



A real-time seam tracking system based on extraction and restoration model in robotic GMAW

Xiaohui Zhao¹ · Bin Yang¹ · Ziwei Li¹ · Yongchang Liang¹ · Yupeng Chi¹ · Yunhao Chen¹ · Hao Wang¹

Received: 12 May 2023 / Accepted: 1 January 2024 / Published online: 9 January 2024
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2024

Abstract

Seam tracking technology is an important part of the intelligent welding field. In this paper, a laser vision-based real-time seam tracking system was built. The system consists of a self-developed laser vision sensor, a six-axis robot, a gas metal arc welding system, and an industrial computer. After building the system, the system calibration was performed. During the seam tracking, the arc light, spatter, and other welding noise have a negative impact on the image processing algorithm to extract the weld feature points, and even lead to system drift and algorithm failure. To this end, a two-stage extraction and restoration model (ERM) was proposed for processing real-time welding images to improve the robustness and accuracy of the seam tracking system. In the ERM, the region of interest was first detected and extracted by the YOLOv5s model, then the extracted images were restored by the conditional generation adversarial network. After using the ERM model, a series of image processing was performed to obtain the coordinates of the weld feature points. The total time consumed by the algorithm is 37 ms per frame on average, which meets the real-time requirement. Moreover, the experimental results show that the seam tracking system based on the ERM can achieve real-time tracking for different types of planar V-bevel welds, and the average error is 0.21 mm, which meets the requirements for actual welding.

Keywords Seam tracking · Image processing · Laser vision sensor · Deep learning

1 Introduction

Welding is one of the most important material joining techniques, which has been widely used in aerospace, marine, automotive, nuclear, and other industrial fields [1]. With the development of welding robots, welding technology has been largely automated [2]. At present, automated welding mainly relies on the traditional “teach and playback” to achieve [3, 4]. This model has the disadvantages of being time-consuming to teach and requiring re-teaching for different types of workpieces or after adjusting the position of the workpiece, resulting in low welding efficiency and a large degree of influence by the level or experience of the operator. With the gradual development of the welding industry from automation to intelligence, seam tracking technology has been proposed and has received widespread attention

from industry personnel [5]. The hardware core of weld seam tracking technology is various types of sensors, which control the welding robot to identify the trajectory of the weld to be applied and automatically track the weld according to the information collected by the sensors. According to different sensor types, weld tracking systems can be divided into contact [6] and non-contact types. Contact-type seam tracking systems mainly rely on mechanical probes to achieve high accuracy and sensitivity tracking, but there are shortcomings such as easy wear and bending of probes. The sensors used in the non-contact type of weld tracking system are ultrasonic sensors [7], arc sensors [8], vision sensors, etc. Ultrasonic sensors make use of the time difference of ultrasonic reflection to achieve the detection of the position of the weld seam, but due to the presence of a large number of spatter, noise, and other complicating factors in the welding process, many times it does not work properly. Arc sensors are also divided into rotary and oscillating types, which make use of the changes in the distance between the welding torch and the weld seam during the welding process to cause changes in the arc parameters to achieve the correction of the welding path. The seam tracking system

✉ Hao Wang
hwang22@mails.jlu.edu.cn

¹ Key Laboratory of Automobile Materials, School of Materials Science and Engineering, Jilin University, Changchun 130025, China

based on an arc sensor has the advantages of good real-time performance and strong anti-interference capability but has the disadvantage of only being suitable for medium and thick plates or grooves with large bevel angles.

With the rapid development of computer vision technology, vision sensors have become the most popular sensors in the field of weld seam tracking due to their advantages such as contactless and adaptability to different environments. Vision sensors are divided into active vision and passive vision, where active vision sensors project structured light onto the object, whereas passive vision relies directly on charge coupled device (CCD) cameras or complementary metal oxide semiconductor (CMOS) cameras for image acquisition from the target. Xu [9] et al. proposed a seam tracking system that uses an improved Canny edge detection algorithm to process the melt pool images captured by the passive vision to achieve a certain accuracy. H.N.M. Shah [10] et al. proposed a seam tracking system that uses a local threshold segmentation algorithm to identify different shapes of butt welds. Compared with passive vision, active vision has the advantages of concentrated information, lower image processing difficulty, and higher accuracy, and has become a mainstream research direction. The hardware core of the active vision system is the laser vision sensor, which can also be divided into many kinds according to the type of laser lines, such as dot [11], single line [12, 13], triple line [14, 15], cross [16], and circle [17].

The accuracy of the image processing algorithm for weld feature point recognition directly affects the accuracy and robustness of seam tracking. The research concern of image processing algorithms for active vision systems is how to accurately extract feature points from laser streaks under the interference of noise such as arc light, spatter, and fume during welding. Zhang [18] et al. scanned the weld seam with a laser sensor and used a second-order derivative algorithm followed by fitting to locate the feature points to achieve tracking of straight and curved weld seams. Xiao [19] et al. improved the Snake model as a stripe feature point extractor and proposed a feature extraction algorithm based on the improved Snake model. Fan [20] et al. used Hough transform and least-squares fitting to extract weld feature points accurately.

In recent years, as machine learning and deep learning are more and more widely used in the field of computer vision, various excellent deep learning algorithms have emerged, and these methods are gradually applied to the field of seam tracking. For example, Chen [21] et al. used a high dynamic range CMOS camera to acquire images and identified the keyhole entrance using the Mask-RCNN model in their study of Keyhole TIG seam tracking which extracted the coordinates of the melt pool centroid and obtained the weld deviation. Zou [22] et al. combined convolutional filters and deep reinforcement learning to locate weld feature points for seam

tracking. Lin [23] et al. used the YOLOv5 model to achieve automatic detection of the keyhole entrance and used the center of the bounding box as the position of the torch in the image. Liu [24] et al. built a CGAN-based restoration extraction network for extracting weld feature points to achieve multi-layer multi-pass seam tracking.

Among them, YOLO and CGAN are representative and promising models. You Only Look Once (YOLO) [25] is a deep learning object detection model with excellent performance, which is characterized by being fast and accurate after training with small samples, enabling real-time object detection. The YOLO series of models are different from the traditional R-CNN series of models in that it only needs to scan the image once, so it is called the one-stage model, while the latter is called the two-stage model. So the YOLO model has a significant advantage in speed compared with other object detection algorithms. There are many versions of the YOLO model, among which YOLOv5 [26] is the most developed and improved version for the current application. Generative adversarial network (GAN) [27] is designed based on the idea of game theory. It is one of the important deep learning algorithms and has been widely used in image generation, but it cannot control the patterns of the images being generated. Therefore, conditional generative adversarial networks (CGAN) [28] are proposed. The generator and discriminator of CGAN are confronted with additional information y , which can be various other forms of data such as labels, so CGAN is widely applied in image restoration. The objective function of CGAN is shown in Eq. (1).

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log (1 - D(G(z|y)))] \quad (1)$$

where $V(D, G)$ is the objective function, \mathbb{E} is the expectation, $p_{\text{data}}(x)$ is the training data set, $p_z(z)$ is the random noise distribution, $D(x)$ is the probability that the sample is from the training dataset, and $D(G(z))$ is the probability that the sample is generated by the generator. y is the additional information as condition.

Overall, YOLO has the advantage of fast extraction of regions of interest (ROI) in applications, but it cannot improve the accuracy of extracted feature points. While CGAN can effectively improve the accuracy of extracted feature points through image restoration, it is very time-consuming itself.

1.1 In order to extract the feature points from the real-time seam tracking images more quickly and accurately

In this paper, in order to ensure sufficient tracking accuracy while meeting real-time requirements (i.e., algorithmic

speed), (1) an extraction and restoration model (ERM) is proposed in this paper, which adopts the two-stage method. Firstly, the YOLOv5s model is used to extract the ROI and crop from the images. Then, the CGAN model is used to perform image restoration on the cropped images. Finally, the welding images are cropped and restored, making extraction of the feature points more quickly and accurately. (2) According to the proposed ERM, a laser vision-based seam tracking system is built. And real-time seam tracking of different weld types with V-bevels in robotic gas metal arc welding (GMAW) is realized.

