ARTICLE

Infrared Fault Detection Method for Dense Electrolytic Bath Polar Plate Based on YOLOv5s

Huiling Yu[1], Yanqiu Hang[2], Shen Shi[1], Kangning Wu[1] and Yizhuo Zhang[1,*]

[1]Software Engineering, Department of Computer Science, Changzhou University, Changzhou, 213146, China

[2]Electrical Engineering, Department of Computer Science, Changzhou University, Changzhou, 213146, China

*Corresponding Author: Yizhuo Zhang. Email: yzzhang@cczu.edu.cn

ABSTRACT

Electrolysis tanks are used to smelt metals based on electrochemical principles, and the short-circuiting of the pole plates in the tanks in the production process will lead to high temperatures, thus affecting normal production. Aiming at the problems of time-consuming and poor accuracy of existing infrared methods for high-temperature detection of dense pole plates in electrolysis tanks, an infrared dense pole plate anomalous target detection network YOLOv5-RMF based on You Only Look Once version 5 (YOLOv5) is proposed. Firstly, we modified the Real-Time Enhanced Super-Resolution Generative Adversarial Network (Real-ESRGAN) by changing the U-shaped network (U-Net) to Attention U-Net, to preprocess the images; secondly, we propose a new Focus module that introduces the Marr operator, which can provide more boundary information for the network; again, because Complete Intersection over Union (CIOU) cannot accommodate target borders that are increasing and decreasing, replace CIOU with Extended Intersection over Union (EIOU), while the loss function is changed to Focal and Efficient IOU (Focal-EIOU) due to the different difficulty of sample detection. On the homemade dataset, the precision of our method is 94%, the recall is 70.8%, and the map@.5 is 83.6%, which is an improvement of 1.3% in precision, 9.7% in recall, and 7% in map@.5 over the original network. The algorithm can meet the needs of electrolysis tank pole plate abnormal temperature detection, which can lay a technical foundation for improving production efficiency and reducing production waste.

KEYWORDS

Infrared polar plate fault detection; YOLOv5; Real-ESRGAN; Marr boundary detection operator; Focal-EIoU loss

1 Introduction

Electrolytic cells are widely used in metal smelting and consist of a tank body, an anode, a cathode, and typically a diaphragm that separates the anode and cathode chambers. The operational principle involves applying direct current to induce oxidation reactions at the anode-solution interface and reduction reactions at the cathode-solution interface, thereby producing the desired products. Current efficiency in electrolysis directly influences both product yield and processing costs. The factors that affect current efficiency include losses caused by short circuits, open circuits, and leaks, with short circuits being the main factor. The site environment is shown in Fig. 1.
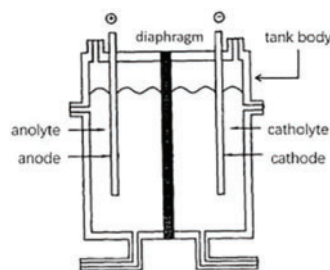
(a) a site plan of the electrolyzer   (b) a brief diagram of the electrolyzer structure

Figure 1: The site environment. (a) A site plan of the electrolyzer; (b) A brief diagram of the electrolyzer structure, which shows only one diaphragm

For a long time, most of the temperature measurement methods used in the workshop were measured by sprinkling water or a handheld thermometer, which requires manual individual checking of each pole plate one by one, so the short-circuit troubleshooting time is longer and the temperature change of the pole plate is not obvious in the early stage, which may lead to missing the best time for troubleshooting. When the next round of troubleshooting is carried out, multi-point short-circuits will occur. If the short-circuit time is more than 4 h, the precipitated crystals are thicker, and a single person cannot lift the plate to deal with them, and the deep short-circuits cannot be effectively dealt with, which will lead to an increase in short-circuits in the late stage of electrolysis. Therefore, developing a rapid and accurate high-temperature target detection algorithm for electrolysis tanks holds significant importance. Such advancements aim to reduce production costs, enhance production efficiency, and facilitate intelligent production processes.

Based on the properties of the pole plate, a short circuit results in unusually high temperatures, necessitating the use of infrared cameras for data acquisition. Currently infrared detection technology is widely used in many industries: aerial surveillance [1], vehicle identification [2], fire monitoring [3], etc. It can be mainly divided into the following two types of algorithms: infrared target detection algorithms based on traditional methods [4] and depth detection algorithms based on deep learning [5]. Due to the differences between the task and the data, the traditional methods will rely more on manual labor, which requires human observation of the data, calculation and experimentation, manual adjustment of the parameters, and the applicability [6] and extrapolation of the model method are relatively poor. The infrared target detection method based on deep learning fills this gap.

Neural networks employed for infrared target detection using deep learning can be broadly classified into two categories: one-stage and two-stage networks. The two-stage algorithm is represented by Faster R-CNN, Sparse R-CNN [7], etc., however, most of these algorithms exhibit a high number of parameters. In contrast, one-stage networks better align with actual industrial production requirements, with the You Only Look Once (YOLO) network being a representative example. Redmon et al. [8] proposed the YOLOv1 network, which directly obtains the specific location information and category classification information of target detection through regression. This approach significantly improves the detection speed. However, it is difficult to predict aspect ratio objects that have not been seen in training data. YOLOv2 [9] improves YOLOv1 from three perspectives: more accurate, faster, and more recognition, adds a regularization method of batch normalization, replaces the classification network with DarkNet19, and the overall convolution operation is more efficient than YOLOv1 few. Compared with YOLOv2, YOLOv3 [10] introduces

the residual module, deepens the network depth, and proposes DarkNet53. At the same time, it draws on the feature pyramid idea to perform predictions at three different scales. Wang et al. [11] proposed YOLOv4, which integrates numerous optimization strategies within its architecture. This includes Mosaic data augmentation for input data, the backbone integration Cross Stage Partial (CSP) [12] structure into DarkNet53, and the introduction of the Mish activation function to enhance detection accuracy. YOLOv5 boasts reduced size, lower deployment costs, greater flexibility, and faster detection speeds compared to earlier iterations (YOLOv1 to YOLOv4). It employs Spatial Pyramid Pooling with Feature Pyramid (SPPF) instead of Spatial Pyramid Pooling (SPP) [13] and utilizes an adaptive approach to anchor box generation during training. While YOLOv6 [14] and YOLOv7 [15] algorithms offer increased accuracy, they also exhibit significantly higher parameter counts compared to YOLOv5, sometimes by several orders of magnitude. Therefore, YOLOv5 is chosen as the base network for this experiment.

