scientific reports

Check for updates

OPEN

High resolution weld semantic defect detection algorithm based on integrated double U structure

Xiaoyan Li[1], Yi Wei[1]✉, Zhigang Lv[1], Peng Wang[2]✉, Liangliang Li[2], Mengyu Sun[3]✉ & Chu Wang[1]

As a key industrial equipment, the welding quality of pressure vessels is directly related to the operation safety, and X-ray nondestructive testing is an important means to evaluate the weld quality. In this study, a high-resolution semantic segmentation algorithm based on a double U-shaped network was proposed to meet the requirements of pressure vessel weld defect detection, and accurate detection was realized through innovative three-stage processing flow. Firstly, the weld area was automatically extracted from the original X-ray films (3000+×800 + pixels) by the Gaussian positioning algorithm. Then, a multi-image hybrid stitching technology was proposed to reconstruct the long weld into a standard size of 1500 × 1500, which not only expanded the WSCR data set, but also effectively improved the data imbalance problem. In terms of model architecture, by improving U2Netp and UNet networks, MC-SPP module (multi-connection spatial pyramid pooling), RMAG module (residual multi-add gating recurrent unit), HDC-CBAM module (hybrid dilated convolution-convolutional block attention) and CCM module (cross-layer connection fusion) were integrated to form a cascade network with multi-level feature fusion capability. The experimental results showed that the model could effectively segment the defects such as cracks, pores, slag inclusion, incomplete fusion and incomplete penetration in the weld film, and the mIOU value reached 78.1, which was significantly superior to the existing methods, and provided an efficient and reliable solution for welding defect detection of pressure vessels, which had important engineering application value.

Keywords Defect detection, Semantic segmentation, Welding seam extraction, Data enhancement

Pressure vessels refer to a closed vessel that can withstand internal pressure and is widely used in manufacturing, aerospace, military industry and other industries. The welding quality of pressure vessels is directly related to production safety and public safety, so it is particularly important to carry out accurate quality inspection of welds. At present, X-ray imaging technology is often used for nondestructive testing of welds. The films of welds are obtained by scanning, and then the defects are identified and processed by professional defect appraisers. Common weld defects include cracks, pores, slag inclusion, incomplete fusion and incomplete penetration[1].

However, in this process, due to the uncertainty of manual detection, there are often missed judgments, misjudgments, inconsistent standards and large judgment errors. In addition, the manual detection process is not only time-consuming and laborious, but also difficult to meet the requirements of green science development and efficient production. Therefore, with the help of advanced scientific and technological means, especially automation and intelligent technology, accurate identification of weld defects has become an important direction in the current scientific research field.

Research on weld extraction

In practical engineering operation, the obtained films of weld defects usually have high resolution, and they contain irrelevant information such as image quality indicators[2], photo date, work-piece serial number, etc., while weld defects usually only account for a very small part of the whole image. If the complete image is directly identified and analyzed, it will consume a lot of computing resources. Therefore, it is necessary to pre-extract the weld area first to improve the processing efficiency.

[1]School of Electronics and Information Engineering, Xi'an Technological University, Xi'an 710021, China. [2]School of Institute for Interdisciplinary and Innovation Research, Xi'an Technological University, Xi'an 710021, China. [3]School of Opto-electronical Engineering, Xi'an Technological University, Xi'an 710021, China. ✉email: weiyi@st.xatu.edu.cn; wp_xatu@163.com; sunmengyu@st.xatu.edu.cn

Based on the lightweight network SOLOv2d, Zou Y et al. proposed a method to extract weld area from weld images with strong noise. This method used the characteristics of three laser bands to train, effectively eliminated the influence of arc and splash in practical engineering, could accurately identify weld stripes, and had high robustness[3]. Rabah Abdelkader and others first defined the weld area as the ROI (Region of Interest), then used the FCM (Fuzzy C- Means) algorithm to extract the initial contour, and finally used the initial curve of the traditional Chan-Vese model to segment the image. This method could improve the adaptability of the algorithm to noise while ensuring segmentation accuracy[4]. Wei Yang et al. proposed a two-stage ROI extraction algorithm for welds, which first roughly determined the ROI of welds based on the distribution of weld centers, and then accurately located the boundaries of ROI through vertical distance curves. Subsequently, feature pyramid network (FPN) and convolution attention mechanism module (CBAM) were added to the feature extraction part, and K-means clustering was used to adjust the prior frame settings. This method had achieved remarkable performance improvement on weld data sets[5]. Xinghui Dong et al. designed an automatic weld detection system in which the center line of the weld was first located by using an RF (Random Forest) regression device, and then the area around the weld was inspected by an RF classifier, so as to detect weld defects. This method could achieve more accurate defect location and reduce the dependence on complex feature extraction[6]. Zhicheng He and others put forward an AFFNet (Attention-Enhanced Feature Fusion Network), which could significantly enhance the ability of feature detection, thus accurately identifying weld details, even reaching pixel-level accuracy. In order to further improve the accuracy of segmentation, the author designed a weighted dice loss function, which made the system achieve the ideal effect more efficiently in the processing process. The experimental results showed that the algorithm had excellent performance in the positioning tests of straight weld and curved weld[7].

Generally speaking, by combining advanced technologies such as deep learning, feature fusion and attention mechanism, these methods not only improved the accuracy of weld area extraction, but also showed strong ability in dealing with complex noise and detail recognition, which provided new ideas and technical support for the development of the weld defect detection field.

Research on data enhancement

The existing X-ray films of weld defects are few, which bring challenges to the subsequent defect identification and detection. Aiming at the problem of small data, many researchers used data enhancement to expand the image data set. Common enhancement methods include mirror image flipping, brightness adjustment, scaling, adding noise and so on[8]. However, traditional augmentation methods are usually simple, with relatively single types of image variations, which can easily lead to over-fitting during the subsequent training process. In order to deal with this problem, many scholars put forward improved methods.

Satish Sonwane et al. pointed out that there was an imbalance in the types of weld defects in the initial data set. Therefore, they enhanced the data by adding Gaussian noise, rotation, shearing and reflection. Among them, rotation randomly selected an angle between 35 and 55 around the center point, and horizontal shearing introduced random horizontal displacement within the range of 0–20. After statistical analysis, the results showed that this method not only expanded the number of data sets, but also balanced the proportion of defect categories, and could better simulate the real camera angle and perspective, thus improving the robustness of the model[9]. Qiang Zhang and others used a sliding crop method to divide the target image into smaller patches suitable for the network, and performed operations such as random rotation, flipping, and adding Gaussian noise on the cropped images to further expand the data set[10]. In the process of cropping, in order to avoid the problems caused by image patching, H. Xu et al. made a preliminary scaling of the whole image before cropping, so as to adapt its size to the subsequent processing, which not only ensured the integrity of the defective area, but also avoided the risk of losing the defective image after cropping[11]. When dealing with high-resolution images, Xin Wang et al. first randomly cut the images and then applied a rotation matrix to rotate the cut images. This approach helps to avoid over-fitting and increases the number of training samples[12]. Weichao Qian and others put forward a weak supervision network. Aiming at the problem of denoising and enhancement of X-ray films of low-light pipeline welds, firstly, the illumination was enhanced by an improved Retinex-Net semi-supervision network, so as to eliminate artifacts, distortion and overexposure. Subsequently, they designed a new denoising network, which could not only remove noise, but also preserve image details, thus generating clearer X-ray images[13]. Liangliang Li et al. proposed two data enhancement modes according to the classification of different defect images: SISD(Single Image Single Defect) and SIMD (Single Image Multi Defects). This method could generate the specified defect image on the defect-free film on demand, balance the distribution of defect categories in the data set, and generate defect images and annotation files at the same time, reducing the workload of subsequent manual labeling[1].

To sum up, many researchers had effectively expanded the data set through innovative data enhancement methods, and improved the generalization ability and robustness of the model.

Research on defect detection

With the wide application of deep learning, the method of neural network in image content detection has become the mainstream day by day, especially in the nondestructive detection of X-ray weld defects, image segmentation technology is widely used. Lei Yang et al. proposed an innovative DCNNs(deep convolution neural network) for nondestructive defect detection. By introducing the attention fusion block(AFB) module to replace the traditional jump connection, the discrimination of features was improved, and the segmentation performance of tiny defects was enhanced. At the same time, they also introduced the RDCCB(residual densely connected convolution block) module, which further enriched the information of local feature maps and improved the overall performance of the model[14]. Li L and others proposed a high-resolution X-ray welding film defect detection model combining mixed expansion convolution and cross-layer feature fusion attention

mechanism. On the basis of the traditional UNet network, the structure of the encoder and jump connection was improved to enhance the feature extraction ability of the model, and the task of segmenting weld defects of solid rocket motor shell was successfully completed[15]. Rui Zhang et al. first constructed a high-performance defect detection model through feature information enhancement and multi-scale modules, and made adaptive optimization based on it, so that the model could obtain the best super-parameter setting. They also introduced HCOA(Hierarchical Cooperative Optimization Algorithm) optimization strategy, which not only reduced the difficulty of model construction, but also improved the adaptive learning ability of the model[16]. Jung Eun Kwon proposed a welding defect detection method based on twin networks. By embedding sub-models with different scales in each output module, this method could effectively detect welding defects with various sizes and aspect ratios. In addition, the model also had the function of image adjustment, which could adaptively predict the optimal filtering parameters of welding defect detection, and significantly alleviate the problems of low brightness and low contrast of data sets[17]. Cheng Ji and others proposed a weld defect detection method based on attention mechanism and deep learning. Based on the two-stage detection network Fast-RCNN, they designed a new visual attention mechanism SPAM (squeeze and position attention mechanism). The model integrated FPN (feature pyramid network) into Faster-RCNN, and introduced SPAM module into ResNet, the backbone network, thus enhancing the ability to detect defects of different sizes. This model could effectively detect small targets with weak contrast in a complex background and balance the detection accuracy of different defects[18]. Aminzadeh A et al. studied how to characterize the geometry and welding quality of aluminum alloy 5052-H32 laser welded plate by real-time 3D scanning technology, and explained that geometric defects (such as excessive welding bulges and shrinkage cavities) will affect the overall quality of the weld. Obtaining detailed geometric data of the weld through 3D scanning was helpful to identify the defect areas in the process of weld segmentation, especially by deep learning or image processing[19] The team also discussed the construction of an intelligent aluminum alloy laser welding plate production plant based on Industry 4.0 technology, focusing on how to improve the intelligent level of the welding process by integrating sensors, Internet of Things, big data analysis and other technologies. By integrating sensors, real-time data analysis and an automatic control system, weld defect segmentation could be closely combined with the production process, and comprehensive monitoring and dynamic optimization of weld quality could be realized. Intelligent factory and automatic control ideas combined real-time monitoring with weld defect detection, which greatly improved the efficiency and accuracy of segmentation. It provided strong support for weld defect segmentation[20].

