Arc bubble edge detection method based on deep transfer learning in underwater wet welding

Bo Guo

guobo651@126.com

Nanchang Institute of Technology

Xu Li

Nanchang Institute of Technology

Article

Keywords: Edge detection, deep transfer learning, arc bubble, underwater wet welding

Posted Date: April 17th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4227560/v1

Additional Declarations: No competing interests reported.

Arc bubble edge detection method based on deep transfer learning in underwater wet welding

Bo Guo [1,*], Xu Li [1]

[1]Nanchang Key Laboratory of Welding Robot & Intelligent Technology, Nanchang Institute of Technology, Nanchang 330099, China

Corresponding author: Bo Guo (e-mail: guobo651@126.com).

ABSTRACT Arc bubble in underwater wet welding reflects the stability of the welding process. An arc bubble edge detection method based on deep transfer learning is proposed to overcome the shortcomings of conventional algorithms in processing underwater wet welding images. The method consists of two training stages: pre-training and fine-tuning. In the pre-training stage, a large source domain dataset is used to train VGG16 as a feature extractor. In the fine-tuning stage, we proposed the Attention-Scale-Semantics (ASS) model, which consists of a Convolutional Block Attention Module (CBAM), a Scale Fusion Module (SCM) and a Semantic Fusion Module (SEM). The ASS model is retrained with the small underwater wet welding target domain dataset to fine-tune the model parameters. The CBAM can adaptively weight the feature maps, focusing on more important feature to better capture edge information. The SCM training method makes extensive use of feature information to simplify the training steps. Additionally, the skip structure of SEM effectively resolves the problem of semantic loss in the high-level network during the training process and improves the accuracy of edge detection. We compare the ASS model to the conventional edge detection model on the BSDS500 dataset and underwater wet welding images, demonstrating that the ASS model is superior to the conventional edge detection model. By comparing with Richer Convolutional Features (RCF), Fully Convolution Network (FCN) and UNet, the excellent performance of the ASS model in arc bubble edge detection method is verified.

INDEX TERMS Edge detection, deep transfer learning, arc bubble, underwater wet welding

I. INTRODUCTION

Image edge detection methods based on deep learning, such as HED [1] ， EGNet [2]， DeepNet [3], etc., are a research hotspot in the field of image edge detection. Wang et al. [4] proposed a rectangular convolution pyramid and edge enhancement network called RENet for accurate and robust pavement crack detection. Experiments conducted on two pavement crack datasets demonstrate that the proposed framework advances other state-of-the-art algorithms in terms of robustness and universality. Xiao et al. [5] adopted the principle of convolutional neural network (CNN) and a Laplacian filter to propose an edge detection model with improved performance. The proposed CNN model successfully detects fuzzy defects on noised X-ray images, and presents better structure similarity of the detected information. In the field of welding image processing, due to the influence of arc light, dust, and splash during the welding process, the welding images have issues such as uneven gray, blurred edges, etc., making it difficult to extract the edge features using conventional methods. However, the image edge detection methods based on deep learning have achieved good results [6]. Ma et al. [7] utilized a combination of Convolutional Neural Network (CNN) and laser vision to detect typical weld defects on galvanized steel sheets. Wang et al. [8] developed a deep learning algorithm for monitoring the width of the welding sedimentary layer. This control algorithm can monitor and control weld width online effectively. Cheng et al. [9] proposed a method for automatically identifying the position of the weld pool using vision technology based deep learning. Cheng et al [10]. proposed an active vision sensing method using a CNN deep learning model to estimate the penetration width.

Numerous public datasets exist for common visual tasks in natural images, such as pedestrian detection, license plate recognition, etc. However, due to the difficulty of acquiring high-quality welding images in an industrial setting and the

dearth of welding image researchers, there is currently no publicly available dataset in this field. In weld image processing based on deep learning, therefore, only limited sample data is typically available. However, a lack of data will result in overfitting when training a deep learning model, which is insufficient for CNN training [11]. Transfer learning is a popular research topic in machine learning. It is applied to the training of neural networks in an effort to solve the small sample problem of welding image datasets. Kumaresan et al. [12] developed a method based on CNN and transfer learning that enhanced the training performance of neural networks, shortened the training time, and enabled the classification of welding defects. Combining transfer learning and MobileNet, Pan et al. [13] proposed a new welding defect model. The accuracy of its predictions reached 97.69%. Jiao et al. [14] developed an algorithm based on Residual Neural Networks (ResNet) using transfer learning, which decreased the training time and improved the accuracy of weld penetration prediction. Wang et al. [15] used the MobileNetV2-based transfer learning model to fit the custom dataset and recognize weld penetration states. Bahador et al. [16] presents a novel application of transfer learning for tool wear detection in turning processes using one-dimensional (1D) convolutional neural network (CNN). The transfer learning model significantly increased the tool wear classification accuracy using the single-axis low-cost MEMS accelerometer from 58% to 85%. Moreover, the proposed model using transfer learning significantly reduced the amount of data required for model development.

In comparison to underwater dry welding and underwater local cavity welding, underwater wet welding has the advantages of simple equipment, low cost, and strong applicability [17]. It is commonly employed in underwater construction and maintenance fields, such as oil pipelines and offshore platforms with water depths of less than 60 meters. The stability of arc bubbles is a crucial factor affecting the underwater wet welding process and the quality of the joint. Tsai et al. [18] discovered the relationship between the heat loss of underwater wet welding pools and the movement of dynamic underwater bubbles. A bubble growth model is developed to predict its dynamic behavior. Yurioka et al. [19] proposed that the thermal cycles applied during welding would alter the mechanical properties of the base metal and that the heat-affected zone of the weld would always increase in hardness and decrease in toughness. Guo et al. [20] studied the underwater wet arc stability using the X-ray transmission method. It demonstrates that a larger bubble size can improve the welding process stability. Wang et al. [11] analyzed the effects of controllable arc bubbles on the stability of the welding process and the microstructural evolution of the joint. The results indicate that by controlling the arc bubble detachment, a more stable wet welding process can be achieved, thereby providing a new method for improving the underwater wet welding process. Wang et al. [21] analyzed the relationships between bubble geometries and welding heat input using a high-speed camera and a dysprosium lamp to study the growth and detachment of bubbles. It offers a novel approach to the advancement of wet welding technology. However, their experiment in an open tank did not replicate the deep-water environment in a closed pressure tank. In addition, the underwater welding images are only processed offline using a simple image processing method, and no online detection method of underwater bubble edge is proposed.

