Journal of Manufacturing Processes 69 (2021) 261–272

Contents lists available at ScienceDirect

Journal of Manufacturing Processes

journal homepage: www.elsevier.com/locate/manpro

Prediction of penetration based on infrared thermal and visual images during pulsed GTAW process

Rui Jiang, Runquan Xiao, Shanben Chen *

Intelligentized Robotic Welding Technology Laboratory, School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, PR China

ARTICLE INFO

Keywords:
Penetration state recognition
Faster R-CNN
Convolutional descriptor selection
Infrared thermal image
GTAW

ABSTRACT

Aiming at improving the recognition accuracy and robustness of the penetration state recognition model, a Dual-input Faster R-CNN (region-convolutional neural network) model with the input of original infrared thermal (IR) and visual (CCD) image was established. For avoiding the negative effects on recognition caused by arc flicker inside CCD image and irrelevant information inside background, synchronous feature extraction, convolutional descriptor selection and recognition based on synthetic features were designed in model. Since the industrial personal computer running the model is usually not equipped with powerful GPU and enough memory space, and sometimes it needs to train a specific model according to the new data set, the model is required to have high recognition accuracy, low recognition time, short training time and small occupation space. By sharing RPN (region proposal network) and ROI (region of interest) Pooling Layer inside Faster R-CNN and introducing Label-integrated Layer, the above requirements can be greatly met. The recognition accuracy of convolutional descriptor selection assisted Dual-input Faster RCNN reached more than 95%, while the recognition time for each IR&CCD-image data pair was less than 270 ms.

1. Introduction

Pulsed gas tungsten arc welding (GTAW) is a common welding process for aluminum alloy. It is widely used in automobile, ship-building, aerospace and other industries. In the actual welding process, there always exist complex, random, uncertain information, such as the random change of penetration state caused by uncertain welding environment and unstable welding parameters. To overcome or restrain the negative effects of these uncertainties on welding quality, it's necessary to promote the intellectualization of welding manufacturing process. Sensing and acquiring welding dynamic process information, as well as the state recognition of welding process, is the one of the significant technical part of intelligent welding manufacturing [1–2]. Information sensing and modeling of welding process are two important parts of intelligent welding [1].

At present, the conventional information sensing for welding process mainly including X-ray method [3–4], ultrasonic sensing method [5–6], infrared sensing method [7], arc voltage, arc light and acoustic signal sensing method [8–9] and visual sensing method [10–13]. The visual sensing method mainly uses CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) camera is used to directly or indirectly obtain the image of welding forming process, and the image feature information is processed and extracted according to different algorithms. Infrared sensing method mainly obtains the temperature field information of welding pool area in the welding process through infrared thermal imaging camera or infrared sensor and then extracts the feature information according to different data processing or image processing methods. Besides, Zhang [14] considered that abilities to sense and adapt to process conditions provide an assurance to actually produce the result from welding manufacturing as expected/predicted by the design, while Khaleghi [15] mentioned that multi-information fusion can enhance the authenticity of information. The multi-sensor technology, which is used in fields like mapping [16], fault diagnosis [17–18], rocket test system [19] and robot system [20], has been applied to welding sense [21–22]. The use of multi-sensor technology can obtain more comprehensive and rich information, can better sense and adapt to the actual welding process.

As for the modeling of welding process especially about penetration state recognition model or prediction model, some scholars have established the prediction model of penetration and width based on the infrared thermal images (IR images) of welding pool. The establish process of the model includes image information acquisition, image

* Corresponding author.
E-mail address: sbchen@sjtu.edu.cn (S. Chen).

https://doi.org/10.1016/j.jmapro.2021.07.046
Received 11 May 2021; Received in revised form 4 July 2021; Accepted 25 July 2021
Available online 5 August 2021

processing, data set making, model training and testing. Subashini [23] used K-means algorithm to segment the key frames of the welding pool's IR images, extracted the length and width of the welding pool, and combined the characteristic information of the welding pool's temperature field to make the data set, then used ANFIS to establish the model. Chandrasekhar [24] used the CA algorithm to segment the IR images, and established the prediction model of welding pool's width and depth based on ANFIS. Ghanty [25] used k-means algorithm to segment IR images, and ANN to establish prediction model of weld width and depth. At the same time, some scholars have established the penetration state recognition model for the visible image, such as Zhang [26] made three-dimensional imaging of the welding pool in GTAW process, and established the linear relationship between the average depression depth of the welding pool and welding pool back width. The feasibility of controlling the full penetration state (root surface weld width) by measuring and controlling the welding surface parameters was expounded. Liu [27] used laser point cloud to reconstruct welding pool morphology, and realized real-time acquisition of weld pool geometric parameters. Taking the geometric parameters of the molten pool as the input and the back width of the welding pool (related with penetration) as the output, the prediction model was established based on ANFIS. Kovacevic [28] used a high-speed shutter camera with pulsed laser illumination to capture the weld pool surface image in GTAW welding process, extracted the edge of the image, then extracted the geometric parameters from the edge, and established the nonlinear functional relationship between the geometric parameters and the welding pool back width (corresponding to the penetration) by using ANN. Jiao [29] by putting the collected passive vision image and the feature information extracted from the image into the CNN (convolutional neural networks) model based on ResNet (residual network), and training by the way of transfer learning. Zhang [30] collected CMOS images of welding pool by high dynamic range imaging technology, and then trained Softmax penetration prediction model based on six features extracted from welding pool. Cheng [31] used active vision sensing to collect the laser line images reflected by welding pool, and established CNN model to realize the penetration state prediction based on the laser line images. Cheng [32] designed a new active vision monitoring method for GTAW, using dual cameras to capture the laser line reflected from the surface of welding pool as well as image of the back of the welding pool, and established a welding pool back width prediction model based on CNN structure. Besides, Cheng [33] also tried to combine the initial state image and current state image of laser line reflected by welding pool surface to enrich the information of input image and established the welding pool back width prediction model based on CNN structure.

The prediction model based on multi-sensor technology has more abundant information input, which could more accurately predict the penetration state. Chen [21] applied multi-sensor information fusion technology to pulsed GTAW. Arc sensor, vision sensor and sound sensor were used to obtain welding current, welding voltage, welding pool images and welding sound information in the welding process at the same time. D-S evidence theory with back-propagation (BP) neural network assignment was used to fuse different signals to predict the penetration state of welding process. Based on DLSTM (dynamic long short-term memory), Chen [22] established a model with welding current, welding voltage, welding pool images and welding acoustic signal as input and penetration state as output. Wang [34] used digital twin method to collect arc image and surface image of welding pool during GTAW welding, and input two kinds of graph into CNN model to predict welding pool back width. Feng [35] collected the active vision image, passive vision image and reverse electrode image of the welding pool, and used the generative adversarial networks to denoise the active vision image. The establishment of the model introduces the idea of ensemble learning. For each input image, five models with good generalization performance are used to predict the penetration state, and the prediction results were obtained by plurality voting. For the prediction results of the three images, the plurality voting was carried out again to finally obtain penetration prediction results.

As a part of multi-sensor technology, CCD and IR camera system has been applied in some fields, such as satellite remote sensing [36], forest fire detection system [37] and urban traffic monitoring system [38]. IR camera can record thermal radiation information (temperature field information) better, while CCD camera can feedback spatial information better [36]. According to the double ellipsoid heat source model [39–42], there exists a close relationship between the welding pool temperature field information and spatial information. Thus, in the welding field, IR image can well reflect the temperature field information of the welding pool zone during the welding process. Through the temperature field data and temperature gradient information inside the IR image, the penetration state of welding pool will be predicted. At the same time, CCD image can feed back the surface morphology information of welding pool. With the change of welding arc, CCD image also has obvious information feedback. Therefore, to improve the robustness and recognition accuracy of the penetration state recognition model, it is necessary to input IR image and CCD image at the same time.

In this paper, Dual-input Faster R-CNN model was established for the recognition of penetration state based on IR and CCD images of welding pool during pulsed GTAW process. The model used IR and CCD original image as input to reduce the time-consuming of data set preparation and decrease the data error in the preprocessing process. For avoiding the negative effects of arc flicker inside CCD image on penetration state recognition, this model was designed to extract the features from both IR and CCD images synchronously, generate the bounding box of the main target respectively, and recognize the state based on above two kinds of feature. By applying the convolutional descriptor selection to IR image's feature map, the anti-interference of the model to the background irrelevant information was obviously improved. By sharing different modules in the Faster R-CNN [43] and introducing the Label-integrated Layer, the Dual-input Faster R-CNN, which enable to extract the features of IR image and CCD image, shows great recognition accuracy and robustness. Finally, the optimal structure of Dual-input Faster R-CNN which can output the bounding box as well as penetration state of IR and CCD images is determined the robustness and recognition rate of Dual-input Faster R-CNN were improved significantly.

2. Setup and experiment

2.1. Setup overview

Fig. 1 shows the schematic diagram of the pulsed GTAW experimental platform. It is mainly composed of FUNAC robot, pulsed GTAW power supply system, visual sensing system, programmable logic controller (PLC) and industrial personal computer (IPC). The IPC uses Modbus TCP protocol to communicate with PLC to control welding current and wire feeding speed. The GTAW torch is installed at the end of the robot arm, and the IPC controls the welding speed by adjusting the motion parameters of the robot. The visible light image (CCD image) and infrared thermal image (IR image) of welding pool are collected by CCD camera and IR camera respectively, and the image information is sent to the IPC.

