ADVANCES IN MATERIALS SCIENCE, Vol. 26, No. 1 (87), March 2026

DOI: 10.2478/adms-2026-0001

Jiaxin Liu[1], Banghe Su[1], Jinqiang Gao[1,*]

[1] *Institute of Materials Joining, Shandong University, MOE Key Lab for Liquid-Solid Structure Evolution and Materials Processing, Jinan, China*

* *jqg@sdu.edu.cn*

DETECTION AND RECOGNITION OF THE SEAM FOR GAP-FREE BUTT JOINTS BASED ON SCHEIMPFLUG IMAGING AND DEEP LEARNING

ABSTRACT

Due to the limited camera depth of field, intense arc interference, and occlusion by fixed weld points, the accurate detection and recognition of the seam for gap-free butt joints confronts substantial technical difficulties. These difficulties are further exacerbated by the limited generalization capability of traditional image processing methods under such noisy and occluded conditions. In response, a vision sensor was designed based on the Scheimpflug imaging principle, utilizing a tilted lens structure to extend the range of clear imaging. The camera calibration process was simplified by modeling the Scheimpflug transformation as specific distortion parameters. An improved lightweight YOLOv3 network was developed, employing a MobileNetV1 backbone combined with an improved Bi-FPN structure, significantly enhancing inference speed while maintaining detection accuracy. Training results demonstrate that the proposed method achieves a mean Average Precision (mAP) of 96.74% for detecting weld seams and fixed points with an inference speed of 63.45 FPS. In the context of thin-sheet, gap-free butt welding with tack weld interference, the proposed method exhibits satisfactory robustness and practical utility. Furthermore, a trajectory fitting algorithm for the weld seam center was developed based on the grayscale distribution characteristics within the detection bounding boxes, enabling effective marking of the weld trajectory and thereby laying the foundation for weld seam tracking.

Keywords: *Gap-free TIG welding; Scheimpflug imaging; lightweight YOLOv3; weld seam detection*

INTRODUCTION

In the automatic welding process, welding position precision is a critical factor determining welding quality, with accurate identification and real-time tracking of the welding seam constituting the core link to achieving this precision [1, 2]. Current research employs various sensing technologies to realize seam tracking, mainly including contact sensing [3], arc sensing [4], vision sensing [5], and acoustic sensing [6]. Beyond spatial localization, sophisticated sensing frameworks are increasingly vital for comprehensive

Received: 4.02.2026; Received in revised form: 16.03.2026; Accepted: 19.03.2026

process monitoring and quality evaluation. For instance, infrared sensing systems have been enhanced for sensitive imperfection detection in TIG welding [7], virtual sensors have been developed for the on-line assessment of weld hardness [8], and machine vision approaches focusing on dynamic arc features have been utilized to accurately evaluate process stability [9]. Benefiting from such multidimensional perception capabilities, vision sensing technologies have been widely adopted in automated/intelligent robotic welding due to its non-contact nature, fast response, and excellent noise immunity [10, 11]. Based on their working principles, vision sensing techniques can be categorized into active and passive types: active vision relies on auxiliary light sources such as lasers or coded patterns [12], while passive vision captures images of the weld pool and weld seam area using ambient light or arc light. Passive vision techniques often exhibit notable advantages in tracking narrow-gap or groove-free weld seams [13]. However, its application is constrained by the camera's depth of field, which limits the range over which objects can be imaged clearly. Traditional methods for extending the depth of field, such as reducing the aperture size or shortening the focal length, can alleviate this issue but often lead to a decrease in resolution. To resolve this contradiction, the Scheimpflug camera employs a tilted lens design, which significantly optimizes the clear focusing of detected points within the sensor's field of view without sacrificing aperture size [14].

The precise establishment of an imaging model for Scheimpflug cameras forms the foundation for improving tracking capabilities. Louhichi et al. [15] proposed the concept of introducing an ideal, non-tilted image plane and constructed an imaging model by describing the transformation between the actual tilted image plane and this ideal plane using a one-dimensional Scheimpflug tilt angle. Subsequently, Zhang et al. [16] extended the representation of the tilt angle to two dimensions, demonstrating that this two-dimensional representation offers higher accuracy in describing the imaging process of tilted cameras. Cornic et al. [17] then developed a comprehensive physical model for Scheimpflug cameras, innovatively integrating the universality of the pinhole model with the practical physical constraints of tilted lenses. This model allows for calibration through the free movement of a calibration target, significantly enhancing operational convenience. However, existing Scheimpflug camera calibration techniques, including the aforementioned methods, commonly suffer from complex calibration procedures and high demands on environmental conditions. Therefore, while maintaining imaging accuracy, this study introduces the Scheimpflug transformation as a specific distortion model within the camera coordinate system to further streamline the calibration algorithm workflow.

It is noteworthy that passive visual weld seam tracking systems are highly susceptible to interference from intense welding arc light, spatter, and material surface glare under actual working conditions. These factors impose severe challenges on optical structure design, image processing algorithms, and the capability to extract feature information, significantly increasing the difficulty of robustly extracting seam information. In thin-plate gap-free welding, the welding fixation process is often employed to pre-fix workpieces in order to mitigate the impact of assembly errors and welding thermal deformation on structural accuracy. However, the presence of welding fixation points obscures seam features, and the difficulty in their identification can interfere with the effectiveness of seam tracking. While traditional visual image processing methods offer certain effectiveness, they often lack generalization capability and struggle to adapt to diverse working conditions. In recent years, breakthroughs in deep learning technology have provided new avenues for addressing these challenges. Deep learning-based vision recognition algorithms, leveraging their strong robustness and ability to automatically extract multi-dimensional seam features, demonstrate considerable potential.

Deep learning-based object detection algorithms are primarily categorized into two-stage and one-stage approaches. The early R-CNN series represents the two-stage methods: R-CNN [18] pioneered the application of convolutional neural networks (CNNs) for regional feature extraction, but required separate convolution for each candidate region, resulting in high computational redundancy; Fast R-CNN [19] significantly improved efficiency by sharing convolutional features; Faster R-CNN [20] further deeply integrated the Region Proposal Network (RPN) with the detection network, enabling end-to-end training and inference. However, the inherent multi-stage nature of two-stage methods limits their inference speed, which restricts their application in real-time detection scenarios. In contrast, the one-stage YOLO series [21] innovatively formulates object detection as a unified regression problem, directly predicting bounding boxes and class labels through a single forward pass, thereby significantly improving speed. In subsequent iterations, YOLOv2 [22] introduced the Anchor Boxes mechanism and incorporated multi-scale feature fusion, optimizing the balance between accuracy and speed; YOLOv3 [23] further integrated the Darknet-53 backbone network, ResNet principles, and the FPN [24], substantially enhancing small object detection capability.

