An Improved Faster Region-based Convolutional Neural Network Algorithm for Identification of Steel Coil End-head

Jian-Zhou Pan,[1,2,3] Chi-Hsin Yang,[2*] Long Wu,[2]
Wen-Hu Tang,[4] and Kung-Chieh Wang[2]

[1]School of Material Science and Engineering, University of Science and Technology,
Beijing, Beijing 100083, China
[2]School of Mechanical and Electrical Engineering, Sanming University,
Sanming, Fujian 365004, China
[3]Fujian Sansteel (Group) Co., Ltd., Sanming, Fujian 365000, China
[4]School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou, Fujian 350108, China

A method that uses machine vision and machine learning technologies to identify the end-head in a steel coil has seldom been proposed. In this study, an improved faster region-based convolutional neural network (F-RCNN) deep learning algorithm is introduced to identify the position of the steel coil end-head for a hardware system set up for image sensing and detection. The feature pyramid network (FPN) and the parallel attention module (PAM), which are both involved in the traditional F-RCNN, are used to increase the detection accuracy. Our experimental results validated the effectiveness of the proposed improved algorithm.

1. Introduction

Because of the importance of steels coils as both products and raw materials in construction and industrial production, the demand for steel coils has recently been increasing. In the production of a steel coil, including the manufacturing processes of heating, rolling, spinning, cooling, and other processes,[1] the internal structural composition of steel coils is not uniform, causing the tensile strength, yield strength, and other mechanical properties of several turns of coil near the steel coil end-head to be unstable. Several turns of coil must then be cut and recycled in accordance with the production regulations of the factory.

The drawbacks of the current production methodology for steel coils are summarized below.[1]

(1) The task of finding and cutting off the steel coil end-head is completed manually. The overall efficiency of production is affected by increases in labor costs.
(2) The working environment of a steel-coil factory involves high temperature and high risk. It is not suitable for the worker to continuously operate the task in (1) over a long duration.

---

(3) Because of (1), the automation of the production process for steel coils is difficult to realize. According to the above analysis, the most critical problem in the production of steel coils is the intelligent identification and positioning of the steel coil end-head.

In current industrial applications, the detection and recognition of objects are achieved by using images of the objects, which are taken by industrial cameras, along with the appropriate object recognition algorithms in machine vision. Object recognition algorithms are basically divided into traditional and deep-learning-based methods.[2] Traditional object recognition algorithms mainly include template or feature matching. The main idea is that some feature points of the detected target are extracted by the designed algorithm, and then, similarity measurement is used to match feature points between target images.[3] With the rapid development of computer network algorithms in recent years, deep learning object detection algorithms have made great progress. Compared with the traditional object recognition algorithms, the detection accuracy and robustness of the deep learning object detection algorithms are improved. They are widely applied in social distance monitoring in response to the COVID-19 pandemic,[4] flight scene analysis,[5] autonomous driving,[6] environmental monitoring,[7] and other fields.

In the literature, the deep learning object detection algorithms are divided into the single-stage and two-stage detection methods. The single-stage detection method includes the you-only-look-once (YOLO) series,[8–11] a single-shot multi-box detector (SSD),[12,13] and MobileNet.[14] The algorithm directly performs convolutional network operations based on a whole image. Then, the category of the detected object and the position of the surrounding border are predicted. However, the single-stage detection method easily enables false and missing detection in small-target detection. The two-stage detection method mainly includes the fast region-based convolutional neural network (RCNN)[15] and improved faster RCNN (F-RCNN).[16,17] The traditional F-RCNN introduces a regional suggestion network to classify and locate the target twice to improve the detection accuracy. In addition, to improve the detection accuracy of the algorithm in small-target detection, many improvements were proposed. The combination of the SSD algorithm with the feature pyramid improved the detection ability of small objects.[18,19] Huang *et al.* modified the network structure of the traditional F-RCNN by adding both low-level and high-level features for target detection.[20]

In previous works, applications of the deep learning object detection algorithms in the field of steel materials were basically for detecting defects on the surfaces of steel products,[21,22] irregular structure weld defects,[23] and weld missing defects.[24] However, the identification of the steel coil end-head by applying the deep learning object detection algorithms is seldom proposed in current studies of the steel industry. The position of the steel coil end-head within the whole bundle of coils is unknown in advance and the distinction is not obvious. The identification of the steel coil end-head is a problem of small-target detection.[20] Motivated by the desire to mitigate the aforementioned drawbacks of the current steel-coil production process, in this study, we aim to present the improved F-RCNN-based algorithm, which is a machine learning technology used in the image detection hardware system of steel coils, for the identification of the steel coil end-head. Our algorithm is based on the traditional F-RCNN, and the feature pyramid network (FPN) is added to the network model to detect fusion features, and

the parallel attention module (PAM) is used to improve the quality of fusion features and the quality of preselected frames generated by the regional suggestion network, so as to improve the detection accuracy of the algorithm.

The rest of the paper is organized as follows. In Sect. 2, the network structure of the traditional F-RCNN is briefly introduced and the problems in the identification of the steel coil end-head using the traditional F-RCNN are described. The main idea of the improved F-RCNN-based algorithm is given and the model training and analysis are described in Sect. 3. In Sect. 4, the details of the identification of the steel coil end-head, including the setup of the steel coil end-head detection hardware system, the augmentation of captured images, the model training of the proposed improved F-RCNN, and the detected results and discussion, are provided. Finally, our conclusions are presented in Sect. 5.

2. Network Structure of Traditional F-RCNN

2.1 Basic principle of F-RCNN

The traditional F-RCNN is divided into four parts, as shown in Fig. 1. The functions of each part are summarized below.[15,16,25]

(1) Feature extraction network

The feature map is extracted from the input image by the feature extraction network, which is constructed from the selected convolutional neural network. The extracted feature map is shared with the regional proposal network (RPN) and the region of interest (RoI) pooling layer. Because only a small number of pixels are involved in small targets, the feature extraction network is built by applying ResNet50[13] to generate a better feature map in this study. After the input image is processed by ResNet50, the feature map with 512 channels is obtained.