The research of the above scholars had effectively improved the accuracy of defect detection through their own innovative methods, and provided valuable reference ideas for the design of network structure in this paper. However, most of these studies focused on image segmentation or image detection. Although they could effectively identify defects, they could not identify defects and sketch their exact shapes at the same time. Therefore, the method of semantic segmentation was chosen in this paper.

Sicheng Liu et al. used UNet network to detect the target and integrated CA (coordinate attention) and ViT (vision transformer) modules, which successfully improved the ability of the traditional UNet model to capture global context information. Through this method, they had achieved higher accuracy with less computational overhead[21]. Mingze Li and others put forward PipeTransUNet model, which combined Transformer and UNet networks, integrated a convolution block attention module, improved the activation function, and further enhanced the ability of defect extraction and location of fusion network[22]. The MSSs (multiple scale spaces) model designed by Tianyuan Liu and others fully applied the concept of multi-scale in different stages of network processing, and preserved the semantic information of context space to the maximum extent through the integration of convolution, down-sampling and attention mechanism. The experimental results proved the effectiveness of its theory[23]. Liu Q and others designed the Fill-UNet model. By introducing two innovative mechanisms, namely Semantic Collaborative Filtering Attention Module (SCFAM) and Multi-scale Semantic Feature Association Mechanism (MSFA), combined with Transformer architecture, the segmentation performance could be significantly improved. Among them, SCFAM module could effectively filter low-level semantic information and enhance perception ability, Transformer structure was good at extracting global context information, and MSFA mechanism realized deep fusion of multi-scale semantic features. Experiments showed that the combination of attention mechanism and Transformer structure was effective in semantic segmentation tasks, especially for capturing local details and global context at the same time, which provided ideas for network design that needs to capture details[24]. Tan J and others innovatively put forward the DarkSegNet model. By combining image pyramid decomposition with spatial low-frequency attention (SLA) and channel low-frequency information enhancement (CLIE) modules, the potential information of low-light images could be effectively mined, which was especially suitable for practical low-light scenes such as rescue of unmanned aerial vehicles after disasters, and had important application value. The research idea of this method was also helpful to segment images with inconspicuous contrast[25].

Although the existing research had made remarkable achievements in seam extraction, data enhancement and defect detection, the data set used in this paper faced some specific challenges, which made the direct application of the existing methods difficult. Firstly, the number of data sets was relatively small, and the image resolution was high, which might cause the problem of insufficient data in the training process, thus affecting the generalization ability of the model. Secondly, the gap between the transverse and longitudinal ratios in the weld area was too large, which might lead to the deviation of the model in detection, especially in the weld area with large difference in the ratio of length to width. In addition, the number of weld defect types in the data set was unbalanced, and the samples of some defect types were far less than others. This imbalance might lead to the lack of recognition ability of the model. Finally, the distribution of defects in the weld was quite different. Some defects might be concentrated in a specific area of the weld, while others were scattered. This uneven distribution of defects increased the difficulty of defect detection. Therefore, based on the inherent problems of these data

sets, the existing weld defect detection methods were difficult to be directly applied, and further optimization and adjustment were needed for the particularity of the data sets.

This paper first used the Gaussian function to locate and cut the weld position in the negative film to ensure the accuracy of subsequent processing. Then, by splicing and cutting a number of welds, the data set WSCR used in this study was made, which not only increased the number of data sets, but also balanced the number of different defect types. Finally, a lightweight integrated double U-shaped network structure was designed, which aimed to capture the context semantic information flexibly and had a large receptive field, so as to complete the segmentation task of various defects in the weld with high precision.

To sum up, the main contributions of this paper are as follows:

(1) An algorithm of seam extraction based on Gaussian function was proposed. The algorithm could effectively extract the seam area in the original data set with low computational cost, which laid the foundation for the subsequent weld identification task.
(2) A data enhancement method based on multi-image hybrid stitching was proposed, and a new data set WSCR was constructed, which not only expanded the number of data sets, but also could fuse different types of weld defects in one image, effectively balancing the difference in the number of defects between different welds and making up for the imbalance of defect types in the data set.
(3) A high-resolution weld semantic defect detection network based on an integrated double U-shaped structure was designed, which combined the improved U2NetP, UNet networks and cascaded them. Specifically, the U2NetP improved network introduced the MC-SPP module and the RMAG module, which enabled the network to pay more attention to the details of tiny defects. At the same time, the structure of UNet_CCM network was designed. On the basis of UNet, HDC-CBAM and a CCM were added, and the backbone network was lightened to better capture the contextual semantic information. The cascade of two innovative networks achieved high-precision segmentation effect with low memory occupation, and compared with other advanced classical networks, it showed better performance, which fully verified the feasibility and excellence of this network.

An efficient and accurate solution for weld defect detection was provided through the above contributions, and its potential and advantages in practical application were shown.

System architecture

System configuration

There are some problems in the open data set of pressure vessel weld defects, which seriously affect the effect of network training and detection accuracy. Specific issues include:

(1) The proportion of weld area is too small: the proportion of weld area in the whole image is usually small. If a complete image is used for training, the algorithm will have to deal with numerous of useless areas, resulting in a waste of training resources. In addition, the position and width of the weld in the image are irregular, which makes it difficult to extract the weld area directly.
(2) Category imbalance: In the public data set, the number of samples of different types of welding defects is quite different, resulting in a serious imbalance of training data. This imbalance may lead to the poor recognition performance of the model for some defect types and affect the overall detection effect.
(3) Small target detection problem: the images in the data set are high-resolution images, and the welding defects to be detected belong to small targets. Therefore, how to accurately segment and detect these small targets under the premise of ensuring the context information is an important challenge in the training process.

In order to solve the above problems, an intelligent algorithm scheme for welding defect detection of high-resolution pressure vessels was proposed. The overall structure of the scheme is shown in (Fig. 1). Specifically, the core technologies of the scheme include the following aspects:

(1) Weld area extraction: in the weld X-ray film containing weld area, non-weld area and character identification, it was necessary to extract the sub-image only containing weld area. This step could effectively reduce the interference of irrelevant areas and provide more accurate input for subsequent defect detection and segmentation.
(2) Data enhancement: Due to the small number of samples in the public data set, in order to enhance the diversity of data and improve the generalization ability of the model, the data set was expanded by the data enhancement technology designed in this paper, and the training effect of the model was improved without increasing the manual labeling data.
(3) Segmentation and detection of weld defects: In the case of varied types of weld defects, the segmentation accuracy of different types of defects was effectively improved by adopting an improved deep learning model.

Gaussian function based on welding seam extraction algorithm for two-end cropping mapping

The data set in this paper was selected from the public data set GD X-ray[26]. In the original data set of weld defects, the image resolution was high and the content information was rich, but the weld and its internal defects occupied a small area. Directly using a deep learning network for training might lead to problems such as low recognition accuracy and excessive memory occupation. Therefore, the weld should be extracted first. The complete data set of welding defects is shown in (Fig. 2), including welding areas, non-welding areas and various

