

Article

Real-Time Recognition of Molten Pools Based on Improved DeepLabV3+ in Keyhole Tungsten Inert Gas Welding Applications

Dayong Yang , Penghui Dai, Shuwan Cui *, Huzhe Song, Feiyang Liu and Xuan Zhou

School of Mechanical and Automotive Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China; dyyang@gxust.edu.cn (D.Y.); 221068134@stdmail.gxust.edu.cn (P.D.); 221068240@stdmail.gxust.edu.cn (H.S.); 221076902@stdmail.gxust.edu.cn (F.L.); 221076958@stdmail.gxust.edu.cn (X.Z.)

* Correspondence: swcui@gxust.edu.cn; Tel.: +86-13597066615

Abstract: During the Keyhole Tungsten Inert Gas (K-TIG) welding process, a significant amount of information related to the weld quality can be obtained from the weld pool and the keyhole of the topside molten pool image, which provides a vital basis for the control of welding quality. However, the topside molten pool image has the unstable characteristic of strong arc light, which leads to difficulty in contour extraction. The existing image segmentation algorithms cannot satisfy the requirements for accuracy, timing, and robustness. Aiming at these problems, a real-time recognition method, based on improved DeepLabV3+, for identifying the molten pool more accurately and effectively was proposed in this paper. First, MobileNetV2 was selected as the feature extraction network with which to improve detection efficiency. Then, the atrous rates of atrous convolution layers were optimized to reduce the receptive field and balance the sensitivity of the model to molten pools of different scales. Finally, the convolutional block attention module (CBAM) was introduced to improve the segmentation accuracy of the model. The experimental results verified that the proposed model had a fast segmentation speed and higher segmentation accuracy, with an average intersection ratio of 89.89% and an inference speed of 103 frames per second. Furthermore, the trained model was deployed in a real-time system and achieved a real-time performance of up to 28 frames per second, thus meeting the real-time and accuracy requirements of the K-TIG molten pool monitoring system.



Citation: Yang, D.; Dai, P.; Cui, S.; Song, H.; Liu, F.; Zhou, X. Real-Time Recognition of Molten Pools Based on Improved DeepLabV3+ in Keyhole Tungsten Inert Gas Welding

Applications. *Electronics* **2024**, *13*, 283.

<https://doi.org/10.3390/electronics13020283>

Received: 14 December 2023

Revised: 24 December 2023

Accepted: 25 December 2023

Published: 8 January 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: K-TIG welding; molten pool; DeepLabV3+ model; MobileNetV2; attention module

1. Introduction

As a welding method with high energy input, K-TIG welding plays a crucial role in the field of medium–thick plate welding. Medium–thick material plates, like titanium alloy, duplex stainless steel, and low carbon steel, can perform a single-sided weld and a double-sided form in a single pass without opening grooves [1–3]. It is often combined with robots for automatic welding in practical industrial production, but it is still difficult to achieve precise real-time regulation during the complex welding process. To improve the intelligent automation level of welding technology, scholars have paid much attention to monitoring the welding process [4].

Welding process monitoring is a key component in intelligent welding [5]. During the welding process, the molten pool contains a wealth of welding quality information. By observing the shapes and characteristics of the molten pool, skilled welders can adjust the welding parameters in real time to maintain good welding quality. However, the shape of the molten pool is dynamic, and it is a very subjective task to monitor its formation manually. To realize the closed-loop control of intelligent welding, visual sensing technology can be used to simulate the welders in monitoring the morphological change in the molten pool and obtain its shape parameters, which are used as indirect control variables fed back

to the control system so that the welding parameters can be adjusted to greatly improve welding quality and efficiency. Visual sensing technology is considered to be the most promising welding sensing technology due to its advantages of convenient operation, strong anti-interference, mass information, high-speed acquisition, and non-contact with the workpiece.

In the visual monitoring of K-TIG welding, image processing is an indispensable step through which to obtain the surface morphology of the weld pool and keyhole clearly and accurately. It is typically required to go through the following four stages in image processing: the selection of the region of interest (ROI), the preprocessing of the image, the segmentation of the image, and feature extraction [6–8]. Its most crucial and challenging part is image segmentation, which can be used to segment the captured objects, like the weld pool, keyhole, weld line, arc, and the like [9]. At present, image segmentation methods for molten pools can be divided into two categories, including methods based on traditional image algorithms and those based on active contour models [10]. Liu et al. [11] analyzed the results of six different edge detection operators and finalized the Canny operator (with a threshold between 0.1 and 0.5) to obtain ideal edge features for a molten pool. Chen et al. [12] utilized a traditional image algorithm to extract the molten pool's width, trailing length, and surface height, which were used as the input signals of the data-driven model to predict the backside width. Liu et al. [13] proposed an active contour model to extract the edges of the groove accurately, and the accuracy of the model was proven by the experiment results. These methods have the obvious advantages of low computational complexity and high efficiency, which make it possible to obtain stable and accurate segmentation results under certain tasks. However, during the K-TIG welding process, the size, shape, and position of the weld pool and keyhole are affected by different welding process parameters, which leads to difficulty in the segmentation of the molten pool image within the actual welding environment [14]. Furthermore, the segmentation methods mentioned above have the problem of poor segmentation performance due to the interference from the complex reflection characteristics of the molten pool's surface, the intense light intensity of the arc light, and the other noise sources in the environment.