(2) RPN

The RPN generates anchor boxes of different sizes in accordance with the set values of parameters in the network. Furthermore, it also determines whether each anchor box belongs
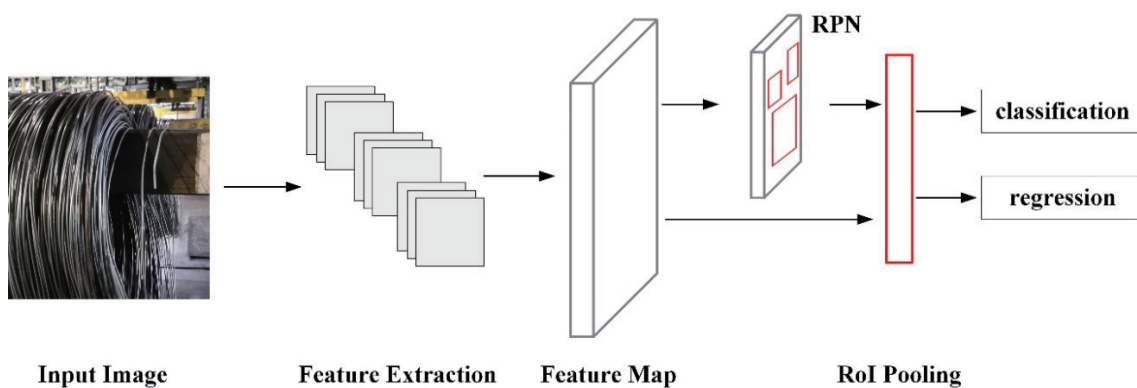

Fig. 1.   (Color online) Structure of traditional F-RCNN.

to the foreground or background. Then, all anchor boxes are divided into positive and negative samples. Finally, the appropriate candidate frame is obtained by correcting the offset of the anchor box by regression.

(3) RoI pooling layer

The target candidate regions are extracted by feeding the feature map in (1) and the anchor boxes from (2) into the RoI pooling layer.

(4) Classification and regression layers

The obtained target candidate regions are further classified by the classification and regression layers. The final detected images, where the position of the steel coil end-head is accurately detected, are the outputs of the layers.

The model optimization of the F-RCNN is accomplished by iteratively updating the network parameters to minimize the loss function. The loss function for optimizing the model of the F-RCNN is defined as

$$L\left(\{p_i\},\{t_i\}\right) = \frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i, p_i^*\right) + \lambda \frac{1}{N_{reg}}\sum_i p_i^* L_{reg}\left(t_i - t_i^*\right), \tag{1}$$

where λ is the weighting balance value, i is the anchor index, p_i is the probability that the anchor is predicted as the object, p_i^* is the category label, t_i is the predicted value of position regression, and t_i^* is the value of the label of position regression.

The category labels $p_i^* = 1$ for the positive sample and $p_i^* = 0$ for the negative sample, where the sample is classified as a positive sample when the value of the intersection over union (IoU) is greater than 0.7. Otherwise, the sample is classified as a negative sample when IoU is lower than 0.3. N_{cls} and N_{reg} are the normalized loss values of classification and regression, respectively, where the index *cls* is the size of the batch and the index *reg* is the number of anchor locations. L_{cls}, the loss function of cross entropy for classification, is defined as

$$L_{cls}\left(p_i, p_i^*\right) = -\log\left[p_i^* p_i + \left(1 - p_i^*\right)\left(1 - p_i\right)\right]. \tag{2}$$

$L_{reg}(x)$ is the positional regression function, which is defined as

$$L_{reg}\left(x\right) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5. & \text{otherwise} \end{cases} \tag{3}$$

2.2 Problems in identification of steel coil end-head by F-RCNN

Because the position of the steel coil end-head cannot be known in advance, the images captured by the cameras must cover the whole bundle of steel coil such that a captured image can include all possible locations of the steel coil end-head. Generally, the steel coil end-head occupies a very small pixel area in the captured image. Then, the identification problem of the steel coil end-head is a problem of small target detection.[26]

When using the traditional F-RCNN to identify the position of the steel coil end-head in a captured image, the following drawbacks are found:

(1) The feature map from the last layer of the feature extraction network is only detected in the traditional F-RCNN. With the increases in the number of feature extraction network layers and down-sampling and pooling operation times, the feature map contains less detailed information about the steel coil end-head in the last layer of the feature extraction network. This is the cause of the poor detection and recognition results when directly applying the traditional F-RCNN for the identification.

(2) When the pixels of the steel coil end-head are similar to the environmental pixels in the captured image, fewer distinguishing features are extracted by the traditional F-RCNN, leading to a greatly reduced detection resolution. When the traditional F-RCNN is applied to determine the location of the end-head from captured images with a large difference between pixels of the steel coil end-head and the environment, the detection resolution is high, as shown in Figs. 2(a) and 2(b). However, for the captured image with very similar pixels between the steel coil end-head and the environment, as shown in Fig. 2(c), the detection resolution is poor.

3. Improved F-RCNN-based Detection Algorithm

3.1 Improvement of traditional F-RCNN

To overcome the aforementioned drawbacks of underlying information loss during down-sampling when applying the traditional F-RCNN in the identification of the steel coil end-head, the improved F-RCNN-based detection algorithm is introduced in the study. The structure of the improved F-RCNN is shown in Fig. 3. The main difference between the traditional F-RCNN and the proposed improved F-RCNN is the inclusion of the FPN and PAM in the latter.

