptive mechanism was introduced to process images with high signal-to-noise ratio and insufficient feature details. At the same time, aiming at the problem of large background interference, the Gaussian mixture model is used to classify the temperature features in the red outflow data, and the pixel-level defects are segmented by this model to improve the processing speed. After filtering and de-noising, the segmentation accuracy of defect area can be effectively

enhanced, and storage requirements and labor costs can be reduced.

The research mainly includes five parts. Firstly, the background and significance of UAV II detection are mainly introduced. Secondly, an overview of UAV II testing is provided. Next are the research methods, which mainly include two sections. In section 2.1, the study proposes a defect detection and recognition method for infrared photos based on YoloV3 transfer learning. In section 2.2, a defect detection and recognition method for infrared photos is proposed by integrating YoloV3 transfer learning and Gaussian mixture model. The fourth part is about verifying the effectiveness of the research model. Finally, there is a summary of the research methods and an analysis of the experimental results. And the shortcomings of research methods and future research directions are proposed.

## 2 RELATED WORKS

When UAVs rapidly develop, they are widely used in monitoring, inspection, and search and rescue. Among them, the application of infrared camera technology on UAVs has attracted widespread attention. Infrared sensors can capture thermal radiation information of targets, helping to detect and identify various objects, such as buildings, equipment, vegetation, etc. Many scholars have conducted a series of studies on this issue. Peng X. et al. proposed a method for detecting insulators using UAV II, aiming to improve the automation level of collecting a large number of images at once. These experiments confirmed that compared to other methods, this method had a faster detection speed of approximately 28 images per second [8]. Xingguo T. et al. proposed a UAV II method for detecting surface defects in wind turbines, aiming to reduce the cost of detecting blade defects in wind turbines and address the harsh working environment. These experiments confirmed that this method had high accuracy when detecting in harsh environments [9]. Lee H. et al. proposed a temperature recording method based on infrared detection wavelet mechanism, aiming to quantitatively analyze it. These experiments confirmed that the detection method had high accuracy [10]. Ranjit S. et al. proposed an infrared thermal imaging technology, aiming at achieving early fault detection research on rotational speed and bearing load. These experiments confirmed that this method could quantitatively evaluate the temperature characteristics of bearings and achieve fault monitoring and diagnosis. This method could quantitatively evaluate the temperature characteristics of bearings based on their operating conditions, further achieving fault monitoring and diagnosis [11].

Deep learning (DL) and transfer learning, as two methods of machine learning, have achieved significant success in fields such as computer vision and natural language processing. They include image classification, object monitoring, speech recognition, and more. There are currently many studies related to both. Subramanian M. et al. proposed a method for optimizing maize leaf disease recognition using DL, aiming to improve recognition accuracy. These experiments confirmed that this method had an accuracy of over 93% for the classification of harmful parts of corn [12]. Wang Y. et al. proposed a

selective pseudo labelling method based on DL, aiming to achieve transfer learning between small samples. These experiments confirmed that this method improved the transfer accuracy of the algorithm after transfer learning by 0.5% [13]. Wei W. et al. proposed a dynamic transfer learning model for optimizing state function parameters, aiming to improve the stability and convergence of the learning model. These experiments confirmed that the optimized parameters improved the stability and convergence time of the DL model [14]. Chen C. et al. proposed a method of using transfer learning for robot grinding and polishing skills learning, aiming to improve the efficiency of robot data feature learning and the problem of multi-system conversion. These experiments confirmed that this method compensated for the shortcomings of traditional methods and further improved the learning ability of robots [15].

In summary, the current infrared defect detection technology is mainly divided into two categories, namely, traditional image processing methods and deep learning-based methods. Traditional image processing methods include filtering, edge detection, threshold processing and so on. These methods help identify defects by enhancing key features of the image. These methods often rely on artificially set parameters and thresholds, which makes them of limited adaptability in different application scenarios and environmental conditions. In addition, they are less efficient at processing complex or noisy images. Deep learning-based methods include convolutional neural networks, which can effectively identify and classify defects by learning features in large amounts of data. Transfer learning uses pre-trained networks to fine-tune new data sets, which can improve detection performance in the absence of large amounts of labelled data. However, deep learning methods require a large amount of labelled data for training, which may be difficult to obtain in some application scenarios. In addition, these methods have relatively high computational costs and high hardware requirements, and their applications in UAV defects are still limited. Infrared imaging technology can accurately measure pixel temperature, but the high resolution temperature data makes the file large and increases the storage and transmission cost. Therefore, a defect detection and recognition method for UAV infrared photos based on DL and transfer learning has been proposed in the study. It is used to effectively reduce the size of Infrared image file and improve the defect detection accuracy.

## 3 RESEARCH METHODS

The study proposes EfficientNet-YoloV3 to accurately locate and identify defects on II, based on infrared non-destructive testing technology, combined with mature UAVs, using digital image processing and DL. And it is also proposed to use a fully convolutional neural network (FCN) to calibrate the II temperature. Aiming at the large angle of the defect area and the susceptibility to temperature analysis interference, an optimized and accelerated Gaussian mixture model is proposed for pixel level segmentation of the defect area.

### 3.1 Infrared Photo Defect Detection Method based on YoloV3 Transfer Learning

Infrared thermal imaging technology can reflect the information of surface thermal radiation and internal heat loss of an object. After signal processing and imaging, it forms an image that can intuitively reflect the temperature information of the object [16, 17]. Infrared thermal

imaging systems can convert the thermal radiation distribution of objects into electrical signals, which are then quantified into images through signal amplifiers, enabling people to visually see objects at night. Fig. 1 shows the operation and imaging principles of an infrared thermal imaging system.

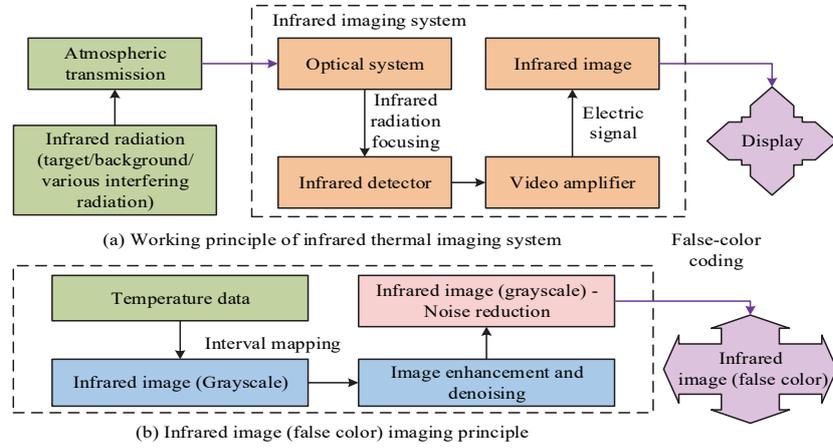


Figure 1 Working principle and imaging principle of infrared thermal imaging system

To observe the temperature distribution more intuitively, the temperature matrix is usually mapped to the grayscale range of the image, and the grayscale image is used to visually display the temperature distribution. Due to the poor resolution of grayscale images in human vision, single channel  $I$  is usually pseudo color encoded using a manually set colorimetric card [18]. Based on Stefan Boltzmann's law, when the temperature of the blackbody is  $T$ , Eq. (1) represents the total radiation intensity of the blackbody [19].