In recent years, numerous researchers have processed images containing bubbles. Fu et al. [22] proposed a Bubble Generative Adversarial Network (BubGAN), which significantly improved the accuracy of bubble image processing in comparison to the traditional GAN method. Zhao et al. [23] designed a bubble machine vision detection system for the transparent layer of quartz crucible, employing a morphological contour extraction algorithm to detect the bubbles, which has a higher detection efficiency. Saha et al. [24] developed an algorithm to detect the bubbles from pseudo-occlusion. The algorithm employed for bubble detection comprises Adaptive Threshold-based Image Segmentation which comprises segmentation, edge detection, segment grouping, and ellipse fitting. Yang et al. [25] proposed a vision detection technology to identify the sapphire bottom bubbles, which offers a technical solution for obtaining high-quality sapphire crystals. Misra et al. [26] investigated the detection method of bubble size and distribution in cake batter using machine vision technology, providing a benchmark for bubble detection in the food industry.

In conclusion, there have been limited studies that utilize deep learning image edge detection methods for underwater welding bubbles. This research successfully detects the bubble edges in the arc area of underwater wet welding using deep learning image edge detection. Furthermore, it accurately determines the size, position, and shape of the bubbles. This study holds great significance in understanding the arc mechanism involved in underwater wet welding, controlling the welding process, optimizing the process parameters, and ultimately improving the overall quality of underwater wet welding.

II. UNDERWATER WET WELDING IMAGE ACQUISITION SYSTEM

The underwater wet welding image acquisition system includes an underwater pressure simulation cabin, a welding system, and a visual system. Fig.1 shows the underwater pressure simulation cabin, which has design pressure of the 1.2MPa, allowing it to simulate the pressure at underwater depth of 100 meters. The cabin has sufficient space to accommodate the visual system and welding system, with a volume of 6.5 m^3. Welding power, wire feeder, and underwater self-protected flux-cored wire comprise the welding system. The welding power is a Panasonic YD-500KR with a maximum welding voltage of 60V and a maximum welding current of 500A. The visual system consists of a camera with a wide dynamic range, an actuator with three degrees freedom, and a quartz tank. The wide dynamic range camera is a NIT MC1003 industrial camera with

logarithmic dynamic range of up to 140 db. Since the camera is placed outside the water tank and images are captured through the water tank's inner wall, ordinary glass cannot meet the experimental requirements for transparency. The inner wall of the water tank is made of quartz glass due to its excellent optical conductivity over a broad wavelength range. The visual system is shown in Fig. 2.

Welding images captured by an underwater wet welding image acquisition system are shown in Fig.3. As illustrated in Fig. 3-a, the wire, arc, and bubble are clear. The underwater welding bubble is situated between the arc and the end of the welding wire, which serves as the paper's edge detection target.

FIGURE 1. Underwater wet welding experimental system

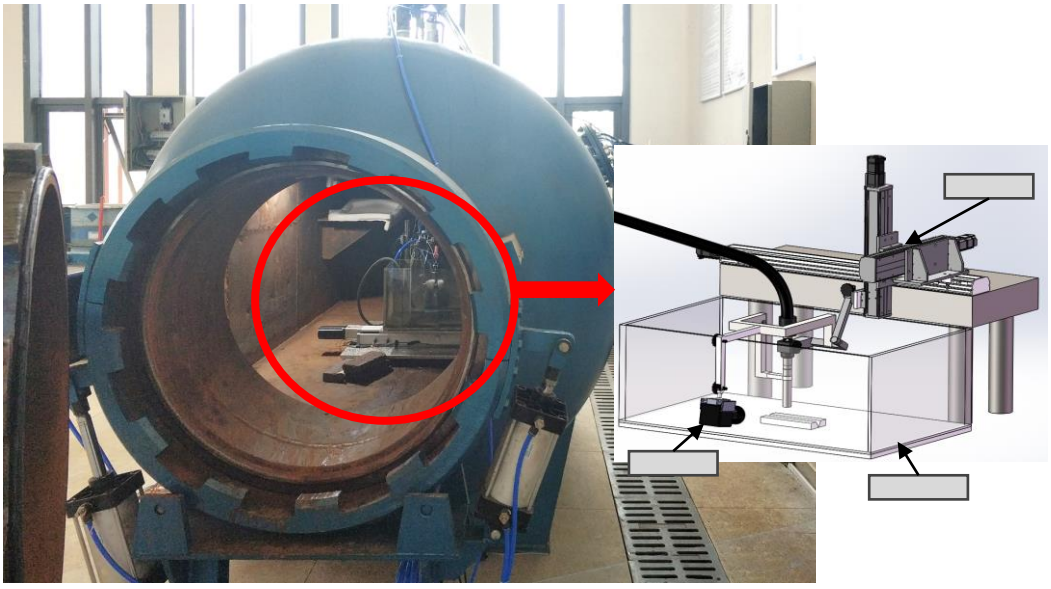

FIGURE 2. Visual system

FIGURE 3. Underwater wet welding images

III. BUBBLE EDGE DETECTION METHOD

Transfer learning is a new machine learning method that employs existing knowledge to solve problems in distinct but related fields [27]. According to the technical characteristics, transfer learning methods can divide into migration based on feature selection, instance-based migration, and migration based on model sharing parameters [28]. The transfer learning method based on model sharing parameters utilizes the correlation between source domain model parameters and target domain tasks. Consequently, the structure and parameter values in the source domain model are used as the initial conditions for the design of the target domain model, allowing the target domain model to train better generalization with a small sample size of target domain data.

The flowchart of the arc bubble edge detection method based on deep transfer learning is shown in Fig. 4. In this paper, VGG16 is used as the source domain model, and the ASS model is used as the target domain model. The large public dataset ImageNet is used to pre-train the source domain model, whose parameters are then migrated to the target domain model. After image pre-processing, the acquired underwater welding images are fed into a target domain model containing prior domain knowledge for training. Image pre-processing include edge enhancement (Linear enhancement) and image denoising (Bilateral filter). The preprocessed image is input into the model for prediction, and the final edge extraction image is obtained.

