

Analysis of Weld Pool Region Constituents in GMAW for Dynamic Reconstruction through Characteristic Enhancement and LSTM U-net Networks

Tianpu Li, Yue Cao, YuMing Zhang*

Department of Electrical and Computer Engineering and Institute for Sustainable Manufacturing, University of Kentucky, Lexington, KY 40506, USA

* Corresponding author email: yuming.zhang@uky.edu

Abstract: This study aims to extract critical scenes/continents in the weld pool region during gas metal arc welding (GMAW). The scenes considered include the wire, arc, and weld pool, while other secondary ones such as oxides are temporarily excluded. They are critical to understanding, analyzing, monitoring and controlling the welding process, in particular the critical correlation how the welding parameter, arc and weld pool are dynamically correlated. Unfortunately, such fundamental correlation has not been studied and lack of effective ways to simultaneously monitor/extract these scenes is responsible. With the development of optoelectronic devices, weld pool regions can be better imaged. However, because of the nature of the scenes in particular the arc which is formed by ionized gas without a clear boundary and highly dynamic, detecting them using computer vision is challenging. Deep learning is an effective method, but model training usually needs a large number of labels. As manually labeling is expensive, we propose an approach to address this challenge that can train a model from a small, inaccurately labeled dataset. This approach is designed, per the characteristics of the scenes and their dynamics All-in-One Network (AOD-Net) was deployed first for defogging, and then the YOLOX network was utilized to identify regions of interest to reduce the impact of molten metal splashes on image sharpness. Subsequently, we developed a timed segmentation network incorporating the Long Short-Term Memory (LSTM) mechanism into U-Net, which can be used to extract more accurate information about the weld pool by combining the temporal and spatial information in the continuous process of welding at a low cost because our scene of interest is in a continuous and dynamic evolutionary process. After defogging and removing the effects of molten metal spatter, we can obtain information on the dynamics of the weld pool and the weld arc at the same time. Experimental results verified that the trained network could extract the critical boundaries accurately under various welding conditions despite the highly dynamic changes and fuzziness of the views.

Keywords: Machine Learning, Welding, Computer Vision, Weld Pool, GMAW.

1 . Introduction

Gas Metal Arc Welding (GMAW) stands out as an exceptionally efficient welding technique where the filler material is continuously and automatically fed and melted through the arc. Its high level of automation makes it easily adaptable to robotic systems, rendering it the preferred choice for various applications such as additive manufacturing, cladding, and general welding processes [1].

The weld pool region is the epicenter for all intricate welding phenomena, necessitating real-time monitoring and control, particularly in robotic welding scenarios where human intervention is

absent. While numerous techniques exist for monitoring the weld pool, including infrared sensors [2,3], vision-based methods emerge as the most direct surrogate for human observation. However, visual sensing of the weld pool region poses challenges, primarily due to the obstructive nature of arc radiation during welding processes. Richardson and Gutow introduced the coaxial observation method, employing a camera positioned coaxially above the tungsten electrode to observe the weld pool area effectively [4]. In another effort, by strategically utilizing a monochromatic or single-color laser as an illumination source, the reflection from the object to the imaging sensor can be enhanced, overcoming the dominance of arc radiation [5-9]. Furthermore, the illumination laser has been pulsed to reach very high power and synchronized with the shutter of the camera [10]. Since the specular reflection from the liquid weld pool is not collected by the camera's sensor, and the surrounding solid metal glows due to diffuse reflection to form a bright area, the dark area in the image is the weld pool. In this way, the boundaries of the weld pool can be clearly observed. This intricate synchronization ensures that the illumination laser, pulsating at high power, effectively enhances the contrast between the weld pool and its surroundings. By selectively filtering out specular reflections and capitalizing on the diffuse reflection from the surrounding metal, the resulting image distinctly highlights the weld pool's boundaries, enabling precise observation and analysis.

High Dynamic Range (HDR) cameras offer another avenue to observe the weld pool region. They utilize advanced imaging technology to capture a broader range of luminance levels in a single image, thereby providing enhanced detail and clarity in both the brightest highlights and the darkest shadows. Cui employed an HDR camera to monitor welding processes, specifically studying the dynamic behavior of the molten pool and locking holes in real-time welding processes [13]. The ease of deployment of HDR cameras in manufacturing environments, coupled with their ability to produce high-clarity images distinguishing melt pool contours and arcs simultaneously, facilitates a comprehensive understanding of melt pool dynamics. This capability makes them ideal for our purpose of capturing all critical scenes in a single image.

Even in high-quality images such as those captured by HDR cameras, the weld pool boundary is still easier to identify than that of the arc. Extracting the weld pool boundary from images requires target detection through image processing. There have been extensive works to detect weld pool boundary using conventional segmentation algorithms [4-14]. They focused on feature extraction from a single image, making it difficult to obtain advanced semantic information. Rother et al. extended the MRF segmentation method [15] and proposed for the first time the concept of collaborative segmentation, which refers to the extraction of common foreground regions from multiple images without human intervention. The foreground features of the seed images are extracted as a priori knowledge by utilizing the classical segmentation method and then used to process a set of images containing the same or similar objects. Vicente et al. [16] proposed to extend the Boykov-Jolly method using multiscale decomposition based on MRF collaborative segmentation, as well as the L1 criterion model [15], the L2 criterion model [17], and the reward model [18]. Alexe et al. [19] proposed a measure based on object scoring by quantitatively calculating the probability of any category of objects in an image window and using the highest scoring window as a feature calibration for each category of objects. Vincente et al. [20] used foreground objects and measured the similarity between the objects, extracting the highest scoring features from multiple candidate object categories in the iCoseg dataset with good

experimental results. However, conventional algorithms require task-specific efforts to design the algorithm details per image characteristics. Although there have been various algorithms developed to detect the weld pool boundaries, they are all difficult to generalize. For arc boundaries, which are much more dynamic and fuzzier, there have been no reported successes. This underscores the need for innovative approaches in overcoming these challenges and developing generalizable robust algorithms capable of accurately delineating both weld pool and arc boundaries.

Convolutional neural networks (CNNs) based deep learning models provide generalized solutions for robust algorithms. Segmentation is achieved through training models using labeled boundaries, minimizing the need for task-specific efforts. LeNet-5 [21], AlexNet [22], and VGG [23] have demonstrated success in image classification, where the output layer categorizes the image. However, deep learning models typically require large labeled datasets for training. Long et al [24] introduced Fully Convolutional Networks (FCNs) and designed an encode-decode structure to extract high-level semantic information and map it back to the source image size. Skip connections or shortcut connections improve coarse pixel localization, and U-Net [25] leverages skip connections and feature cascading from encoder to decoder layers to obtain fine-grained image details. This approach yields promising results even with small datasets.

In terms of enabling deep learning techniques to process GMAW images, Baek et al. [26] proposed a model based on semantic segmentation of residual neural networks to effectively predict the depth of a molten weld pool, which is different from the weld pool boundary. Wang et al. [27] designed a multiscale feature fusion semantic segmentation network, Res-Seg, that can detect the weld pool boundary. The accuracy of the information obtained depends heavily on the dataset used for training, which requires a large number of datasets and accurate data labeling. Wang et al. [28] supplemented the melt pool dataset with a Deep Convolutional Generative Adversarial Network (DCGAN) [29] after designing the Res-Seg network. As such, there have been no generalizable successes in detecting the weld pool boundary, nor any mention of the arc boundary, in the highly dynamic and violent GMAW process without the availability of a large set of manual labels.

The continuous welding process in automated welding is characterized by high temperatures, spatter, smoke, and intense brightness, which often result in captured images of the weld pool being unclear. Accurately labeling large datasets under such conditions poses a significant challenge. Because image-based methods for detecting the weld pool boundary are susceptible to interference from welding arc brightness and deformation, previous studies have mainly focused on mitigating the effects of the arc. Bea et al. [30] addressed this issue by capturing weld pool images at the moment of the short circuit of the arc welding process, thereby circumventing the interference of arc light. However, due to the high brightness and variable shape of the arc, traditional CCD cameras struggle to capture clear images, leading to a lack of research on image-based welding arc boundary detection. Moreover, obtaining the boundary of the welding arc while simultaneously capturing the welding molten pool has not been adequately explored, largely due to the overwhelming brightness of the arc light.

In this study, we utilize an HDR welding camera to capture usable images of the weld pool and the arc simultaneously. We propose a composite U-Net network integrated with LSTM-Cell [31] to

extract key information of both the weld pool and the arc. Prior to processing, images undergo preprocessing using AOD-Net [32] and a Reinteresting module inspired by the YOLOX [33] network to mitigate the effects of spatter and smoke during welding. Leveraging the uniquely designed preprocessing module and feature recognition capabilities of the U-Net network, our model achieves precise and simultaneous extraction of the weld pool and welding arc boundaries, even with a limited training dataset. Experimental results demonstrate the robustness of the model, marking a significant advancement as it achieves the simultaneous detection of weld pool and welding arc boundaries, in addition to the wire, allowing us to automatically monitor the robotic welding process to assure process operation in the absence of human presence and analyze the welding process at higher level from big data.

2 . Welding Process :

Two 6-degree-of-freedom (DOF) industrial robots, UR10e and UR16e, were used as a welding robot and a monitoring robot, respectively (Fig. 1). On the monitoring robot, the Xiris' HDR weld camera XVC-1100, a digital weld camera for open-arc welding, was mounted to capture images of the arc and the molten pool. A bandpass filter centered at 650nm is mounted on the lens to filter out strong arc light. We use a PC to control the speed and direction of the welding robot and monitor the robot to follow the control to control the robot movement.

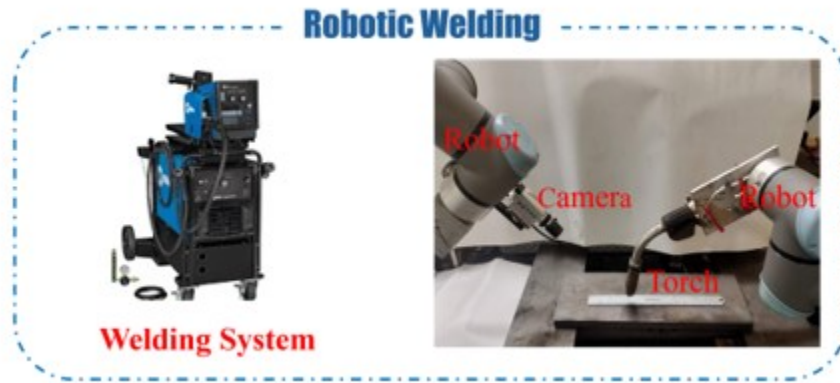


Figure 1 Automated Welding Systems

We set a constant welding speed during each experiment and the welding parameters are shown in Table 1. The welding power supply was operated in constant voltage mode with a high voltage of 40.8 V to ensure a spray transfer mode with less spatter and good bead appearance, thus making it suitable for bead-on-plate experiments on thick workpieces. The welding parameters were controlled by a computer using a PCI-6229 National Instruments data acquisition card running in C++.