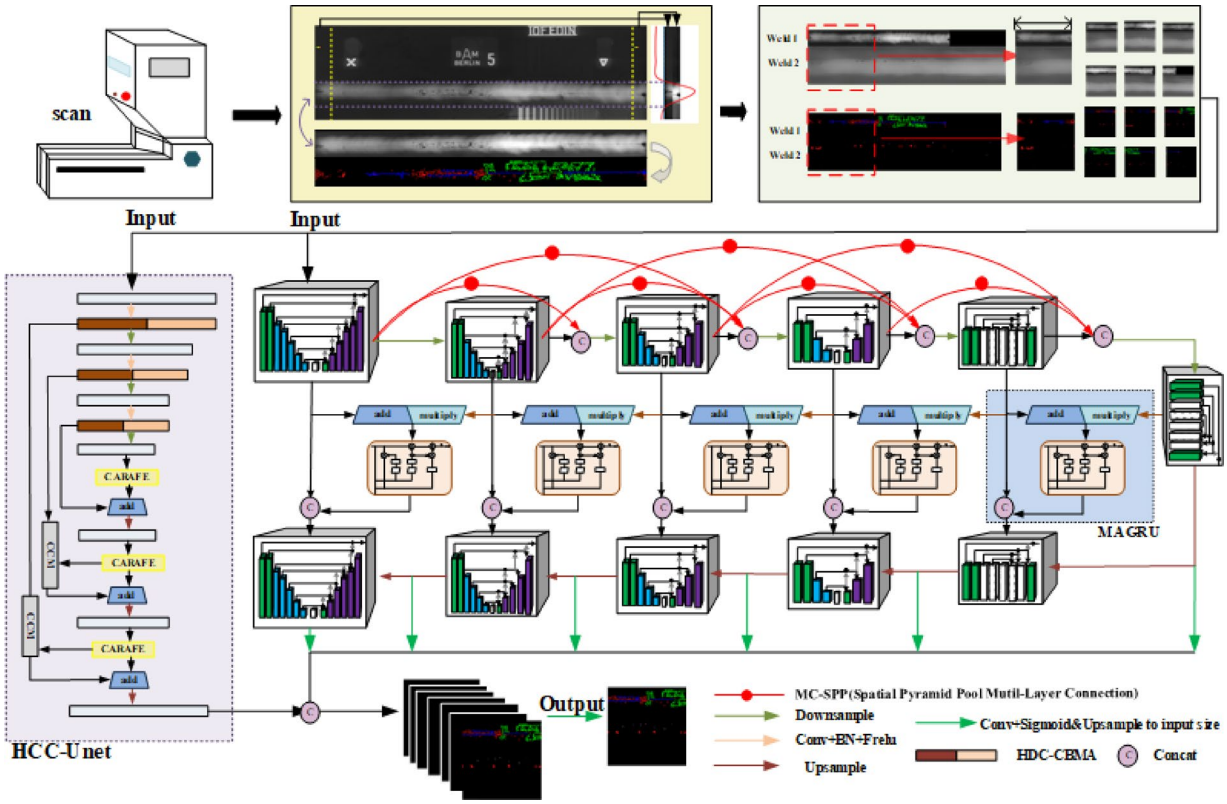


Fig. 1. Overall design structure (original artwork by the author).

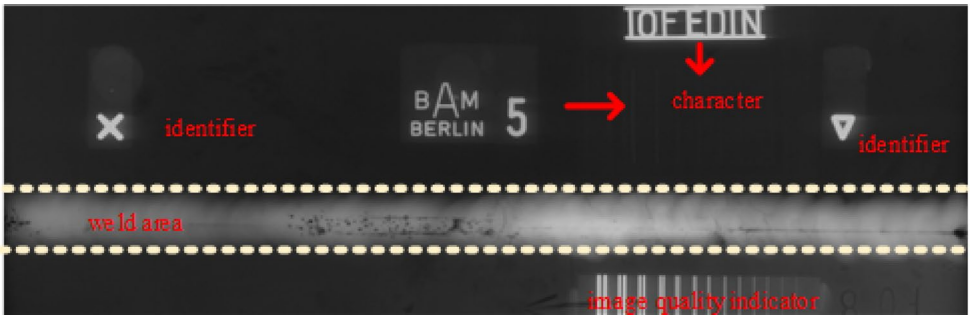


Fig. 2. Demonstration of the original weld defect data set.

identifiers. The existence of these identifiers interferes with the extraction of weld, and the position of weld in the image was not fixed[27], so it was difficult to extract it directly. However, through observation, it was found that the left and right end regions of the image had no identifiers, and these regions included the head and tail of the weld. Therefore, in this paper, the starting and ending positions of the weld were extracted by cutting these areas respectively, and this position information was mapped back to the original image, thus completing the extraction of the weld[28].

Because the left and right sides of some original images were very close to the identifier area, through observation and statistics, this paper finally determined 2.7% of the length of the original image as the cropping length, which was marked as w, and the specific operation is shown in (Fig. 3). Firstly, the regions with length w were cut out from the left and right sides of the image, and the two parts were spliced to obtain a vertical bar graph P1 with the same height as the original image, length 2w and no identifier. Then, the image was averaged to improve the contrast and brightness difference in the image. Because the welding area was generally distributed horizontally in the negative image, the longitudinal characteristics of the image were preserved during the processing, and only the lateral information was adjusted. The specific operation formula is as follows:

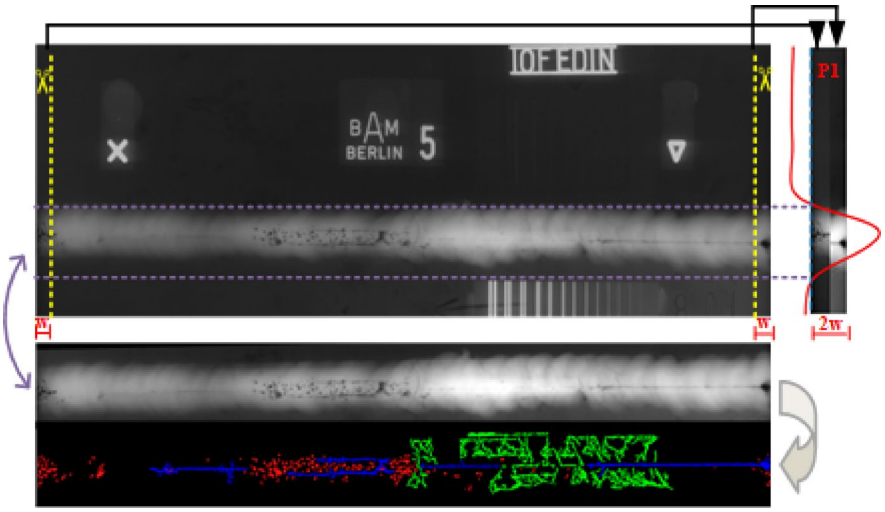


Fig. 3. The operation schematic diagram.

$$y_{i_f} = \frac{y_{i_m} \cdot y_\sigma}{y_{i_m} - \frac{1}{N}\sum_{i=1}^{N} y_{i_m}} \tag{1}$$

Among them:

$$y_\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{i_m} - \frac{1}{N}\sum_{i=1}^{N} y_{i_m}\right)^2} \tag{2}$$

$$y_{i_m} = y_i - y_m \tag{3}$$

$$y_m = \frac{1}{N}\sum_{i=1}^{N} y_i,\ \mathrm{y}_i = \sum_{i=1}^{N} x_i \tag{4}$$

Where, x_i is the pixel value of each line in the image, N is the number of vertical pixel points of the image, and y_{i_f} is the final processed result value.

After averaging the image, use Gaussian function to make statistics on the obtained results[29]. The expression of Gaussian function is as follows:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} y_{i_f} \tag{5}$$

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_{i_f} - \mu\right)^2} \tag{6}$$

$$G(y_{i_f}) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_{i_f}-\mu)^2}{2\sigma^2}} \tag{7}$$

Where $G(y_{i_f})$ is the final output Gaussian function curve.

The peak of the calculated Gaussian function curve corresponded to the weld area, thus determining the upper and lower limits of the weld position in the image P1. Then, these two boundary values were mapped back to the original image to complete the extraction and cutting of weld area.

A data enhancement method based on hybrid splicing of multiple images

Because the number of films in the open data set of weld defects was small, it couldn't meet the needs of deep learning network training[30], so this paper adopted the method of data enhancement to expand the data set. The aspect ratio of the extracted weld image was about 10:1, but in deep learning network training, the aspect ratio of the image was usually required to be 1:1[31]. Therefore, it was necessary to cut the welding film. In order to achieve this, this paper chose the common sliding cutting method. However, not all the defects in weld images were evenly distributed, and simply using a sliding window to cut would lead to many images without defects, which had no

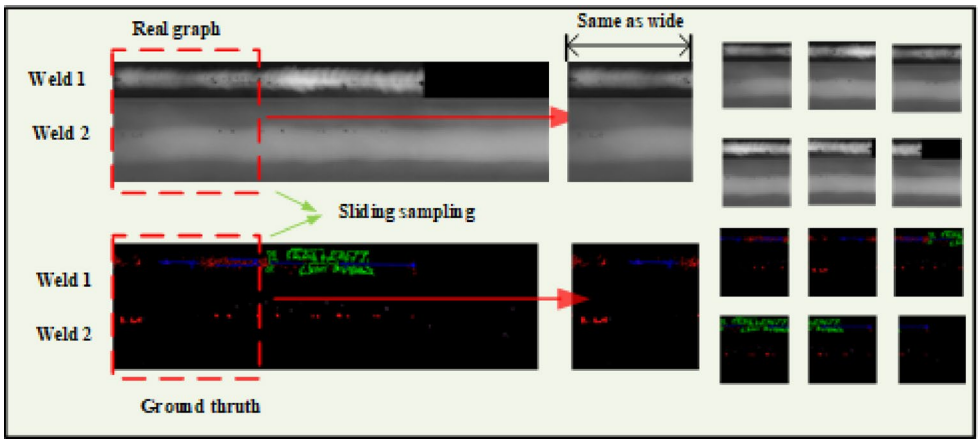


Fig. 4. Demonstration of data set creation.

Input: image1 and image 2 **Output: cropped image**
1: size_height = max[image1. height, image2. height]
2: size_wide = image1.wide + image2.wide //Use the function to create a new image with size_height * size_wide and an initial color of black
3: new_img = creating Images(size_height, size_wide, 'black') //Paste image1 and image2 onto the new image
4: new_img = paste image (image1, image2)
5: height, width = new_img.shape
6: stepSize = int(0.5 * height)
7: windowSize = height
8: **for** y<- 0 to height
9: **for** x<- 0 to width
10: **if** (y + windowSize) < = height and (x + windowSize) < = width:
11: cropped image= [y: y + windowSize, x:x + windowSize]
12: **save cropped image**
13: **if** x + windowSize > width:
14: **end if**
15: x<- x + windowSize
16: **if** (y + windowSize) > height:
17: **end if**
18: y<- y + windowSize
19: **end function**

Table 1. Clipping pseudo-code display.

practical significance for training. In the acquisition stage, all the images have been excluded from the problems such as poor brightness and overexposure. Therefore, the data before training has not undergone further image normalization, contrast adjustment or filtering. We believe that this data processing method effectively ensures the quality of training data and reduces unnecessary operational interference. In the acquisition stage, all the images have been excluded from the problems such as poor brightness and overexposure. Therefore, the data before training has not undergone further image normalization, contrast adjustment or filtering. We believe that this data processing method effectively ensures the quality of training data and reduces unnecessary operational interference.