With the rapid development of deep learning, image semantic segmentations based on the convolutional neural network (CNN) have made significant progress, and various semantic segmentation models have also appeared, one after another. Because of its two characteristics of local connection and weight sharing, the CNN can make the network deeper and obtain higher accuracy in complex tasks by greatly reducing the parameters of the network model. The fully convolutional network (FCN), proposed in 2015 [15], improved based on CNN. The FCN is a seminal contribution to deep learning for semantic segmentation, which establishes a general network model framework. It leverages the feature map generated by the final convolutional layer in the CNN to perform up-sampling, and it classifies and predicts each pixel throughout this procedure. However, the network of the FCN is relatively large, it is not sensitive to the details of the image, and the target boundary is blurred due to the low correlation between pixels. To improve the accuracy of segmentation results, many researchers have tried a variety of methods. Peng et al. [16] proposed an improved global convolutional network method, which uses large separable filters to reduce the number of parameters and optimize the segmentation boundary. PSPNet [17], developed by Zhao, extracts appropriate global features for scene parsing and semantic segmentation tasks. It uses a pyramid pooling module to combine local and global information and proposes an optimization strategy with moderate supervision loss. The DeepLab model [18–21] is a deep convolutional neural network (DCNN) model proposed by Chen et al. Its core concept is to use atrous convolution, which can not only clearly control the resolution of the response when calculating the feature response but also expand the receptive field of the convolutional kernel, as well as integrate more feature information without increasing the number of parameters and calculations.

In recent years, some researchers have attempted to apply deep learning models to the monitoring of the K-TIG welding process. Xia et al. [22] designed a novel visual monitoring system for K-TIG welding, employing an XVC-1000e camera to observe the weld pool and keyhole. The residual network (ResNet) was developed for welding state classification based on the molten pool image, achieving an impressive recognition rate of 98% in experiments. To guarantee a good welding seam tracking effect, Chen et al. [23] proposed an image processing algorithm using Mask-RCNN to accurately identify the keyhole entrance and weld centerline. The accuracy of the algorithm was verified by experiments, and the inference speed was 1.53 frames per second. Wang et al. [24] introduced a segmentation-LSTM model, segmenting the regions of the weld pool and keyhole. Deployed in an embedded system, this model achieved a running time of 50 ms. The above research results show that although these models have shown excellent performance, their model structures were too large and complex, resulting in slow processing speeds and high memory usage. Moreover, challenges persisted in accurately segmenting the weld pool and keyhole areas, with room for improvement in the accuracy and stability of the weld pool and keyhole edge.

To address these issues, an improved DeepLabV3+ model is put forward, aiming to boost processing speed and improve the accuracy and robustness of image segmentation. A visual sensing system with a high-dynamic-range (HDR) camera was designed and implemented to capture weld pool and keyhole images simultaneously, reducing interference from strong arc light. MobileNetV2 serves as the backbone network where its atrous rates are adjusted to diminish model parameters and computational complexity to classify each pixel in the molten pool image and extract the information of the weld pool and keyhole quickly and efficiently from the image. The proposed model lays a solid foundation for the practical application of proven deep learning techniques in the K-TIG welding process control.

2. System and Experiments

2.1. Experiment System

As shown in Figure 1, the experimental system is composed of four subsystems: the K-TIG welding system, robot system, visual sensing system, and control system. The K-TIG welding system includes a welding power source (model HTIG-P10), a water cooler, a specialized welding torch equipped with a 6.4 mm diameter tungsten electrode, and pure argon with a purity of 99.99% serving as the shielding gas for K-TIG experiments. The robot system, which mainly consists of a six-DOF KUKA robot (model KR16), controller cabinet, and smart PAD, is responsible for the movement of the welding torch. The HDR camera of the visual sensing system is utilized to observe the topside molten pool, forming an approximately 60-degree angle with the workpiece to aim at the keyhole entrance and welding pool area. The systems considered above are all controlled by an industrial computer within the control system.

The HDR camera and welding torch are installed together at the end of the KUKA robot through a clamping device. The workpieces to be welded are fixed on the workbench before welding, with the robot moving forward at the welding speed while the welding torch and camera remain stationary. The HDR camera was chosen for its capability to capture high-dynamic-range visual information, including weld pools, keyholes, and arcs. Moreover, it effectively suppresses the impact of welding arc light on the front end of a molten pool. The parameters of the HDR camera are listed in Table 1.