Table 1. Welding parameters.

Welding Parameters	Values
Welding method	GMAW
Welding type	bead on plate
Polarity	DCEP
Welding speed	3-12 mm/s
Voltage	40.8 V
Wire feeding speed	310 in./min
Gas flow rate (95% Ar and 5% CO ₂)	15 ft ³ /h
Contact tip to workpiece	20 mm
Wire diameter	1.2 mm
Workpiece thickness	20 mm

Through several experiments, over 10 sets of videos of the welding process were captured. Each of them was segmented into images by frames for us to observe/screen. It was found that the visibility of the weld pool boundary in the captured images is often obscured by the smoke (Fig. 2). During GMAW process, smoke is generated due to the intense heat generated by the arc spots causing wire, workpiece, and workpiece coating to overheat to vaporize. The vapors are highly dynamic and fluctuate within the weld pool view. They will condense into fine particles to produce welding fumes. The composition of these fumes may vary depending on the welding material, the type of electrode used, and any coatings on the metal.

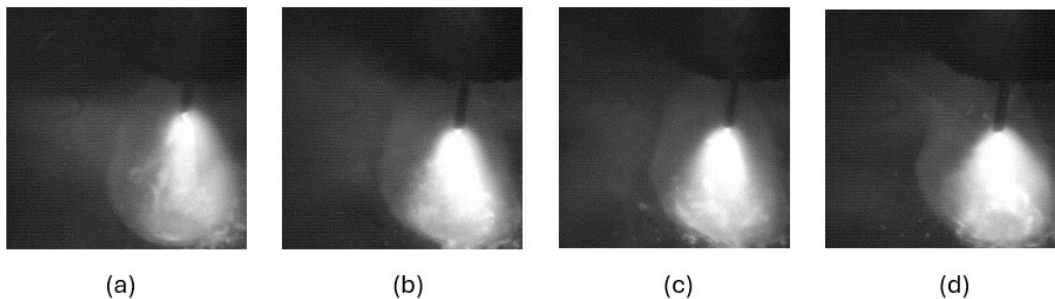


Figure 2 Smoke observable during the welding process

At the same time, we have found that spatters may arise during the welding process (Fig. 3). For open arc GMAW process, the melted wire is transferred into the workpiece through spray or globular transfer mode. Spray transfer minimizes spatters, but the current is high. At higher current densities, the metal may overheat and burst, forming fine-grained spatters.

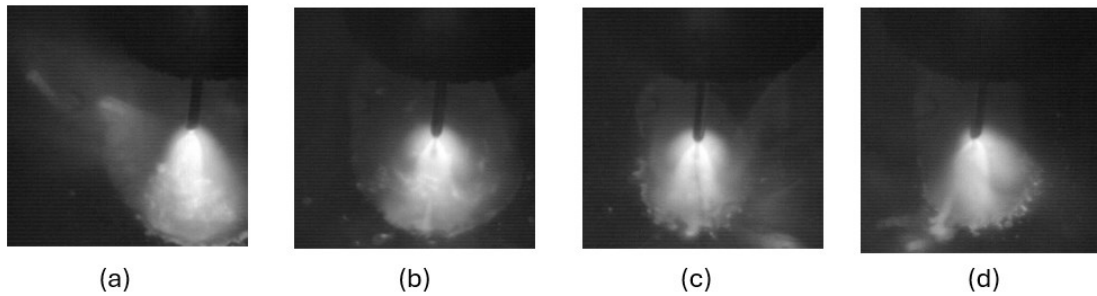


Figure 3 Molten spatters from the welding process

What we are primarily concerned with and aim to extract are the shape and width of the weld pool. Additionally, we need to extract information about the arc and torch. All complex welding phenomena in arc welding originate from the arc, which serves as the input of the welding process, while all welding phenomena occurring on the workpiece, which we are concerned with, constitute the resultant output. Welding parameters (such as current, voltage, speed, etc.) are the nominal and easily quantified inputs of the welding process. However, the actual inputs that directly control the welding process are the heat and force distribution of the arc. Welding parameters determine the integration of these distributions, but the actual distributions also require the center and standard deviation for a complete description. Therefore, in addition to the total heat/power of the arc, which is determined by welding parameters, the location of the arc center and the extent of arc spread are also critical in producing the welding output. Thus, monitoring and extracting this information are crucial for understanding and analyzing the welding process in greater depth.

Arc behaviors have been widely studied. Wu [34] et al. analyzed the effect of SiO₂-activated TIG arc plasma morphology, arc space electric field strength, arc temperature, arc current density, and arc on the geometry of the weld. Hua et al. [35] showed that the wire extension, arc length and welding current all affect the arc behaviors and instability. They analyzed the reasons for the increase in the arc deviation, which causes an increase in arc length and thus arc interruption. Fang et al. [36] showed that the welding arc shape can be judged during the welding current over zero to determine whether the arc extinguishing or re-arcing phenomenon occurs. Zhang [37] and others showed that the magnetic pipe will have an impact on the arc. Combined with Zhu [38] and others, they analyzed the magnetic bias blowing in the welding process on the welding arc, and the reasons for the welding arc can be inferred from welding to the magnetic field. Arc morphology helps determine the welding process's magnetic field distribution and other information. As such, the welding arc contains key information about the welding state, and obtaining the shape of the welding arc is one of the keys to understanding the welding process. Unfortunately, there have not been works to monitor and automatically extract the arc.

Analysis of the welding process and captured videos can guide us in deciding what is critical and understanding the challenges in extracting it. It is evident that using conventional image processing approaches to extract the necessary critical information is challenging, and a deep learning approach is required. Additionally, manually labeling a larger number of images for critical information is unrealistic. A significant issue is that critical scenes are fuzzy, making them difficult

for humans to identify clearly. Consequently, we will encounter the issue of limited labels and labeling errors. the dataset labeling, a set for labeling the molten pool, a set of arc and torch information for labeling. Labeling using Labelme open source labeling software, using polygonal labeling, labeling information obtained for the json format, this time you need to convert the json format data, converted to the form of the target mask. After the completion of the data set production, conversion is complete.

3 . Methodology:

We present a combinatorial neural network for scene extraction from weld pool (area) images. As shown in Fig.4, the network first preprocesses the captured image, and then inputs the image into "Denoising Module" to denoise the weld pool part of the image. The image is then fed into the "Reinteresting Module" to identify the region of interest to minimize the effect of smoke and spatter. The next step is to extract the shape and size of the weld pool from the preprocessed image using a time series-based segmentation model in the "Timing Segmentation Module" and in the "Wire and Arc Segmentation" respectively, where the "Segmentation Module" obtains the Wire and Arc information. Once this process is completed, key information about the weld pool during the welding process is acquired in real time.

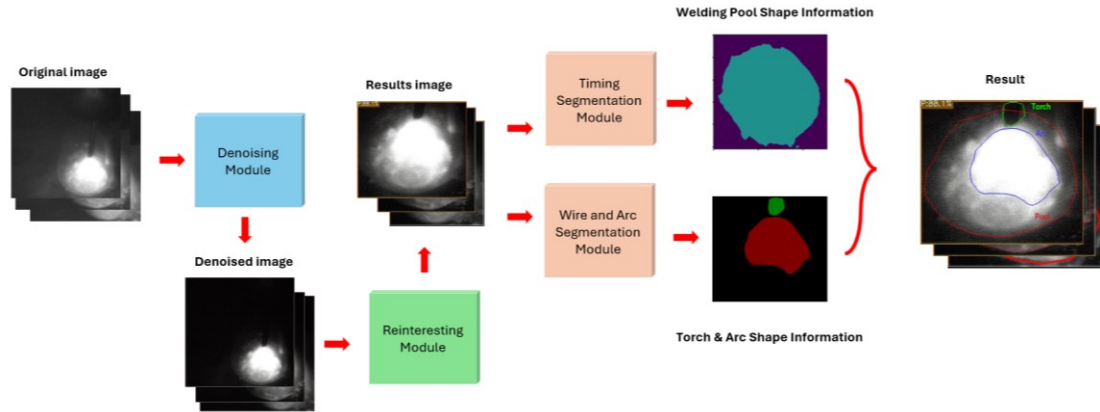


Figure 4 Welding Process Melt Pool Critical Information Detection Network Architecture

3.1 Preprocessing

3.1.1 AOD-Net Defogging

Since the smoke and spatter will also be captured, which would adversely affect the accuracy in extracting the weld pool, a pre-processing of the captured image may benefit. The pre-processing consists of Denoising and Reinteresting. A defogging model constructed using AOD-Net (an All-in-One Network) [32] is used for image denoising. After inputting the image into the K-estimation module (Fig. 5), the K-value at each pixel in the image is output, which is substituted into the new atmospheric diffraction model to calculate the resultant image after defogging.