This paper is organized as follows: Section 2 introduces the establishment and calibration of the experiment system. Section 3 proposes a two-stage extraction and restoration model to extract ROI and restore images. Section 4 presents the subsequent image processing process and results after ERM. Section 5 introduces the experiments and analyses to verify the established system's feasibility, accuracy, and robustness. The conclusion of this paper is shown in Section 6.

2 Seam tracking system

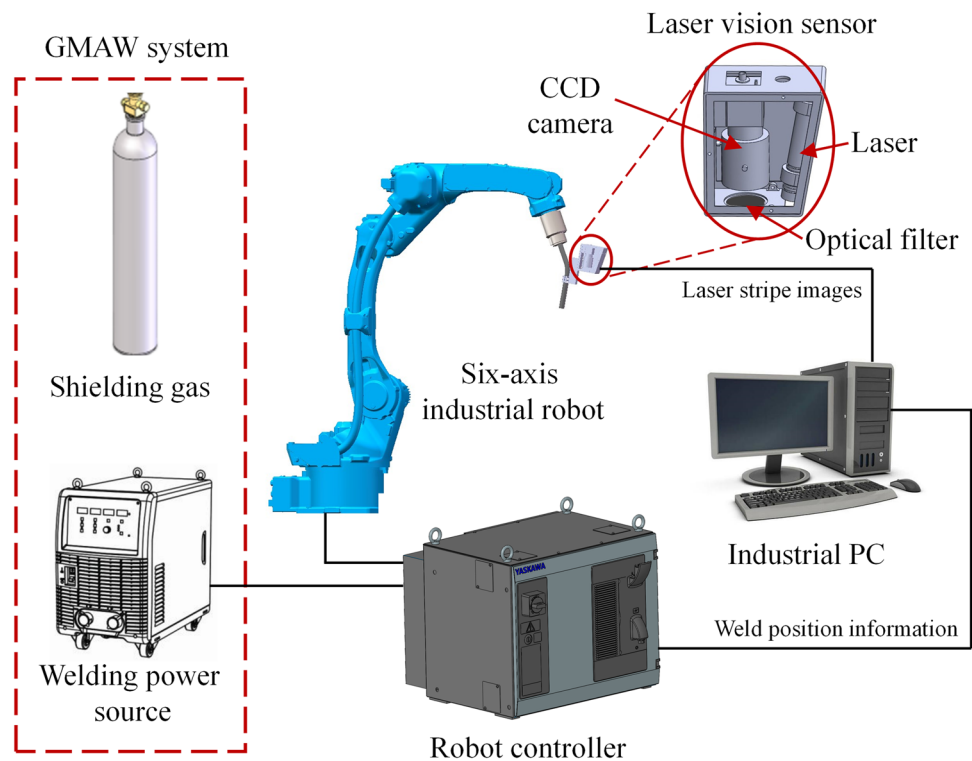
2.1 Robotic GMAW seam tracking system

The schematic diagram of the robotic seam tracking system built in this paper is shown in Fig. 1. The system consists of

a six-axis industrial robot, a gas metal arc welding (GMAW) system, an independently developed laser vision sensor (LVS), and an industrial personal computer (IPC). Among them, the LVS is the most critical and core hardware component of the whole system. As shown in Fig. 2, the LVS is mounted on the welding torch at the end of the robot arm using a fixture. The LVS mainly consists of a single-line laser, a CCD camera, and an optical filter, with the laser and the CCD camera placed in an oblique-direct manner. In the actual process of welding, the CCD camera is about 150 mm high from the workpiece, and the laser streak is 40 mm away from the welding wire. The detailed parameters of the laser are shown in Table 1. The CCD camera captures the weld image at a maximum frame rate of 60 frames per second (fps) with a maximum resolution of 2592×2048 to meet the requirements of real-time seam tracking. According to the spectral characteristics of the arc light of the GMAW, a narrow band filter with a central wavelength of 635 nm is selected and placed at the CCD camera lens for filtering to reduce the interference of arc and natural light.

The flow chart of the seam tracking system built is shown in Fig. 3. Before weld seam tracking begins, the parameters obtained from the system calibration need to be set in the parameter settings. Afterward, the initial welding position is arrived at according to the preset robot teaching trajectory. Then, arc on, and the laser vision sensor sends the laser stripe images collected in real-time to the IPC during the welding process. The IPC extracts the weld seam feature points through

Fig. 1 Schematic diagram of the robotic GMAW seam tracking system



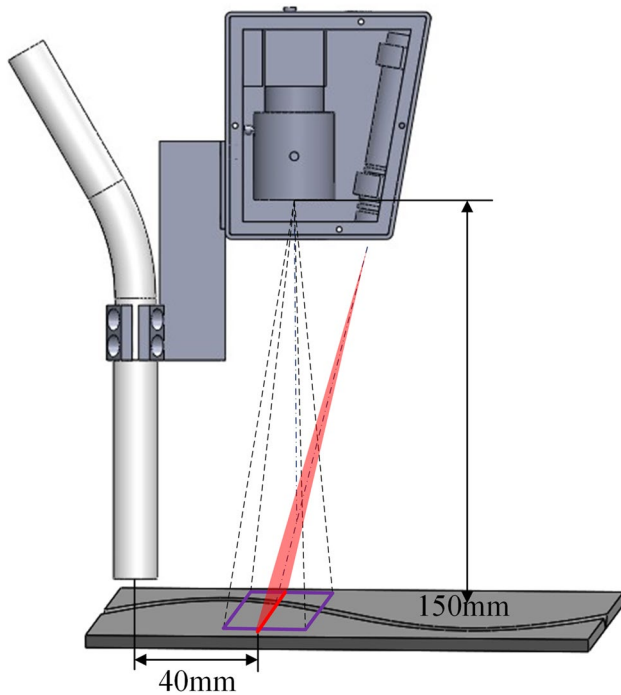


Fig. 2 Laser vision sensor

the image processing algorithms and transforms the feature points into robot coordinates for transmission back to the robot control cabinet. The robot is controlled to move according to the coordinate. And, the process is repeated for the next frame. The robot realizes smooth movements by interpolating and fitting between the coordinates. Eventually, the final welding position is arrived, arc off, and the weld seam tracking is finished.

2.2 System calibration

The whole system needs to be calibrated after establishment. The system calibration mainly includes camera calibration, light plane calibration, and hand-eye calibration. The purpose of the system calibration is to convert the pixel coordinates on the image to the spatial 3D coordinates in the robot base coordinate system. The transformation is achieved by modeling multiple sets of coordinate transformation relationships and solving for the camera internal parameters, the light plane equations, and the hand-eye matrix.

Firstly, to simplify the model, camera lens distortion is not considered. The transformation relationship between the pixel coordinates (u, v) of the image and the camera coordinates (X_C, Y_C, Z_C) can be obtained based on the rectilinear projection model, as shown in Eq. (2).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \frac{1}{Z_C} \begin{bmatrix} f/dx & 0 & u_0 \\ 0 & f/dy & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} = K \begin{bmatrix} X_C/Z_C \\ Y_C/Z_C \\ 1 \end{bmatrix} \quad (2)$$

where f is the camera focal length. dx and dy are the actual distances represented by one pixel in the x -axis and y -axis directions, respectively, and (u_0, v_0) are the pixel coordinates of the camera's optical center in the imaging plane. K is the internal parameter matrix of the camera. The K can be obtained by the calibration method proposed in [29].

Because Eq. (2) lacks a constraint, the light plane equation is introduced to obtain the exact mapping relationship, as shown in Eq. (3).

$$Z_C = AX_C + BY_C + C \quad (3)$$

where A , B , and C can be calculated from [30].