In recent years, a variety of improved YOLO models have been proposed and applied to welding visual perception. Yang et al. [25] introduced a simplified YOLOv3 architecture, utilizing Darknet-53 as the backbone and integrating a Feature Pyramid Network (FPN), combined with anchor box clustering and a dynamic sample updating strategy, their model achieved a high accuracy of 94.7%. Li et al. [26] enhanced the YOLOv3 framework using a ResNet34-d backbone and a Path Aggregation Network (PANet) with residual structures. This model attained a precision of 95.03% mAP50 at 49.43 FPS, effectively ensuring real-time and accurate detection of the weld pool morphology in thin-plate TIG welding. Additionally, Song et al. [27] performed lightweight improvements on YOLOv4 by integrating a MobileNetv3 backbone, a Bidirectional Feature Pyramid Network (BiFPN), and depthwise separable convolutions. Their approach realized a mean Average Precision (mAP) of 98.84% at 62.22 FPS, providing a reliable solution for the real-time detection of weld seam features in environments with strong noise interference. Dai et al. [28] refined the YOLOv3 framework. Their approach incorporated MobileNetV3 as a lightweight backbone, devised a cross-scale feature pyramid structure, and introduced the Complete Intersection over Union (CIoU) loss function. This enhanced model achieved precise localization in small-target spot weld detection, thereby overcoming the limitations of traditional methods which often suffer from insufficient accuracy under complex interference.

This paper focuses on the accurate detection and recognition of weld seams and fixation points in gap-free butt joint welding, where challenges such as limited depth of field, intense arc interference, and occlusion hinder conventional vision-based methods. To address these issues, a visual sensor based on the Scheimpflug imaging principle is designed, which extends the clear imaging range and simplifies camera calibration by modeling the Scheimpflug transformation as specific distortion parameters. On this hardware foundation, a lightweight improved YOLOv3 network is constructed for robust feature detection. The model replaces the original Darknet-53 backbone with MobileNetV1 and incorporates an enhanced Bi-FPN structure to strengthen multi-scale feature fusion, thereby significantly increasing inference speed while maintaining high detection accuracy. Furthermore, a trajectory extraction algorithm for the weld seam center is proposed based on grayscale distribution characteristics within the detected bounding boxes, enabling stable and precise acquisition of the seam position even under occlusion. This work provides an effective vision-based solution for seam tracking in complex welding environments.

EXPERIMENT DETAILS

Experimental system

The experimental system primarily consists of two components: the TIG welding system and the vision sensing system, as illustrated in Figure 1.

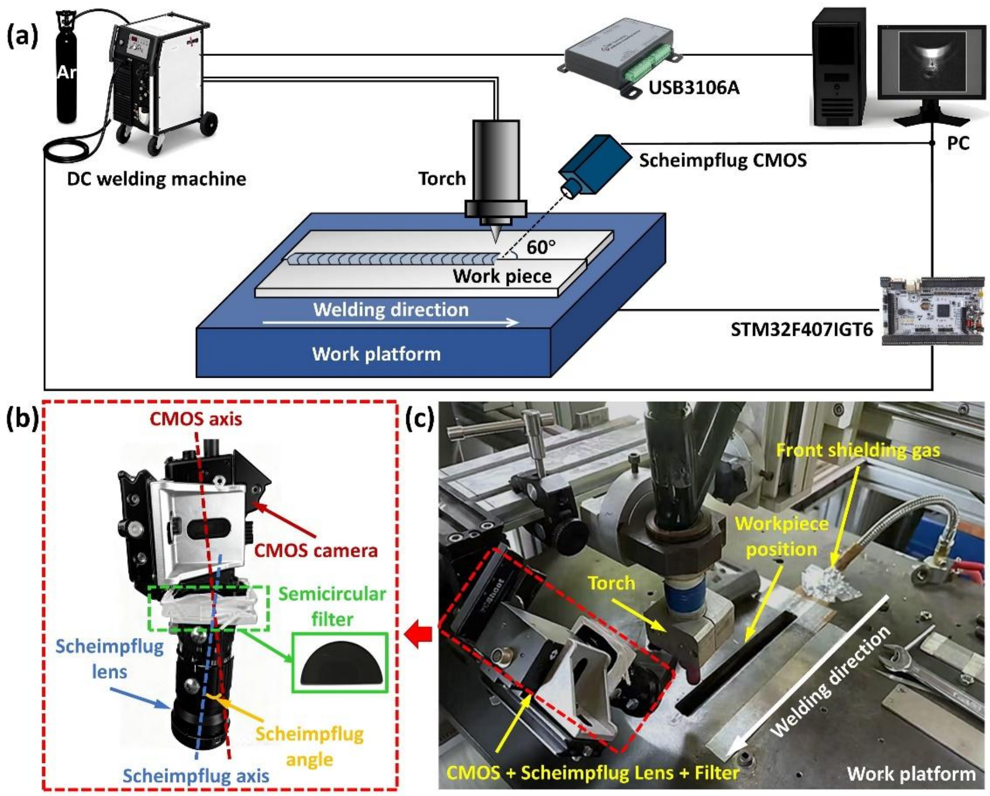


Fig. 1. Experimental system: (a) system composition, (b) vision sensing system, (c) work platform configuration

The TIG welding system includes a welding power source, a three-dimensional motion platform, a water-cooled TIG welding torch, and a gas shielding device. It is noteworthy that the vision sensing system is composed of a CMOS camera equipped with a Scheimpflug lens and a semi-circular filter with 25% transmittance. The camera is installed at a 60° inclination on the front side of the workpiece, a configuration that effectively suppresses intense interference from the welding arc. The two systems are integrated via a data acquisition card to enable real-time monitoring and control, while the captured image data are stored in a PC for subsequent experimental analysis.

