Adv. Manuf.
https://doi.org/10.1007/s40436-025-00569-6

# Pyramid-based anti-fisheye feature enhancement preprocessing algorithm in torpedo can electrical devices: application in steel rolling process

**Tian-Jie Fu[1] · Shi-Min Liu[2,3] · Pei-Yu Li[1] · Ruo-Xin Wang[2,3]**

Received: 21 May 2024 / Revised: 5 July 2024 / Accepted: 21 May 2025


**Abstract** The steel manufacturing industry currently urgently needs highly accurate detection algorithms for electrical connection devices to slow down the time and danger of electrical connections to torpedo cans during high-temperature operations. The fisheye effect and fuzzy features of industrial cameras seriously affect accuracy and effectiveness, hindering the widespread application of object detection algorithms in the manufacturing industry. We propose a feature enhancement preprocessing algorithm for torpedo can electrical devices based on the pyramid structure that resists fisheye effects and serves to detect and locate electrical connection devices. With the aid of this preprocessing algorithm, the detection efficiency and accuracy of state-of-the-art (SOTA) object detection models are significantly improved. Experimental validation confirms the superiority of our method over other SOTA methods. With the application of our preprocessing algorithm, the production capacity of the steel plant increased by 31.8%, and material wastage caused by transportation decreased by 10.9%.

**Keywords** Anti-fisheye effect (AFE) · Feature enhancement · Machine vision · Steel smelting · Torpedo can

✉ Shi-Min Liu
shimin.liu@polyu.edu.hk

[1] School of Mechanical Engineering, Zhejiang University, Hangzhou 310058, People's Republic of China

[2] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, People's Republic of China

[3] State Key Laboratory of Ultra-precision Machining Technology, Department of Industrial Systems and Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, People's Republic of China

## 1 Introduction

Recently, the rapid development in such areas as the HE series steel industry is one fraught with numerous safety hazards during the steel smelting process. Air pollution caused by the smelting process can result in health issues such as respiratory and skin irritation. It is therefore paramount for steel factories to bolster the implementation and maintenance of safety facilities and procedures to mitigate risks associated with these hazardous processes. Industrial automation and other smart industry applications are increasingly becoming an integral part of the steel industry. Intelligent automation technologies are being employed to automate manual processes, reduce costs, enhance production efficiency, and ensure safety [1–4]. Furthermore, machine learning algorithms are widely used for data analysis and detection functions, enabling better decision-making and more efficient production processes. Advanced analytics are utilized to optimize production and inventory management, enabling factories to produce goods with fewer delays and at lower costs. Consequently, the steel industry is reaping the benefits of smart industry applications, providing effective opportunities for enhancing productivity and reducing costs.

Figure 1 presents a flowchart of the smelting process, where the blast furnace is the primary iron smelting equipment. Iron ore, limestone, and coal must undergo smelting in the blast furnace to yield pig iron or molten iron for various applications [5]. The blast furnace smelting is the first step in steel smelting, where at high temperatures, the carbon in coke reacts with the oxygen in the injected air to reduce iron oxides in the iron ore to metallic iron. The molten iron, heated to extremely high temperatures post-combustion, is poured into torpedo or molten iron cans for transportation to the next smelting stage. The exceedingly high temperature of

Published online: 02 September 2025

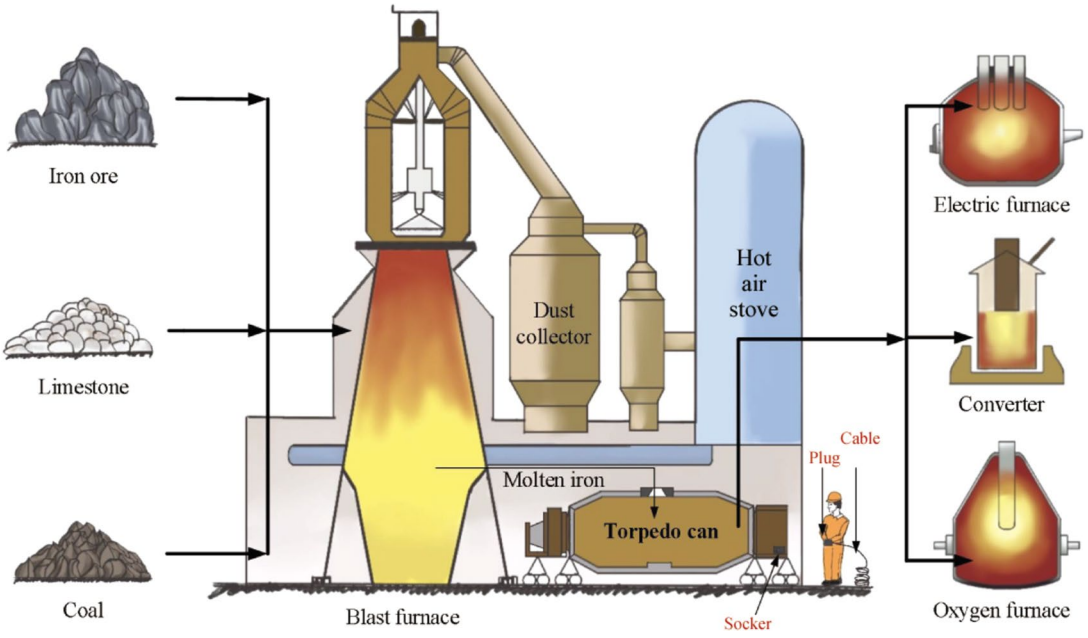


**Fig. 1** Smelting process diagram

molten iron poses significant safety risks to electrical workers [6]. Additionally, the dim environment during iron folding and the fisheye effect of industrial cameras significantly impedes target detection. However, the rapid advancement of smart industry and neural network technology offers the potential for automating the electrical connection operation in torpedo cans.

### 1.1 Object detection (OD) technology

In recent years, OD technology has become a rapidly developing field in computer vision. Through the efforts of scholars, OD technology has made significant strides in terms of accuracy and inference speed [7–12]. Early OD primarily relied on handcrafted features and models. However, these methods had limited scalability and lacked generalization ability, performing poorly on unseen data. With the enhancement in computational system drive capabilities in recent years, the evolution of deep learning models has facilitated the development of OD technology architectures. Dong et al. [13] obtained the optimal object anchor scale for high-speed rail inspection system (HSRIS) OD based on adaptive object scale learning operators and designed a detection method for HSRIS objects based on convolutional neural networks (CNNs). Ren et al. [14] amplified the subtle differences between objects and the background, and established multiple texture perception refinement modules to learn texture perception features in deep CNNs for target OD. Liu et al. [15] proposed a triple-supervised dual-task network for objects, background, and boundaries, accurately detecting the location and detailed boundaries of objects. However, to date, no scholars have proposed relevant OD algorithms for environments that exhibit both fisheye effects and low illumination. Furthermore, there is currently no corresponding state-of-the-art (SOTA) detection algorithm for torpedo can's electrical connection devices in the steel industry.

### 1.2 Anti-fisheye effect (AFE) technology

In recent years, fisheye effect detection technology has been extensively studied. With the advancement of machine learning and computer vision, various algorithms have been proposed and improved for fisheye effect detection. Comparative experiments have shown that methods based on deep learning can significantly enhance the accuracy of fisheye effect detection compared to traditional methods [16]. Fan et al. [17] proposed that the correction results of distorted images of the same scene shot with different lenses should be identical, and designed a self-supervised image rectification (SIR) method based on neural networks. Chao et al. [18] introduced pixel level distortion flow and internal distortion consistency features, and validated its effectiveness through experiments. Li et al. [19] proposed the no-prior fisheye representation method based on the FisheyeDet contour-based object detector, which had good generalization ability. Additionally, some scholars have separated the peripheral area from the central area when learning features to adapt to the shape and proportion of face anchor points in the AFE network [20]. Currently, the AFE is often achieved through multi-camera positioning. However, considering the impacts of cost and environmental

factors, the iron folding process tends to favor single-camera positioning. Furthermore, to date, there are no existing algorithms that resist the fisheye effect in low-illumination environments.