For infrared target detection, the long-term problems include small targets, unclear grayscale changes, and low image resolution. In the collected dataset, the arrangement of targets is very dense, which exacerbates the problem of unclear grayscale changes at the boundaries of infrared images, and the target aspect ratio is large, making it difficult for the YOLO network to directly adapt. Based on the above issues, the summary of the work in this article is as follows:

(1) Aiming at the problem of slow grey scale change of target edges in infrared pictures and relatively dense arrangement of targets, which makes it difficult to distinguish the target boundaries, the experiment combines Focus and Marr operators and replaces the first Convolutional Block Separable (CBS) in the net with this new module.

(2) Aiming at the problem that the size of the electrolytic tank bezel in the lens increases and decreases at the same time, and the CIOU is unable to adapt to it, and, the difficulty of detecting different temperature segments is different, but CIOU cannot distinguish difficult to detect targets, the experiment uses the Focal-EIOU loss to replace the CIOU.

(3) Aiming at the problem that the infrared camera is higher from the ground and captures smaller pictures with lower pixels, the infrared pictures are preprocessed after modifying the U-Net in Real-ESRGAN to Attention U-Net.

2 Related Work

2.1 Infrared Target Detection

In the complex environment of electrolysis tanks, where the design of the plates tends to be compact and less spaced from each other, this dense arrangement creates a smoother grey scale transition [16] of the infrared radiation in the edge region of the plates when using infrared thermography for fault detection. The ambiguity of this edge grey scale change poses a significant problem for computer vision algorithms that rely on image edge information to identify object boundaries. The above problems, combined with the inherently low-resolution nature of infrared images, further exacerbate the risk of missed detections. Infrared images, due to the limitations of their imaging principle, tend to have a low pixel density, i.e., each pixel represents a relatively large actual spatial dimension, which results in a relatively small size of the target (e.g., polar plate boundary) presented in the image. With this dual challenge, subtle variations in polar plate boundaries may be almost submerged in background noise or similar grey values of neighboring pixels in infrared images, making the already difficult-to-capture edge information even more difficult to identify. Jiao et al. [17] proposed an effective weld bevel edge recognition technique tailored to industrial scenarios. This technique incorporates median

filtering, Sobel operator edge detection, and a defined narrow region edge search, aimed at reducing guidance errors and ensuring system accuracy. Li et al. [18] introduced a boundary enhancement operator to adaptively extract boundary features from the limit points of the boundary to enhance existing features. Based on the characteristics of the target instance and the improved network model, the complete intersection-union ratio function was introduced to deal with the YOLOv3 loss. Notably, although the proposed method yields pronounced enhancement outcomes on low-resolution and high-noise remote sensing images, however, the images processed by the boundary operator are black and white. If the network has tasks other than positioning, it also needs to combine color information to determine abnormalities. To address this, the experiment suggests utilizing a boundary detection operator to extract boundary features. Subsequently, these features are superimposed onto the color feature maps and transmitted to the back layer of the network for comprehensive analysis.

Given the problem that the electrolytic cell is rectangular and has a very large aspect ratio, and the currently used loss function CIoU cannot adapt, Hamid et al. [19] proposed Generalized IoU (GIoU). This proposition adds an item after IoU to calculate the minimum circumscribed rectangle of two boxes, which is used to characterize the distance between the two boxes, thereby solving the problem of zero gradients when the two objectives do not intersect. However, when the prediction box lies within the ground truth box and their areas are equal, GIoU fails to discern the relative positional relationship between them. To overcome this limitation, Zheng et al. [20] introduced Distance-IoU (DIoU), which incorporates center distance within the GIoU framework. In scenarios where the prediction box is contained within the ground truth box, both frames possess equal area and center distance, and the DIoU metric struggles to discern the aspect ratio relationship between them. To address this limitation, Zheng et al. [21] introduced CIoU, which incorporates an aspect ratio factor, however, the definition of the aspect ratio in CIoU can be ambiguous. In response to the problem of varying degrees of difficulty in testing samples, Lin et al. [22] proposed Focal loss to solve the category imbalance loss function, which is used in the image field to solve the problem of positive and negative sample imbalance and difficult classification sample learning in one-stage target detection. However, Focal loss uses basic IoU and is applied to actual problems, there will be situations where the value is 0 when the predicted box and the real box do not intersect.

2.2 Super-Resolution Preprocessing

In target detection using deep learning, especially Convolutional Neural Networks (CNNs), image information is passed through the convolutional layers layer by layer for feature extraction. However, as the number of network layers deepens, while higher-level abstract features can be captured, it inevitably leads to the loss of a large amount of shallow detail information. This shallow information often contains key features such as the edges and textures of the target, which are crucial for accurate positioning and identification of small targets. Therefore, the loss of shallow information can seriously affect the performance of target detection and increase the risk of missed detection and false alarms. To alleviate this problem, super-resolution pre-processing techniques have been introduced into the target detection process. Zhao et al. [23] first added the Convlutional Block Attention (CBAM) to Super-Resolution Generative Adversarial Networks (SRGAN) and then cascades the improved SRGAN with YOLOv5 for final detection. Xiao et al. [24] proposed a novel adaptive squeezing excitation module that adds a Reception Field Block (RFB) [25] module to the YOLO network. The module uses a super-resolution model to increase the expressive power of receptive fields and feature images. Most of the above methods focus on pre-processing using SRGAN, however, the enlarged details processed by SRGAN [26] are usually accompanied by artifacts, Enhanced Super Resolution

