



Composite Vision Method for Measurement of 3D Bead Profile — Part 2: Image Processing

A novel, as-welded, three-dimensional bead profile measurement method is presented for multilayer and multipass welding

BY J. JIAO, Z. LI, W. XU, C. LI, S. WANG, F. CHENG, Z. WANG, AND S. WU

Abstract

In multilayer, multipass (MLMP) welding, the first pass in each layer establishes the reference elevation for subsequent passes and determines the layer height. On the other hand, the last pass, which completes the layer, critically affects its filling quality. In such an equal-height path-planning strategy, the online identification of the profiles of the just-finished beads is of great significance for real-time path adjustment in MLMP welding. However, due to the special positions of the first and last weld beads, identifying their profile features poses multiple challenges, including light occlusion, welding spatter, silicide precipitates, and brightness variation. To address the difficulty of identifying the first and last as-welded beads in MLMP welding, this study proposes an improved image segmentation method based on a Dual-UNet. In addition to the passive vision image, the edge mask map extracted from the active vision images is employed as the second input to the model. An Edge-Attention module and Edge-Aware loss are incorporated into the network architecture, along with channel-wise feature concatenation and a dynamic attention mechanism, to enhance the model's ability to perceive edge features. Then, a Dual-UNet model based on a dual-branch encoding structure and attention fusion is constructed. The experimental results demonstrate that the improved model achieves maximum errors of less than 0.93 mm in width and 0.38 mm in height for identifying the first and last as-welded beads in MLMP welding.

Keywords

- Image Segmentation Method
- Edge Mask Map
- Dual-UNet
- Edge-Attention Module

Background

In Part 1, a composite vision method combining a passive and an active vision sensor was proposed for real-time measurement of the 3D profile of as-welded beads during MLMP welding (Ref. 1). This previous work describes the image processing strategy as a process which involves constructing a U-Net model to extract as-welded-bead-edge features from passive vision images and combining them with laser-line information from active vision images to achieve high-precision identification of the 3D profile of intermediate as-welded beads. Figure 1 presents the overall architecture and image processing strategy of the composite vision system. In Fig. 1A, the layout of the passive and active vision systems is illustrated, where the passive vision camera captures 2D images of the as-welded bead surface, as shown in Fig. 1A-I. The active vision camera, on the other hand, acquires the 3D profile information of the as-welded bead through laser projection, with the result displayed in Fig. 1A-II. Figure 1B demonstrates the extraction process of the as-welded bead ROI. In the methodology proposed in Part 1, the passive vision image is processed by this conventional U-Net (Ref. 2) model to extract features of the as-welded bead edge, as shown in Fig. 1B-I. The coordinates of Feature a_1 and Feature a_2 are extracted and subsequently transformed into the active vision image coordinate system after centerline extraction, as shown in Fig. 1B-II. An ROI region is then defined in the active vision image to reconstruct the 3D profile of the as-welded bead at the current welding location, as shown in Fig. 1B-III. The width and height errors of the method are maintained within 5% and 7%,

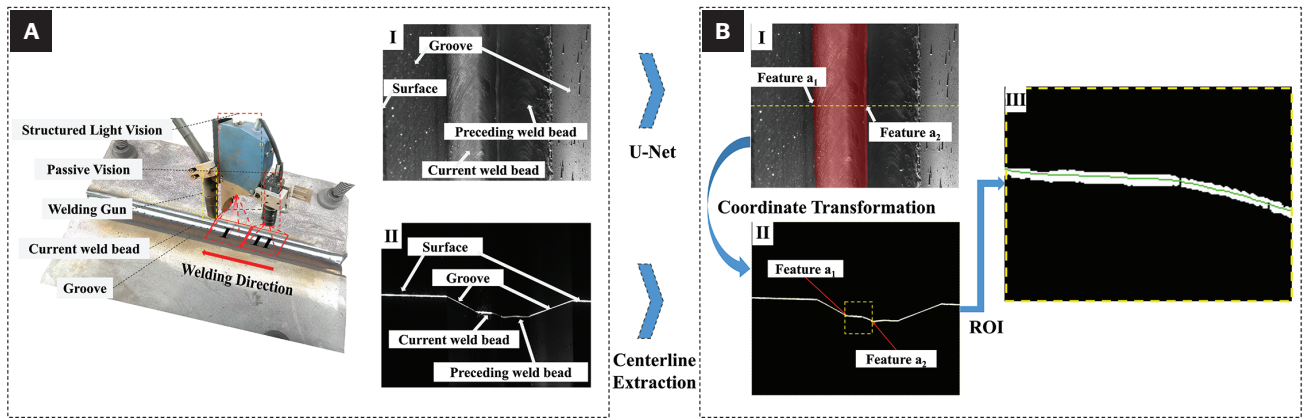


Fig. 1 – Method in Part 1: A – Compound vision architecture; B – the weld bead ROI extraction process.

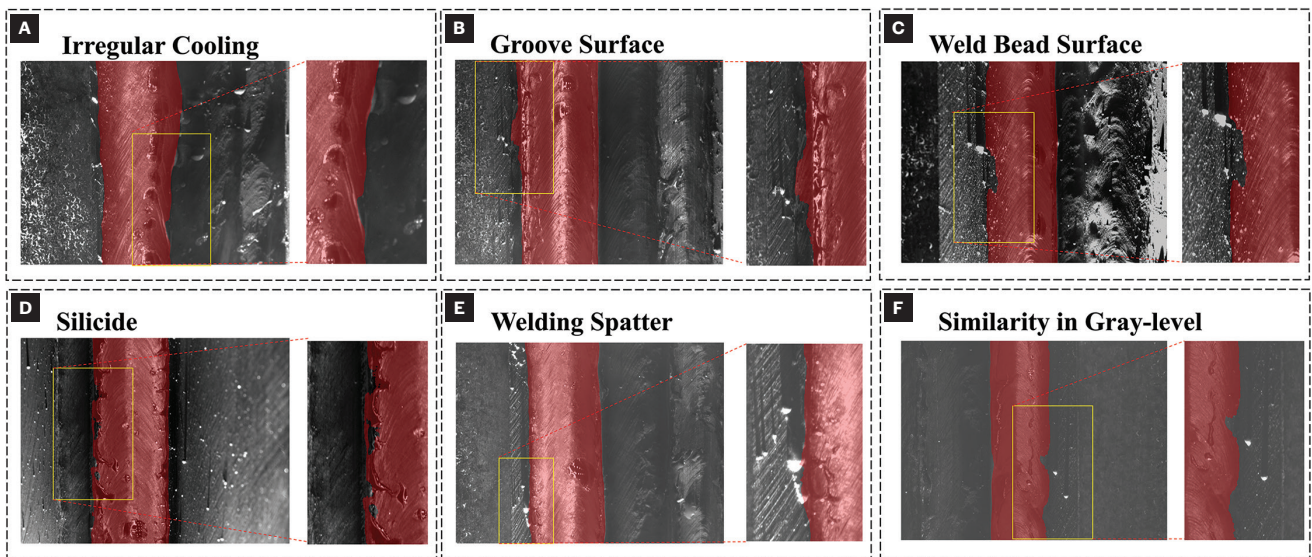


Fig. 2 – Difficult-to-recognize scenarios: A – Irregular cooling interference; B – groove surface interference; C – weld bead surface interference; D – silicide interference; E – welding spatter interference; F – similarity in gray-level interference.

respectively, which indicates that the approach is feasible and stable for the recognition of intermediate as-welded beads in MLMP welding scenarios.

To evaluate the method introduced in Part 1 for the first and last as-welded beads in a single layer of MLMP welding, we carried out a large number of experiments to investigate the identification of the first and last as-welded beads in a single layer of MLMP welding. The results indicated that the U-Net model performed poorly in these specific scenarios, yielding unstable and inaccurate recognition outcomes. This suboptimal performance was mainly attributed to the special positioning of these weld beads within the welding sequence, which makes their passive vision images particularly susceptible to occlusion by groove sidewalls, interference from welding spatter, and unclear surface textures (Refs. 3–5).

The complexity and challenges inherent in these scenarios make the U-Net model mentioned in Part 1 prone to mis-segmentation and incomplete edge detection when pro-

cessing images of the first and last as-welded bead. Figure 2 illustrates the model's performance differences in identifying as-welded beads. Specifically, Fig. 2A illustrates false edge detection caused by surface textures resulting from irregular cooling of the as-welded bead. Figure 2B illustrates the incorrect boundary detection caused by burn marks on the groove surface that result from the welding process. Figure 2C highlights the incomplete identification of the as-welded bead region due to the similar surface textures between the as-welded bead and the groove. Figure 2D presents feature occlusion caused by silicide precipitation on the as-welded bead surface, which obscures the as-welded bead and makes it difficult to determine its boundaries. Figure 2E shows the inaccurate boundary localization resulting from interference by welding spatter during the boundary extraction process. Finally, Fig. 2F shows boundary recognition errors due to the high gray-level similarity between the as-welded bead and the surrounding background. These segmentation results indi-

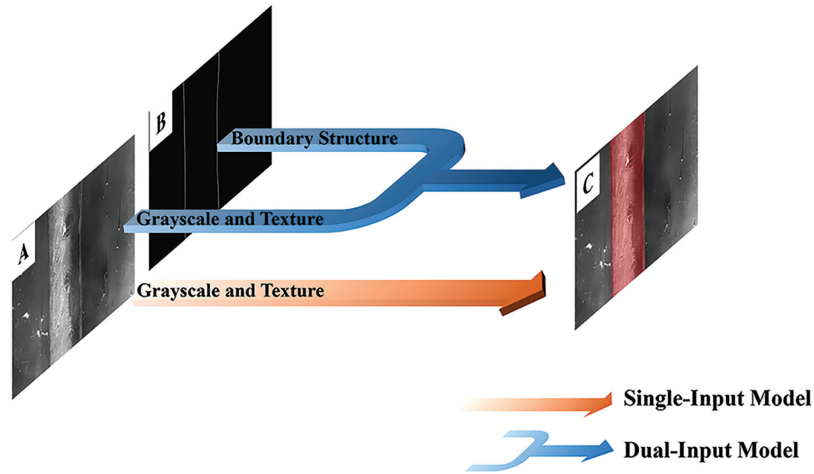


Fig. 3 – Image recognition algorithm models: A – Passive image; B – edge mask map; C – result.