The simultaneous equations (Eq. 4) are obtained by considering the above equations.

$$\begin{cases} X_C = Z_C(u - u_0)dx/f \\ Y_C = Z_C(v - v_0)dy/f \\ Z_C = Cf / (f + Adx(u - u_0) + Bdy(v - v_0)) \end{cases} \quad (4)$$

Finally, the transformation relationship between the world coordinates (X_W, Y_W, Z_W) and the camera coordinates (X_C, Y_C, Z_C) is obtained after the hand-eye calibration, as shown in Eq. (5).

$$\begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} = \begin{bmatrix} R_W & T_W \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} R & T \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = P_W P \begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} \quad (5)$$

where P_W , P are composed of rotation matrix and translation vector, P_W are obtained from the demonstrator reading, and the hand-eye matrix P can be calculated from [31].

After the system calibration is completed, the transformation relationship between 2D pixel coordinates and 3D spatial coordinates has been established, so the actual coordinates of the weld feature points can be obtained from the real-time welding images. The system calibration results are shown in Table 2.

Table 1 Laser parameters

Size	Laser type	Wavelength	Linewidth	Lens type	Lens angle
Φ16 mm×90 mm	Uniform and straight	638 nm	0.08 mm	Powell lens	30°

Fig. 3 Flow chart of the seam tracking system

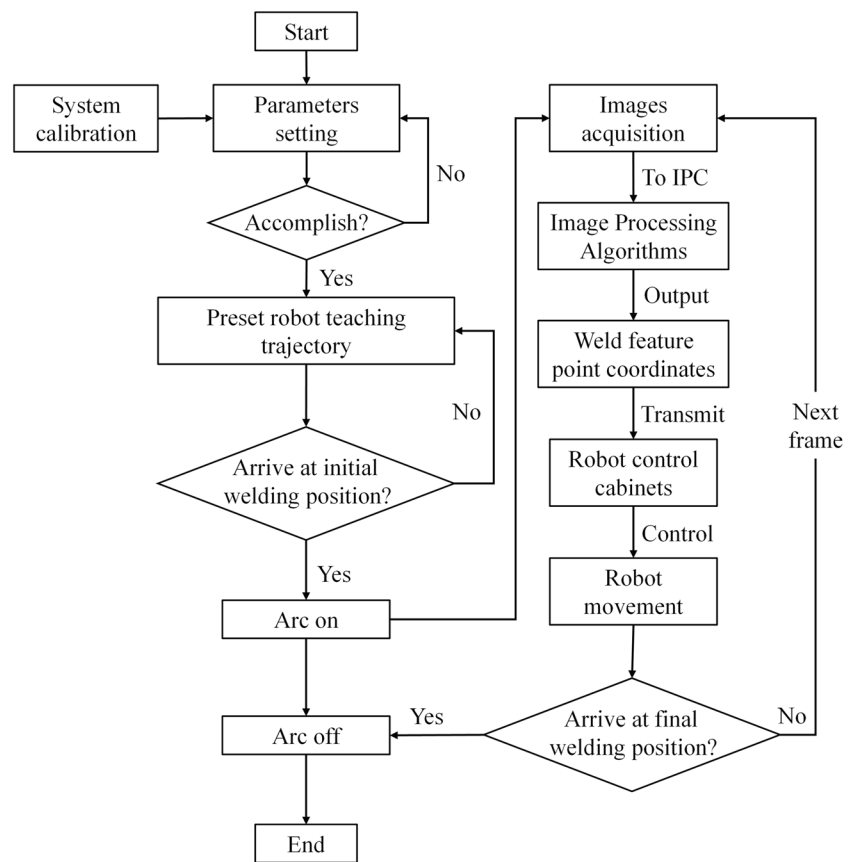


Table 2 System calibration results

Parameter matrix	Calibration result
Camera internal parameters matrix	$K = \begin{bmatrix} 6.1341 \times 10^3 & 0 & 1.3075 \times 10^3 \\ 0 & 6.1329 \times 10^3 & 1.0333 \times 10^3 \\ 0 & 0 & 1 \end{bmatrix}$
Hand-eye matrix	$P = \begin{bmatrix} 5.707308 \times 10^{-2} & -9.9796973 \times 10^{-1} & 2.8268192 \times 10^{-2} & 80.250241 \\ 9.9823469 \times 10^{-1} & 5.6576309 \times 10^{-2} & -1.807301 \times 10^{-2} & -0.7982302 \\ 1.643701 \times 10^{-2} & 2.9249772 \times 10^{-2} & 9.9943698 \times 10^{-1} & -150.04151 \\ 0 & 0 & 0 & 1.0 \end{bmatrix}$
Laser plane equation	$Z_c = 0.044715X_c + 2.166656Y_c + 209.899414$

3 The two-stage extraction and restoration model

Due to the extraction of weld feature points from real-time welding images by image processing algorithms would be greatly affected by noises such as arc light, spatter, reflection, and fume, (as shown in Fig. 4). A two-stage extraction restoration model (ERM) is proposed for processing real-time welding images before image processing. The model is divided into two stages. Firstly, the first stage extracts the ROI from the original images acquired by LVS to minimize the influence of noise such as arc light,

fume, and splash. Then the second stage carries out image restoration on the images with ROI has been extracted to eliminate the remaining noises such as splash and fume. Finally, the new image was obtained. The schematic diagram of ERM is shown in Fig. 5.

3.1 Region of interest extraction

The YOLOv5s model was used in the first stage to complete the ROI extraction. Depending on the width and depth multiples of the model, YOLOv5 can be classified into five versions from small to large: YOLOv5n, YOLOv5s, YOLOv5m, YOLOv5l, and YOLOv5x. The performance

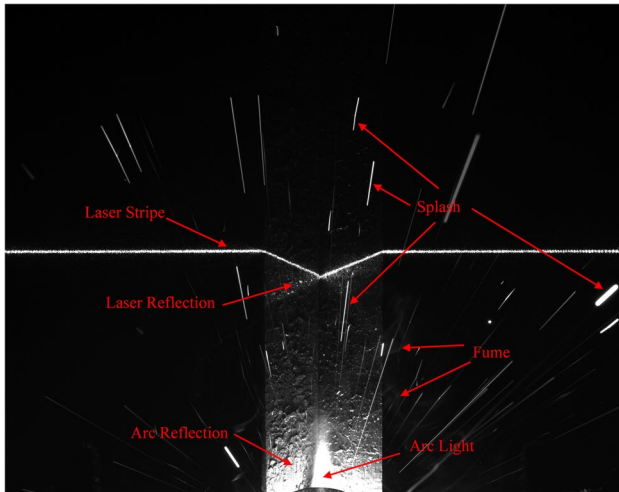


Fig. 4 Real-time welding images

of each version is not the same, so we tested based on a small number of datasets. The smaller the width and depth multiples, the faster the detection speed, but the lower the detection accuracy. Except for YOLOv5n, which has insufficient accuracy, the detection accuracy of all other versions meets our requirements. Therefore, the fastest detection speed YOLOv5s is chosen to reduce the processing time of the whole model and to satisfy the real-time capability.