Generative Adversarial Networks (ESRGAN) [27] makes improvements on this basis, and the operations include removing the BN layer, and the basic structure is replaced with RDDB. For example, Wang et al. [28] proposed a network combining YOLOv5 and ESRGAN for the problem of too small labels in digital recognition by replacing the backbone network of YOLOv5 with MobileNetv3 [29] to improve the image quality and achieve digital recognition. However, in the real world, the problem of recovering degraded low-resolution images is still unsolved. Wang et al. proposed Real-ESRGAN [30], which replaces Visual Geometry Group (VGG) [31] with U-Net [32] to enhance the adversarial learning on the details of the pictures and introduces spectral normalization to stabilize the complex dataset and reduce the instability caused by the training, the above operations enable the network to produce more realistic and clear images. Yang et al. [33] utilized Real-ESRGAN technology to enhance image resolution, effectively addressing the challenge of identifying minor defects in the dataset. Furthermore, Deformable Convolutions (DCN) [34] was seamlessly integrated into YOLO, alongside the design of a module based on CBAM and Concurrent Spatial and Channel Squeeze and Excitation (SCSE) [35] attention mechanisms, which improved feature map representation. Zheng et al. [36] proposed a two-stage recognition method that integrates the YOLO algorithm with Real-ESRGAN. Firstly, YOLO utilizes GhostNet [37] as its backbone to reduce computational load. The head integrates the attention-guided module to enhance attention capabilities on small targets. Secondly, the flame region identified by YOLO undergoes cropping, stitching, and enhancement using Real-ESRGAN for improved clarity.

In summary, the article chooses YOLOv5 as the base network, preprocesses the image after modifying Real-ESRGAN, then modifies the loss function of YOLOv5, and finally replaces the first CBS with Focus and adds the Marr boundary detection operator to it. The network obtained by the above modification can be well adapted to the characteristics of the electrolysis tank and efficiently detect abnormal targets.

3 Methods

3.1 Model Introduction

YOLOv5 can achieve a good balance of accuracy and number of parameters, while YOLOv5s is the smallest in the series and is suitable for edge devices, so the experiments use YOLOv5s as the base network. To further improve the detection performance of YOLOv5s, the loss function and the first CBS of the network are modified, and the image is preprocessed with super-resolution. Its structure is shown in Fig. 2.

The backbone in the above figure is the backbone of YOLOv5s itself, the difference is that Real-ESRGAN is used to super-resolve the image, the loss function is replaced with Focus-EIOU Loss, and the first CBS is replaced with the Focus module, and the Marr operator is added to the Focus module. The above processing can help the network to obtain clearer shallow information, which helps to solve the problems of low pixels in infrared images, slow change of target edges, large target aspect ratio, and samples varying in difficulty of detection.

3.2 Focus Combined with Marr Boundary Detection Operator

In the context of convolutional neural networks for target feature extraction, as the network depth increases, shallow information diminishes progressively while abstract features become more pronounced. In tasks like target detection, especially in scenarios such as infrared imagery characterized by smooth grayscale transitions and subtle edges, retaining shallow details such as object edges and textures is critical for accurate localization and identification. However, deep convolutional layers
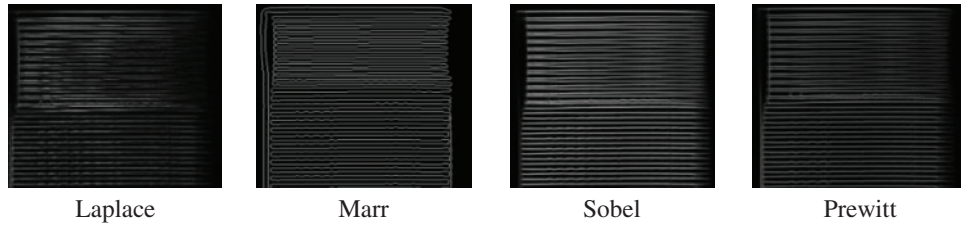
often struggle to preserve such crucial shallow information, potentially leading to missed detections and reduced recall rates. To address this issue and enhance the network's capacity to retain shallow details, we replace the initial CBS module with a Focus module. The Focus module utilizes a slicing approach, dividing the input image into four complementary sub-images. This operation reduces the image dimensions for subsequent processing while mitigating the loss of shallow information typical in standard convolution operations. Introducing the Focus module enriches the network's reservoir of shallow details, significantly improving the precision and recall of target detection, especially in challenging contexts like infrared imagery. This transformation is illustrated in Fig. 3.

Figure 2: YOLOv5s-RMF structural flowchart, with modifications made to sections "a. Modified Focus," "b. Focus-EIOU-Loss," and "c. Real-ESRGAN Image Preprocessing."

Figure 3: Focus schematic

However, the introduction of Focus was not enough, as it only retained the information and did not highlight important information; CBS picked out information from the images that were worth learning, but Focus did not focus on it, and the experimental results showed that the recall did not improve much. To mitigate this limitation and enhance recall effectiveness, we integrated a boundary detection operator within the Focus module to accentuate boundary information more prominently. The selection of the detection operator holds particular significance at this stage. Comparative analyses were conducted using various operators, as depicted in Fig. 4, to evaluate their effectiveness.

| Laplace | Marr | Sobel | Prewitt |

Figure 4: Processing effect of different boundary detection operators

It is clear to see that the Marr operator works better than the others. The overall structure of the network is shown in Fig. 2. The formula for the Marr boundary operator is shown in Eq. (1):

$$\nabla^2 G(x, y) = \left(\frac{x^2 + y^2 - 2\sigma^2}{\sigma^4} \right) e^{\frac{-(x^2 + y^2)}{2\sigma^2}} \tag{1}$$

where $\nabla^2 G$ is the filter, ∇^2 is the Laplace operator, and G is a two-dimensional Gaussian function with standard deviation σ. Algorithm 1 below demonstrates the detailed process:

Algorithm 1: Add the Marr boundary detection operator to the Focus module
---
Input: $3 \times 1280 \times 1280$ (C × H × W)
Output: $32 \times 640 \times 640$ (C × H × W)
1: x1 = x [..., ::2, ::2]; x2 = x [..., 1::2, ::2]; x3 = x [..., ::2, 1::2]; x4 = x [..., 1::2, 1::2];
2: x2_marr = marr(x2); x3_marr = marr(x3);
3: x2_final = x2 * x2_marr.expand_as(x2); x3_final = x3 * x3_marr.expand_as(x3)
4: Result = conv(torch.cat([x1, x2_final, x3_final, x4], 1))

The module integrating the Marr operator and Focus can make the network pay more attention to the boundary information while reducing the information loss and retaining more shallow information for the network, which is a good way to alleviate the problem of inconspicuous grey scale changes in the infrared picture itself.

3.3 Focal-EIoU Loss

CIoU is used in YOLOv5s, which uses penalties including overlap area, centroid distance, and aspect ratio. Notably, if the aspect of the detected target can increase or decrease at the same time, then the CIOU will be invalid, but the object of this experiment will increase and decrease at the same time the length and width of the places far away from the camera will become smaller together, and the length and width of the places close to the camera will become larger together. Consequently, the CIoU loss function may not precisely regress to the aspect ratio of the ground truth box. This leads to a decrease in model training performance, including poor performance in metrics such as precision and recall, as well as an increased risk of overfitting or underfitting. To address this limitation, the EIoU introduces an additional penalty term based on the edge length.

Additionally, detecting objects at different temperature ranges poses varying levels of difficulty. In infrared cameras, objects with higher temperatures are more pronounced. We define temperatures exceeding 80°C as abnormal targets in the dataset, where objects approaching 79°C exhibit colors similar to those at 80°C. Such samples are more challenging to distinguish compared to targets at 90°C or 60°C. Existing metrics like EIOU and CIOU do not differentiate between these cases, prompting the introduction of the Focal loss to address this issue. The Focal loss strategy is designed to diminish

the impact of straightforward samples while augmenting the training emphasis on challenging ones. When the loss of a large number of simple samples is accumulated, the number of complex samples is relatively small, so it is difficult to learn the complex samples that the network is interested in, that is, the simple samples swamp the complex samples. Combining Focal loss with EIoU, the resulting formula is presented in Algorithm 2.

Algorithm 2: Focal-EIOU loss

Input: The ground truth and predicted box $A, B \subseteq S \subseteq R^n$

Output: LFocal $- EIOU$

1: For A and B, $\left(x_1^A, y_1^A\right)$, $\left(x_2^A, y_2^A\right)$ represent coordinates of the top left corner and lower right corner of A, $\left(x_1^B, y_1^B\right)$, $\left(x_2^B, y_2^B\right)$ represent coordinates of the top left corner and lower right corner of B.

2: $b = \dfrac{x_1^B + x_2^B}{2}$, $b^{gt} = \dfrac{x_1^A + x_2^A}{2}$, $\rho = \| b - b^{gt} \|$

3: $w^c = \max\left(x_2^A - x_2^B\right) - \min\left(x_1^A, x_1^B\right)$, $h^c = \max\left(y_2^A - y_2^B\right) - \min\left(y_1^A, y_1^B\right)$

4: $1 - IOU = 1 - \dfrac{(A \cap B)}{(A \cup B)} \rightarrow LIOU$

5: $\dfrac{\rho^2\left(b, b^{gt}\right)}{(w^c)^2 + (h^c)^2} \rightarrow Ldis$

6: $\dfrac{\rho^2\left(w, w^{gt}\right)}{(w^c)^2} + \dfrac{\rho^2\left(h, h^{gt}\right)}{(h^c)^2} \rightarrow L_{asp}$

7: $L_{IOU} + L_{dis} + L_{asp} \rightarrow L_{EIOU}$

8: $L_{\text{Focal}} - EIOU = IoU^\gamma L_{EIoU}$

The modification of the loss function can make the network more adaptable to the current dataset, and can better detect targets with too large aspect ratios and poor temperature differentiation, bringing about an improvement in the quality of detection.

3.4 Real-ESRGAN Super-Resolution Pre-Processing

In the current sampling environment, extensive water vapor and acidic gases affect the camera during image capture, resulting in blurred and noisy images. Additionally, transmission losses further degrade image clarity. Moreover, the densely packed arrangement of targets increases detection challenges, leading to potential missed detections and decreased detection accuracy. To mitigate these issues, we preprocess the images using Real-ESRGAN. However, Real-ESRGAN is not inherently suitable for such densely packed targets, and the substantial noise in the images can lead to artifacts in the trained network. Therefore, improvements to the network architecture are necessary.

Initially, during training, we augmented the dataset with 15 cartoon images. Inspired by datasets presented in the Real-ESRGAN literature, this augmentation is motivated by the low noise levels inherent in cartoon images. In neural network training, small datasets can easily lead to overfitting, where the model excessively relies on specific noise patterns in the dataset rather than learning a general mapping from input to output. Introducing a small number of low-noise cartoon images into the dataset serves as a form of regularization, helping the model reduce its reliance on noise patterns and thereby mitigating the risk of overfitting.

Furthermore, we replace the U-Net architecture with Attention U-Net, which introduces grid-based gating mechanisms to enhance attention focus on local regions. This modification is well-suited for densely packed targets, as illustrated in Table 1 of the network structure.