Through observation, it was found that the content and quantity of defects in each welding film were quite different, and the proportion of different types of defects was also significantly unbalanced. In order to solve this problem, this paper proposed a hybrid mosaic method based on multiple images. Specifically, firstly, the welding films with different defect types were spliced up and down to generate a composite image containing multiple defect types. Then, these spliced images were clipped by sliding, so as to obtain data images for training. The specific steps of data set making were shown in (Fig. 4), and the pseudo-code of cutting is shown in (Table 1).

In this experiment, we selected two weld images for stitching up and down, and constructed the WSCR data set used in the experiment through weld extraction and data expansion (finally scaled to 1500 × 1500). Through

statistical analysis, the distribution ratio of various defects in the original data set and the newly constructed WSCR data set is shown in (Fig. 5).

Design of a high-resolution weld semantic defect detection network based on an integrated double-U structure

This chapter mainly introduced the network model designed in this paper, which aimed to realize the task of small target defect segmentation and detection in high-resolution weld images. The network drew lessons from the classic U2NetP and UNet structures, and improved on U2NetP, retaining all RSU(Recurrent Residual U-Net) modules in the original network. Specifically, this paper referred to the dense connection operation, and combined it with the multi-scale spatial pyramid to design the MC-SPP module. This module not only effectively retained more semantic information, but also significantly expanded the scope of the receptive field.

In addition, in the jump connection part, the RMAG module was introduced. The module adopted two different image fusion methods, supplemented by residual structure connections, which further enriched the semantic information of the intermediate feature map and effectively avoided the problem of gradient disappearance. Finally, combined with the GRU module, the segmentation results were more flexible and accurate by using time information and spatial structure information in time series.

On the basis of the classic UNet network, this paper also designed HDC-CBAM module by combining inflation convolution and attention mechanism, and introduced classic modules such as AFF and CARAFE. By cascading the above-mentioned improved networks and combining the multi-angle attention of different networks to the segmentation target, the network in this paper could more accurately complete the task of segmentation and detection of high-resolution weld defect images. The overall network structure is shown in (Fig. 6).

Multi-connection Spatial pyramid pooling module design (MC-SPP)

The original U2NetP network extracts features through step-by-step down-sampling. However, due to the high resolution of the images processed in this paper, direct down-sampling may result in significant loss of image information. This can negatively impact the final segmentation results, particularly leading to missed detections.

To address this issue, we drew inspiration from HRNet (high-resolution network)[32], which connects high-resolution and low-resolution feature maps. In this paper, we connect two adjacent RSU modules to supplement and refine image information over a certain range.

Furthermore, to further expand the receptive field and enhance the network's ability to capture global information, we introduce a three-layer pyramid structure during the dense connection process. The dimensions of the three branches of the pyramid structure are 32, 64, and 128, respectively. After each convolution layer, we add a Batch Normalization (BN) layer and a ReLU activation function to eliminate variance offset, ensuring the stability and consistency of the network's output. The design structure of the MC-SPP module is shown in (Fig. 7).

Residual multi-add gating recurrent unit module design (RMAG)

In the traditional U2NetP network, the skip connection between RSU modules directly linked the middle feature maps of the same level[33], which limited the capture of semantic information and made the weight updating process relatively fixed. This approach struggled to effectively capture more spatial information.

To address this, we proposed an improved RMAG module, consisting of three key parts:

(1) Inter-layer feature map fusion: This part combines the intermediate feature maps from the current and adjacent lower layer encoders. First, we perform pixel-wise addition to create a feature map (f1), then con-

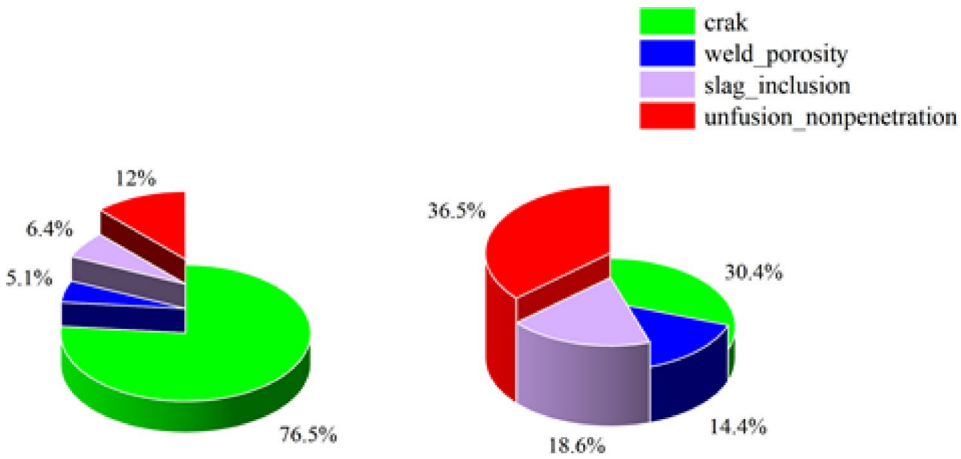


Fig. 5. Distribution of defects in different datasets (left panel is the original data set, right panel is the WSCR data set).

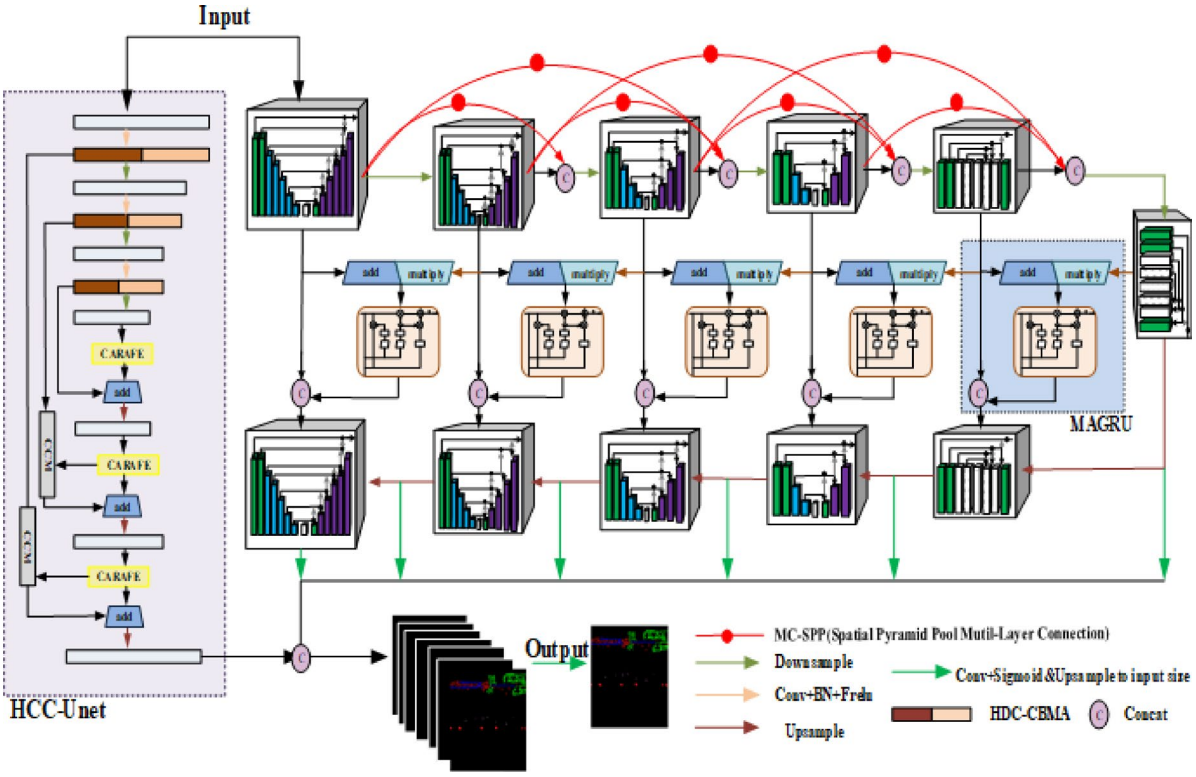


Fig. 6. Overall network structure (original artwork by the author).

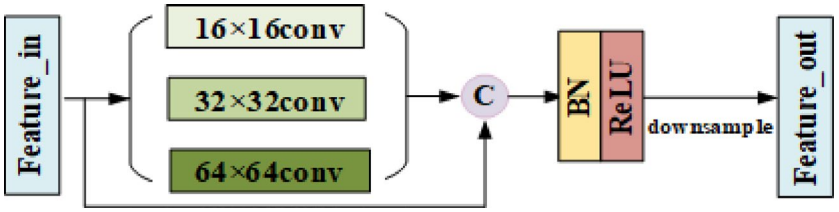


Fig. 7. MC-SPP module design.

catenate the two feature maps to obtain (f2). Finally, f1 and f2 are concatenated again to expand spatial dimensions, preserving contextual information and enhancing feature relationships.

(2) GRU module: A variant of LSTM (long-term memory network)[34], GRU improves training efficiency and reduces memory consumption. By controlling information flow, especially in long sequences, GRU mitigates gradient explosion or disappearance. It enhances the network's flexibility in handling context features, improving segmentation accuracy.