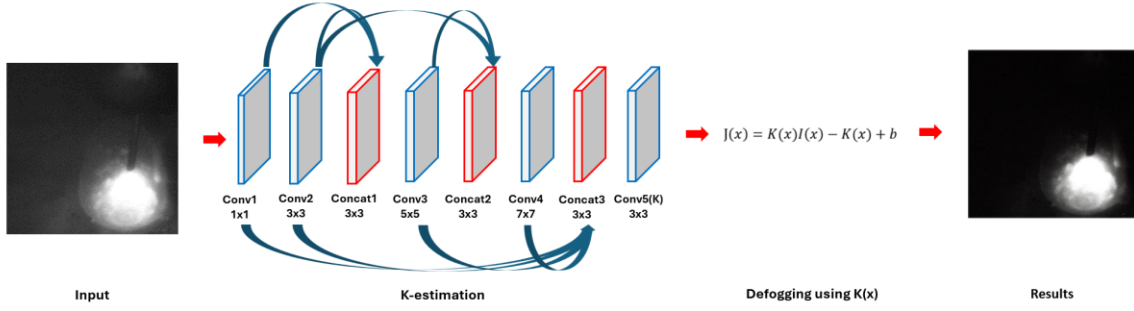


Figure 5 AOD-Net de-smoking structure for welding process

The model was designed based on a reformulated atmospheric scattering model that was transformed to yield the fog depth and concentration levels at each pixel in the evaluated image. Equation (1) represents the original atmospheric scattering model:

$$I(x, \lambda) = e^{-\beta(\lambda)d(x)}R(x, \lambda) + L_{\infty}[1 - e^{-\beta(\lambda)d(x)}] \quad (1)$$

where $I(x, \lambda)$ is the original unfogged image obtained by the detection system, and $R(x, \lambda)$ denotes the defogged image that needs to be recovered. The parameter x denotes the position of the pixel point in the image, λ is the wavelength of the light, and L_{∞} denotes the value of the atmospheric light at infinity (the source of the light comes from the sky at infinity). $t = e^{-\beta(\lambda)d(x)}$ denotes the transfer function, the physical meaning of which is the proportion of light that can reach the detection system after particle attenuation, where $\beta(\lambda)$ denotes the scattering coefficient, which is used to characterize the scattering ability of the medium for different wavelengths of light, and $d(x)$ is the light transmission distance. We first neglect the λ variable in Equation (1), the wavelength of light, so that the transmission function becomes the transmission matrix $t(x)$, and the recovered de-fogged image will be represented by $J(x)$, and the value of atmospheric light at infinity is denoted using A . Equation (2) can be obtained. At this point we can get the expression Equation (3) for the defogged image $J(x)$, where b is a constant bias defaulted to 1.

$$I(x) = J(x)t(x) + A[1 - t(x)] \quad (2)$$

$$J(x) = I(x)\frac{1}{t(x)} - A\frac{1}{t(x)} + b \quad (3)$$

At this point we want to get the image $J(x)$ after defogging. This requires to calculate the transmission matrix $t(x)$ and A . However, for these two parameters we can only estimate. A separate estimation of these two parameters will lead to the accumulation of error and amplification. As such, AOD-Net is proposed to integrate these two parameters. Using $K(x)$, Equation (3) is rewritten as Equation (4).

$$J(x) = I(x)K(x) - K(x) + b \quad (4)$$

To calculate the parameter $K(x)$, AOD-Net is designed with a K -estimation module, which is structured to utilize five convolutional layers to process the input data. Multi-scale information is

integrated by combining concatenated layers of varying sizes. Illustrated in Fig. 6, the convolutional component of this model involves concatenation operations denoted as concat1, concat2, and concat3. Concat1 connects features extracted from conv1 and conv2 layers, concat2 combines features from conv2 and conv3 layers, and concat3 integrates features from conv1, conv2, conv3, and conv4 layers. The final convolutional layer, conv5, is responsible for generating the atmospheric attenuation coefficient (K) of this image. This architecture ensures the incorporation of multi-scale information and the effective utilization of features extracted at different levels of abstraction, ultimately contributing to the accurate determination of the atmospheric conditions affecting the captured image.

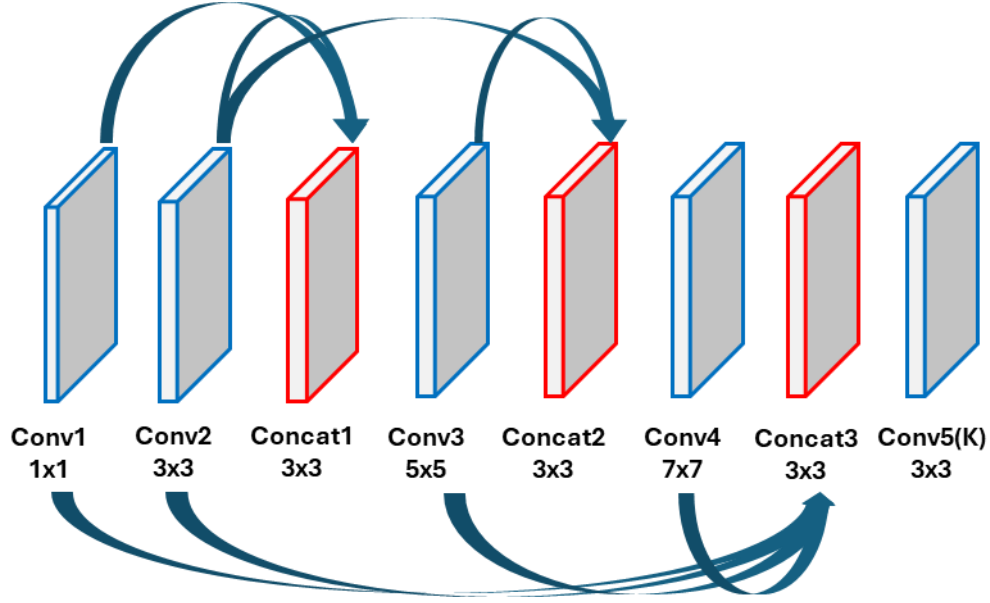


Figure 6 K -estimation module network structure

Since the K estimation module used is improved based on the atmospheric diffraction model, the training dataset of this network does not need to use weld pool images, but only images with different degrees of fog depths, and the use of this type of dataset makes the AOD-Net more robust and adaptable when dealing with weld pool images. The dataset we used is ground-truth images with depth meta-data from the indoor NYU2 Depth Database [32], which has 27256 data sets, a total of 10 epochs of training, and is used for weld pool defogging with good results. The obvious de-fogging effect can be seen (Fig.7).

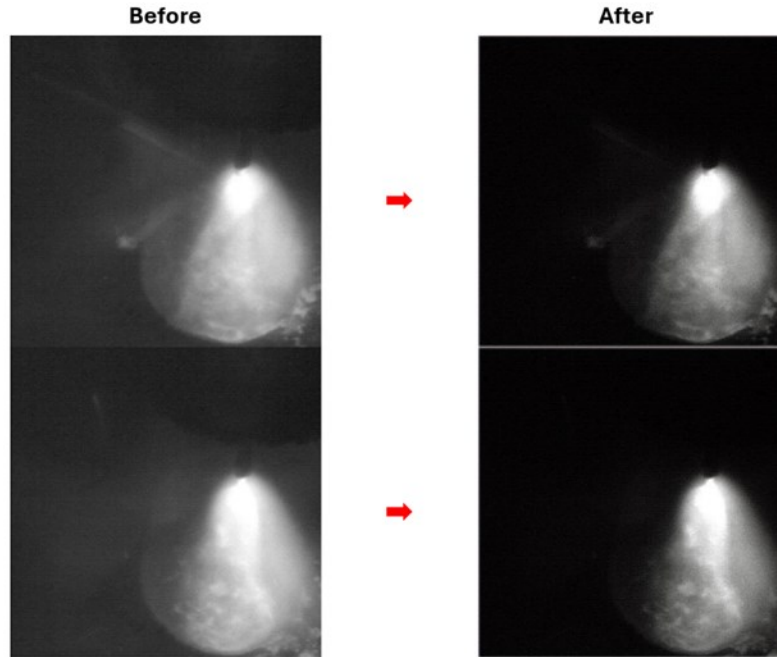


Figure 7 Comparison of the effect before and after applying the denoising module to the weld pool image

3.1.2 YOLOX-s Reinteresting Weld Pool

Given the interference of spatter during welding, which can often obscure the weld pool details in captured images, we mitigate this effect by amplifying the weld pool information. This is achieved through a process we term 'Reinteresting' the weld pool. To execute this, we refocus the image specifically on the weld pool region. This involves training the YOLOX-s network to accurately identify the weld pool amidst other elements in the image. Subsequently, we selectively eliminate extraneous information, leaving behind only the weld pool details. By implementing this 'Reinteresting' approach, we enhance the clarity and prominence of the weld pool within the image, facilitating more accurate analysis and interpretation. In Fig.8, only the lower right corner of the original image has the information we need. After we use the "Reinteresting" module, the image will be automatically cropped to the weld pool, effectively avoiding the effects of molten metal spatter in other areas.

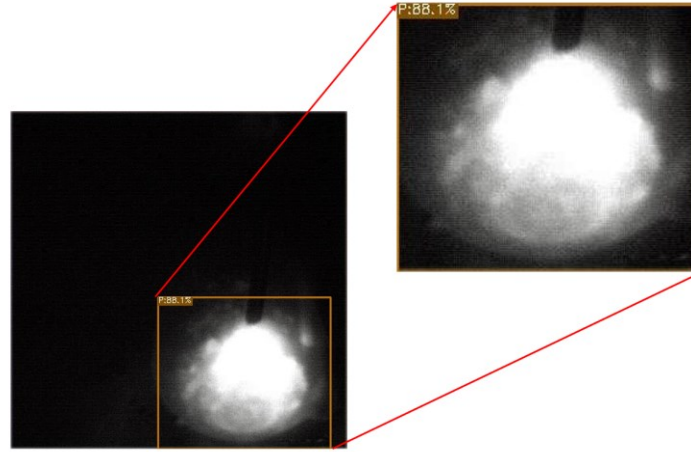


Figure 8 "Reinteresting" the weld pool.

YOLOX-s (Fig. 9) [33], introduced by Cavity Technologies in 2021, builds upon the foundation of the YOLOv5-s model with significant improvements. One notable enhancement lies in its utilization of two data augmentation techniques, namely Mosaic and Mixup, at the input stage. The Mosaic algorithm, a derivation of the CutMix algorithm [33], stands out for its effectiveness in detecting small targets. Unlike CutMix, Mosaic integrates multiple images by employing random scaling, random cropping, and random arrangement. Additionally, it ensures that the changing relationships of each image are preserved in their corresponding image labels. This approach proves particularly beneficial for accurate detection in scenarios involving small targets. On the other hand, MixUp data augmentation focuses on enhancing class diversity within the training dataset. By employing linear interpolation to blend images from different classes, MixUp generates new training samples, effectively expanding the dataset. This technique promotes robustness and generalization of the model by exposing it to a wider range of class variations during training.