### 1.3 Feature enhancement technology

During OD, there are often many negative factors, such as lighting, noise, camera precision, and shooting environment, which can lead to unclear and indistinct features. Therefore, feature enhancement technology has gradually become a hot research topic in the field of machine learning in recent years [21–23]. For instance, Wang et al. [24] proposed an adaptively fused attention module (AFAM) applied in the textile manufacturing industry to enhance spatial and channel features. To address the issue of color distortion and detail blurring in underwater shooting images, Qi et al. [25] introduced semantic information as high-level guidance and proposed an underwater image enhancement network. Guo et al. [26] proposed a multi-scale Retinex algorithm with a color protection (MSRCP) image enhancement algorithm to compensate for the lack of lighting and slow detection speed in underwater environments. Additionally, some scholars have proposed tracking methods that combine feature enhancement and template updates to address issues such as trackers in videos not being able to focus on global information and not adapting well to target changes [27]. Due to the singular nature of the scenes captured by the camera during the iron folding process, the presence of a large number of noise points, and the minimal contrast between the electrical connection device and the background, it is difficult to distinguish between them. Therefore, it is challenging to directly utilize previous enhancement algorithms based on channel or spatial features.

### 1.4 Research gaps

From the above analysis, we can identify the following issues in the current automatic recognition process of steel rolling.

(i) Through the analysis of the torpedo can iron folding process, it was found that the electrical connection operation of the torpedo can was dangerous and difficult to ensure human safety. There is currently a lack of an effective automated electrical connection detection algorithm to be combined with a robotic arm to replace manual operations, thereby ensuring human safety and instability.

(ii) The use of fisheye cameras ensures the field of view for detection but sacrifices detection accuracy. Therefore, the current conventional OD algorithms cannot meet the detection accuracy requirements of actual torpedo can's electrical connection devices.

(iii) The on-site environment during the electrical connection of the torpedo can is relatively dark, so the captured images have issues of blurring and low contrast. The current anti-fisheye detection algorithms cannot effectively extract the features of the electrical connection device. A highly stable, fast, and efficient disguised OD algorithm needs to be combined with a robotic arm to achieve process automation.

The steel smelting process is becoming more intelligent, and this paper will explain this phenomenon in detail. The remainder of this paper is structured as follows. Chapter 2 introduces the automatic recognition strategy for the steel rolling process. Chapter 3 presents the proposed AFE and feature enhancement (AFE-FE) preprocessing algorithm for torpedo can's electrical connection devices based on a pyramid structure. Chapter 4 experiments on the proposed model and verifies its effectiveness. Finally, the main contributions of this paper are summarized and prospects are discussed.

## 2 Automatic identification strategy for the steel rolling process

The transportation of high-temperature molten iron is a crucial step in steel production. The transportation process includes an additional iron folding step, where the molten iron inside the torpedo can is poured into the molten iron can for subsequent processes, as shown in Fig. 2. The torpedo can, filled with molten iron, is rotated, pouring the molten iron from the torpedo can into a molten iron can. Transferring molten steel using a torpedo can pose inherent risks, including high voltage connections, potential tipping during pouring, and the hazardous proximity of personnel to the gas cylinder and tipping point. Workers handling heavy electrical plugs and controlling torpedo can tipping in the real-time face a severe workload, increasing the potential for accidents, as shown in Fig. 3. The main processes of torpedo car iron transfer include: (i) the train pulls the torpedo car into the steel plant; (ii) workers carry and install the plug; (iii) workers operate the torpedo car to tilt and carry out iron tapping.

In applications requiring a wide field of view, fisheye cameras are becoming increasingly popular. These cameras offer a very large field of view with minimal distortion, as shown in Fig. 4, representing the performance of fisheye cameras under low light conditions. Fisheye cameras, compared to other wide-angle types, offer simpler, cost-effective implementation and comprehensive scene monitoring for improved situational awareness. Given the torpedo can's variable positioning during transportation, with up to a meter of

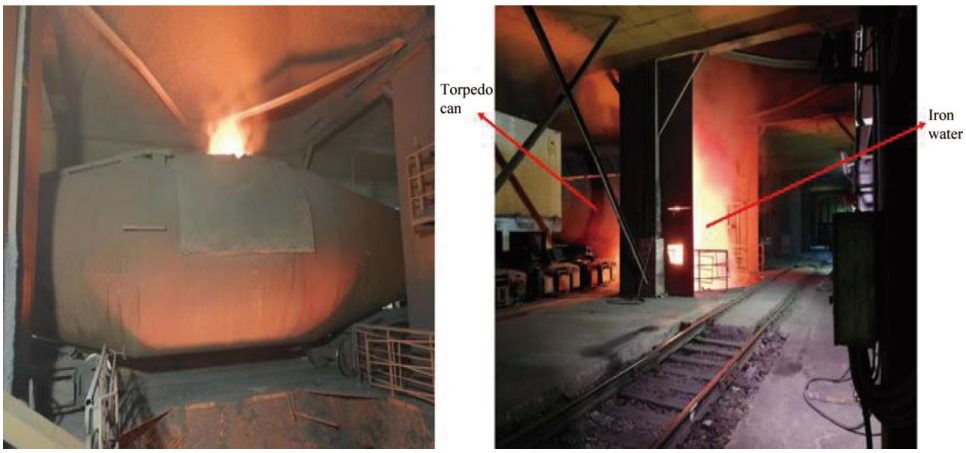


**Fig. 2** Iron folding site

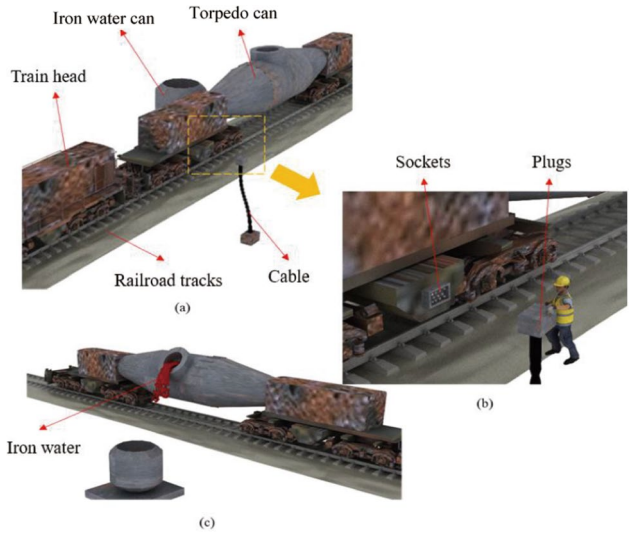


**Fig. 3** Torpedo can iron folding process

**Fig. 4** Fisheye camera shooting effect

**Fig. 5** Recognition and localization of fisheye images and normal images

deviation, the fisheye camera's broader field of view ensures effective recognition.

As shown in Fig. 5, the red box represents the positioning under the fisheye effect, while the gray box signifies the actual positioning. The disparity between the two exceeds 1−2 cm. Successful power connection for the torpedo can only be achieved when the robotic arm guides the socket to the precise location. The fisheye camera introduces substantial positioning errors. Deviations exceeding 2 mm can result in damage to the robotic arm structure and the power-receiving device, posing a risk of serious accidents and operational inconvenience in production. In steel production, the torpedo can iron folding process impacts production efficiency and product quality, yet current methods are inefficient, risky, and involve harsh working conditions. Combining machine learning with robotic arms, where a neural network detection algorithm provides coordinates for a robotic arm to automate iron folding, can address these issues.