The arrangement of IR camera and CCD camera is shown in Fig. 2. The CCD camera is located 350 mm away from the GTAW torch, and the horizontal angle with the workpiece is 45°. The IR camera is fixed on the center line of the workpiece's side parallel to the weld direction. In the horizontal direction, the vertical distance from IR camera to the weld seam is 500 mm, while the horizontal angle between them is 45°. To obtain the correlation between the image and the world coordinate system, several square grid templates are used to calibrate the CCD as well as the IR camera, respectively. The image resolution of CCD sensor is 0.021 mm/pixel, and the image resolution of IR sensor is 0.405 mm/pixel when it is calibrated in visible camera mode. The sensor system can realize the synchronous acquisition of IR and CCD images of welding pool. The detailed technical specifications of IR camera and CCD camera

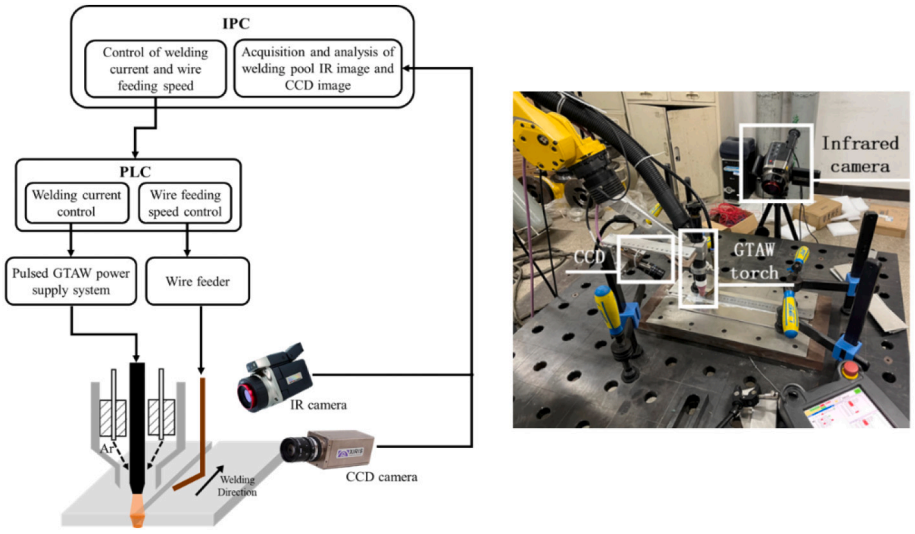


Fig. 1. Schematic of pulsed GTAW experimental platform.

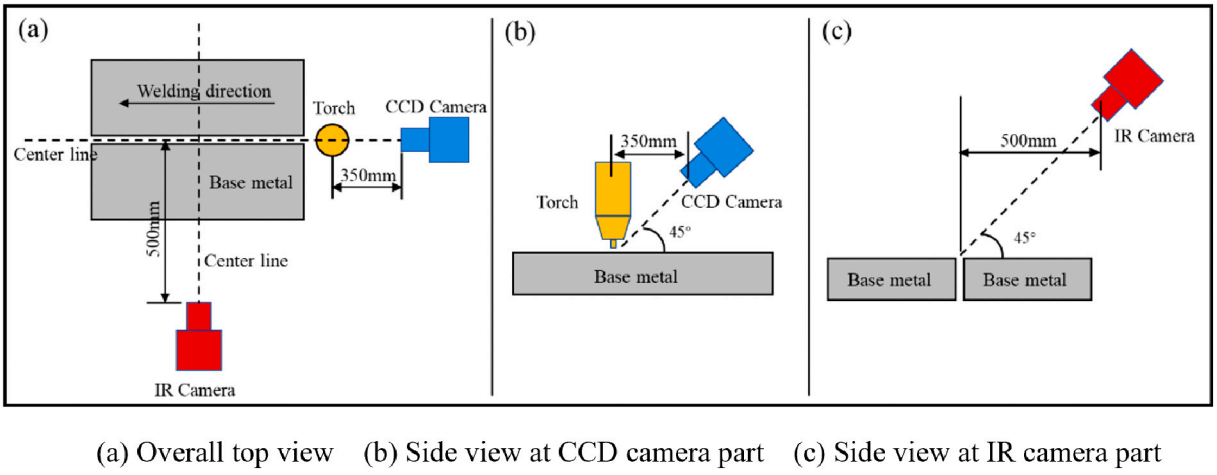


Fig. 2. Schematic diagram of CCD and IR camera placement.

Table 1
The detailed technical specifications of the IR and CCD camera.

Type	Sensitivity (NETD)	Precision	Size of recorded image	Spatial resolution	Sampling frequency
AVIO R500EX-Pro	0.025 °C	±1 °C	640 pc × 480 pc	0.87 mrad	30 Hz
XIRIS XVC-1000	–	–	1280 pc × 1024 pc	–	30 Hz

are given in Table 1.

2.2. Experiment design

Table 2 presents the general welding conditions. The welding experiment is designed as follows: the welded joints were made L300 × W50 × H4 LF6 Aluminum alloy plates and the welding position is butt-weld in the downhand position. Two experimental groups were set up by changing the welding current (including basic current and peak current) to get the image of welding pool with different penetration state. In the whole welding process, it was necessary to ensure that the CCD and the IR camera could collect the molten pool synchronously in real time.

2.3. Preparation for data sets

During the welding experiment, the two groups of welding parameters were repeated twice respectively, and totally 1548 IR&CCD data pairs were collected. After screening out the irrelevant frames (such as the blank frames before arcing), 1416 data pairs were obtained.

Based on the collected IR images and CCD images, to establish the model, three kinds of data set need to be established: IR-image data set, CCD-image data set and IR&CCD data set. Among them, the IR-image data set and CCD-image data set are respectively used for the establishment of the recognition model of penetration state based on CCD images of welding pool and the recognition model of penetration state based on IR images of welding pool. The IR&CCD-image data set is composed of IR&CCD-image data pairs. In each data pair, two kinds of images are required to correspond to the penetration state of the welding

Table 2
Welding parameters.

Welding condition	Parameter	Welding condition	Parameter	
Current polarity	AC Pulse	Argon flow (L/min)	12	
AC frequency (Hz)	50	Max base current (A)	120	
Pulse frequency (Hz)	10	Max peak current (A)	250	
Retention of start time (s)	3	Peak current (A)	120–140–160–180–200–220	140–160–180–200–220–240
Welding speed (mm/s)	3	Base current (A)	60–70–80–90–100–110	70–80–90–100–110–120
Wire feed rate (cm/mm)	50	AC balance ratio (%)	−65	

pool at the same time. The above three data sets need label, bounding box selection and data augmentation.

For IR images and CCD images, they are labeled in three states: incomplete penetration, complete penetration and burn-through according to Fig. 3.

The next is the selection of bounding box. For the IR images, the bounding box is used to select the effective temperature field in the infrared images, that is, the temperature field in the welding pool area. The size of IR images' bounding box is 60 px × 60 px. While for the CCD images, the welding pool area is selected as bounding box. The size of CCD images' bounding box is 600 px × 600 px.

Each bounding box is saved as:

$$bounding\ box_i = \left[x_{1,i}, y_{1,i}, x_{2,i}, y_{2,i}\right] \tag{1}$$

Inside Eq. (1), $x_{1,\,i}$, $y_{1,\,i}$ means the abscissa and ordinate of the upper left corner of the bounding box, respectively. While $x_{2,\,i}$, $y_{2,\,i}$ means the abscissa and ordinate of the lower right corner of the bounding box.

Finally, the data augmentation is processed as two steps:

(1) Every image as well as the image's corresponding bounding box inside the data set will be randomly rotated from −10° to 10°, then the rotation results will be cropped and filled to the original size (640 px × 480 px for IR images and 1280 px × 1080 px for CCD images), finally the rotation results will be saved to the training mini-batch.
(2) Randomly pick half of the images from the data set, then add 6% salt and pepper noise points to the original image.

Each step above was finished on OpenCV-Python (version 4.4.0). The effects of steps (1) and (2) are shown in Fig. 4. Step (1) made the number of data set doubled, while step (2) did not influence the number of data set. Thus, 2832 sets of IR&CCD-image data pairs were obtained.

Each kind of data set will be shuffled and then divided into three parts: 60% original data set is divided to train-part data set, 40% original data set is equally divided between validation-part data set and test-part data set, that means the train-part, validation-part and test-part data set contains 1700, 226 and 227 sets data pairs, respectively. Train-part data set is used to train the new model, validation-part data set is used to detect whether the model is over fitted after each epoch update and test-part data set is used to test the generalization performance of the model.