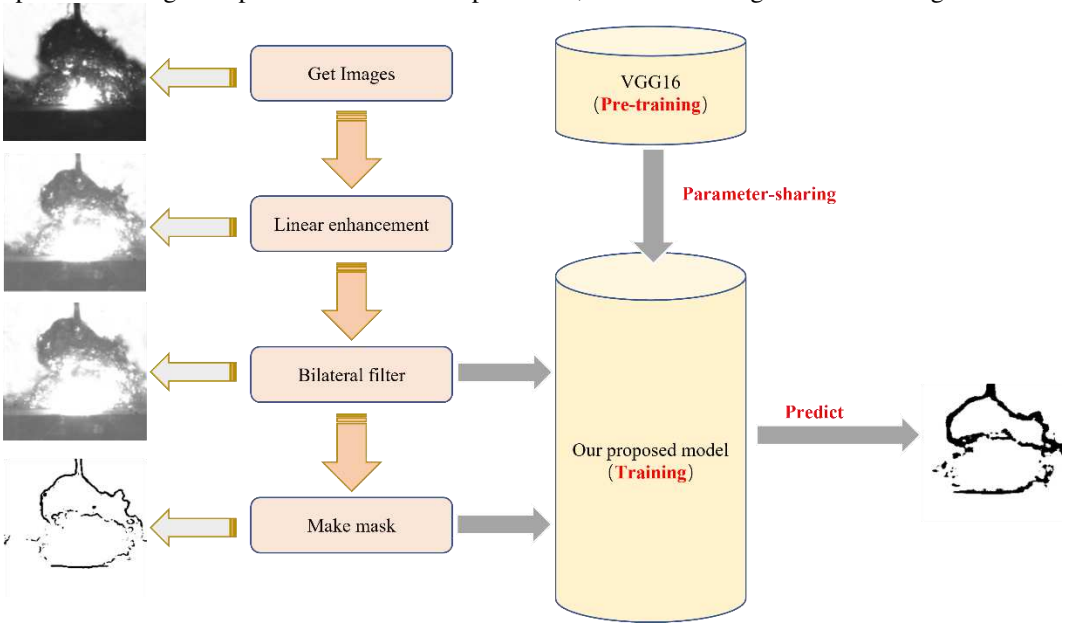

FIGURE 4. Flowchart of arc bubble edge detection method based on deep transfer learning

A. IMAGE PRE-PROCESSING

In order to achieve optimal results in bubble edge detection, the bubble edges in the welding images are enhanced through image pre-processing prior to being fed into the training model. First, convert the RGB images to grayscale images. Then, the bubble edges are then improved using a piecewise linear transformation. However, the piecewise linear transformation not only emphasizes the bubble region but also increases the noise. Therefore, the bilateral filtering algorithm [29] is utilized to denoise the image. As shown in Fig.5, after pre-processing, the upper and lower edges of the bubbles in the welding image are more pronounced, and noise is diminished.

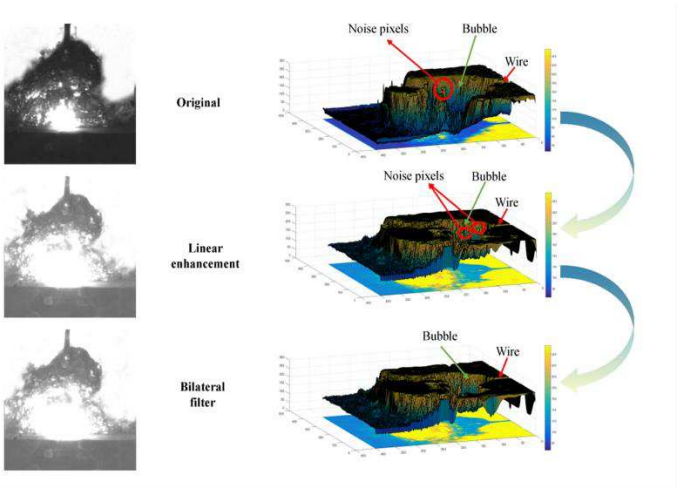

FIGURE 5. Images pre-processing results

B. DEEP TRANSFER LEARNING MODEL

Inspired by prior work in deep learning[30,31], we design an edge extraction model for underwater wet welding images based on deep learning and transfer learning. its structure is shown in Fig. 6. The ASS model uses the pruned VGG16 as the backbone, adds Convolutional Block Attention Module (CBAM) [32] at the shallow layer of the network, constructs a scale fusion module (SCM) consisting of five different scale features and a semantic fusion module (SEM) composed of four skip structures, and then fuses all features. From low level to high level, the SCM outputs are labeled o1, o2, o3, o4, and o5. The model outputs are labeled Predict. During model training, the backbone was first pre-trained, and then the constructed deep transfer learning model was fine-tuned.

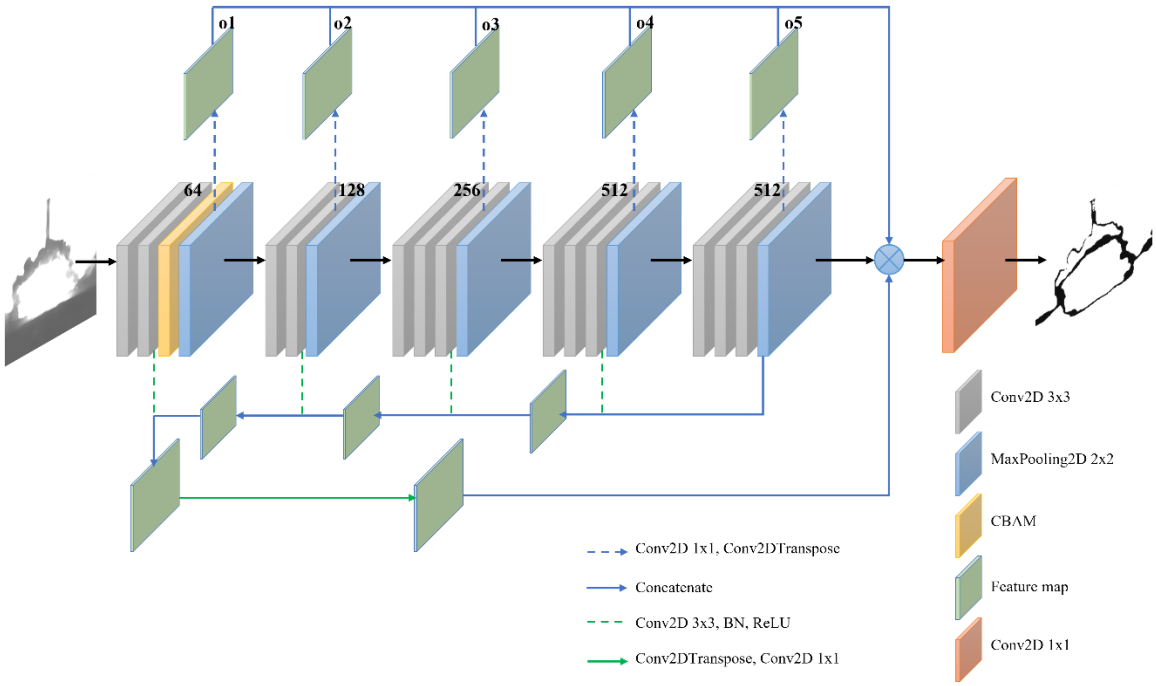

FIGURE 6. The structure of the ASS model

CBAM combines channel and spatial attention mechanisms to enhance the feature representation capability of convolutional neural networks. In the channel attention module, each channel of the feature map is treated as a feature detector, and the importance of different channels for specific features is determined by learning channel weights. It focuses more attention on channels that are helpful for the current task, thereby extracting more discriminative features. In contrast to channel attention, spatial attention focuses on "where" the effective information in the feature map is located, placing more emphasis on local weights and directing more attention to spatial positions that contribute to the task. CBAM infers attention maps along two

independent dimensions (channel and spatial), and then multiplies the attention maps with the input feature map for adaptive feature optimization. Since the model can only extract basic low-level features such as edges and textures in the shallow layers, we apply CBAM to the shallow layers to enhance the perception of local details, thus better capturing the details of arc bubble edges and improving edge clarity and accuracy.