Experimental methods

In this study, 304 stainless steel plates with dimensions of 200 mm × 40 mm × 3 mm were used as the base material, as shown in Figure 2. The plates were fabricated by laser cutting, and the edges were subsequently ground to remove burrs. Prior to welding, the plate surfaces were polished with sandpaper and cleaned with ethanol to eliminate surface contaminants that could adversely affect weld quality. To prevent plate displacement due to thermal distortion or external forces during welding and to ensure accurate alignment of the weld seam, tack welding was performed at both ends of the plate and along the welding zone prior to the actual welding process. A tungsten electrode with a diameter of 3.2 mm was used, and ground to a tip angle of about 35°. High-purity argon (99.99%) was employed as the shielding gas on both the front and back sides, with flow rates set to 15 L/min and 8 L/min, respectively. The welding parameters were set as follows: welding current 110-140 A and welding speed 120-150 mm/min.

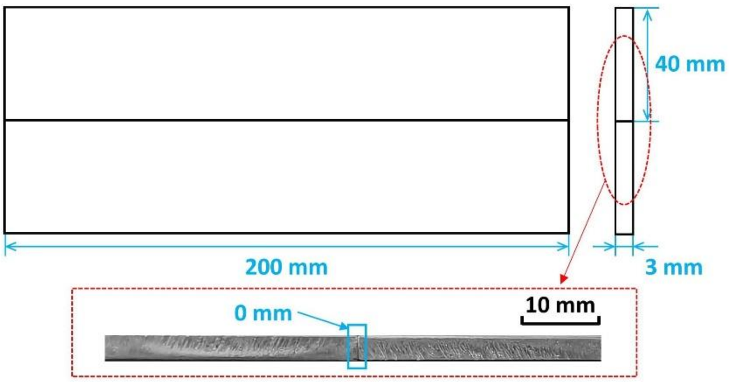


Fig. 2. Schematic diagram of welding test specimen

Camera calibration

The unique optical characteristics of Scheimpflug lenses preclude the direct use of conventional calibration methods. Existing calibration schemes for Scheimpflug imaging systems are relatively complex and exhibit poor adaptability under welding conditions. Therefore, this paper models the Scheimpflug transformation as a special distortion parameter and incorporates it into the pinhole imaging model for optimization. Within this model, four coordinate systems are involved [29]: the world coordinate system (O_W−$X_WY_WZ_W$), the camera coordinate system (O_C–$X_CY_CZ_C$), the image coordinate system (O_{xy}–xy), and the pixel plane coordinate system (O_{uv}–uv). The checkerboard calibration plate employed in the calibration process has overall dimensions of 10.0 mm × 10.0 mm, with individual checker size of 1.0 mm × 1.0 mm, arranged in a uniform 10 × 10 grid pattern in both horizontal and vertical directions.

The transformation from the world coordinate system to the camera coordinate system via a rigid-body transformation is given by Eq. (1):

$$\begin{bmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{bmatrix} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} X_W \\ Y_W \\ Z_W \\ 1 \end{bmatrix} \tag{1}$$

where R and T denote a 3×3 rotation matrix and a 3×1 translation vector, respectively.

From the camera coordinate system, the camera homogeneous coordinates $(X_C', Y_C', 1)$ are obtained as expressed in Eq. (2):

$$\begin{bmatrix} X_C' \\ Y_C' \end{bmatrix} = \begin{bmatrix} X_C/Z_C \\ Y_C/Z_C \end{bmatrix} \tag{2}$$

Given that the images from the Scheimpflug imaging system in this study are predominantly concentrated in the central region, and the distortion at the edges has minimal impact on weld feature extraction, the distortion correction is performed directly in the camera coordinate system to reduce computational cost. The corresponding distortion model is shown in Eq. (3):

$$\begin{bmatrix} X_C'' \\ Y_C'' \end{bmatrix} = \begin{bmatrix} X_C'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + 2p_1 X_C' Y_C' + p_2(r^2 + 2(X_C')^2) \\ Y_C'(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) + p_1(r^2 + 2(Y_C')^2) + 2p_2 X_C' Y_C' \end{bmatrix} \tag{3}$$

where $r = \sqrt{(X_C')^2 + (Y_C')^2}$ is the radial distance, and k_1, k_2, k_3 and P_1, P_2 are the radial and tangential distortion coefficients, respectively.

Based on the characteristics of the Scheimpflug transformation, the ideal imaging plane is converted to the Scheimpflug imaging plane, as illustrated in Figure 3. The rotation matrix R_τ of the Scheimpflug lens and the Scheimpflug matrix S_τ can be expressed as Eqs. (4) and (5):

$$R_\tau = \begin{bmatrix} \cos\sigma & 0 & \sin\sigma \\ 0 & 1 & 0 \\ -\sin\sigma & 0 & \cos\sigma \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\rho & -\sin\rho \\ 0 & \sin\rho & \cos\rho \end{bmatrix} \tag{4}$$

$$S_\tau = \begin{bmatrix} \cos\rho\cos\sigma & 0 & \sin\sigma \\ 0 & \cos\rho\cos\sigma & -\cos\rho\cos\sigma \\ 0 & 0 & 1 \end{bmatrix} \tag{5}$$

where (ρ, σ) denote the two-dimensional clockwise rotation angles about its x-axis and y-axis.

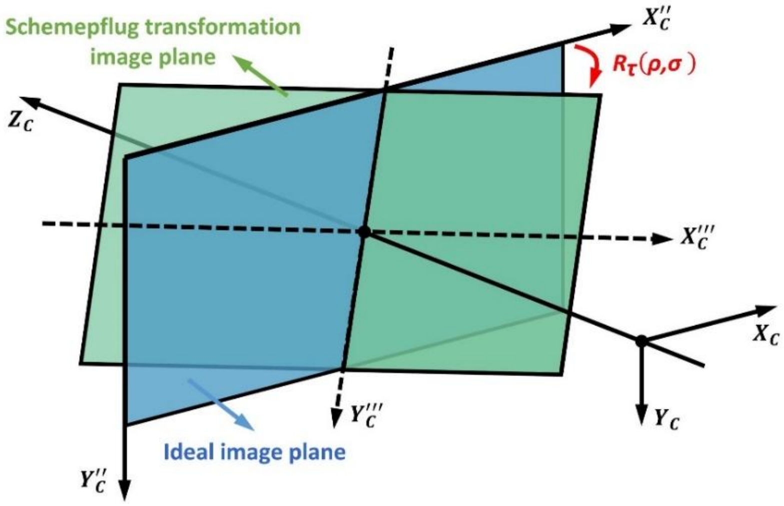


Fig. 3. Scheimpflug imaging model

The homogeneous camera coordinate system $(X_C''', Y_C''', 1)$ after the Scheimpflug transformation is given by Eq. (6):

$$\beta \begin{bmatrix} X_C''' \\ Y_C''' \\ 1 \end{bmatrix} = S_\tau R_\tau^T \begin{bmatrix} X_C'' \\ Y_C'' \\ 1 \end{bmatrix} \tag{6}$$

where the constant β is the scale factor associated with homogenization.