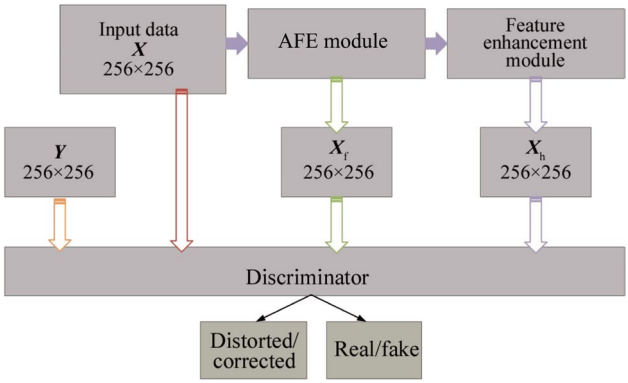


**Fig. 6** Overall network structure

## 3 Algorithm for automatic identification of rolling processes

### 3.1 Overall network structure

In response to the aforementioned engineering problem of torpedo can electrical connection devices, this paper proposes an AFE-FE model based on the pyramid structure. Figure 6 shows the overall network structure of the AFE-FE model, which includes three modules: the AFE module, the feature enhancement module, and the discriminator module. Different modules perform different functions; the AFE module and the feature enhancement module together constitute the main part of the image generation, while the discriminator judges whether the generated image meets the requirements.

The network of the AFE-FE model is divided into four processes.

Step 1: Image $\boldsymbol{X}$ is input into the AFE module and output as $\boldsymbol{X}_{\mathrm{f}}$ to the discriminator.

Step 2: Image $\boldsymbol{X}$ is input into the AFE module, then passed through the feature enhancement module, and output as $\boldsymbol{X}_{\mathrm{h}}$ to the discriminator.

Step 3: Image $\boldsymbol{X}$ is directly input into the discriminator.

Step 4: Image $\boldsymbol{Y}$ is input into the discriminator.

In this context, Image $\boldsymbol{X}$ refers to image data captured on-site using an industrial camera with a fisheye effect. Image $\boldsymbol{Y}$ refers to image data captured under manual lighting conditions using a standard high-definition camera with no distortion when smelting operations are stopped, meeting the requirements for no distortion and clear features.

The four processes occur simultaneously, achieving both the correction of fisheye distortion and the clarity of blurred feature information under low light conditions. Each module of the network is fully trained, improving training quality while saving training time. Figure 7 shows the structural diagrams of the AFE module and the feature enhancement module. Both modules use an attention module to ensure that features are not lost and clear.

### 3.2 AFE module

Figure 7 shows the structure diagram of the AFE module, which learns the mapping function between the distorted image $\boldsymbol{X}$ captured by the industrial camera, the undistorted image $\boldsymbol{Y}$, and the feature-enhanced image $\boldsymbol{X}_{\mathrm{h}}$. This structure corrects the distorted images captured by the fisheye camera. In the model, we used the CSPParkNet53 tiny [28] module to achieve feature-down sampling of the backbone network. Utilizing a 1×1 convolutional layer of micro feature pyramid network (FPN) to fuse features at different scales enhances feature representation while reducing computational complexity. The module inputs the distorted image captured by the industrial camera and outputs an image and distortion flow. The image size is $H\times W\times 3$, and the distortion flow is a feature map of $H\times W\times 2$. $H$ and $W$ represent the height and width of the input image, respectively. The distortion flow represents the pixel-level image coordinate mapping in the predicted undistorted image. Combining the distortion flow with the image yields an undistorted image. This module is designed based on the principles of cross-rotation consistency and warping consistency inherent in fisheye images.

(i) Cross-rotation consistency. Borrowing from the concept of geometric consistent generative adversarial networks (GCGAN) [29], the image $\boldsymbol{X}$ is rotated by 45°, 90°, and 135°and simultaneously input into the prediction module for training. That is, it is rotated by an angle $R_{\theta}$, where $\theta\in\{45°, 90°, 135°\}$. The fisheye distortion flow is mirror-symmetric, so for all input images, including those at the original angle and those at the rotated angle, the rotation flow should remain consistent.

(ii) Warping consistency. The distortion flow of each pixel in the image is outward or inward relative to the image center. It also exhibits radial symmetry, meaning that the distortion flow is the same for pixels equidistant from the image center. Moreover, the distortion flow has a smoothness characteristic, meaning that the difference in distortion flow size for adjacent pixels along the horizontal or vertical direction is minimal.

### 3.3 Feature enhancement module

Figure 7 also shows the structural diagram of the feature enhancement module. This structure adopts a pyramid configuration, progressively enhancing and merging weaker features at each level, resulting in a feature representation that is rich and discriminative [30]. The input is the corrected image output by the anti-fisheye module, and the output is

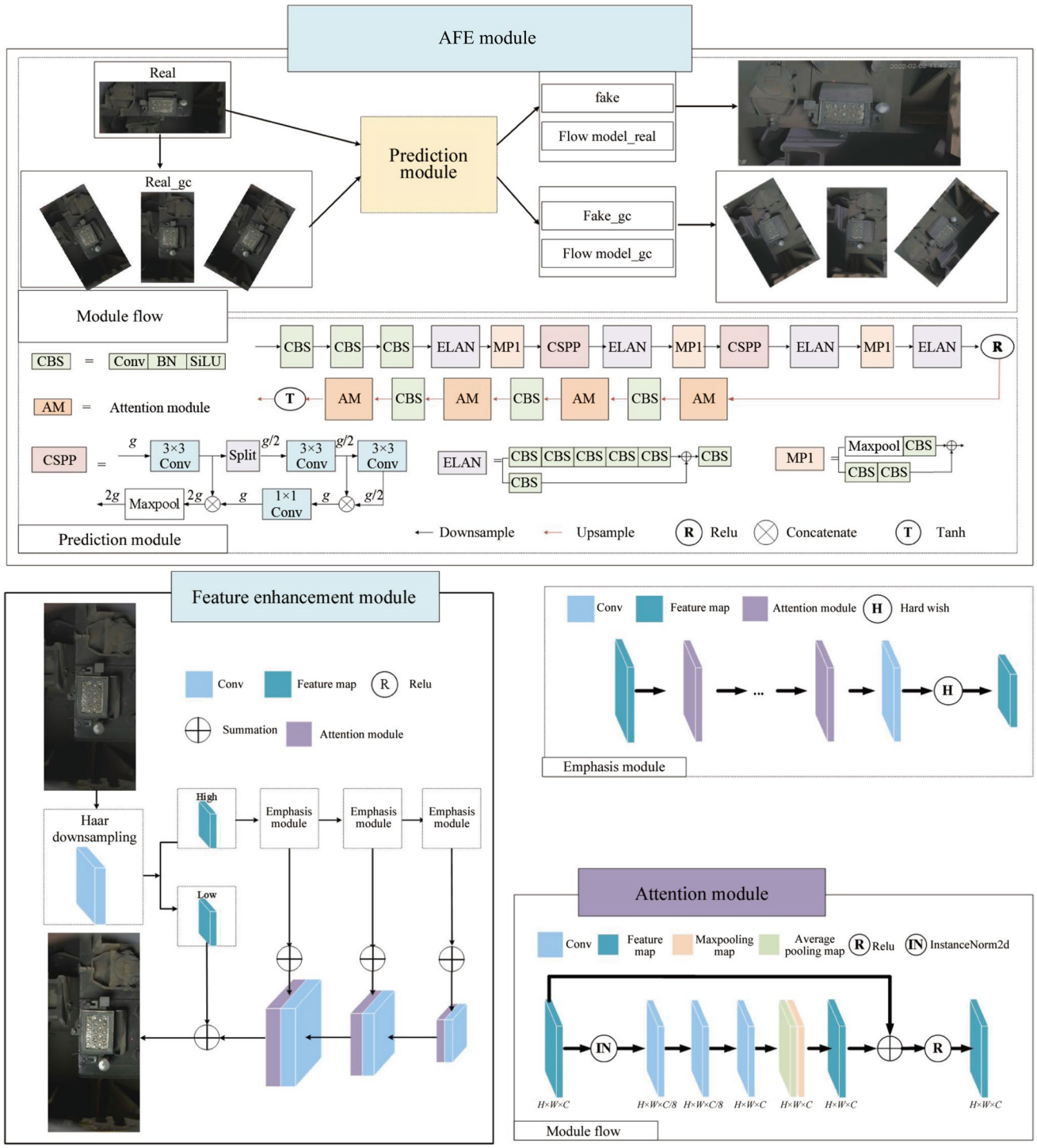


**Fig. 7** Structure of anti-fisheye and feature enhancement modules

the feature-enhanced undistorted image. Initially, the Haar wavelet is utilized to gather high-frequency and low-frequency features within the image. The emphasis module then enhances high-frequency features, which are progressively overlaid with low-frequency features to yield an image of the electrical connection device with enhanced features. The use of an attention module ensures that detailed features are not lost during the feature extraction process. This approach is particularly beneficial in scenarios where the image may contain a wealth of complex or subtle features that are essential for accurate OD. By enhancing high-frequency features and preserving detailed features, the AFE-FE model can

generate a more comprehensive and discriminative representation of the electrical connection device, thereby improving the performance of the subsequent detection model.