Table 1 – Specific Types

Categories	Passive Image	Active Image	Details
Weld Bead Texture of the First Pass			
Weld Bead Texture of the Last Pass			
Silicides on the First Pass			
Silicides on the Last Pass			
Spatter Adjacent to the First Pass			
Spatter Adjacent to the Last Pass			

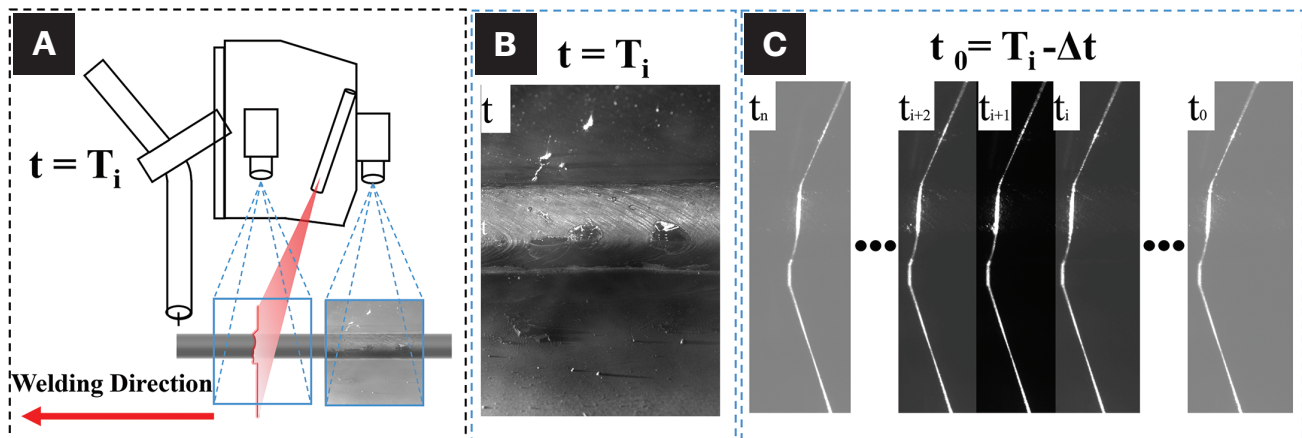


Fig. 4 – Temporal relationship: A – System architecture; B – passive image; C – active image sequence.

cate that the task of identifying the as-welded bead in MLMP welding is difficult for the image recognition model proposed in Part 1 to delineate the bead's boundaries accurately.

To address the aforementioned challenges, Part 2 will focus on optimizing the identification algorithm for the first and last as-welded beads, aiming to improve the model's performance in complex interference environments (Refs. 6–8). Specifically, a Dual-UNet is designed, in which the edge mask map extracted from active vision images is used together with the passive vision image as a pair of inputs. This image-processing strategy is expected to improve the accuracy and robustness of recognition of the as-welded bead boundary. The remainder of this paper is organized as follows: the Dual-Input Strategy and Dataset Preparation section introduces the dual-input strategy and outlines the dataset preparation process. The Weld Bead Segmentation Model section presents the detailed architecture of the as-welded bead segmentation model, including the Backbone, Neck, and Head subsections, as well as the loss functions. Additionally, the training and evaluation metrics of the Dual-UNet are described and compared with those of UNet and UNet+EA. The Weld Bead Recognition Experiment section presents the accuracy validation of the as-welded bead recognition experiments. First and last as-welded beads of different dimensions were obtained by intentionally offsetting the weld path. Recognition accuracy for bead width and height was compared with reference measurements. The 3D offset data for the as-welded bead is also verified. The conclusion summarizes the study's content.

Dual-Input Strategy and Dataset Preparation

Dual-Input Strategy

In the image recognition task of Part 1, the passive image is independently processed by a single-input UNet model to produce the segmented as-welded-bead image, as shown in Fig. 3. The model's performance is not ideal when identifying passive images of the first and last as-welded beads

affected by disturbances. This is because the passive images provide only grayscale and texture information, which the aforementioned interferences can easily disrupt during the welding process (Refs. 9, 10).

Compared to passive vision images, active vision images exhibit stronger anti-interference capabilities due to their laser stripe reflection imaging mechanism (Refs. 11, 12). As shown in Table 1, even under interfering conditions such as blurred texture, silicides, and spatter, the active image still preserves a laser stripe, despite slight interruptions, as shown in the second column. The stripe can be fitted by using a series of active image processing algorithms. The enlarged view of the disturbed area and the result after centerline extraction are displayed in the fourth and fifth columns of Table 1, respectively. It can be observed that the active vision image can extract the morphological structure of the as-welded bead at the same location. The structure possesses strong anti-interference capability and contains spatial positional information of the as-welded bead. Therefore, we use structural information to extract pseudo-edge points from the as-welded bead and generate an edge mask map registered to the passive image. The edge mask map, together with the passive image, serves as the input to the dual-input model, as illustrated in the Dual-Input UNet (Dual-UNet) structure depicted in Fig. 3. Data preprocessing will be detailed in the following sections.

Passive and Active Image Pairs

Passive images capture regional features of an as-welded bead on the weld plate surface, with each image corresponding to a specific time interval during welding. In contrast, active images reflect the surface contour of the as-welded bead at a single instant. The temporal relationship is illustrated in Fig. 4. Figure 4A shows that at a certain time, T_i , the welding torch equipped with both active and passive vision systems completes the image acquisition. However, the active image captured at this moment does not correspond to the region described in the passive image. Therefore, it is necessary to combine the welding speed, camera frame rate, and calibrated field-of-view dimensions to determine the corresponding

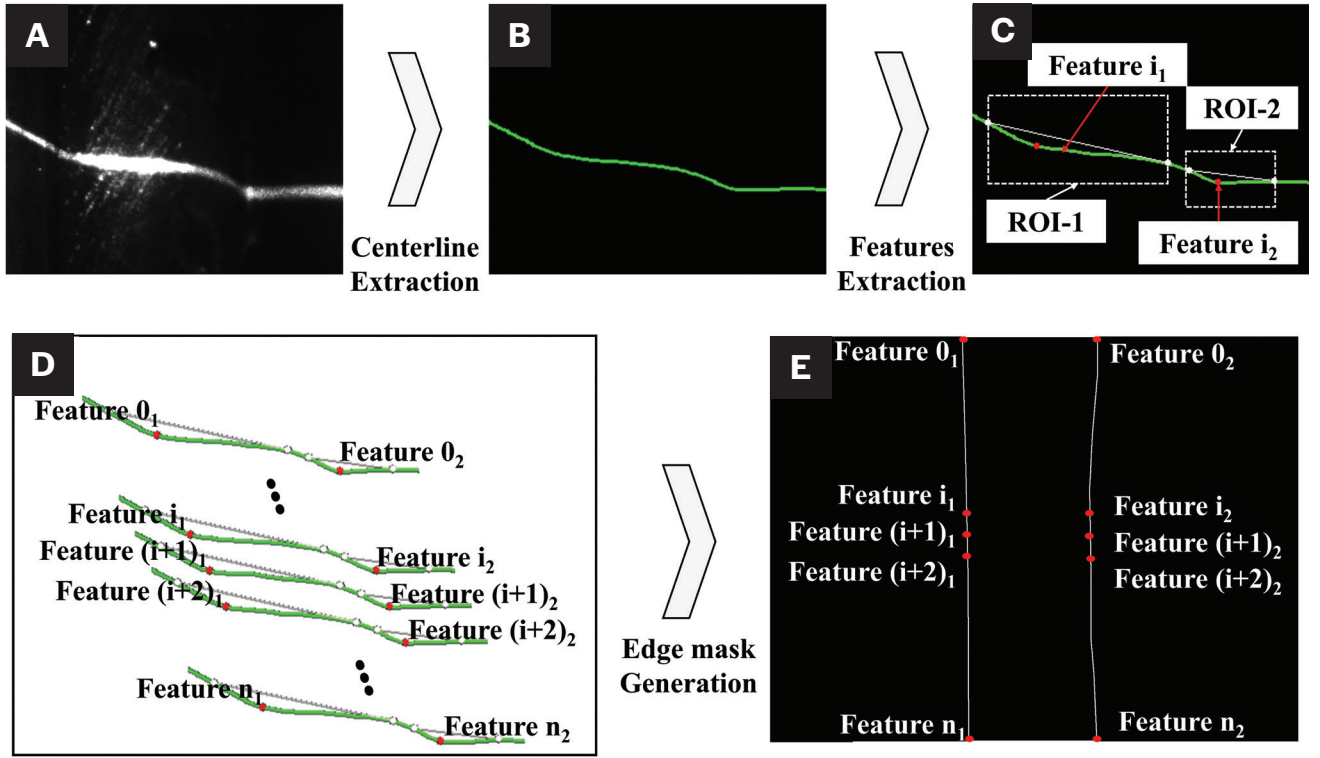


Fig. 5 – The edge mask map Proc: A – The original image of active vision; B – the centerline extraction result; C – the feature point extraction results; D – the feature point array; E – the edge mask map.