3. The establishment of Dual-input Faster R-CNN

3.1. Faster R-CNN

The structure diagram of Faster R-CNN is shown in Fig. 5. Faster R-CNN mainly consist of Feature Extractor, Region Proposal Network (RPN), ROI (Region of Interest) Pooling, Classification & Regression Layer. This model has three main features:

(a) Faster R-CNN integrates the generation of anchor box, anchor box labeling and position correction, and region proposals output through RPN.
(b) Through the ROI Pooling Layer, the region proposals information is introduced into the feature map to realize the ROI processing of the feature map.
(c) The RPN Layer directly uses the feature map output by feature extractor, and finally replaces the SVM (for classification and regression) in the original R-CNN [44] by two parallel fully connected layer groups to output image labeling results and bounding box coordinates, which makes Faster R-CNN adopt the

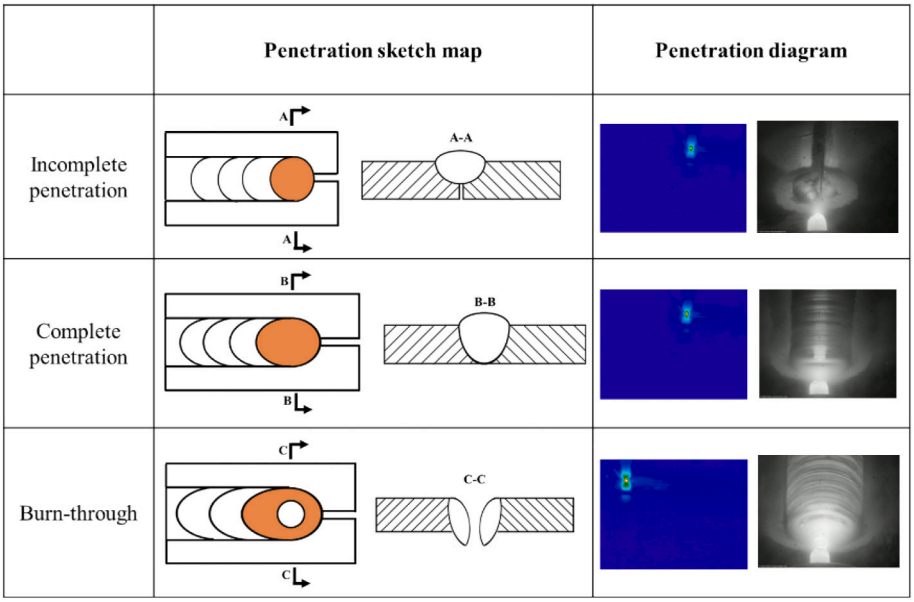
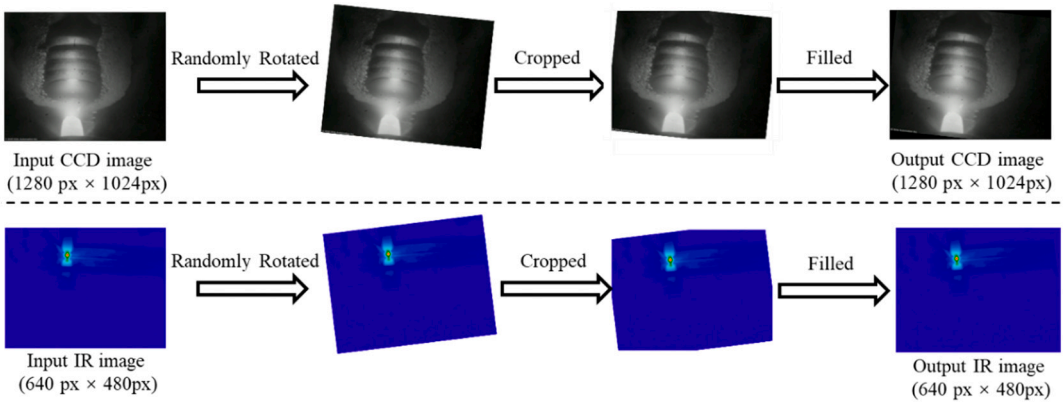
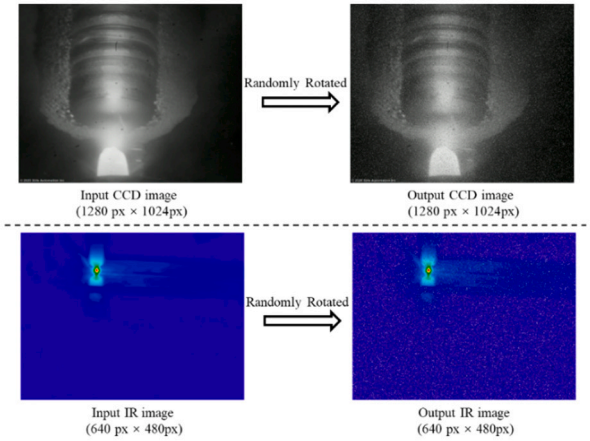


Fig. 3. Penetration state schematic diagram.



(a) The effects of step (1) in data augmentation



(b) The effects of step (2) in data augmentation

Fig. 4. The effects of data augmentation.

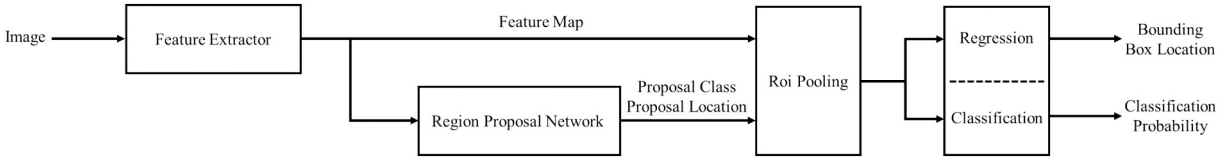


Fig. 5. Structure diagram of Faster R-CNN.

end-to-end training mode, reducing hardware and time consumption.

The functions of Feature Extractor are extracting the feature of input image and then showing the corresponding feature map. Currently, the networks like ResNet18, ResNet50 and VGG16 are commonly used as the backbone of Feature Extractor.

The structure diagram of RPN is shown in Fig. 6. RPN is built to generate the region proposals for the feature map from Feature Extractor.

For each feature map, RPN will firstly enrich information by using a 3×3 convolutional layer, then, a preset number of anchor boxes will be generated, these anchor boxes will be divided into two kinds: foreground and background, which is positive sample and negative sample, respectively, based on the rules:

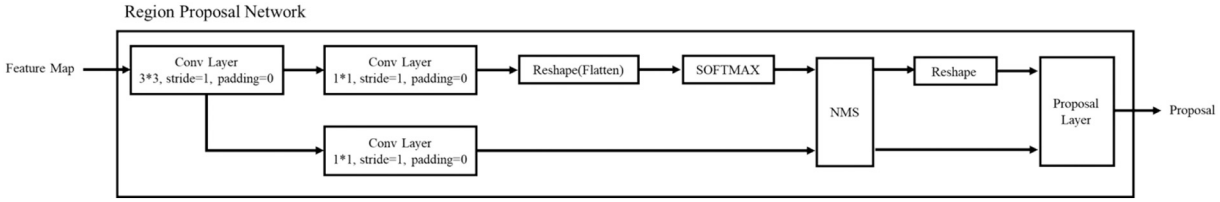


Fig. 6. Structure diagram of RPN.

(a) If the anchor box and ground truth box have the largest intersection over union (IOU) value, they are marked as positive samples, label = 1.
(b) If the IOU of anchor box and ground truth box is >0.7, it is marked as positive sample, label = 1.
(c) If the IOU of anchor box and ground truth box is <0.3, it is marked as negative sample, label = 1.

After that, the difference between anchor box and ground truth box will be calculated. By using non-maximum suppression (NMS) process, the overlapping boxes can be removed. Finally, a few proposal boxes will be gained.

The loss function of Faster R-CNN is:

$$L(\{p_i\},\{t_i\}) = \frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i, p_i^*\right) + \lambda\frac{1}{N_{reg}}\sum_i L_{reg}\left(t_i, t_i^*\right) \tag{2}$$

$$L_{cls}\left(p_i, p_i^*\right) = -\sum_{j=1}^{T} p_j log p_j^* \tag{3}$$

$$L_{reg}\left(t_i, t_i^*\right) = \begin{cases} 0.5*\left(t_i - t_i^*\right)^2 & \left|t_i - t_i^*\right| < 1 \\ \left|t_i - t_i^*\right| - 0.5 & \text{otherwise} \end{cases} \tag{4}$$

Inside Eq. (2):
p_i: prediction classification probability of No. i anchor;
p_i*: the real label of No. i anchor (when anchor is positive sample, p_i* = 1);
t_i: parametric coordinates of bounding box predicted by No. i anchor;
t_i*: the real parametric coordinates of bounding box predicted by No. i anchor;
N_{cls}: the size of mini-batch;
N_{reg}: the number of anchor location;
λ: weight balance parameters.