### 3.4 Loss function

The discriminator primarily consists of two modules, Identify and Classification, i.e., $\boldsymbol{D} = \left(D_{\text{ide}}, D_{\text{cl}}\right)$. $D_{\text{ide}}$ differentiates between the generated images, namely $\boldsymbol{X}_{\text{f}}$ and $\boldsymbol{X}_{\text{h}}$, and the undistorted real image $\boldsymbol{Y}(y)$. $D_{\text{cl}}$ outputs are Distorted (distorted or unenhanced feature images) and Corrected (undistorted and feature-enhanced images). $D_{\text{cl}}$ defines undistorted images as True and generated images as False, and adversarial training is conducted using the principle of adversarial networks. $\boldsymbol{A}(x)$ represents the image generated by the anti-fisheye module. $\boldsymbol{F}(\boldsymbol{x}_{\text{f}})$ represents the image generated by the feature enhancement module. $\boldsymbol{x}_{\text{f}}$ represents the image generated by the anti-fisheye module. $L_{\text{ide}}$ represents identity authentication loss, used to distinguish the authenticity of real and undistorted images from generated images. $E_{y\sim\text{date}(y)}$ represents the expected distribution of true undistorted images $(y)$ . $E_{x\sim\text{date}(x)}$ represents the distribution of the original distorted input image $(x)$ . $L_{\text{cl}}$ represents classification loss, used to identify image distortion attributes. $L_{\text{D}}$ represents the total loss function of the discriminator, which is used to optimize the ability of identity discrimination and distortion classification, driving the discriminator to accurately distinguish.

$$\begin{aligned} L_{\text{ide}} =& E_{y\sim\text{data}(y)} \log_{10}(D_{\text{ide}}(\boldsymbol{y})) \\ &+ E_{x\sim\text{data}(x)} \log_{10}(1 - D_{\text{ide}}(\boldsymbol{A}(\boldsymbol{x}))) \\ &+ E_{x\sim\text{data}(x)} \log_{10}(1 - D_{\text{ide}}(\boldsymbol{F}(\boldsymbol{x}_{\text{f}}))) \end{aligned}, \tag{1}$$

$$\begin{aligned} L_{\text{cl}} =& E_{y\sim\text{data}(y)} \log_{10}(D_{\text{cl}}(\boldsymbol{y})) \\ &+ E_{x\sim\text{data}(x)} \log_{10}(1 - D_{\text{cl}}(\boldsymbol{x})) \\ &+ E_{x\sim\text{data}(x)} \log_{10}\left(1 - D_{\text{cl}}(\boldsymbol{A}(\boldsymbol{x}))\right) \\ &+ E_{x\sim\text{data}(x)} \log_{10}\left(1 - D_{\text{cl}}(\boldsymbol{F}(\boldsymbol{x}_{\text{f}}))\right), \end{aligned} \tag{2}$$

$$L_{\text{D}} = L_{\text{ide}} + L_{\text{cl}}. \tag{3}$$

Since the images input into the AFE-FE model is not labeled, auxiliary training is conducted based on the special correspondence relationship of image types. This paper adopts cross-rotation consistency and in-image warping consistency features to achieve the function of correcting the fisheye effect. Equations (4) and (5) are the cross-rotation consistency feature loss functions between the original image and the image generated by the AFE module. The subscript 1 denotes the L1 norm (Manhattan norm). Equations (6) and (7) are the feature loss functions for achieving warping consistency. In the equations, $\boldsymbol{P}(\boldsymbol{x})$ is the output of the prediction module, and $\boldsymbol{f} = \boldsymbol{P}(\boldsymbol{x})$. With (0, 0) as the centre, $f_{\text{p}}$ represents the flow condition of the pixel at the image coordinate $p \in R^2$ in the flow-map. $\boldsymbol{R}_{\theta}(\boldsymbol{x})$ represents the image post-rotation, while $\boldsymbol{A}\left(\boldsymbol{R}_{\theta}(\boldsymbol{x})\right)$ denotes the image generated by the anti-fisheye module after rotation. $\boldsymbol{R}_{\theta}(\boldsymbol{A}(x))$ represents anti-fisheye output $A(x$ rotated by $\theta$. $\|\bullet\|_1$ measures the difference between $\boldsymbol{P}$(x) and $\boldsymbol{P}(\boldsymbol{R}_{\theta}$ (x)). $\boldsymbol{D}(x_f)$ represents discriminator output for generated image $\text{x}_f$. $O$ represents overall optimization objective combining min-max over networks ($\boldsymbol{A}$, $\boldsymbol{F}$, $\boldsymbol{D}$). $L_{\text{rad}}(\boldsymbol{A})$ ensures that the anti-fisheye module $\boldsymbol{A}$ preserves geometric consistency under image rotation. $L_{\text{flow}}(\boldsymbol{A})$ maintains consistency in optical flow predictions before and after rotation. $L_{\text{geo}}(\boldsymbol{A})$ enforces radial consistency in flow vectors to align with fisheye distortion patterns. $L_{\text{smo}}(\boldsymbol{A})$ promotes spatial smoothness in the flow field. $L_{\text{mse}}(\boldsymbol{A})$ supervises the feature enhancement module $\boldsymbol{F}$ to preserve content. $L_{\text{A}}$ aggregates geometric, flow, radial, and smoothness constraints for $\boldsymbol{A}$. $L_{\text{F}}$ combines reconstruction fidelity and adversarial loss. $p$ represents the pixel coordinates in the image

$$\begin{aligned} L_{\text{geo}}(\boldsymbol{A}) =& E_{x\sim\text{data}(x)}\left(\left|\left|\boldsymbol{R}_{\boldsymbol{\theta}}(\boldsymbol{A}(\boldsymbol{x})) - \boldsymbol{A}\left(\boldsymbol{R}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)\right|\right|_1\right) \\ &+ E_{x\sim\text{data}(x)}\left(\left|\left|\boldsymbol{A}(\boldsymbol{x}) - \boldsymbol{R}_{-\boldsymbol{\theta}}\left(\boldsymbol{A}\left(\boldsymbol{R}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)\right)\right|\right|_1\right), \end{aligned} \tag{4}$$

$$L_{\text{flow}}(\boldsymbol{A}) = E_{x\sim\text{data}(x)}\left(\left|\left|\boldsymbol{P}(\boldsymbol{x}) - \boldsymbol{P}\left(\boldsymbol{R}_{\boldsymbol{\theta}}(\boldsymbol{x})\right)\right|\right|\right), \tag{5}$$

$$L_{\text{rad}}(\boldsymbol{A}) = E_{x\sim\text{data}(x)}\left(\int_f \left|\frac{\boldsymbol{f}_{\text{p}} \cdot p}{\left|\left|\boldsymbol{f}_{\text{p}}\right|\right| \cdot ||p||}\right| + \int_r \text{Var}\left(\left\{\left|\left|\boldsymbol{f}_{\mathbf{p}}\right|\right| \,|||p|| = r\right\}\right)\right), \tag{6}$$

$$L_{\text{smo}}(\boldsymbol{A}) = E_{x\sim\text{data}(x)}\left(\int_f \left|\left|\nabla_{\boldsymbol{x}}\boldsymbol{f}_{\mathbf{p}}\right|\right|_1 + \left|\left|\nabla_{\boldsymbol{y}}\boldsymbol{f}_{\mathbf{p}}\right|\right|_1\right), \tag{7}$$

$$L_{\text{mse}}(\boldsymbol{F}) = E_{x_{\text{f}}\sim\text{data}(x_{\text{f}})}\left(\left|\left|\boldsymbol{x}_{\mathbf{f}} - \boldsymbol{F}\left(\boldsymbol{x}_{\text{f}}\right)\right|\right|\right), \tag{8}$$

$$L_{\text{A}} = L_{\text{geo}}(\boldsymbol{A}) + L_{\text{flow}}(\boldsymbol{A}) + L_{\text{rad}}(\boldsymbol{A}) + L_{\text{smo}}(\boldsymbol{A}), \tag{9}$$

$$L_{\text{F}} = L_{\text{mse}}(\boldsymbol{F}) + \alpha(L_{\text{bce}}\left(\text{sigmoid}\left(\boldsymbol{D}\left(\boldsymbol{x}_{\text{f}}\right)\right), 1\right), \tag{10}$$

$$O = \arg\min_{\boldsymbol{A}}\min_{\boldsymbol{F}}\max_{\boldsymbol{D}} L(\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{D}). \tag{11}$$