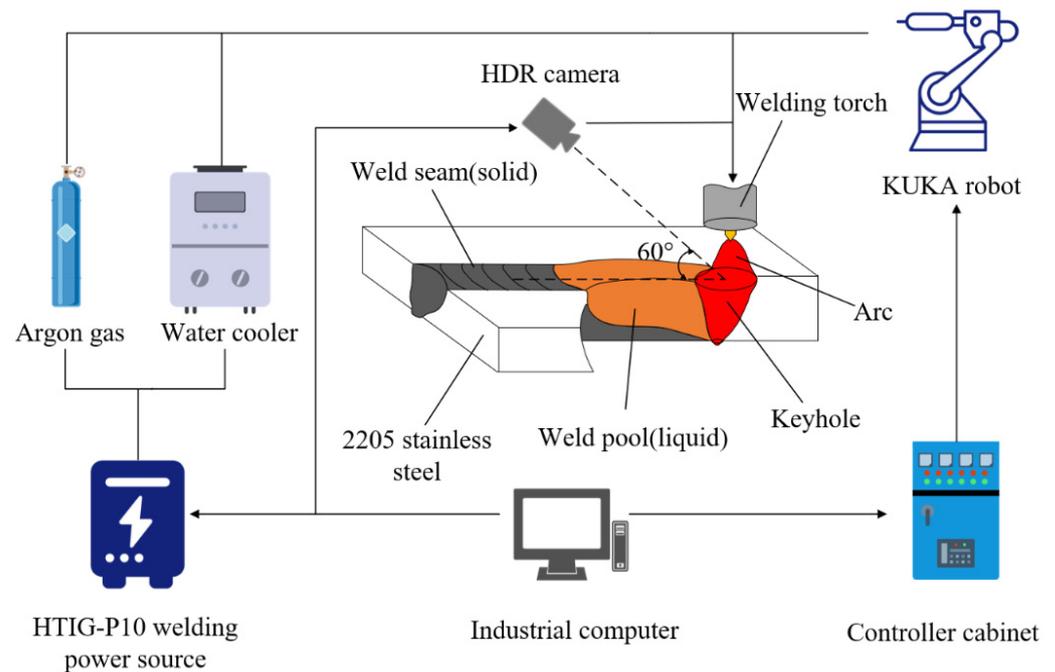


Figure 1. The schematic diagram of the K-TIG welding experimental platform.

Table 1. The parameters of the HDR camera.

Parameter	Value
Model	LITS-FS3H
Resolution	1920 (H) × 1080 (V)
Pixel size	3.75 μm (H) × 3.75 μm (V)
Dynamic range	120 dB
Frame rate	30 fps
Exposure mode	Rolling Shutter

2.2. Experimental Procedure

In this paper, experiments were performed on 300 mm × 100 mm × 8 mm 2205 stainless steel as the experimental object due to its wide application in industry. The butt welding was adopted, with no obvious wrong edge or gap between the workpieces. The tungsten tip was 3 mm away from the surface of the workpieces.

Figure 2 shows the flowchart of the experiment procedure. When started, signals were output from the industrial computer to the welding power source and controller cabinet, controlling the weld current and movement speed. Under the action of the heat source, the workpieces began to melt from the surface into the interior, forming the weld pool and keyhole. Simultaneously, as the robot started to move, the HDR camera was triggered to monitor and capture images of weld pools as well as keyholes. These images were then transmitted to the industrial computer for use as inputs in the segmentation model. During the training process of the model, the areas of the weld pools and keyholes were manually labeled in the captured images, as shown in Figure 3. The trained segmentation model was subsequently utilized to recognize the actual edges of the weld pool and keyhole.

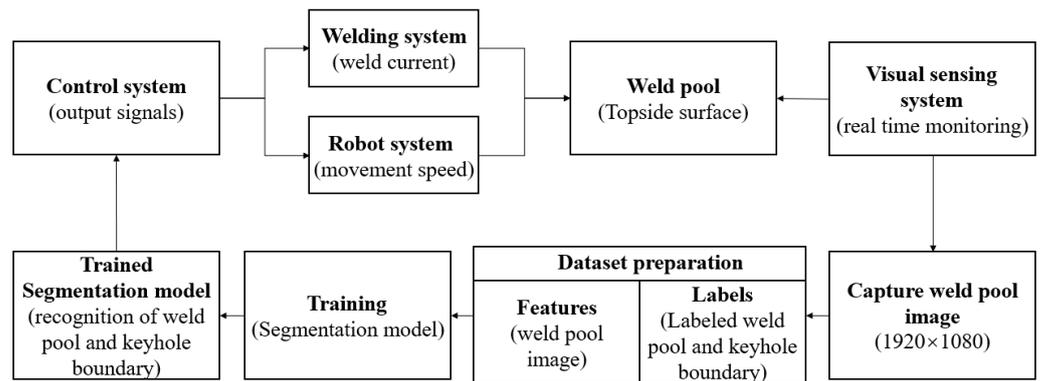


Figure 2. The flowchart of the experiment procedure.

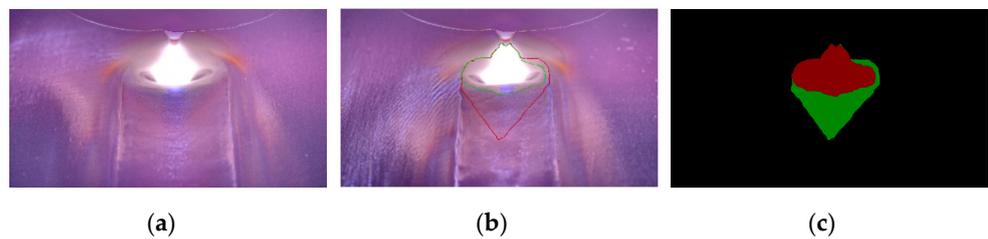


Figure 3. Dataset preparation for training segmentation model. (a) Input image; (b) Labeled image; (c) One-hot encoding.

2.3. Data Collection

Given the absence of public molten pool datasets, this experiment used an HDR camera to capture the molten pool images. The shapes of the weld pool and keyhole vary with different welding currents and speeds. To ensure the reliability and robustness of the segmentation model, the training dataset should be collected in as many different welding conditions as possible. Welding currents were randomly varied below 480 A, while welding speeds ranged between 280 mm/min and 360 mm/min. Figure 4 illustrates images of the molten pool obtained under different welding currents and speeds.