YOLOv5s consists of four sections, namely input, backbone, neck, and head. The input section preprocesses the images with pixel and channel transformations and uses Mosaic data enhancement to extract four images randomly scaled, cropped, and lined up for stitching. In this way, not only the diversity of the dataset can be enriched but also the training burden can be reduced. Later, the image enters the backbone section, which consists of the focus structure, CBL module, Cross Stage Partial Network (CSPNet) [32], and spatial pyramid pooling (SPP) [33]. The focus structure is to slice the images, sample the pixels at intervals, get four images with similar and complementary features, and concat them before performing the CBL operation, obtaining a feature map with no information loss. Where the CBL module is the integration of convolution (Conv), batch normalization (BN), and activation function Leaky rectified linear unit (LeakyReLU). CSPNet splits the feature map for processing and then concat with the unprocessed part, which

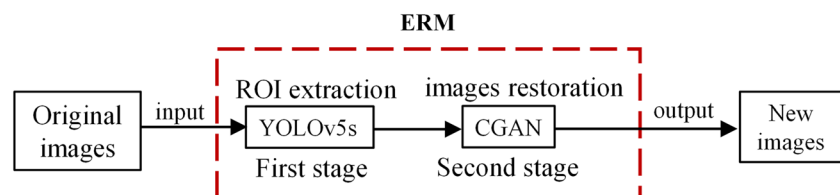
can improve the model operation speed without affecting the result. There are two CSPNet structures in YOLOv5: CSP1×n and CSP2×n, which are used in backbone and neck, respectively. The neck section that follows mainly consisted of Feature Pyramid Network (FPNet) and Path Aggregation Network (PANet) [34]. FPNet is top-down to transmit the feature information of the upper layer by upsampling for fusion, and PANet is bottom-up to transmit the localization information of the lower layer by downsampling for fusion, cooperating with each other to realize the feature map fusion of different layers. The body of the head section is mainly detect, which uses anchors to detect objectives on feature maps of different scales. The structure of YOLOv5s used in our work is shown in Fig. 6.

In order to train the YOLOv5s model, a large number of real-time welding images are required as a dataset. The dataset of real-time welding images is obtained by recording through LVS and then splitting the frames. We used Labelling to label 2000 images, of which 1500 images were used for training and 500 images for validation. The image size is 2592×2048. The hardware settings of the computer used for training are Intel Core i5-12400F CPU, NVIDIA GeForce RTX 3070 Ti 8G GPU, and 16GB RAM. The model training results are shown in Table 3. Precision, Recall, F1 Score, and mAP are the common metrics used to evaluate object detection models. Precision means the proportion of all predictions that are correctly predicted. Recall means the proportion of all true values that are correctly predicted. F1 score is the combination of precision and recall into one metric, numerically defined as their harmonic mean. mAP@0.5 is the mean of Average Precision values when the intersection of union (IoU) is greater than or equal to 0.5. After the model has been trained, ROI detection is performed on the new real-time welding images which are 2592×2048 in size. The average detection speed is 7 ms per frame, and the maximum detection frame rate is up to 142 fps. ROI detection and extraction results are shown in Fig. 7.

3.2 Images restoration

CGAN model was used in the second stage to complete the image restoration. Based on the basic principles of GAN and CGAN, in order to restore the real-time welding image after the extraction of ROI, the input of random noise z is canceled and fed into the generator G as the model condition

Fig. 5 Schematic diagram of the extraction and restoration model



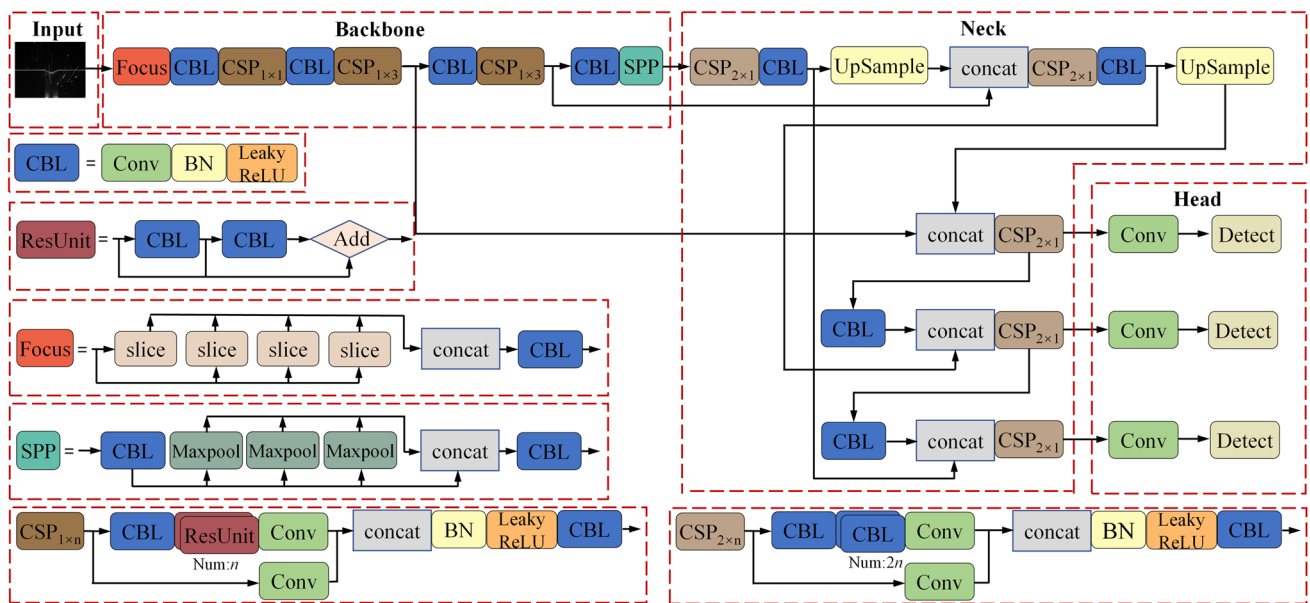


Fig. 6 The structure of YOLOv5s used in the ERM

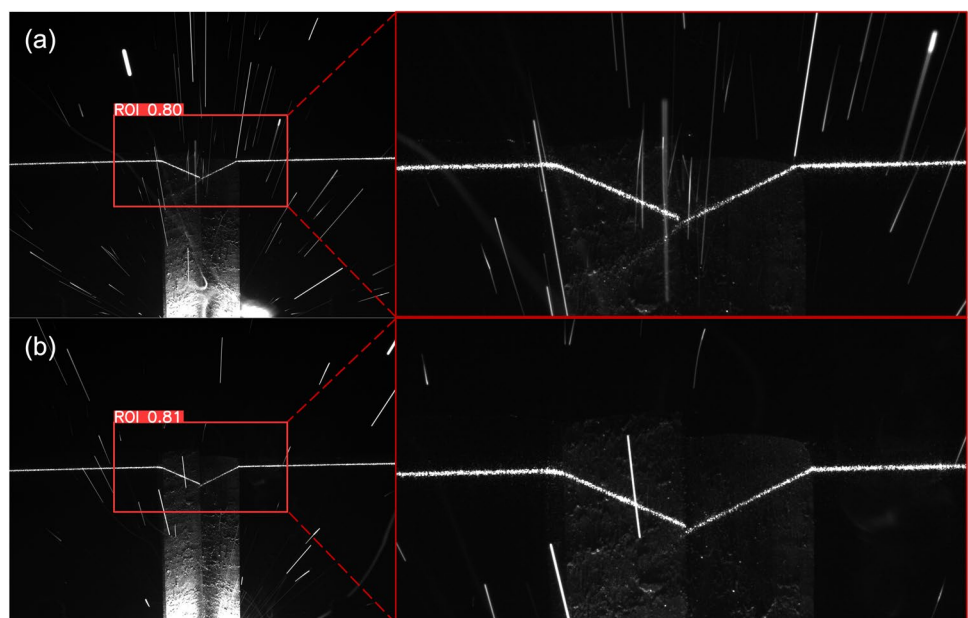
Table 3 Training results of the YOLOv5s model

Object	Precision	Recall	F1 score	mAP@0.5	Average detection speed
ROI	0.979	0.94	0.959	0.977	7 ms/frame

y , as shown in Fig. 8. The laser streak images without noise such as arc light and splash are input to the discriminator D as x along with $G(y)$ generated by G .