By utilizing the relationship of similar triangles to transform the camera coordinates into the image coordinate system, Eq. (7) is derived from Eqs. (2)–(6):

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \frac{X_C}{Z_C} f \\ \frac{Y_C}{Z_C} f \end{bmatrix} = \begin{bmatrix} X_C''' f \\ Y_C''' f \end{bmatrix} \tag{7}$$

where f is the focal length, defined as the distance between the image plane and the origin of the camera coordinate system.

The transformation relationship from the image coordinate system to the pixel plane coordinate system is given by Eq. (8):

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{8}$$

where (u_0, v_0) are the coordinates of the origin O_{xy} of the image coordinate system within the pixel plane coordinate system, and f_u and f_v are the effective focal lengths in the u and v directions.

Given the fixed relative positions and constant spatial scale between the camera, lens, and the welding plane, this paper constructs a fixed perspective transformation matrix for image conversion. This matrix enables the transformation of corner points in the corrected image into an ideally arranged, standard array of corner points. The perspective transformation matrix is defined by Eq. (9):

$$\begin{bmatrix} A & B & C \\ D & E & F \\ G & H & 1 \end{bmatrix} \tag{9}$$

where A, B, D, and E are linear transformation parameters; C and F are translation parameters; G and H are perspective deformation parameters; and the 1 serves as a normalization term of the matrix to preserve the homogeneity of the matrix transformation.

After perspective transformation, the coordinates (x', y') of the ideally arranged array of corner points in standard form are defined by Eq. (10):

$$\begin{cases} x' = \frac{Ax+By+C}{Gx+Hy+1} \\ y' = \frac{Dx+Ey+F}{Gx+Hy+1} \end{cases} \tag{10}$$

The scaling factors K_x and K_y between the pixel values in the standard corner-point array and the checkerboard dimensions are given by Eq. (11):

$$K_x = K_y = 0.0483 \text{ mm/pixel} \tag{11}$$

METHODOLOGY

Dataset construction and preprocessing

A total of 1,691 welding images were collected under various welding conditions, with the detailed parameters of these conditions presented in Table 1. All images are grayscale with a resolution of 500×600 pixels. As YOLOv3 is a supervised learning model, manual annotation of the images was required prior to training. The annotation work was completed using the EasyData data service platform. In the annotation process, two types of target bounding boxes were defined: the welding seam and the welding fixation point. Figure 4 presents an example of the annotated images, where the red boxes indicate welding fixation points and the white boxes indicate the welding seam. The entire dataset was split into training, validation, and test sets in an 8:1:1 ratio, resulting in 1,353, 169, and 169 images, respectively.

Table 1. *Welding parameter variations*

Process parameter	Value range
Welding speed	120-150 mm/min
Torch height	2-4 mm
Welding current	110-140 A
Welding fixation point size	4-7 mm

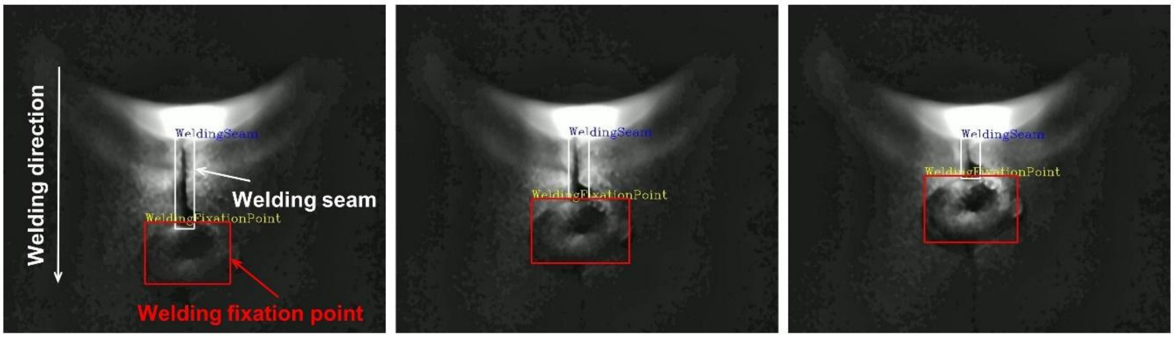


Fig. 4. Annotation results

To mitigate potential overfitting caused by the limited dataset size and insufficient sample diversity, several data augmentation techniques were applied, including flipping, random occlusion, scaling, and brightness adjustment. The specific augmentation operators and their corresponding parameters are summarized in Table 2. After data augmentation, the size of the training set was increased to 15,000 images.

Table 2. *Data augmentation operators and parameters*

Operator	Probability	Intensity
Rotate_Only_BBoxes	0.3	9°
Equalize_Only_BBoxes	0.1	-
Flip_Only_BBoxes	0.5	-
Contrast	0.3	0.64
Brightness	0.5	0.46
Sharpness	0.2	0.46

Lightweight YOLOv3 model

In the gap-free welding process of thin plates, factors such as surface grinding of the workpiece and the reflectivity of stainless steel materials often lead to insufficiently distinct features in weld seam images, posing significant challenges for visual recognition. To address this, this study employs the deep learning algorithm YOLOv3, specifically designed for real-time object detection, to enhance the robustness and generalization capability of the image processing stage. As a general object detection framework, the YOLOv3 architecture, illustrated in Figure 5, primarily consists of three components: the Darknet-53 backbone network, the multi-scale feature fusion network FPN, and the detection head.

Under the constraints of welding costs, edge computing devices often face resource limitations when performing real-time inference tasks. To address this, this study reconstructs the YOLOv3 network, with the core objectives of improving the model's detection accuracy for subtle features while reducing its computational resource requirements for this specific welding scenario. Following the lightweight design philosophy proposed by Cai et al. [30], the computationally more efficient lightweight backbone network MobileNetV1 is adopted to replace the original Darknet-53.

