Automated Recognition of Weld Pool Characteristics from Active Vision Sensing

Deep learning was used to automatically extract weld pool characteristics from active vision images to predict weld penetration in GTAW

BY Y. C. CHENG, Q. Y. WANG, W. H. JIAO, J. XIAO, S. J. CHEN, AND Y. M. ZHANG

ABSTRACT

While penetration occurs underneath the workpiece, the raw information used to detect it during welding must be measurable to a sensor attached to the torch. Challenges are apparent because it is difficult to find such measurable raw information that fundamentally correlates with the phenomena occurring underneath. Additional challenges arise because the welding process is extremely complex such that analytically correlating any raw information to the underneath phenomena is practically impossible; therefore, handcrafted methods to propose features from raw information are human dependent and labor extensive. In this paper, the profile of the weld pool surface was proposed as the raw information. An innovative method was proposed to acquire it by projecting a single laser stripe on the weld pool surface transversely and intercepting its reflection from the mirror-like weld pool surface. To minimize human intervention, which can affect success, a deep-learning-based method was proposed to automatically recognize features from the single-stripe active vision images by fitting a convolutional neural network (CNN). To train the CNN, spot gas tungsten arc welding experiments were designed and conducted to collect the active vision images in pairs with their actual penetration states measured by a camera that views the backside surface of the workpiece. The CNN architecture was optimized by trying different hyperparameters, including kernel number, kernel size, and node number. The accuracy of the optimized model is about 98% and the cycle time in the personal computer is ~ 0.1 s, which fully meets the required engineering application.

KEYWORDS

• Convolutional Neural Network (CNN) • Deep Learning
• Gas Tungsten Arc Welding (GTAW) • Weld Penetration
• Weld Pool Image

Introduction

Gas tungsten arc welding (GTAW) has been widely used in industrial manufacturing, especially for some critical cases such as pressure vessels and aerospace. In these critical application cases, complete joint penetration is a basic requirement to obtain high-quality weld joints. Generally, the back-bead width is taken as the criterion of weld joint penetration but is usually unavailable to sense in practical manufacturing due to the spatial limitation. To address this problem, some sensing methods, including spectral, infrared, audio, arc voltage, vision, x-ray, and molten pool oscillation sensing, have been developed to monitor the topside joint penetration state and build a foundation for automatic and intelligent control of welding processes.

Li and Zhang (Ref. 1) proposed a method to measure the arc length using a particular wavelength of arc light in GTAW, and the accuracy of the experiments was about 0.2 mm for arc lengths in the range of 0.5 to 5 mm. Alfaro and Franco (Ref. 2) presented an evaluation using an infrared sensor to monitor the penetration of the weld pool and analyzed the defects through graphics and statistic methods. Lv et al. (Ref. 3) established a prediction model using the statistic features of the arc sound signal through artificial neural networks to monitor the state of the weld pool in GTAW. Zhang et al. (Ref. 4) proposed an arc voltage sensing method to detect the penetration in pulsed GTAW (GTAW-P). The results showed that the fluctuation amplitude of the arc voltage can reflect the weld penetration status and a certain threshold as the feature signal that judges the weld pool state between incomplete joint penetration and complete joint penetration. However, the accuracy of the arc voltage sensing method was influenced by the experimental conditions, such as the oscillation of the weld pool and arc stability. Feng et al. (Ref. 5) proposed a method to quantify the height between the weld pool surface and the reversed electrode image and used a passive vision to capture the distance between the electrode tip and the virtual reversed electrode image, arc length, and weld pool geometry to calculate the surface height of the weld pool. This method has been experimentally verified and applied to monitor the occurrence of complete joint penetration. Huang et al. (Ref. 6) used spectral analysis based on data mining and empirical mode decomposition to detect the porosity in alternating current GTAW-P and used the x-ray analysis to verify that the welding conditions affected the size and distribution of pores. However, x-ray equipment is expensive and protec-