Table 2 – Database Statistical

Category	Numbers	Proportion of the Database
Grayscale Interference	115	11.5%
Silicides on the Bead	123	12.3%
Spatter Adjacent to the Bead	147	14.7%
Normal	615	61.5%
Total	1000	

active image sequence. Figures 4B and C show the result of a single passive image captured at T_i and its corresponding active image sequence, where $t_0 = T_i - \Delta t$. The value of Δt is determined by the relative position between the active and passive cameras and the welding speed. The calculation is given by Eq. 1:

$$\Delta t = \frac{L}{W_h} \quad (1)$$

Here, L denotes the distance between the active and passive cameras, and W_h represents the welding speed. As shown in Fig. 4C, this calculation provides a temporal alignment basis for image pair matching, ensuring accurate spatiotemporal correspondence between passive images and the active image sequence during the dynamic welding process.

After obtaining the active image sequence, an edge mask map must be extracted from it to serve as the model's second input. The specific processing workflow is shown in Fig. 5. Each active image (Fig. 5A) is first processed using the centerline extraction strategy discussed in Part 1, yielding the laser-stripe centerline shown in Fig. 5B. Subsequently, ROI regions 1 and 2 are defined based on the as-welded bead location. The polygon approximation method is then applied to extract as-welded features, namely $\text{Feature } i_1$ and $\text{Feature } i_2$, as illustrated in Fig. 5C. In this figure, the white lines denote the ROI boundaries obtained through the polygon approximation, and the red points represent the extracted feature points of the as-welded bead. Next, all feature points from the active image sequence are collected to form a feature point array, as shown in Fig. 5D. Finally, these feature points are mapped to their corresponding spatial positions in the passive image, generating an edge mask map shown in Fig. 5E. This process can be implemented using Python scripts to obtain edge-mask maps for all passive images, ensuring one-to-one naming correspondence to form active-passive image data pairs used for subsequent dataset construction.

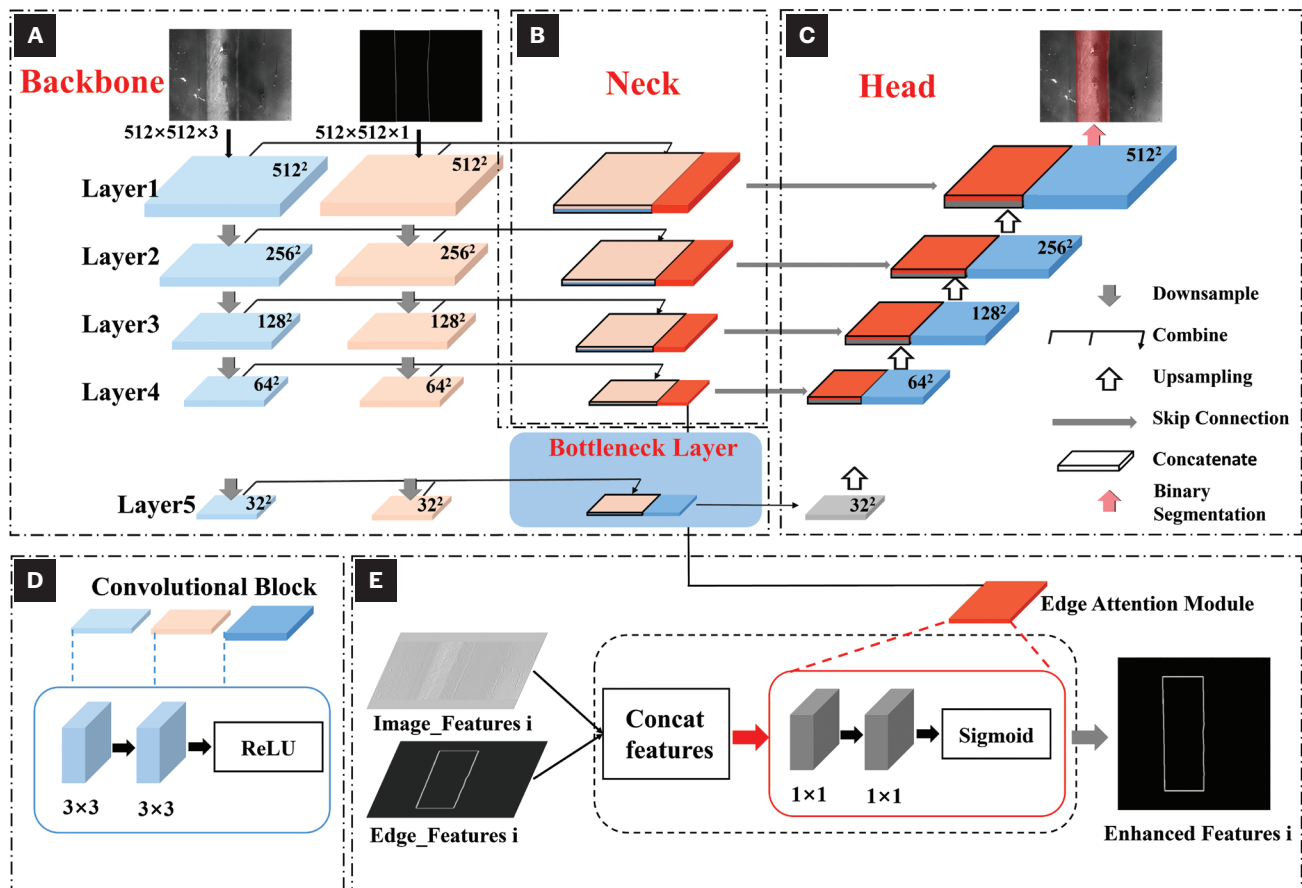


Fig. 6 – The architecture of the Dual-UNet model: A – Backbone; B – neck; C – head; D – the structure of the convolutional block; E – the feature fusion process.

Dataset Preparation

The welding experiments in this study were conducted on S355 steel plates measuring 300 mm $\times$ 400 mm $\times$ 32 mm, using a 60 deg. V-groove and multilayer multipass (MLMP) filler welding. A total of 17 steel plates were tested. To simulate the common challenges in real-world welding scenarios, such as light occlusion, brightness variation, welding spatter, and silicide precipitates, we adjusted the welding parameters and environmental conditions. The welding current was set in the range of 140 A to 240 A, based on typical values used in industrial thick-plate welding. For certain weld beads, the shielding gas flow rate was intentionally reduced from the standard 40 L/min to 20 L/min, leading to increased welding spatter. As the number of welding layers increased, heat accumulation and silicide deposition became more pronounced, leading to a greater number of silicide precipitates in the upper weld beads. Additionally, due to variations in welding position and the influence of arc light and ambient lighting, the captured weld bead images exhibited significant brightness variation. The active-passive vision system synchronously captured 164 sets of images for the first and last as-welded beads, yielding approximately 40,000 passive and 40,000 active images, respectively. From these, 1000 passive images were manually selected to construct a weld-bead image

database, which was used to validate the robustness and adaptability of the proposed composite vision method in complex, variable welding environments. Among these, 385 images contain interference factors, including gray-scale blurring, silicide formation, spatter, and groove texture. The database sample distribution is presented in Table 2. Subsequently, based on the corresponding relationships, the associated active image sequences were obtained, and edge mask maps were extracted accordingly. After annotating the passive images, the database was further augmented using techniques such as translation, mirroring, Gaussian noise addition, and brightness adjustment (Ref. 13). Ultimately, 5000 image pairs were obtained as the final dataset.

Weld Bead Segmentation Model

This section provides a detailed description of a Dual-UNet model designed to address the identification challenges posed by the first and last as-welded beads. The model adopts a dual-input configuration in contrast to the U-Net introduced in Part 1. The model uses an Edge-Attention (EA) module (Ref. 14) and an Edge-Aware loss function (Ref. 15). Compared to other UNet variants (Refs. 16, 17), the Dual-UNet model is specifically designed with respect to its architecture and feature fusion strategy, focusing

on processing dual-source images captured by the composite vision system in real time. This design is tailored for industrial welding applications, aiming to ensure robust, continuous extraction of the geometric contour of as-welded beads in complex, dynamic environments.

The overall architecture of the model is illustrated in Fig. 6, comprising three core modules: the Backbone, Neck, and Head. Figure 6A shows the structure of the Backbone module, which receives paired passive images and edge mask maps as input. The module encodes these inputs and passes the encoded features to the subsequent stage. Figure 6B shows the Neck module, which corresponds to the skip connection mechanism in the U-Net network. In this stage, the encoded features from the first four layers are fused and enhanced and then transmitted to the next stage via the skip connection. Figure 6C depicts the Head module, where the image resolution is gradually restored. After receiving the encoded features from the deepest layer, the module employs four decoding stages to progressively upscale the feature maps to the original input size, ultimately producing the segmentation output. Additionally, Figs. 6D and E illustrate the structure of the convolutional block and the feature fusion process. The details will be elaborated in the subsequent sections.