$$W_b = \sigma T_b^4 \tag{1}$$

In Eq. (1),  $\sigma$  is the Stefan Boltzmann constant. When the temperature of the measured object is  $T$ , Eq. (2) represents the total radiation intensity.

$$W_b = \varepsilon(T)\sigma T^4 \tag{2}$$

In Eq. (2),  $\varepsilon(T)$  represents the total emissivity of the measured object. Eq. (3) represents the temperature of the measured object.

$$T = T_b / \sqrt[4]{\varepsilon(T)} \tag{3}$$

Due to the fact that the emissivity of objects is less than 1.0, the temperature of the object is always higher than the blackbody temperature  $T$  [20]. According to Planck's radiation law, Eq. (4) is the surface temperature measured by the infrared thermal imaging system.

$$T_r^n = \tau_a [\varepsilon T_0^T + (1 + a) T_u^n] + \varepsilon_a T_a^n \tag{4}$$

In Eq. (4),  $\varepsilon$  represents the emissivity of the measured object.  $\varepsilon_a$  represents the atmospheric emissivity.  $\tau_a$

represents atmospheric transmittance.  $T_0$  represents the surface temperature of the object.  $T_u$ ,  $T_a$  represent environmental temperature and atmospheric temperature, respectively.  $T_r$  represents the temperature measured through an infrared thermal imaging system. Eq. (5) represents the true temperature of the measured surface.

$$T_0 = \left\{ \frac{1}{\varepsilon} \left[ \frac{1}{\tau_a} T_r^n - (1 - \alpha) T_u^n - \frac{\varepsilon_a}{\tau_a} T_a^n \right] \right\}^{1/n} \tag{5}$$

In Eq. (5),  $n$  represents the internal parameters of the system. The material and quality of the infrared imager itself have a significant impact on the measurement accuracy of temperature, often resulting in stripe noise. Polynomial regression can be used to fit the distribution of data points, and the function in Eq. (6) can be used to express the distribution of data.

$$y(x, W) = w_0 + w_1 x + \dots + w_m x^m = \sum_{i=0}^m w_i x^i \tag{6}$$

In Eq. (6),  $m$  stands for arbitrary constant.  $w_0, w_1, \dots, w_m$  represent the coefficients of the polynomial.

If  $W = \begin{bmatrix} w_0 \\ w_1 \\ \dots \\ w_m \end{bmatrix}$ , the linear algebraic form shown in Eq. (7)

can be obtained.

$$y(x, W) = XW \tag{7}$$

To evaluate the quality of function fitting, it is necessary to use a loss function to calculate the error between the predicted value of the sample and the true value. The study uses root mean square error in Eq. (8).

$$E_{RMS} = \sqrt{\frac{E(W^*)}{N}} \tag{8}$$

Convolutional Neural Networks (CNN) use convolutional operations as the basic computing unit. At the output end of CNN, the features are linked from the fully connected layer to the global and detailed features, and the probability of each category is calculated using a step function. For color images, convolution and are independently and incoherent on the RGB channels. Eq. (9) is the calculation of two-dimensional convolution.

$$conv_{x,y} = \sum_i^{N \cdot N} \omega_i P_i \tag{9}$$

In Eq. (9),  $(x, y)$  represents the spatial coordinates on the digital image.  $N \cdot N$  represents the size of the convolutional kernel.  $\omega$  is the convolutional kernel weight.  $P$  represents the pixel value of the image. Pooling is a down sampling operation of CNN, which reduces the scale of the feature map and resolution to reduce parameters [21]. At

present, convolutional layers with a step size greater than or equal to 2 are commonly used to achieve downsampling of feature maps. CNN usually adds layer related biases to convolutional calculations and uses nonlinear activation functions to enhance the network's function expression ability. The Leaky ReLU activation function only has a small contribution rate to the overall population and can solve the large-scale neuronal death in ReLU. Eq. (12) is the Leaky ReLU activation function.

$$Leaky\ ReLU : f(x) = \begin{cases} x & x \geq 0 \\ ax & x < 0 \end{cases} \tag{10}$$

In Eq. (10),  $a$  is a decimal close to 0. Due to the excellent results achieved by CNN in image classification tasks, after structural modification, it is transferred to the target detection task [22]. YoloV3 has been optimized for targets of different scales, especially for small target recognition [23, 24]. YoloV3 preset 9 anchors and designed different scales for large, medium, and small targets, with 3 anchors for each scale. To improve the accuracy of target detection in the network, YoloV3 adopts a design similar to an image feature pyramid, which integrates global and local features to stretch into the same dimension. Fig. 2 shows multi-scale prediction.

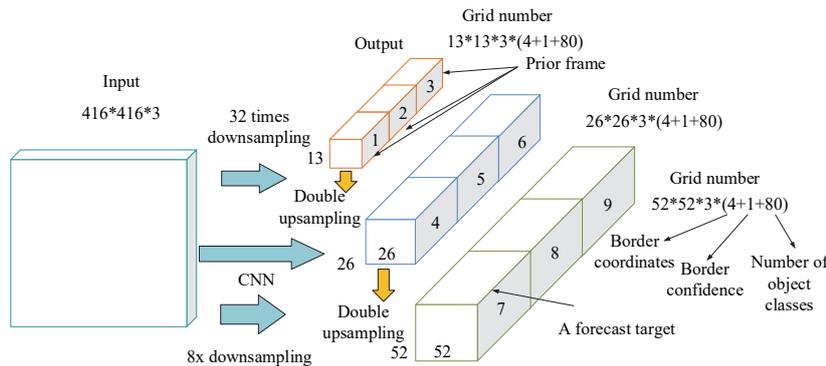


Figure 2 Framework of multi-scale prediction

In Fig. 2, to cope with identifying image targets of different sizes, Darknet-53 ultimately uses three different scales to identify targets of different scales. The scale of  $13 \times 13$  has richer global features for identifying large targets. The  $26 \times 26$  scale convolution results are also fused with the  $13 \times 13$  scale upsampling results for identifying medium-sized targets on the same scale. The scale of  $52 \times 52$  is also integrated with the upsampling results of  $26 \times 26$  on the convolution results of the same scale for identifying small-sized targets. YoloV3 divides each feature map into fixed size prediction units, with each prediction unit predicting three bounding boxes. YoloV3 replaces Softmax with Logistic in the loss function. Eq. (11) represents the total loss function.

$$loss = l_{box} + l_{ds} + l_{obj} \tag{11}$$

In Eq. (11),  $l_{box}$ ,  $l_{ds}$ ,  $l_{obj}$  represent the prediction loss of bounding box position, target confidence loss, and classification prediction loss, respectively.

### 3.2 Infrared Photo Defect Detection and Recognition Method Combining YoloV3 Transfer Learning and Gaussian Mixture Model

The Gaussian mixture model can use multiple Gaussian distributions to model the motion trend of each pixel in the image [25]. Due to noise interference in image sampling, pixel values at the same location will fluctuate up and down on the basis of stable pixel values. Therefore, this pixel value's distribution can be described by a single Gaussian distribution model centered on the pixel mean. However, a single model cannot accurately indicate the distribution changes, so a mixed Gaussian distribution model is proposed. Assuming image pixel's distribution is  $(x_0, y_0)$ , the changes within  $0 - t$  time can be modelled using a mixed model, and the current pixel's probability distribution is represented by Eq. (12).