The SCM simultaneously extracts and trains multiple-scale convolution features, which maximizes the use of feature data and simplifies the training process. This module, however, employs the method of directly fusing the same scale feature map, while RCF [29] fuses all features after convolution. All convolution features perform another convolution operation with 21 channels, which makes the model too complex and requires too many model parameters to be trained. When the model is overly complex, insufficient training samples frequently result in insufficient model training and make it challenging to improve model accuracy. When there are too many parameters to be trained, the benefits of transfer learning based on prior knowledge cannot be fully utilized, resulting in slow or difficult convergence of the model. Therefore, SCM avoids the aforementioned issues through direct fusion. The architecture order is as follows: perform a Concatenate operation to fuse the convolution features of the same scale as the backbone, perform a Convolution operation on the fusion feature with a channel number of 1, perform a Deconvolution operation to recover features that are smaller than the original scale, and obtain different scale feature maps via a Loss function.

Visualization is accomplished by outputting feature maps at various scales, and the results are shown in Fig. 7. It is demonstrated that the VGG16 model introduces too many pooling layers, which leads to too much spatial position information loss due to too large step size during deconvolution, resulting in too low resolution of the feature map output by the high layer of the network, which has a substantial impact on the image's fineness. In addition, the semantic features of the high-level feature map of the convolutional neural network are more prosperous, allowing them to better filter out the texture edges within the target compared to the low-level, resulting in outputs that are more concentrated on the target contour. Combining the characteristics of low-level and high-level network semantic information, the ASS model employs the SEM, which is comprised of four skip structures. It is able to perform local prediction while adhering to global prediction, solve the problem of high-level semantic loss of skeleton network, and generate high-quality images. Before each scale pooling, the module extracts the convolution layer features as input. Its construction sequence is as follows: Convolution, BatchNormalization, and Activation are successively applied to the high-level convolution features. Then, perform the Concatenate operation with the next layer of convolution features to iterate the previous step until the convolution features are restored to the original scale. The recovered features are then convoluted with a channel number of 1, fused with the output features of the SCM. The fused features are fed into the loss function, and the model is iteratively trained to output the optimal prediction results.

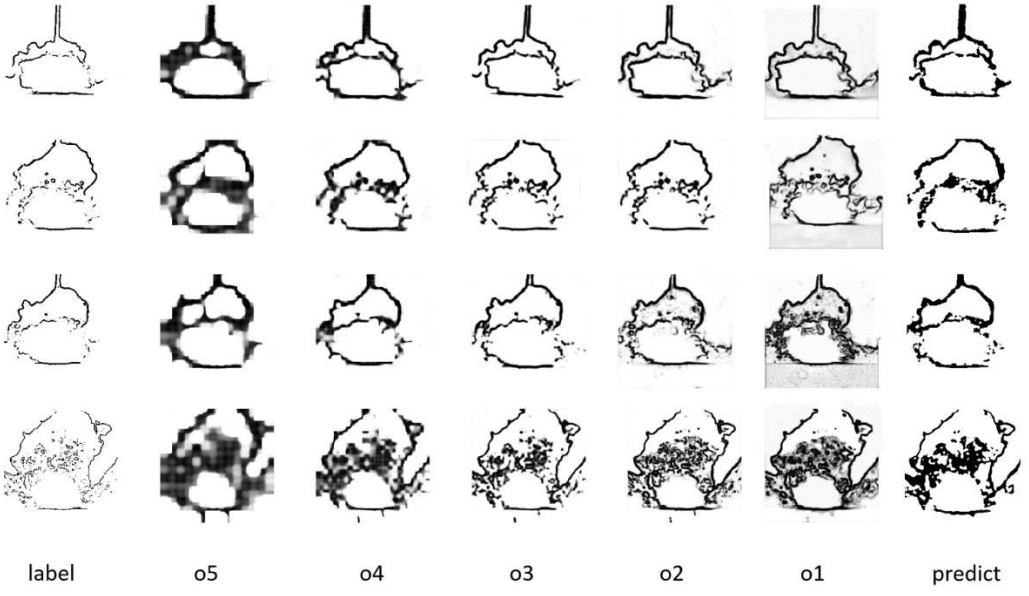

FIGURE 7. The visualization of feature map

C.TRAINING METHOD OF DEEP TRANSFER LEARNING MODEL

The overall training process of ASS model is shown in Fig.8, which consists of two stages. First, the VGG16 feature extractor is pre-trained on a large source domain dataset called ImageNet. During this stage, the activation value X_b of the

source domain dataset is input into the feature extraction function $f(\theta)$. Then the extracted feature is input into the classifier function $c(.|w_n)$. The outputting image activation value X_{pred} is input into the Loss function L_{pred} to Constantly train to obtain the optimal model. Sharing the feature extractor parameters with the ASS model will drastically reduce training time of the model and improve the generalization ability of the model. Second, the pre-trained feature extractor is combined with the CBAM, SCM and the SEM to construct a deep transfer learning model and finish the model's training on a small target domain Underwater wet welding dataset. During this stage, the activation value X_n of the target domain dataset is input into the feature extraction function $f(\theta)$, and the features that are more useful for arc bubble edge recognition are determined by CBAM. Then the extracted feature X^K and X^Z are then input into the SCM and the SEM, respectively. Finally, the outputting fusion feature image activation value X^{fuse} from two modules is input into the Loss function $L(W)$ to Constantly train to obtain the optimal model.

The training process of the ASS model mainly includes forward propagation and backward propagation. The forward propagation entails extracting convolutional features from VGG16, optimizing and selecting the features using CBAM, and then inputting them to the microcontroller and SEM. Finally, the outputs of both modules are combined to predict the edges of underwater welding bubbles. The backward propagation process compares the results of forward propagation prediction with the labels and continuously seeks the global minimum loss function value via the Adam optimizer.