Backbone

The Backbone module is the network's fundamental feature-extraction component, as shown in Fig. 6A. The module comprises two parallel VGG16 encoder paths (Ref. 2), which are specifically designed to process the passive image and the edge mask map at a resolution of 512×512 . The passive vision image is processed via the image path to extract multi-scale grayscale and texture features of the as-welded bead, ranging from local textures to global shapes. The edge mask map is processed via the edge path to encode boundary information. Each encoding path consists of five encoding stages. Each encoding stage consists of five convolutional blocks. The convolutional block consists of two consecutive 3×3 convolutional layers followed by a ReLU activation function, as shown in Fig. 6D. The feature channels of the passive vision image and the edge mask map gradually increase after each convolutional block. Each path produces five feature results. These features maintain consistent spatial resolution and channel dimensions across each encoding stage and provide structurally aligned, hierarchically rich features for subsequent feature fusion. The first four feature results were passed to the Neck module via skip connections, where they are fused and subsequently enhanced before being passed to the Head module. Multimodal features from local details to mid-scale features are preserved in the as-welded bead image through this process. The feature results from the fifth stage are concatenated channel-wise in the Bottleneck Layer, processed by a convolutional block, and then fed into the bottom layer of the Head module. The fifth stage integrates multiscale features, generating a high-level, abstracted, and low-resolution spatial feature representation. This resulting feature encoding captures the global structure and semantic infor-

mation of the as-welded bead and serves as the contextual basis for the subsequent decoding stages.

Neck

The Neck module consists of four feature fusion layers, as shown in Fig. 6B. In each layer, corresponding layers of the Backbone are concatenated along the channel dimension to form a fused feature map that integrates texture and structural information. The as-welded bead boundary is enhanced by this feature concatenation mechanism, which enables the Dual-UNet to fuse image and edge features at different scales. An Edge-Attention module is introduced at each layer of the Neck module to dynamically modulate the fusion weights between image and edge features. This mechanism adaptively enhances the representation of edge regions, improving the recognition of as-welded bead boundary features. The feature fusion process is illustrated in Fig. 6E: First, `image_feat_i` and `edge_feat_i` are concatenated along the channel dimension to generate a fused feature map. Subsequently, two 1×1 convolutional layers are applied, which compress channels and transform features to extract edge attention features. The transformed feature map is fed into a Sigmoid activation function to generate attention weights. Finally, an element-wise multiplication is performed to integrate the attention weights, resulting in a weight-enhanced image feature map. The enhanced feature maps are passed to the decoding stage via skip connections. This mechanism allows the model to focus on edge regions through edge mask maps in complex welding scenarios.

Head

The Head module uses the U-Net decoder architecture, consisting of four decoding stages, as depicted in Fig. 6C. Each decoding stage comprises an upsampling operation, a skip connection mechanism, and two 3×3 convolutional layers with ReLU activation functions. First, the bottom decoding module receives the global features from the Bottleneck Layer and performs an upsampling operation. Then, it concatenates the weight-enhanced image features with the upsampled results via a skip connection. Finally, two convolutional layers with ReLU activation are applied to further restore image resolution and enhance the quality and edge integrity of the segmentation map. The resulting feature maps are then fed into the next decoding stage. The subsequent three decoding stages repeat these three operations, progressively mapping the feature maps from the deep semantic space back to the image space. At the topmost stage, the Head module performs binary segmentation, mapping the feature maps into a 2-channel output that distinguishes the as-welded bead region from the background. The final output resolution is 512×512 , consistent with the input image.

Loss Function

A multi-task joint loss function is introduced to ensure that the image segmentation model focuses on the edge infor-

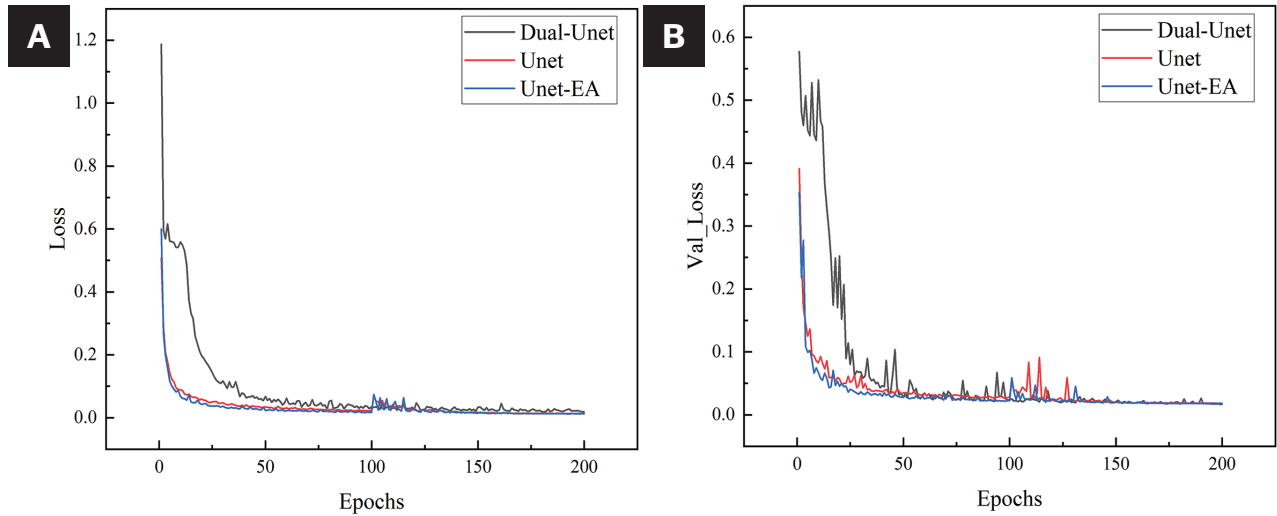


Fig. 7 – Model training curves: A – Training loss; B – val loss.

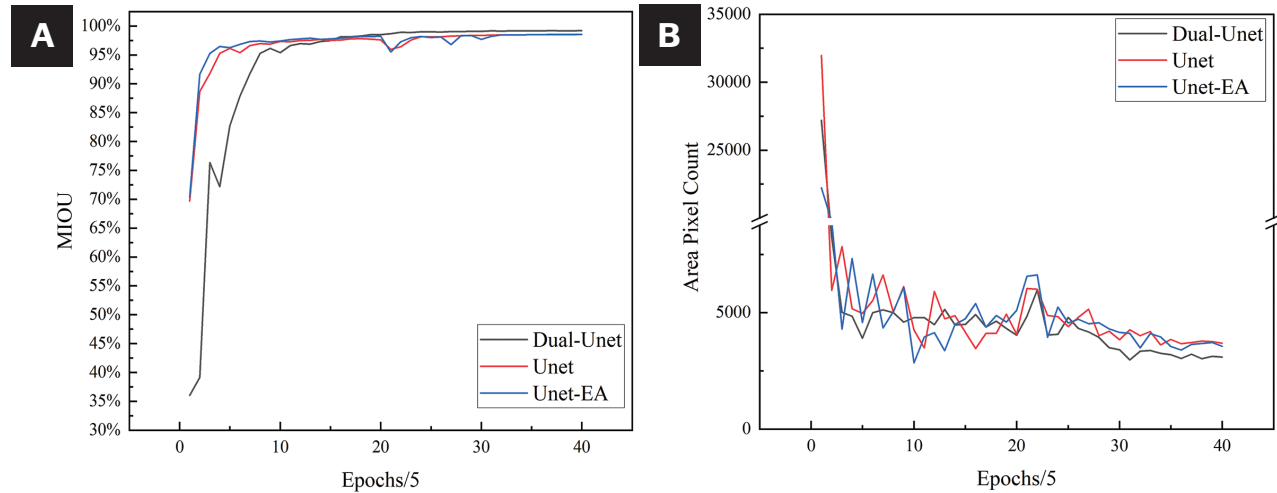


Fig. 8 – Evaluation metric curves: A – MIOU; B – FN Area.

mation provided by the edge mask map input during training. This loss function integrates Cross-Entropy (CE) loss, Dice loss, and Edge-Aware (EA) loss, which respectively impose constraints on the global distribution of the as-welded bead region, the overlap between the predicted segmentation map and the ground truth, and the preservation of edge details (Ref. 15). Among these, CE loss is one of the most commonly used loss functions in image segmentation. Its formulation is given by Eq. 2:

$$L_{CE} = - \sum_{i=1}^N y_{gt,i} \log(y_{pred,i}) \quad (2)$$

where $y_{gt,i}$ denotes the ground truth label of the i -th pixel, $y_{pred,i}$ in $[0,1]$ represents the predicted probability for the i -th pixel, and N is the total number of pixels in the image. CE loss provides a fundamental constraint on the model's classification and ensures that the predicted map aligns well with the distribution of the as-welded bead region. In addition, to further improve the model's ability to identify the overall

as-welded bead area, Dice loss is introduced. Its formulation is given by Eq. 3:

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N y_{pred,i} \cdot y_{gt,i}}{\sum_{i=1}^N y_{pred,i} + \sum_{i=1}^N y_{gt,i} + \varepsilon} \quad (3)$$

where ε is a smoothing term to avoid division by zero. This loss improves the overlap between predicted and ground-truth maps, particularly in small or blurred regions. Finally, Edge-Aware loss is adopted based on the Dice loss framework, using an edge mask map y_{edge} as input. The formulation is given by Eq. 4:

$$L_{Edge} = 1 - \frac{2 \sum_{i=1}^N y_{pred,i} \cdot y_{edge,i}}{\sum_{i=1}^N y_{pred,i} + \sum_{i=1}^N y_{edge,i} + \varepsilon} \quad (4)$$

where $y_{edge,i}$ represents the edge property of the i -th pixel. The final training loss function is defined as a linear combination of the three individual losses. The formulation is as follows:

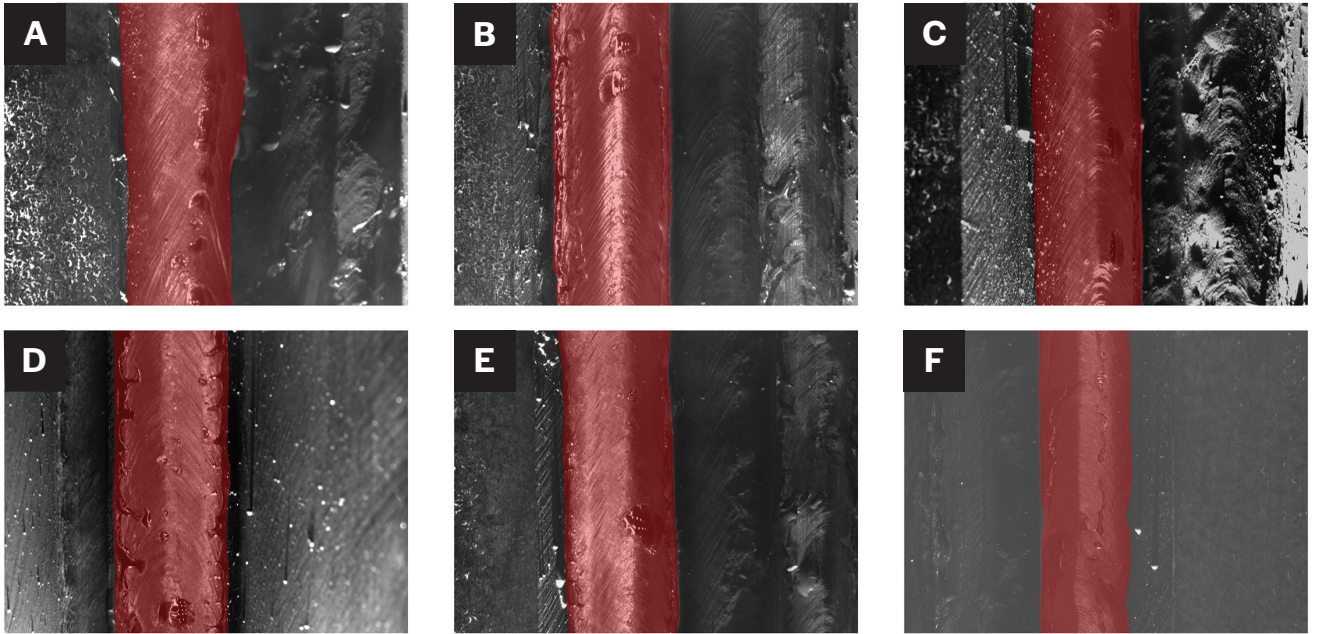


Fig. 9 – Identification results: A to F – Images mentioned in Fig. 2.

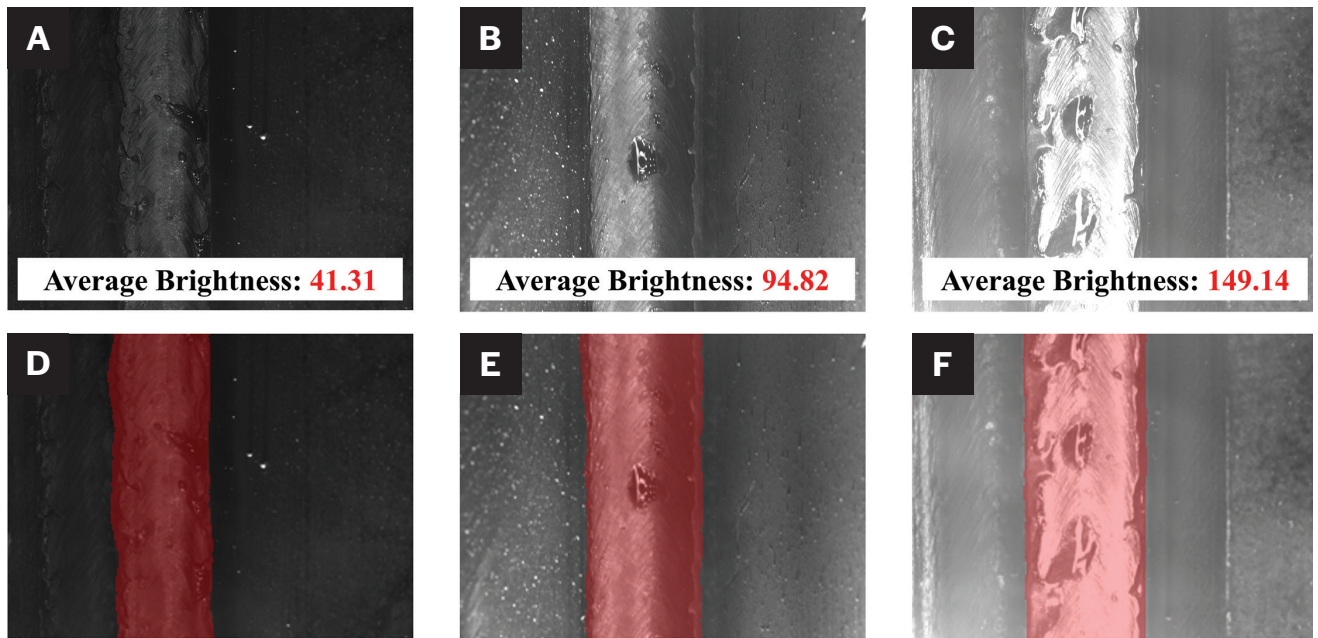


Fig. 10 – Different brightness conditions: A to C – Original weld bead images; E to F –their segmented results.

$$L_{total} = L_{CE} + L_{Dice} + L_{Edge} \quad (5)$$

In this work, all loss terms are equally weighted to balance the model's focus on both the global structure and edge details of the as-welded bead.

Training and Evaluation

The training platform for the proposed algorithm is configured with an Intel i9-13900 CPU, an NVIDIA RTX 4090

GPU with 24 GB of VRAM, and 64 GB of system memory. The software environment includes PyTorch 2.2.0, Python 3.8, and CUDA 11.8. For training, the Adam optimizer is used with an initial learning rate of 10^{-4} , which is decayed using a cosine annealing schedule to adapt to different training stages. The batch size is set to 4, balancing GPU memory constraints and data parallel efficiency to avoid out-of-memory errors. The total number of training epochs is set to 200 to ensure sufficient model training. During training, the training and validation losses, Mean Intersection over Union (MIoU) (Ref. 18), and the false negative (FN) area were recorded. We conducted comparative experiments among the Dual-UNet model, the U-Net

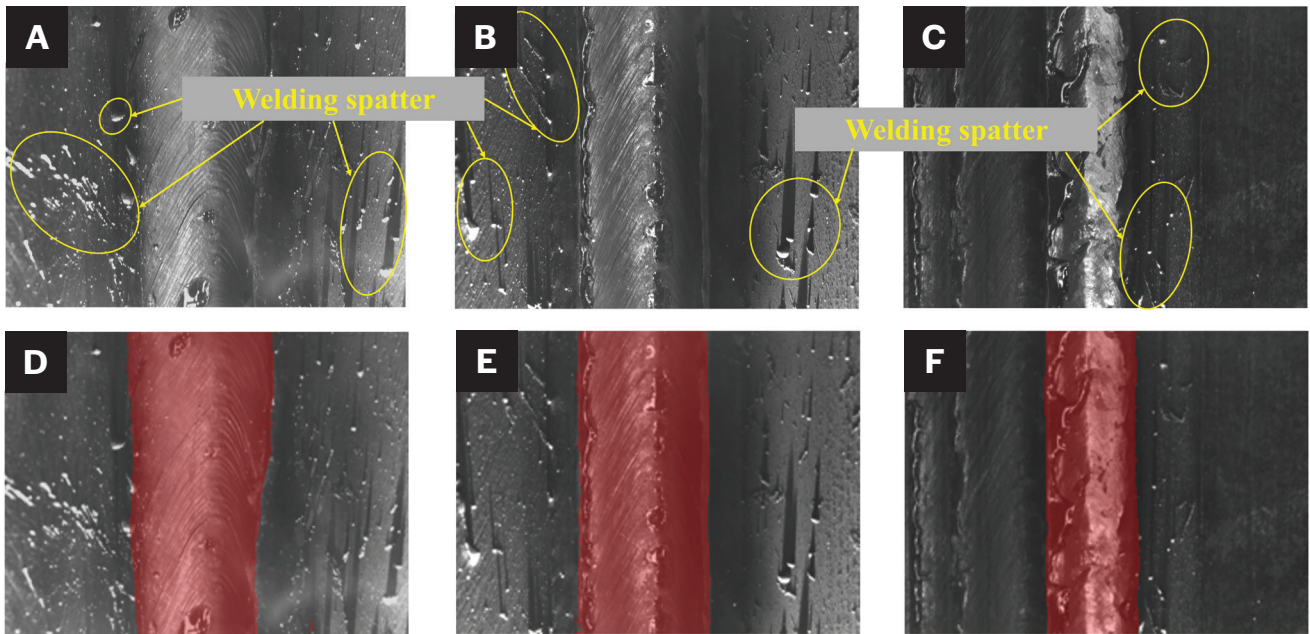


Fig. 11 — Welding spatter in passive images: A to C — Original weld bead images; E to F — their segmented results.