Table 1: Attention U-Net backbone

Backbone:	
#[from, respects, module, args]	
−[−1, 1, Conv2d, [64, 3, 1]]	# 0
−[−1, 1, Conv2d, [128, 4, 2]]	# 1
−[−1, 1, Conv2d, [256, 4, 2]]	# 2
−[−1, 1,Conv2d, [512, 4, 2]]	# 3
−[[2, 3], 1, Attention block, [128]]	# 4
−[3, 1, Conv2d, [256, 3, 1]]	# 5
−[[1, 5], 1, Attention block, [64]]	# 6
−[2, 1, Conv2d, [128, 3, 1]]	# 7
−[[0, 7], 1, Attention block, [32]]	# 8
−[1, 1, Conv2d, [64, 3, 1]]	# 9
−[−1, 1, Conv2d, [64, 3, 1]]	# 10
−[−1, 1, Conv2d, [64, 3, 1]]	# 11
−[−1, 1, Conv2d, [1, 3, 1]]	# 12

Fig. 5 shows the comparison chart of various super-resolution networks, we first use the weights obtained from the training of each network on the public dataset, the direct input to the output image, we can see that the image obtained by SRGAN is still very fuzzy, the image obtained by ESRGAN changes the hue of the original image, and Real-ESRGAN is processed, such as want to become clear, and also does not change the hue, but artifacts appeared, for this point, we improved accordingly, and finally got the image shown in the lower right corner. The modified network is only used as a preprocessing method and does not participate in the training of the YOLO model.
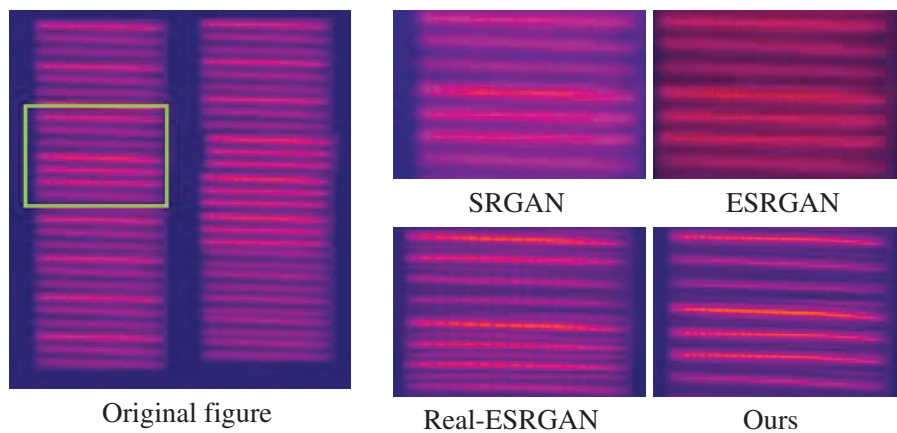
Figure 5: Comparison of various super-resolution networks

4 Experiments

The experiment was implemented based on the pyTorch framework and the code compiler used was PyCharm. The system used a 12th Gen Intel(R) Core(TM) i9-12900H 2.50 GHz CPU, an NVIDIA GeForce RTX 3060 Laptop GPU, and a hard drive with 512 G. The server was accessed using MobaXterm_23 0, and the Ubuntu version was Ubuntu 18.04.6 LTS (GNU/Linux 5.4.0-144-generic x86_64). MobaXterm_23.0 to access the server, Ubuntu version is Ubuntu 18.04.6 LTS (GNU/Linux 5.4.0-144-generic x86_64). The basic data of the network is shown in Table 2.

Table 2: Experimental parameters

Component	Name/value
Input image size	1280×1280
Epoch	300
Training batch size	32
Initial learning rate	0.01
Momentum	0.937
Optimizer	SGD

4.1 Experimental Data

The dataset used in the experiments is homemade and is obtained through the system shown in Fig. 6. The system consists of several infrared cameras, model MAG32 of Shanghai Magnity Electronics Co., Ltd. (Shanghai, China) with pixels of 384×288 and a pixel size of 17 um. Temperature data of the polar plates are firstly collected by these infrared cameras and then transmitted to the server through a switch and fiber optic. Switch optical fiber transmission to the upper-level network, all the data are aggregated and then arrive at the server, and finally recorded in real-time in the software ThermoGroup.
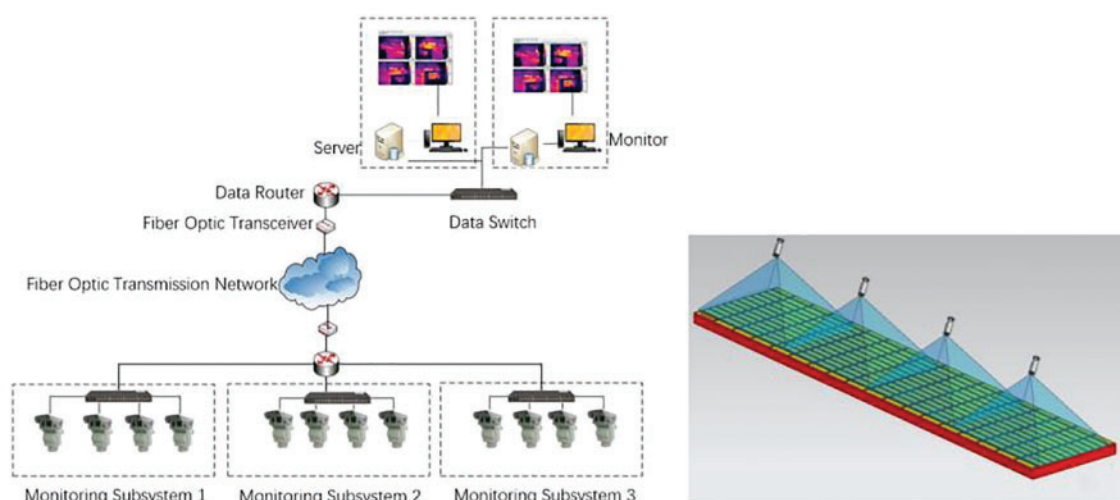
Figure 6: Brief architecture of image acquisition system

A total of 1200 images were collected for the experiment. The image format is BMP, and the dataset is labeled with 'right' and 'wrong' using Vott, above 80°C is noted as 'Wrong', and below 80°C is noted as 'Right'. Due to the overall small size of the dataset, the dataset was expanded by adding noise and random brightness adjustment, and the final number of datasets used for training was 3112, and the number of datasets used for testing was 816.