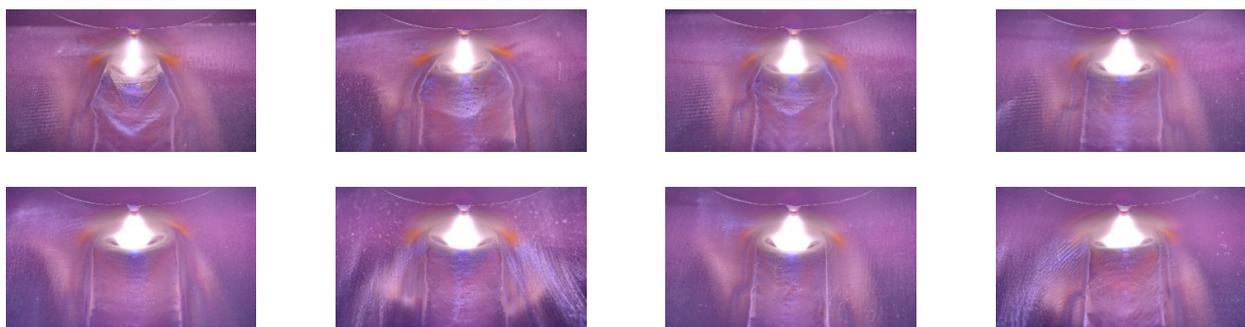


Figure 4. Images of molten pool obtained under different welding currents and speeds.

A total of 3015 clear molten pool original images were amassed. To effectively mitigate the overfitting problem in the segmentation model, it is necessary to enrich the dataset by the method of image augmentation, which includes three geometric transformations: rotation, scaling, and flipping. The dataset was randomly divided into two sub-datasets: 10,854 as the training dataset and 1206 as the validation dataset. To make it easier to label the weld pool and keyhole edge, the Improved LabelMe image annotation software [24] was used to label both the training dataset and validation dataset images with the labels “weld pool” and “keyhole”.

3. Design of Segmentation Model

3.1. Improved DeepLabV3+ Model

The DeepLabV3+ model is improved based on DeepLabV1-3, which has achieved good results in public datasets such as Pascal VOC and CitySpace. The DeepLabV3+ model consists of an encoder and decoder. The encoder part includes the Xception network, atrous spatial pyramid pooling (ASPP) module, and 1×1 convolutional layer. The ASPP module is comprised of a 1×1 ordinary convolution layer, three atrous convolution layers with different atrous rates, and a pooling layer. Two feature maps with different levels are produced by the Xception network of the encoder after extracting the features of the input image. The features of the low-level feature map are extracted in the ASPP module, which is spliced in the channel dimension and compressed through the 1×1 convolution to obtain a feature map with high-level semantic information. In the decoder part, first, the 1×1 convolution is used to adjust the channel dimension of the high-level feature map. After that, the adjusted feature map can fuse with the feature maps upsampled by four times. Subsequently, the fused feature map employs a 3×3 convolution to recover spatial information and then is upsampled four times to enhance the target boundary, producing the final segmentation result.

In the experiment, the complicated structure of the Xception network will result in some blurring and confusion problems in detail extraction for the weld pool and keyhole. The large number of parameters in Xception necessitates a significant investment of time and computing power, so MobileNetV2 is selected as the backbone feature extraction network. MobileNetV2, with fewer parameters and faster training speed, accelerates feature extraction in molten pool images. This facilitates improved detail capture for the weld pool and keyhole, rendering it more suitable for molten pool extraction tasks. At the same time, given the differing characteristics of the weld pool and keyhole, it is not suitable to apply the DeepLabV3+ model directly to the molten pool segmentation. The improved DeepLabV3+ model architecture is shown in Figure 5.

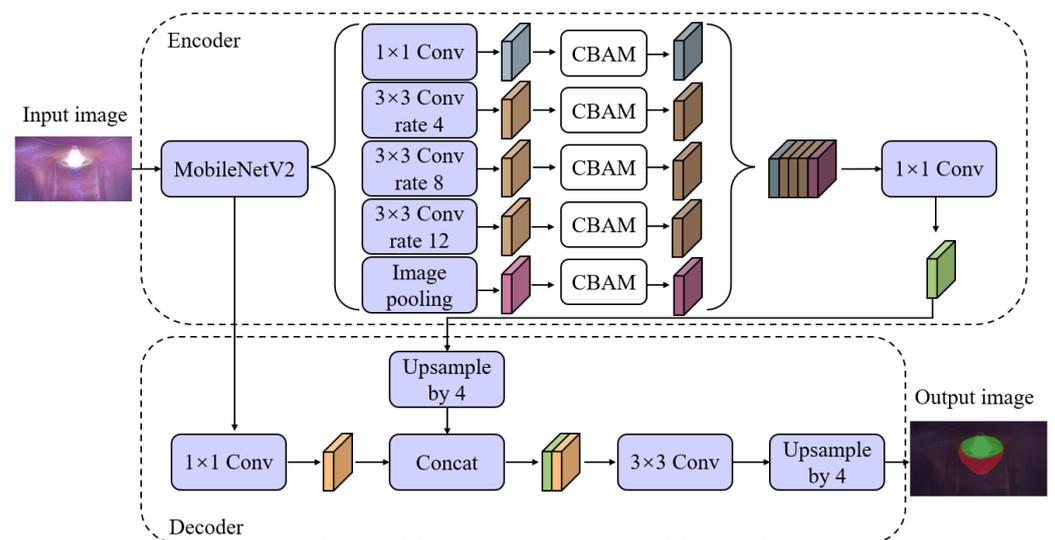


Figure 5. Improved DeepLabV3+ model architecture.