The structure of the CGAN model used in this paper consists of a generator D and a discriminator D . The trained generator enables the restoration of real-time welding images and eliminates the effect of noise on subsequent weld feature point extraction. Due to the more layers of the model, the slower the model training and convergence speed will be, and even the restoration speed will be affected. And the fewer layers will lead to poor final restoration results. Therefore, the generator is constructed using six convolutional layers and six deconvolutional layers, as shown in

Fig. 7 ROI detection (a, b) and extraction results



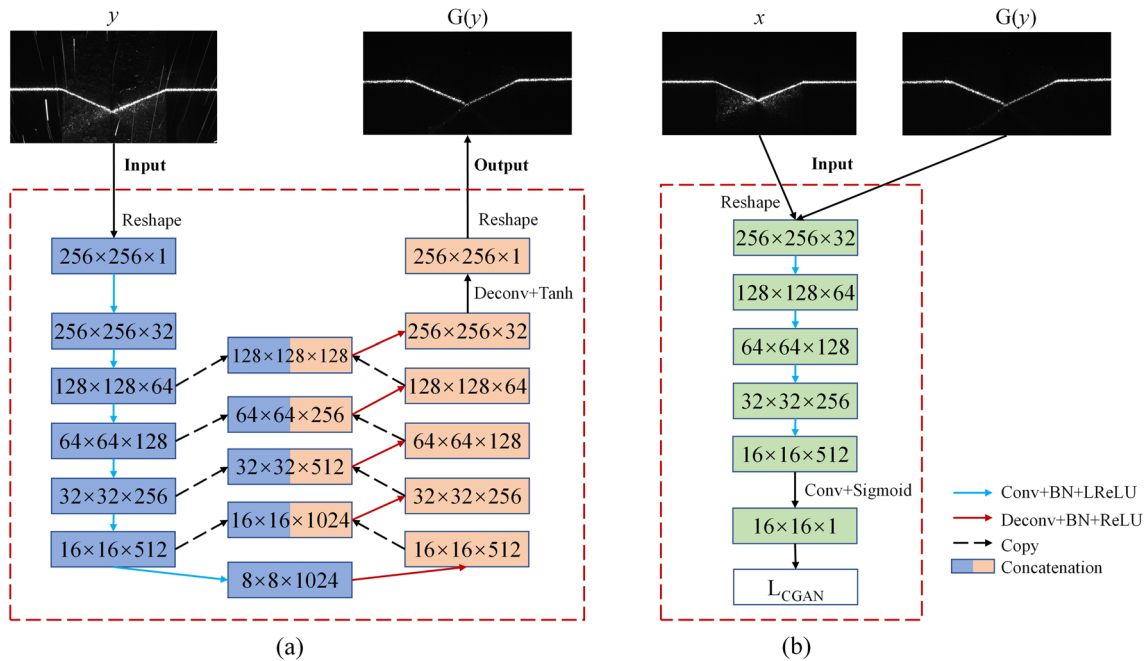


Fig. 8 The structure of the generator G (a) and the discriminator D (b) used in the ERM

Fig. 8(a). Each convolutional layer is after convolution, and then BN and LeakyReLU activation are performed. Whereas the deconvolution layers are different: the first five layers are deconvolved followed by BN and ReLU activation. And the last deconvolution layer, no more BN is performed after deconvolution and Tanh activation is used. The discriminator uses five convolutional layers, as shown in Fig. 8(b). The structure of the first four layers is the same as the convolution layer of the generator, but the last layer is no longer BN after convolution and sigmoid activation is used. In the entire CGAN model for each convolution and deconvolution layer, the size of the kernel used for the convolution process is 3×3 and the stride is 1.

In order to train the generator and discriminator of the CGAN model, a large number of real-time welding images and laser streak images in the noise-free case need to be acquired as two types of datasets for the model condition y and discriminator data x , respectively. The two types of images were obtained by robot teaching and playback, using LVS for video recording without and with arc start welding, and then splitting the frames. We respectively selected 1000 images as the required datasets for training and performed ROI extraction based on the trained YOLOv5s from 3.1 for the two datasets. ROI extraction is performed on the datasets in order to do the following:

1. keep the trained generator from changing the shape and position information of the laser stripes in the original image;

2. cropping the images can reduce the training burden and increase the restoration speed in order to meet the real-time performance; and

3. make the trained generator which can directly restore the ROI extracted by YOLOv5s to realize the coupling of ROI extraction-image restoration in order to build the extracted and restoration model.

The hardware settings of the computers used for training are shown in Section 3.1. After the model has been trained, the real-time welding images after ROI detection are restored which are about 950×500 in size. The average restoration speed was of 25 ms per frame and the maximum restoration frame rate of 40 fps. The image restoration results of the generator after being trained are shown in Fig. 8, and the discriminator is only involved in the training process.

4 Image processing after ERM

The transformation of the real-time welding images captured by the LVS into spatial coordinates for controlling the robot motion is the essential step in the seam tracking technology. This step is achieved by extracting the pixel coordinates of the weld feature points in the image through image processing techniques, then transforming them into spatial coordinates according to the transformation relations obtained in Section 2.2. The image processing process is the most critical step in this process, and the accuracy of the processing

result directly affects the tracking accuracy and the robustness of the tracking system.

As shown in Fig. 9, the image processing process of the seam tracking system after ERM is divided into two parts: pre-processing and post-processing. The original images (Fig. 4) acquired by the LVS in real-time are first entered into the proposed ERM in this paper for ROI extraction and image restoration, and the results are shown in Fig. 10(a). After this, a series of processing is performed to remove interference information. The images need to be grayscale transformation, smooth filtering, threshold segmentation, and morphological processing. The CCD camera within the

LVS used in this paper is a grayscale camera, so the grayscale transformation is not required. The smooth filtering process used a median filter with a filter kernel size of 3×3 , as shown in Fig. 10(b). The threshold segmentation used the OTSU method to determine the threshold value, and the segmentation results are shown in Fig. 10(c). Morphological processing was performed using an opening operation with a kernel size of 3×3 and region connecting. Among them, the opening operation eliminates reflections, etc. that cannot be filtered out by the filtering, as shown in Fig. 10(d); the region connecting reconnects the disconnected parts caused by the corrosion in the opening operation, as shown in Fig. 10(e).