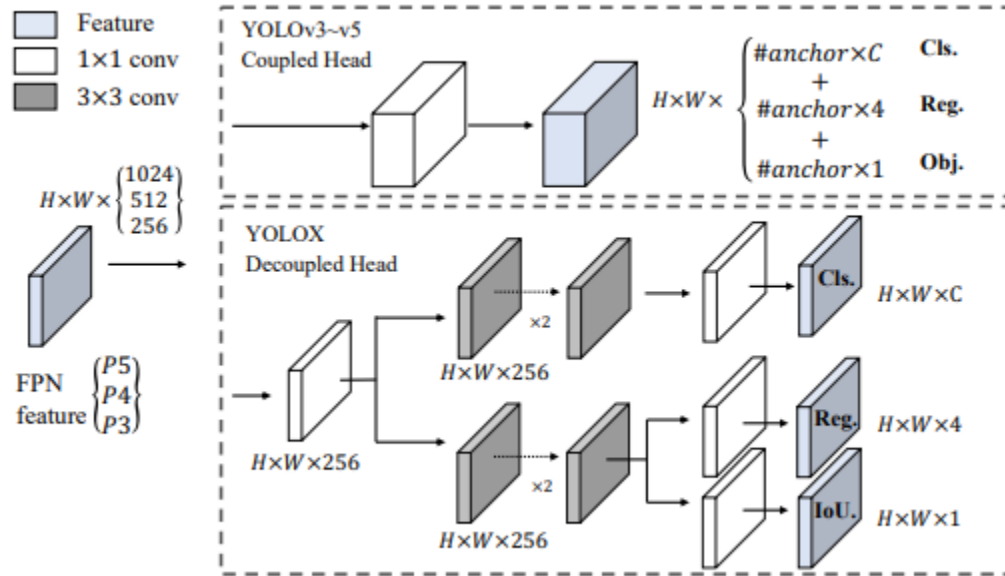


Figure 9 YOLOX network improvements over previous YOLO networks in terms of decoupling headers

The backbone of YOLOX-s is basically the same as that of YOLOv5-s, except that the activation function is changed to the SiLU function. That is, the Neck section is the same as YOLOv5-s, but with the activation function replaced by SiLU. The main improvement comes from changing YOLO head to Decoupled Head in Predication, since the "expressiveness" of YOLO head is poorer compared to that of Decoupled Head.

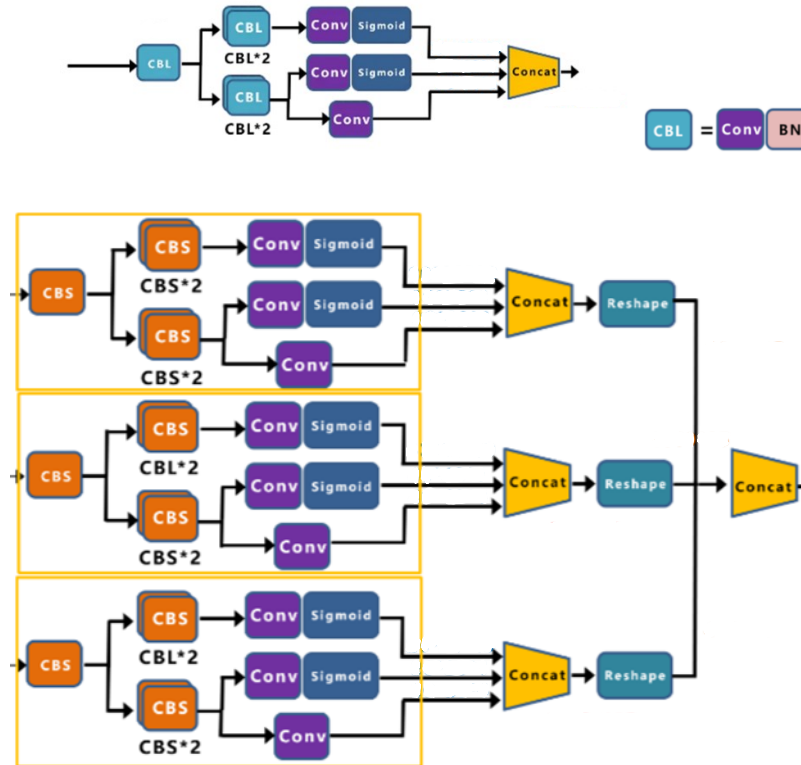


Figure 10 Design of the decoupled head section of the YOLOX network

In Fig.10, there are a total of three branches before the final Concat in the Decoupled Head structure). The first branch is used to categorize the target category in each frame and to predict each score, which is then convolved by a CBL of size 1x1 and two CBLs of size 3x3 and activated using the Sigmoid function. (Here CBL stands for Conv-BN-LeakyReLU, which is a common convolutional block structure. This structure consists of three parts: a convolutional layer (Conv), a batch normalization layer (BN, Batch Normalization), and a LeakyReLU activation function.) The second branch is used to determine whether the target frame is foreground or background, similar to the first branch. The last branch is to predict the coordinate information (x, y, w, h) of the target frame, which does not require Sigmoid function activation compared to the second branch. In the end these three branches are fused together by Concat. The network uses 3 Decoupled heads and outputs the results after the output Concat.

3.2 Wire and Arc image segmentation

3.2.1 U-Net Network

The left of the U-Net network [25] (Fig. 11) acts as feature extraction and the right as upsampling. Such a structure is also called Encoder-Decoder structure in some literature. This network is named U-Net because the overall structure of the network resembles the capital letter U. Each blue box (except for the first input) corresponds to a multichannel feature map, where the number of channels is labeled at the top of the box. $x - y$ magnitude is located at the bottom left corner of the box. The blue line boxes indicate the copied and cropped (Concat) feature maps. Arrows indicate the different operations. The network consists of a contracting path (contracting path) and an expanding path (expanding path). Shrinkage path (downsampling process) is used to obtain contextual information, extract the features of the feature map through multiple convolution operations in each layer to generate a new feature map and increase the number of features in the feature map, and reduce the size of the feature map through the max pool process to reduce the size of the feature map and connect each layer. When the feature map is small enough and the number of channels is large enough, the feature map is "expanded", i.e., the expansion path (up-sampling process) is used to locate the image accurately, which is similar to the structure of the contraction path, except that the max pool process is replaced by the inverse convolution process to expand the image and superimpose the feature maps corresponding to the contraction path layer in front of each layer. Therefore, the two paths are symmetrical to each other. The structure of the U-Net network makes it possible to use valid labeled data more efficiently from a very small number of training images. U-Net uses a completely different approach to feature fusion: concatenation, where U-Net uses features that are spliced together in the channel dimensions to form thicker features. FCN fusion, on the other hand, uses a summation of the corresponding points and does not form thicker features.

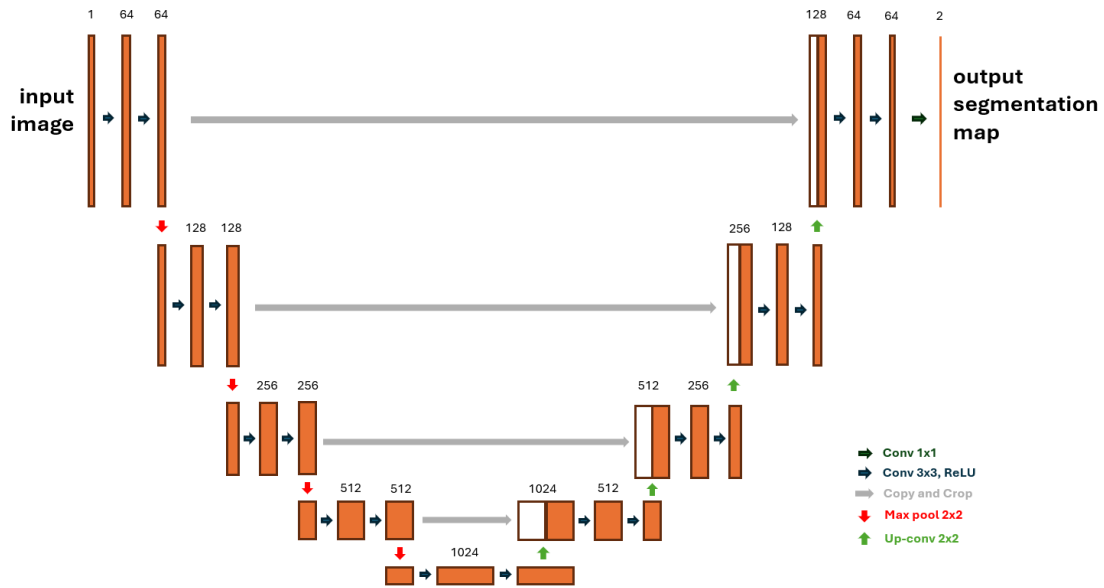


Figure 11 U-Net network structure[25]

In the weld pool image we obtained, there is information about the wire and arc which is also needed for us to combine with the extracted weld pool to completely depict the process. Since the features of the two are obvious after the image is preprocessed by our de-fogging and Reinteresting modules, the wire and arc can be easily segmented using a multi-objective 5-layer

U-Net network. Because of the obvious features and the code-encode structure of the U-Net network, the data annotation does not need to be very accurate, and the training dataset does not need to be very large to obtain good results. As such, we only use 118 sets of data for training and validation. After training 200 epochs, good results are achieved that the trained model can segment the wire with arc (Fig. 12). It is interesting to note that the segmentation results are even better than labeling which may be inaccurate as the result of a manual process.

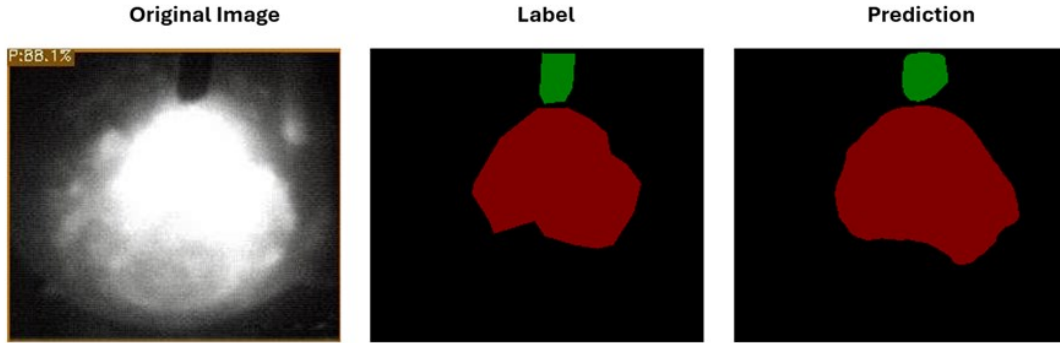


Figure 12 Weld Pool Data and Results

3.3 Segmentation of Weld Pool Images Based on Time Series

3.3.1 Convolution LSTM

As an improvement of the classical FC-LSTM network, Convolution LSTM (ConvLSTM) was proposed by Xingjian Shi et al [39]. First of all LSTM (Long Short-Term Memory) is a kind of RNN (Recurrent Neural Network). The LSTM structure consists of four parts: memory cells, input gates, output gates and forgetting gates (Fig. 13).