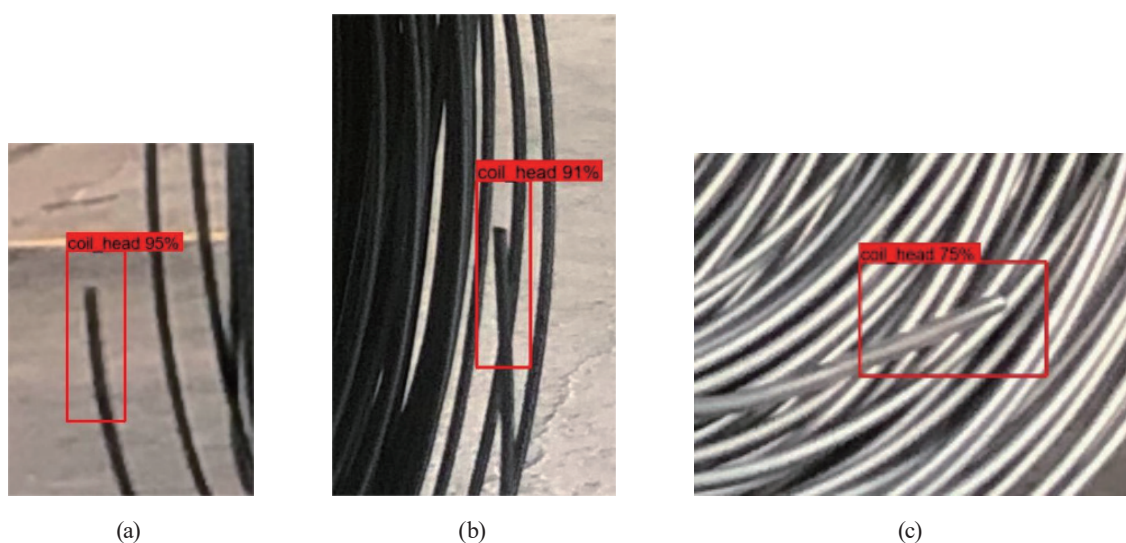| (a) | (b) | (c) |

Fig. 2.   (Color online) Results of steel coil end-head detection by traditional F-RCNN when pixels of the steel coil end-head and environment are (a), (b) different and (c) similar.
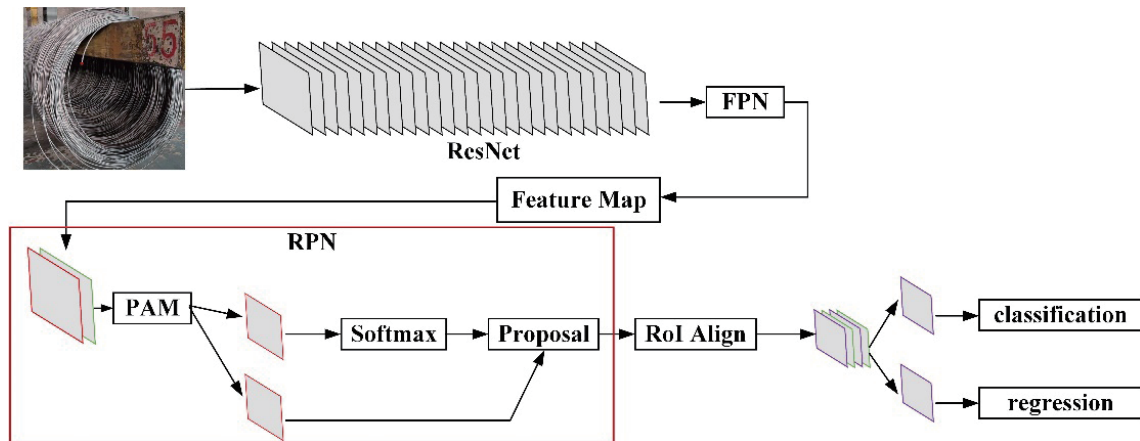
Fig. 3.   (Color online) Structure of the improved F-RCNN model.

The purpose of adding the FPN[27,28] into the structure of the improved F-RCNN is to increase the detection and recognition accuracy of small targets by detecting multiple fused feature maps. The purpose of adding the PAM into the structure of the improved F-RCNN is to strengthen the feature extraction of the target region. In Fig. 3, the PAM is added after the first convolution layer of the RPN to obtain a high-quality preselection frame, which improves the detection and positioning accuracy of the network. In the RoI Align network, the fusion feature and the preselected frame are pooled to obtain a 7×7 feature map. Finally, the classification and regression layers are used for identification and border positioning.

3.2   FPN-based feature extraction network

When the number of layers of the traditional F-RCNN is increased, the size of feature pixels for the small target gradually decreases during feature extraction. This further reduces the detection accuracy of the final feature map. The function of the FPN in the improved F-RCNN is to extract and fuse the features of the different layers in the feature extraction network so as to improve the detection performance of small target objects. The schematic diagram of the FPN is shown in Fig. 4. It consists of a bottom-up model (left side), a top-down model (right side), and lateral connections.

The structure of the FPN-based feature extraction network based on ResNet50[13] is shown in Fig. 5. The functions of the FPN are applied in the proposed improved F-RCNN to obtain bottom-up semantic features through horizontal connections of the features of five convolution layers in the ResNet50 network. After the double up-sampling, the upper layer features are fused with the shallow layer features, and five layers of fusion features are obtained finally. The feature map is converted by the last 1×1 convolution layer into the 256-dimensional feature map. In addition, the 5-layer fusion features are fed into the RPN. Then, the detection accuracy of the proposed improved F-RCNN for small targets, such as the steel coil end-head, is improved by applying the rich detail information in the feature maps of different scales.
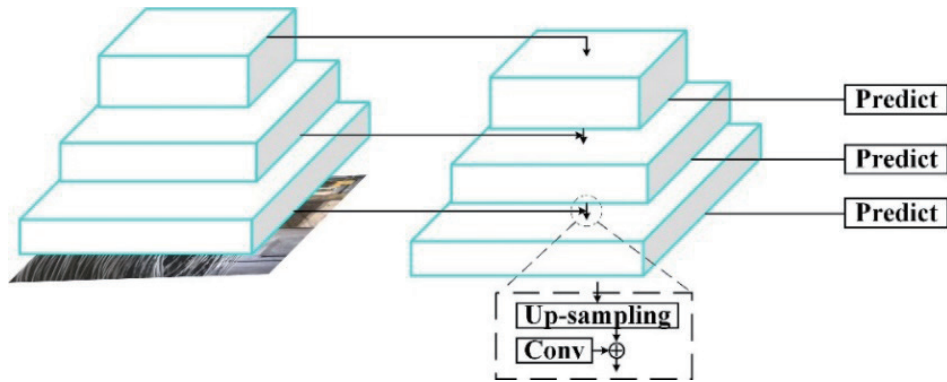
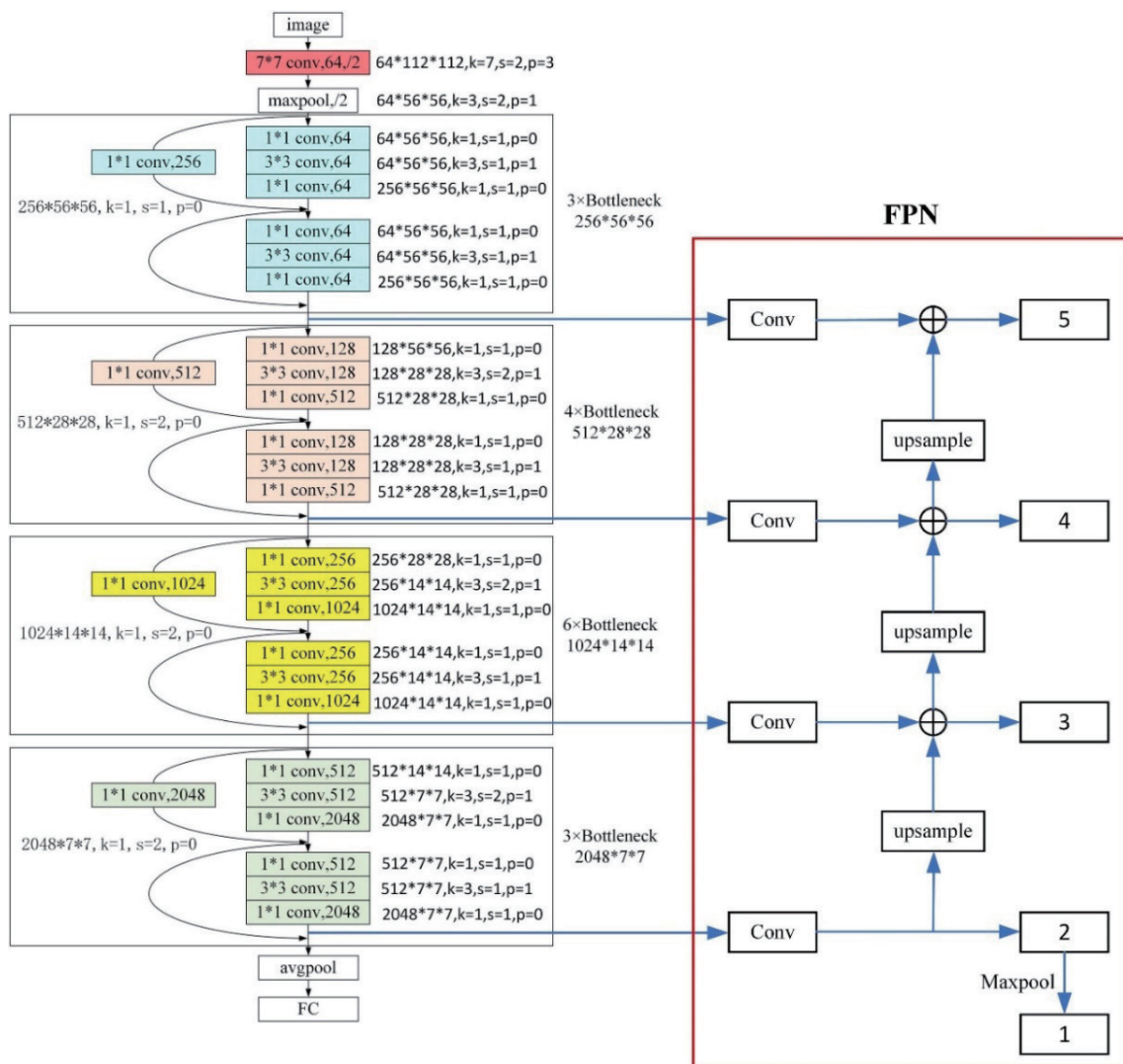Fig. 4.   (Color online) Schematic diagram of FPN.