(3) Residual connection: This module combines the processed feature maps from the first two parts with the intermediate feature map from the original layer, preserving accuracy and improving gradient flow, thereby strengthening the network's expressive capability.

These improvements make the RMAG module highly effective in semantic segmentation, particularly for weld images, by better capturing and utilizing spatial and contextual information. The design structure of the RMAG module is shown in (Fig. 8).

Design of lightweight UNet network model based on composite modules (UNet_HCC)

Based on the classic UNet network, our module was lighter and performed better. Specifically, the original five-layer encoder and four-layer decoder architecture was reduced to four-layer encoder and three-layer decoder[35], thus reducing the depth of the network, and halving the number of output channels of the intermediate feature map, significantly reducing the computational complexity.

In this study, the original UNet architecture (Fig. 9a) was improved in many dimensions, and the improved network structure is shown in (Fig. 9b). In the part of the network encoder, we innovatively introduced the HDC-CBAM module (Fig. 9d), which significantly expanded the receptive field of the network and enhanced

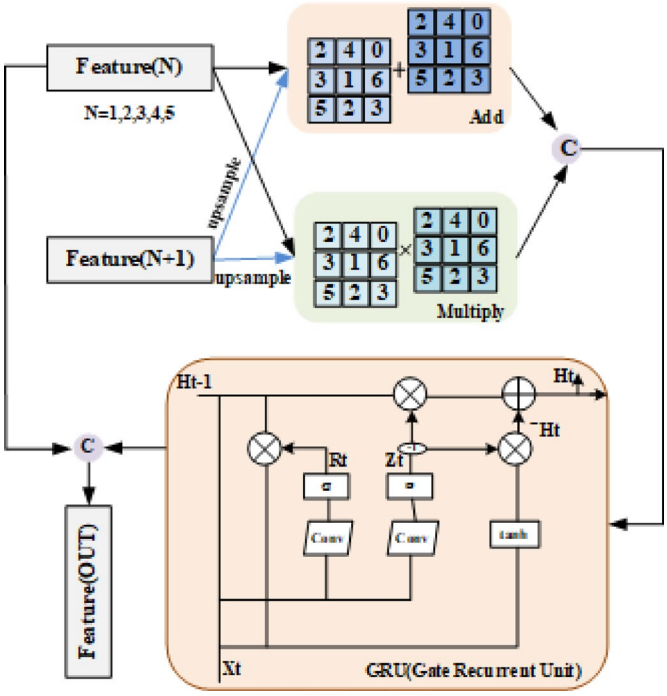


Fig. 8. RMAG module design.

the ability of context feature extraction by integrating extended convolution, residual connection and attention mechanism, thus improving the accuracy and richness of feature expression.

In the jump connection part, we designed the feature cross fusion module CCM (Fig. 9c). The module adopted cascade feature processing flow: firstly, the feature maps of adjacent layers were spliced, and then feature fusion was realized by 1×1 convolution, FReLU activation and 3×3 convolution in turn. In particular, the residual connection mechanism introduced in the module not only retained the detailed information, but also ensured the effective transmission of key features.

In the decoder part, in order to solve the problem of limited receptive field of traditional up-sampling operation, we adopted CARAFE (content-aware receptive field expansion)[36] up-sampling module with content awareness (Fig. 9e). The module realized more accurate feature reconstruction through content-related dynamic weight calculation, and significantly improved the resolution quality of the output image. Experiments showed that these optimizations not only improved the overall performance of the network, but also made it especially suitable for real-time processing scenarios and resource-constrained hardware environments through optimization of computational efficiency. The detailed architecture of each improved module is shown in (Fig. 9c–e), and the complete network structure is shown in (Fig. 9b).

Network integration

This study focused on high-resolution small target detection. In view of the problem that some networks with simple structures could not effectively complete the segmentation task or the segmentation accuracy was too low (less than 30%), two U-shaped networks with excellent segmentation effects were selected as the basic networks. Through experimental observation, we found that both the classic U2NetP, UNet networks and the improved network on this basis all focused on the small target segmentation in this study. Therefore, our team tried to cascade these two networks and achieved remarkable accuracy improvements. This attempt at network combination not only improved the segmentation accuracy, but also provided a new network optimization idea for our team. In the future, we will further study how to optimize the cascade mode and how to determine the appropriate number of cascaded networks to find the best balance between accuracy improvement and network scale.

Loss function

The loss function measures the gap between the true value and the predicted value and is an important part of the parameters of the network model[37]. In this paper, for the multi-classification task of weld defects, thus choosing the cross-entropy loss function suitable for multitask classification. The loss function expression is shown in the following equation, where N denoted the number of task categories in the classification task, n denoted the number of tensors in the data set, y_i denoted the true labeling value, and $f(x_i)$ denoted the predicted value of the segmentation result.

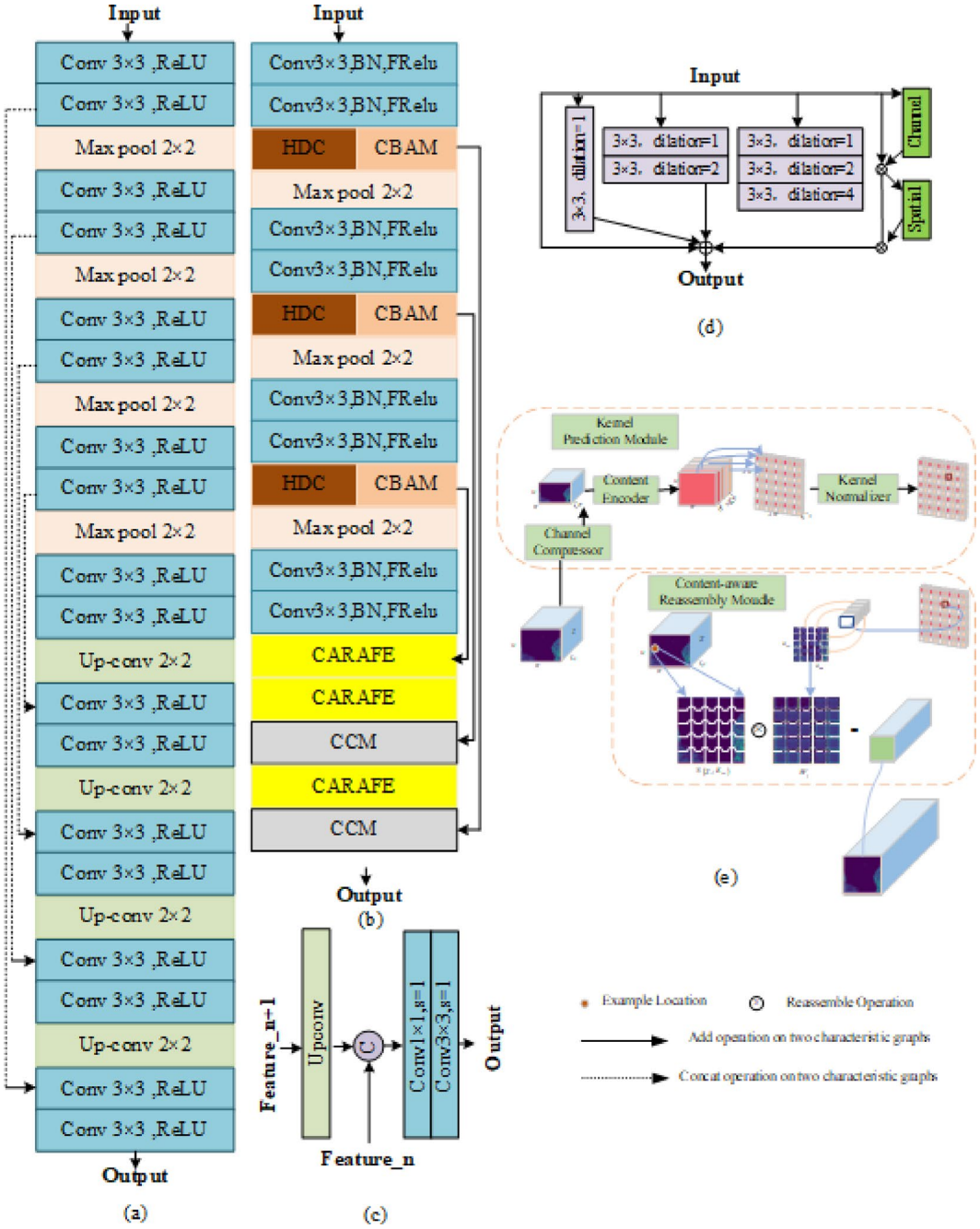


Fig. 9. UNet_HCC structure design diagram.

$$Loss = -\frac{1}{N}\sum_{i=1}^{n}[y_i \log f(x_i) + (1-y_i)\log(1-f(x_i))] \tag{8}$$

Experimental results and analysis

In this section, the test of the network structure designed in this paper was completed, and the feasibility and superiority of the data set designed in the previous section and the designed network structure were discussed.

Environmental parameter settings

The training platform configuration and network training parameter settings used in this paper are shown in (Table 2).

Name	Parameter
Operating system	Windows10
Graphics card	NVIDIA Quadro M6000
CUDA	12.1
CUDNN	8.0.5
Deep learning framework	Pytorch 2.2.0
Python	3.8.8
Opencv	4.9.0
Traindataset	1518
Testdataset	169
Epoch	100
Batch_size	4
Learning rate	1e-4
Num_workers	2

Table 2. Parameter configuration.