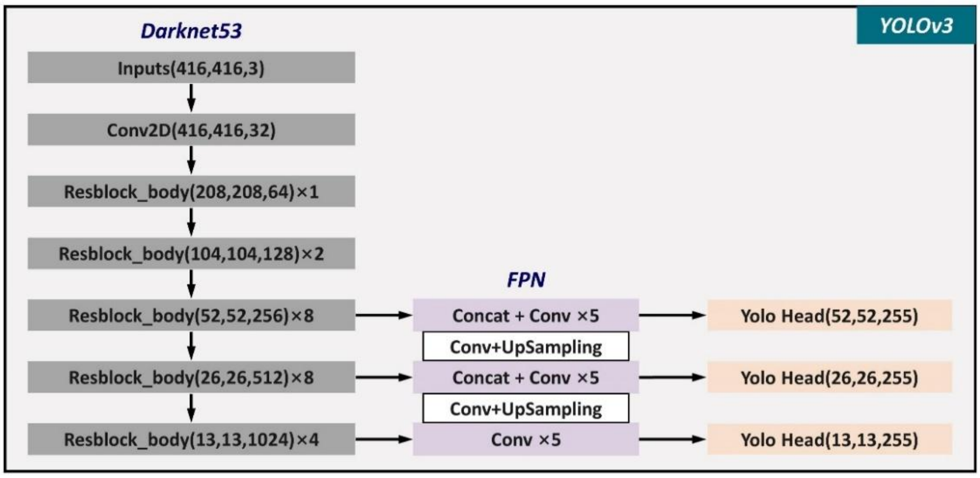
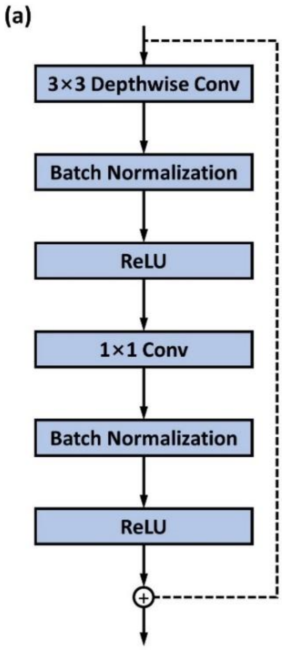


Fig. 5. YOLOv3 network architecture



(b) MobileNetV1		
Type/Stride	Filter Shape	Input Size
Conv-s2	3×3×3×32	224×224×3
Conv dw-s1	3×3×32dw	112×112×32
Conv-s1	1×1×32×64	112×112×32
Conv dw-s2	3×3×64dw	112×112×64
Conv-s1	1×1×64×128	56×56×64
Conv dw-s1	3×3×128dw	56×56×128
Conv-s1	1×1×128×128	56×56×128
Conv dw-s2	3×3×128dw	56×56×128
Conv-s1	1×1×128×256	28×28×128
Conv dw-s1	3×3×256dw	28×28×256
Conv-s1	1×1×256×256	28×28×256
Conv dw-s2	3×3×256dw	28×28×256
Conv-s1	1×1×256×512	14×14×256
5×Conv-s1	3×3×512dw	14×14×512
5×Conv dw-s1	1×1×512×512	14×14×512
Conv dw-s2	3×3×512dw	14×14×512
Conv-s1	3×3×512×1024	7×7×512
Conv dw-s2	3×3×1024dw	7×7×1024
Conv-s1	1×1×1024×1024	7×7×1024

Fig. 6. Architecture of MobileNetV1: (a) convolutional structure, (b) network architecture

This network effectively reduces both the parameter count and computational load through depthwise separable convolutions. To maintain compatibility with the YOLOv3 detection head, the global pooling layer and fully connected layer at the end of MobileNetV1 are removed. The structure of MobileNetV1 is shown in Figure 6.

YOLOv3 enhances its ability to detect small objects by incorporating the FPN architecture. However, the classical FPN suffers from the limitation of semantic information gradually decaying during the feature propagation process. To address this, this paper introduces optimizations based on Bi-FPN. The feature hierarchy is streamlined to three levels, and a dual-stage weighted fusion mechanism is incorporated. This effectively compensates for information loss by strengthening cross-scale semantic correlations, thereby precisely meeting the requirements of the YOLOv3 detection head while reducing computational complexity. The improved Bi-FPN network structure is illustrated in Figure 7.