Table 3 — Welding Parameters

No.	Type	Welding Speed	Welding Current	Torch Position
18-2	Last Pass	6 mm/s	180A	Right
18-9	Last Pass	6 mm/s	150A	Right
18-10	First Pass	6 mm/s	200A	Right
19-5	Last Pass	6 mm/s	160A	Right
19-14	Last Pass	6 mm/s	140A	Left
19-15	First Pass	6 mm/s	220A	Left

model, and the U-Net+EA model to validate the recognition capability of the dual-input architecture.

The loss variation trends of the three models during both the training and validation phases are shown in Figs. 7A and B. The results indicate that the U-Net model demonstrates relatively good performance from the early stages of training, with a rapid decrease in loss. In contrast, the Dual-UNet model exhibits a higher initial loss, suggesting some difficulty in understanding the task at the beginning of training. However, the Dual-UNet model gradually converges and ultimately achieves a loss level comparable to that of the traditional model as the training proceeds. This suggests that although the dual-input structure introduces a slightly higher learning burden in the initial phase, it captures the task's key features more effectively and exhibits strong learning capability. This observation is also reflected in the validation loss curve.

As shown in Fig. 8A, the MIoU of all three models increases gradually during training. In the middle to later stages of train-

ing, the MIoU of the Dual-UNet model surpasses that of the other two models, reaching a final value of 99.1%. The FN Area is introduced as a supplementary metric to quantify the model's performance in terms of region coverage. Figure 8B presents the comparison results of the three models under the FN Area metric. The results demonstrate that all three models converge rapidly as the number of training epochs increases. Ultimately, the FN Area values of the traditional U-Net and U-Net+EA models approach 3500, while that of the Dual-UNet model approaches 3000. This indicates that the Dual-UNet model can achieve more accurate as-welded-bead identification.

The as-welded bead samples introduced in Fig. 2 are processed using the Dual-UNet model, and the segmented results are presented in Fig. 9. It can be observed that the Dual-UNet model achieves more accurate segmentation compared to the UNet model mentioned in Part 1. Furthermore, we collected representative weld bead images from

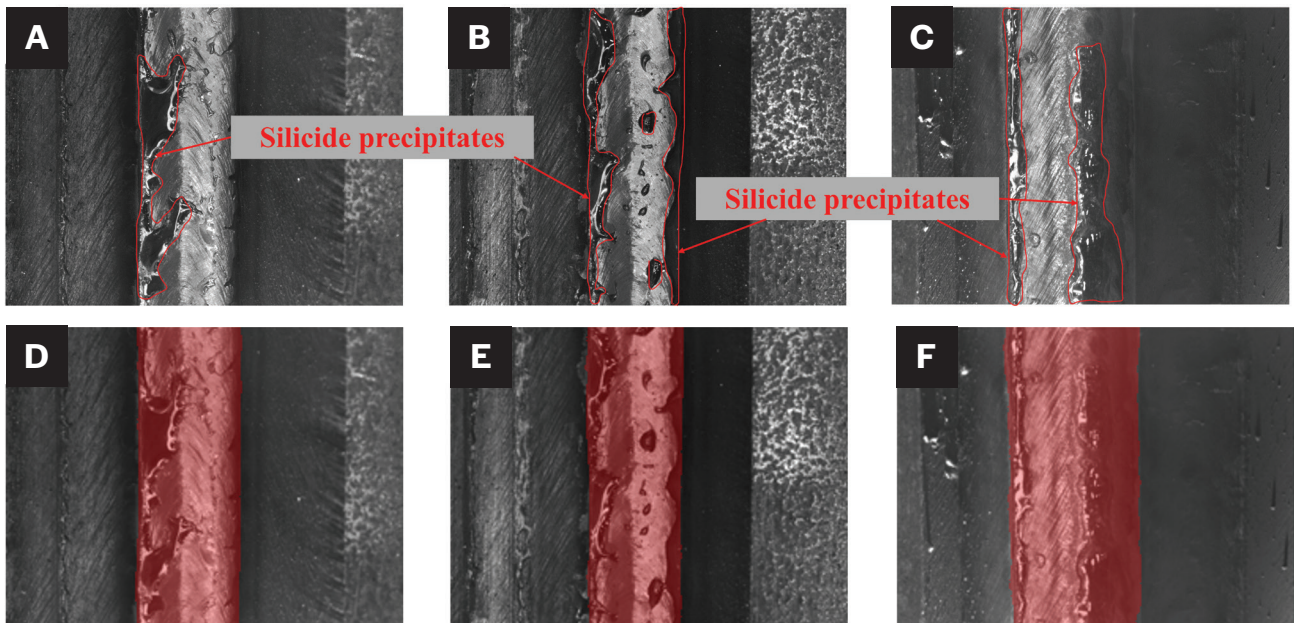


Fig. 12 — Silicide precipitates on the weld beads: A to C — Original weld bead images; E to F — their segmented results.

Table 4 — Accuracy Results

No.	Actual Width (mm)	Actual Height (mm)	Predicted Width (mm)	Predicted Height (mm)	Width Error (%)	Height Error (%)
18-2	8.02	3.30	7.92	3.30	1.25	0.00
18-9	8.54	2.26	8.16	2.48	4.45	9.73
18-10	9.45	6.54	9.31	7.06	1.48	7.95
19-5	9.80	3.72	8.87	4.02	9.49	8.06
19-14	7.67	2.86	7.88	2.89	2.74	1.05
19-15	11.65	3.07	11.17	3.45	4.12	12.38

the validation set, including those with light occlusion and brightness variation, welding spatter, and silicide precipitates. The segmentation results from the Dual-UNet model are provided to illustrate its performance further. Figure 10 shows weld-bead images and their segmented results under different brightness conditions resulting from light occlusion and brightness variation. Figure 11 presents images with prominent welding spatter and the associated segmented results. Figure 12 displays images containing extensive silicide precipitates along with the corresponding segmented results. The results indicate that by fusing edge information obtained from active vision, the Dual-UNet model is capable of accurately identifying the as-welded bead regions and effectively suppressing missegmentation caused by interfering factors. It achieves more complete and stable detection of as-welded bead contours, validating the superiority of the multi-modal input strategy in complex welding scenarios.

Weld Bead Recognition Experiment

Two sets of welding test plates (Plate 18 and 19) were designed using a layer-wise, pass-wise welding strategy. The arc initiation point of the welding torch was intentionally offset to produce irregular geometric variations in the as-welded beads during the welding process. Specifically, the arc initiation points of the subsequent layer were shifted backward by a specified distance to prevent overlap with the 3D morphology of the designated as-welded beads. This approach yielded six weld beads (18-2, 18-9, 18-10, 19-5, 19-14, and 19-15) with distinct geometric profiles for verification tests. The welding process parameters recorded for these weld beads are summarized in Table 3.

Sampling was conducted on the weld beads to obtain their actual geometric dimensions, with the sampling positions

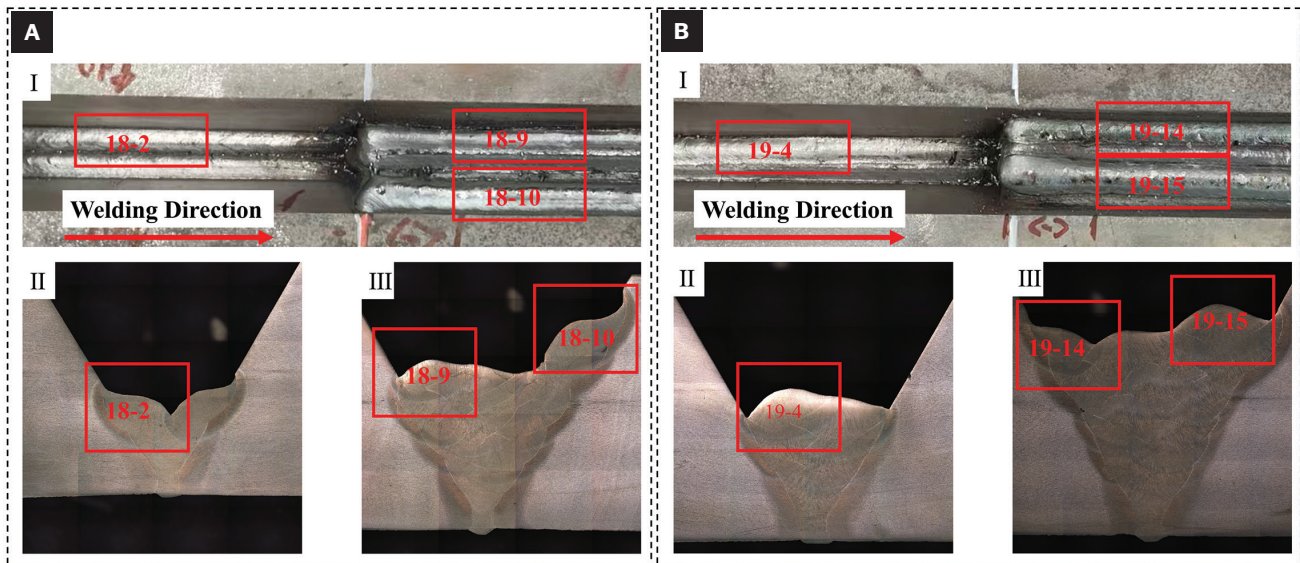


Fig. 13 – The sampling positions of the six weld beads: A – Welding plate No. 18; B – welding plate No. 19.