https://doi.org/10.29391/2021.100.015

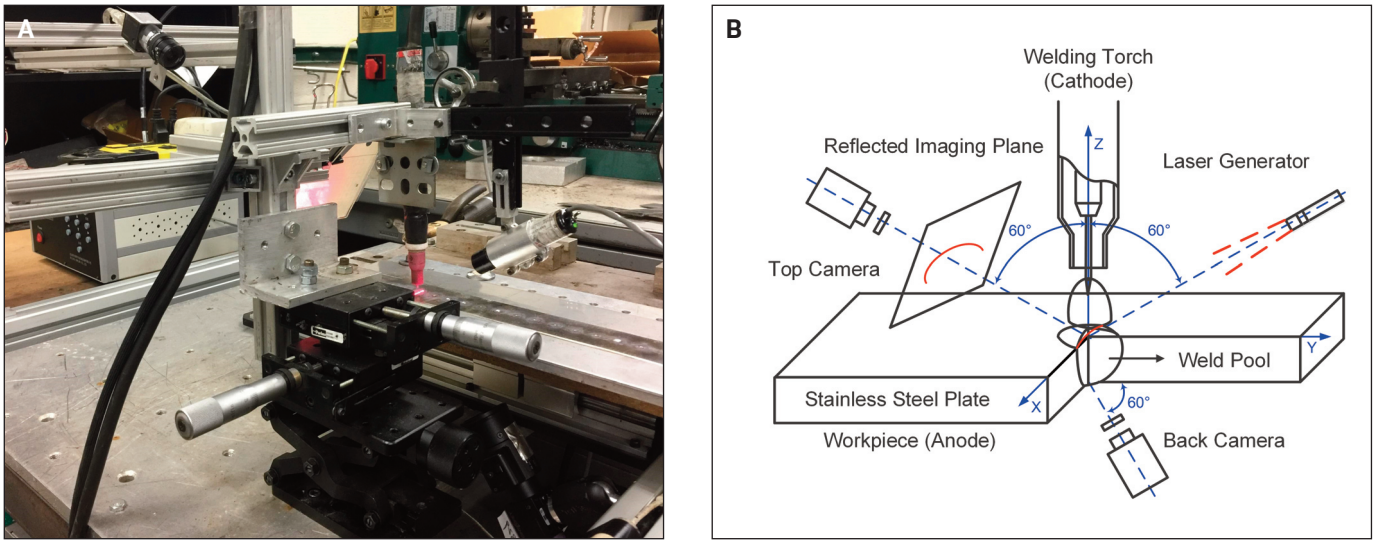

Fig. 1 — System configuration: A — Sensing system configuration; B — experiment schematic diagram.

tion against gamma radiation must be executed. Li et al. (Ref. 7) used a photoelectric conversion chamber to analyze the captured images in which an illumination laser with a five-line pattern was projected onto the surface of the weld pool. The voltage variation had a relationship with weld pool oscillation, and the oscillation frequency of incomplete joint penetration was higher than complete joint penetration with an abrupt transition.

Among these sensing methods, arc sound, infrared, x-ray, spectral, and molten pool oscillation sensing are phenomenological methods where the data-driven mapping models between sensed information and joint penetration state are built. The physical theories behind these methods are unknown or unclear. The other sensing methods, including arc voltage and vision sensing, are developed based on physical principles that complete joint penetration results in weld pool surface concaving, which is preferred for practical applications. Compared with the arc voltage sensing method as demonstrated by Zhang et al. (Ref. 8), where there is a single information source and penetration state identification depends on arc voltage dynamics instead of instantaneous value, vision sensing collects/uses abundant information. It is also the main method for human welders perceiving welding processes.

The welding penetration state characterized as the backbead width of the weld joint needs to be monitored to apply the optimal welding parameters (welding current, arc length, traveling speed, etc.). In production, the backside of the workpieces being welded is not available due to space limitation and difficulty in synchronizing the torch and sensor. Therefore, weld penetration state estimation from available information is the basis of this projection. In the welding process, the information provided by the weld pool has been proven to be a useful source to predict weld penetration states. The weld pool surface is believed to contain sufficient information to determine the weld penetration. Many parameters of the weld pool profile as reported by Zhang et al. (Ref. 9) can indirectly/directly indicate the penetration status, such as the pool width, length, and surface convexity, etc., so the topside surface information is significant to represent the welding penetration process. Different penetration states have different weld pool topside surfaces. With this physical relationship as the foundation, the topside surface information of the weld pool can be used to describe the penetration state. The image information can be applied to predict the penetration state.

Vision-sensing methods used to identify the weld joint penetration state include active and passive vision sensing and have been studied extensively in previous research. An innovative laser-vision-based sensing method was presented by Shi et al. (Ref. 10) to measure the oscillation frequency in GTAW, and the frequency was verified through experiments. However, this method is only used to analyze the feasibility of the extracted frequency by algorithm, but real-time monitoring and control need to be tested and demonstrated further. This method was established through data-driven mapping without the real physical principle to support it. For passive vision, Liang et al. (Ref. 11) established a biprism stereo vision system to sense and reconstruct the surface of the weld pool. However, the reconstructed model takes many calculations in real-time penetration sensing. Compared with passive vision sensing, active vision sensing can minimize interferences from the welding process, like high-intensity arc brightness, spatter, and smoke, which may destroy the effectiveness of the information sensed by passive vision. By adding and capturing structured light (usually a laser pattern) projected on the weld pool surface, geometrical information about the weld pool surface is collected. In previous studies, the characteristic features were extracted first, and a mapping model was built between the extracted features and joint penetration state as explained by Zhang et al. (Ref. 12). The characteristic features are defined and extracted artificially and are usually some simple geometrical dimensions of the weld pool, such as length, width, and height. The mapping models between the joint penetration state and extracted features are designed artificially. The performance of such an approach depends greatly on the designed characteristic features and models. In addition, some other useful information may be lost during processing raw images. To address such issues, we propose to