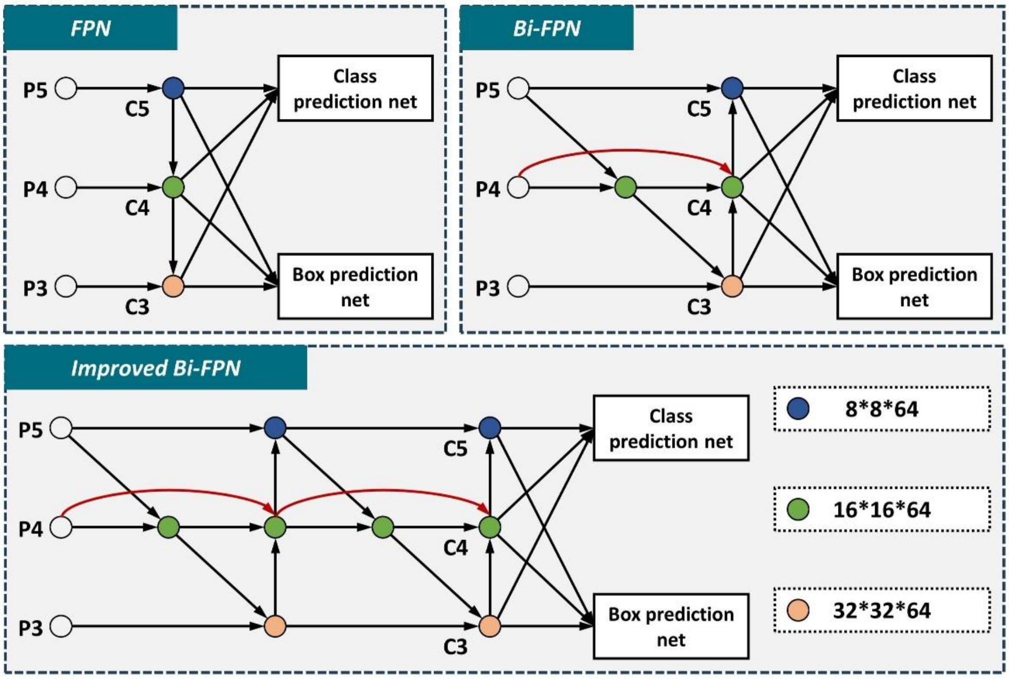


Fig. 7. Improvement process of the Bi-FPN network structure

Weld seam trajectory fitting method

Analyzing image features is fundamental for subsequent weld seam tracking. The grayscale information perpendicular to the weld seam direction is extracted from the images, with the relevant analysis results presented in Figure 8. It can be observed that the grayscale values of the weld pool region and the bounding boxes are significantly higher than other areas, and their contours are distinct. Notably, the grayscale value at the center of the weld seam within the bounding boxes is markedly lower than that on both sides, reaching its minimum precisely at the weld centerline. Based on this characteristic, a row-by-row scan is performed across the bounding box region of the weld seam detected by the deep learning model. For each row, the pixel position with the lowest grayscale value is marked. By connecting all marked points, an initial trajectory line is formed. This trajectory is then smoothed to eliminate outliers that deviate significantly from the average position. The effect of the weld seam trajectory fitting is illustrated in Figure 9. To improve the accuracy of trajectory fitting, a threshold is applied to the confidence scores returned by the deep learning model. Marks with a confidence level below 0.7 are considered untrustworthy and are filtered out. Furthermore, since excessively small bounding boxes may be affected by shadows from solidified weld points, scratches, or oil stains on the workpiece itself, the trajectory fitting is constrained to bounding boxes with a height greater than 100 pixels and an area exceeding 300 square pixels. When the presence of a welding fixation point occludes the weld seam, preventing successful trajectory fitting, the system, upon detecting the fixation point, controls the welding torch to continue moving along the current weld seam trajectory. This ensures the continuity of the welding process.

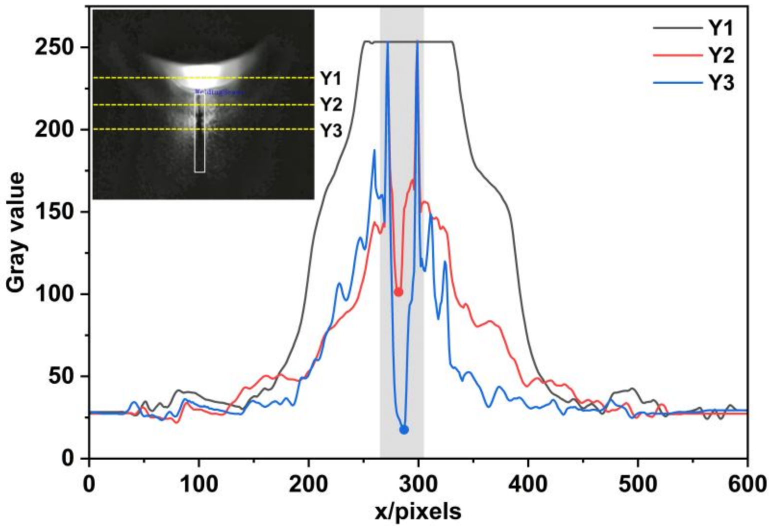


Fig. 8. Grayscale value analysis of the weld seam region

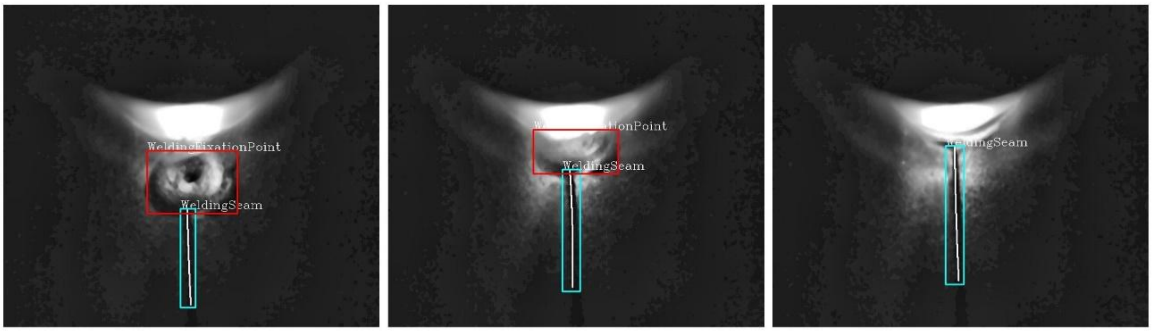
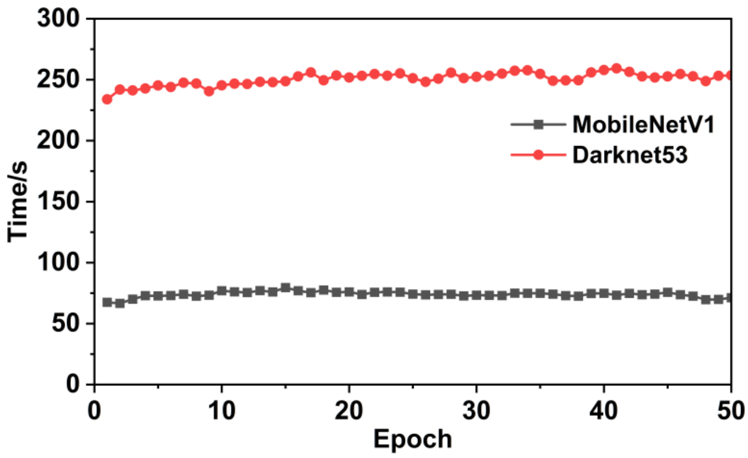


Fig. 9. Weld seam trajectory fitting result

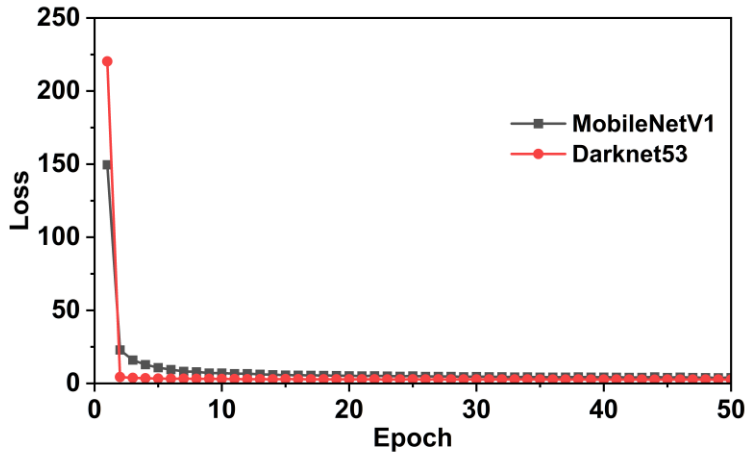
RESULTS AND DISCUSSION

The model training and validation in this study were conducted on a computer equipped with an Intel Core i7-12700 2.10 GHz processor and an NVIDIA GeForce GTX 3080 Super GPU. All models were developed and validated based on the PaddleDetection framework. During training, the input image size was uniformly resized to 600×700 pixels, with a batch size of 8, an initial learning rate of 0.01, and the training duration set to 50 epochs.