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} \cdot \eta\left(X_t, \mu_{i,t}, \sum_{i,t}\right) \quad (12)$$

In Eq. (12),  $K$  represents the number of distributions, and its value is usually 3 - 5.  $\omega_{i,t}$ ,  $\mu_{i,t}$  represent the  $t$ -th distribution weight and mean at time  $i$ .  $\eta$  represents the distribution probability density function.  $\sum_{i,t}$  represents the covariance matrix. To further simplify the calculation and improve the algorithm running speed, assuming that three channels' pixel values of color image exist independently, the covariance matrix in Eq. (13) can be obtained.

$$\sum_{k,t} \sigma_k^2 I \quad (13)$$

In Eq. (13),  $k$  stands for arbitrary constant. Usually, the threshold can be set to 2.5 times the current mixed Gaussian distribution's standard deviation. If a mixed model composed of  $k$  single Gaussian distribution models cannot match the current distribution changes, the mixed Gaussian distribution model will be updated [26]. The current pixel value is used as the new distribution mean, the initial distribution variance is the new distribution variance, and the new distribution weight is the lowest weight of all the current distributions. This replaces the existing distribution with the lowest weight of the mixed Gaussian distribution, and updates it. Therefore, Eq. (14) is the new weight.

$$\omega_{k,t} = (1 - \alpha) \omega_{k,t-1} + \alpha (M_{k,t}) \quad (14)$$

In Eq. (13),  $\alpha$  represents the model's learning rate, which is used to control the update speed. For successfully

matched Gaussian distribution models,  $M_{k,t}$  is 1, and the rest is 0. For mismatched distribution models,  $\mu$ ,  $\sigma$  remain unchanged and the parameters of the complementary model for the matching results are updated. To better separate moving targets from the background, it should determine that this model can match this background better. So, it is necessary to prioritize all single Gaussian distribution models, and Eq. (15) is the calculation of priority  $rank$ .

$$rank = \omega / \sigma \quad (15)$$

For those single Gaussian distribution models that can better match the background, their weights are relatively high while ensuring a low variance, so the priority of this Gaussian distribution model is higher. In Eq. (16), the first  $B$  Gaussian distribution is used as the background matching model.

$$B = \arg \min_b \left( \sum_{k=1}^b \omega_k > T_B \right) \quad (16)$$

In Eq. (16),  $T_B$  represents the ratio of the background to entire image. The study uses the original FCN to calibrate the II temperature, but it encounters problems such as slow network training speed and difficulty in converging the loss curve [27]. To this end, a batch normalization layer (BN) is added after the convolutional layer of VGG-19. This layer can accelerate network's convergence speed by normalizing the previous layer's output, solve the gradient vanishing in backpropagation, and enhance the network's generalization ability. Fig. 3 shows an improved FCN.

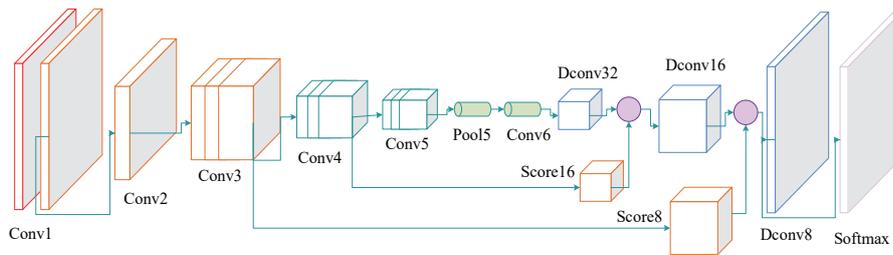


Figure 3 Improved FCN network structure

In Fig. 3, the improved FCN adjusts the network input and output size based on the dataset image size. After adding a BN layer to the convolutional layer, the ReLU activation function is replaced with Leaky ReLU. And two fully connected layers are removed from the full convolution section, and the output of Pool5 layer is

convolutionally processed to obtain  $8 \times 10 \times \text{Class}$ . In addition, the proposed singularity elimination algorithm smoothes the regions where pixel value mutations occur in the original II, further reducing calibration errors and improving prediction accuracy. Fig. 4 shows the overall framework of this algorithm.

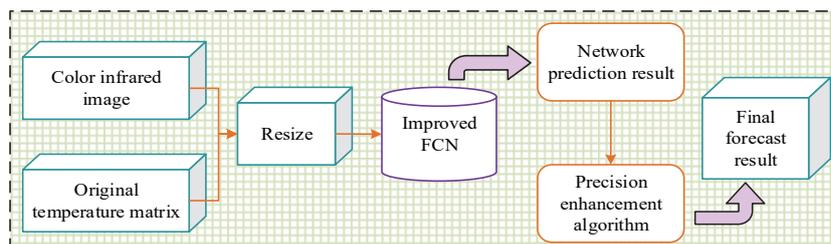


Figure 4 Overall algorithm framework

Due to the relatively small number of images containing defects, it is not entirely possible to rely on neural network training to identify defects when the samples are insufficient. Therefore, the research on defect detection of heating UAV II includes two steps. One is to annotate the position of the IIUAV, train and predict the UAV area using the YoloV3 network, and preliminarily identify the defects. Adjust the hyperparameters repeatedly according to the performance of the model until the optimal combination is found. Sometimes, the initially set hyperparameter range may need to be adjusted based on the experimental results. The second is to extract the manually

confirmed health images and compare them with the tested images in the same coordinate to detect defects. EfficientNet, a classification model developed by Google in 2019, uses a coordinated approach of width, depth, and resolution to enhance feature extraction and improve accuracy. Although improving a single dimension such as network depth or image sharpness can increase model performance, accuracy at 80 is difficult to improve. Therefore, we choose to combine three methods to achieve a breakthrough in precision. Fig. 5 shows the EfficientDet structure.

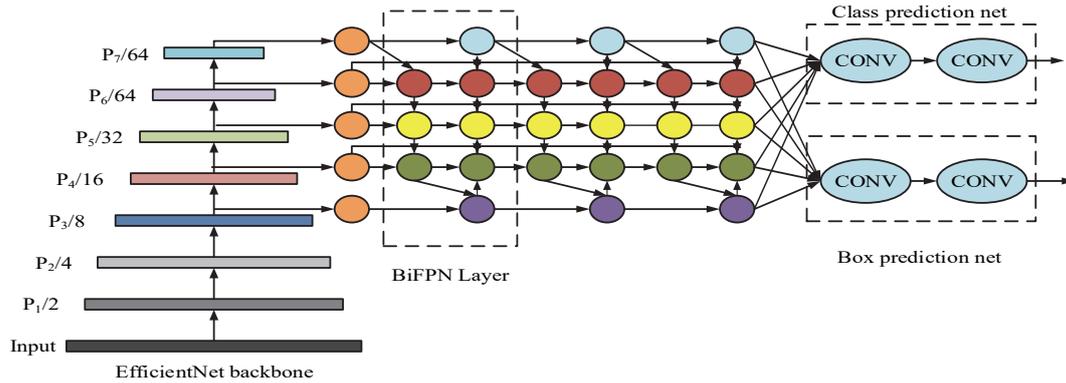


Figure 5 EfficientDet structure diagram

EfficientDet is an object detection network built on EfficientNet, which uses EfficientNet as its backbone to extract image features. EfficientDet introduced a weighted bidirectional feature pyramid network (BiFPN) that allows fast and efficient multi-scale feature fusion, so that more accurate target detection can be achieved by fusing feature maps of different resolutions. In addition, EfficientDet also uses composite scaling to jointly adjust the network width, depth and resolution, as well as the size of BiFPN, to achieve efficient model scaling. In this study, a series of mixed coefficients are used to control the magnification of the three dimensions, and the specific calculation process is shown in Eq. (17).