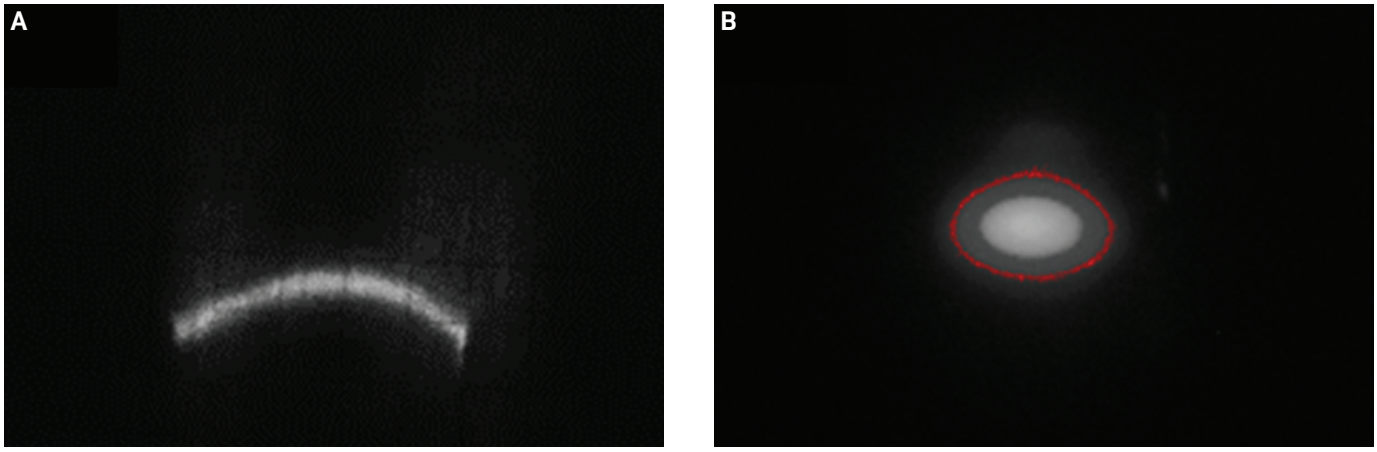

Fig. 2 — Images illustrating the vision-sensing structure. A — Topside image of weld pool; B — backside image of weld pool.

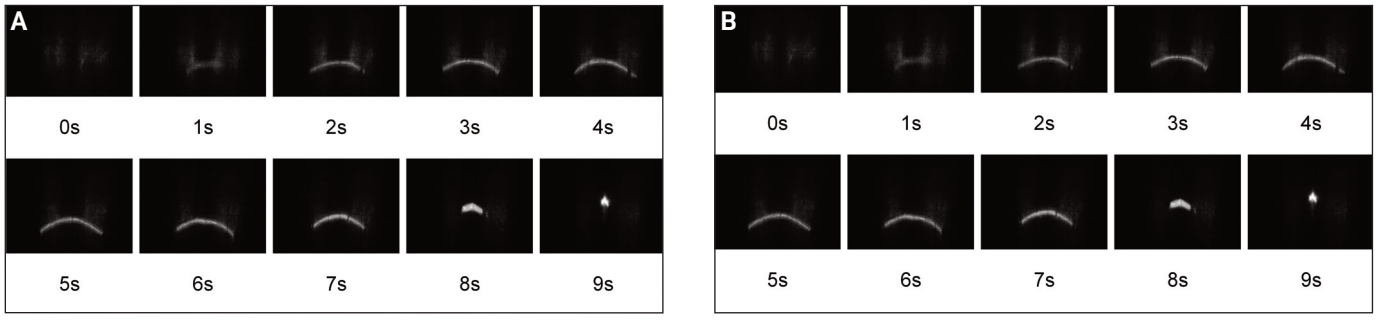

Fig. 3 — Topside and backside images to establish the dataset. A — Topside images; B — backside images.

use a single-stripe laser as the structured light to project on the weld pool and implement the CNN to process the raw images and then extract the useful features automatically. This approach will decrease the intensity of human involvement, and the automatically extracted features will keep useful information as much as possible compared with traditional methods.

The CNN, as demonstrated by Lecun et al. (Ref. 13) as one of the most popular models to process grid-shape data, has been frequently studied and successfully applied in natural language processing and computer vision, including image or video recognition and image classification. In the CNN, convolutional kernels are implemented to extract low-level features, and full-connected layers are used to do high-level reasoning. After being trained by large-volume data, the CNN can learn useful features automatically, increasing developmental efficiency.

In the research area of CNNs, some great contributions have been made that effectively promoted the development of CNNs. LeCun et al. (Ref. 14) presented the first CNN and defined the basic structure. The resultant LeNet structure was applied in written character recognition, which was deployed commercially in business and personal checks. However, this model was not popular in the following years because other algorithms, such as support vector machines, can often achieve better results. AlexNet, developed by Krizhevsky et al. (Ref. 15), prompted the popularity of CNNs again, and its contribution reflected new characteristics such as rectified linear unit (ReLU) active function, dropout, overlapping max pooling, local response normalization, graphics processing unit acceleration, and data augmentation. Afterward, Zeiler and Fergus (Ref. 16) developed ZF Net to optimize the classification of ImageNet by adjusting parameters but not improving the network structure. With the development of CNNs, its structure became deeper. The visual geometry group at the University of Oxford proposed the VGG Nets as reported by Simonyan and Zisserman (Ref. 17), which pushed the depth to 16–19 weight layers. GoogLeNet, which was presented by Szegedy et al. (Ref. 18), improved the utilization of the computing resources and, meanwhile, obtained a top-5 error of 6.67% on both the validation and testing data. With the depth of the neural network layer increasing, computing resources were largely occupied, and even the training process could not be executed using current computer hardware. To resolve this problem, ResNet, developed by He et al. (Ref. 19), presented a residual learning framework to ease the training of networks, gain accuracy from considerably increased depth, and achieve 3.57% error on the ImageNet test set.

This paper presents a novel active vision system that can identify the weld joint penetration state based on the CNN model. Firstly, experimental procedures are presented here, where images of the front and backside weld pool corresponding to different penetration states were collected as the dataset. Then, principles including the weld penetration analysis and reflection modes are introduced. The structure of the CNN is established, and a data augmentation method is proposed to expand the dataset. Finally, the accuracies with different parameters are analyzed.