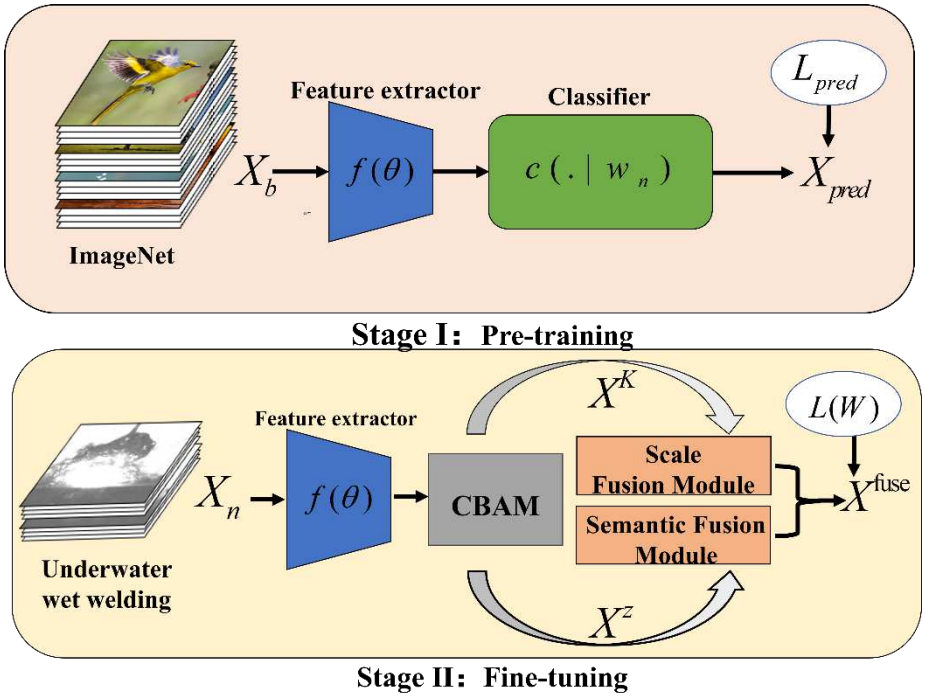

FIGURE 8. Flowchart of the overall training process

CBAM weights the extracted features in both channel and spatial dimensions. It generates a one-dimensional channel attention map Mc (Mc ∈ C×1×1) and a two-dimensional spatial attention map Ms (Ms ∈ 1×H×W). Here, C represents the number of channels, H is the height of the feature map, and W is the width of the feature map. The input features are denoted as F, while the output features are represented as F', as shown in equation (1), (2) and (3) respectively.

$$M_c(F) = Sigmoid\,(MLP(AvgPool\,(F)) + MLP\,(MaxPool\,(F))) \quad (1)$$

$$M_s(F) = Sigmoid\,(conv\,([AvgPool\,(F);\ MaxPool\,(F)])) \quad (2)$$

$$F' = M_C\,(F) \otimes F, \quad F'' = M_s\,(F') \otimes F' \quad (3)$$

SCM extracts convolution features from VGG16, performs successive Concatenate, Convolution, and Deconvolution operations, and then inputs the feature map into the Loss function. Therefore, the forward propagation formula of the module is as follows:

$$X^K = trans^{k\neq 1}\{conv[concat < \sum_{j=1}^{n}\sum_{i=1}^{|l_k|} X_{i,j}^k * W_{i,j}^k >]\} \quad (4)$$

Where $|l_k|$ is the pixels number of the k -th scale feature map, k is the number of different scale features (equals to 5 here). k =1 represents that the output image scale is equal to the image original scale. Therefore, X^K is the activation value of all neurons in the k -scale feature output image; $X_{i,j}^K$ is the activation value of the i -th neuron of the extracted j -th feature map; and $W_{i,j}^k$ is the parameters that the neuron can learn. $trans\{\bullet\}$, $conv[\bullet]$ and $concat<\bullet>$ are the deconvolution function, convolution function, and fusion function, respectively.

Before the Pooling layers, SEM extracts convolution features from VGG16 and performs Deconvolution, BathNormalization, Activation, and Concatenation operations in succession. The output of the features of the SEM is then fused with features of the SCM. It will input the loss function to obtain the model prediction result. Therefore, the forward propagation formula of the module is as follows:

$$X^s = conv[\sum_{z=1}^{4} concat < relu(bn[trans\{\sum_{i=1}^{|l_z|} X_i^z * W_i^z\}]), \sum_{i=1}^{|l_{z+1}|} X_i^{z+1} * W_i^{z+1} >] \quad (5)$$

$$X^{fuse} = concat < \sum_{K=1}^{5} X^K, X^s > \quad (6)$$

Where z is the number of skip connections (equals to 4 here). Where z is the number of skip connections (equals to 4 here). $|l_z|$ is the number of pixels in the input feature map of the z -th skip connection, and X_i^z 、 W_i^z are the activation value and learnable parameters of the i -th neuron of the input feature map of the jump connection, respectively. Similarly, $|l_{z+1}|$ is the number of pixels in the input feature map of the $(z+1)$-th skip connection, X_i^{z+1} 、 W_i^{z+1} are the activation value and learnable parameters of the i -th neuron in the input feature map of the skip connection; X^s is the semantic fusion feature output image all neuron activation value; $bn[\bullet]$ is the regularization function, $relu(\bullet)$ is the ReLU activation function, respectively; and X^{fuse} is the activation value of all feature fusion output feature maps.

Edge detection can be formulated as a pixel-level binary classification problem [33]. Because the proportion of edge pixels is much lower than that of non-edge pixels, if an ordinary loss function is used, the supervised learning process will tend to segment a large proportion of non-edge pixels, causing model instability and affecting training results. Moreover, there is a greater disparity between the number of edge and non-edge pixels in the edge detection problem of underwater welding images. In order to further balance the edge pixels and non-edge pixels, the parameter λ must be added. In conclusion, the ASS model must calculate the loss of each pixel relative to the pixel label at each size-side output layer and then perform backward propagation. The formula for backward propagation is as follows:

$$l(X_i;W) = \begin{cases} \alpha \log(P(X_i;W)) & y_i = 0 \\ 0 & \\ \beta \log(P(X_i;W)) & y_i = 1 \end{cases} \quad (7)$$

$$\begin{cases} \alpha = \lambda . \dfrac{|Y^+|}{|Y^+|+|Y^-|} \\ \beta = \lambda . \dfrac{|Y^-|}{|Y^+|+|Y^-|} \end{cases} \quad (8)$$

Where $|Y^+|$ 、 $|Y^-|$ are the number of edge pixels and non-edge pixels in the image, respectively. The hyper-parameter λ is to balance edge pixels and non-edge pixels. X_i is the activation value at pixel i ; y_i is the probability

that the pixel i is the edge point in label; W is the learnable parameters in the neural network; $P(\bullet)$ represents the $sigmoid$ activation function.

$$L(W)=\sum_{i=1}^{|I|}(\sum_{k=1}^{K}l(X_i^{(k)};W))+l(X_i^{(k)};W)) \tag{9}$$

$$w_i^n=w_i^o-\alpha(1+\frac{\partial L(W)}{\partial w}) \tag{10}$$

Where $X_i^{(k)}$ is the activation value of the i-th pixel in the output image representing the k-th scale feature. X_i^{fuse} is the activation value of the i-th pixel in all feature fusion output images. $|I|$ is the total number of pixels per image. w_i^n is them i-th pixel learning parameter in the feature map updated by backward propagation, and w_i^o is the i-th pixel learnable parameter in the feature map updated by input backward propagation. α is the learning rate, which is related to the Adam optimizer and gradually decreases with the training process.