Evaluation indicators

In this paper, IOU, Recall, Precision, and model size were used as the evaluation indexes for this semantic segmentation of weld defects. The higher IOU indicated that the result predicted by the model is closer to the real value, which could be used to judge the model[38]; Recall was suitable for evaluating whether or not to correctly recognize a certain class of instances, and the focus was on how many pixels were correctly recognized by the model in this kind of recognition. The higher the value, the less miss detection[39]; Precision referred to the ratio of correctly predicted pixels to all pixels, the higher the value, the fever false detections[40]. In this paper, in addition to the comparison of the various indexes of the network to predict the image, our team also compared the size of each network model, and whether the network was light enough was also used as one of the criteria for judging the network as a whole. (mIOU, mRecall, mPrecision referred to the corresponding above performance indicators after taking the average of all categories after statistics.)

$$IOU = \frac{TP}{FN + FP + TP} \tag{9}$$

$$\mathrm{Re}\,call = \frac{TP}{TP + FN} \tag{10}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{11}$$

Where TP stands for the portion of the predicted value that overlaps with the true value, FP is the portion of the predicted value that does not overlap with the true value, TN the portion of the predicted value other than the predicted value that overlaps with the true value other than the predicted value, and FN refers to the portion of the predicted value that overlaps with the true value other than the predicted value[41].

Performance comparison

Qualitative inorganic analysis

In order to verify the superiority of the network designed in this paper, we compared the training results with classic traditional networks (such as UNet, U2NetP, DeepLabV3+, etc.) and the network structure designed by scholars such as Liangliang Li, Sicheng Liu and Qinghe Ma[42] in recent years. The data set used in the experiment was WSCR. After training and testing, we visualized the test results, and the specific results are shown in the figure.

On the whole, UNet, DeepLabV3 + and UNext networks had obvious shortcomings in detail processing, resulting in false detection and missed detection, and had different performances in different types of defect detection. First of all, from Fig. 10a, it could be found that the network designed by Liangliang Li and Qinghe Ma had a big deviation in the overall shape control, while U2Net and U2NetP were stuck in the details processing, which was not as good as the network designed in this paper. In Fig. 10b, the network designed in this paper showed the detection effect closer to the real shape when dealing with densely distributed defects, and had stronger consistency, showing its advantages in shape reduction. Figure 10c further showed that the network designed in this paper not only had higher recognition integrity, but also effectively reduced redundant adhesion parts and improved the accuracy of defect detection, compared with the excellent networks such as U2NetP, Liangliang Li and Sicheng Liu. Figure 10d showed that most networks perform well in dealing with this defect, but for a large number of defects with dense arrangements, the network designed in this paper could not only accurately describe the details of larger defects, but also maintain high accuracy in the detection of smaller defects.

To sum up, by comparing the details of different defect detection, the network designed in this paper had almost no false detection, and could identify subtle defects, greatly reducing the phenomenon of missing

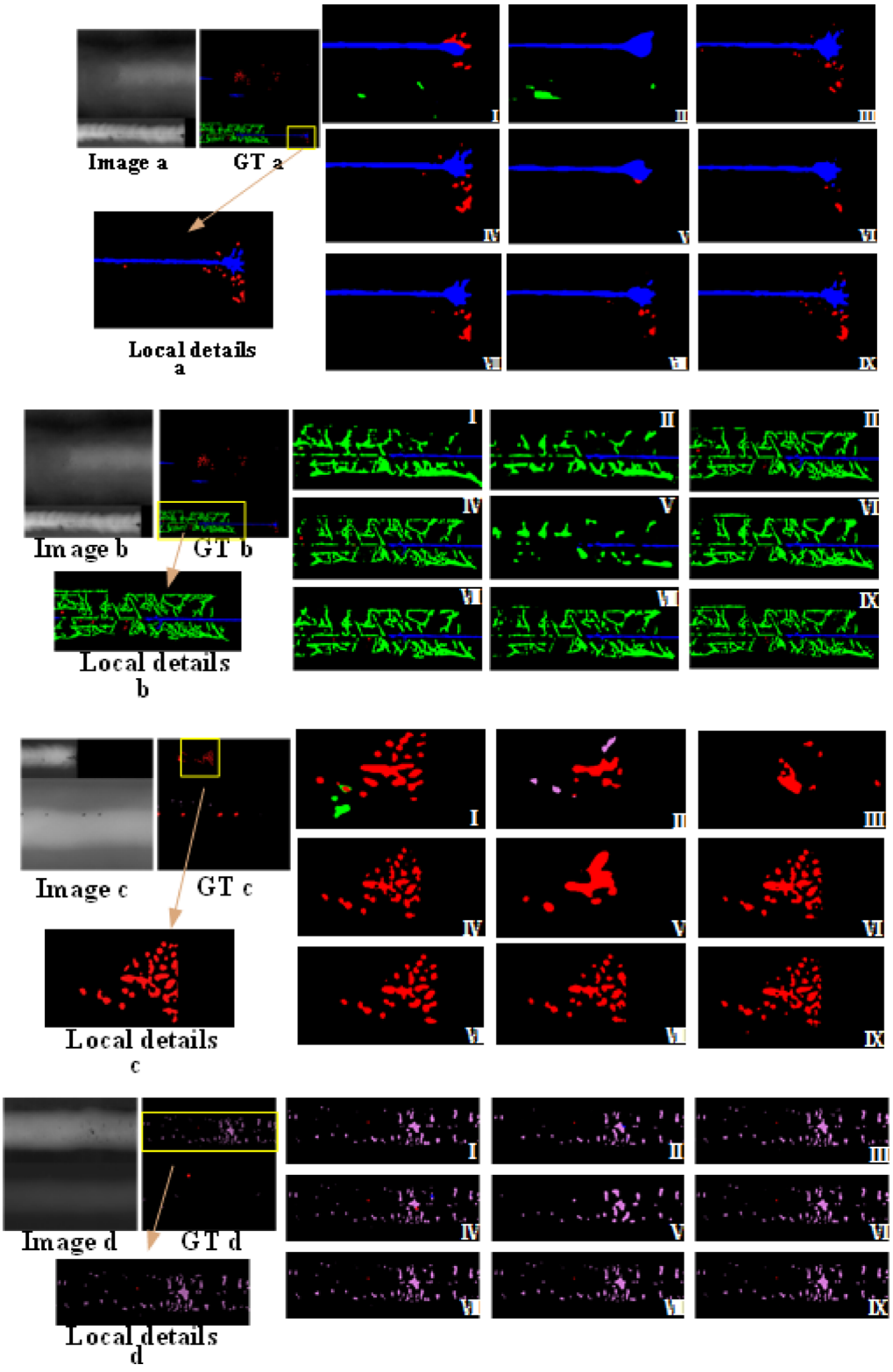


Fig. 10. Comparative visualization of different networks (I. UNet, II. DeepLabV3+, III. U2Net, IV. U2NetP, V. UNext, VI. Liangliang Li's, VII. SichengLiu's, VIII. Qinghe Ma's, IX. OURS).

detection. In the aspect of detail processing, this network showed stronger advantages, which could more completely retain the morphological characteristics of the original defects, especially in the changeable and fine defect processing.

Quantitative analysis

According to the performance of different models provided in the Table 3; Fig. 11 on WSCR data sets, the network designed in this paper performs well in many evaluation indexes, especially in mIOU index, which was obviously superior to other network models. This showed that the network in this paper had strong comprehensive performance in semantic segmentation tasks, and could achieve a good balance between accuracy and recall. Specifically, the network in this paper was also outstanding in the indicators of mRecall(83.22) and mPrecision(91.61), which were close to the optimal model Sicheng Liu's network[21] (84.47 and 89.45) respectively, with a small gap of less than 1%. This showed that although the network designed in this paper is slightly inferior to the most advanced model in some evaluation indexes, its overall performance was still competitive. In addition, in terms of model size, the parameters of this network were only 1/12 of the traditional U2Net network (168 MB), only 13.4 MB, and far smaller than most other models (such as 209 MB of DeepLabv3 + and 168 MB of U2Net). This showed that the network in this paper had successfully achieved lightweight while ensuring high performance, and had obvious advantages, especially suitable for application scenarios that require low computing resources and storage space.

To sum up, the network designed in this paper had achieved lightweight through optimization design on the basis of ensuring high accuracy, which embodied excellent comprehensive performance and has strong application potential.

Ablation experiments

Data set ablation

In this paper, an algorithm of seam extraction based on the Gaussian function was introduced. In order to verify the feasibility and effectiveness of the algorithm, this paper conducted experimental training and testing on the data set of non-extracted welds and the data set of extracted welds in the traditional U2NetP network and the network designed in this paper. Statistics of experimental results are presented in (Table 4; Fig. 12). It could be clearly observed that after the weld seam extraction was completed, the evaluation indexed of the two networks had been significantly improved. These real data results fully proved the necessity and effectiveness of the weld extraction method proposed in this paper.

In Chap. 2, this paper proposed a data enhancement method based on multi-image mosaic. In this method, a new data set WSCR was generated by stitching different numbers of weld images up and down and cutting them. In order to verify the detection effect of data sets under different weld numbers, this paper made a single

Name	Year	mIOU	mRecall	mPrecision	Size(MB)
UNet	2015	66.18	74.68	82.62	150.92
DeepLabV3+	2018	55.32	59.69	83.37	209
U2Net	2020	**77.61**	**82.73**	**91.75**	168
U2Netp	2020	66.35	71.82	88.67	**4.66**
UNext	2022	51.00	54.77	83.28	**5.65**
Liangliang Li's	2023	71.66	77.39	89.31	**3.28**
SichengLiu's	2024	**77.60**	**84.47**	89.45	94.9
Qinghe Ma's	2024	72.10	76.15	**92.00**	12.2
OURS	——	**78.01**	**83.22**	**91.61**	13.4

Table 3. Performance of different models on data set WSCR data set (red first, blue second, green third).