As shown in Fig. 7, the data set is processed by adding noise and adjusting the brightness, which can mimic the interference of the field environment to the infrared camera.

(a) Original image   (b)Adding noise and adjusting brightness
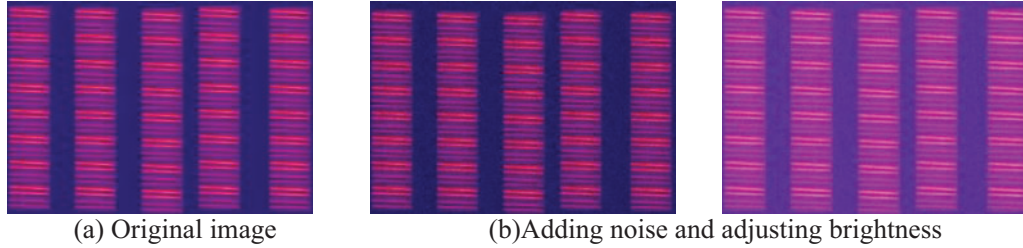
Figure 7: Dataset expansion status

4.2 Evaluation Metrics

Based on the characteristics of the project task, recall, precision, mAP, and GFLOPs amount are used to evaluate the improved performance of the network. The evaluation criteria are expressed as follows:

$$P = \frac{T_P}{T_P + F_P} \tag{2}$$

$$R = \frac{T_P}{T_P + F_N} \tag{3}$$

$$mAP = \frac{\sum_{i=1}^{K} APi}{K} \tag{4}$$

$$\text{FLOPs} = 2 \times C_{out} \times H_{out} \times W_{out} \times (C_{in} \times k^2 + bias) \tag{5}$$

where $1 GLOPs = 10^9 FLPOs$, GLPOs are used as a metric, which is measured in billions of floating point operations that can be performed per second. This metric not only reveals the computational requirements of the model but also reflects the hardware specifications required for its operation and the complexity of the network architecture.

4.3 Experimental Comparison with Other Models

Based on the above metrics, the results of YOLOv5s, YOLOv6s, YOLOv7, and YOLOv5s-RMF on the homegrown datasets were compared, each one of which was obtained from the autonomous dataset trained with the pre-trained models yolov5s.pt, yolov6s.pt, yolov7.pt, and yolov5s.pt. Results are shown in Table 3.

As can be seen from the above table, the precision of YOLOv5 + DenseNet is higher than our method, but the recall is low, confirming the analysis earlier in this paper that simply deepening the depth and width of the network without shallow information protection is detrimental to the recall

improvement. and although the recall of ours is lower than that of YOLOv7, the GLOPs are only 0.2 higher than those of YOLOv5s, and much smaller than others.

Table 3: Comparison of different algorithms

Name	Precision/%	Recall/%	mAP@.5/%	GFLOPs
YOLOv5s	92.7	61.1	78.4	15.8
YOLOv5 + DenseNet [38]	94.5	69.9	83.3	208.6
YOLOv5 + Soft-NMS [39]	93.2	60.0	78.0	15.8
YOLOv5 + DoubleHead [40]	93.0	68.3	80.6	66.2
YOLOv6s	82.9	59.2	82.9	42.8
YOLOv7	90.3	83.1	90.3	105.1
Ours	94.0	70.8	83.6	16.0

4.4 Ablation Experiments

To further verify the feasibility of the improved YOLOv5 network algorithm and the different effects of each module on the detection results, ablation experiments were carried out on each module, firstly comparing the effects before and after the super-resolution processing, i.e., rows one, two and two in the table, the network precision rate after super-resolution increased by 3% and recall increased by 0.3%, which proves that super-resolution processed images can indeed provide the network with more feature information. Next, training on the super-resolution processed dataset, modifying the loss function in the network, aiming to help the network better adapt to the target with larger length and width, we can see that the precision rate of the network increases by 2.5%, and the recall rate increases by 1.3%, and finally, modifying the first CBS in the network, because convolution of the image will reduce the shallow information, so the Foucs operation is used to reduce the fine-grained information of these dilution, and at the same time, to improve the recall, the Marr operator is added to Focus, which provides clearer boundary information for the network, and it can be seen that the precision is still higher than the initial network although it has been decreased, and at the same time, the recall is increased by 8.2%. The details of the ablation experiment are shown in Table 4.

Table 4: Ablation study

YOLOv5s	Real-ESRAGN	Focal-EIOU	Focus	Precision/%	Recall/%
✓				92.7	61.1
✓	✓			95.7	61.4
✓	✓	✓		98.2	62.6
✓	✓	✓	✓	94.0	70.8

The comparative results are shown in Fig. 8, the left figure is the result plot obtained by the YOLOv5s network, running a 384 × 288 image, and it can be seen that in the figure, there are some boxes with more than one target inside them, and some targets are miss-detected. The right figure is the result obtained by the YOLOv5-RMF network running a four-times magnified image, and there

were no missed detections. Therefore, this network helps to improve the overall inspection results and helps factories to automate production, increase productivity, and optimize staffing.
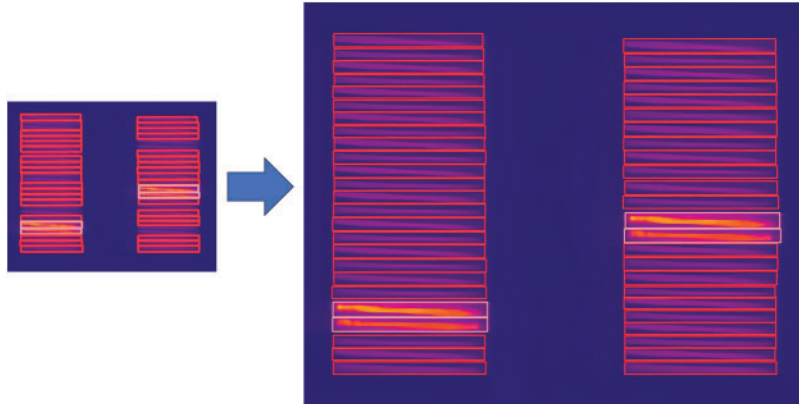