The objective function $L(\boldsymbol{A}, \boldsymbol{F}, \boldsymbol{D})$ incorporates the loss functions from Eq. (1) to Eq. (10). This paper's model addresses the complete optimization objective function as shown in Eq. (11).

## 4 Implementation details

### 4.1 Computer configuration

The method proposed in this paper collects images through an industrial camera with a resolution of 3 840×2 160 pixels. These are initially processed to a size of 256×256, serving as the original sample images. These original images are used as the basic input, and the proposed method is implemented using a Windows 10 operating system, a NVIDIA GeForce RTX 3070 Super graphics card, and a device with 32 GB of video memory. The model's effectiveness is validated using Hikvision CS-MV-CS200-10GM video detection equipment in combination with different SOTA OD models.

### 4.2 Design of the test

#### *4.2.1 Picking out*

The model in this paper requires two types of samples.

(i) Input samples for supervision (Image ***Y***). Therefore, ongoing smelting operations need to be halted, and under the premise of ensuring safety, a lighting device with sufficient brightness is manually held up. A high-definition camera without distortion effects is used to capture a set of undistorted samples with clear features.
(ii) Real original samples (Image ***X***) captured by an industrial camera in real scenarios, forming a set of original samples.

#### *4.2.2 Establishment of a sample library*

Sample library required for the proposed model: After screening, the original dataset contains a total of 22 000 sample images (Image ***X***), and there are 8 100 sample images for supervision (Image ***Y***). The images are stored in the corresponding folder paths of the AFE-FE model.

Sample library required for validation: As the proposed model's effectiveness is validated using SOTA OD models, a comparative sample set needs to be established for this paper.

Sample set 1: A total of 20 000 sample images are selected from the original sample set that has not been trained using the AFE-FE model. These are divided into training, testing, and validation sets at a ratio of 6:2:2. After labeling, the images are stored in their respective folder paths.

Sample set 2: After the AFE-FE model has been trained, this model generates a total of 20 000 images that have been corrected for fisheye distortion and have enhanced features. These are divided into training, testing, and validation sets at a ratio of 6:2:2. After labeling, the images are stored in their respective folder paths.

All the above samples were collected from the smelting site and corresponding databases were constructed.

**Table 1** Model performance

| | Model size/MB | Params/M | FLOPs | MAC/MB |
|---|---|---|---|---|
| AFE-FE | 13.5 | 4.14 | 2.61 GFlops | 76.28 |

#### *4.2.3 AFE-FE model training*

The industrial camera used in this paper captures images of size 3 840×2 160, so the collected images are first proportionally reduced to 256×144 using the Bicubic interpolation method. The edges are then padded to expand them to a size of 256×256.

This can achieve the aforementioned cross-rotation consistency and warping consistency characteristics, helping to improve the accuracy and performance of the model and reduce the risk of overfitting. In addition, the above processing can improve the training speed of the model without sacrificing accuracy or losing features. It can also effectively enhance the model's generalization ability, which is conducive to improving the model's usability. Therefore, it can provide significant advantages for the neural network model in this paper.

We propose an AFE-FE model to implement fisheye correction and feature enhancement. Using a learning rate of $10^{-5}$ and Adam as the optimizer, the model is compiled based on the Pytorch framework, with a batch size set to 8. The aforementioned samples are input into this model for training, and when the training reaches the optimal indicator, the final model parameters of this method are saved. The algorithms are then tested using the test set. In Table 1, we analyzed the model size, parameters (Params), floating-point operations per second (FLOPs), and memory access cost (MAC) measurement to reflect the efficiency of the network.

#### *4.2.4 Object recognition*

Two comparative sample sets required for validation are trained using excellent SOTA OD models. After training, a control group is set up to validate the effectiveness of the proposed model. The OD models selected in this paper include: Focus-and-Detect [31], USTD [32], YOLODrone [33], YOLOv4-MN3 [34], Faster R-CNN [35], and SSD. A control group is set up to validate the superiority of the combination of the aforementioned SOTA models and the AFE-FE model proposed in this paper. In addition, the SIR model,

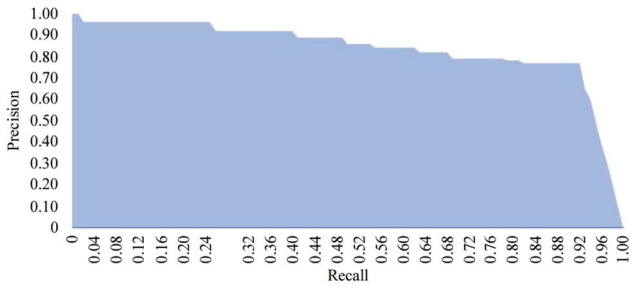


**Fig. 8** $A_P$ computation

FE-GAN model, AFAM model, and SGUIENet model are combined with SOTA object detection models to test the effects of fisheye correction and feature enhancement.

### 4.3 Evaluation metrics

This paper selects the following parameters as evaluation indicators: precision ($P_{re}$), recall ($R_{ec}$), average precision ($A_P$), $A_{P50}$, $A_{P75}$, and accuracy ($A_{ccuracy}$). These indicators can evaluate the positioning accuracy of the OD model. $P_{re}$ represents the ratio of the number of correctly detected targets to the total number of predicted bounding boxes. $R_{ec}$ represents the proportion of correctly detected targets to total targets. $A_P$ is the area between the precision curve and the coordinate axis within the range of [0,1], as shown in Fig. 8. $A_{P50}$ represents the $A_P$ value when IoU is fixed at 50%, and $A_{P75}$ represents the $A_P$ value when IoU is fixed at 75%. $A_{ccuracy}$ represents the proportion of correctly predicted samples in a given test sample set. The formulas are shown as follows

$$P_{re} = \frac{T_P}{T_P + F_P}, \tag{12}$$

$$R_{ec} = \frac{T_P}{T_P + F_N}, \tag{13}$$

$$A_P = \int_0^1 P(R)\mathrm{d}R, \tag{14}$$

$$A_{ccuracy} = \frac{T_P + T_N}{T_P + T_N + F_P + F_N}. \tag{15}$$

In OD, the predicted intersection over union (IoU) between the predicted bounding box and the "ground-truth" bounding box in the corresponding image is defined. If the IoU exceeds a defined threshold, the predicted bounding box in this prediction is considered a true positive ($T_P$). Conversely, if the IoU is below the defined threshold, the predicted bounding box in this prediction is deemed a false positive ($F_P$). $F_P$ represents the total number of negative samples incorrectly predicted as positive samples, i.e., false alarms. False negatives ($F_N$) represent the number of positive samples incorrectly predicted as negative samples, i.e., missed detections. True negative ($T_N$) represent correctly predicted negative samples, i.e., true non-targets identified as negative. The evaluation indicators used in this paper all indicate that the higher the score, the better the performance of the detection model.