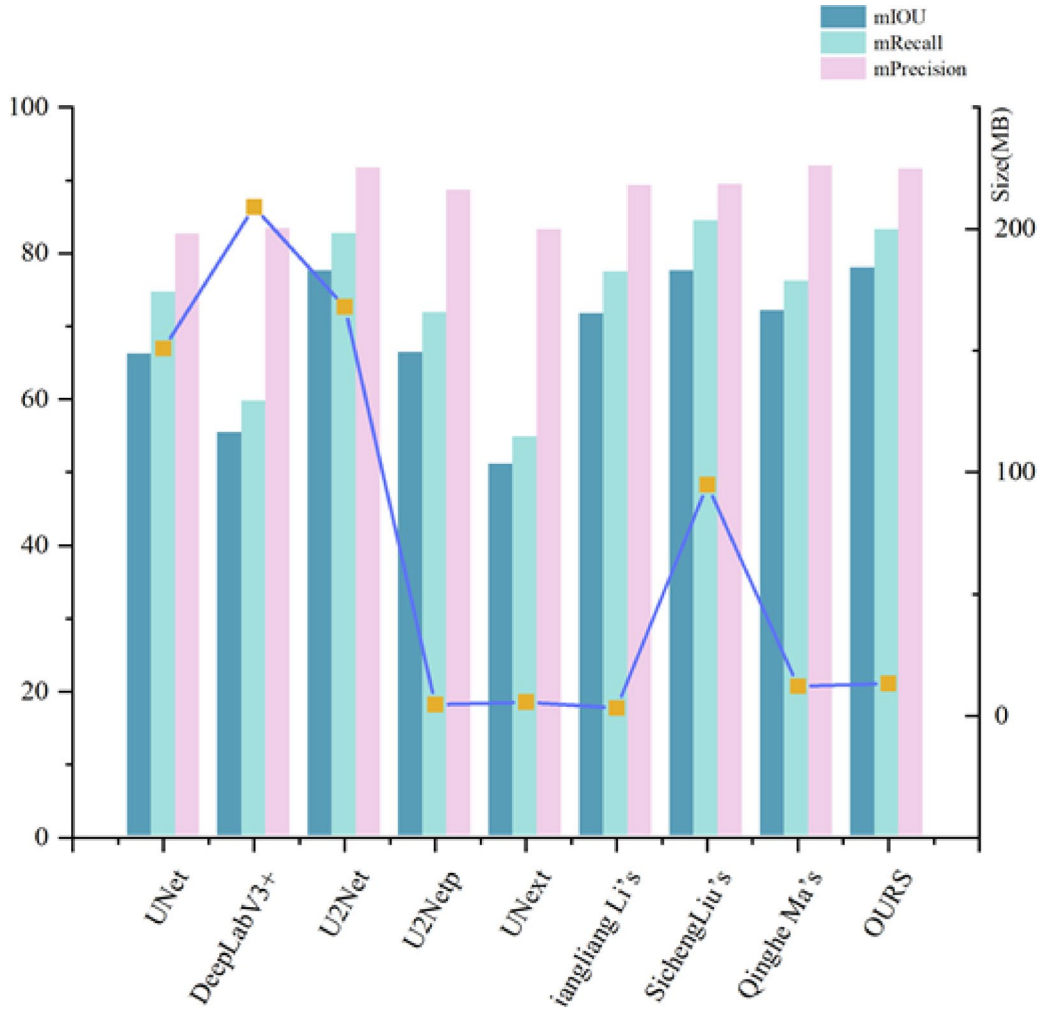


Fig. 11. Presentation of data from qualitative analysis of different networks.

Net	datasets	mIOU	mRecall	mPrecision
U2NetP	HRWDS	46.2	49.36	78.93
U2NetP	WDS	55.64	60	83.54
U2NetP	WSCR	66.35	71.82	88.67
OURS	HRWDS	55.78	59.79	86.41
OURS	WDS	64.51	69.12	88.28
OURS	WSCR	78.01	83.22	91.61

Table 4. Impact of different datasets on network performance.

weld cutting data set, two weld splicing cutting data sets and three weld splicing cutting data sets respectively, and tested them in the network. The experimental results showed that in the data set of two-seam splicing and cutting, each precision index reached the highest value, showing a trend similar to the inflection point. Therefore, this paper finally chose the cutting method of double weld stitching to build the data set.

Network structure ablation

In order to further verify the effectiveness of the network designed in this paper, firstly, the newly designed modules in this paper were added to the basic U2NetP network in turn, and trained and tested on WSCR data sets, and the evaluation indexes were counted. In addition, in order to prove the superiority of the improved UNet_HCC network in network cascade, the cascade experiment of the improved U2NetP network and UNet network was also carried out. Through this experiment, this paper provided strong data support, which proved the feasibility of UNet_HCC network and the superiority of cascade network structure. Specific manifestations are shown in (Table 5; Fig. 13).

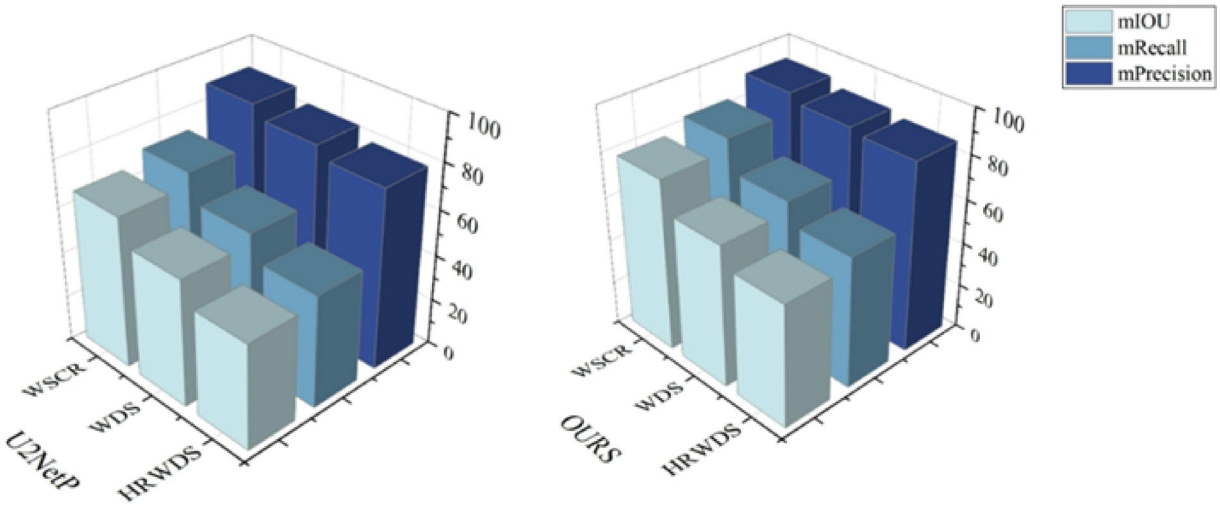


Fig. 12. Statistical graph of the impact of different datasets on network data performance.

Net	Datasets	mIOU	mRecall	mPrecision	Size
U2NetP	WSCR	66.35	71.82	88.67	4.66
U2NetP + MCSPP	WSCR	71.94	77	90.64	5.52
U2NetP + MCSPP + MAGRU	WSCR	76.08	81.57	90.89	7.5
U2NetP + MCSPP + MAGRU + UNet	WSCR	77.12	83.02	90.52	157
OURS	WSCR	78.01	83.22	91.61	13.4

Table 5. Performance after adding different modules.

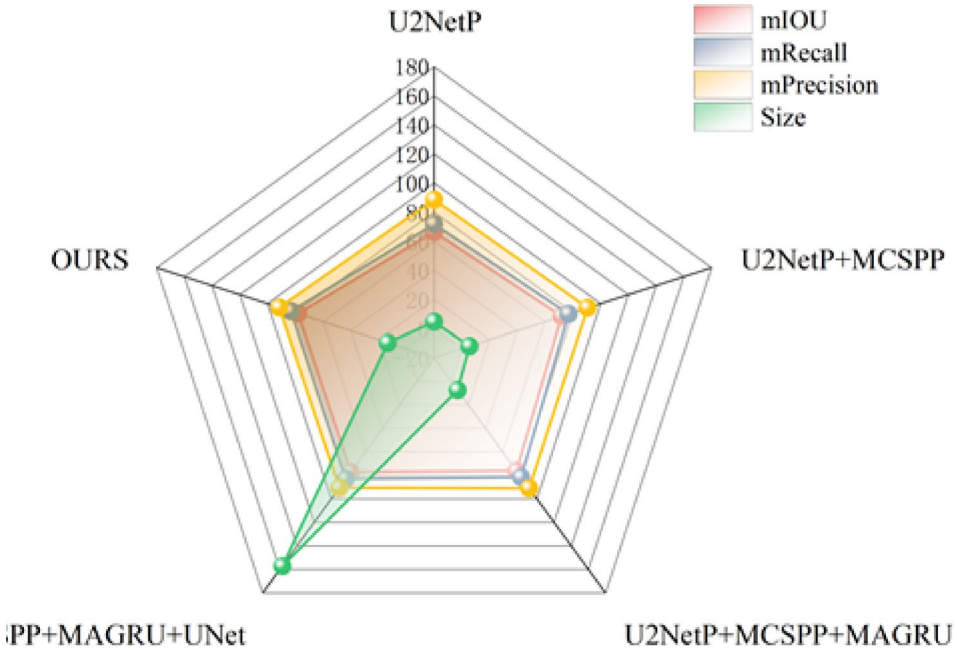


Fig. 13. Performance graph after adding different modules.