Fig. 9 Flow chart of image processing

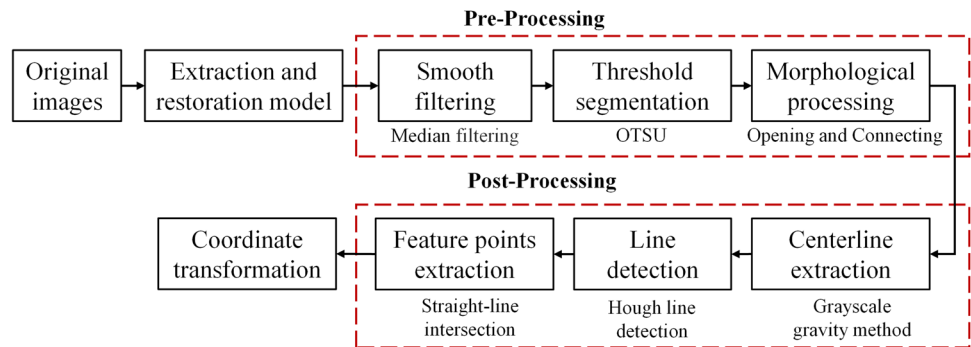
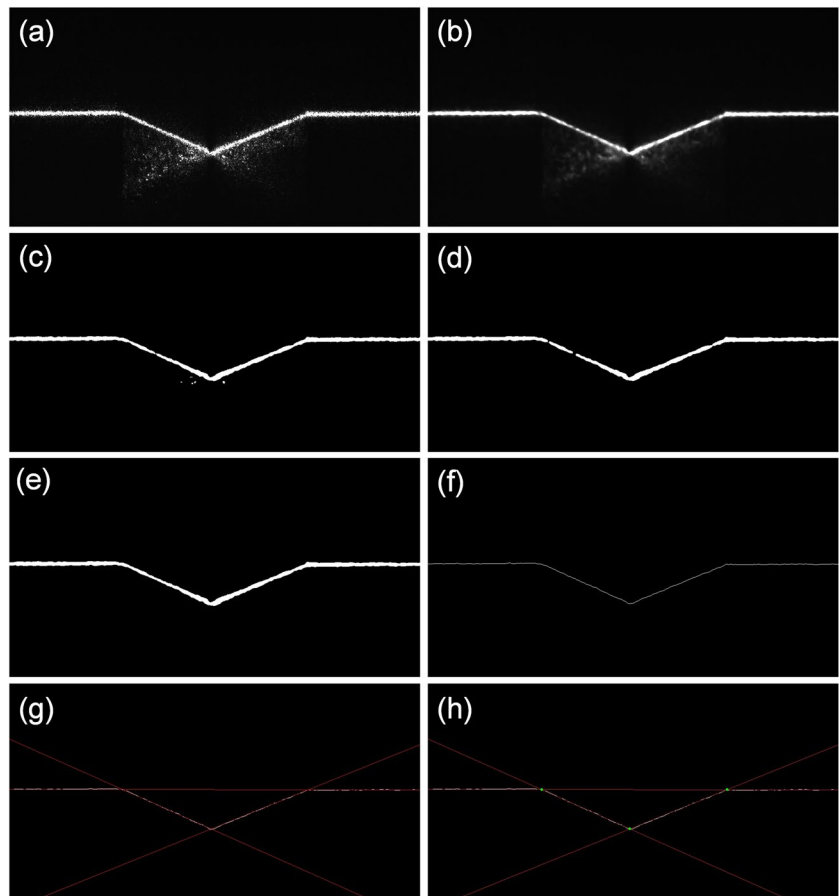


Fig. 10 Image processing results of **a** ERM, **b** smooth filtering, **c** threshold segmentation, **d** opening, **e** connecting, **f** centerline extraction, **g** line detection, and **h** feature points extraction



After the pre-processing, all the noise has been removed except for the backbone part of the laser stripe. The average running speed is 2 ms per frame for image pre-processing.

The image post-processing section implements the extraction of weld feature points for the image after pre-processing (Fig. 10(e)). We performed three steps of centerline extraction, line detection, and feature points extraction for the laser stripes. The centerline extraction used the grayscale gravity method, which converts the laser stripes into single-pixel lines according to the grayscale distribution without losing the feature information, as shown in Fig. 10(f). After getting the single-pixel lines, the Hough straight-line detection method was used to detect it in the straight-line, and the detection threshold was adjusted to obtain three straight lines as in Fig. 10(g). After that, the feature points are extracted by the method of intersection of three lines two by two, and the three feature points shown in Fig. 10(h) are obtained, which are the bevel feature points on the left, right, and bottom. The average running speed of image post-processing is 3 ms per frame.

Eventually, the image processing results are shown in Fig. 10. The whole image processing process transforms the real-time welding image into feature point coordinates with an average transformation speed of 37 ms per frame in total and a maximum frame rate of about 27 fps, which basically meets the real-time requirements. And the frame rate of the image processing algorithm can also be improved by turning down the camera resolution in the LVS to reduce the number of pixels in the original image.

5 Experiments and analysis

5.1 Model comparison experiments

The model parameters are fixed in the actual seam tracking after the ERM training is finished. We compared the ERM proposed in this paper, YOLOv5s, CGAN, and no model in some comparative experiments under the same welding image dataset. The images of the dataset are 2592×2048 in size. The process of the comparison experiments is as follows: firstly, the welding images are processed using different types of models, and subsequently, the images processing in Section 4 (Called IP in Table 4 and Fig. 11) was performed to extract the weld feature points.

The time-consuming of the different methods are shown in Table 4. Since ERM and YOLOv5s extracted the ROI and the image size was significantly reduced, the time-consuming was much less compared to CGAN and none model to extract the weld feature points. And the results of the different methods to extract the weld feature points are shown in Fig. 11. The accuracy of ERM and CGAN to extract weld feature points was higher compared to YOLO and none

Table 4 Time-consuming of the different methods

Items	Methods	Image size	Average time/ms
1	ERM+IP	~950×500	37
2	YOLOv5s+IP	~950×500	13
3	CGAN+IP	2592×2048	182
4	None+IP	2592×2048	32

model. Because they restored the images so that the effects of noise such as arc light and splash were reduced to a low level. From Fig. 11(b), (d), it can be seen that YOLOv5s and none model extract feature points can even be said to fail. To summarize, the ERM has a significant advantage in terms of accuracy and robustness in extracting weld feature points, even though the average time is slightly longer than that of the YOLOv5s and none model.

5.2 Seam tracking experiments

The experimental platform is the robotic seam tracking system built in this paper, as shown in Fig. 12. The six-axis industrial robot used is a floor-mounted Yaskawa robot, MOTOMAN-GP12/AR1440, with ±0.08 mm repeatable positioning accuracy, equipped with a robot controller YRC1000. The welding machine used in the GMAW system is RD350S, and the shielding gas is 80% Ar+20% CO₂.

Real-time seam tracking experiments were performed on the proposed system using V-bevel workpieces with three different types of welds. The workpiece includes three types of straight welds, curved welds, and folded welds, as shown in Fig. 13. The specific welding parameters of GMAW are shown in Table 5. The image acquisition frame rate of the camera was adjusted to 20 Hz in order to reduce the pressure on the algorithm while ensuring real-time performance. Since the tracking experiments for all three welds were performed on a near-horizontal welding platform with essentially no variation in the Z-axis direction, only the variation of the trajectory point coordinates in the horizontal X-O-Y plane was studied. The tracking trajectories of the three types of welds are shown in Fig. 14(a1), (b1), and (c1), which are basically consistent with the weld geometry. The tracking error in Y-direction is shown in Fig. 14(a2), (b2), and (c2). The error of all three kinds of welds is concentrated within ±0.4 mm, and the average error is 0.21 mm, which meets the seam tracking accuracy requirement. The results of the seam tracking are shown in Fig. 15, which proves that the seam tracking system established in this paper has a good tracking result for different shapes of planar V-bevel workpieces, and the welding results are also ideal and meet the actual welding requirements.

Fig. 11 The results of the **a** ERM+IP, **b** YOLOv5s+IP, **c** CGAN+IP, and **d** none+IP to extract the weld feature points