$$\begin{cases} d = \alpha^\phi \\ f = \beta^\phi \\ r = \gamma^\phi \\ s.t. \alpha \cdot \beta^\phi \cdot \gamma^\phi \approx 2; \alpha \geq 1, \beta \geq 1, \gamma \geq 1, \phi = 1 \end{cases} \quad (17)$$

In Eq. (17),  $d, r, f$  represents network width, network depth and image resolution respectively.  $\gamma, \beta$  indicates learnable parameters. The study borrowed from MnasNet and replaced the YoloV3 backbone Darknet-53 with the EfficientNet series, retaining the output of YoloV3 at three scales of  $13 \times 13, 26 \times 26,$  and  $52 \times 52,$  respectively. Fig. 6 shows the steps of II defect monitoring.

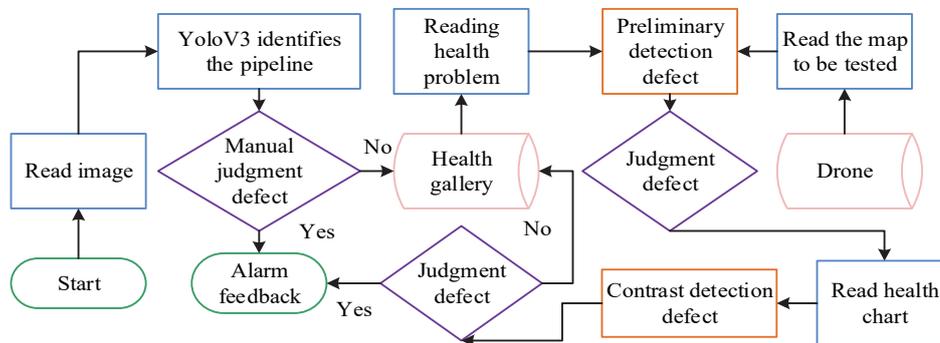


Figure 6 Steps of infrared image defect monitoring

In Fig. 6, in defect detection, YoloV3 is first used to predict the UAV region, and the image and data are saved in pairs. Test images are input and subjected to grayscale, Gaussian smoothing, dynamic threshold binarization, morphological denoising. Based on edge features, area shape features, and aspect ratio features, possible defect

areas are selected and matched with known UAV areas to provide preliminary detection results. If there are problems with the initial detection, then deep detection should be performed, otherwise it is a healthy image.

## 4 RESULTS AND DISCUSSION

In this experiment, the study conducted model training on an Intel i7 server equipped with an Intel Core i7-7700K processor and 32 GB of memory. In addition, the study used Nvidia GTX 1080Ti (11 GB) as the GPU.

### 4.1 Experimental Analysis of II Defect Detection based on YoloV3

#### 4.1.1 Performance Comparison Experiment Results of the Improved YoloV3 Algorithm

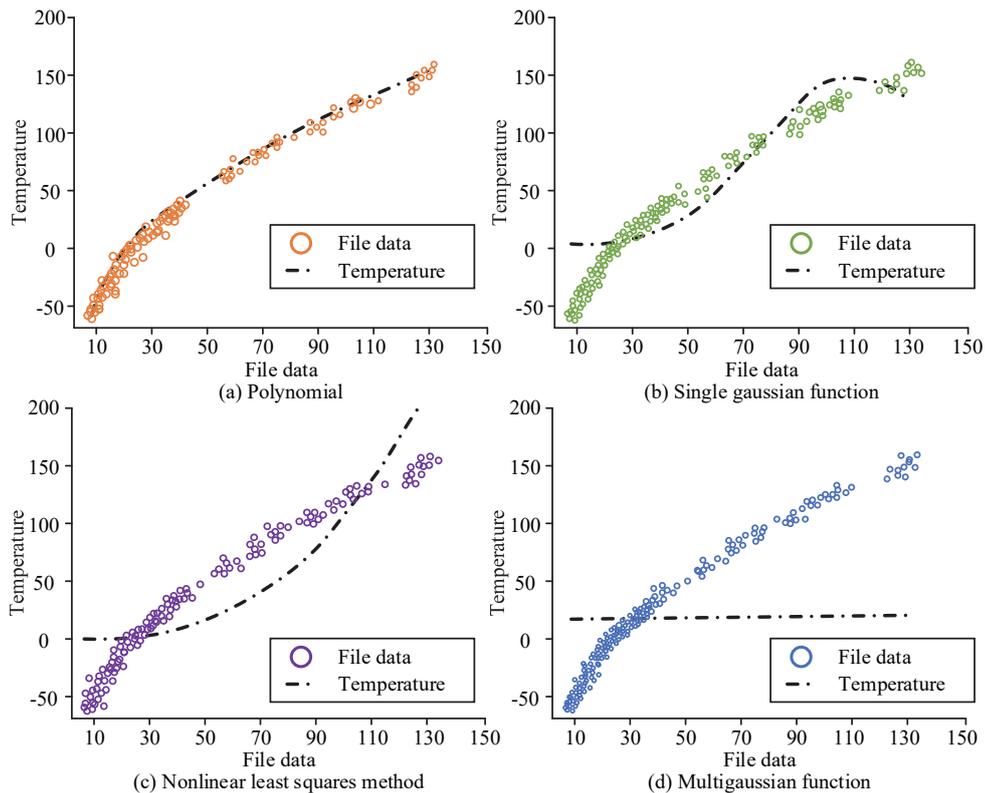
In the experiment, Zen XT2 dual-optical thermal imaging camera carried by DJI drone was used to capture infrared images of heating pipes. The camera, developed by DJI in collaboration with FLIR, can operate in high-gain mode in the range of  $-25^{\circ}$  to  $135^{\circ}\text{C}$ . A total of 1960 infrared images were collected at different times and locations to form a dataset. The data was divided into 1600 training sets, 200 validation sets, and 160 test sets. After careful data screening, 824 high-quality thermal images of the pipeline were retained. The images were all recorded in different places and times using the same thermal imaging lens and camera system, with a resolution of  $640 \times 512$  pixels. At the same time, three main defects of the pipeline

were identified, including road\_defects, floor\_defects and underground\_defects. Using labelme annotation tool, these defects are accurately annotated, and the corresponding XML annotation documents are generated. After data annotation, the number of images containing defects was 452.