3.2. MobileNetV2 Network

MobileNetV2 is used as the backbone feature extraction network, and its network structure is shown in Figure 6. MobileNetV2 adopts an inverted residual block structure, which gives the network a good ability for feature extraction while maintaining a small number of parameters and computational complexity. First, the dimension of the input feature map is expanded by 1×1 convolution, followed by the utilization of ReLU6 for feature non-linearization. Then, 3×3 depthwise separable convolution and the ReLU6 activation

function are employed to comprehensively extract and non-linearize features. Finally, feature fusion and dimensionality reduction are carried out through 1×1 convolution.

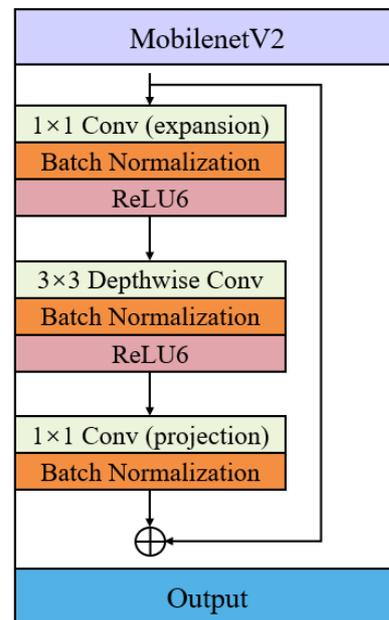


Figure 6. MobileNetV2 network structure.

3.3. The Optimization of ASPP Module

In the DeepLabV3 + model, the ASPP module plays a crucial role in capturing multi-scale context information from the image. It achieves this by utilizing convolution kernels of different scales to optimize segmentation results and enhance segmentation accuracy. Therefore, the ASPP in this paper used three atrous convolutional layers with varying atrous rates to extract information from the weld pool and keyhole. If a larger atrous rate is selected, the receptive field expansion can reduce the loss of feature information during the convolution process. However, it may result in misidentification as the background and poor extraction effects when the receptive field exceeds the size of the weld pool or keyhole to be extracted. With the increase of the atrous rate, the attenuation of the expansion convolution becomes invalid when it is greater than 24, leading to a loss of the resolution in the feature map and subsequent loss of detailed information at the boundaries of the weld pool and keyhole. Conversely, if the atrous rate is too small, the receptive field becomes smaller than the feature area, and too much local information will be obtained, which will lead to the loss of global information. To address these considerations and leverage the characteristics of different scales and irregular edges of the weld pool and keyhole, this paper optimized the parameters through multiple experiments and finally reduced the atrous rate appropriately. The atrous rates in the ASPP were adjusted to 4, 8, and 12, striking a balance to improve the model's recognition ability for molten pool images.

3.4. Attention Module

The CBAM is a lightweight attention module proposed by Woo et al. [25]. It infers the attention weights independently along the two dimensions: channel and space. The resulting weights are then multiplied by the input feature to adaptively adjust the features. The introduction of CBAM into the CNN can greatly improve the performance of the model, and it brings a very small number of parameters and computations, which can be embedded into most networks. The structure is shown in Figure 7.

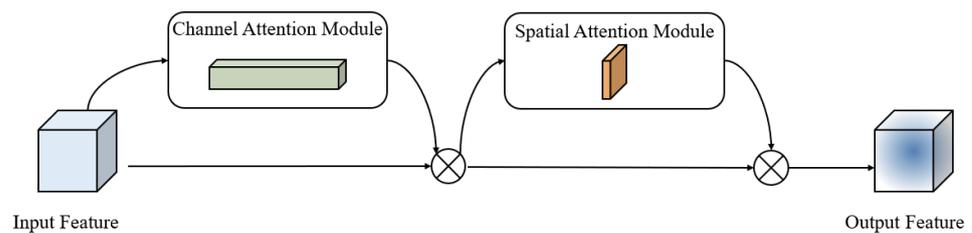


Figure 7. Structure of CBAM.

The CBAM is composed of a channel attention module and a spatial attention module. In the channel attention module, a pooling module is formed by the mean-pooling and max-pooling via parallel connection. Two feature maps, including the maximum and average information, are derived through the pooling module. Subsequently, both feature maps pass through a Multilayer Perceptron (MLP), in which the first fully connected layer reduces dimensionality to mitigate computational complexity while the second fully connected layer elevates dimensionality. The outputs from the MLP are summed, and the channel weights are generated by the Sigmoid activation function. These weights are then multiplied with the input features to adjust the feature response of each channel, generating the input feature needed for the spatial attention module. In the spatial attention module, mean-pooling and max-pooling operations are applied to the feature map produced by the channel attention module. The results are combined to form a two-channel feature map, which is then converted into a single-channel feature map through convolutional operations. The Sigmoid activation function generates spatial weights, which are multiplied with the input feature map, weighting the feature response at each spatial location to obtain the final feature map.