Fig. 5.   (Color online) Structure of FPN-based feature extraction network.

3.3 PAM

To improve the quality of the generated preselection frame of feature extraction for the area of the steel coil end-head, the PAM is added to the proposed improved F-RCNN. The traditional F-RCNN already includes the cascade attention module (CAM), in which the spatial attention (SA) and channel attention (CA) modules are connected in cascade.[29–31] The principle of the SA and CA modules is that, in accordance with the required focus, the human visual system will focus on certain parts of a specific field of vision. The corresponding algorithms were developed. In previous studies,[29–31] the CAM is applied to improve the effectiveness of the F-RCNN. In this study, the PAM, where the SA and CA modules are parallelly placed, is added into the proposed improved F-RCNN to increase the detection accuracy for the steel coil end-head.

The structure of the PAM in the proposed improved F-RCNN is shown in Fig. 6. The CA, SA, and input feature are arranged in parallel. In the PAM, the input feature F is fed to the CA and SA after convolution processing. Then, the outputs of the CA and SA are spliced with the original input feature by adjusting the number of feature channels to output the refined feature.

The function of the CA module is to obtain the internal relationships between the feature channels. The operation of the CA module is depicted in Fig. 7. Firstly, global average pooling, $AvgPool(F)$, and maximum pooling, $MaxPool(F)$,[32] are used to obtain the feature spatial information. Then, the multilayer perceptron (MLP) is applied to process the two results of global average pooling and maximum pooling. The output feature vector M_c is performed with pixel-by-pixel aggregation. Furthermore, each channel weight of the input feature is obtained
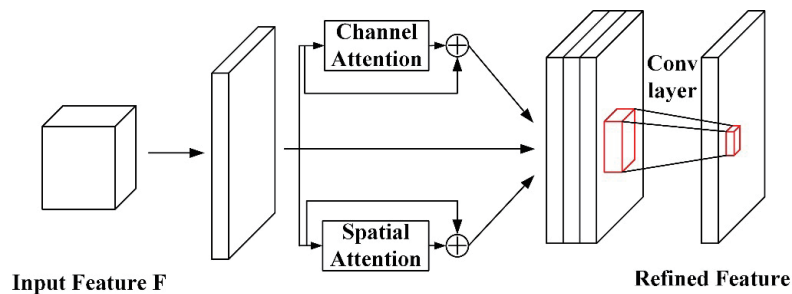

Fig. 6.   (Color online) PAM in the improved F-RCNN.
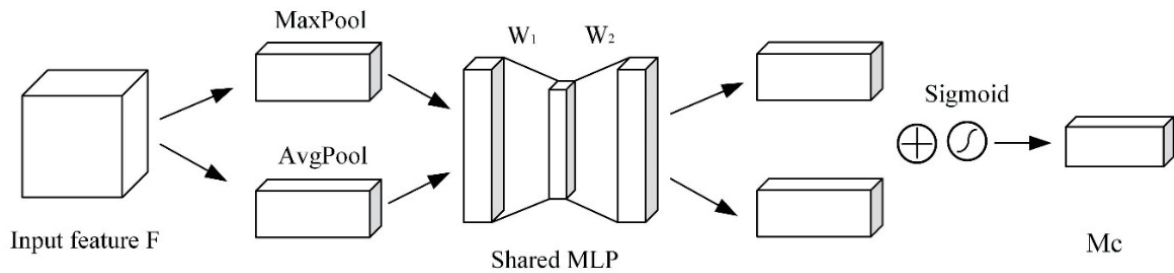

Fig. 7.   Schematic diagram of the operation of CA module.

after processing by the sigmoid activation function. Equation (4) is the evaluated formula of the operation of the CA module.

$$
\begin{aligned}
M_c &= \sigma\big(MLP\big(AvgPool(F)\big) + MLP\big(MaxPool(F)\big)\big) \\
&= \sigma\big(W_2\big(W_1\big(Avg(F)\big)\big) + W_2\big(W_1\big(Max(F)\big)\big)\big)
\end{aligned}
\tag{4}
$$

Here, σ represents the sigmoid function and F is the input feature. W_1 and W_2 are the weights connected by the ReLU activation function.[33] $Avg(F)$ and $Max(F)$ are for the averaging operation and taking the maximum of F, respectively.