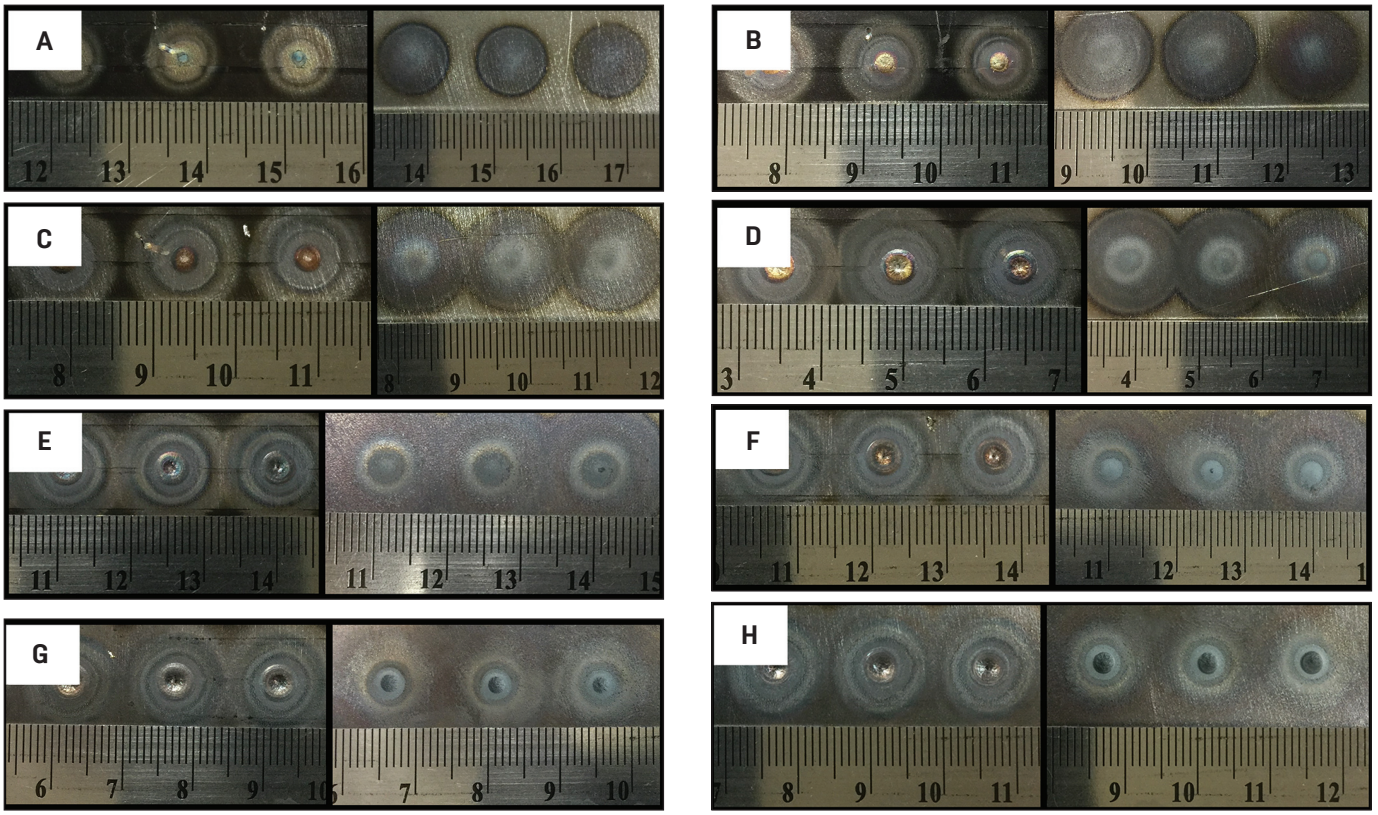

Fig. 4 — Macroscopic surfaces of welded workpieces at different times: A — 1 s; B — 2 s; C — 3 s; D — 4 s; E — 5 s; F — 6 s; G — 7 s; H — 8 s.

Experimental Procedures

Experimental System

A series of experiments were performed using GTAW to acquire the dataset. The diagram of the experiment platform is shown in Fig. 1A, including the welding system, motion system, and vision-sensing system. GTAW experiments were conducted using the parameters listed in Table 1 with the power supplied from a Miller® Maxstar® 210 welding power source and 99.99% pure argon as the shielding gas. The motion system was driven by a step motor, which allows the workpiece to move along one dimension. As shown in Fig. 1B, the vision-sensing system included a 50-mW/685-nm single-stripe laser generator, an imaging plane, and two Point Grey charge-coupled device cameras whose modes are FL3-FW-03S1C and FL3-FW-20S4C-C with 685-nm filters to sense the information of the topside and backside of the weld pool. Both the motion and vision-sensing systems were controlled through a computer with a National Instruments PCI-6229 data acquisition card using a Python™ environment to capture the images, control the current and time of the welding power, and move the workpiece.

During the welding process, the single-stripe laser was projected on the topside center of the weld pool and continuously reflected by the weld pool, whose topside surface is like a mirror. The reflected pattern was captured by a camera and characterizes the dynamic surface of the weld pool. The second camera was configured to monitor the backside surface of the weld pool synchronously where the captured images depend on the radiation of the weld joint and characterize the joint penetration state. A pair of typical images captured with the synchronously triggered cameras is shown in Fig. 2, in which the red circle of the backside images is calibrated for the brightness area through a certain threshold value. Hence, we can quantify the information from the backside image.