IV. MODEL TEST AND ANALYSIS

The training and testing are executed on Intel i9-9900K, 64G RAM. GPU is NVIDIA GeForce RTX 3080Ti.

To evaluate the performance of our proposed model, we compared it with commonly used edge detection models such as RCF, FCN8s, and UNet. Figure 9 displays the detection results of our model in comparison with RCF, FCN8s, and UNet. From Figure 9, it can be observed that UNet often fails to accurately extract the edges of arc bubbles. FCN8s produces unclear contours with jagged edges. Although RCF can detect the correct edges to a certain extent, its recognition performance is poorer in the transitional areas affected by arc light, small bubbles, and impurities. In contrast, our model achieves continuous and clear detection of bubble edges, and exhibits better recognition performance in the challenging regions where arcs and bubbles intersect. In addition, since the lower edge of the bubble in the model detection diagram is discontinuous, the least square algorithm is used to fit the lower edge of bubbles in the image to obtain the final result image. The upper boundary edge of the underwater welding bubble is marked with a red line, while the lower boundary edge of the underwater welding bubble is marked with a blue line.

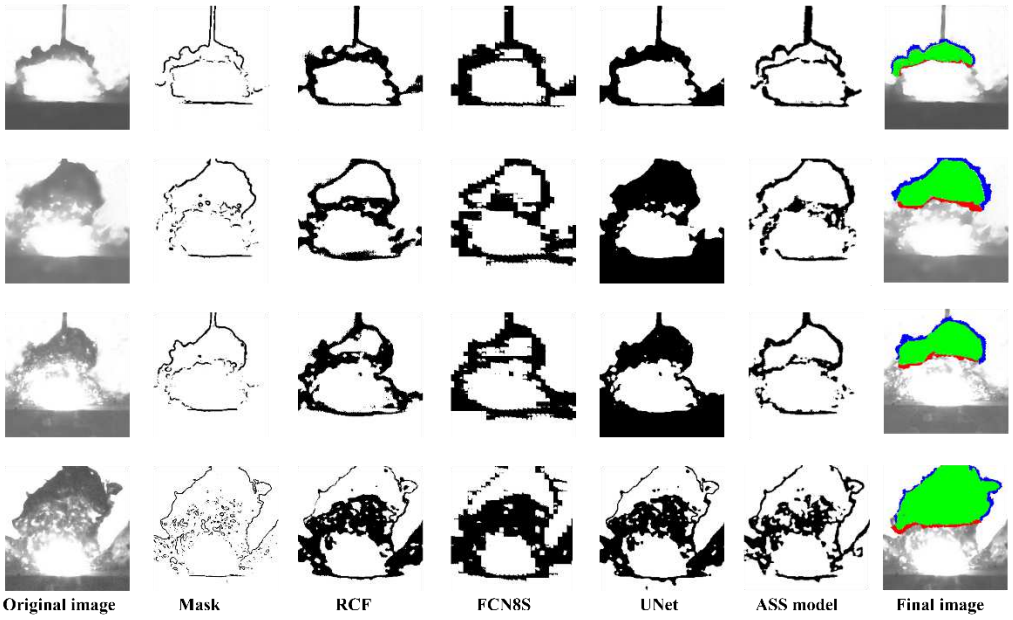

FIGURE 9. The results of ASS model and other edge detection models

In addition, three standard metrics (F-measure, Mean Absolute Error (MAE), accuracy and Pixel error (PE)) are utilized to quantitatively assess the image edge detection capability of the proposed method.

The F-measure is the harmonic mean of average precision and recall rate. According to the mathematical relationship between average precision, average recall rate, TP, FP, and FN, the following is the formula for F-measure:

$$F-measure=\frac{(1+\beta^2)\ TP}{(1+\beta^2)\ TP+\beta^2 FN+FP}*100\% \qquad (8)$$

Where β^2 is equal to 0.3, TP is the number of correctly identified defect samples, FP is the number of non-defect samples that are incorrectly identified as defects, and FN is the number of incorrectly identified defect samples. If the F-measure is high, which indicates the defect detection ability of the approach is good.

MAE [34] is a metric that measures the average difference between prediction and ground-truth maps. If the MAE is low, then the stability of the method is high. Let P and R are the saliency map and ground truth values, respectively; then, the MAE is as follows:

$$MAE=\frac{1}{row*col}\sum_{i=1}^{row}\sum_{j=1}^{col}|\,P(i,j)-R(i,j)\,| \qquad (9)$$

The row and col represent the row and width of the image, respectively. The accuracy is calculated as follows:

$$Accuracy=\frac{TP+TN}{row*col}*100\% \qquad (10)$$

PE represents the number of incorrectly predicted pixels divided by the total number of pixels. A lower value indicates that the model's prediction is closer to the true mask, indicating better model performance.

As shown in Table.1, we evaluate and compare the ASS model with other traditional edge detection algorithms using MAX F-measure (MAXF), MAE, and Accuracy. We can see that our method has a great effect on underwater wet welding images (228 images and masks). The MAF of the conventional algorithm when processing the Underwater wet welding image data set is approximately 0.91. However, the MAXF of our model increased to 0.95, indicating that the ASS model achieves optimal edge extraction. Our model achieved the best scores for MAE, Accuracy and PE, with 0.116 0.936 and 0.077, respectively.

The accuracy of the RCF, FCN8s, UNet and our model using the Underwater wet welding image dataset during training and validation are presented in Fig. 10. It is evident from the figure that the change curve of the ASS model within 50 epochs is always higher than others curves. When the model accuracy curve tends to smooth, the ASS model accuracy curve is approximately 10% greater than FCN8s, approximately 8% greater than RCF and approximately 6% greater than UNet. During model training, the ASS model is highly stable and accurate, and its convergence speed is rapid. The overall training process of ASS model is superior to RCF, FCN8s and UNet.