The function of the SA module is to obtain the internal relationships between the feature spaces. The operation of the SA module is depicted in Fig. 8. The SA module is applied to stack the maximum value and the average value of the feature points. Then, the number of channels of the feature map is adjusted by convolution processing, which is also used to strengthen and suppress the features in different positions of the feature map. The weights of each feature point of the input steel coil end-head feature map are also obtained by processing the sigmoid activation function. Equation (5) is the evaluated formula of the operation of the SA module.

$$
M_s = \sigma\big(f^{7\times7}\big(\big[AvgPool(F); MaxPool(F)\big]\big)\big) = \sigma\big(f^{7\times7}\big(\big[Avg(F); Max(F)\big]\big)\big)
\tag{5}
$$

σ represents the sigmoid function, F is the input feature, and $f^{7\times7}$ stands for a convolution operation with a convolution kernel of 7.[34,35]

In the PAM, channel attention, spatial attention, and the input feature are arranged in parallel. Channel attention, which focuses on the features, acquires the internal information of the feature channel. Spatial attention, which focuses on the positions of feature processing, obtains the internal information of the feature space. To prevent the loss of the original input, the original input is multiplied pixel by pixel with the processing of the channel and spatial attentions. At the
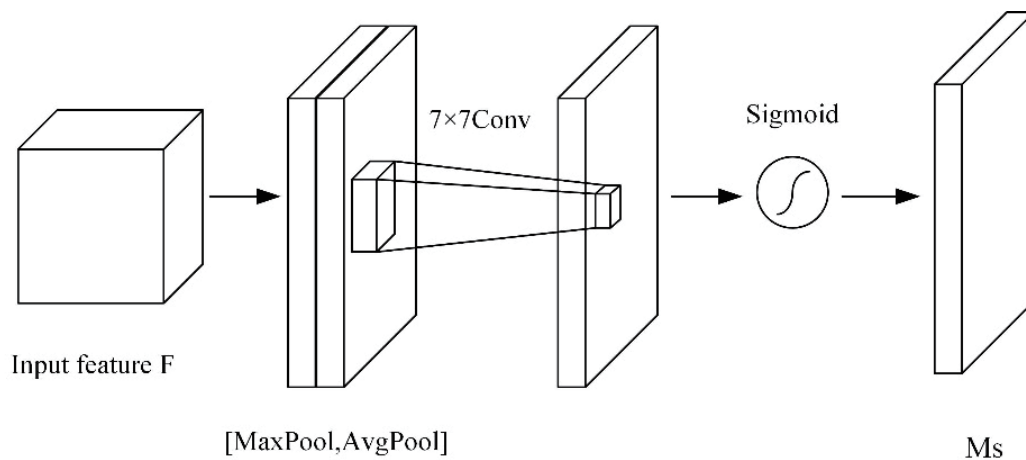

Fig. 8.   Schematic diagram of the operation of the SA module.

same time, to avoid damage of the feature structure, the processing result is stacked with the input and fed to the convolution layer to adjust the number of channels and then output. The evaluated formula of the operation of the PAM is

$$M_{PAM} = W_3 [M_c \otimes F; M_s \otimes F; F],$$
$$F = W_0(x),$$

(6)

where $\otimes$ represents multiplying the elements one by one and F is the feature after the first convolution. M_c and M_s are the operational formulas of the CA and SA modules defined by Eqs. (4) and (5), respectively. x is the model input. W_3 is the weight of the last convolution layer, and W_0 is the weight of the first convolution layer of the improved F-RCNN.

3.4 Model training and analysis of results

The experimental running environment is the Windows system with the GPU RTX3080, 10 Gb of memory, and Intel i9-10920x CPU. The GPU accelerates the use of CUDA10.0 and CUDNNv7.6.4. The framework is PaddlePaddle and the programming language Python is applied to develop the whole software program.

To verify the effectiveness of the improved F-RCNN with the added FPN and PAM in the study, the Pascal VOC2012 dataset is applied to train the proposed model. In addition, the traditional F-RCNN and the traditional F-RCNN with the FPN are also trained at the same time to compare the improvement effect of model accuracy before and after model training. The number of iterations for model training is set to 50 and the learning rate is set to 0.002. The stochastic gradient descent (SGD) optimization algorithm is used in model training. The learning rate decreases gradually, the batch size is 3, and the number of iterations for warming up is set to 10.

Pascal VOC2012 is an open dataset containing a total of 20 categories, for example, people, aircraft, and birds. In the study, in accordance with the numbers of training set:test set:verification set data strokes, the Pascal VOC2012 dataset is divided into 8:1:1. The loss functions of the improved F-RCNN model with the FPN and PAM, the F-RCNN model with the FPN, and the traditional F-RCNN model obtained by transfer learning using coco pretraining parameters with 50 iterations of training are shown in Fig. 9.

It is shown that the loss functions of the three types of F-RCNN models for each learning epoch decreased rapidly at the beginning of the training process and tended to be stable at the end. This means that model learning is effective and fit to a stable degree. Furthermore, for the proposed improved RCNN with the FPN and PAM, the loss function converges to about 0.14 after 20 iterations of training.

The three types of F-RCNN models are set with the optimal weights obtained after model training, and the optimal models are applied to the test dataset. The mean average precision (mAP) values of the three types of F-RCNN models, obtained by applying the test dataset, are listed in Table 1, where mAP$^{IoU=0.5}$, mAP$^{IoU=0.75}$, and mAP$^{IoU=0.5:0.95}$ have the selected IoU thresholds of 0.5 and 0.75, and IoU= 0.5:0.05:0.95, respectively.
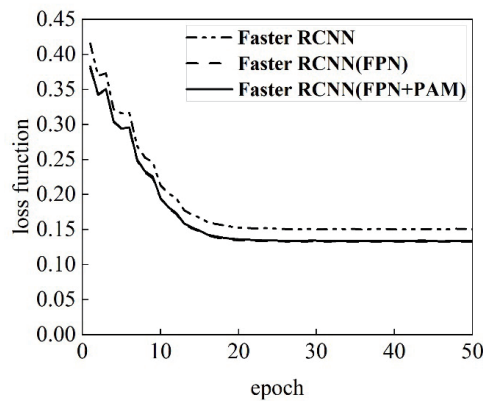
Fig. 9.   Loss functions of three types of F-RCNNs.

Table 1
mAP values of three types of F-RCNN model.

	mAP$^{IoU=0.5}$	mAP$^{IoU=0.75}$	mAP$^{IoU=0.5:0.95}$
F-RCNN	60.97	37.64	35.69
F-RCNN (FPN)	62.23	41.42	38.24
F-RCNN (FPN + PAM)	63.11	41.79	38.73

Compared with the traditional F-RCNN, mAP$^{IoU=0.5}$ improves from 60.97 to 63.11%, mAP$^{IoU=0.75}$ from 37.64 to 41.796%, and mAP$^{IoU=0.5:0.95}$ from 35.69 to 38.73% for the proposed improved F-RCNN. The addition of the FPN improves the accuracy of the traditional F-RCNN by about 3% on average and that of the PAM improves the accuracy of network detection by about 1% on average. The proposed improved F-RCNN network with different thresholds improves by about 4% on average.

4.   Identification of Steel Coil End-head

The improved F-RCNN described in the previous section is a kernel network algorithm used to identify the steel coil end-head. For the deep learning of the improved F-RCNN, the open image dataset is not yet available for the identification. To train the improved F-RCNN, the image detection hardware system is built to collect the steel coil pictures in actual production in the factory. After removing the images that are fuzzy and unintelligible, have no steel coil end-head, and do not cover the whole bundle of steel coils, 3653 image samples of steel coils are obtained. The Labeling software is used to label the steel coil end-head to mark the coil head in the image sample.