Figure 8: Comparison of YOLOv5s and YOLOv5-RMF running results

5 Conclusion

For YOLOv5s in dealing with infrared dense targets, there are problems: the input infrared image resolution is not high, the details are not clear; infrared picture grey scale change is not obvious, and the target arrangement is very dense, resulting in difficulty to detect the target boundaries; the network's inherent loss function struggles adapt to the goal of increasing and decreasing the size of the edges in the same way, and the difficulty of different samples in the dataset varies, so the loss function cannot be balanced effectively. We propose several methods for improvement. Firstly, after modifying Real-ESRGAN, the images are preprocessed with a super-resolution network; then the first CBS of the network is modified and replaced with Focus, and at the same time the Marr boundary detection operator is introduced in Focus; finally, the loss function of the network is modified. The experimental results show that this method effectively reduces the phenomenon of missed detection, improves recall and accuracy, makes up for the insufficiency of the original network, and can achieve the fast detection of infrared-dense targets.

Based on the current operation of 216 electrolysis cells per day, each with a capacity of 0.46 tons per cell per day, and considering an average of 3 electrode plate short circuits and interruptions (with 40% periodic energization), the system aims to increase effective electrolysis of electrode plates from 28 plates per cell per day to 31 plates. With this system, the single-cell capacity is expected to increase to 0.483 tons per cell per day. The system also achieves a 1.44% increase in current efficiency. In summary, adopting this system enables timely detection of short circuit faults, reduces overall energy consumption, optimizes personnel deployment, and achieves intelligent electrolysis cell monitoring.

In our experiments, the network can be used in all electrolyzer production plants, but the experiments have only investigated the effects of short circuits and not the problems caused by disconnection and leakage losses. At the same time, we observed that employing super-resolution techniques enhances the visual clarity of images, however, we noticed a potential issue where the images appear excessively smooth. Addressing the issues of smoothness, disconnection, and leakage will be our next research direction.

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Huiling Yu, Yanqiu Hang; data collection: Yanqiu Hang; analysis and interpretation of results: Yanqiu Hang; draft manuscript preparation: Yanqiu Hang, Shen Shi, Kangning Wu, Yizhuo Zhang. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] Y. Hu, M. Xiao, K. Zhang, Q. Kong, G. Han and P. Ge, "Aerial infrared target tracking in severe jamming using skeletal tracking technology," *Infrared Phys. Technol.*, vol. 113, 2021, Art. no. 103545. doi: 10.1016/j.infrared.2020.103545.

[2] W. Zhang, X. Fu, and W. Li, "The intelligent vehicle target recognition algorithm based on target infrared features combined with lidar," *Comput. Commun.*, vol. 155, pp. 158–165, 2020. doi: 10.1016/j.comcom.2020.03.013.

[3] J. Zhu, W. Li, D. Lin, H. Cheng, and G. Zhao, "Intelligent fire monitor for fire robot based on infrared image feedback control," *Fire Technol.*, vol. 56, pp. 2089–2109, 2020. doi: 10.1007/s10694-020-00964-4.

[4] S. S. Rawat, S. K. Verma, and Y. Kumar, "Review on recent development in infrared small target detection algorithms," *Proc. Comput. Sci.*, vol. 167, pp. 2496–2505, 2020. doi: 10.1016/j.procs.2020.03.302.

[5] Y. He *et al.*, "Infrared machine vision and infrared thermography with deep learning: A review," *Infrared Phys. Technol.*, vol. 116, 2021, Art. no. 103754. doi: 10.1016/j.infrared.2021.103754

[6] D. Zhang, J. Zhang, K. Yao, M. Cheng, and Y. Wu, "Infrared ship-target recognition based on SVM classification," *Infrared Laser Eng.*, vol. 45, no. 1, pp. 1–6, 2016.

[7] P. Sun *et al.*, "Sparse R-CNN: End-to-end object detection with learnable proposals," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14454–14463. doi: 10.48550/arXiv.2011.12450.

[8] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, Real-time object detection," in *2016 IEEE Conf. on Comput. Vis. and Pattern Recognit.*, Las Vegas, NV, USA, IEEE, 2016, pp. 779–788.

[9] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *IEEE Conf. on Comput. Vis. & Pattern Recognit.*, Honolulu, HI, USA, IEEE, 2017, pp. 6517–6525.

[10] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[11] C. Wang, A. Bochkovskiy, and H. M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network," in *IEEE/ CVF Conf. on Comput. Vis. and Pattern Recognit.*, Nashville, TN, USA, IEEE, 2021, pp. 13029–13038.

[12] C. Wang, H. M. Liao, Y. Wu, P. Chen, J. Hsieh and I. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Seattle, WA, USA, Jun. 2020, pp. 390–391.

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, 2015. doi: 10.1109/TPAMI.2015.2389824.

[14] C. Li *et al.*, "YOLOv6: A single-stage object detection framework for industrial applications," Sep. 07, 2022, *arXiv:2209.02976*.

[15] C. Wang, A. Bochkovskiy, and H. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. of the IEEE/ CVF Conf. on Comput. Vis. and Pattern Recognit.*, Vancouver, BC, Canada, IEEE, 2023, pp. 7464–7475.

[16] Y. Cheng, X. Lai, Y. Xia, and J. Zhou, "Infrared dim small target detection networks: A review," *Sensors*, vol. 24, no. 12, 2024, Art. no. 3885. doi: 10.3390/s24123885.