Table1 Quantitative comparison including MAXF, MAE, and accuracy on the Underwater Welding dataset

Methods	MAXF	MAE	Accuracy	PE
RCF	0.918	0.160	0.856	0.138
FCN8s	0.902	0.190	0.834	0.157
UNet	0.927	0.162	0.870	0.116
ASS model	0.957	0.116	0.936	0.077

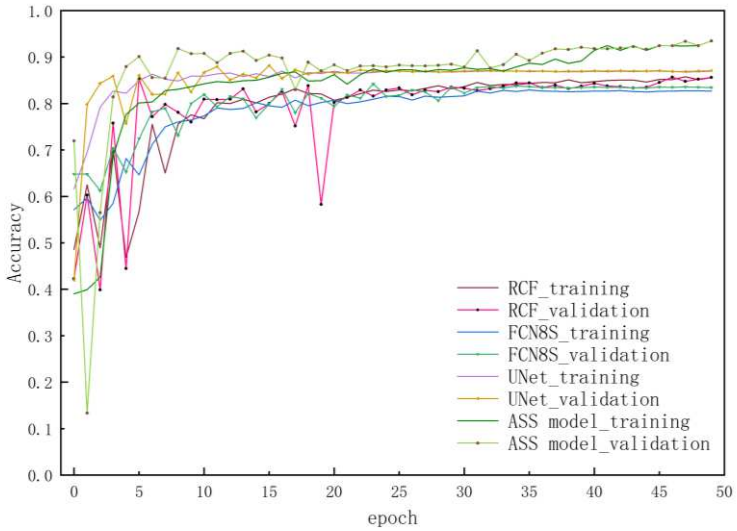

FIGURE 10. Change of ASS Model Training Accuracy and Test Accuracy in each model

In addition, we conducted ablation experiments to analyze our model. The detection results of our model without using CBAM and without using SEM are shown in Figure 11, and the corresponding evaluation metrics are presented in Figure 12.

Without the inclusion of SEM, the model that only performs multi-scale fusion has already lost a significant amount of high-level semantic information, resulting in inaccurate detection of arc bubbles. After adding SEM, the regions of arc bubbles can be roughly identified, but further improvement is needed for the detection accuracy in the detailed areas. The integration of CBAM allows the model to focus on more important edge features, thereby enhancing the accuracy of edge detection in complex scenes. The precision curve graph also demonstrates that the utilization of SEM and CBAM improves the model's accuracy by approximately 10%.

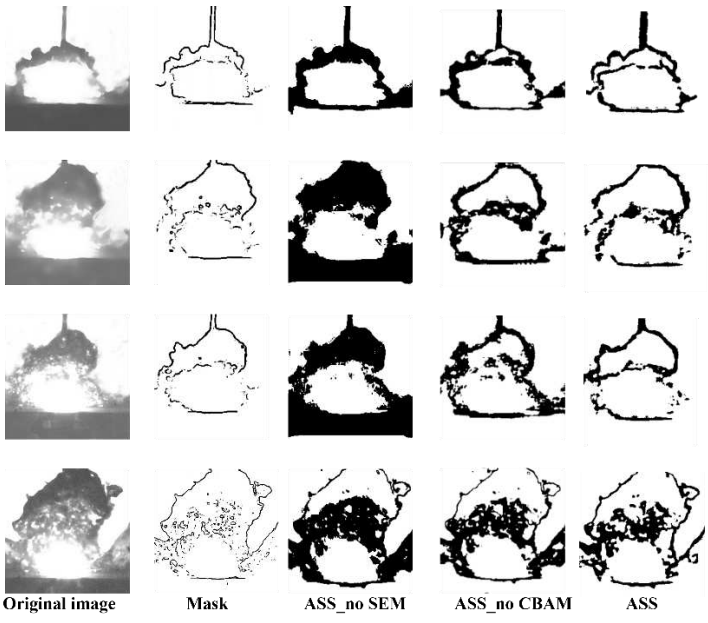

FIGURE 11. The results of ASS model without using CBAM and without using SEM

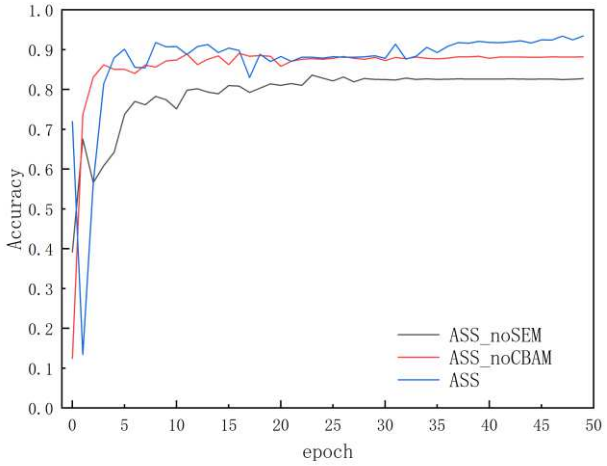

FIGURE 12. The accuracy of ASS model without using CBAM and without using SEM

The evaluation metrics of the different methods from the BSDS500 dataset (200 images) is calculated, as shown in Table.2. The ASS model has a higher average accuracy (0.781) than other models. In terms of PE, the model detection result is 0.168, which is lower than that of other models. The stability of the model processing results is also greater than that of other models, as demonstrated by the fact that when the MAE is smaller, the more stable it is.

TABLE 2 Quantitative comparison including MAE, accuracy and PE on BSDS500 dataset

Edge detection method	MAE	ACCURACY (%)	PE
UNet	0.303	69.57	0.198
RCF	0.281	67.89	0.205
ASS model	0.159	78.14	0.168

Finally, we measure that the average prediction time of each frame of ASS model is 93 ms, which meets the real-time requirements under the assumption of high accuracy.

V. CONCLUSION

This article proposes a Transfer Learning-based ASS model for detecting the edges of underwater wet welding arc bubbles. This method addresses the issue of insufficient sample size in underwater wet welding images. With this approach, the edges of underwater wet welding arc bubbles can be obtained, and their sizes, positions, and morphologies can be inferred. It is of great significance to study arc mechanisms, control processes, optimize parameters, and improve the quality of underwater wet welding. The main contributions of this paper are as follows:

The ASS model adopts the transfer learning method based on model-sharing parameters in order to obtain prior knowledge by pre-training the feature extractor and to solve the problem of insufficient training samples. Furthermore, this model presents CBAM, SCM and SEM. The introduction of the attention mechanism allows the model to focus more on the edge features we need and improve the detection accuracy. By simultaneously training scale features, it maximizes feature information and simplifies training steps. The problem of high-level semantic loss of backbone is solved by combining low-level and high-level network semantic information via the SEM skip structure.

We compared the performance of our ASS model with RCF, FCN8s, and UNet models for underwater wet welding arc bubbles edge extraction. The experiments demonstrated that our edge detection method is more effective and stable than other

conventional edge detection models on two different datasets. The detected arc bubble edges are more precise and clearer. In addition, our method achieves a detection speed of 93 ms per frame, which meets the requirement of real-time monitoring.

Data availability

The datasets generated and analyzed during the current study are not publicly available due the confidentiality of the data but are available from the corresponding author on reasonable request.