4.1   Setup of steel coil end-head detection hardware system

In the study, a schematic diagram of the steel coil end-head image detection hardware system is shown in Fig. 10. The hardware of the detected system is mainly composed of a type C hook, a drum expansion mechanism, and a machine-vision system module. The newly produced steel
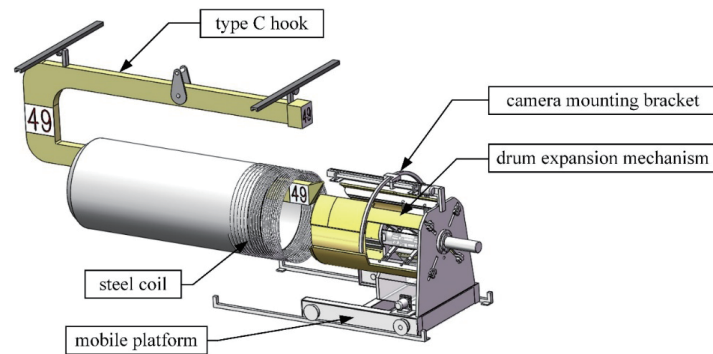
Fig. 10.   (Color online) Schematic diagram of the steel coil end-head detection system.

coil is suspended on the type C hook and moved from the coil collecting station to the front of the drum expansion mechanism.

The drum expansion mechanism is composed of a drum expansion and a mobile platform with a moving track. The machine-vision system module mainly consists of five industrial cameras and the camera mounting bracket with a moving guide rail. The camera mounting bracket with the moving guide rail is moved along the central axis of the drum expansion mechanism to scan the steel coil, and images of the steel coil end-head are captured by the five mounted industrial cameras. Then, the images are spliced and fed to the proposed improved F-RCNN to evaluate and identify the location of the steel coil end-head.

A schematic diagram of the five industrial cameras around the drum expansion mechanism is shown in Fig. 11. The origin of the coordinate system (the x–y plane) is located at the center of the drum expansion mechanism and the z-axis coincides with the central axis of the drum expansion mechanism. The coordinates (in mm) of the five cameras 1 to 5 are (887.08, −700), (935.64, 648.14), (0, 1130), (−925.64, 648.14), and (−887.08, −700), respectively, in the x–y plane.

The procedures for the identification of the steel coil end-head are summarized as follows.

(1) The steel coil suspended on the type C hook is transported from the coiling station and stops in front of the drum expansion mechanism.

(2) To scan and detect the steel coil end-head, the drum expansion mechanism is moved forward and backward by the mobile platform with a prespecified number of expanding and shrinking movements.

(3) After the scanning and detecting tasks, the images taken by the five industrial cameras, which are arranged around the circumference of the drum expansion mechanism, are spliced as the input image and fed to the proposed improved F-RCNN for the identification of the steel coil end-head.

(4) If the steel coil end-head is not identified, return to the previous step and repeat the scanning and detecting tasks. Otherwise, if the steel coil end-head is detected, three-dimensional position coordinates are obtained.
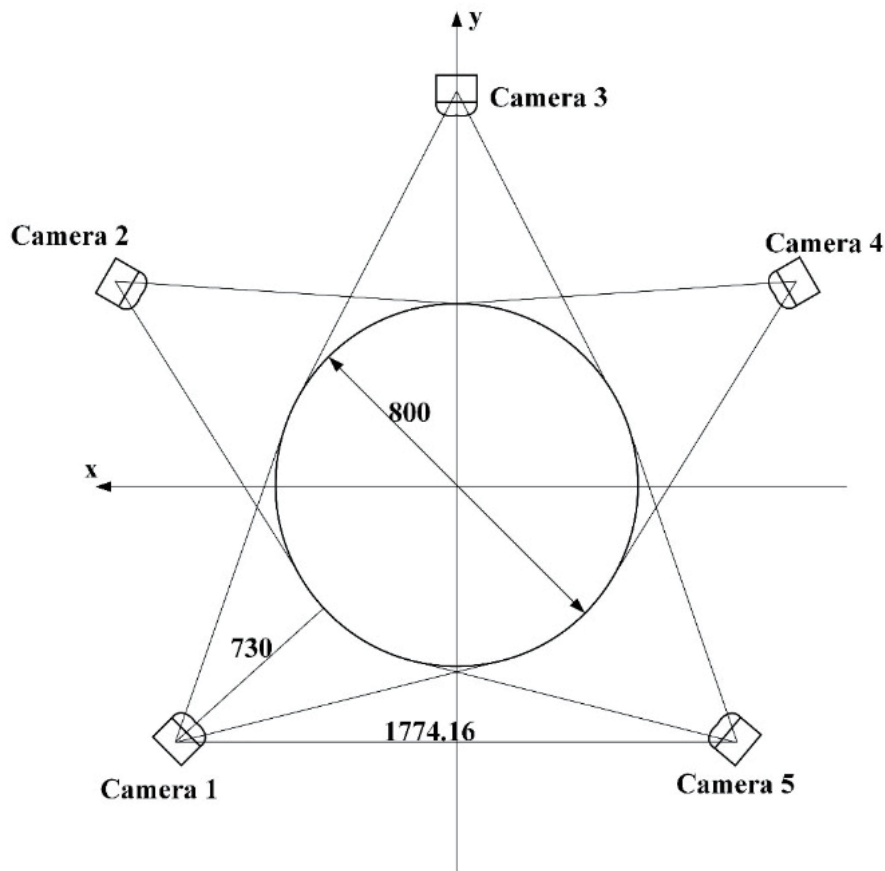
Fig. 11.   Schematic diagram of the five industrial cameras.

4.2   Augmentation of captured images of steel coil end-head

Captured images only include one steel coil end-head, and there is not enough pixel data for identification. Besides, there are only a few features that can be used to distinguish between the steel coil end-head and the coil. Thus, a poor generalization ability of the overfitting network model easily results during network training.

In the captured images, the pixels of the steel coil end-head and the edge of the steel coil clearly differ. This causes the gradient information of the edge of the steel coil to be greatly different. The gradient information refers to the pixel difference in the image. The pixel difference is calculated to scan the input image through the window function and further evaluate the convolution. The gradient information of the edge of the steel coil, named the edge gradient, is an important feature for distinguishing the steel coil end-head from the steel coil. In the study, the edge gradient is extracted and added to the original captured image, and combined with the horizontal and vertical flips of the original steel coil end-head image to achieve the augmentation of the captured images.

Figure 12 shows selected augmented captured images. The augmented images can enhance the robustness and detection performance of the trained improved F-RCNN. The steps for the argumentation of the captured image of the steel coil end-head shown in Fig. 12(a) are as follows.