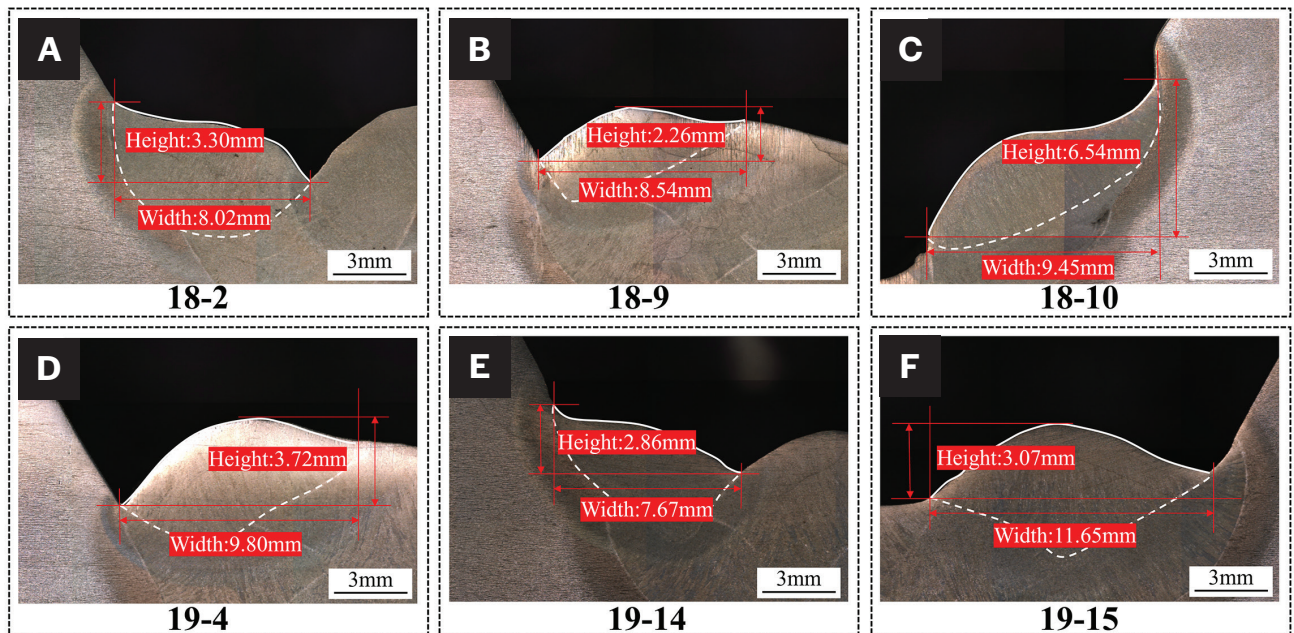


Fig. 14 – Results of the actual contour measurement: A – No. 18-2; B – No. 18-9; C – No. 18-10; D – No. 18-4; E – No. 18-14; F – No. 18-15.

illustrated in Fig. 13. The actual width and height of the six weld beads were then measured based on the metallographic test results. The actual weld-bead geometry is shown in Fig. 14. Following weld bead recognition by the Dual-UNet, surface contour extraction, and coordinate transformation on the corresponding weld bead images, the contour fitting results of the six weld beads from the image processing algorithm are visualized in Fig. 15.

Statistical error results comparing the fitted value and actual value are presented in Table 4. It can be observed that the maximum errors in the width and height of the algorithm's recognition results are 0.93 mm and 0.38 mm, respectively. The average error rates for the predicted width

and height are 4.09% and 6.53%, while the maximum error rates are 10.51% and 12.38%, respectively. The overall recognition accuracy reaches the millimeter level, enabling high-precision 3D contour measurement for various weld beads in MLMP welding processes. These results provide reliable dimensional references for subsequent optimization of the welding process.

The position information of the first weld bead No. 19-15 and the last weld bead No. 18-2, which exhibit typical three-dimensional morphology features, was extracted and used for 3D contour reconstruction, as shown in Fig. 16. Specifically, Figs. 16A and C present the 3D contour reconstruction results of the last weld bead, No. 18-2, and the first

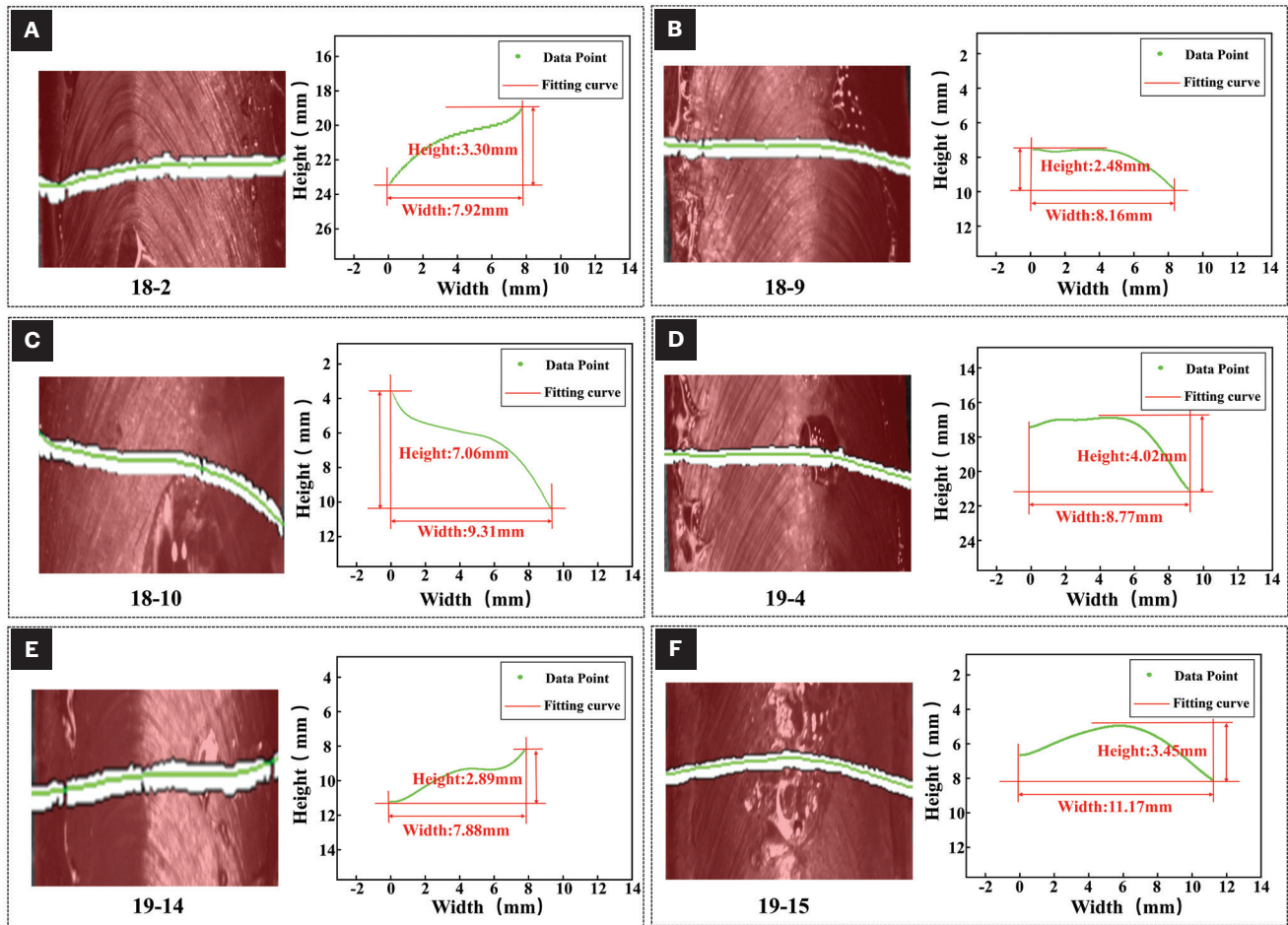


Fig. 15- Results of the fitted process: A – No. 18-2; B – No. 18-9; C – No. 18-10; D – No. 18-4; E – No. 18-14; F – No. 18-15.