**Table 1** Parameters of infrared imager and model training parameters

Infrared imager parameters		Model training parameter	
External optical transmission	1.00	Number of images/batch	20
Relative humidity	30.0%	Learning rate	$1e-6$
distance	0.0 m	Maximum number of iterations	40000
Atmospheric temperature	$20.0^{\circ}\text{C}$	Gradient descent algorithm	Adam Optimizer
External optical temperature	$20.0^{\circ}\text{C}$	Neuron retention rate	0.85
emittance	1.00	Batch Size	32 (First 100 rounds)
Reflection temperature	$20.0^{\circ}\text{C}$		8 (last 200 rounds)

This experiment compared the fitting effects of different methods in curve fitting. Fig. 7 showed several common curve fitting methods.



**Figure 7** Fitting results of various methods

According to Fig. 7, in the fitting of the relationship between the two, polynomials performed the best, while single Gaussian functions and nonlinear least squares methods had relatively poor fitting effects, while multi-Gaussian functions had the worst fitting effects. Therefore, the fitting effect of polynomials proved to be the best. After

repeated training, the optimal model parameters were selected for prediction on the test set, and the recognition results were shown in Fig. 8. Fig. 8 showed the accuracy, recall, and mAP of the original and improved YoloV3 algorithm models after training.

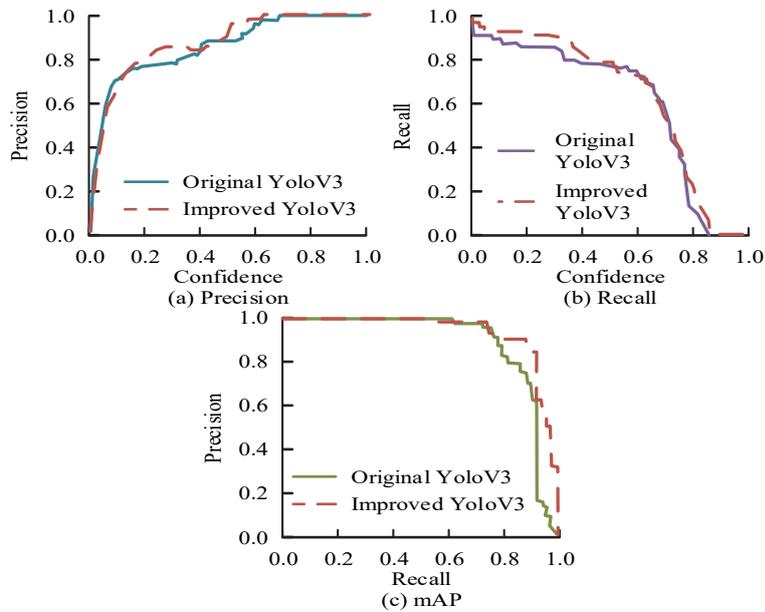


Figure 8 Accuracy, recall rate and mAP results of YoloV3 training before and after improvement

According to Fig. 8, the improved YoloV3 model had good convergence performance and had a certain improvement in the speed of object detection. To further prove the superiority of the research model, 1000 infrared image test sets were randomly selected to compare the proposed algorithm with other network models, and the results are shown in Tab. 2.

Table 2 Performance comparison results of each model

Index	Faster RCNN	YoloV3	YoloV4	EfficientNet -YoloV3(B1)
Floor defects	1.00	0.77	0.61	0.13
Road defects	0.85	0.69	0.58	0.15
Underground defects	0.96	0.71	0.82	0.33
Floor defects	0.00	0.30	0.49	0.92
Road defects	0.22	0.48	0.61	0.90
Underground defects	0.08	0.56	0.38	0.84
mAP / %	9.99	44.92	49.03	91.32

From Tab. 2, Log-average miss rate measures the average false detection rate of defects. A high value means a low detection accuracy. mAP is the average accuracy, and a high value means a high detection accuracy. The data showed that the EfficientNet-YoloV3 of the study design outperformed other networks on AP and mAP of all categories, indicating that it outperformed DarkNet53 and ResNet50 in feature extraction of pipeline infrared images. To further test the detection performance of the improved YOLOv3 algorithm trained model, the mAP of Fast-RCNN, Faster-RCNN, SSD, YOLOv3 and other algorithms were compared in Fig. 9.

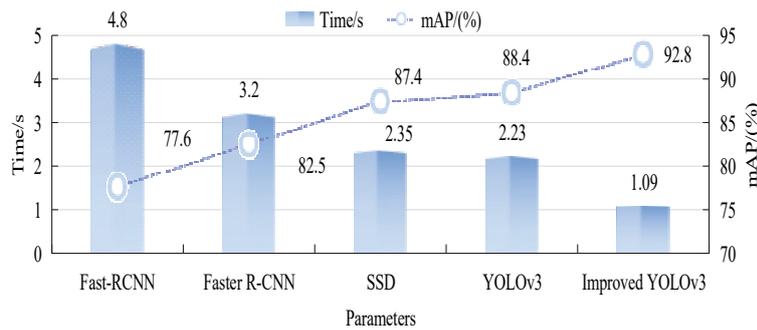


Figure 9 Comparison of training model data

According to Fig. 9, the mAP of the improved YOLOv3 algorithm reached 92.8%, which was improved compared to other algorithms. In terms of detection speed, the improved YOLOv3 only took 1.09s, which was less than other algorithms. In summary, the improved structure could effectively improve the detection accuracy. BN layer and Leaky ReLU activation function were further used in this study.

#### 4.1.2 Temperature Calibration Experiment Results Based on Full Convolutional Neural Network

In terms of temperature label division, based on the temperature measurement range of the infrared imager and considering the actual temperature range of the image, a 1% ratio was attempted for division. Fig. 10 showed the error distribution under these different partitioning strategies.

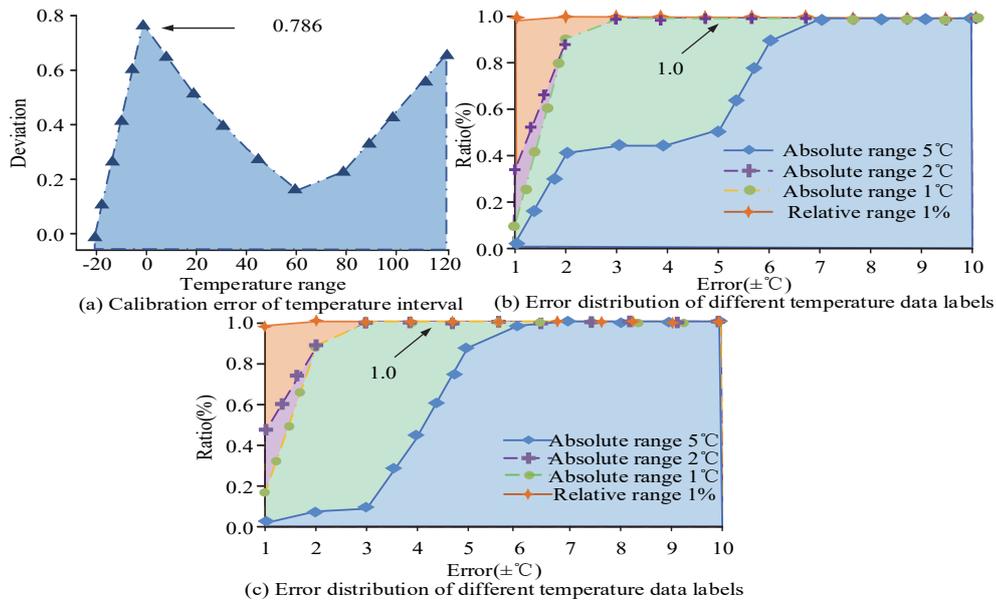


Figure 10 Temperature interval calibration error and different temperature data label error distribution