(1) The gradient information in the *x*- and *y*-directions of the captured image is extracted to form the gradient maps by the Sobel operator.[36,37]

(2) OTSU threshold processing[38–40] is performed for the gradient maps in the *x*- and *y*-directions, and the pixels of the two maps are added item by item.

(3) The 3 × 3 kernel function is applied to perform the open operation processing on the gradient map to eliminate image noise. A typical result is shown in Fig. 12(b).

(4) Pixels are added item by item to the gradient map and the original image to enhance the edge features of the steel coil end-head in the original captured image.

(5) Flip the picture, as shown in Fig. 12(c) for a vertical flip and in Fig. 12(d) for a horizontal flip.

(a)
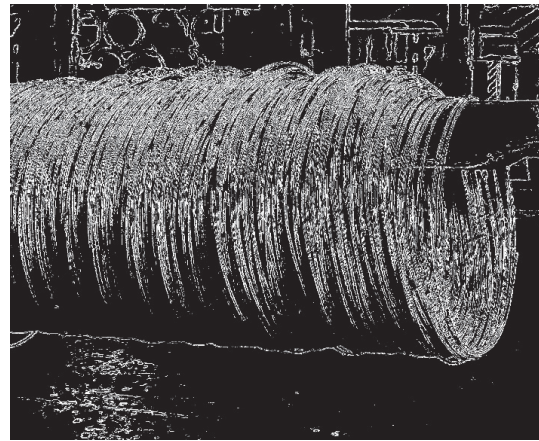
(b)

(c)

(d)

Fig. 12. (Color online) Augmentation of the captured images. (a) Original captured image. (b) Gradient image by eliminating noise (c) Vertical flip image. (d) Horizontal flip image.

4.3   Training-improved F-RCNN for steel coil end-head identification

After the aforementioned argumentation, the augmented image dataset of the steel coil end-head is divided into 8:1:1 for training set:verification set:test set. During the training of the F-RCNNs, the picture size of each steel coil image is converted to 800 × 800. Three F-RCNNs are trained in 50 iterations with coco pretraining parameters, and the values of loss function are recorded. The curves of loss functions during network training are depicted in Fig. 13. It is shown that, with the progress of training, the loss function of the improved F-RCNN for the steel coil end-head decreases continuously before 50 iterations of training. Furthermore, the improved F-RCNN tends to be stable after more than 20 iterations of training. This indicates that the detection model tends to converge.

The curves of detection accuracy in the verification set for the three F-RCNNs after model training are depicted in Fig. 14. It is shown that the improved F-RCNN has higher mAPs than the other two F-RCNNs with different thresholds in the verification set. This means that the proposed improved F-RCNN can better improve the image recognition accuracy in the
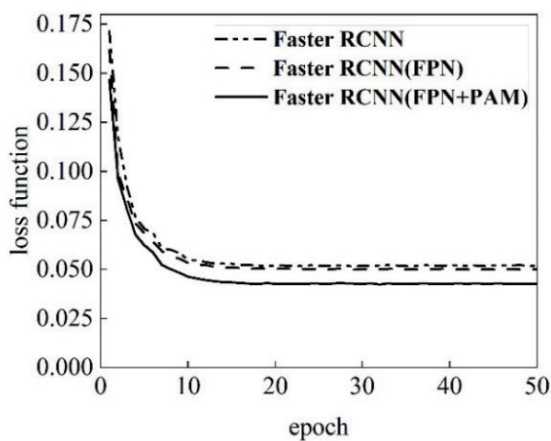

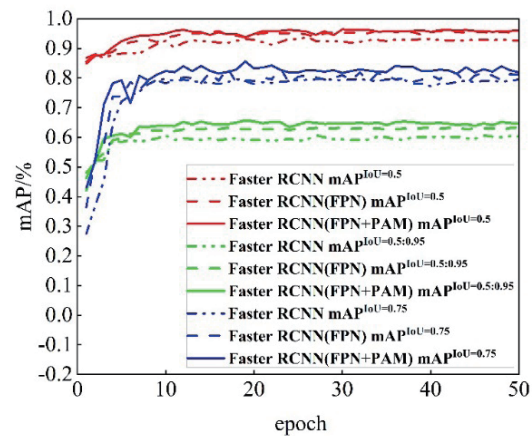Fig. 13.   Loss functions of F-RCNNs.

Fig. 14.   (Color online) Curves of detection accuracy for F-RCNNs.
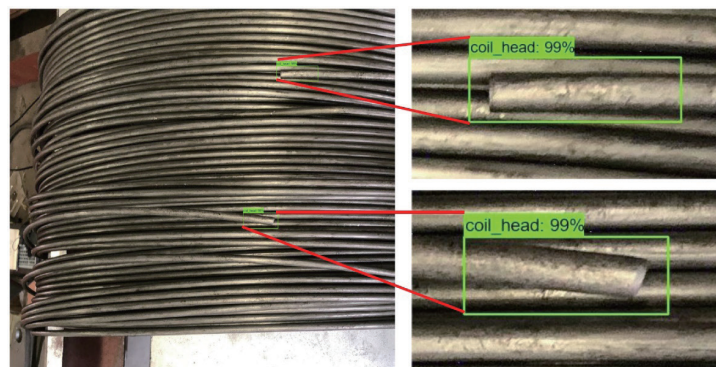

Fig. 15.   (Color online) Results of identification of steel coil end-heads by proposed improved F-RCNN.

identification of the steel coil end-head. For the proposed improved F-RCNN, the results of the identification of the steel coil end-heads in the sampled pictures are shown in Fig. 15. Compared with the detection results of the traditional F-RCNN[16,17] shown in Fig. 2(c), the proposed F-RCNN exhibits a high resolution for the recognition of the steel coil end-heads.

5. Conclusions

For the image detection hardware system of steel coils, an improved F-RCNN-based algorithm for the identification of the steel coil end-head was addressed in this study. In the traditional F-RCNN, the FPN and PAM are added to the network model to improve the detection accuracy of the algorithm for small target detection. Additionally, since the captured steel coil-end image data is not enough for model training, three types of image are processed to achieve the augmentation of captured images. That is, the image is mapped to the original captured image by adding the edge gradient, and the images are flipped to the original image by horizontal and vertical. The experimental results indicate that the proposed algorithm can further improve the image recognition accuracy in the identification of steel coil end-heads.