Conclusion

Detection of small defects in high-resolution weld films was a key problem in intelligent negative analysis. Aiming at the problems of invalid data training and uneven distribution of defects in the existing sliding window method, an innovative algorithm framework of "locating first and then detecting" was proposed in this study. Firstly, accurate positioning and effective area extraction were realized based on Gaussian function and longitudinal characteristics of weld, and then a balanced defect data set was constructed by multi-image stitching technology to optimize data distribution while maintaining resolution. In the aspect of network architecture, by combining the double U-shaped structure of MC-SPP, RMAG, CCM and CARAFE, the receptive field was expanded and the detection accuracy was improved on the basis of U2NetP and UNet, while maintaining the lightweight characteristics of the model. Experiments showed that the algorithm provided an effective solution for high-resolution film defect detection.

The overall effect of the algorithm proposed in this study was good, but when the slag inclusion and cracks were particularly dense, because the resolution of the input film was too large, it could only be roughly segmented, and the end shape could not be completely and finely obtained. Our team will continue to optimize the detection model to achieve higher segmentation accuracy. In addition, it is essential to explore better data augmentation strategies and develop more intelligent data synthesis methods to address extreme sample imbalances.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 3 March 2025; Accepted: 13 May 2025
Published online: 22 May 2025

References

1. Li, L. et al. Synthetic data augmentation for high-resolution X-ray welding defect detection and classification based on a small number of real samples. *Eng. Appl. Artif. Intell.* **133**, 108379 (2024).
2. Liu, M. et al. A lightweight and accurate recognition framework for signs of X-ray weld images. *Comput. Ind.* **135**, 103559 (2022).
3. Zou, Y. & Zeng, G. Light-weight segmentation network based on SOLOv2 for weld seam feature extraction. *Measurement* **208**, 112492 (2023).
4. Abdelkader, R. et al. Segmentation of x-ray image for welding defects detection using an improved Chan-Vese model. *Mater. Today: Proc.* **42**, 2963–2967 (2021).
5. Yang, W. et al. Generalized weld bead region of interest localization and improved faster R-CNN for weld defect recognition. *Measurement* **222**, 113619 (2023).
6. Dong, X., Taylor, C. J. & Cootes, T. F. A random forest-based automatic inspection system for aerospace welds in X-ray images. *IEEE Trans. Autom. Sci. Eng.* **18** (4), 2128–2141 (2020).
7. He, Z. et al. An image segmentation-based localization method for detecting weld seams. *Adv. Eng. Softw.* **194**, 103662 (2024).
8. Hongyu, G. Yongyi, M. Richong, Z. MixUp as locally linear out-of-manifold regularization. *Proceedings of the AAAI Conference on Artificial Intelligence* (2019).
9. Sonwane, S. & Chiddarwar, S. *Enhancing Weld Defect Detection and Classification with MDCBNet: A multi-scale Dense Cross Block Network for Improved Explainability* 142103029 (NDT & E International, 2024).
10. Zhang, Q. et al. Image defect classification of surface Mount technology welding based on the improved ResNet model. *J. Eng. Res.* (2024).
11. Xu, H. et al. Defect detection in welding radiographic images based on semantic segmentation methods. *Measurement* **188**, 110569 (2022).
12. Wang, X. et al. High-resolution remote sensing images semantic segmentation using improved UNet and SegNet. *Comput. Electr. Eng.* **108**, 108734 (2023).
13. Qian, W. et al. *Image Enhancement Method for low-light Pipeline Weld X-ray Radiographs Based on Weakly Supervised Deep Learning* 143103049 (NDT & E International, 2024).
14. Yang, L. et al. A nondestructive automatic defect detection method with pixelwise segmentation. *Knowl. Based Syst.* **242**, 108338 (2022).
15. Li, L. et al. A pixel-level weak supervision segmentation method for typical defect images in X-ray inspection of solid rocket motors combustion chamber. *Measurement* **211**, 112497 (2023).
16. Zhang, R. et al. Research on X-ray weld seam defect detection and size measurement method based on neural network self-optimization. *Eng. Appl. Artif. Intell.* **133**, 108045 (2024).
17. Kwon, J. E. et al. *Context and scale-aware YOLO for Welding Defect Detection* 139102919 (NDT & E International, 2023).
18. Ji, C., Wang, H. & Li, H. *Defects Detection in Weld Joints Based on Visual Attention and Deep Learning* 133102764 (NDT & E International, 2023).
19. Aminzadeh, A. et al. A real-time 3D scanning of aluminum 5052-H32 laser welded blanks; geometrical and welding characterization. *Mater. Lett.* **296**, 129883 (2021).
20. Aminzadeh, A. et al. Toward an intelligent aluminum laser welded blanks (ALWBs) factory based on industry 4.0; a critical review and novel smart model. *Opt. Laser Technol.* **167**, 109661 (2023).
21. Liu, S. et al. High-throughput measurement method for rice seedling based on improved UNet model. *Comput. Electron. Agric.* **219**, 108770 (2024).
22. Li, M. et al. PipeTransUNet: CNN and transformer fusion network for semantic segmentation and severity quantification of multiple sewer pipe defects. *Appl. Soft Comput.* **159**, 111673 (2024).
23. Liu, T., Zheng, P. & Liu, X. *A Multiple Scale Spaces Empowered Approach for Welding Radiographic Image Defect Segmentation* 139102934 (NDT & E International, 2023).
24. Liu, Q. et al. Fill-UNet: extended composite semantic segmentation. *Appl. Soft Comput.* **172**, 112891 (2025).
25. Tan, J. et al. DarkSegNet: Low-light semantic segmentation network based on image pyramid. *Sig. Process. Image Commun.* 117265. (2025).
26. Mery, D. et al. GDXray: the database of X-ray images for nondestructive testing. *J. Nondestruct. Eval.* **34** (4), 1–12 (2015).
27. Jang, S. et al. Machine learning-based weld porosity detection using frequency analysis of Arc sound in the pulsed gas tungsten Arc welding process. *J. Adv. Join. Processes*. **10**, 100231 (2024).
28. Li, L. et al. An adaptive false-color enhancement algorithm for super-8-bit high grayscale X-ray defect image of solid rocket engine shell. *Mech. Syst. Signal Process.* **179**, 109398 (2022).

29. Rajeswari, S. et al. Detection and classification of various types of leukemia using image processing, transfer learning and ensemble averaging techniques. *2022 2nd Asian Conference on Innovation in Technology (ASIANCON)* 1–6 (IEEE, 2022).
30. Nabiee, S. et al. Hybrid U-Net: semantic segmentation of high-resolution satellite images to detect war destruction. *Mach. Learn. Appl.* **9**, 100381 (2022).
31. Wang, Z. et al. Semantic segmentation of high-resolution remote sensing images based on a class feature attention mechanism fused with Deeplabv3+. *Comput. Geosci.* **158**, 104969 (2022).
32. Sun, K. et al. Deep high-resolution representation learning for human pose estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 5693–5703 (2019).
33. Li, X. et al. You only train twice: A lighter and faster method for industrial weld defect detection based on dynamic kernel network. *Measurement* **231**, 114642 (2024).
34. Cheng, C. & Xu, H. A 3D motion image recognition model based on 3D CNN-GRU model and attention mechanism. *Image Vis. Comput.* **146**, 104991 (2024).
35. Zuo, Y. & Li, W. An improved UNet lightweight network for semantic segmentation of weed images in corn fields. *Comput. Mater. Contin.* **79** (3). (2024).
36. Zeng, W. & He, M. Rice disease segmentation method based on CBAM-CARAFE-DeepLabv3+. *Crop Prot.* **180**, 106665 (2024).
37. Yeung, M. et al. Unified focal loss: generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. *Comput. Med. Imaging Graph.* **95**, 102026 (2022).
38. Shi, X. & Zhang, Z. MAST-UNet: more adaptive semantic texture for segmenting pulmonary nodules. *Biomed. Signal Process. Control.* **99**, 106804 (2025).
39. Sarkar, M., Mandal, A. & Tudu, A. DC-UNet: looking for follicles in the ovarian ultrasound images. *Frankl. Open.* **8**, 100149 (2024).
40. Ma, C., Wang, Z. & Semi-Mamba-UNet Pixel-level contrastive and cross-supervised visual Mamba-based UNet for semi-supervised medical image segmentation. *Knowl. Based Syst.* **300**, 112203 (2024).
41. Meng, W., Liu, S., Wang, H. & AFC-Unet Attention-fused full-scale CNN-transformer Unet for medical image segmentation. *Biomed. Signal Process. Control.* **99**, 106839 (2025).
42. Ma, Q. et al. Constructing and exploring intermediate domains in mixed domain semi-supervised medical image segmentation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 11642–11651 (2024).

Acknowledgements

This work was supported by National Natural Science Foundation of China (62171360), Shaanxi Science Fund for Distinguished Young Scholars(2024JC-JCQN-57), The Qinchuangyuan Platform Science and Technology Achievement Transformation Project of Weiyang District in Xi'an (202416), 2022 Shaanxi University Youth Innovation Team Project, Engineering Technology Research Center of Shaanxi Province for Intelligent Testing and Reliability Evaluation of Electronic Equipments (2023-ZC-GCZX-0047), Xi'an Key Laboratory of Military Civilian Dual-Use Intelligent Evaluation Technology, Xi'an Automation Society Young Talent Support Program.

Author contributions

X. L.: Conceptualization, Methodology, Software, Writing - original draft. Y.W.: Visualization, Funding acquisition, Writing - review and editing. Z.L.: Funding acquisition. P. W.: Data collation and visualization. L.L.: Supervision. M. S.: Supervision. C. W.: Investigation.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.W., P.W. or M.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025