3.2. Structure of Dual-input Faster R-CNN model

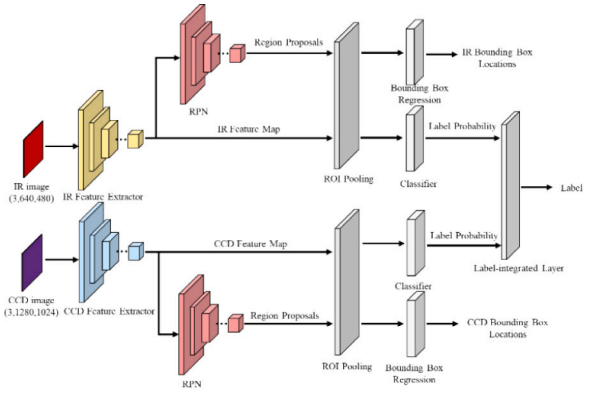
Four kinds of Dual-input Faster R-CNN models (as shown in Fig. 7) were established in order to know the way to achieve the best performance by sharing modules. All models used IR and CCD original image as input, which can (1) reduce the preparation time of data set and (2) reduce the redundant recognition error caused by image segmentation or other preprocessing methods. For convenience, the models above were named as DFR-1 to DFR-4. Besides, the models, based on original Faster R-CNN and trained with IR-image data set and CCD-image data set respectively, were established to illustrate the advantages of Dual-input Faster R-CNN in the recognition of penetration state. These two models were named as FR-IR and FR-CCD.

The loss function of Dual-input Faster R-CNN can be described as:

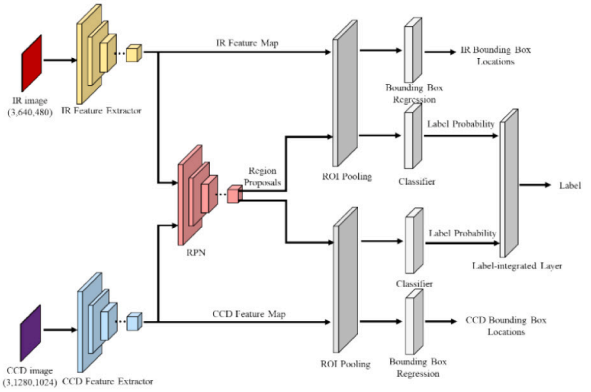
$$L\left(\{p_i\},\{t_i^{IR}\},\{t_i^{CCD}\}\right) = \frac{1}{N_{cls}}\sum_i L_{cls}\left(p_i, p_i^*\right) + \lambda_1\frac{1}{N_{reg}^{IR}}\sum_i L_{reg}\left(t_i^{IR}, t_i^{IR*}\right) + \lambda_1\frac{1}{N_{reg}^{IR}}\sum_i L_{reg}\left(t_i^{IR}, t_i^{IR*}\right) \tag{5}$$

$$L_{cls}\left(p_i, p_i^*\right) = -\sum_{j=1}^{T} p_j log p_j^* \tag{6}$$

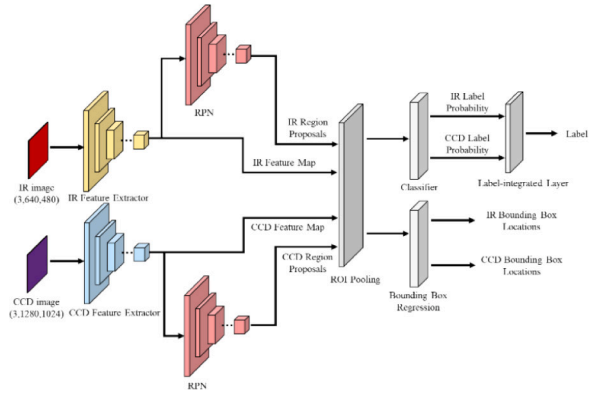
$$L_{reg}\left(t_i, t_i^*\right) = \begin{cases} 0.5*\left(t_i - t_i^*\right)^2 & \left|t_i - t_i^*\right| < 1 \\ \left|t_i - t_i^*\right| - 0.5 & \text{otherwise} \end{cases} \tag{7}$$



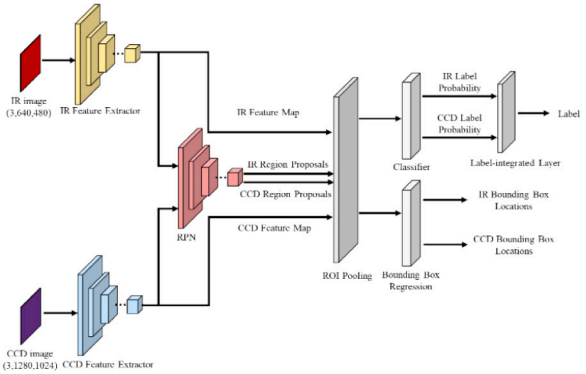
(a) DFR-1



(b) DFR-2



(c) DFR-3



(a) DFR-4

Fig. 7. Structure diagrams of Dual-input Faster R-CNN models.

Inside Eq. (5):

p_i: prediction classification probability of No. i anchor;

p_i^*: the real label of No. i anchor (when anchor is positive sample, $p_i^* = 1$);

t_i^{IR}: parametric coordinates of bounding box predicted by No. i anchor came from IR images' feature map;

t_i^{IR*}: the real parametric coordinates of bounding box predicted by No. i anchor came from CCD images' feature map;

t_i^{CCD}: parametric coordinates of bounding box predicted by No. i anchor came from CCD images' feature map;

t_i^{CCD*}: parametric coordinates of bounding box predicted by No. i anchor came from CCD images' feature map;

N_{cls}: the size of mini-batch;

N_{reg}^{IR}: the number of anchor location came from IR images' feature map;

N_{reg}^{CCD}: the the number of anchor location came from CCD images' feature map;

λ_1, λ_2: weight balance parameters.

The Label-integrated Layer inside Dual-input Faster R-CNN, which is obviously different from the original Faster R-CNN, is to integrate the classification probability results of IR and CCD images to ensure the consistency of classification results. Label-integrated Layer can be described as:

$$p_i = \sum_{i=1}^{c} \left(p_{IR}[i] + w_1 * p_{CCD}[i]\right) \Big/ 2 \tag{8}$$

Inside Eq. (8), $p_{IR}[i]$ and $p_{CCD}[i]$ means the classification probability results of IR and CCD images, respectively, w_1 is weight balance parameter and it will be trained with the model, p_i is the prediction classification probability of No. i anchor, it's also the input of Eq. (5).

Considering that the detected target (the temperature field of welding pool zone) in the original IR image only accounts for a small part of the whole image, there will be two negative effects on the recognition of the model: Firstly, the selection of bounding box may be affected by the irrelevant information of the image's background, thus affecting the final recognition results; the second is the model may be insensitive to the tiny changes of the temperature field information inside the welding pool zone. In order to avoid the above negative effects, it's essential to select the main objects inside the original feature map to ensure the model will output bounding box as well as recognition result of the correct object. Thus, convolutional descriptor selection, which can achieve the unsupervised selection of main objects from feature maps, is taken into consideration. This method can achieve the unsupervised selection of main objects. The use of convolutional descriptor selection, which is helpful for fine-grained image recognition, is inspired by SCDA (selective convolutional descriptor aggregation) [45]. This method can be mainly divided into three steps: generating the activation map; obtaining the mask map; selecting the descriptors of original feature map, and its mathematical expression is as follows:

$$\widehat{F}_{ch} = \left\{ F_{(i,j,ch)} \,|\widetilde{M}_{i,j} = 1 \right\} \tag{9}$$

$$\widetilde{M}_{i,j} = LLC\left(M_{i,j}\right) \tag{10}$$

$$M_{i,j} = \begin{cases} 1 \ if \ A_{i,j} > \overline{a} \\ 0 \ \text{otherwise} \end{cases} \tag{11}$$

$$\overline{a} = \frac{\sum_i \sum_j A_{i,j}}{m \bullet n} \tag{12}$$

$$A = \sum_{n=1}^{d} F_{ch} \tag{13}$$

Inside Eqs. (9)–(13):

$\widehat{F}_{ch}$: the No. ch channel's feature map after convolutional descriptor selection;

F_{ch}: the original No. ch channel's feature map;

$\widetilde{M}_{i,j}$: the largest connected component (LLC) of the mask map ($M_{i,j}$);

$\overline{a}$: selection threshold of each descriptor;

A: the activation map of input feature map.

Two more Dual-input Faster R-CNN models, named as SSCD-DFR and DSCD-DFR respectively, were established. In the former one, The convolutional descriptor selection was applied to process IR image's feature map output by Feature Extractor of DFR-4, while in the latter one, the IR and CCD image's feature map were processed at the same time of DFR-4.

3.3. Training hyperparameters

The hyperparameters of the above models during the training process are shown in Table 3, where anchor ratios and anchor scales are used to specify the number of anchor boxes with different sizes and aspect ratios generated for each pixel. In this training, a total of 9 anchor boxes are generated for each pixel. Train_num_before_NMS and Train_num_after_NMS means the number of anchor box before and after NMS process respectively during training, while Test_num_before_NMS and Test_num_after_NMS means the number of anchor box before and after NMS process respectively during testing. Although data augmentation is used, the training data set still belongs to small data set, which would not be good for training. To avoid this problem, transfer learning is used to train the model. The specific training method is: before training, the Feature Extractor is set to ResNet18 pre-trained by ImageNet, which will lead to better feature extraction capability of Feature Extractor [14]. Set the training epoch to 40, freeze the feature extractor in the first 20 epochs, only train the RPN, ROI Pooling Layer, fully connected layer (Classification Layer and Bounding Box Regression Layer in Fig. 5) and Label-integrated Layer in the model, then thaw the Feature Extractor and train the whole model in the last 20 epochs. During the whole training, L2 regularization is used to suppress the occurrence of over fitting. The training method can speed up and optimize the learning efficiency of the model while avoiding the over fitting problem of the model. All models were trained and tested by PC shown in Table 4 and with PyTorch (version 1.2.0) + CUDA (version 10.0) platform.