**Table 2** Comparison of precisions of different models

| | Precision with AFE-FE model/% | Precision without AFE-FE model/% |
|---|---|---|
| Focus-and-Detect | 95.456 | 84.562 |
| USTD | 94.826 | 88.463 |
| YOLODrone | 96.135 | 87.354 |
| Faster R-CNN | 84.452 | 80.654 |
| SSD | 87.475 | 82.465 |

The $P_{SNR}$ and $S_{SIM}$ [36] indicators are selected to evaluate the differences in pixel values and structure. $P_{SNR}$ represents the ratio of the maximum signal value of an image to the noise in the background. $R$ denotes the maximum possible pixel value of the image, representing the dynamic range of the signal. Mean squared error ($M_{SE}$) quantifies the average squared difference between the original and reconstructed images, measuring pixel-level distortion. The larger the $P_{SNR}$, the higher the image quality. $S_{SIM}$ measures the similarity between images based on luminance, contrast, and structure. Luminance ($\boldsymbol{l}$) compares the mean intensities of images $\boldsymbol{x}$ and $\boldsymbol{y}$, adjusted by exponent $\alpha$. Contrast ($\boldsymbol{c}$) evaluates the standard deviation of pixel values, reflecting texture variations, weighted by $\beta$. Structure ($\boldsymbol{s}$) measures the cross-correlation between $\boldsymbol{x}$ and $\boldsymbol{y}$, capturing spatial relationships, with $\gamma$ as its exponent. The larger the $S_{SIM}$, the smaller the image distortion. The formulas are shown as follows

$$P_{SNR} = 10\log_{10}\left(\frac{R^2}{M_{SE}}\right), \tag{16}$$

$$S_{SIM}(\boldsymbol{x},\boldsymbol{y}) = (\boldsymbol{l}(\boldsymbol{x},\boldsymbol{y}))^{\alpha}(\boldsymbol{c}(\boldsymbol{x},\boldsymbol{y}))^{\beta}(\boldsymbol{s}(\boldsymbol{x},\boldsymbol{y}))^{\gamma}. \tag{17}$$

### 4.4 Testing & evaluation

#### *4.4.1 Comparison with SOTA OD model combination*

Two comparative sample sets required for validation are trained using excellent SOTA OD models. The excellent SOTA OD models selected in this paper include: Focus-and-Detect, USTD, YOLODrone, YOLOv4-MN3, Faster R-CNN, and SSD. After training, control groups are set up:

**Table 3** Test metric scores for the different methods on each test set

| | Our detaset | | | | | | MCindoor20000 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Time/ms | $R_{ec}$/% | $A_P$ | $A_{P50}$ | $A_{P75}$ | $A_{ccuracy}$ | Time/ms | $R_{ec}$/% | $A_P$ | $A_{P50}$ | $A_{P75}$ | $A_{ccuracy}$ |
| Focus-and-Detect+ours | 26.6 | 95.136 | 41.7 | 62.1 | 48.9 | 78.63 | 30.1 | 96.145 | 40.3 | 61.9 | 47.9 | 79.15 |
| Focus-and-Detect | 26.3 | 94.241 | 41.1 | 61.7 | 48.1 | 76.82 | 29.8 | 94.836 | 36.6 | 58.8 | 43.8 | 75.48 |
| USTD+ours | 24.2 | 94.864 | 41.7 | 59.8 | 45.7 | 73.94 | 29.6 | 94.215 | 40.4 | 57.9 | 41.3 | 74.18 |
| USTD | 23.9 | 93.415 | 41.4 | 59.2 | 45.3 | 71.61 | 29.3 | 92.678 | 34.1 | 56.1 | 40.1 | 70.89 |
| YOLODrone+ours | 25.3 | 98.435 | 43.9 | 62.3 | 47.8 | 88.42 | 27.6 | 98.351 | 48.2 | 67.1 | 49.6 | 89.46 |
| YOLODrone | 24.9 | 97.682 | 43.2 | 64.8 | 47.1 | 84.67 | 27.5 | 96.153 | 41.6 | 62.3 | 43.4 | 83.41 |
| Faster R-CNN+ours | 26.9 | 87.552 | 37.4 | 55.8 | 44.2 | 55.75 | 29.4 | 87.115 | 38.4 | 56.6 | 41.5 | 59.15 |
| Faster R-CNN | 26.4 | 87.132 | 36.7 | 54.9 | 43.5 | 52.34 | 29.2 | 83.646 | 35.5 | 53.4 | 40.2 | 53.45 |
| SSD+ours | 27.6 | 89.341 | 39.8 | 57.8 | 44.8 | 72.65 | 24.4 | 88.845 | 37.4 | 56.8 | 41.3 | 73.15 |
| SSD | 27.4 | 88.454 | 38.9 | 57.3 | 44.1 | 70.64 | 24.3 | 84.643 | 34.6 | 51.6 | 40.1 | 70.48 |

**Table 4** Average positioning errors of different algorithms

| | Error with AFE-FE model/mm | Error without AFE-FE model/mm |
|---|---|---|
| Focus-and-Detect | 0.98 | 5.24 |
| USTD | 1.34 | 11.73 |
| YOLODrone | 0.79 | 4.68 |
| Faster R-CNN | 1.64 | 16.84 |
| SSD | 1.98 | 20.16 |

**Table 5** Comparison of $P_{SNR}$ and $S_{SIM}$ of different models

| | $P_{SNR}$/ dB | $S_{SIM}$ |
|---|---|---|
| SIR | 28.46 | 0.75 |
| FE-GAN | 27.15 | 0.68 |
| AFE-FE (ours) | 31.42 | 0.97 |

**Table 6** Comparison of precisions of different models

| | $P_{re}$/% |
|---|---|
| YOLODrone | 87.354 |
| YOLODrone+SIR | 92.456 |
| YOLODrone+FE-GAN | 89.254 |
| YOLODrone+AFE-FE (ours) | 96.135 |

Control group 1. The trained SOTA models are directly used to detect images of torpedo can electrical connection devices collected on-site. Control group 2: The trained AFE-FE model is combined with the SOTA models before detecting images of the torpedo can's electrical connection devices collected on-site. A total of 5 000 images of torpedo can electrical connection devices collected on-site are used as the test set. As shown in Table 2, the precision indicator of the SOTA OD model combined with the AFE-FE model proposed in this paper is significantly higher than the indicator detected directly by the model. This demonstrates that combining the AFE-FE model proposed in this paper with the SOTA models can significantly improve the detection accuracy of torpedo can's electrical connection devices. Among them, the combination of the YOLODrone model and the AFE-FE model achieved the highest score.

Table 3 calculates the test metric scores before and after combining different SOTA models with our proposed model. We conducted experimental verification on the self-made dataset and the MCindoor20000 dataset [37], respectively. The prediction time of the OD model slightly increases after adding the preprocessing algorithm in this paper, but it is still within an acceptable threshold. Compared to the increase in recognition performance, the increased time cost can be ignored. It can be concluded that the preprocessing algorithm in this paper significantly improves the recognition accuracy and ability of each OD model.

Combine different algorithms with the robotic arm to calculate the error between each recognition positioning coordinate and the actual position. Table 4 calculates a total of 100 experiments for each method and calculates the average positioning error. It can be seen that the preprocessing algorithm proposed in this article meets the needs of actual production and can greatly improve the probability of successful power connection.

#### *4.4.2 Comparison with the SOTA anti-fisheye model*

This paper selects the SIR model and FE-GAN model, two SOTA fisheye correction models, for comparison with the AFE-FE model, to test the fisheye correction effect of the

**Table 7** Proportions of successful power connections

| | Proportion/% |
|---|---|
| YOLODrone | 94 |
| YOLODrone+SIR | 95 |
| YOLODrone+FE-GAN | 95 |
| YOLODrone+AFE-FE (ours) | 99 |

model in this paper. As can be clearly seen from Table 5, compared to other SOTA models, the AFE-FE model in this paper achieves higher scores on all evaluation indicators. This indicates that the images corrected for fisheye effects using the AFE-FE model achieve the best results in terms of quality and distortion.