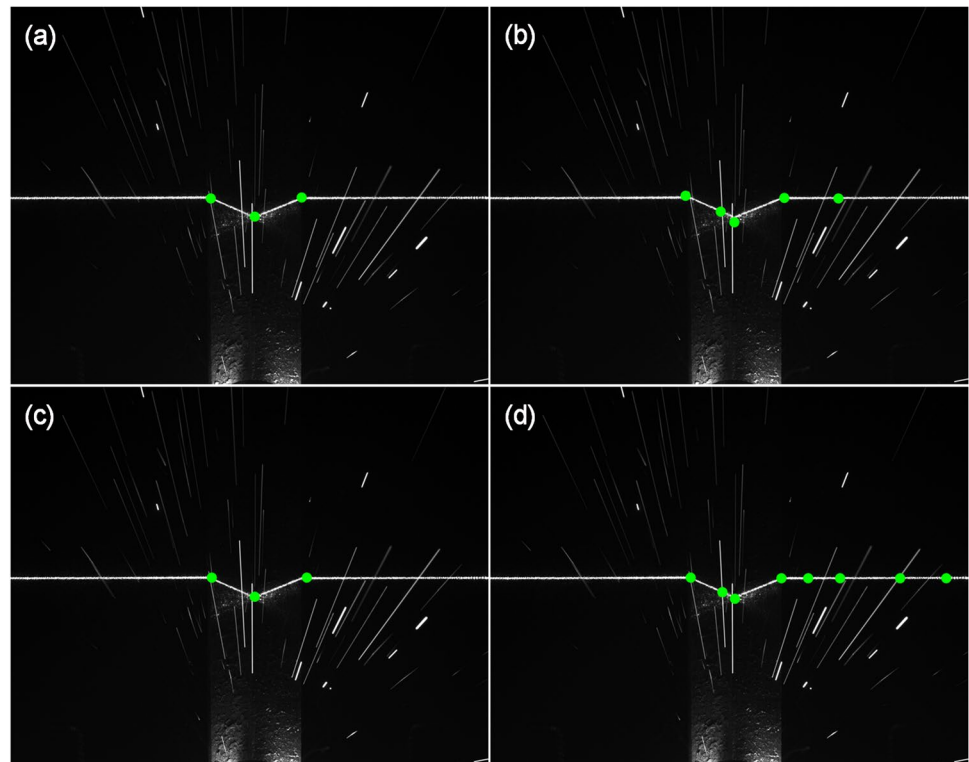
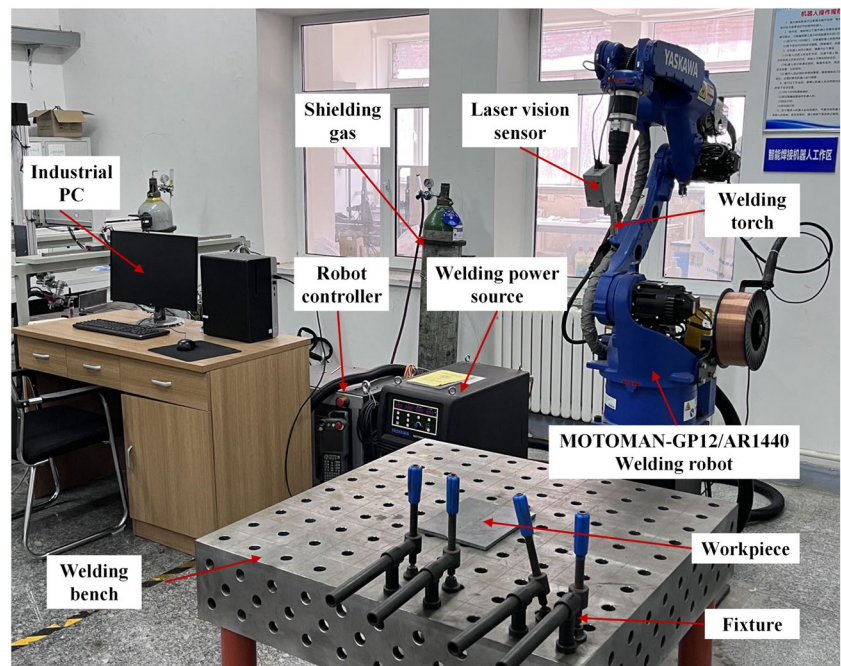


Fig. 12 Experimental platform



6 Conclusion

In this paper, the research on laser vision-based seam tracking technology started from three modules: hardware system construction, system calibration, and image

processing. A deep learning-based seam tracking system was proposed and experimentally verified. The conclusions of this paper are as follows:

1. This paper built a laser vision-based seam tracking system for robotic GMAW, independently designed and developed a laser vision sensor, and calibrated the whole tracking



Fig. 13 Experimental workpieces of **a** straight, **b** curved, and **c** folded types

Table 5 Welding parameters of GMAW

Welding method	Current	Voltage	Welding speed	Wire feeding speed	Gas flow rate
GMAW	80 A	16.7 V	30 cm/min	1.9 m/min	15 L/min

system to realize the transformation from two-dimensional pixel coordinates to three-dimensional spatial coordinates.

2. In order to solve the problem of the influence of various noises on real-time welding images on the extraction of weld feature points, a deep learning-based extraction and restoration model was proposed, which uses the two-stage method

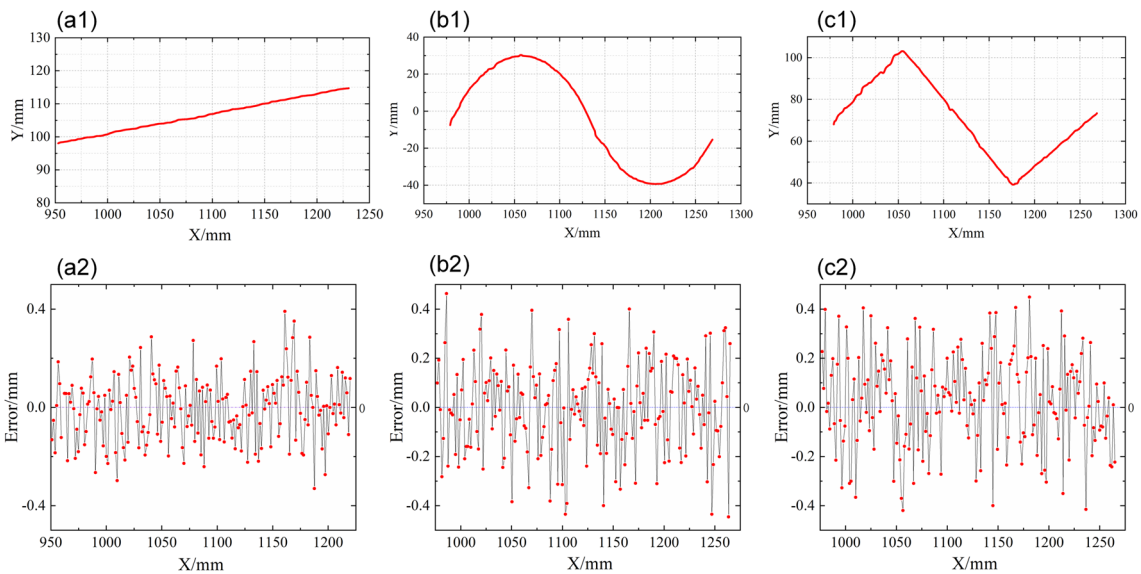


Fig. 14 Seam tracking trajectories and errors



Fig. 15 Seam tracking results

of YOLOv5s-CGAN to realize ROI extraction and image restoration of real-time welding images. The model reduces the influence of noise such as arc light, splash, and smoke on subsequent algorithms and ensures the tracking accuracy and robustness of the seam tracking system.

3. The image processing algorithms after ERM are divided into two parts: pre-processing and post-processing. The pre-processing used smooth processing, threshold segmentation, and morphological processing. A series of processing was performed on the images after ERM to remove interference information. The post-processing used center-line extraction, straight-line detection, and intersection to extract the weld feature points. For the whole image processing process, the average processing frame rate is about 27 fps, and the algorithms meet the real-time requirements of tracking.

4. After experimental verification, the built seam tracking system can achieve real-time tracking of different types of planar V-bevel welds, and the average error is 0.21 mm, which meets the actual welding requirements.

Author contribution Xiaohui Zhao: validation original draft, conceptualization, project administration, and supervision. Bin Yang: background research and investigation, analysis, writing and validation original manuscript, methodology, software, editing. Ziwei Li: editing, validation original draft, supervision. Yongchang Liang: validation original draft, software, investigation. Yupeng Chi: investigation, data curation. Yunhao Chen: investigation. Hao Wang: validation original draft, conceptualization, supervision. All authors read and approved the final manuscript.

Funding This work was supported by the National Natural Science Foundation of China (No. 52275338) and the Innovation Platform and Talent Specialization of Jilin Provincial Department of Science and Technology (No. 20230508039RC).

Data availability Not applicable.

Declarations

Ethics approval The authors confirm that they have abided to the publication ethics and state that this work is original and has not been used for publication anywhere before.

Consent to participate The authors are willing to participate in journal promotions and updates.

Consent for publication The authors give consent to the journal regarding the publication of this work.

Conflict of interest The authors declare no competing interests.