4. Results and discussion

4.1. Performance of Dual-input Faster R-CNN

The loss value-epoch curves of FR-IR and FR-CCD are shown in Fig. 8. The loss value-epoch curves as well as w_1-epoch curves of Dual-input Faster R-CNN models are included in Fig. 9. It is considered that all loss value-epoch curves show a normal downward trend, and the fluctuation gradually becomes smaller. For the loss value-epoch curves of Dual-input Faster R-CNN, because of the thawing of feature extractor, there exists a sudden increasement in the loss value when epoch = 20. After training with epoch = 40, the w_1 values of each Dual-input Faster R-CNN model are listed in Table 5.

The output results of Dual-input Faster R-CNN models are shown in

Table 3
Training hyperparameters of models.

Hyperparameter	Value	Hyperparameter	Value
Anchor ratios	[0.5, 1, 1.5]	Feature Extractor backbone	Resnet18
Anchor scales	[8, 16, 32]	Epoch	40
Train_num_before_NMS	12,000	Learning rate	Initial value: 1e−4
Train_num_after_NMS	2000		With StepLR: step size = 1, gamma = 0.95
Test_num_before_NMS	3000	Optimizer	Adam
Test_num_after_NMS	300		L2 regularization parameters: 5e−4

Table 4
Model training equipment configuration.

Component	Model number
CPU	Intel i7-10700
GPU	NVIDIA GeForce RTX 2060S
RAM	DDR4 16 Gb (dual channel)
Hard disk	1 Tb (7200 rpm)

the Fig. 10. According to the partial enlarged images in Fig. 8, it is considered that the above six models enable to generate bounding boxes of different sizes and aspect ratios for both IR and CCD images. Besides, each bounding box can accurately select the area corresponding to the temperature field for IR images and welding pool zone for CCD images, and the useless background box in the original image is rarely selected. During testing, the recognition accuracy of DFR-1 to DSCD-DFR is 95.58%, 93.69%, 92.34%, 94.47%, 95.90% and 96.10% respectively, while the recognition time cost is 230 ms, 243 ms, 256 ms, 246 ms, 320 ms and 264 ms per IR & CCD-image data pair.

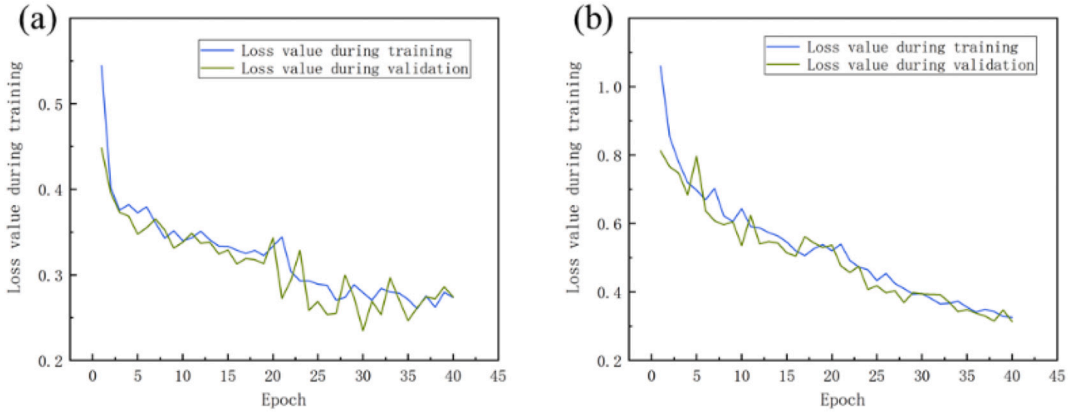


Fig. 8. Loss value-epoch curves of FR-IR (a) and FR-CCD (b).

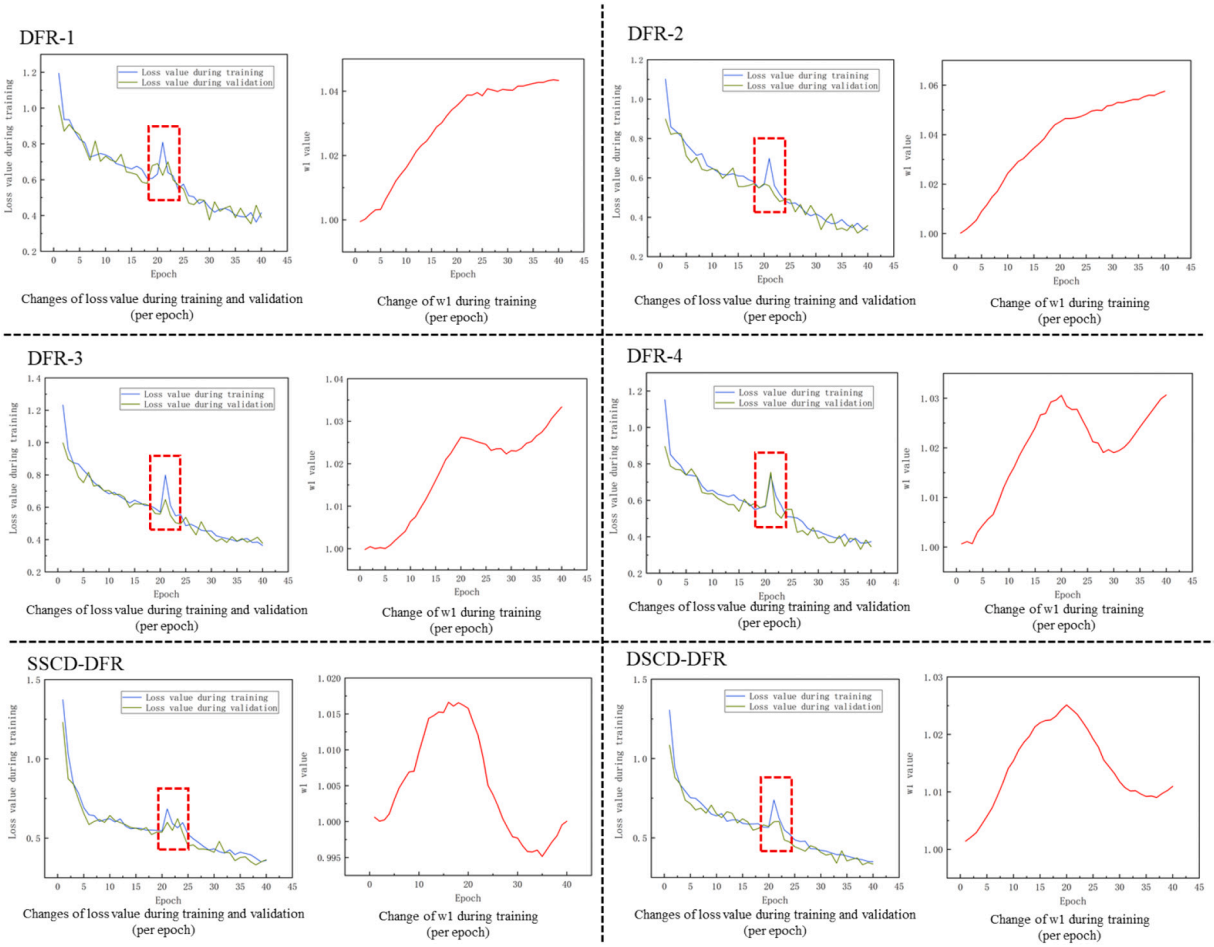


Fig. 9. Loss value-epoch curves and w_1-epoch curves of each model.

Table 5
w_1 values of each model (epoch = 40).

Model name	w_1 value after training (epoch = 40)
DFR-1	1.0433037
DFR-2	1.0575647
DFR-3	1.0333524
DFR-4	1.0306288
SSCD-DFR	1.0000456
DSCD-DFR	1.0109748

The recognition accuracy is defined as:

$$\text{acc} = \frac{\text{Number of samples correctly predicted}}{\text{Total number of samples}} \tag{14}$$

As for FR-IR and FR-CCD, the recognition accuracy and recognition time cost is 90.84% and 93.69% and 138 ms and 243 ms respectively.

4.2. Advantages of Dual-input Faster R-CNN in penetration state recognition

A model comparison table (listed in Table 6) is established according to the performance and related parameters of the models. From a macro perspective, for IR images, CCD images and IR & CCD-images data pairs, the Dual-input Faster R-CNN and Faster R-CNN can output bounding box as well as the corresponding category. The recognition accuracy of all models can reach more than 90% while the recognition time cost of all models will be less than 0.3 s. Compared with the three-classification recognition model in Ref.21, whose recognition accuracy was 94.4%, DFR-1, DFR-4, SSCD-DFR and DDSC-DFR all showed higher accuracy.