From Fig. 10a, the temperature calibration errors of the function were all within 1°. At the center of the temperature range, the error was relatively small, while at both ends of the range, the error increased relatively. According to Fig. 10b and Fig. 10c, as the temperature range was subdivided, the error was more distributed in smaller areas, which improved the accuracy of overall temperature calibration. According to the data in the figure, when the error was within the range of  $\pm 1^\circ\text{C}$ , the calibration using the 2 °C absolute temperature range was better than the 1 °C absolute temperature range. This indicated that more classification would increase model complexity. But due to the limitations of II, DL was difficult to learn more detailed features. In addition, using a 1% relative temperature interval division showed better results, with over 99% of data points having errors within 1 °C, which was superior to absolute temperature interval division. Therefore,

combining 1% relative temperature interval division with deep neural networks was an effective temperature calibration method. To further validate the excellent performance of the research method, 100 II samples were randomly selected as the test set, and two II temperature calibration methods proposed in this article were compared with other calibration methods. By analyzing the test set, it is found that there are large error points in the temperature prediction of some infrared images within 1% relative temperature range, which affects the overall prediction accuracy. The temperature calibration error distribution of some infrared images based on the FCN model, and the smoothing processing of the infrared image pixel distribution model by the sliding window median filtering algorithm using the singular-point elimination algorithm are shown in Fig. 11.

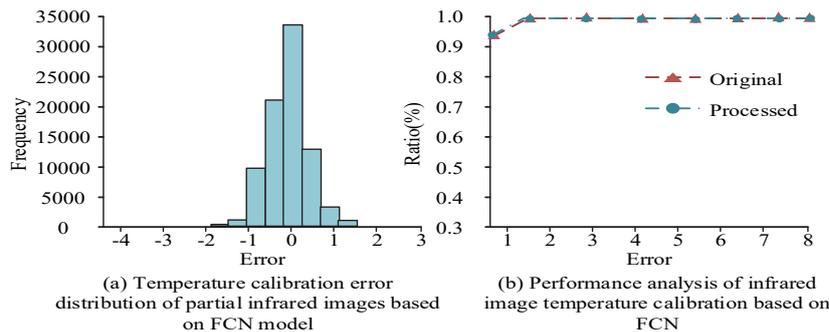


Figure 11 Anomaly distribution of singular point error at 1% relative temperature interval

From 11(a) that most of the errors are concentrated in the region of  $(-2, +2)$ , but there are a few pixel calibration errors are large, and some errors even reach  $-6^\circ\text{C}$ . It is precisely because of the existence of these singular points that the phenomenon of large deviation occurs in the process of model training. It can be seen from Fig. 11b that the temperature calibration accuracy of the processed infrared image is obviously improved, the error distribution is more concentrated, the overall error is reduced, the prediction error of a small part of the singularity is reduced, and the proportion of the prediction

error within  $\pm 2^\circ$  is increased by more than 1%. The sliding window median filtering algorithm can effectively improve the accuracy of pixel temperature prediction in FCN deep learning model infrared image.

#### 4.1.3 Comparison of Experimental Results of Defect Detection Indexes of Different Network Models

Fig. 12a showed the performance comparison results of various II temperature calibration methods. To conduct in-depth research, some representative II files were

selected for temperature calibration using research algorithms to verify their performance in traditional II file compression. Fig. 12b showed the experimental results.

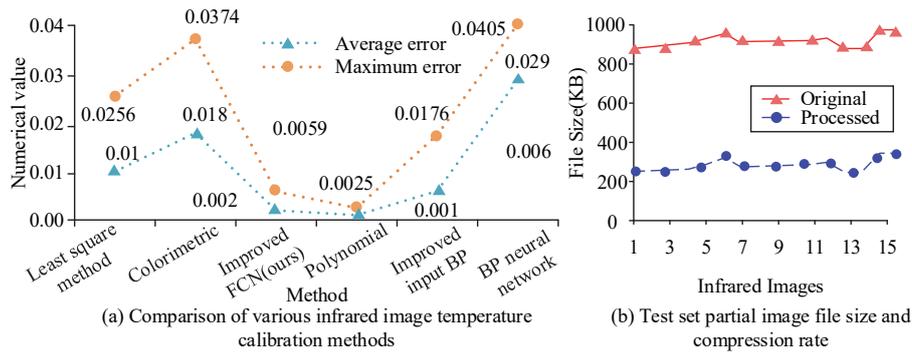


Figure 12 Performance of the algorithm in traditional infrared image file compression

According to Fig. 12a, the polynomial regression fitting method performed best in terms of maximum error and average error, followed by the improved FCN method, which had similar performance to the polynomial regression fitting method. Meanwhile, considering data requirements and temperature calibration intervals, the improved FCN method was suitable for various application scenarios. According to Fig. 12b, the FCN-based II temperature calibration algorithm could significantly reduce the II file size, with an average compression rate of 60.5%. This effectively saved the storage space of II, reduced the resources occupied by image storage, and improved the storage and transmission efficiency of II. 1200 images were selected for the experiment to construct a dataset. After multiple rounds of training, the optimal model parameters were selected and predicted on the test set. The study used Faster RCNN for model training, and Fig. 13 showed the comparison results of relevant indicators between two networks.

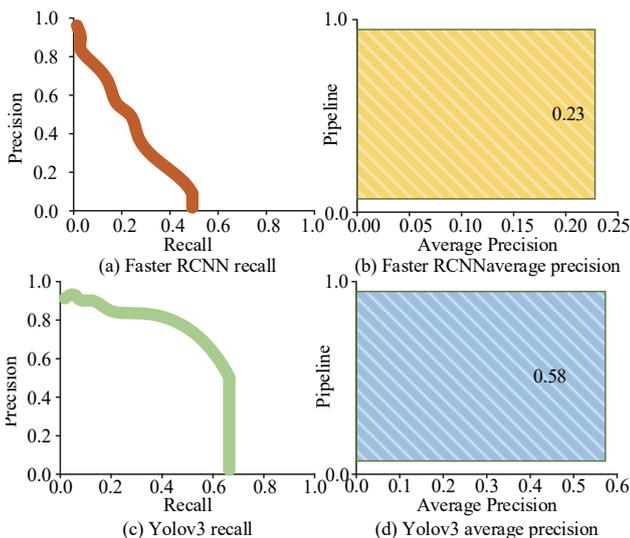


Figure 13 Comparison of relative indicators between Faster RCNN and YoloV3

According to Fig. 13, YoloV3 performed better in recognition accuracy and mAP, reaching 88.56% on the test set, and the recognition speed was faster. In contrast, Faster RCNN was a two-stage detection network. To show the performance effect of the research model comparing processing speed and compression rate in detail, Tab. 3

shows the ANOVA of compression rate and processing speed of the research model and other models.