Table 1 — Welding and Sensing Parameters

Parameters	Type/Value
Material	304L
Thickness (mm)	1.85
Welding Type	Direct Current Electrode Negative
Welding Current (A)	60
Shielding Gas	Argon
Gas Flow (L/min)	7
Arc Length (mm)	4.8
Image Sampling Frequency (fps)	20

Data Collection

Using the welding parameters in Table 1, welding experiments were conducted with the images of the topside and backside as shown in Fig. 3. The macroscopic surfaces of the welded workpieces during different times are illustrated in Fig. 4. For the backside images, the brightness area of the image represents the weld pool state and its computable value at the given time. Thus, at each particular time, a computable value can sufficiently demonstrate one-to-one

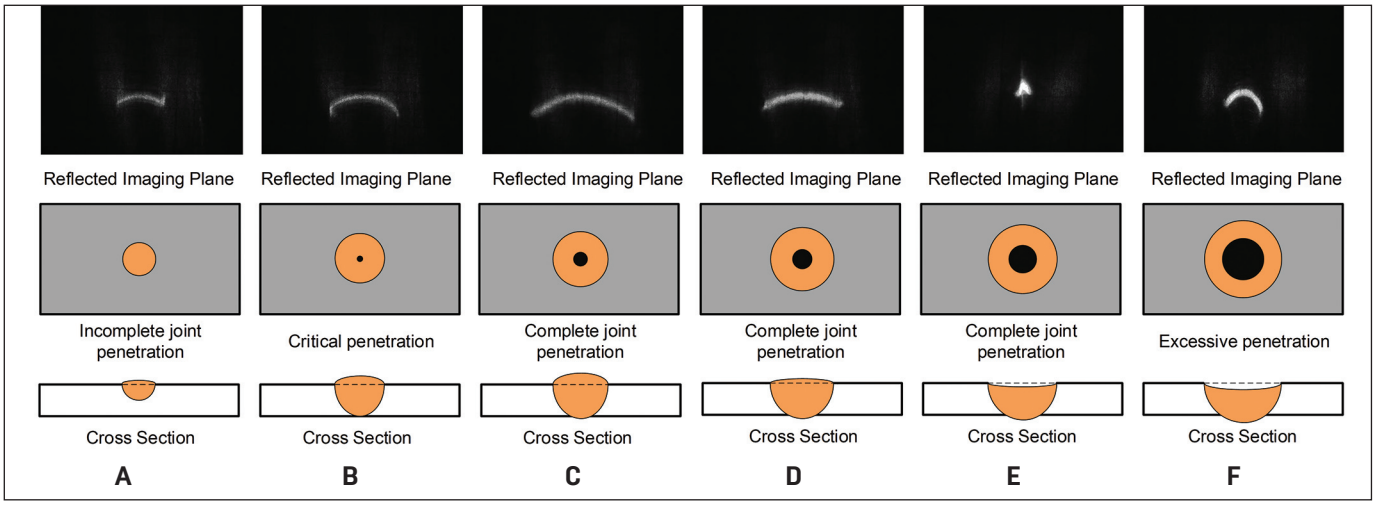

Fig. 5 — Dynamic penetration in the GTAW process (the single-stripe shape in different penetrations): A — Incomplete joint penetration; B — critical penetration; C, D, E — complete joint penetration; F — excessive penetration.

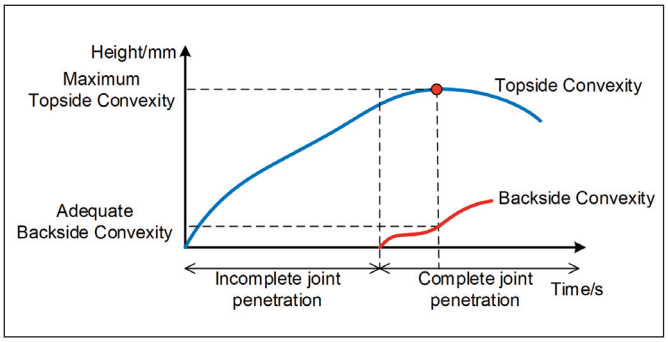

Fig. 6 — Dynamic evolution of the weld pool surface in GTAW.

correspondence with a topside image (a specific penetration state). The quantized penetration information can be delivered to the topside image by calculating the brightness area of the backside image. Therefore, the backside images served as the labels of topside images in the training model.

The real images of the topside and backside of the weld pool are shown in Fig. 2. For the backside image of the weld pool, first, the images were processed using the binary method in which the threshold value was set as 50, and the example of a backside image is shown in Fig. 2B in which the area has been drawn with a red circle. Then, we defined the area in the circle was used as the label for a synchronous topside image. During the calibration process, we used the area that is the number of all pixels in the red circle as the judging criteria. We defined the average value of the circle area via three repetitive welding experiments in 7 s as the designed penetration, and the average value was approximately 61 pixels (1728 pixels per image). The average area percentage, which was calculated using the equation (target pixels/all pixels) × 100%, was about 3.515%. In the welding process, if the number of pixels exceeds the area, we predict that the designed penetration occurs.

For arc spot welding, the penetration state transformed in sequence among incomplete joint penetration, complete joint penetration, and excessive penetration according to the welding time increasing. Figure 4A–H, with three repetitive welding experiments, shows the macroscopic surfaces of the weld pool during the welding time from 1 to 8 s in 1 s increments. In every part image, the left image is the topside surface and the right image is the backside of weld pool. The backside weld pool width of the designed or desired penetration state in 7 s is shown in Fig. 4G.