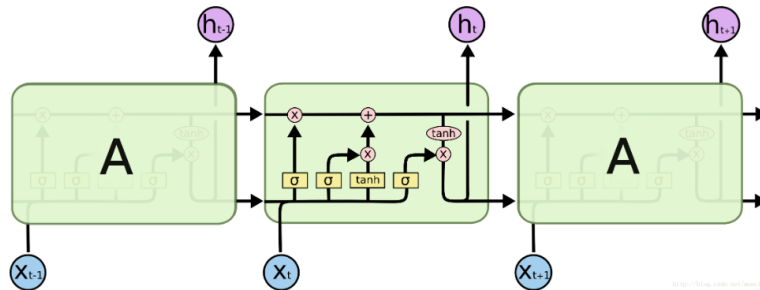


Figure 13 Classical LSTM network structure

The ConvLSTM model changes the fully-connect layer into a convolutional layer and replaces the input-to-state and state-to-state parts of FC-LSTM from feed-forward computation into the form of convolution. As a whole, the model structure of ConvLSTM is mainly divided into two parts: the temporal structure follows the typical RNN network structure; the spatial structure follows the CNN feature extraction method. To put it simply, the ConvLSTM model is equivalent to replacing all fully connected structures in LSTM with

convolutional structures, while adopting a structure based on peep connections. Its only change is in the way the individual gating units are computed as shown in Fig. 14.

$$f_t = \sigma([h_{t-1}, x_t, C_{t-1}] * W_f + b_f)$$

$$i_t = \sigma([h_{t-1}, x_t, C_{t-1}] * W_i + b_i)$$

$$\tilde{C}_t = \tanh([h_{t-1}, x_t] * W_c + b_c)$$

$$C_t = f_t \odot C_{t-1} \oplus i_t \odot \tilde{C}_t$$

$$o_t = \sigma([h_{t-1}, x_t, C_t] * W_o + b_o)$$

$$h_t = o_t \odot \tanh(C_t)$$

Figure 14 ConvLSTM network weight relations

3.3.2 U-Net fusion ConvLSTM

Since the welding process is an industrial process that has a close connection in the temporal relationship, it has a certain correlation in the time dimension. Yin et al [40] used a time-series based model for predicting changes in geographic lake boundaries by fusing LSTM models in a U-Net network. We introduce a special RNN network, LSTM network, into the U-Net classical decode-encode network to convey the temporal information. The LSTM-Cell is constructed to integrate it into the U-Net image segmentation network (Fig. 15).

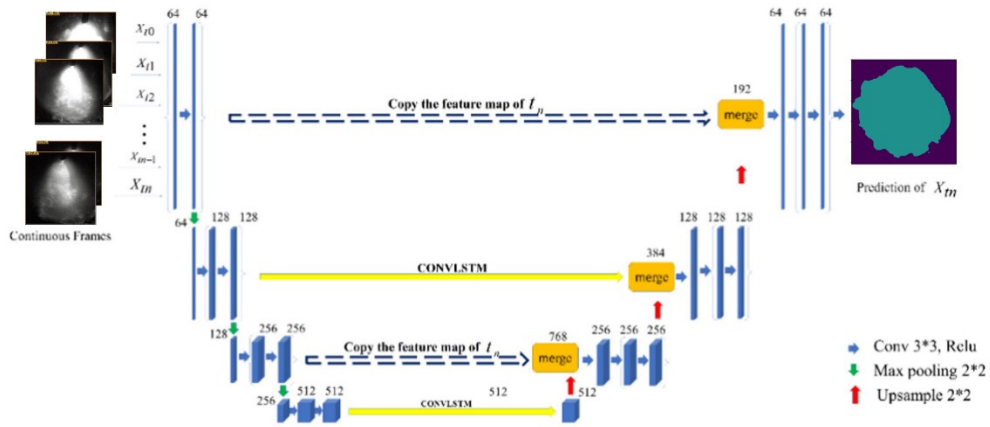


Figure 15 U-Net Fusion ConvLSTM Network Structure

As shown in Fig. 16, the encode module of the network is employed to capture the weld pool information from continuously fed images. Additionally, the ConvLSTM Cell is

integrated into the feature fusion process within each layer of both decode and encode operations. This Convolutional Long Short-Term Memory (ConvLSTM) network is a variant of the traditional LSTM network, utilizing convolution operations instead of matrix multiplication to retain spatial information within the image.

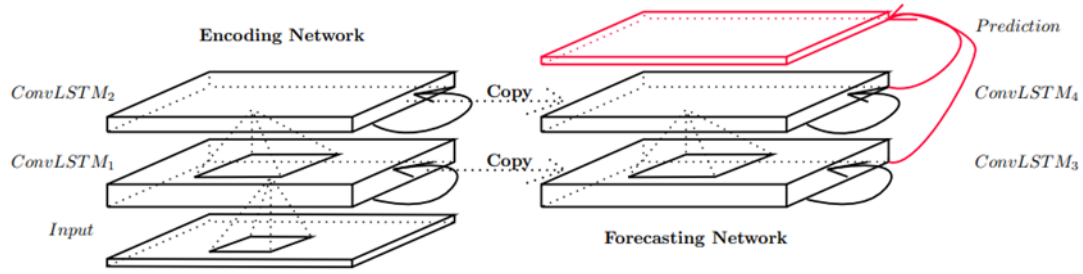


Figure 16 ConvLSTM Structure Fusing Temporal and Spatial Features

In ConvLSTM Cell the LSTM structure is used to preserve the temporal feature information of the image and the convolution operation is used to preserve the spatial features and used in the next frame (Fig. 17).

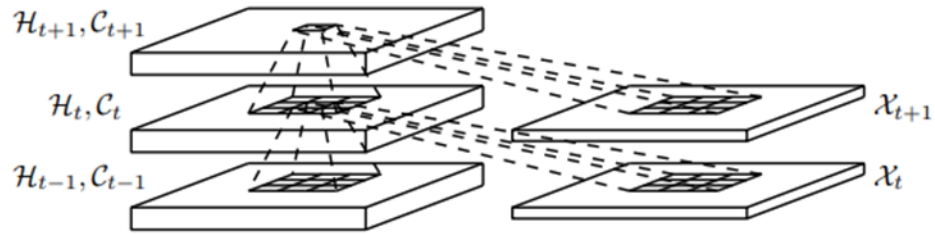


Figure 17 ConvLSTM Cell

At the same time the target features are obtained by downsampling and then upsampling is used to recover the information about the weld pool based on the obtained features. The input of the model is the weld pool image after Denoising and Reinteresting, and the output is the shape and size of the segmented weld pool (Fig. 18).

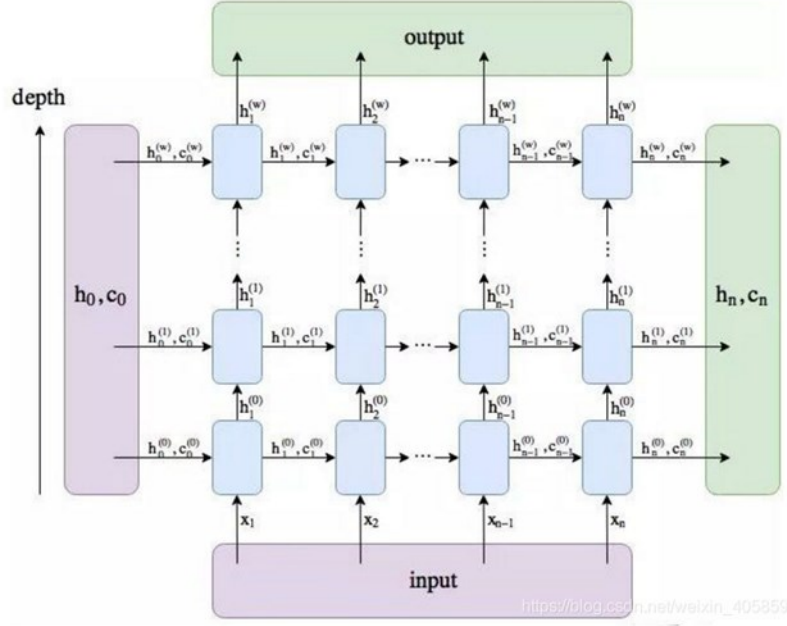


Figure 18 ConvLSTM Cell passes temporal and spatial hidden features in U-Net

In Fig. 18, input corresponds to the input feature maps after the convolution operation of each layer of downsampling, and output is the feature maps added to the corresponding downsampling layer before the convolution operation, and the blue box is the ConvLSTM-Cell, which can be set to more than one (i.e., the depth of the pass), in order to reduce the computational pressure, in the network we design, we currently use only one ConvLSTM-Cell. Its vertical direction is one time feature map passing, and the horizontal direction is time series feature passing. The passings rely on two functions, h and c .

4. Results and Discussion

4.1 Training

We used a computer with an Intel® Core™ i9-13900K CPU, 64GB of random access memory, and an NVIDIA GeForce RTX 4090 to train the model using the Pytorch library. In our model, AOD-Net for Defogging uses Adam optimization algorithm with an initial learning rate of 0.001, $\beta_1=0.9$, $\beta_2=0.999$ and weight decay set to be 1×10^{-8} . Loss function is MSE which is used to compute pixel differences between images before and after defogging. The SSIM structural similarity is used to measure the similarity in structural information between images before and after defogging. The Reinteresting Weld Pool section uses the SGD (Stochastic Gradient Descent) optimization algorithm, with the initial learning rate set to be 0.01, and the weight attenuation set to be 0.0005. The loss function uses IoU loss (Intersection over Union loss) to calculate the bounding box regression loss. The segmentation part of the U-Net network uses the Adam optimization algorithm. The initial learning rate is set to 0.001, the StepLR method is used

to decay, and the learning rate decays 0.1 per 7 steps. The loss function uses Cross-Entropy Loss. The evaluation indicator uses Accuracy.

4.1.1 Preprocessing Network Training

First Denoising Module in AOD-Net, we use ground-truth images with depth meta-data from the indoor NYU2 Depth Database [41], of which 27256 sets of data are used for training and 1449 sets of data are used for training validation. A total of 10 Epochs were trained. In Fig. 19 We can see that the training loss and the testing loss will drop quickly to around 0.02 at the beginning and then stay at that level. The final model is used for weld image defogging with good results.