Table 3 Statistical results of compression ratio and processing speed variance of the model

Item	Discrepancy	Sum of squares	Degree of freedom	Mean square	F	P
Compressibility	Interclass	0.004	2	0.002	74.2	0.013
	Intra-class	0.002	2	0.001	/	/
	Total	0.001	2	/	/	/
Processing speed	Interclass	12.042	4	3.011	2	0.096
	Intra-class	293.5	195	1.505	/	/
	Total	305.59	199	/	/	/

From Tab. 3, the variances of compression rate and processing speed between the research model and other models are 74.2 and 2, respectively, and the P-values are all less than 0.05, indicating that the differences are statistically significant. To further verify the proposed algorithm's excellent performance, this study randomly selected 200 UAV II samples as a test set and conducted detection and statistical analysis. Tab. 4 showed the performance results of the proposed algorithm compared to the indicators of various defect detection methods.

According to Tab. 4, the algorithm had a deep detection accuracy of up to 90% and a preliminary detection time of only 11ms. Compared with other detection methods, its false detection rate was the lowest, only 2.6%, and it took 49 ms, which was 63 ms and 27 ms less than the wavelet texture feature statistical analysis method and spectral residual visual significance method. Taking into account accuracy, error rate, and time required, the proposed method had high defect detection.

Table 4 compares the performance results of the proposed algorithm with those of various defect detection methods

Method		Accuracy rate / %	Time / ms	False detection rate / %
Textual method	Depth detection	90	18	2.6
	Preliminary detection	80	11	2.6
	Pipeline identification	85	20	2.6
Morphology		62	27	12.8
Statistical analysis of small ripple characteristics		95.4	112	6.4
Spectral residual visual significance method		95.3	76	4.7

## 4.2 Experimental Analysis of II Defect Detection Based on Gaussian Mixture Transfer Model

To compare the difference in defect image segmentation performance between the original GMM and the optimized accelerated GMM, a detailed comparison was conducted using the same test set. Due to the advantage of modelling each pixel, GMM slightly outperformed the proposed algorithm in defect

segmentation pixel accuracy. However, in terms of processing speed, this algorithm was significantly superior to GMM. After optimizing and accelerating the segmentation of defect areas in GMM, there were still internal interference noise points and regions that needed to be further removed using morphological and other methods. Fig. 12 showed the comparison of segmentation effects before and after image processing.

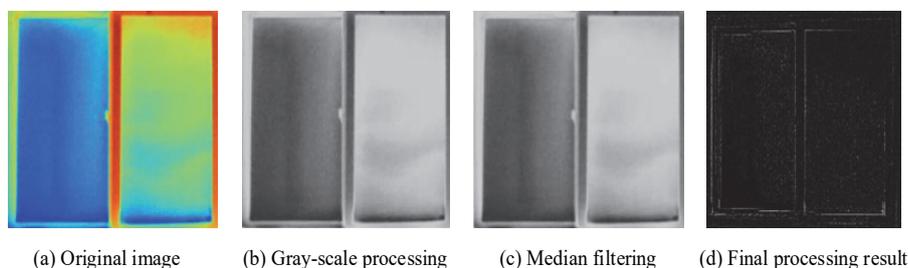


Figure 12 Comparison of segmentation effect before and after image processing

In Fig. 12, after filtering and morphological processing, the defect area became more complete, and noise was basically eliminated, which helped to further analyze the defect area.

## 4.3 Discussion

A method based on FCN for predicting the temperature of UAV II was proposed to address the issues of large storage space and high transmission cost in point by point temperature calibration. These experiments confirmed that the FCN-based II temperature estimation method had a calibration error of up to 99% for II temperatures that are less than  $2^{\circ}\text{C}$  of soil, and effectively saved 60.5% of II storage space. A multi-scale network adaptive mechanism based on YoloV3 had been proposed to address the high signal-to-noise ratio and low detail features of UAV II, achieving deep feature recognition for various sizes of II. And by fusing global and local features into the same dimension, high-precision detection of II defects was achieved. The experiment proved that the research algorithm achieved 88.56% mAP on the test set, which was superior to various classic object detection networks. In addition, the study further improved the accuracy of defect area segmentation by filtering and denoising the processing results of the Gaussian mixture model.

## 5 CONCLUSION

To improve the efficiency and accuracy of this technology, this study proposed a defect detection and recognition method for UAV infrared photos based on DL and transfer learning. It utilized automated image processing and machine learning technology to achieve automatic detection and recognition of defects in UAV infrared photos. The experimental results show that the infrared image temperature calibration algorithm based on FCN can effectively reduce the file size and achieve an average compression rate of 60.5%. The EfficientNet-YoloV3 model, designed in the study, outperformed DarkNet53 and ResNet50 on AP (average accuracy) and mAP, demonstrating superior performance in UAV

infrared image analysis. The temperature calibration accuracy of infrared image is significantly improved, the error is more concentrated, the overall reduction, and the proportion of prediction error within the range of  $\pm 2^{\circ}$  is increased by more than 1%. In addition, the sliding window median filtering algorithm is used to further improve the accuracy of FCN model in pixel temperature prediction. Compared with other defect detection methods, the algorithm proposed in this study performs well in deep detection, with an accuracy of up to 90%, initial detection only 11 ms, false detection rate of 2.6%, and processing time of only 49 ms. This is 63 ms shorter than small ripple feature statistical analysis and 27 ms shorter than spectral residual visual significance method. Considering accuracy, error rate and processing time, the research method shows obvious advantages in defect detection. Although the research has achieved good results, there is still room for improvement. For example, the limited diversity of defect types may affect the model generalization ability, so it is necessary to expand the dataset of defect types. Secondly, the lack of actual hardware testing may limit the performance of the algorithm in practical applications, so it is important to conduct actual hardware testing to verify the robustness and practicality of the model. These problems are important areas for further research.

## 6 REFERENCES

- [1] Ali, M. U., Saleem, S., Masood, H., Kallu, K. D., Masud, M., Alvi, M. J., & Zafar, A. (2022). Early hotspot detection in photovoltaic modules using color image descriptors: An infrared thermography study. *International Journal of Energy Research*, 46(2), 774-785. <https://doi.org/10.1002/er.7201>
- [2] Chen, Y., Li, Z., Li, L., Ma, S., Zhang, F., & Fan, C. (2022). An anti-drone device based on capture technology. *Bionic Intelligence and Robotics*, 2(3), 43-49. <https://doi.org/10.1016/j.birob.2022.100060>
- [3] Li, X. (2022). Infrared image filtering and enhancement processing method based upon image processing technology. *Journal of Electronic Imaging*, 31(5), 1-17. <https://doi.org/10.1117/1.JEI.31.5.051408>
- [4] Wang, K., Yuan, S., Yao, Z., Gao, J., & Feng, J. (2022). Design and Implementation of Infrared Image Classification