The difference between the Dual-input Faster R-CNN and original Faster R-CNN in penetration state recognition is obvious. The recognition accuracy of Dual-input Faster R-CNN model is significantly higher than that of Faster R-CNN model (approximately 3% higher on average). The reason for this phenomenon is Dual-input Faster R-CNN model enables to simultaneously read the features of IR image and CCD image from the welding pool zone, that is, the input information is more abundant.

Because in the process of pulsed GTAW, the welding arc will keep flashing due to the influence of pulse. The flashing of arc will affect the recognition results of the model. When the IR image and CCD image of the same welding pool at over penetration state input FR-IR and FR-CCD respectively, these two models will output different results (as Fig. 11 shows). The results show that the FR-CCD model will consider the welding pool is not penetrated because of arc extinction. Compared with the output of FR-IR and FR-CCD models, Dual-input Faster R-CNN can accurately judge the same input. The biggest feature of Dual-input model is to support the simultaneous input of IR and CCD images and

Table 6
Model performance and related parameter table.

Model name	Accuracy	Recognition time (per frame)	Training time (Epoch = 40)	Storage occupation
DFR-1	95.58%	230 ms	1 h:45 min:35 s	94.98 MB
DFR-2	93.69%	243 ms	1 h:23 min:3 s	52.63 MB
DFR-3	92.34%	256 ms	1 h:20 min:46 s	53.43 MB
DFR-4	94.47%	246 ms	1 h:22 min:20s	52.60 MB
SSCD-DFR	*95.87%*	*264 ms*	*2 h:2 msin:23 s*	*52.60 MB*
DDSC-DFR	96.10%	320 ms	3 h:12 min:14 s	52.60 MB
FR-IR	90.84%	138 ms	46 min:55 s	47.43 MB
FR-CCD	91.57%	182 ms	47 min:10 s	47.43 MB

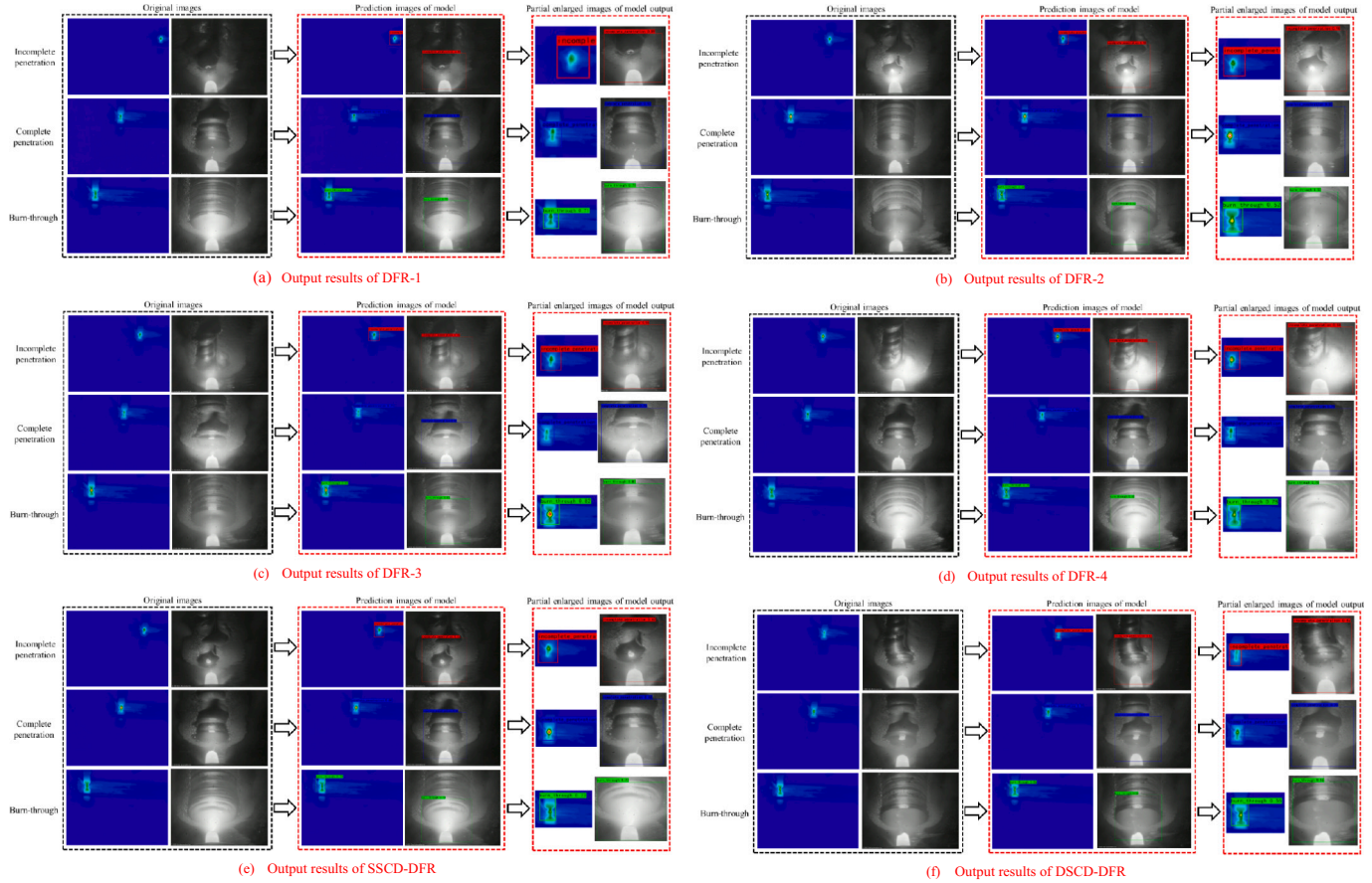


Fig. 10. Output results of Dual-input Faster R-CNN.

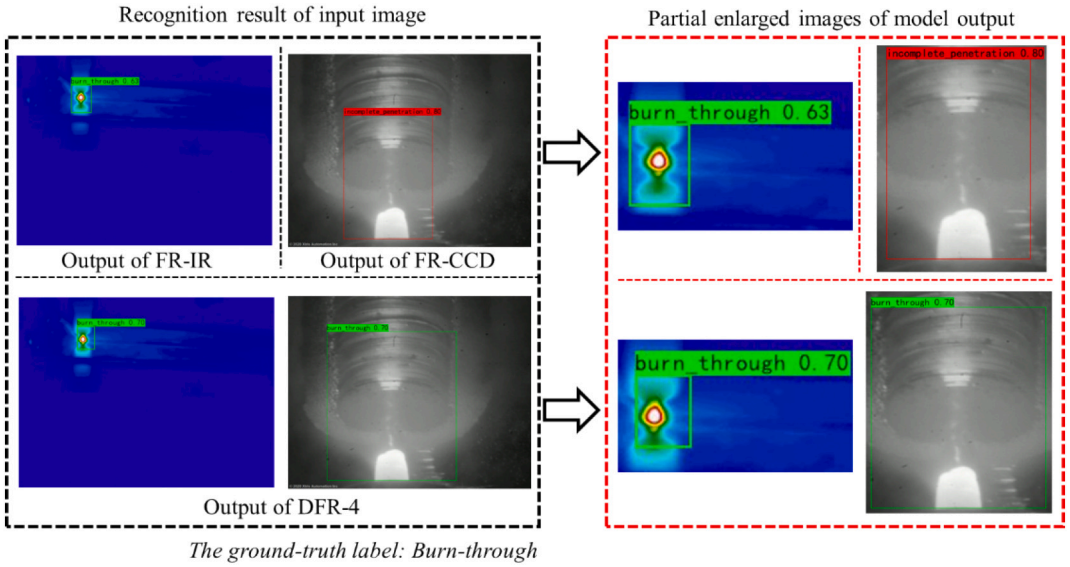


Fig. 11. Output comparison of DFR and FR.

the addition of Label-integrated Layer. The IR image records the temperature field information of the welding pool zone. During welding process, the temperature field of welding pool will appear obvious thermal delay, so that the change of the temperature field is always slower than that of the welding arc (heat source). Moreover, there exists an obvious relationship between the temperature field information of the welding pool and the penetration state (for instance, the position with high central temperature and large temperature gradient is generally the place with the maximum penetration). Thus, the information of penetration state can be fed back through the temperature field (IR image). CCD image mainly records the macro morphology of the welding pool and a small part of the arc state information. Compared with IR image, the information recorded by CCD image will change rapidly, but the rapid-changing data is not conducive to the recognition of the model, so IR image with delay characteristics can alleviate the misjudgment of CCD image. On the other hand, the Label-integrated Layer can make the model synthesize the IR image and CCD to give the final recognition results and improve the robustness of the model.

4.3. The influence of module sharing on model performance

By comparing the data of the first four rows in Table 6, it can be found that the sharing of different modules will have various degrees of influence on the model performance and related parameters.

Compared with the other three models, DFR-1 only adds the label integration layer on the basis of using two Faster R-CNN. DFR-1 has the highest recognition accuracy and the shortest recognition time. Nevertheless, since DFR-1 doesn't share any modules, the training time and storage occupation are also the largest. If VGG16 is used as the backbone of Feature Extractor instead of ResNet18, the storage occupation will increase up to more than 500 MB, which is not convenient for the actual training and use of the model.