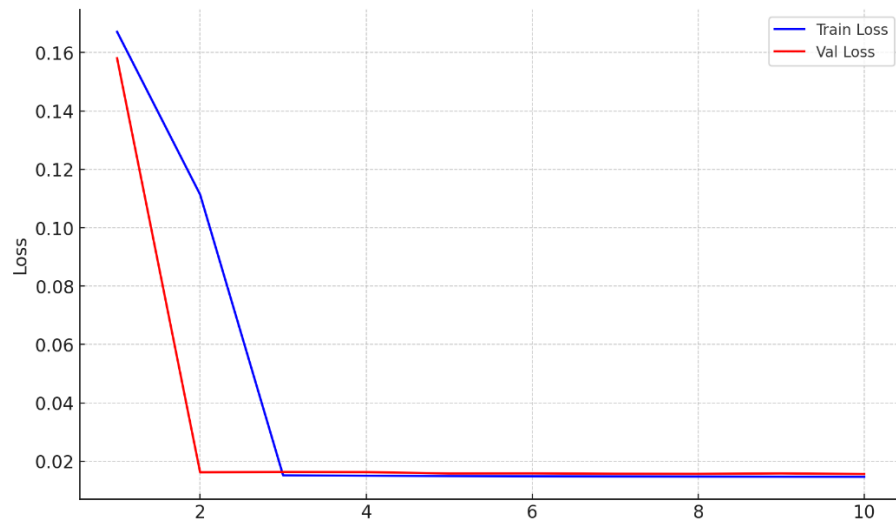


Figure 19 Denoising Module Training Loss

For YOLOX-s Reinteresting Weld Pool section, we use 911 sets of images labeled with the location of the weld pool as the dataset, set 80% of them as the training set and 20% as the validation set, the dataset is randomly sorted before being divided to ensure an even distribution of the samples, and a total of 300 Epochs are trained with the training loss shown in Fig.20. Since the weld pool is easily recognized in the image, our loss quickly drops from about 25% at the beginning to about 5%. The Loss is evaluated by the overlap of the predicted box with its manually labeled one. However, we just need the general location of the weld pool in the image. Hence, for our "Reinteresting" process the accuracy of the results reaches 99%.

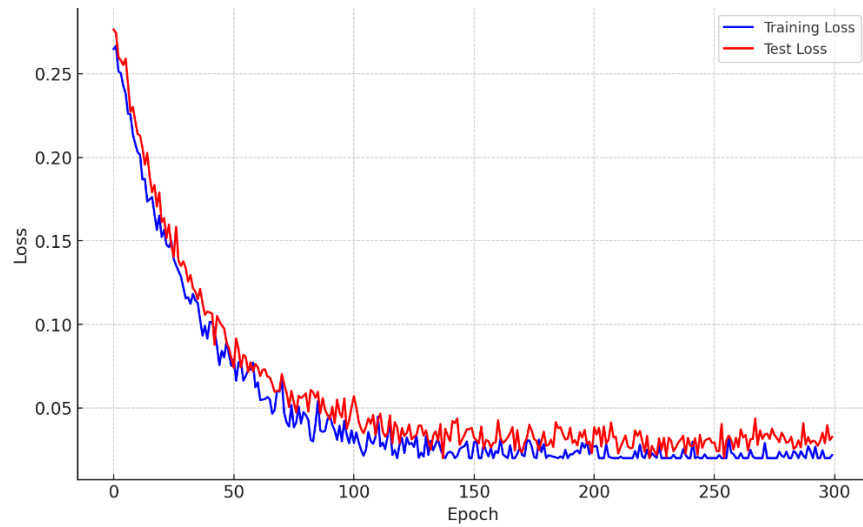


Figure 20 YOLOX-s Reinteresting Weld Pool

4.1.1 Segmentation process training

The image segmentation part is mainly Wire and Arc image segmentation using U-Net and segmentation of weld fusion pool using LSTM U-Net network. In segmentation of Wire and Arc we labeled 118 sets of data for the training dataset and another 20 sets of data for the validation set, for a total of 200 Epochs. We use Cross entropy to evaluate the loss In Fig.20. Cross entropy calculates the difference between a true probability distribution and a model-predicted probability distribution. If the model's prediction is very close to the true distribution, the cross entropy will be low, indicating that the model is performing well; if the model's prediction is very different from the true distribution, the cross entropy will be high, indicating that the model is not performing well. As the training progresses, the cross entropy of the model is decreasing. It indicates that the model is better for segmentation of Wire and Arc.

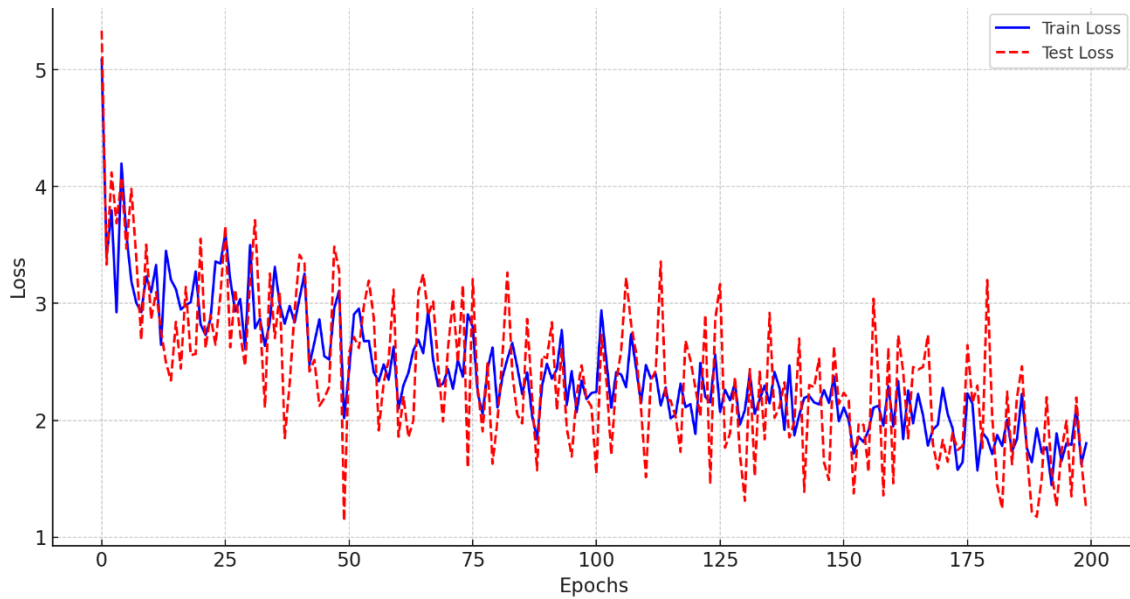


Figure 21 Wire and Arc Segmentation Training Loss

LSTM U-Net Model for Weld Pool Segmentation We labeled 336 sets of data for training and evaluation of the network, and since the network will segment based on temporal features of the process, we additionally chose 77 sets of data for training and evaluation only to verify that the network does not overfit a single process. We chose to use loss to evaluate the training process. In Fig.21 the training set evaluation Loss eventually drops below 1%, and the test set evaluation Loss is only slightly higher than the training set, but also eventually drops below 1%. This proves that the model has superior performance in the weld pool segmentation task.

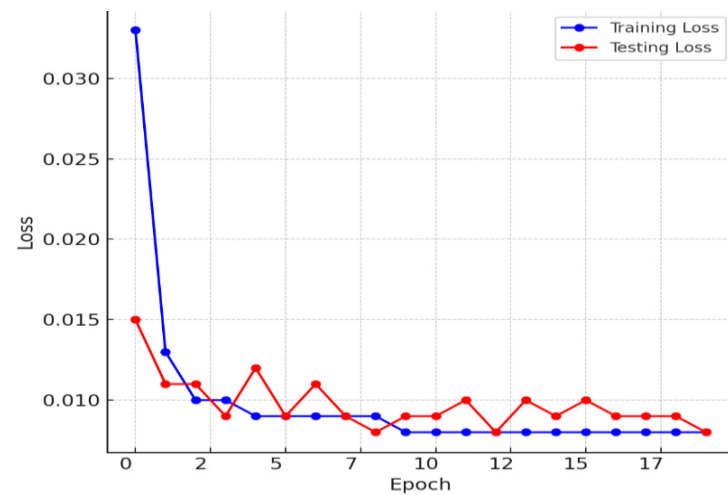


Figure 22 Segmentation of Weld Pool

4.2 Results

After the training, we obtained the model needed for each processing module, and the final results, after inputting the welding process images into each module according to the flow order, are shown in Fig. 22 as the four demonstration results (a), (b), (c), and (d), which come from the welding process under different conditions, where the red line indicates the boundaries of the weld molten pool, the blue line is the boundaries of the Arc, and the green line is the boundaries of the Wire. It can be seen that for all images captured from the various welding conditions/states, the method succeeds in dynamically obtaining the Weld Pool, Arc, and Wire boundaries simultaneously, and the method can process the images at more than 10 frames per second. This speed is sufficient to monitor the GMAW process.

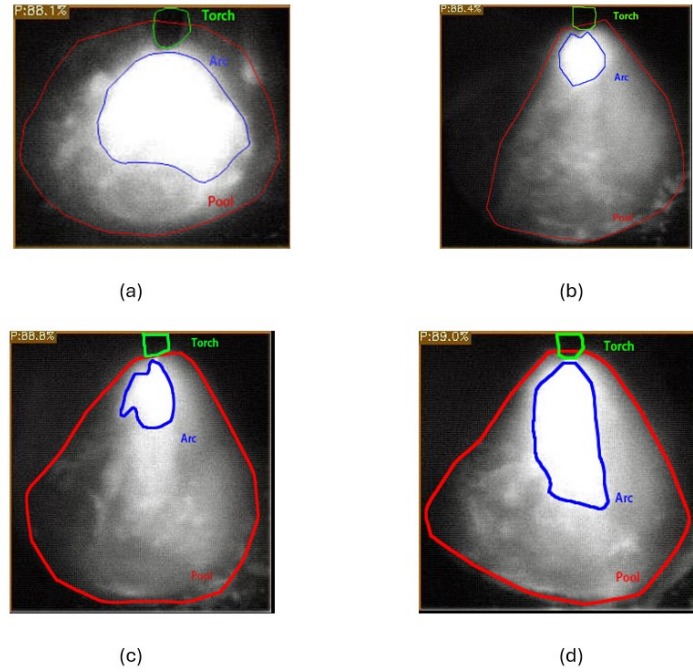


Figure 23 The final output of the Welding Process Melt Pool Critical Information Detection Model result information.

In the image preprocessing part, AOD-Net Defogging uses SSIM values for evaluation. The SSIM value ranges from "0" to "1". The closer to "1", the more similar the defogging image is to the real image. The change of SSIM value with Epoch is shown in Fig. 24. The SSIM value in the training set tends to 0.95, and the SSIM value in the test set tends to 0.93. Judging from the evaluation parameter data, the effect is good.