[17] X. Jiao, Y. Yang, and C. Zhou, "Seam tracking technology for hyperbaric underwater welding," (in Chinese), *Chin. J. Mech. Eng. (CJME)*, vol. 22, no. 2, pp. 265–269, 2009. doi: 10.3901/CJME.2009.02.265.

[18] K. Li, J. Yang, and Z. Huang, "Improved YOLOv3 target detection based on boundary limit point features," *J. Comput. Appl.*, vol. 43, no. 1, pp. 81–87, 2023.

[19] R. Hamid, T. Nathan, J. Gwak, A. Sadeghian, L. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box," in *Proc. of the IEEE/ CVF Conf. on Comput. Vis. and Pattern Recognit.*, Los Angeles, CA, USA, IEEE, 2019, 658–666.

[20] Z. Zheng, P. Wang, W. Liu, J. Li, R. Ye and D. Ren, "Distance-IoU loss: Faster and better learning for bounding box regression," *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 7, pp. 12993–13000, 2020. doi: 10.1609/aaai.v34i07.6999.

[21] Z. Zheng *et al.*, "Enhancing geometric factors in model learning and inference for object detection and instance segmentation," *IEEE Trans. Cybern.*, vol. 52, no. 8, pp. 8574–8586, 2021. doi: 10.1109/TCYB.2021.3095305.

[22] T. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detectin," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 2017, no. 99, pp. 2999–3007, 2017.

[23] Z. Zhao, F. Ni, Y. Song, H. Li, G. Wang and B. Yang, "Substation equipment defect detection method based on improved SRGAN," in *Advances in Machinery, Materials Science and Engineering Application IX*, IOS Press, 2023, pp. 687–693.

[24] X. Zhou, L. Jiang, C. Hu, S. Lei, T. Zhang and X. Mou, "YOLO-SASE: An improved YOLO algorithm for the small targets detection in complex backgrounds," *Sensors*, vol. 22, no. 12, 2022, Art. no. 4600. doi: 10.3390/s22124600.

[25] S. Liu, D. Huang, and Y. Hong, "Receptive field block net for accurate and fast object detection," in *Proc. of the Eur. Conf. on Comput. Vis. (ECCV)*, Munich, Germany, 2018, pp. 385–400.

[26] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. of the IEEE Conf. on Comput. Vis. and Pattern Recognit.*, Tianjin, China, 2017, pp. 4681–4690. doi: 10.48550/arXiv.1609.04802.

[27] X. Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. of the Eur. Conf. on Comput. Vis. (ECCV) Workshops*, Munich, Germany, 2018.

[28] Z. Wang, Y. Dong, D. Niu, M. Liu, Q. Li and X. Chen, "Billet number recognition based on ESRGAN and improved YOLOv5," in *2022 37th Youth Acad. Annu. Conf. of Chin. Assoc. of Automat. (YAC)*, Beijing, China, IEEE, 2022, pp. 1384–1389. doi: 10.1109/YAC57282.2022.10023659.

[29] B. Koonce, "MobileNetV3," in *Convolutional Neural Networks with Swift for Tensorflow*, Berkeley, CA: Apress, 2021, pp. 125–144. doi: 10.1007/978-1-4842-6168-2_11.

[30] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *IEEE/ CVF Int. Conf. on Comput. Vis. Workshops*, Montreal, QC, Canada, 2021, pp. 1905–1914.

[31] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, 2014. doi: 10.48550/arXiv.1409.1556.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. and Comput.-Assist. Intervent.–MICCAI 2015: 18th Int. Conf.*, Munich, Germany, Springer International Publishing, 2015, pp. 234–241.

[33] S. Yang, Z. Zhang, B. Wang, and J. Wu, "DCS-YOLOv8: An improved steel surface defect detection algorithm based on YOLOv8," in *Proc. of the 2024 7th Int. Conf. on Image and Graph. Process.*, Beijing, China, 2024, pp. 39–46. doi: 10.1145/3647649.364765.

[34] J. Dai *et al.*, "Deformable convolutional networks," in *Proc. of the IEEE Int. Conf. on Computer Vision*, Venice, Italy, 2017, pp. 764–773. doi: 10.48550/arXiv.1703.06211.

[35] A. G. Roy, N. Navab, and C. Wachinger, "Concurrent spatial and channel 'squeeze & excitation' in fully convolutional networks," in *Medical Image Comput. and Comput. Assist. Intervent.–MICCAI 2018: 21st Int. Conf.*, Granada, Spain, Springer International Publishing, 2018, pp. 421–429.

[36] H. Zheng, S. Dembele, Y. Wu, Y. Liu, H. Chen and Q. Zhang, "A lightweight algorithm capable of accurately identifying forest fires from UAV remote sensing imagery," *Front. For. Glob. Change*, vol. 6, 2023, Art. no. 1134942. doi: 10.3389/ffgc.2023.1134942.

[37] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu and C. Xu, "GhostNet: More features from cheap operations," in *Proc. of the IEEE/CVF Conf. on Comput. Vis. and Pattern Recognit.*, 2020, pp. 1580–1589. doi: 10.48550/arXiv.1911.11907.

[38] X. Zhang, M. Yan, D. Zhu, and G. Yang, "Marine ship detection and classification based on YOLOv5 model," *J. Phys. Conf. Ser.*, vol. 2181, no. 1, 2022, Art. no. 012025. doi: 10.1088/1742-6596/2181/1/012025.

[39] H. Wang, J. Jin, H. Ke, and X. Zhang, "DDH-YOLOv5: Improved YOLOv5 based on double IoU-aware decoupled head for object detection," *J. Real Time Image Process.*, vol. 19, no. 6, pp. 1023–1033, 2022. doi: 10.1007/s11554-022-01241-z.

[40] J. Shi, L. Li, F. Liu, and C. Xu, "Remote sensing image objects detection algorithm based on improved YOLOv5," in *Int. Conf. on Mech. and Robot. (ICMAR 2022)*, Zhuhai, China, 2022, vol. 2331, pp. 1206–1213. doi: 10.1117/12.2652220.