The data of DFR-2 and DFR-3 show that merely sharing RPN and ROI Pooling Layer can significantly reduce the storage occupation and training time of the model (compared with RPN-1, the storage occupation and training time are reduced by approximately 55.8% and 22.3% respectively). Yet the recognition accuracy of DFR-2 and DFR-3 will also decrease, which is approximately 2.7% lower than that of DFR-1.

On the premise of high recognition accuracy (merely 1.2% lower than that of DFR-1), the storage occupation, training time and recognition time of DFR-4 are close to those of DFR-2 and DFR-3. Therefore, it is considered that the DFR model sharing RPN and ROI Pooling Layer will achieve the best comprehensive performance when applied to the recognition of penetration state.

4.4. The influence of convolutional descriptor selection

According to the data of the first six rows in Table 6, it is found that convolutional descriptor selection improves the recognition accuracy of the model by ~2% and ~6% compared with that of DFR-4 and FR-IR, respectively. Fig. 10 illustrates the output of FR-IR, DFR-4, SSCD-DFR and DSCD-DFR for the same burn-through state IR image. It is found that due to the influence of background irrelevant information of the original image, FR-IR and DFR-4 cannot locate the bounding box in the welding pool zone, which leads to the wrong feature selection and the penetration state recognition further. Compared with FR-IR and DFR-4, SSCD-DFR and DSCD-DFR enable to suppress the interference of background irrelevant information, and then select the correct bounding box.

Considering the improvement of recognition accuracy and the output listed in Fig. 12, convolutional descriptor selection is thought to make positive effect on bounding box selection of the Dual-input Faster R-CNN. By selecting the main target of the feature map output from Feature Extractor, the model is able to more accurately select the bounding box. What's more, inside the ROI Pooling Layer, the proposal box generated by RPN is used to intercept the original feature map, and the intercepted feature map is resized to the size before interception, so that the model can detect tiny changes in the feature map.

By comparing the relevant parameters of SSCD-DFR and DSCD-DFR in Table 6, it is considered that applying convolutional descriptor selection to the feature map of IR image and CCD image at the same time will lead to the longer training time of the model (134% longer than of DFR-4), and the recognition time of model will increase to more than 300 ms. Nevertheless, when only applying convolutional descriptor selection to the feature map of IR image, the recognition time will be controlled within 270 ms, while the recognition accuracy achieves more than 95%. Thus, with the basis of the DFR model sharing RPN and ROI Pooling Layer, the best comprehensive performance of model will be obtained after applying convolutional descriptor selection to the IR image's feature map.

5. Conclusions

A Dual-input Faster R-CNN model was established to identify the penetration state of welding pool in pulsed GTAW. The model takes IR

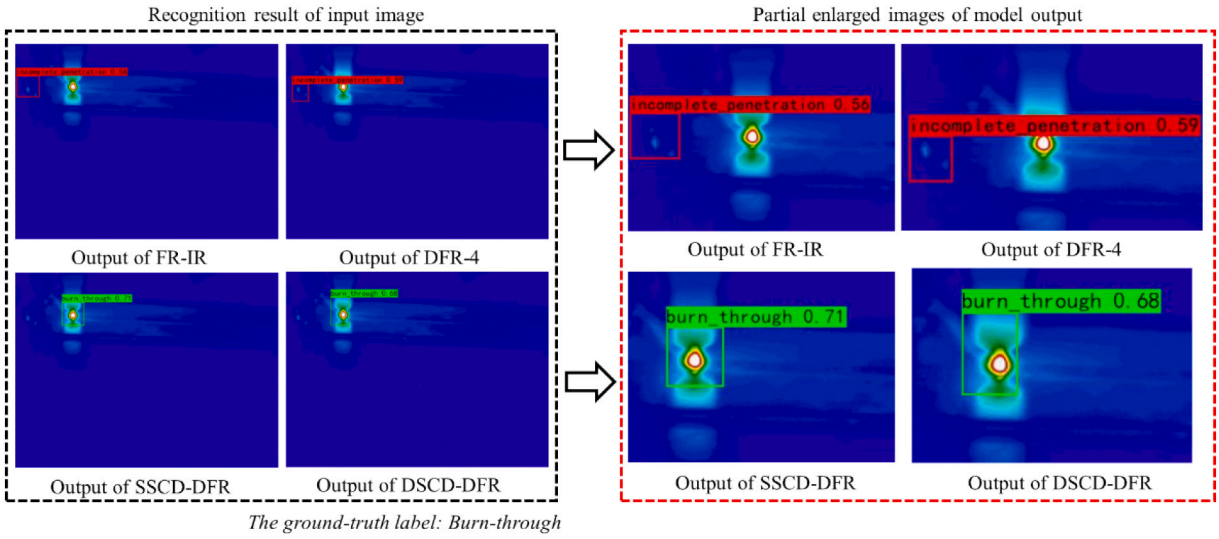


Fig. 12. Output comparison of FR-IR, DFR-4, SSCD-DFR and DSCD-DFR.

and CCD original image as input, which shortens the data set production cycle and avoids the extra error during image segmentation or other preprocessing. Compared with the penetration state recognition model based on Faster R-CNN with IR or CCD images as input, Dual-input Faster R-CNN extracts features from both IR and CCD images, and combines the features of both images for recognition, which reduces the misjudgment rate of the model. And the output bounding box can be used for ROI selection in other fields to reduce labor. By sharing the RPN and ROI Pooling Layer, the best comprehensive performance of the Dual-input Faster R-CNN model was obtained. The recognition accuracy of the model for IR and CCD data pairs was more than 94%, and the recognition time was 246 ms. Moreover, by applying convolutional descriptor selection to IR image's feature map output from Feature Extractor, the Dual-input Faster R-CNN model enables to confirm the location of the main target better, and it's not easy to be affected by the irrelevant information in the background, which leads to the high recognition accuracy of the model while maintaining a small recognition time. Therefore, by extracting the feature map of IR image and CCD image at the same time, using convolutional descriptor selection, the model can avoid the influence of arc flicker in CCD image and irrelevant information in IR image on the selection of bounding box and recognition accuracy (more than 95%), and improve the robustness of the model significantly.

However, there exists some deficiencies: (1) The frame rate of CCD camera and IR camera is required to be the same, otherwise it is difficult to ensure that the CCD and IR images can be collected in the same sequence. (2) IR camera has a large volume and quality, which is difficult to be placed next to the torch and moves with the torch. Therefore, it is difficult to collect IR and CCD images simultaneously for the welding process with long distance or changeable welding track. (3) Although the model can select bounding box and label the CCD as well as IR images of different sizes input simultaneously, the prediction time of the model is still a little long. To overcome them, in the next step, CCD and IR sensor systems need to be better integrated into the welding system. It is necessary to develop appropriate pan tilt so that the two sensors can synchronously follow the welding torch and accurately collect the image and related information of the molten pool area. Optimize the image acquisition module in the welding system to reduce the number of blank frames. To improve the prediction efficiency and accuracy of the model, the existing model should be optimized. It includes fusing the features of the two images in other ways, simplifying the model structure, reducing the model parameters, and developing more efficient attention method.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under the Grant No. 61873164. The authors would like to thank the editor and anonymous reviewers for their careful review and constructive comments on the earlier versions of this article.

References

[1] Chen SB, Lv N. Research evolution on intelligentized technologies for arc welding process. J Manuf Process 2014;16(1):109–22. https://doi.org/10.1016/j.jmapro.2013.07.002.
[2] Chen SB. Visual information acquirement and real-time control methodologies for weld pool dynamics during pulsed GTAW1. THERMIC 2006 pt.4. Shanghai: Welding Engineering Institute, Shanghai Jiaotong University; 2007.
[3] Rokhlin SI, Guu AC. Computerized radiographic sensing and control of an arc welding process. Weld J 1990;69(3):83–95. https://www.osti.gov/biblio/6816308.
[4] Rokhlin SI, Cho K, Guu AC. Closed-loop process control of weld penetration using real-time radiography. NDT&E Int 1996;29(3):188. https://doi.org/10.1016/0963-8695(96)84917-6.
[5] Mi B, Ume C. Real-time weld penetration depth monitoring with laser ultrasonic sensing system. J Manuf Sci Tech 2006;128(1):280–6. https://doi.org/10.1115/1.2137747.
[6] Kita A. Measurement of weld penetration depth using non-contract ultrasound method. Atlanta: Georgia Institute of Technology; 2005.
[7] Fan H, Ravala NK, Wikle HC, Chin BA. Low-cost infrared sensing system for monitoring the welding process in the presence of plate inclination angle. J Mater Process Tech 2003;140(1–3):668–75. https://doi.org/10.1016/S0924-0136(03)00836-7.
[8] Chandrasekhar N, Vasudevan M, Bhaduri AK, Jayakumar T. Intelligent modeling for estimating weld bead width and depth of penetration from infra-red thermal images of the weld pool. J Intell Manuf 2015;26:59–71. https://doi.org/10.1007/s10845-013-0762-x.
[9] Lv N, Zhong JY, Chen HB, Lin T, Chen SB. Real-time control of welding penetration during robotic GTAW dynamical process by audio sensing of arc length. Int J Adv Manuf Tech 2014;74(1–4):235–49. https://doi.org/10.1007/s00170-014-5875-7.
[10] Lv N, Xu YL, Li SC, Yu XW, Chen SB. Automated control of welding penetration based on audio sending technology. J Mater Process Tech 2017;250(12):81–98. https://doi.org/10.1016/j.jmatprotec.2017.07.005.
[11] Chen SB, Zhao DB, Wu L, Lou YJ. Intelligent methodology for sensing, modeling and control of pulsed GTAW: part 2: butt joint welding. Weld J 2000;74(1–4):235–49.
[12] Chen SB, Lou YJ, Wu L, Zhao DB. Intelligent methodology for sensing, modeling and control of pulsed GTAW: part 1: bead-on-plate welding. Weld J 2000;79(6):151–63.