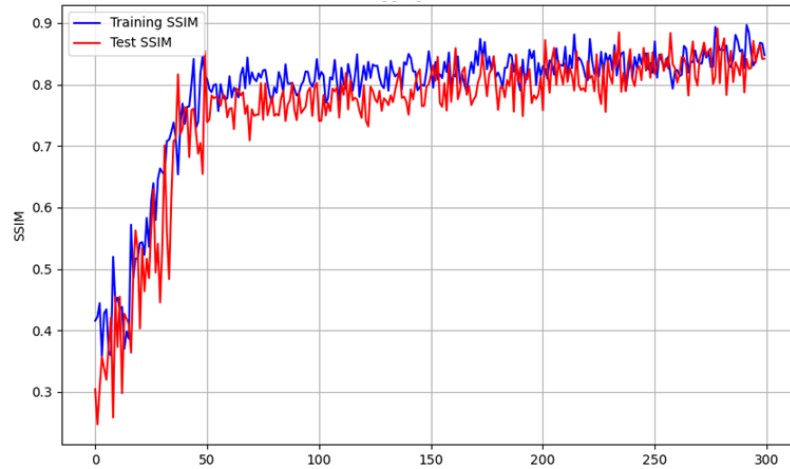


Figure 24 Defogging SSIM Value

YOLOX-s Reinteresting Weld Pool is evaluated using mAP values, and the variation of mAP values with Epoch is shown in Fig. 25. mAP values converge to 0.95 in the training set, and the test set is around 0.9. It indicates that the model works well.

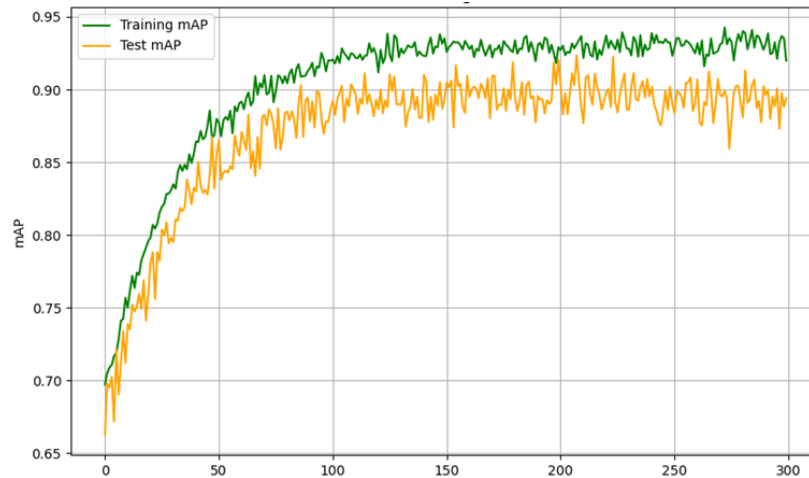


Figure 25 Reinteresting Weld Pool mAP

In the Wire and Arc image segmentation section we use U-Net for segmentation and use the Accuracy value as an evaluation. The variation of Accuracy value with Epoch is shown in Fig. 26. The Accuracy value of the training set tends to be close to 0.7 and the Accuracy of the test set also tends to be close to 0.7, but with a large fluctuation.

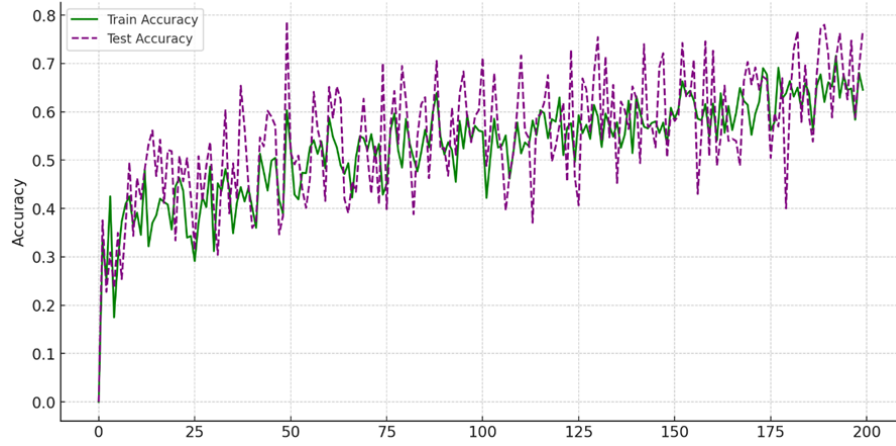


Figure 26 Wire and Arc image segmentation Accuracy

We also use the Accuracy value as the evaluation criterion in Segmentation of Weld Pool Images Based on Time Series. The change of Accuracy value with Epoch is shown in Fig. 27. The Accuracy value of the training set is above 0.8 and changes little. The Accuracy value of the test set is also above 0.8 and tends to be stable.

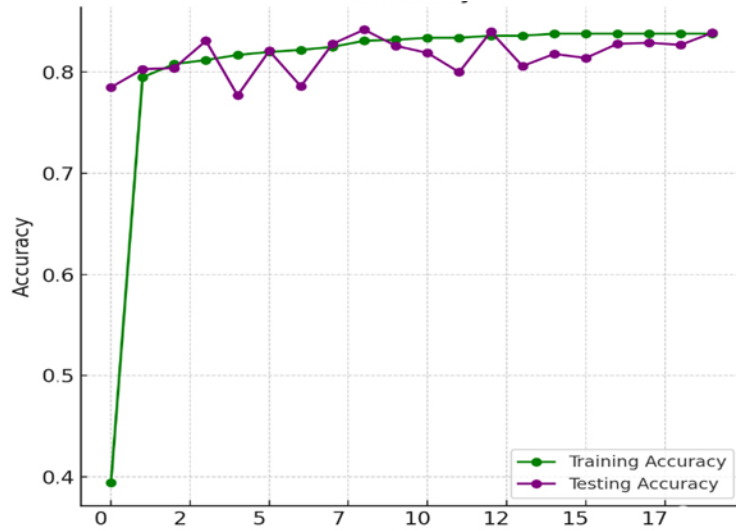


Figure 27 Segmentation of Weld Pool Images Based on Time Series Accuracy

4.3 Module influence

The modules in the model are combined separately to prove that each module in our proposed model has an effective influence on the final result. First, the result of using only the U-Net model for weld pool segmentation (Fig. 28) shows that the weld pool segmentation is incomplete, and the judgment of using spatter as the boundary segmentation of the weld pool is incorrect.

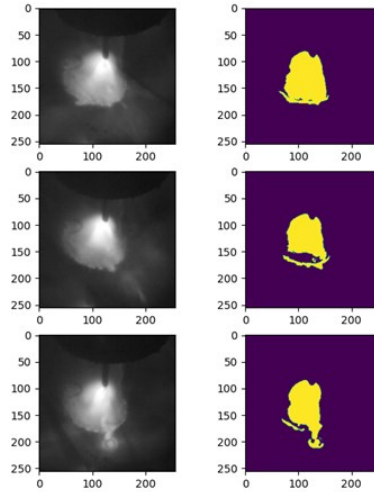


Figure 28 U-Net Segmentation Results

After using the U-Net model fused with LSTM, the segmentation can be almost complete, but due to the influence of the smoke generated during welding, the darker weld pool at the edge cannot be identified, as shown in Fig. 29.

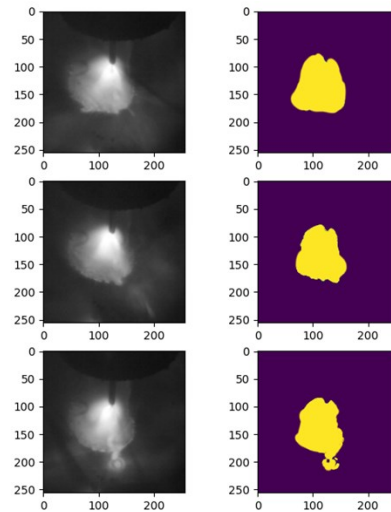


Figure 29 LSTM U-Net Segmentation Results

Then we added the AOD-Net Defogging module to the U-Net model fused with LSTM to remove the effects of welding smoke. As shown in Fig. 30, the weld pool with difficult edges is successfully segmented. However, there is still molten metal splashes that affect the segmentation of the weld pool.

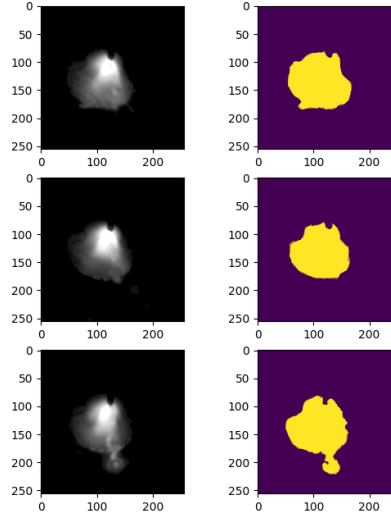


Figure 30 LSTM U-Net & AOD-Net Segmentation Results

Finally, YOLOX-s Reinteresting Weld Pool module was added to remove the influence of molten metal splashes on weld pool segmentation. As shown in Fig. 31, the final result is good.

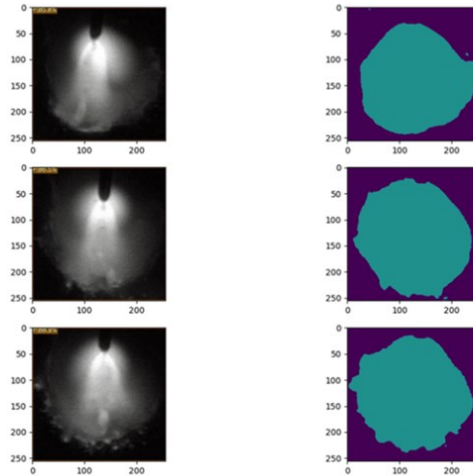


Figure 31 LSTM U-Net & AOD-Net, YOLOX-s Segmentation Results

The average SSIM value of the three images of each model is obtained and calculated, as shown in Table 2. It can be seen that the SSIM value of segmentation using only the U-Net network is the lowest, only 0.693. Then, as LSTM Cell, defogging module and Reinteresting Weld Pool module are added to the model, the SSIM value reaches 0.901, which shows that the modules designed in our network are very important for this segmentation task.