The SIR model, FE-GAN model, and AFE-FE model are combined with the YOLODrone OD model for a comparison in terms of detection accuracy. As shown in Table 6, the fisheye correction model in this paper has a clear advantage in detection accuracy compared to other models. The aforementioned different algorithms are combined with a robotic arm. If the robotic arm cannot completely fit the electrical connection device, it is judged as a failure; if it can completely fit, it is judged as a success. Table 7 calculates the proportion of successful connections in a total of 100 experiments for each method. It can be seen that the preprocessing algorithm proposed in this paper greatly increases the probability of successful electrical connection compared to other fisheye correction algorithms.

In summary, the model proposed in this paper is superior to the SIR model and the FE-GAN model in terms of fisheye correction effects, and the effectiveness of this algorithm in OD has been experimentally validated.

#### *4.4.3 Comparison with the SOTA feature enhancement model*

We compared several SOTA feature enhancement models, including AFAM model, SGUIENet model, MEGF model, and FDENet model, with AFE-FE model to test their feature enhancement effect. As can be seen from Table 8, compared to other SOTA models, the AFE-FE model in this paper achieves higher scores on all evaluation indicators. This indicates that the images enhanced with features using the AFE-FE model achieve the best results in terms of quality and distortion. The AFAM model, SGUIENet model, and AFE-FE model are combined with the YOLODrone OD model for a comparison in terms of detection accuracy. As shown in Table 9, compared to other models, the detection accuracy of the model in this paper has a clear advantage over other feature enhancement models.

In summary, the model proposed in this paper is superior to the AFAM model and the SGUIENet model in terms of feature enhancement effects, and the effectiveness of this algorithm in OD has been experimentally validated.

**Table 8** Comparison of $P_{SNR}$ and $S_{SIM}$ of different models

| | $P_{SNR}$/dB | $S_{SIM}$ |
|---|---|---|
| AFAM | 27.12 | 0.76 |
| SGUIENet | 26.42 | 0.74 |
| MEGF [38] | 25.94 | 0.78 |
| FDENet [39] | 27.35 | 0.81 |
| AFE-FE (ours) | 31.42 | 0.97 |

**Table 9** Comparison of precisions of different models

| | $P_{re}$/% |
|---|---|
| YOLODrone | 87.354 |
| YOLODrone+ AFAM | 91.445 |
| YOLODrone+ SGUIENet | 87.548 |
| YOLODrone+ MEGF | 91.132 |
| YOLODrone+ FDENet | 89.985 |
| YOLODrone+AFE-FE (ours) | 96.135 |

#### *4.4.4 Practical application effect*

We obtained production capacity data from the internal production records and annual reports of our partner factory. We analyzed the average monthly production capacity changes from June 2022 to June 2023, ensuring that the production environment and production demand remained constant during this period. Table 10 calculates the production capacity and waste volume before and after the combination of the FE-GAN model with SOTA models applied to the steel rolling process. It can be seen that after using the method proposed in this paper, the monthly production capacity has been significantly improved while reducing the amount of smelting waste. This indicates that the method proposed in this paper can improve smelting efficiency and reduce smelting costs and waste.

Upon analysis, it can be found that the AFE-FE model proposed in this paper has obvious advantages in the detection of torpedo can electrical connection devices, which are shown as follows.

(i) The AFE-FE model is robust, can be combined with different SOTA OD models, and enhances the performance of the target monitoring model.
(ii) The AFE-FE model has excellent fisheye effect correction capabilities, ensuring the positioning accuracy of the electrical connection device to the greatest extent.

**Table 10** Comparison of practical application effects

| | Monthly production capacity/t | Wastage/t |
|---|---|---|
| None | 6 421 | 341 |
| YOLODrone | 6 802 | 338 |
| YOLODrone+AFE-FE(ours) | 8 462 | 304 |

(iii) The AFE-FE model enhances the blurred target features in dim environments, improving the detection accuracy of the electrical connection device.

## 5 Conclusions and future work

Addressing the challenge of recognizing and locating torpedo can electrical connection devices, we propose an AFE-FE model preprocessing algorithm, comprising an anti-fisheye module to correct fisheye camera distortions, and a feature enhancement module to enhance blurred features under low illumination. Combined with SOTA OD models, the AFE-FE model performs well in detecting these devices. An image database of these devices is established, and after multiple training iterations, optimal hyperparameters are fixed based on validation set results. Evaluation indicators are calculated for both training and test sets to assess image quality and detection performance. Finally, real-time detection is performed on on-site video footage, with target confidence and location marked on the image.

This paper experimentally demonstrates the effectiveness of a proposed algorithm for positioning torpedo can electrical connection devices. Traditional manual operations are time-consuming, labor-intensive, and pose safety risks. The proposed algorithm, combined with a robotic arm, accelerates the connection process, ensures safety, and improves efficiency and cost-effectiveness. After implementing this preprocessing algorithm, the monthly production capacity in 2023 rose significantly to 8 462 t, a 31.8% increase from 2022's 6 421 t. Material waste decreased by 10.9%, from 341 t to 304 t, enhancing smelting efficiency and reducing costs and waste.

The preprocessing algorithm in this paper can improve the overall efficiency and cost-effectiveness of the steel manufacturing industry. The algorithm has a certain robustness and can be applied to other scenes that need fisheye correction and feature enhancement, and it has good application prospects. However, the algorithm still has some shortcomings. It is necessary to customize specific model design indicators according to different scenarios. In the future, more factors need to be considered in the algorithm to enhance the usability and robustness of the algorithm, such as processor computing power, related shooting equipment, more target features that need to be positioned, etc. When integrating the algorithm into the existing steel smelting production line, the scalability of the algorithm and the cost of maintenance and debugging also need to be considered.

**Funding** Open access funding provided by The Hong Kong Polytechnic University.



## References

1. Conrad LF, Oliver ML, Jack RJ et al (2021) Quantification of 6-degree-of-freedom chassis whole-body vibration in mobile heavy vehicles used in the steel making industry. J Low Freq Noise Vib Active Control 31(2):85–104. https://doi.org/10.1260/0263-0923.31.2.85
2. Fu T, Li P, Liu S (2024) An imbalanced small sample slab defect recognition method based on image generation. J Manuf Process 118:376–388. https://doi.org/10.1016/j.jmapro.2024.03.028
3. Fu T, Liu S, Li P (2024) Digital twin-driven smelting process management method for converter steelmaking. J Intell Manuf 36:2749–2765. https://doi.org/10.1007/s10845-024-02366-7
4. Fu T, Li P, Liu S (2025) A method for quality inspection of continuous casting billet based on infrared flaw detection and 2-D image. IEEE Trans Instrum Meas 74:1–14. https://doi.org/10.1109/TIM.2024.3502874
5. Gao X, Zhang R, You Z et al (2022) Use of hydrogen–rich gas in blast furnace ironmaking of V–bearing titanomagnetite: mass and energy balance calculations. Materials 15(17):6078. https://doi.org/10.3390/ma15176078
6. Semenov YS, Shumel'chik EL, Gorupakha VV et al (2017) Monitoring blast furnace lining condition during five years of operation. Metallurgist 61(34):291–297. https://doi.org/10.1007/s11015-017-0491-z
7. Ni J, Shen K, Chen Y et al (2023) An improved SSD-like deep network-based object detection method for indoor scenes. IEEE Trans Instrum Meas 72:1–15. https://doi.org/10.1109/TIM.2023.3244819
8. Guo J, Chen H, Liu B et al (2023) A system and method for person identification and positioning incorporating object edge detection and scale-invariant feature transformation. Measurement 223:113759. https://doi.org/10.1016/j.measurement.2023.113759
9. Zhai D, Zhang X, Li X et al (2023) Object detection methods on compressed domain videos: an overview, comparative analysis, and new directions. Measurement 207:112371. https://doi.org/10.1016/j.measurement.2022.112371
10. Sun Y, Song K, Zhou T et al (2023) A shared method of metal object detection and living object detection based on the quality factor of detection coils for electric vehicle wireless charging. IEEE Trans Instrum Meas 72:1–17. https://doi.org/10.1109/TIM.2023.3277132