The excessive atrous rate of the ASPP module will make the model unable to extract the feature information well, which will also affect the correlation between the local features of the molten pool and reduce the segmentation accuracy. Specifically, the CBAM attention module is introduced after each atrous convolution in the ASPP module, therefore adjusting the weight of the feature channel and enhancing the perception of edge features in the weld pool and keyhole.

4. Training and Results

The proposed improved DeepLabV3+ network is implemented in the PyTorch framework. The training was conducted in a computer environment with support for the Intel i7-9750H CPU (Intel Corporation, Santa Clara, CA, USA) and NVIDIA GTX 1650 (Nvidia Corporation, Santa Clara, CA, USA) graphics card. To expedite training and ensure quick convergence, the network was initialized with pretraining weights and then finetuned through transfer learning. Transfer learning is divided into two stages. In the first stage, the parameters of the network model were initialized on the Pascal VOC dataset. By freezing the batch normalization layer of the ASPP module and decoder part, it was not updated during feature migration, which can reduce the model error, ensure the migration effect, and obtain the pretraining weight. The second stage involved training on the dataset of the molten pool. Since the features extracted by the backbone network, MobileNetV2 was common in the whole model. The backbone network was frozen at the beginning of training to speed up the model training and prevent the weights from being destroyed during the training process. In later training epochs, the backbone network was unfrozen to participate in the training of the entire model. The model was trained for 100 epochs with a batch size of 8, utilizing the stochastic gradient descent (SGD) algorithm with 0.9 momentum and a weight decay of 1×10^{-4} . The learning rate was 1×10^{-4} .

After 100 epochs of training, the change in the loss value during model training and validation is shown in Figure 8. As we can see in the figure, with the increase in training time, both training and validation loss curves rapidly converge and then stay stable. When the time of training reaches 60, the loss value reaches 0.03. To assess the effectiveness of the DeepLabV3+ network, the mean intersection over union (MIOU) was adopted to calculate and evaluate the percentage overlap between the target and prediction, which is defined in Equation (1).

$$MIOU = \frac{1}{K + 1} \sum_{i=0}^k \frac{p_{ii}}{\sum_{i=0}^k p_{ij} + \sum_{i=0}^k (p_{ji} - p_{ii})} \quad (1)$$

where K denotes the number of target categories, p_{ii} is the number of pixels predicted correctly, p_{ij} is the number of pixels of class i predicted to be class j , p_{ji} is the number of pixels of class j predicted to be class i . The MIOU of the model for the segmentation of the molten pool image is up to 89.89%.

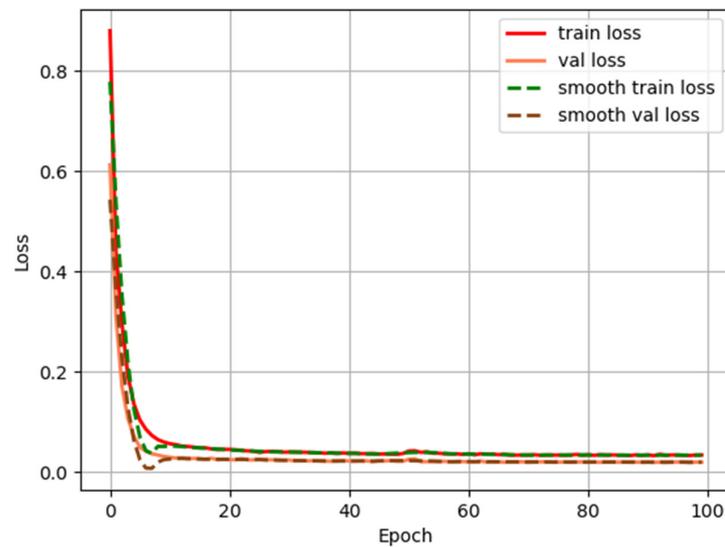


Figure 8. The change in the loss value during model training and validation.

The best-performing network model from the verification set was employed for testing on additional molten pool images. The images before and after the segmentation are shown in Figure 9. It can be observed in the figure that the boundaries of the weld pool and keyhole are successfully recognized by the trained DeepLabV3+ network. The network can process the images at 103 frames per second, i.e., the inference time of a single image is only 9.62 ms. Such a speed ensures efficient deployment and application for monitoring the K-TIG welding process.

The trained model was deployed in the K-TIG experiment system, as shown in Figure 1. During welding, the molten pool images were collected by the HDR camera and were transferred to the industrial computer, where the trained model was then applied to segment the weld pool and keyhole. Due to the camera's low frame rate, it will cause problems with duplicated frames. Therefore, before processing, it is necessary to determine whether the images acquired by the camera are the same from the pixel level. Although the hardware limitations affected the model performance display to a very great degree, the results demonstrated that the proposed model achieved a better balance between segmentation accuracy and speed, in which the average frame rate of online detection is about 28 frames per second.