References

1 M. Lin: Fujian Metallurgy **5** (2021) 21. https://doi.org/10.19574/j.cnki.issn1672-7665.2021.05.006
2 S. Bandong, Y. Y. Nazaruddin, and E. Joelianto: Heliyon **9** (2023) e1321. https://doi.org/10.1016/j.heliyon.2023.e13213
3 S. Qiu, C.-H. Yang, L. Wu, K.-C. Wang, and J. Pan: Sens. Mater. **34** (2022) 789. https://doi.org/10.18494/SAM3638
4 U. Ahuja, S. Singh, M. Kumar, K. Kumar, and M. Sachdeva: Multimedia Tools Appl. **82** (2023) 7553. https://doi.org/10.1007/s11042-022-13718-x
5 J. Ni, T. Gong, Y. Gu, J. Zhu, and X. Fan: Comput. Intell. Neurosci. **2020** (2020) 7490840. https://doi.org/10.1155/2020/7490840
6 J. Ni, Y. Chen, J. Zhu, D. Ali, and W. Cao: Appl. Sci.-Basel **10** (2020) 2749. https://doi.org/10.3390/app10082749
7 P. Shi, X. Fang, J. Ni, and J. Zhu: Appl. Sci.-Basel **11** (2021) 4001. https://doi.org/10.3390/app11094001
8 J. Redmon, S. Divvala, R. Girshick, and A. Farhadi: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 779. https://doi.org/10.1109/CVPR.2016.91
9 J. Redmon and A. Farhadi: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 6517. https://doi.org/10.1109/CVPR.2017.690

10 X. Zhu, S. Lyu, X. Wang, and Q. Zhao: Proc. 2021 IEEE/CVF Int. Conf. Computer Vision Workshops (ICCVW) 2778. https://doi.org/10.1109/ICCVW54120.2021.00312

11 B. Rahmat, Y. V. Via, A. Wasian, I. Y. Purbasari, N. K. Sari, W. Wurjani, S. Bandong, E. Joelianto, and P. I. Siregar: Video-based Container Tracking System Using Deep lLearning: Cyber Physical, Computer and Automation System, Advances in Intelligent Systems and Computing **1291**. (Springer, Singapore, 2021) 81. https://doi.org/10.1007/978-981-33-4062-6_8

12 W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg: SSD: Proc. 2016 European Conf. Computer Vision (ECCV) **9905**, 21. https://doi.org/10.1007/978-3-319-46448-0

13 K. He, X. Zhang, S. Ren, and J. Sun: Proc. 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 770. https://doi.org/10.1109/CVPR.2016.90

14 A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam: Mobilenets: arXiv:1704.04861 (2017). https://doi.org/10.48550/arXiv.1704.04861

15 R. Girshick: ICCV arXiv: 1504.08083 (2015). https://doi.org/10.48550/arXiv.1504.08083

16 S. Ren, K. He, R. Girshick, and J. Sun: IEEE Trans. Pattern Anal. Mach. Intell. **39** (2017) 1137. https://doi.org/10.1109/TPAMI.2016.2577031

17 M. Lokanath, K. S. Kumar, and E. S. Keerthi: IOP Conf. Ser. Mater. Sci. Eng. **263** (2017) 052028. https://doi.org/10.1088/1757-899X/263/5/052028

18 G. Cao, X. Xie, W. Yang, Q. Liao, G. Shi, and J. Wu: Proc. 2018 Int. Conf. Graphic and Image Processing (ICGIP) 106151. https://doi.org/10.1117/12.2304811

19 P. Kaur, B. S. Khehra, and A. P. S. Pharwaha: Math. Probl. Eng. **2021** (2021) 5565561. https://doi.org/10.1155/2021/5565561

20 J. Huang, Y. Shi, and Y. Gao: J. Comput. Res. Dev. **56** (2019) 319. https://doi.org/10.7544/issn1000-1239.2019.20170749

21 Y. Xu, Z. Ding, W. Li, K. Zhang, and L. Tong: J. Electr. Comput. Eng. **2023** (2023) 5399616. https://doi.org/10.1155/2023/5399616

22 B. Tang, Z. K. Song, W. Sun, and X.-D. Wang: IET Image Proc. **17** (2023) 1334. https://doi.org/10.1049/ipr2.12715

23 W. Liu, Z. Wen, J. Wang, S. Wang, H. Wang, W. Zhang, S. Huang, and J. Hong: Shock Vib. **2022** (2022) 1427417. https://doi.org/10.1155/2022/1427417

24 C. Chen, S. Wang, and S. Huang: Meas. Control **56** (2023) 832. https://doi.org/10.1177/00202940221092030

25 R. Girshick, J. Donahue, T. Darrell, and J. Malik: Proc. 2014 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 580. https://doi.org/10.1109/CVPR.2014.81

26 N. Bodla, B. Singh, R. Chellappa, and L. S. Davis: Soft-NMS: Proc. 2017 IEEE Int. Conf. Computer Vision (ICCV) 5562. https://doi.org/10.1109/ICCV.2017.593

27 T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie: Proc. 2017 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) 2117. https://doi.org/10.1109/CVPR.2017.106

28 H. Wang and T. Wang: Electronics **12** (2023) 93. https://doi.org/10.3390/electronics12010093

29 H. Liu, D. Wang, K. Xu, P. Zhou, and D. Zhou: Autom. Constr. **146** (2023) 104692. https://doi.org/10.1016/j.autcon.2022.104692

30 W. Qiao, H. Guo, E. Huang, H. Chen, and C. Lian: J. Mar. Sci. Eng. **11** (2023) 793. https://doi.org/10.3390/jmse11040793

31 H. Yin, L. Sun, X. Yu, X. Liu, and W. Yu: Neural Process. Lett. (2023) Online. https://doi.org/10.21203/rs.3.rs-2573179/v1

32 M. Lin, Q. Chen, and S. Yan: arXiv:1312.4400 (2014). https://doi.org/10.48550/arXiv.1312.4400

33 V. Nair and G. F. Hinton: Proc. 2010 the 27th Int. Conf. Machine Learning (ICML-10) 807.

34 A. Krizhevsky, L. Sutskever, G. E. Hinto: Proc. 2012 the 25th Int. Conf Neural Information Processing Systems 1097.

35 B. Wang, W. Pei, B. Xue, and M. Zhang: arXiv:2211.15143 (2022). https://doi.org/10.48550/arXiv.2211.15143

36 L. Han and A. Han: J. Comput. Theor. Nanosci. **12** (2015) 1121. https://doi.org/10.1166/jctn.2015.3860

37 L. Wang, Y. Shen, H. Liu, and Z. Guo: Cognit. Syst. Res. **58** (2019) 160. https://doi.org/10.1016/j.cogsys.2019.06.002

38 N. Otsu: IEEE Trans. Syst. Man Cybern.: Syst. **9** (1979) 62. https://doi.org/10.1109/TSMC.1979.4310076

39 Y. Liu, J. Sun, H. Yu, Y. Wang, and X. Zhou: Appl. Sci.-Basel **10** (2020) 6343. https://doi.org/10.3390/app10186343

40 S. Hong, Z. Jiang, L. Liu, J. Wang, L. Zhou, and J. Xu: Appl. Sci.-Basel **12** (2022) 11701. https://doi.org/10.3390/app122211701