To further evaluate the performance of the backbone networks, a detailed analysis of the training process was conducted for both MobileNetV1 and DarkNet-53. The training time and loss variation across epochs were compared, with the results shown in Figure 10. As shown in Figure 10a, the training time for MobileNetV1 exhibits smaller fluctuations and is significantly lower overall than that of DarkNet-53, indicating higher training efficiency. The loss curves in Figure 10b reveal that the training loss of Darknet-53 is notably higher than that of MobileNetV1, suggesting that MobileNetV1 possesses a stronger fitting capability for the data in this task. In summary, MobileNetV1 demonstrates superior training efficiency and lower training loss in this study, and its lightweight architecture is more suitable for the requirements of the current task.



(a) Training time per epoch



(b) Loss function per epoch

Fig. 10. Training process of the MobileNetV1 and Darknet-53 backbone networks

To investigate the influence of different network architectures on the recognition performance of the YOLOv3 algorithm, three network architectures were trained and subjected to comparative analysis. The results are presented in Table 3. FPS refers to the inference speed of the deep learning model on an NVIDIA GeForce GTX 3080 GPU, and it does not include the time consumed by preprocessing, post-processing, and trajectory extraction. Experiments indicate that while the YOLOv3 model with DarkNet-53 as its backbone network achieves high detection accuracy, its high computational complexity leads to slower inference speed, making it difficult to meet the real-time requirements of the weld seam tracking task.

Replacing DarkNet-53 with the lightweight MobileNetV1 as the backbone network significantly improved the model's inference speed while maintaining detection accuracy comparable to the baseline model. Furthermore, replacing the original FPN structure with an improved Bi-FPN effectively enhanced the model's accuracy. Although the introduction of improved Bi-FPN slightly reduced the inference speed, it remained substantially faster than the original DarkNet-53 model and fully met the requirements for real-time control applications.

In summary, by incorporating MobileNetV1 and an improved Bi-FPN, the modified YOLOv3 model achieved enhancements in both inference speed and detection accuracy. This provides a more practical and efficient solution for the weld seam tracking task during the welding process, while ensuring real-time performance.

Table 3. *Comparison of object detection models with different architectures*

YOLOv3 Network Model	mAP	FPS
DarkNet-53_FPN	96.35%	43.15
MobileNetV1_Bi-FPN	95.42%	77.86
MobileNetV1_improved Bi-FPN	96.74%	63.45

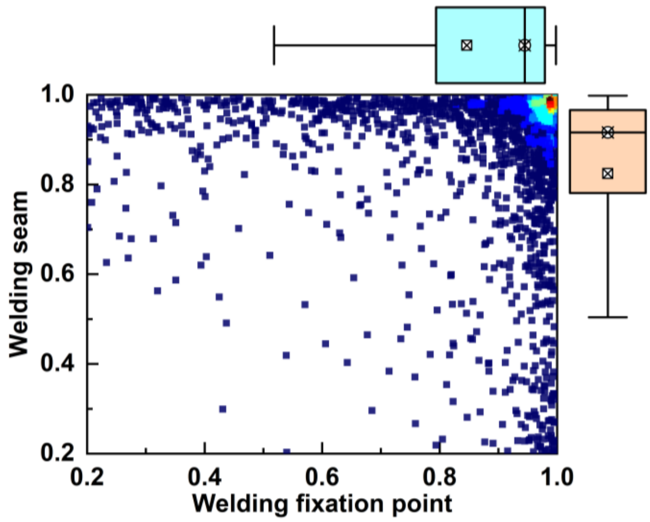


Fig. 11. Marginal box plot of confidence scores

In this study, the performance of the model was validated using the confidence scores output by the statistical model. Confidence is a value within the range of [0, 1], reflecting the probability that a detected bounding box contains the target object and the IoU quality between the detected and ground truth boxes. The validation dataset consisted of 3,818 images that were not included in the training set. The statistical results of the confidence scores are presented in Figure 11. As shown, the detection confidence scores of the proposed model on this dataset were all above 0.50, with the vast majority of bounding boxes exhibiting high confidence levels. Specifically, the median confidence for weld seam detection was 0.92, with an average of 0.82. For welding fixation point detection, the median confidence was 0.95, with an average of 0.85. These results indicate that the model possesses strong recognition capability for both weld seams and welding fixation points.

CONCLUSIONS

- A visual sensing method based on Scheimpflug imaging was proposed, which achieves large-depth-of-field clear imaging of the welding process. The Scheimpflug transformation is modeled as specific distortion parameters and incorporated into the pinhole imaging model, thereby simplifying the camera calibration process.

- A lightweight weld seam feature detection model was constructed. By replacing the original Darknet-53 backbone of YOLOv3 with MobileNetV1 and introducing an improved Bi-FPN structure for multi-scale feature fusion, the model significantly increased the inference speed to 63.45 FPS while largely maintaining high detection accuracy (mAP of 96.74%). This achieved a good balance between accuracy and real-time performance, making it suitable for resource-constrained edge computing environments.
- The improved model demonstrated significantly superior speed compared to the original YOLOv3 in comparative experiments involving different network architectures. Comparison of the training processes indicated that the lightweight backbone network offers higher training efficiency and lower loss. Furthermore, statistical analysis of confidence scores on the validation set showed that the model achieved high confidence scores in recognizing both weld seams and welding fixation points, indicating reliable performance within the tested scope.
- A weld seam trajectory extraction algorithm based on grayscale distribution was proposed. By analyzing the grayscale distribution characteristics within the detected weld seam bounding boxes, the pixel with the minimum grayscale value in each row is identified and connected to fit an initial weld seam centerline. This process is optimized by incorporating confidence score filtering and bounding box size constraints, enabling stable and continuous extraction of the weld seam trajectory even under challenging conditions such as occlusion by welding fixation points.