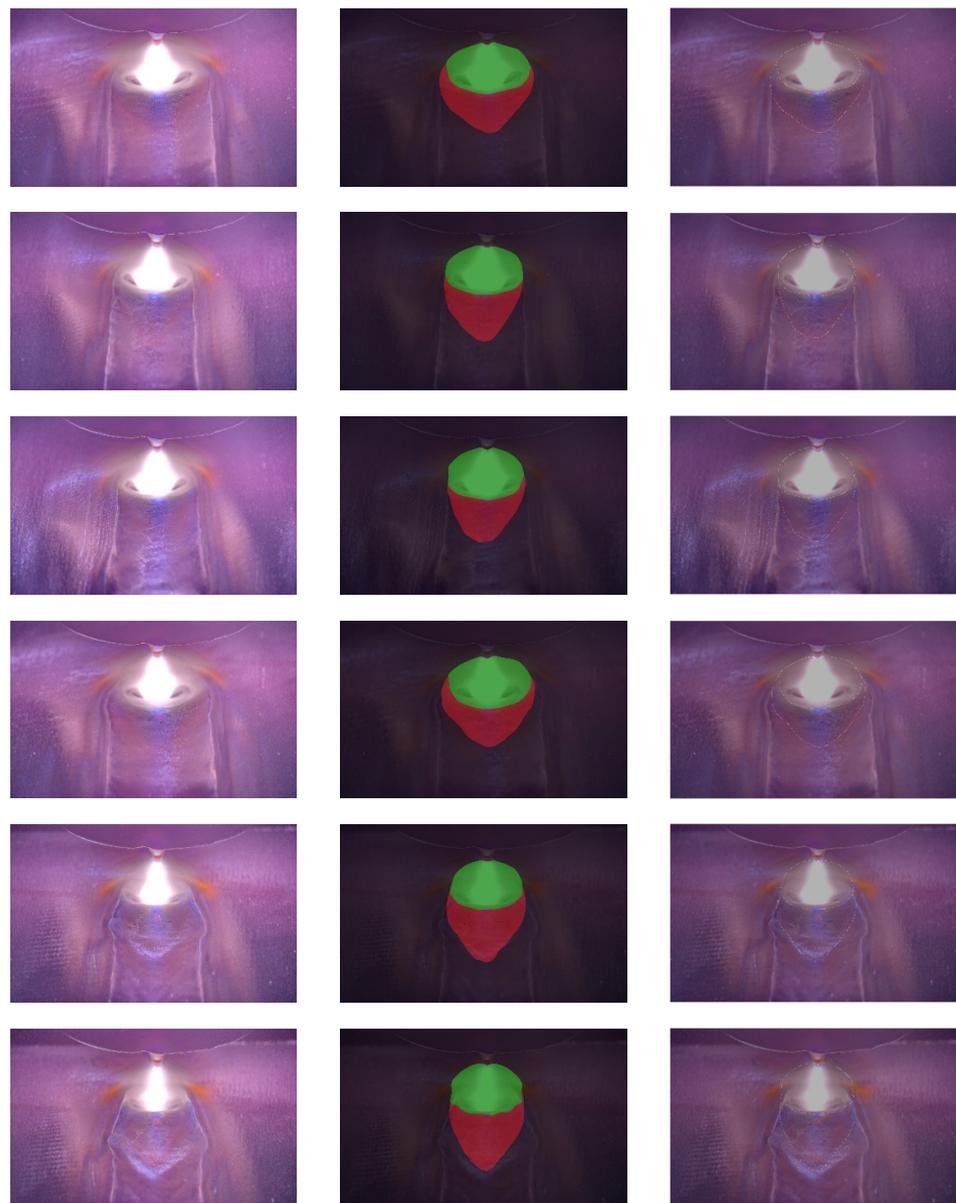


Figure 9. The images before and after the segmentation, boundary of the detected molten pool.

5. Conclusions

In this paper, aiming at the engineering requirements of weld pool image segmentation in K-TIG welding, a low-parameter and lightweight DeepLabV3+ network was built. Based on the DeepLabV3+ model, the Xception backbone network is replaced with the MobileNetV2 backbone network, significantly reducing the number of parameters of the model and minimizing its dependence on hardware conditions, therefore accelerating the convergence speed of the network. The segmentation speed of the molten pool reaches 103 frames per second, with an MIOU value of 89.89%, which realizes real-time high-precision segmentation. The improved model in this paper reduces manual participation in the manufacturing process, offering significant benefits in terms of reducing the defective rate of the project, cutting costs, and enhancing the safety of the project. Furthermore, for melt pool monitoring in K-TIG welding, the improved model enables more precise segmentation of the weld pool and keyhole, extracting their contour edges comprehensively and obtaining characteristic parameters, such as area and length-width ratio, which can be instrumental in monitoring welding quality and provide a research basis for the subsequent realization of the closed-loop control in K-TIG welding.

Author Contributions: Conceptualization, D.Y., P.D. and S.C.; methodology, D.Y., P.D., S.C. and H.S.; software, P.D. and F.L.; validation, D.Y., P.D., S.C. and F.L.; investigation, D.Y., P.D., S.C. and H.S.; data curation, P.D., H.S., F.L. and X.Z.; writing—original draft preparation, P.D.; writing—review and editing D.Y. and S.C.; funding acquisition, S.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by [Special Project for Central Government Guiding Local Science and Technology Development (Liuzhou City)] grant number [2022JRZ0101] and [Guangxi University of Science and Technology Doctoral Fund] grant number [19Z27] and [Guangxi University of Science and Technology Graduate Education Innovation Program Project] grant number GKYC202316.