References

- Wang B, Hu S, Sun L, Freiheit T (2020) Intelligent welding system technologies: State-of-the-art review and perspectives. *J Manuf Syst* 56:373–391. <https://doi.org/10.1016/j.jmsy.2020.06.020>
- Xu F, Xu Y, Zhang H, Chen S (2022) Application of sensing technology in intelligent robotic arc welding: a review. *J Manuf Process* 79:854–880. <https://doi.org/10.1016/j.jmapro.2022.05.029>
- Lei T, Rong Y, Wang H, Huang Y, Li M (2020) A review of vision-aided robotic welding. *Comput Ind* 123:103326. <https://doi.org/10.1016/j.compind.2020.103326>
- Yang L, Li E, Long T, Fan J, Liang Z (2018) A high-speed seam extraction method based on the novel structured-light sensor for arc welding robot: a review. *IEEE Sens J* 18:8631–8641. <https://doi.org/10.1109/JSEN.2018.2867581>
- Rout A, Deepak BBVL, Biswal BB (2019) Advances in weld seam tracking techniques for robotic welding: a review. *Robot Comput Integr Manuf* 56:12–37. <https://doi.org/10.1016/j.rcim.2018.08.003>
- Lei T, Huang Y, Shao W, Liu W, Rong Y (2020) A tactual weld seam tracking method in super narrow gap of thick plates. *Robot Comput Integr Manuf* 62:101864. <https://doi.org/10.1016/j.rcim.2019.101864>
- Mahajan A, Figueroa F (1997) Intelligent seam tracking using ultrasonic sensors for robotic welding. *Robotica* 15:275–281. <https://doi.org/10.1017/S0263574797000313>
- Le J, Zhang H, Chen X (2018) Realization of rectangular fillet weld tracking based on rotating arc sensors and analysis of experimental results in gas metal arc welding. *Robot Comput Integr Manuf* 49:263–276. <https://doi.org/10.1016/j.rcim.2017.06.004>
- Xu Y, Fang G, Lv N, Chen S, Zou J (2015) Computer vision technology for seam tracking in robotic GTAW and GMAW. *Robot Comput Integr Manuf* 32:25–36. <https://doi.org/10.1016/j.rcim.2014.09.002>
- Shah H, Sulaiman M, Shukor A, Kamis Z, Ab Rahman A (2018) Butt welding joints recognition and location identification by using local thresholding. *Robot Comput Integr Manuf* 51:181–188. <https://doi.org/10.1016/j.rcim.2017.12.007>
- Mao Y, Xu G (2022) A real-time method for detecting weld deviation of corrugated plate fillet weld by laser vision sensor. *Optik* 260:168786. <https://doi.org/10.1016/j.ijleo.2022.168786>
- Zhao X, Zhang Y, Wang H, Liu Y, Zhang B, Hu S (2022) Research on trajectory recognition and control technology of real-time tracking welding. *Sensors* 22:8546. <https://doi.org/10.3390/s22218546>
- Deng L, Lei T, Wu C, Liu Y, Cao S, Zhao S (2023) A weld seam feature real-time extraction method of three typical welds based on target detection. *Measurement* 207:112424. <https://doi.org/10.1016/j.measurement.2022.112424>
- Zou Y, Chen X, Gong G, Li J (2018) A seam tracking system based on a laser vision sensor. *Measurement* 127:489–500. <https://doi.org/10.1016/j.measurement.2018.06.020>
- Shao W, Huang Y, Zhang Y (2018) A novel weld seam detection method for space weld seam of narrow butt joint in laser welding. *Optics Laser Technol* 99:39–51. <https://doi.org/10.1016/j.optlastec.2017.09.037>
- Kiddee P, Fang Z, Tan M (2016) An automated weld seam tracking system for thick plate using cross mark structured light. *Int J Adv Manuf Technol* 87:3589–3603. <https://doi.org/10.1007/s00170-016-8729-7>
- Xu P, Xu G, Tang X, Yao S (2008) A visual seam tracking system for robotic arc welding. *Int J Adv Manuf Technol* 37:70–75. <https://doi.org/10.1007/s00170-007-0939-6>
- Zhang G, Zhang Y, Tuo S, Hou Z, Yang W, Xu Z, Wu Y, Yuan H, Kyoosik S (2021) A novel seam tracking technique with a four-step method and experimental investigation of robotic welding oriented to complex welding seam. *Sensors* 21:3067. <https://doi.org/10.3390/s21093067>

19. Xiao R, Xu Y, Hou Z, Chen C, Chen S (2021) A feature extraction algorithm based on improved Snake model for multi-pass seam tracking in robotic arc welding. *J Manuf Process* 72:48–60. <https://doi.org/10.1016/j.jmapro.2021.10.005>
20. Fan J, Jing F, Yang L, Long T, Tan M (2019) A precise seam tracking method for narrow butt seams based on structured light vision sensor. *Optics Laser Technol* 109:616–626. <https://doi.org/10.1016/j.optlastec.2018.08.047>
21. Chen Y, Shi Y, Cui Y, Chen X (2021) Narrow gap deviation detection in keyhole TIG welding using image processing method based on Mask-RCNN model. *Int J Adv Manuf Technol* 112:2015–2025. <https://doi.org/10.1007/s00170-020-06466-5>
22. Zou Y, Chen T, Chen X, Li J (2022) Robotic seam tracking system combining convolution filter and deep reinforcement learning. *Mech Syst Signal Process* 165:108372–108085. <https://doi.org/10.1016/j.ymsp.2021.108372>
23. Lin Z, Shi Y, Wang Z, Li B, Chen X (2023) Intelligent seam tracking of an ultranarrow gap during K-TIG welding: a hybrid CNN and adaptive ROI operation algorithm. *IEEE Trans Instrum Meas* 72:1–14. <https://doi.org/10.1109/TIM.2022.3230475>
24. Liu C, Shen J, Hu S, Wu D, Zhang C, Yang H (2022) Seam tracking system based on laser vision and CGAN for robotic multi-layer and multi-pass MAG welding. *Eng Appl Artif Intell* 116:105377. <https://doi.org/10.1016/j.engappai.2022.105377>
25. Redmon J, Divvala S, Girshick R, Farhadi A (2016) You only look once: unified, real-time object detection. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 779–788. <https://doi.org/10.1109/CVPR.2016.91>
26. GitHub (2021) YOLOV5-Master. <https://github.com/ultralytics/yolov5.git>. Accessed 14 Oct 2022
27. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2020) Generative adversarial nets. *Commun ACM* 63(11):139–144. <https://doi.org/10.1145/3422622>
28. Mirza M, Osindero S (2014) Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784v1*. <https://doi.org/10.48550/arXiv.1411.1784>
29. Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22:1330–1334. <https://doi.org/10.1109/34.888718>
30. Fan J, Jing F, Fang Z, Liang Z (2016) A simple calibration method of structured light plane parameters for welding robots. In: *Proceedings of the 35th Chinese Control Conference*, pp 6127–6132. <https://doi.org/10.1109/ChiCC.2016.7554318>
31. Tsai R, Lenz R (1989) A new technique for fully autonomous and efficient 3d robotics hand/eye calibration. *IEEE Trans Robot Autom* 5:345–358. <https://doi.org/10.1109/70.34770>
32. Wang C, Liao H, Wu Y, Chen P, Hsieh J, Yeh I (2020) CSPNet: a new backbone that can enhance learning capability of CNN. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 390–391. <https://doi.org/10.1109/CVPRW.2020.00203>
33. He K, Zhang X, Ren S, Sun J (2015) Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Machine Intell* 37(9):1904–1916. <https://doi.org/10.1109/TPAMI.2015.2389824>
34. Liu S, Qi L, Qin H, Shi J, Jia J (2018) Path aggregation network for instance segmentation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 8759–8768. <https://doi.org/10.1109/CVPR.2018.00913>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.