- Algorithm for Defective Power Equipment Based on Deep Learning. *Nonlinear Optics, Quantum Optics*, 56(1/2), 83-95.
- [5] Yang, X., Yin, C., Dadras, S., Lei, G., Tan, X., & Qiu, G. (2022). Spacecraft damage infrared detection algorithm for hypervelocity impact based on double-layer multi-target segmentation. *Frontiers in Information and Electronic Engineering: English*, 23(4), 571-586. <https://doi.org/10.1631/FITEE.2000695>
- [6] Kenry, Duan, Y. K., & Liu, B. (2018). Recent Advances of Optical Imaging in the Second Near-Infrared Window. *Advanced Materials*, 30(47), 1802394. <https://doi.org/10.1002/adma.201802394>
- [7] Ouyang, Q., Wang, L., Park, B., Kang, R., & Chen, Q. (2021). Simultaneous quantification of chemical constituents in matcha with visible-near infrared hyperspectral imaging technology. *Food Chemistry*, 350(6), 1-8. <https://doi.org/10.1016/j.foodchem.2021.129141>
- [8] Peng, X., Liang, F., Qian, J., Yang, B., & Zheng, X. (2019). Automatic Recognition of Insulator from UAV Infrared Image Based on Periodic Textural Feature. *Gaodianya Jishu/High Voltage Engineering*, 45(3), 922-928.
- [9] Tan, X. & Zhang, G. (2022). Research on Surface Defect Detection Technology of Wind Turbine Blade Based on UAV Image. *Chinese Journal of Instrument: English*, 9(1), 41-48.
- [10] Lee, H. & Kim, W. (2020). Quantitative Analysis on Phase Contrast of Defect Detection Mechanism by Wavelet Infrared Thermography. *Journal of the Korean Society for Nondestructive Testing*, 40(2), 85-90. <https://doi.org/10.7779/jksnt.2020.40.2.85>
- [11] Wu, J. & Liao, S. (2022). Intelligent Detection of Dangerous Goods in Security Inspection Based on Cascade Cross Stage YOLOv3 Model. *Tehnički vjesnik*, 29(3), 888-895. <https://doi.org/10.17559/TV-20220222034206>
- [12] Ranjit, S. & Kim, W. T. (2021). Quantitative Evaluation for Early Defect Detection of Contaminated Ball Bearing by Temperature Mapping in Infrared Thermography. *International Journal of Applied Engineering Research*, 9(21), 9401-9409.
- [13] Subramanian, M., Shanmugavadivel, K., & Nandhini, P. S. (2022). On fine-tuning deep learning models using transfer learning and hyper-parameters optimization for disease identification in maize leaves. *Neural Computing and Applications*, 34(16), 13951-13968. <https://doi.org/10.1007/s00521-022-07246-w>
- [14] Wang, Y., Liu, X., Bin, L. I., & Chang, Q. (2021). The manifold embedded selective pseudo-labeling algorithm and transfer learning of small sample dataset. *Journal of Northwestern Polytechnical University*, 39(5), 1122-1129. <https://doi.org/10.1051/jnwpu/20213951122>
- [15] Jiang, Z. L., Chen, Y. N., Wang, K., Yang, B. W., & Song, G. H. (2023). A Graph-Based PPO Approach in Multi-UAV Navigation for Communication Coverage. *International Journal of Computers Communications & Control*, 18(6), 1-14. <https://doi.org/10.15837/ijccc.2023.6.5505>
- [16] Wei, W. U., Jing, X., Wencai, D. U., & Chen, G. (2021). Learning dynamics of gradient descent optimization in deep neural networks. *Science in China: Information Science (English)*, 64(5), 13-27. <https://doi.org/10.1007/s11432-020-3163-0>
- [17] Chen, C., Wang, Y., Gao, Z. T., Peng, F. Y., Tang, X. W., Yan, R., & Zhang, Y. K. (2022). Intelligent learning model-based skill learning and strategy optimization in robot grinding and polishing. *Science in China: Technical Sciences English*, 65(9), 1957-1974. <https://doi.org/10.1007/s11431-022-2112-4>
- [18] Brashler, K. & Alshahrani, Y. (2021). Use infrared thermal imaging for pump system condition monitoring and early failure detection. *Hydrocarbon Processing*, 100(5), 57-59.
- [19] Žiljak, J., Jurečić, D., & Žiljak, V. (2018). Packaging Design with Hidden Near Infrared Colour Separation. *Tehnički vjesnik*, 25(Supplement 1), 211-215. <https://doi.org/10.17559/TV-20170705114921>
- [20] Xu, Z. L. (2023). The Origin of Cosmic Microwave Background Radiation. *Modern Physics (English)*, 14(4), 534-551. <https://doi.org/10.4236/jmp.2023.144030>
- [21] Adams, S., Bucknall, T., & Kouzani, A. (2021). An initial study on the agreement of body temperatures measured by infrared cameras and oral thermometry. *Scientific Reports*, 11(1), 1-10. <https://doi.org/10.1038/s41598-021-91361-6>
- [22] Wang, P., Song, L., Wang, H., Han, C., Guo, X., & Cui, L. (2022). Grouping sparse filtering: a novel down-sampling approach toward rotating machine intelligent diagnosis in 1D-convolutional neural networks. *Measurement Science & Technology*, 33(6), 1-10. <https://doi.org/10.1088/1361-6501/ac4ce6>
- [23] Wu, B., Zhang, T., & Mao, L. (2023). Large-scale Image Classification with Multi-perspective Deep Transfer Learning. *Computer Science and Information Systems*, 20(2), 743-763. <https://doi.org/10.2298/CSIS220714015W>
- [24] Xiao, J. (2021). exYOLO: A small object detector based on YOLOv3 Object Detector. *Procedia Computer Science*, 188(10), 18-25. <https://doi.org/10.1016/j.procs.2021.05.048>
- [25] Ma, J., Chen, J., & Yang, C. (2023). Using optimized gaussian mixture model rules and global tracking graph for feature extraction and tracking in time-varying data. *The Visual Computer*, 39(5), 1869-1892. <https://doi.org/10.1007/s00371-022-02451-z>
- [26] Maina, C. B., Weke, P. G. O., Ogotu, C. A., & Ottieno, J. A. M. (2022). A Normal Weighted Inverse Gaussian Distribution for Skewed and Heavy-Tailed Data. *Applied Mathematics (English)*, 13(2), 163-177. <https://doi.org/10.4236/am.2022.132013>
- [27] Chen, Z. (2022). Research on internet security situation awareness prediction technology based on improved RBF neural network algorithm. *Journal of Computational and Cognitive Engineering*, 1(3), 103-108. <https://doi.org/10.47852/bonview/JCCE149145205514>

**Contact information:****Rong MENG**

(Corresponding author)

State Grid Hebei Extra High Voltage Company,  
Hebei, Shijiazhuang, 050070, China  
E-mail: mr19830311@126.com**Zhilong ZHAO**State Grid Hebei Extra High Voltage Company,  
Hebei, Shijiazhuang, 050070, China  
E-mail: 401448030@qq.com**Mengtian HE**State Grid Hebei Extra High Voltage Company  
Hebei, Shijiazhuang, 050070, China  
E-mail: Lsacr1107@hotmail.com**Zhiyuan WANG**State Grid Hebei Extra High Voltage Company,  
Hebei, Shijiazhuang, 050070, China  
E-mail: 995025584@qq.com**Yanbo DUAN**State Grid Hebei Extra High Voltage Company,  
Hebei, Shijiazhuang, 050070, China  
E-mail: 610049855@qq.com