REFERENCES

1. Pham, D.-A., Bui, D.-Q., Le, T.-D., Tran, D.-H., Nguyen, T.-H.: Automatic welding seam tracking and real-world coordinates identification with machine learning method. Results in Engineering 23 (2024), 10.

2. Shang, G., Xu, L., Li, Z., Zhou, Z., Xu, Z.: Digital-twin-based predictive compensation control strategy for seam tracking in steel sheets welding of large cruise ships. Robotics and Computer-Integrated Manufacturing 88 (2024), 16.

3. Rout, A., Deepak, B.B.V.L., Biswal, B.B.: Advances in weld seam tracking techniques for robotic welding: A review. Robotics and Computer-Integrated Manufacturing 56 (2019), 12-37.

4. Lin, J., Jia, A., Huang, W., Wen, Z., Hong, B., Hong, Y.: Weld seam tracking method of root pass welding with variable gap based on magnetically controlled arc sensor. The International Journal of Advanced Manufacturing Technology 126 (2023), 5227-5243.

5. Xu, F., He, L., Hou, Z., Xiao, R., Zuo, T., Li, J., Xu, Y., Zhang, H.: An automatic feature point extraction method based on laser vision for robotic multi-layer multi-pass weld seam tracking. The International Journal of Advanced Manufacturing Technology 131 (2024), 5941-5960.

6. Wenji, L., Zhenyu, G., Xiao, J., Li, L., Jianfeng, Y.: Research on the seam tracking of narrow

gap P-GMAW based on arc sound sensing. Sensors and Actuators A: Physical 292 (2019), 205-216.

7. Górka, J., Jamrozik, W.: Enhancement of Imperfection Detection Capabilities in TIG Welding of the Infrared Monitoring System. Metals 11 (2021), 1624.
8. Górka, J., Jamrozik, W., Wyględacz, B., Kiel-Jamrozik, M., Ferreira, B.G.: Virtual Sensor for On-Line Hardness Assessment in TIG Welding of Inconel 600 Alloy Thin Plates. Sensors 24 (2024), 3569.
9. Jamrozik, W., Górka, J.: Assessing MMA Welding Process Stability Using Machine Vision-Based Arc Features Tracking System. Sensors 21 (2020), 84.
10. Guo, Q., Yang, Z., Xu, J., Jiang, Y., Wang, W., Liu, Z., Zhao, W., Sun, Y.: Progress, challenges and trends on vision sensing technologies in automatic/intelligent robotic welding: State-of-the-art review. Robotics and Computer-Integrated Manufacturing 89 (2024), 45.
11. Xu, F., Xu, Y., Zhang, H., Chen, S.: Application of sensing technology in intelligent robotic arc welding: A review. Journal of Manufacturing Processes 79 (2022), 854-880.
12. Nguyen, A.H., Ly, K.L., Lam, V.K., Wang, Z.Y.: Generalized Fringe-to-Phase Framework for Single-Shot 3D Reconstruction Integrating Structured Light with Deep Learning. Sensors 23 (2023), 18.
13. Yang, L., Deng, J., Shen, J.: A new passive vision weld seam tracking method for FSW based on K-means. The International Journal of Advanced Manufacturing Technology 128 (2023), 3283-3295.
14. Sun, C., Liu, H.B., Jia, M.N., Chen, S.Y.: Review of Calibration Methods for Scheimpflug Camera. Journal of Sensors 2018 (2018), 15.
15. Louhichi, H., Fournel, T., Lavest, J.M., Aissia, H.B.: Self-calibration of Scheimpflug cameras: an easy protocol. Measurement Science and Technology 18 (2007), 2616-2622.
16. Zhang, X., Zhou, T.: Generic Scheimpflug camera model and its calibration. IEEE International Conference on Robotics and Biomimetics (ROBIO), Zhuhai, Peoples R China, 2015, 2264-2270.
17. Cornic, P., Illoul, C., Cheminet, A., Le Besnerais, G., Champagnat, F., Le Sant, Y., Leclaire, B.: Another look at volume self-calibration: calibration and self-calibration within a pinhole model of Scheimpflug cameras. Measurement Science and Technology 27 (2016), 15.
18. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2014, 580-587.
19. Girshick, R.: Fast r-cnn. Proceedings of the IEEE international conference on computer vision, 2015, 1440-1448.
20. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE transactions on pattern analysis and machine intelligence 39

(2016), 1137-1149.

21. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: Unified, real-time object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 779-788.

22. Redmon, J., Farhadi, A.: YOLO9000: better, faster, stronger. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 7263-7271.

23. Hussain, M.: YOLO-v1 to YOLO-v8, the Rise of YOLO and Its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines 11 (2023), 25.

24. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 2117-2125.

25. Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., Liang, Z.: Automatic Detection and Location of Weld Beads With Deep Convolutional Neural Networks. IEEE Transactions on Instrumentation and Measurement 70 (2021), 1-12.

26. Li, S., Gao, J., Zhou, E., Pan, Q., Wang, X.: Deep learning-based fusion hole state recognition and width extraction for thin plate TIG welding. Welding in the World 66 (2022), 1329-1347.

27. Song, L., Kang, J., Zhang, Q., Wang, S.: A weld feature points detection method based on improved YOLO for welding robots in strong noise environment. Signal, Image and Video Processing 17 (2022), 1801-1809.

28. Dai, W., Li, D., Tang, D., Jiang, Q., Wang, D., Wang, H., Peng, Y.: Deep learning assisted vision inspection of resistance spot welds. Journal of Manufacturing Processes 62 (2021), 262-274.

29. Deng, H., Liu, Y., Zhang, G., Qu, X., Ge, L., Yang, S.: Multi-angle Scheimpflug projection 3D microscope: Design, calibration, and three-dimensional reconstruction. Measurement 222 (2023), 22.

30. Cai, K.W., Miao, X.Y., Wang, W., Pang, H.S., Liu, Y., Song, J.Y.: A modified YOLOv3 model for fish detection based on MobileNetv1 as backbone. Aquacultural Engineering 91 (2020), 9.