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declared that they do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

1. Cui, S.; Shi, Y.; Zhu, T.; Liu, W. Microstructure, texture, and mechanical properties of Ti-6Al-4V joints by K-TIG welding. *J. Manuf. Process.* **2018**, *37*, 418–424. [\[CrossRef\]](#)
2. Cui, S.; Pang, S.; Pang, D.; Zhang, Z. Influence of Welding Speeds on the Morphology, Mechanical Properties, and Microstructure of 2205 DSS Welded Joint by K-TIG Welding. *Materials* **2021**, *14*, 3426. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Cui, S.; Yu, Y.; Tian, F.; Pang, S. Morphology, microstructure, and mechanical properties of S32101 duplex stainless-steel joints in K-TIG welding. *Materials* **2022**, *15*, 5432. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Fan, X.; Gao, X.; Liu, G.; Ma, N.; Zhang, Y. Research and prospect of welding monitoring technology based on machine vision. *Int. J. Adv. Manuf. Technol.* **2021**, *115*, 3365–3391. [\[CrossRef\]](#)
5. Zhang, Y.M.; Yang, Y.-P.; Zhang, W.; Na, S.-J. Advanced Welding Manufacturing: A Brief Analysis and Review of Challenges and Solutions. *J. Manuf. Sci. Eng.* **2020**, *142*, 110816. [\[CrossRef\]](#)
6. Gao, J.-Q.; Qin, G.-L.; Yang, J.-L.; He, J.-G.; Zhang, T.; Wu, C.-S. Image processing of weld pool and keyhole in Nd:YAG laser welding of stainless steel based on visual sensing. *Trans. Nonferrous Met. Soc. China* **2011**, *21*, 423–428. [\[CrossRef\]](#)
7. Wu, D.W.; Xiong, Z.Y.; Gu, W.P.; Shan, J. Molten pool image processing and feature extraction based on multiple visions. In Proceedings of the 3rd International Conference on Advanced Engineering Materials and Technology (AEMT), Zhangjiajie, China, 11–12 May 2013. [\[CrossRef\]](#)
8. Johan, N.F.; Shah, H.N.M.; Sulaiman, M.; Naji, O.A.A.M.; Arshad, M.A. Weld seam feature point extraction using laser and vision sensor. *Int. J. Adv. Manuf. Technol.* **2023**, *127*, 5155–5170. [\[CrossRef\]](#)
9. Cai, W.; Jiang, P.; Shu, L.; Geng, S.; Zhou, Q. Real-time identification of molten pool and keyhole using a deep learning-based semantic segmentation approach in penetration status monitoring. *J. Manuf. Process.* **2022**, *76*, 695–707. [\[CrossRef\]](#)
10. Ai, Y.; Han, S.; Lei, C.; Cheng, J. The characteristics extraction of weld seam in the laser welding of dissimilar materials by different image segmentation methods. *Opt. Laser Technol.* **2023**, *167*, 109740. [\[CrossRef\]](#)
11. Liu, X.G.; Zhao, B. Based on the CO₂ gas shielded welding molten pool image edge detection algorithm. *Appl. Mech. Mater.* **2013**, *437*, 840–844. [\[CrossRef\]](#)
12. Chen, Z.; Chen, J.; Feng, Z. Welding penetration prediction with passive vision system. *J. Manuf. Process.* **2018**, *36*, 224–230. [\[CrossRef\]](#)
13. Liu, J.; Fan, Z.; Olsen, S.; Christensen, K.; Kristensen, J. Using active contour models for feature extraction in camera-based seam tracking of arc welding. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 5948–5955. [\[CrossRef\]](#)
14. Wang, Z.; Shi, Y.; Hong, X.; Zhang, B.; Chen, X.; Zhan, A. Weld pool and keyhole geometric feature extraction in K-TIG welding with a gradual gap based on an improved HDR algorithm. *J. Manuf. Process.* **2021**, *73*, 409–427. [\[CrossRef\]](#)
15. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. [\[CrossRef\]](#)
16. Peng, C.; Zhang, X.; Yu, G.; Luo, G.; Sun, J. Large kernel matters—Improve semantic segmentation by global convolutional network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1743–1751. [\[CrossRef\]](#)
17. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid scene parsing network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2881–2890. [\[CrossRef\]](#)
18. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Semantic image segmentation with deep convolutional nets and fully connected CRFs. *arXiv* **2014**, arXiv:1412.7062. [\[CrossRef\]](#)
19. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *arXiv* **2016**, arXiv:1606.00915. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* **2017**, arXiv:1706.05587. [\[CrossRef\]](#)

21. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018. [[CrossRef](#)]
22. Xia, C.; Pan, Z.; Fei, Z.; Zhang, S.; Li, H. Vision based defects detection for Keyhole TIG welding using deep learning with visual explanation. *J. Manuf. Process.* **2020**, *56*, 845–855. [[CrossRef](#)]
23. Chen, Y.; Shi, Y.; Cui, Y.; Chen, X. Narrow gap deviation detection in Keyhole TIG welding using image processing method based on Mask-RCNN model. *Int. J. Adv. Manuf. Technol.* **2021**, *112*, 2015–2025. [[CrossRef](#)]
24. Shi, Y.-H.; Wang, Z.-S.; Chen, X.-Y.; Cui, Y.-X.; Xu, T.; Wang, J.-Y. Real-time K-TIG welding penetration prediction on embedded system using a segmentation-LSTM model. *Adv. Manuf.* **2023**, *11*, 444–461. [[CrossRef](#)]
25. Woo, S.; Park, J.C.; Lee, J.Y.; Kweon, I.S. CBAM: Convolutional block attention module. In Proceedings of the 15th European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.