Principles

Dynamic Analysis of Welding Penetration

In one spot welding cycle, the penetration state shown in Fig. 5 changed from incomplete, critical, complete, to excessive joint penetration corresponding to the different geometrical profiles of the weld pool surface and different reflected laser patterns. We can observe that the reflected laser single stripe in the imaging plane, the top surface of weld pool, and the cross section are demonstrated in Fig. 5A–F. The imaging plane was used to receive the reflected laser from the weld pool. The diameters of the orange circle and black circle in the middle image of every part are the weld pool widths of the topside and backside, respectively. The cross sections of different penetration states show the evolution process of the weld pool. Before one spot welding cycle started, the camera captured the fully diffused reflection of the laser in the imaging plane due to the solid surface of the workpiece. When the weld pool started to form and grow, the top surface of the molten metal liquid was specular and became convex due to thermal expansion. As shown in Fig. 5A, the laser stripe curved after the specular reflection by the convexity of the molten metal liquid. With the metal heated continuously, the size of the weld pool increased such that the convexity and area of the molten metal increased. This resulted in the size of the reflected laser stipe increasing continuously in length and height and reaching critical penetration as shown in Fig. 5B. In the critical penetration state, the backside of weld pool began to increase. After passing critical penetration, the convex molten metal in the backside started to affect the top surface of the weld pool and the shape of the reflected laser stripe. While the thermal

Table 2 — The Parameters of the CNN Model

Layer Number	Layer Type	Method	Kernel Size	Kernel Number	Input Shape	Output Shape	Parameters
1	Convolution	Dot Product	5 × 5	128	48 × 36	44 × 32	3328
	Activation	ReLU	—	—	44 × 32	44 × 32	0
2	Pooling	Max Pooling	2 × 2	128	44 × 32	22 × 16	0
3	Convolution	Dot Product	3 × 3 × 128	128	22 × 16	20 × 14	147584
	Activation	ReLU	—	—	20 × 14	20 × 14	0
4	Pooling	Max Pooling	2 × 2	128	20 × 14	10 × 7	0
	Flatten	—	—	128	10 × 7	8960	0
5	Full-Connected	—	—	128	8960	128	1147008
	Activation	ReLU	—	—	128	128	0
6	Full-Connected	—	—	1	128	1	129
	Activation	Sigmoid	—	1	1	1	0

expansion of the weld pool tended to increase the convexity of the weld pool surface in the topside, the molten metal dropping in the backside tended to decrease that. In the beginning, the former factor dominated such that the length and height of the reflected laser stripe still increased as shown in Fig. 5C. Then, the latter dominated this dynamic change so that both the length and height of the laser stripe decreased as shown in Fig. 5D where the weld pool surface on the topside began to transfer from convexity to concavity. With the heat input increasing, because at some point the topside surface concavity of the weld pool focused the single stripe to form it as shown in Fig. 5E, the reflected single-stripe curve in the imaging plane became a focal point, like a concave mirror reflects light inward to one focal point. Finally, in the excessive penetration state as shown in Fig. 5F, the topside weld pool surface concaved significantly, and then the focal point from the imaging plane moved closer to the weld pool. Hence, the reflected laser stripe in the imaging plane became much longer than that in Fig. 5E, which is addressed and explained in detail in the Reflection Modes section.

Figure 6, as demonstrated by Li et al. (Ref. 20), illustrates the dynamic change in the geometry of the weld pool surface on both sides in GTAW during the increase of weld penetration. The topside convexity evolved and then shrunk when the workpiece changed from incomplete joint penetration to complete joint penetration, and backside convexity evolved in the complete joint penetration stage. The maximum height value of the topside convexity did not occur in the critical-penetration stage because the melting metal volume per second was still larger than the liquid metal volume flowing to the backside. When both volumes balanced dynamic equilibrium, the height of the topside convexity reached the maximum value exactly as a stage in Fig. 5C and the red point in Fig. 6.

Reflection Modes

As shown in Fig. 7A–D, both the convex and concave reflection modes occured as determined by the geometrical profiles of the weld pool topside surface. When the weld pool surface was convex due to thermal expansion of the molten metal, the convex reflection mode occurred in Fig. 7A and C. When the molten metal dropping in the backside dominated over the thermal expansion, the concave mode occured as shown in Fig. 7B and D. The two reflection modes are described as follows: According to the reflection principle, the projected laser is reflexed to a different direction according to the variation surface of the weld pool, and the imaging plane can capture and show the changing reflected single stripe that represents the penetration information. The topside surfaces of the weld pool in Fig. 7A and B are convex and concave, respectively. For the convex reflection mode in Fig. 7A and the concave reflection mode in Fig. 7B, three main points have been demonstrated to describe the reflection principle of the weld pool. When the metal was heated by the thermal arc, the topside of the weld pool generated a convexity due to the expansion of metal from solid to liquid in the incomplete joint penetration process. Meanwhile, the central reflection of the b point in the weld pool surface in Fig. 7C transfered to the top-right corner b_1 point, and the b' point in the reflected imaging plane of Fig. 7C transfered to the b_2 point in comparison with the nonmelted plate surface, so an arc-shaped curve appeared in the plane. Afterward, when the molten metal transferred to the backside of the plate, the topside of the weld pool generated a concavity during complete joint penetration or excessive penetration. In the meantime, the central reflection b point in Fig. 7D transferred to the bottom-left corner b_1 point in comparison with the nonmelted plate surface, and the b' point in the reflected imaging plane of Fig. 7D transfered to the b_2 point, so an arc-shaped curve appeared in the plane. The two curves, in which the topside surface information was captured to demonstrate the features including the penetration state, were different due to the convex and concave reflected modes. Therefore, the reflected single stripe in the imaging plane presented the penetration information directly and efficiently.