REFERENCES:

[1] Xie, S. and Z. Tu, Holistically-Nested Edge Detection. International Journal of Computer Vision, 2015. 125(1-3): p. 3-18.

[2] Zhao, J., et al. EGNet: Edge Guidance Network for Salient Object Detection. in Proceedings of the IEEE/CVF international conference on computer vision. 2019. Seoul: IEEE.

[3] Bertasius, G., J. Shi and L. Torresani. DeepEdge: A Multi-Scale Bifurcated Deep Network for Top-Down Contour Detection. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. Ithaca: Cornell University Library, arXiv.org.

[4] Wang Y, Song K, Liu J, et al. RENet: Rectangular convolution pyramid and edge enhancement network for salient object detection of pavement cracks[J]. Measurement, 2021, 170: 108698.

[5] Xiao Z, Song K Y, Gupta M M. Development of a CNN edge detection model of noised X-ray images for enhanced performance of non-destructive testing[J]. Measurement, 2021, 174: 109012.

[6] Hong, Y., et al., In-process monitoring of lack of fusion in ultra-thin sheets edge welding using machine vision. Sensors, 2018. 18(8): p. 2411.

[7] Ma, G., et al., A vision-based method for lap weld defects monitoring of galvanized steel sheets using convolutional neural network. Journal of Manufacturing Processes, 2021. 64: p. 130-139.

[8] Wang, Y., et al., Active disturbance rejection control of layer width in wire arc additive manufacturing based on deep learning. Journal of Manufacturing Processes, 2021. 67: p. 364-375.

[9] Cheng, Y.C., et al., Automated recognition of weld pool characteristics from active vision sensing. Weld J, 2021. 100(5): p. 183S-192S.

[10] Cheng, Y., et al., Dynamic estimation of joint penetration by deep learning from weld pool image. Science and Technology of Welding and Joining, 2021. 4(26): p. 279-285.

[11] Wang, J., et al., Characterization of the underwater welding arc bubble through a visual sensing method. Journal of Materials Processing Technology, 2018. 251: p. 95-108.

[12] Kumaresan, S., et al., Transfer Learning With CNN for Classification of Weld Defect. IEEE access, 2021. 9: p. 95097-95108.

[13] Pan, H., et al., A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects. IEEE access, 2020. 8: p. 119951-119960.

[14] Jiao, W., et al., End-to-end prediction of weld penetration: A deep learning and transfer learning based method. Journal of Manufacturing Processes, 2021. 63: p. 191-197.

[15] Wang, Z., et al., Recognition of GTAW weld penetration based on the lightweight model and transfer learning. Welding in the World, 2022: p. 1-14.

[16] Bahador A, Du C, Ng H P, et al. Cost-effective classification of tool wear with transfer learning based on tool vibration for hard turning processes[J]. Measurement, 2022, 201: 111701.

[17] Rowe, M.D., S. Liu and T.J. Reynolds, The effect of ferro-alloy additions and depth on the quality of underwater wet welds. WELDING JOURNAL-NEW YORK-, 2002. 81(8): p. 156-S.

[18] Tsai, C.L. and K. Masubuchi. Mechanisms of rapid cooling and their design considerations in underwater welding. in Offshore Technology Conference. 1979: OnePetro.

[19] Yurioka, N. and Y. Horii, Recent developments in repair welding technologies in Japan. Science and Technology of Welding and Joining, 2013. 11(3): p. 255-264.

[20] Guo, N., et al., Study of underwater wet welding stability using an X-ray transmission method. Journal of materials processing technology, 2015. 225: p. 133-138.

[21] Wang, J., et al., Experimental study of arc bubble growth and detachment from underwater wet FCAW. Welding in the world, 2019. 63(6): p. 1747-1759.

[22] Fu, Y. and Y. Liu, BubGAN: Bubble generative adversarial networks for synthesizing realistic bubbly flow images. Chemical Engineering Science, 2019. 204: p. 35-47.
[23] Zhao, Q., R. Li and Q.U. Qian, Research on statistical detection method of micro bubbles in transparent layer of quartz crucible based on image processing. Journal of Crystal Growth, 2021. 556: p. 125966.
[24] Saha P K, Pal R, Sarkar S, et al. A novel image processing technique for detection of pseudo-occluded bubbles and identification of flow regimes in a bubble column reactor[J]. Measurement, 2022, 189: 110568.
[25] Yang, X., et al., Research on image recognition and detection method of sapphire bubbles. Journal of instrumentation, 2019. 14(12): p. P12013-P12013.
[26] Misra, N.N., R. Phalak and A. Martynenko, A microscopic computer vision algorithm for autonomous bubble detection in aerated complex liquids. Journal of Food Engineering, 2018. 238: p. 54-60.
[27] Zhuang, F., P. Luo and Q. He, Survey on transfer learning research. Journal of Software, 2015. 1(26): p. 26-39.(in Chinese).
[28] Zhang, X., Y. Zhuang and F.W. Yan, Status and development of transfer learning based category-level object recognition and detection. Acta Automatica Sinica, 2019. 7(45): p. 1224−1243.
[29] Tomasi, C. and R. Manduchi, Bilateral filtering for gray and color images. 1998: p. 839-846.
[30] Liu, Y., et al., Richer Convolutional Features for Edge Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. 41(8): p. 1939-1946.
[31] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[32] Woo S, Park J, Lee J Y, et al. Cbam: Convolutional block attention module[C] Proceedings of the European conference on computer vision (ECCV). 2018: 3-19.
[33] Kaganami, H.G. and Z. Beiji. Region-based segmentation versus edge detection. in 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing. 2009: IEEE.
[34] Borji, A., et al., Salient Object Detection: A Benchmark. IEEE Trans Image Process, 2015. 24(12): p. 5706-22

Author contributions

B.G. and X.L. wrote the main manuscript text and X.L. prepared figures. All authors reviewed the manuscript.

Competing Interests

The authors declare no competing interests.

Acknowledgements

This work was supported in part by Jiangxi Provincial Natural Science Foundation under grant number 20224BAB202033，in part by Jiangxi Provincial Key Research and Development Program under grant number 20192BBE50015, and in part by Science and Technology Project of Jiangxi Provincial Education Department under grant number GJJ201912.

Bo Guo received the M.S. degree in control theory and control engineering from the Northwestern Polytechnical University and the Ph.D. degree in mechatronics from South China University of Technology. He is currently the director of Nanchang Key Laboratory of Welding Robot & Intelligent Technology and an Associate Professor with the School of Electrical Engineering, Nanchang Institute of Technology. His main research interests include machine learning, computer vision, and welding robot.

Xu Li received the B.S. degree in electrical engineering and automation from Hubei Normal University, in 2020. He is currently pursuing the M.S. degree in Nanchang Institute of Technology. His research interests include intelligent control and machine learning.