weld bead, No. 19-15, respectively. Figs. 16B and D correspond to the post-welding-formed weld beads captured by a camera. As indicated by the coordinates marked in Fig. 16A, the left-side edge of the last weld bead No. 18-2 extracted by the algorithm shows a height deviation of 1.34 mm at 0.99 mm from the adjacent weld bead, and a width deviation of 3.48 mm. This deviation gradually decreases along the X-direction, leading to the formation of underfill and undercut defects in this weld segment, with a total length of 210.42 mm. This result is consistent with the welding path design in which the torch was intentionally moved close to the groove. Similarly, the coordinates marked in Fig. 16C indicate that the left-side edge of the first weld bead No. 19-15 extracted by the algorithm exhibits a height deviation of 2.28 mm at 0.99 mm from the adjacent groove, and a width deviation of 4.13 mm. This deviation also decreases along the X-direction, resulting in underfill, incomplete fusion, and undercut defects in this region, with a total length of 238.18 mm. This outcome aligns with the welding path design in which the torch was intentionally moved away from the groove. This method allows for the accurate determination of the weld offset dimensions and the identification of the positions where these offsets affect the welding quality. In practical welding processes, groove deviations due to thermal accumulation, stress superposition, and clamping errors may

alter the weld bead morphology. Such changes can significantly affect the arrangement of subsequent weld passes, potentially leading to underfill, undercuts, and incomplete fusion defects. The proposed 3D contour measurement technique enables real-time acquisition of accurate weld bead profiles, supporting welding engineers in developing path-planning algorithms for MLMP welding and conducting studies on weld correction methods to achieve intelligent path planning under closed-loop control and improve the level of automation and intelligence in the welding process. The investigation into real-time path planning and correction strategies will be further detailed in Part 3.

Conclusion

A Dual-UNet image segmentation algorithm was proposed to address the challenges in identifying the first and last weld beads in MLMP welding. This method integrates the edge mask map extracted from active vision images as the second input, which is temporally matched and fused with passive vision images to form a multi-modal image processing framework that combines 2D texture information with 3D geometric features. Experimental results demonstrate that the proposed method achieves average identification

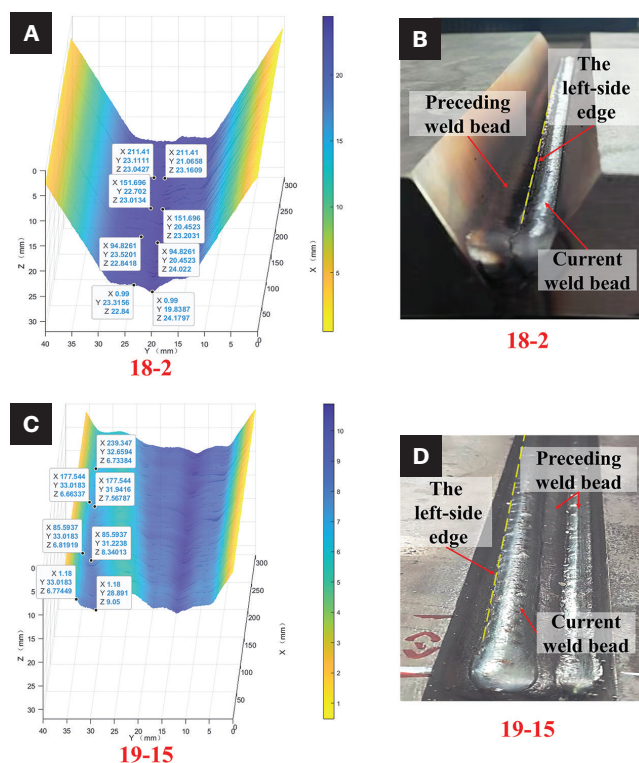


Fig. 16 — The 3D contour reconstruction results and actual morphologies of Nos. 18-2 and 19-15: **A** — 3D contour reconstruction result of No. 18-2; **B** — actual morphology of No. 18-2; **C** — 3D contour reconstruction result of No. 19-15; **D** — actual morphology of No. 19-15.

errors of 4.09% for bead width and 6.53% for bead height, with maximum errors of 0.93 mm and 0.38 mm, respectively. The identification accuracy reaches the millimeter level.

Two sets of welding test plates were used to verify the accuracy of the proposed algorithm with six weld beads. The 3D-reconstructed contour clearly reflects the geometric variations in the weld beads under different welding parameters. For the first bead No. 19-15, a height deviation of 2.28 mm at 0.99 mm from the adjacent groove and a width deviation of 4.13 mm were observed. For the last bead, No. 18-2, a height deviation of 1.34 mm at 0.99 mm from the adjacent weld bead, and a width deviation of 3.48 mm were observed.

References

- Wu, S., Li, C., Cheng, F., Wang, Z., and Jiao, J. 2025. Composite vision method for measurement of 3D bead profile, part 1—Principles. *Welding Journal* 104(9): 335S–350S. DOI: 10.29391/2025.104.026
- Ronneberger, O., Fischer, P., and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*: 234–241. DOI: 10.1007/978-3-319-24574-4_28
- Hu, Z., Qin, X., Li, Y., and Ni, M. 2020. Welding parameters prediction for arbitrary layer height in robotic wire and arc additive manufacturing. *Journal of Mechanical Science and Technology* 34(4): 1683–1695. DOI: 10.1007/s12206-020-0331-0
- Zhao, B. 2022. Multi-layer and multi-channel welding trajectory control method of welding robot. *Jordan Journal of Mechanical & Industrial Engineering* 16(1).

5. Cho, D.-W. 2019. Algorithms of mathematical models for fast multi-pass weld-bead drawing technique in V-groove welding. *Journal of Mechanical Science and Technology* 33(11): 5347–5352. DOI: 10.1007/s12206-019-1027-1

6. Ramadan, R., and Aly, S. 2022. CU-net: A new improved multi-input color U-net model for skin lesion semantic segmentation. *IEEE Access* 10: 15539–15564. DOI: 10.1109/ACCESS.2022.3148402

7. Wang, J., Wang, S., Liang, W., Zhang, N., and Zhang, Y. 2022. The auto segmentation for cardiac structures using a dual-input deep learning network based on vision saliency and transformer. *Journal of Applied Clinical Medical Physics* 23(5): e13597. DOI: 10.1002/acm2.13597

8. Khan, M. W. 2014. A survey: Image segmentation techniques. *International Journal of Future Computer and Communication* 3(2): 89. DOI: 10.1007/978-981-13-0761-4_105

9. Yang, X., Cheng, L., Yuan, G., and Wu, H. 2024. Texture-aware neural network for edge detection. *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*: 254–268. DOI: 10.1007/978-981-97-8505-6_18

10. Dong, R., Wang, B., Li, S., Zhou, Z., Li, S., and Wang, Z. 2016. Interactive image segmentation with color and texture information by region merging. *Chinese Control and Decision Conference (CCDC)*: 777–783. DOI: 10.1109/CCDC.2016.7531090

11. Lu, J., Zhao, Y., Chen, X., Han, J., and Zhao, Z. 2023. Detection of reinforcement of multi-bead and multi-layer weld in additive manufacturing based on on-line visual information of weld pool. *Journal of Materials Research and Technology* 23: 4678–4690. DOI: 10.1016/j.jmrt.2023.02.101

12. Han, Y., Fan, J., and Yang, X. 2020. A structured light vision sensor for on-line weld bead measurement and weld quality inspection. *The International Journal of Advanced Manufacturing Technology* 106(5): 2065–2078. DOI: 10.1007/s00170-019-04450-2

13. Yuan, T., and Hu, B. 2025. REU-net: a remote sensing image building segmentation network based on residual structure and the edge enhancement attention module. *Applied Sciences* 15(6). DOI: 10.3390/app15063206

14. Al Qurri, A., and Almekkawy, M. 2023. Improved UNet with attention for medical image segmentation. *Sensors* 23(20): 8589. DOI: 10.3390/s23208589

15. El Bahi, O., Qaraai, Y., and El Allaoui, A. 2025. Deformation field prediction based on a modified loss function with U-net for non rigid image registration. *Signal, Image and Video Processing* 19(10): 857. DOI: 10.1007/s11760-025-04452-6

16. Wang, Z., Fu, S., Zhang, H., Wang, C., Xia, C., Hou, P., Shun, C., and Shun, G. 2025. Dual-branch dynamic hierarchical U-net with multi-layer space fusion attention for medical image segmentation. *Scientific Reports* 15(1): 8194. DOI: 10.1038/s41598-025-92715-0

17. Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P., and Johansen, H. D. 2020. Doubleu-net: a deep convolutional neural network for medical image segmentation. *IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS)*: 558–564. DOI: 10.1109/CBMS49503.2020.00111

18. Behera, S. K., Rath, A. K., and Sethy, P. K. 2021. Fruits yield estimation using Faster R-CNN with MIOU. *Multimedia Tools and Applications* 80(12): 19043–19056. DOI: 10.1007/s11042-021-10704-7

JUNFU JIAO, CHUANG LI, SENYU WANG, and SHAOJIE WU (shaojie@tju.edu.cn) are with the School of Materials Science and Engineering, Tianjin University, China. **FANGJIE CHENG, ZHIJIANG WANG, and SHAOJIE WU** are with Tianjin Key Laboratory of Advanced Joining Technology, Tianjin, China. **ZHE LI** is with China Nuclear Power Engineering Co. Ltd., Beijing, China. **WEI XU** is with Offshore Oil Engineering Co. Ltd., China.