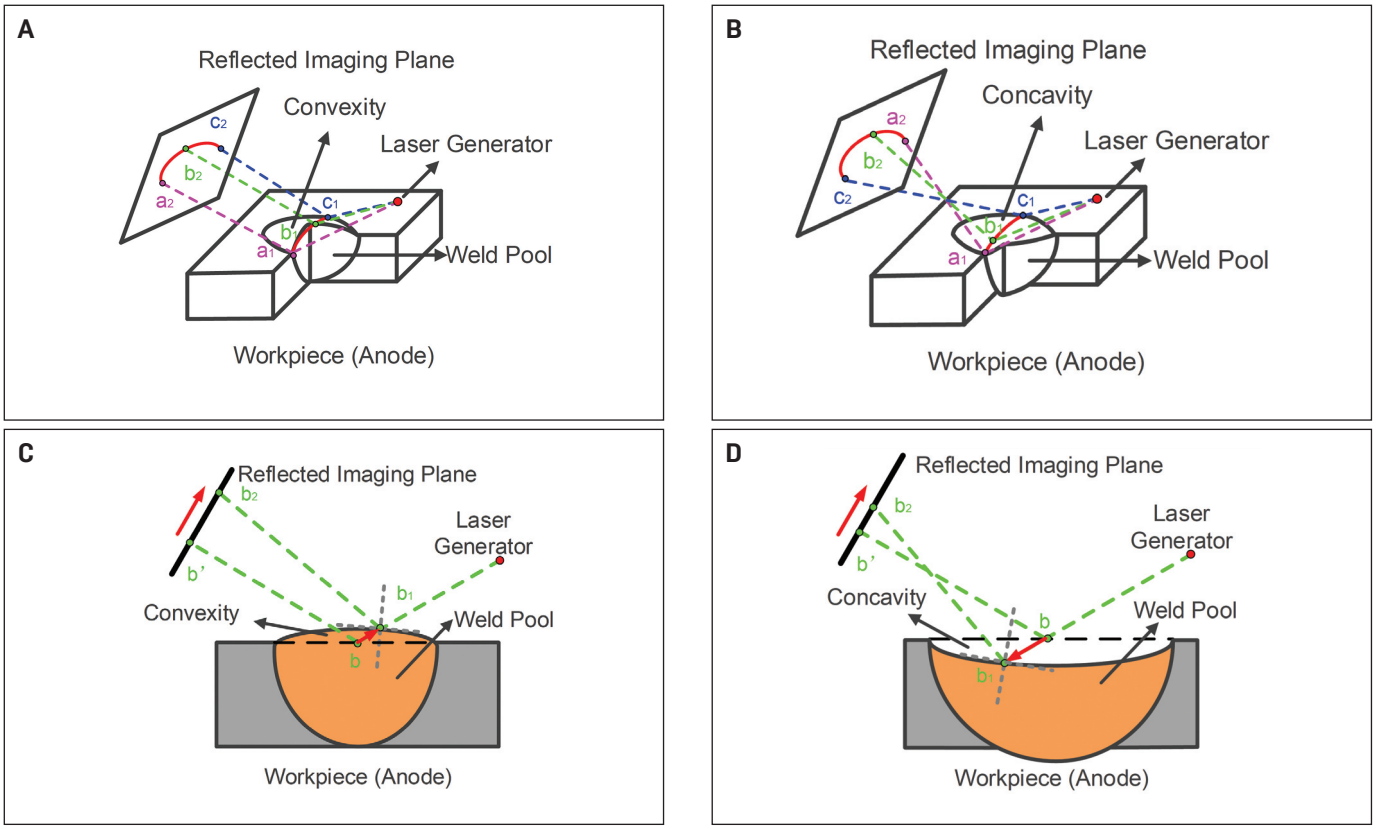

Fig. 7 — Two reflection modes: A, C — Convex reflection mode; B, D — concave reflection mode.

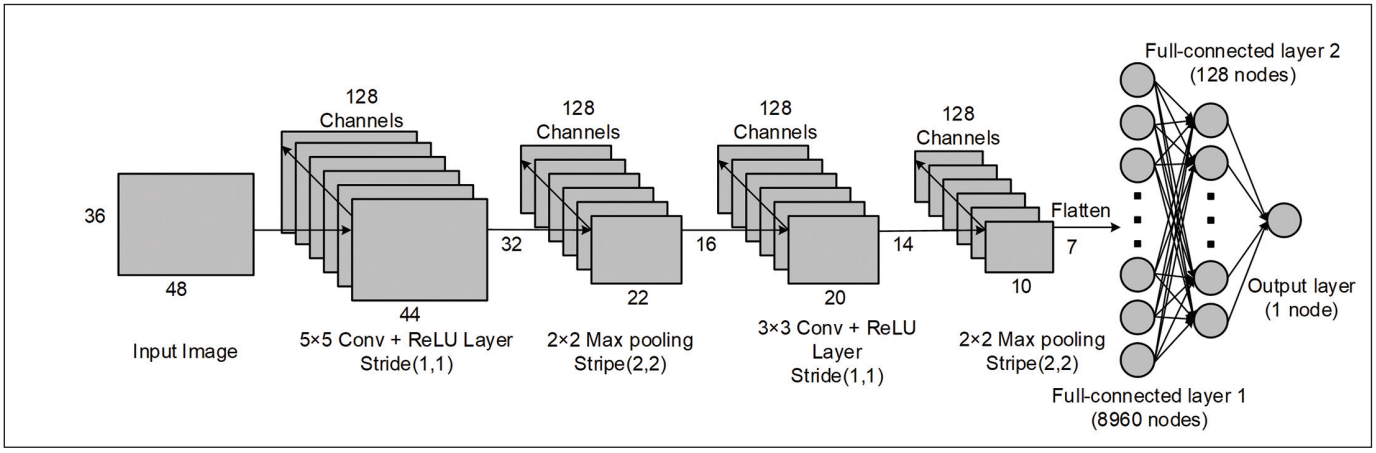

Fig. 8 — CNN architecture.