Table 2 Average SSIM Values

Model	Average SSIM value
U-Net	0.693
LSTM U-Net	0.795
LSTM U-Net & AOD-Net	0.831
LSTM U-Net & AOD-Net, YOLOX-s	0.901

5. Conclusion and Future Work

Due to challenges in acquiring clear weld images and accurately labeling data in GMAW, there are hurdles in directly applying state-of-the-art deep learning models to reconstruct its complete weld pool scenes. Effect reconstruction requires understanding of their characteristics so that appropriate deep learning networks can be designed and chosen accordingly despite the constraints in the image quality and label size. To this end, we analyzed the constituents in the weld pool scene and their dynamic behaviors, characteristics and inter-correlations. This empowered us to propose an innovative procedure to sequentially extract the constituents in the scene in an effective order that helps maximally migrate the effect from the inter-correlations. For respective constituents, we chose effective deep learning networks accordingly. They include U-Net that excels in accurate target segmentation with small datasets and LSTM network that can effectively utilize the strong temporal relationship in welding processes. They also include methods to effectively pre-process (denoise and reinterest) the image to migrate the effect of smoke and spatters on the subsequential recognition of the unclear weld pool boundary and dynamically fluctuating boundary of the arc during GMAW. As such, we designed an effective combinatorial deep learning networks to overcome various challenges from the physical process and data limitation and successfully reconstruct the complex weld pool scene in the highly dynamic gas metal arc welding process. The model we proposed has achieved good results in the welding system we built. Experiments show that each module designed in the model has a significant effect on the extraction of key information of the weld pool.

Future work includes analyzing the role of the extracted key information in regulating the welding process, and then combining it with automated welding production equipment to apply advanced control algorithms to meet the challenges of complex weldments and high-precision welding tasks.

Acknowledgement: This work is partially funded by the National Science Foundation under grant CMMI-2024614 and the University of Kentucky Institute for Sustainable Manufacturing.

References

- [1] Y. Liu, W. Zhang and Y. Zhang, "Dynamic Neuro-Fuzzy-Based Human Intelligence Modeling and Control in GTAW," in *IEEE Transactions on Automation Science and Engineering*, vol. 12, no. 1, pp. 324-335, Jan. 2015, doi: 10.1109/TASE.2013.2279157.
- [2] Saeed, G.; Lou, M.J.; Zhang, Y.M. Computation of 3D weld pool surface from the slope field and point tracking of laser beams. *Meas. Sci. Technol.* 2004, 15, 389–403.
- [3] Sun, M., Yang, M., Wang, B., Qian, L., & Hong, Y. (2021, August). Applications of molten pool visual sensing and machine learning in welding quality monitoring. In *Journal of Physics: Conference Series* (Vol. 2002, No. 1, p. 012016). IOP Publishing.
- [4] Richardson, R., and Gutow, D., 1983, "Coaxial Arc Weld Pool Viewing for Process Monitoring and Control," *Weld. J.*, 63(3), pp. 43–50.
- [5] Węglowski, M. S., 2007, "Investigation on the Arc Light Spectrum in GTA Welding," *J. Achiev. Mater. Manuf. Eng.*, 20, p. 519–522.
- [6] Li, P. J., and Zhang, Y. M., 2000, "Analysis of an Arc Light Mechanism and Its Application in Sensing of the GTAW Process," *Weld. J.*, 79(9), pp. 2525-s.
- [7] Wang, Z., Zhang, C., Pan, Z., Wang, Z., Liu, L., Qi, X., ... & Pan, J. (2018). Image segmentation approaches for weld pool monitoring during robotic arc welding. *Applied Sciences*, 8(12), 2445.
- [8] Kiddee, P., Fang, Z., and Tan, M., 2016, "An Automated Weld Seam Tracking System for Thick Plate Using Cross Mark Structured Light," *Int. J. Adv. Manuf. Technol.*, 87(9–12), pp. 3589–3603.
- [9] Xu, P., Xu, G., Tang, X., and Yao, S., 2008, "A Visual Seam Tracking System for Robotic Arc Welding," *Int. J. Adv. Manuf. Technol.*, 37(1–2), pp. 70–75.
- [10] Kovacevic, R., Zhang, Y. M., and Ruan, S. (May 1, 1995). "Sensing and Control of Weld Pool Geometry for Automated GTA Welding." *ASME. J. Eng. Ind.* May 1995; 117(2): 210–222. <https://doi.org/10.1115/1.2803297>.
- [11] Zhang, Y. M., Song, H. S., and Saeed, G., 2006, "Observation of a Dynamic Specular Weld Pool Surface," *Meas. Sci. Technol.*, 17(6), pp. L9–L12.
- [12] Zhang, W., Wang, X., and Zhang, Y., 2013, "Analytical Real-Time Measurement of a Three-Dimensional Weld Pool Surface," *Meas. Sci. Technol.*, 24(11), p. 115011.
- [13] Cui, Yan-Xin, et al. "Investigation into keyhole-weld pool dynamic behaviors based on HDR vision sensing of real-time K-TIG welding process through a steel/glass sandwich." *Advances in Manufacturing* 9 (2021): 136-144.
- [14] Cheng, Y., Yu, R., Zhou, Q., Chen, H., Yuan, W. and Zhang, Y., 2021. Real-time sensing of gas metal arc welding process—A literature review and analysis. *Journal of Manufacturing Processes*, 70, pp.452-469.
- [15] Rother, C.; Minka, T.P.; Blake, A.; Kolmogorov, V. Cosegmentation of Image Pairs by Histogram Matching—Incorporating a Global Constraint into MRFs. In *Proceedings of the IEEE Computer*

Society Conference on Computer Vision and Pattern Recognition (CVPR), New York, NY, USA, 17–22 June 2006; pp. 993–1000.

[16] Vicente, S.; Kolmogorov, V.; Rother, C. Cosegmentation Revisited: Models and Optimization. Lecture Notes in Computer Science. In Proceedings of the Computer Vision (ECCV), Crete, Greece, 5–11 September 2010; pp. 465–479.

[17] Mukherjee, L.; Singh, V.; Dyer, C.R. Half-integrality-based Algorithms for Cosegmentation of Images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 2028–2035.

[18] Hochbaum, D.S.; Singh, V. An Efficient Algorithm for Co-segmentation. In Proceedings of the 12th IEEE International Conference on Computer Vision (ICCV), Kyoto, Japan, 29 September–2 October 2009; pp. 269–276.

[19] Alexe, B.; Deselaers, T.; Ferrari, V. What Is an Object? In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 73–80.

[20] Vicente, S.; Rother, C.; Kolmogorov, V. Object cosegmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011.

[21] Lecun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based Learning Applied to Document Recognition. *Proc. IEEE* 1998, 86, 2278–2324.

[22] Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* 2017, 60, 84–90.

[23] Karen, S.; Andrew, Z. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv* 2014, arXiv:1409.1556.

[24] Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. *IEEE Trans. Pattern Anal. Mach. Intell* 2017, 39, 640–651.

[25] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III* 18. Springer International Publishing, 2015.

[26] Baek, D., Moon, H. S., & Park, S. H. (2024). In-process prediction of weld penetration depth using machine learning-based molten pool extraction technique in tungsten arc welding. *Journal of Intelligent Manufacturing*, 35(1), 129-145.

[27] R. Yu, J. Kershaw, P. Wang and Y. Zhang, "Monitoring of Backside Weld Bead Width from High Dynamic Range Images Using CNN Network," 2022 8th International Conference on Control, Decision and Information Technologies (CoDIT), Istanbul, Turkey, 2022, pp. 39-44, doi: 10.1109/CoDIT55151.2022.9804041.

[28] Wang, Y., Han, J., Lu, J., Bai, L., & Zhao, Z. (2020). TIG stainless steel molten pool contour detection and weld width prediction based on Res-Seg. *Metals*, 10(11), 1495.

- [29] Kim, Daeun Dana, et al. "Generating pedestrian training dataset using DCGAN." Proceedings of the 2019 3rd International Conference on Advances in Image Processing. 2019.
- [30] Bae, K. Y., Lee, T. H., & Ahn, K. C. (2002). An optical sensing system for seam tracking and weld pool control in gas metal arc welding of steel pipe. *Journal of Materials Processing Technology*, 120(1-3), 458-465.
- [31] Arbelle, Assaf, and Tammy Riklin Raviv. "Microscopy cell segmentation via convolutional LSTM networks." 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019.
- [32] Li, B., Peng, X., Wang, Z., Xu, J., & Feng, D. (2017). An all-in-one network for dehazing and beyond. arXiv preprint arXiv:1707.06543.
- [33] Ge, Z., Liu, S., Wang, F., Li, Z., & Sun, J. (2021). YOLOX: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430.
- [34] Wu Jun, Zou Zeng-Da, WANG Xin-Hong, LI Qing-Ming. Effect of SiO₂ activator on the arc phenomenon of tungsten arc welding of stainless steel. *Journal of Welding*, 2007. Available at: <http://www.cqvip.com/qk/94943x/200704/24317575.html>
- [35] Hua Xuming, Wang Fei, Ma Xiaoli, Wu Yixiong. Analysis of arc interruption phenomenon during three-wire GMAW welding. *Journal of Shanghai Jiao Tong University*, 2010. Available at: <http://www.cqvip.com/qk/92944x/201004/33867447.html>
- [36] Fang Chen-Fu, Chen Shujun, LIU Jia, YIN Shu-Yin, SONG Yong-Lun, LI Huan, HOU Run-Shi. Comparative analysis of SACTIG and VPTIG arc stability (I). *Journal of Welding*, 2005. Available at: <http://www.cqvip.com/qk/94943x/200512/20953448.html>
- [37] Zhang Bo. Analysis of magnetic bias blowing phenomenon in pipeline welding and its solution. *Chemical Management*, 2013. Available at: <http://www.cqvip.com/qk/80982x/201310/46043352.html>
- [38] Zhu, Y.-G. Overcoming and eliminating the effect of magnetic bias blowing on the welding arc. *Petrochemical Construction*, 2007. Available at: <http://www.cqvip.com/qk/95372b/200703/24953349.html>
- [39] Shi, Xingjian, et al. "Convolutional LSTM network: A machine learning approach for precipitation nowcasting." *Advances in neural information processing systems* 28 (2015).
- [40] Yin, Lirong, et al. "U-Net-LSTM: time series-enhanced lake boundary prediction model." *Land* 12.10 (2023): 1859.
- [41] Silberman, Nathan, et al. "Indoor segmentation and support inference from rgb-d images." *Computer Vision—ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part V* 12. Springer Berlin Heidelberg, 2012.