11. Zou Y, Liu C (2023) A light-weight object detection method based on knowledge distillation and model pruning for seam tracking system. Measurement 220:113438. https://doi.org/10.1016/j.measurement.2023.113438
12. Xie Q, Li D, Yu Z et al (2020) Detecting trees in street images via deep learning with attention module. IEEE Trans Instrum Meas 69(8):5395–5406. https://doi.org/10.1109/TIM.2019.2958580
13. Dong Z, Liu Y, Feng Y et al (2022) Object detection method for high resolution remote sensing imagery based on convolutional neural networks with optimal object anchor scales. Int J Remote Sens 43(7):2698–2719. https://doi.org/10.1080/01431161.2022.2066487
14. Ren J, Hu X, Zhu L et al (2023) Deep texture-aware features for camouflaged object detection. IEEE Trans Circuits Syst Video Technol 33(3):1157–1167. https://doi.org/10.1109/TCSVT.2021.3126591
15. Liu Z, Zhang Z, Tan Y et al (2022) Boosting camouflaged object detection with dual-task interactive transformer. Proceed - Int Confer Pattern Recog 2022:140–146. https://doi.org/10.1109/ICPR56361.2022.9956724
16. Xue Z, Xue N, Xia GS et al (2019) Learning to calibrate straight lines for fisheye image rectification. Proc IEEE/CVF Conf Comput Vis Pattern Recognit (CVPR) 2019:1643–1651
17. Fan J, Zhang J, Tao D (2020) SIR: self-supervised image rectification via seeing the same scene from multiple different lenses. IEEE Trans Image Process 32:865–877. https://doi.org/10.1109/TIP.2022.3231087
18. Chao C, Hsu P, Lee H et al (2020) Self-supervised deep learning for fisheye image rectification. In: Proc 2020 IEEE international conference on acoustics, speech and signal processing (ICASSP) 2020: 2248−2252 https://doi.org/10.1109/icassp40776.2020.9054191
19. Li T, Tong G, Tang H et al (2020) FisheyeDet: a self-study and contour-based object detector in fisheye images. IEEE Access 8:71739–71751. https://doi.org/10.1109/ACCESS.2020.2987868
20. Yang CY, Chen HH (2021) Efficient face detection in the fisheye image domain. IEEE Trans Image Process 30:5641–5651. https://doi.org/10.1109/TIP.2021.3087400
21. Wang W, Wang H, Chen L et al (2023) A novel soft sensor method based on stacked fusion autoencoder with feature enhancement for industrial application. Measure J Int Measure Confeder 221:113491. https://doi.org/10.1016/j.measurement.2023.113491
22. Li B, Wang T, Zhai Y et al (2022) RFIENet: RGB-thermal feature interactive enhancement network for semantic segmentation of insulator in backlight scenes. Measure J Int Measure Confeder 205:112177. https://doi.org/10.1016/j.measurement.2022.112177
23. Nirmala DM, Sankar V (2023) Graph based heterogeneous feature extraction for enhanced hardware trojan detection at gate-level using optimized XGBoost algorithm. Measure J Int Measure Confeder 220:113320. https://doi.org/10.1016/j.measurement.2023.113320
24. Wang J, Yang J, Lu G et al (2023) Adaptively fused attention module for the fabric defect detection. Adv Intell Syst 5(2):2200151. https://doi.org/10.1002/aisy.202200151
25. Qi Q, Li K, Zheng H et al (2022) SGUIE-Net: semantic attention guided underwater image enhancement with multi-scale perception. IEEE Trans Image Process 31:6816–6830. https://doi.org/10.1109/TIP.2022.3216208
26. Guo T, Wei Y, Shao H et al (2021) Research on underwater target detection method based on improved MSRCP and YOLOv3. In: Proc 2021 IEEE international conference on mechatronics and automation, ICMA 2021: 1158–1163. https://doi.org/10.1109/ICMA52036.2021.9512827
27. Hu X, Liu H, Hui Y et al (2022) Transformer feature enhancement network with template update for object tracking. Sensors 22(14):5219. https://doi.org/10.3390/s22145219
28. Wang CY, Liao HYM, Wu YH et al (2020) CSPNet: a new backbone that can enhance learning capability of CNN. Proc IEEE Conf Comput Vis Pattern Recognit Workshop (CVPR Workshop) 2020:390–391
29. Fu H, Gong M, Wang C et al (2019) Geometry-consistent generative adversarial networks for one-sided unsupervised domain mapping. In: IEEE conference on computer vision and pattern recognition (CVPR) 2422−2431. https://doi.org/10.1109/CVPR.2019.00253.
30. Zang X, Li G, Gao W (2022) Multidirection and multiscale pyramid in transformer for video-based pedestrian retrieval. IEEE Trans Industr Inf 18(12):8776–8785. https://doi.org/10.1109/TII.2022.3151766
31. Koyun OC, Keser RK, Akkaya İB et al (2022) Focus-and-detect: a small object detection framework for aerial images. Signal Process Image Commun 104:116675. https://doi.org/10.1016/j.image.2022.116675
32. Qi S, Du J, Wu M et al (2022) Underwater small target detection based on deformable convolutional pyramid. In: IEEE international conference on acoustics, speech and signal processing (ICASSP), Singapore, 2022, pp 2784−2788 https://doi.org/10.1109/ICASSP43922.2022.9746575
33. Sahin O, Ozer S (2021) YOLODrone: improved YOLO architecture for object detection in drone images. In: 2021 44th international conference on telecommunications and signal processing (TSP), pp 361–365. https://doi.org/10.1109/TSP52935.2021.9522653
34. Liao X, Lv S, Li D et al (2021) Yolov4-mn3 for PCB surface defect detection. Appl Sci 11(24):11701. https://doi.org/10.3390/app112411701
35. Ren S, He K, Girshick R et al (2015) Faster R-CNN: towards real-time object detection with region proposal networks. arXiv preprint arXiv:1506.01497
36. Wang Z, Bovik AC, Sheikh HR et al (2004) Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4):600–612. https://doi.org/10.1109/TIP.2003.819861
37. Bashiri FS, LaRose ER, Peissig PL et al (2018) Mcindoor20000: a fully-labeled image dataset to advance indoor objects detection. Data Brief 17:71–75. https://doi.org/10.1016/j.dib.2017.12.047
38. Jin H, Li L, Su H et al (2024) Learn to enhance the low-light image via a multi-exposure generation and fusion method. J Vis Commun Image Represent 100:104127. https://doi.org/10.1016/j.jvcir.2024.104127
39. Gao F, Li L, Wang J et al (2023) A lightweight feature distillation and enhancement network for super-resolution remote sensing images. Sensors 23(8):3906. https://doi.org/10.3390/s23083906




**Tian-Jie Fu** is a doctoral student at School of Mechanical Engineering, Zhejiang University. She obtained master's degree from Zhejiang University in 2019. She joined School of Mechanical Engineering at Zhejiang University in 2021. Her main research focuses on machine vision in the steel manufacturing process and automatic control technology in manufacturing, as well as digital twin technology. She has published five research papers in internationally renowned journals and conference proceedings.

**Shi-Min Liu** is a postdoctoral fellow in Department of Industrial and Systems Engineering at The Hong Kong Polytechnic University in Hongkong, China. He received his Ph.D. degree in Mechanical Engineering at Donghua University, China, in 2022. His research interests include digital twin, intelligent manufacturing system, intelligent machining, system adaptability, bionic manufacturing, etc.

**Pei-Yu Li** graduated from Department of Mechanical Engineering at Zhejiang University in 1988 with bachelor's degree. He obtained master's degree in Mechanical Manufacturing from Zhejiang University in 1991, Ph.D. degree in Mechanical Manufacturing in 1995. In 1997, he was appointed as an associate professor. His research focuses on structural dynamics, mechatronics integration technology, embedded technology, etc.

**Ruo-Xin Wang** received bachelor's degree in Mechanical Design, Manufacture and Automation from University of Shanghai for Science and Technology, Shanghai, in 2015. She received master's degree in Mechanical Engineering from Southwest Jiaotong University, Chengdu, in 2018. She received Ph.D. degree from The Hong Kong Polytechnic University in 2023. Her current research interests include machine learning-based surface characterization, smart measurement, and knowledge graph.