[13] Fan C, Lv F, Chen SB. Visual sensing and penetration control in aluminum alloy pulsed GTA welding. Int J Adv Manuf Tech 2009;42(1):126–37. https://doi.org/10.1007/s00170-008-1587-1.
[14] Zhang YM, Yang YP, Zhang W, Na SJ. Advanced welding manufacturing — an analysis and review of challenges and solutions. J Manuf Sci Tec 2020;142(11):1–33. https://doi.org/10.1115/1.4047947.
[15] Khaleghi B, Khamis A, Karray FO, Razavi SN. Multisensor data fusion: a review of the state-of-the-art. Inform Fusion 2013;14(1):28–44. https://doi.org/10.1016/j.inffus.2011.08.001.
[16] Giordan D, Notti D, Villa A, et al. Low cost, multiscale and multi-sensor application for flooded area mapping. Nat Hazard Earth Sys 2018;18(5):1493–516. https://doi.org/10.5194/nhess-2017-420.
[17] Ming D, He D. Hidden semi-Markov model-based methodology for multi-sensor equipment health diagnosis and prognosis. Eur J Oper Res 2007;178(3):858–78. https://doi.org/10.1016/j.ejor.2006.01.041.
[18] Liu J, Hu Y, Wang Y, Wu B, Fan JK, Hu ZX. An integrated multi-sensor fusion-based deep feature learning approach for rotating machinery diagnosis. Meas Sci Technol 2018;29(5):12. https://doi.org/10.1088/1361-6501/aaaca6.
[19] Banerjee P, Raj RA. Multi-sensor data fusion strategies for real-time application in test and evaluation of rockets/missiles system. IEEE Int Conf Indust Technol 2000;1:723–8.
[20] Luo RC, Su KL. Multilevel multisensor-based intelligent recharging system for mobile robot. IEEE T Ind Electron 2008;55(1):270–9. https://doi.org/10.1109/TIE.2007.903989.
[21] Chen C, Xiao RQ, Chen HB, Lv N, Chen SB. Arc sound model for pulsed GTAW and recognition of different penetration states. Int J Adv Manuf Tech 2020;108(1–4):3175–91. https://doi.org/10.1007/s00170-020-05462-z.
[22] Chen B, Wang JF, Chen SB. Prediction of pulsed GTAW penetration status based on BP neural network and D-S evidence theory information fusion. Int J Adv Manuf Tech 2010;48(1–4):83–94. https://doi.org/10.1007/s00170-009-2258-6.
[23] Subashini L, Vasudevan M. Adaptive Neuro-Fuzzy Inference System (ANFIS)-based models for predicting the weld bead width and depth of penetration from the infrared thermal image of the weld pool. Metall Mater Trans B 2012;43(1):145–54. https://doi.org/10.1007/s11663-011-9570-x.
[24] Chandrasekhar N, Vasudenan M, Bhaduri AK, Jayakumar T. Intelligent modeling for estimating weld bead width and depth of penetration from infra-red thermal images of the weld pool. J Intell Manuf 2015;26(1):59–71. https://doi.org/10.1007/s10845-013-0762-x.
[25] Ghanty P, Vasudevan M, Mukherjee DP, Pal NR, Chandrasekhar N, Maduraimuthu V, et al. An artificial neural network approach for estimating weld bead width and depth of penetration from infrared thermal image of weld pool. Sci Technol Weld Joi 2008;13(4):395–401. https://doi.org/10.1179/174329308X300118.
[26] Zhang YM, Wu L, Chen DH, Walcott BL. Determining joint penetration in GTAW with vision sensing of weld face geometry. Weld J 1993;72:10.
[27] Liu YK, Zhang YM. Model-based predictive control of weld penetration in gas tungsten arc welding. IEEE T Contr Syst T 2014;22(3):955–66. https://doi.org/10.1109/TCST.2013.2266662.
[28] Kovacevic R, Zhang YM, Li L. Monitoring of weld joint penetration based on weld pool geometrical appearance. Weld J 1996;75(10):317–29.
[29] Jiao WH, Wang QY, Cheng YC, Zhang YM. End-to-end prediction of weld penetration: a deep learning and transfer learning based method. J Manuf Process 2021;63:191–7. https://doi.org/10.1016/j.jmapro.2020.01.044.
[30] Zhang BR, Shi YH, Cui YX, Wang ZS, Hong XB. Prediction of keyhole TIG weld penetration based on high-dynamic range imaging. J Manuf Process 2021;63:179–90. https://doi.org/10.1016/j.jmapro.2020.03.053.
[31] Cheng YC, Wang QY, Jiao WH, Xiao J, Chen SJ, Zhang YM. Automated recognition of weld pool characteristics from active vision sensing. Weld J 2021;100:183–92. https://doi.org/10.29391/2021.100.015.
[32] Cheng YC, Chen SJ, Xiao J, Zhang YM. Dynamic estimation of joint penetration by deep learning from weld pool image. Sci Technol Weld Joi 2021;26(4):279–85. https://doi.org/10.1080/13621718.2021.1896141.
[33] Cheng Y, Wang Q, Jiao W, Yu R, Chen S, Zhang Y, et al. Detecting dynamic development of weld pool using machine learning from innovative composite images for adaptive welding. J Manuf Process 2020;56:908–15. https://doi.org/10.1016/j.jmapro.2020.04.059.
[34] Wang QY, Jiao WH, Zhang YM. Deep learning-empowered digital twin for visualized weld joint growth monitoring and penetration control. J Manuf Syst 2020;57:429–39. https://doi.org/10.1016/j.jmsy.2020.10.002.
[35] Feng YH, Chen ZY, Wang DL, Chen J, Feng ZL. DeepWelding: a deep learning enhanced approach to GTAW using multisource sensing images. IEEE T Ind Inform 2020;16(1):465–74. https://doi.org/10.1109/TII.2019.2937563.
[36] Ahrari AH, Kiavarz M, Hasanlou M, Marofi M. Thermal and visible satellite image fusion using wavelet in remote sensing and satellite image processing. Int Arch Photogram Rem Sens Spat Inform Sci 2017;XLII-4/W4(4):11–5. https://doi.org/10.5194/isprs-archives-XLII-4-W4-11-2017.
[37] Sun F, Yang YZ, Lin CW, Liu ZH, Chi LH. Forest fire compound feature monitoring technology based on infrared and visible binocular vision. J Phys 2021;1792(1):012022 (9 pp), https://doi.org/10.1088/1742-6596/1792/1/012022.
[38] Chen JY, Yang XM, Lu L, Li QL, Li ZY, Wu W. A novel infrared image enhancement based on correlation measurement of visible image for urban traffic surveillance systems. J Intell Transport S 2020;24(3):290–303. https://doi.org/10.1080/15472450.2019.1642753.
[39] Bibby MJ, Goldak JA, Shing GY. A model for predicting the fusion and heat-affected zone sizes of deep penetration welds. Can Metall Quart 1985;24(1):101–5. https://doi.org/10.1179/000844385795448894.
[40] Boo KS, Cho HS. Transient temperature distribution in arc welding of finite thickness plates. P I Mech Eng B-J Eng 1900;204(3):175–83. https://doi.org/10.1243/PIME_PROC_1990_204_005_01.
[41] Fachinotti VD, Amilcar AA, Cardona A. Analytical solutions of the thermal field induced by moving double-ellipsoidal and double-elliptical heat sources in a semi-infinite body. Int J Numer Meth Bio 2011;27(4):595–607. https://doi.org/10.1002/cnm.1324.
[42] Nguyen NT, Ohta A, Matsuoka K, Suzuki N. Analytical solutions for transient temperature of semi-infinite body subjected to 3-D moving heat sources. Weld J 1999;83(3):82–93.
[43] Ren SQ, He KM, Ross G, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. IEEE T Pattern Anal 2015;39(6):1137–49. https://doi.org/10.1109/TPAMI.2016.2577031.
[44] Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. 2014 IEEE conference on computer vision and pattern recognition, doi:https://doi.org/10.1109/CVPR.2014.81;2014 [accessed 28 March 2014].
[45] Wei XS, Luo JH, Wu JX, Zhou ZH. Selective convolutional descriptor aggregation for fine-grained image retrieval. IEEE T Image Process 2017;26(6):2868–81. https://doi.org/10.1109/TIP.2017.2688133.