Results and Discussion

CNN Model

The result of this study is an effective model that can process the active vision images to classify the state of the weld penetration. This is a trained CNN model with three main layers that are the convolutional with ReLU, pooling, and full-connected layers, which were used to build the proposed CNN architectures as shown in Fig. 8 and Table 2. A benefit of using the Keras software library is that it is built on top of symbolic mathematical libraries for fast and efficient computation. A Keras sequential model that is a high-level neural network was selected as the linear stack of layers. It has three backend implementations available that are TensorFlow, Theano, and Microsoft Cognitive Toolkit. The TensorFlow backend developed by Google was selected in this model to complete data training and testing.

The convolution layer is a key part of the convolutional network, which computes dot products between the filter and the input. When we slid the filter over the width and height of the input volume through a certain method, two-dimensional mapping was produced, and we stacked these mappings along the depth dimension to produce the output volume. The specific calculation principle of the convolution

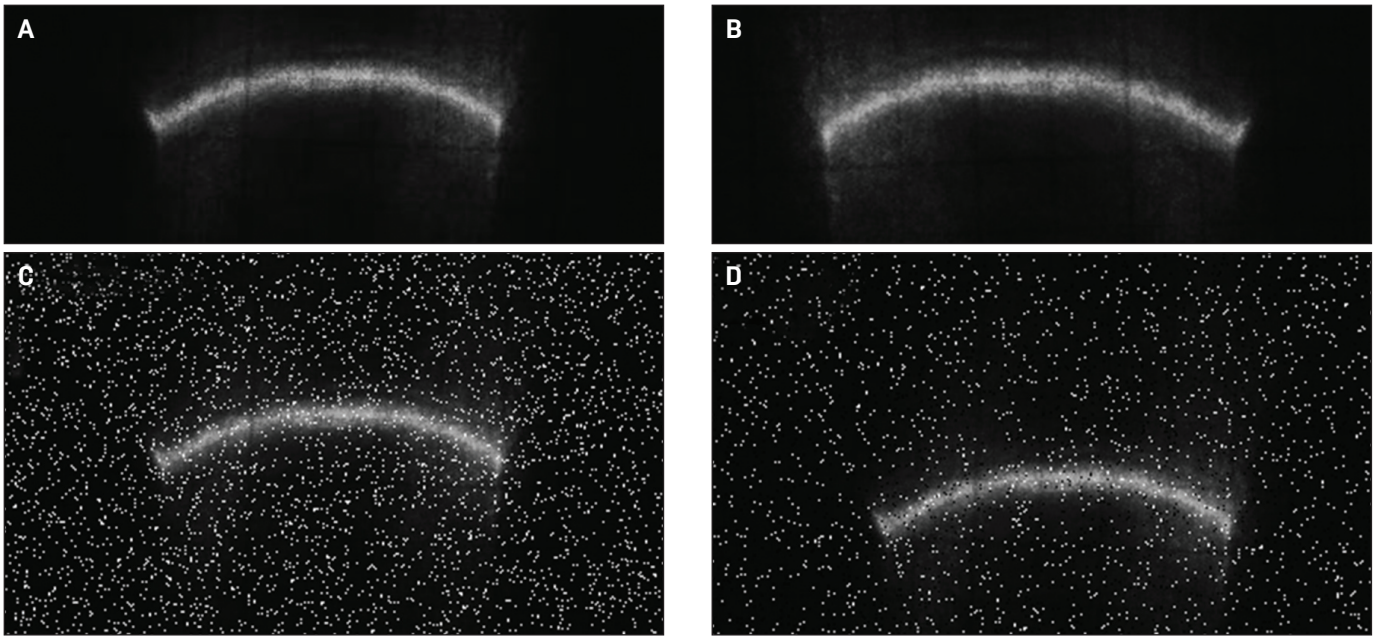

Fig. 9 — Data argumentation: A — Original image; B — horizontal flip; C — Gaussian noise; D — salt-and-pepper noise.

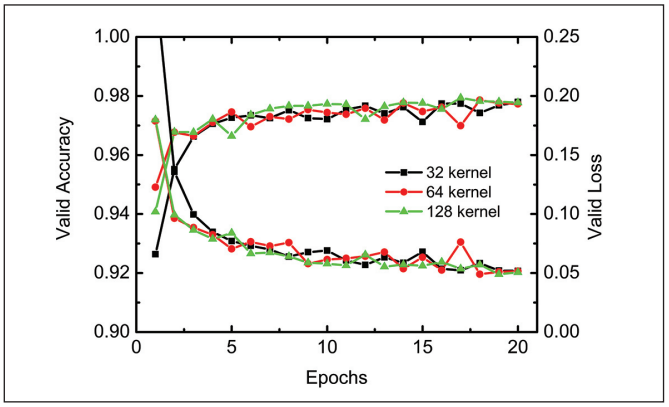

Fig. 10 — CNN model optimization of kernel numbers.

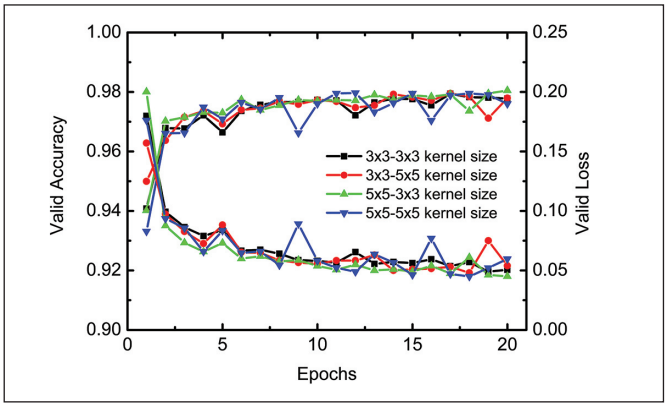

Fig. 11 — CNN model optimization of kernel sizes.

layer is shown as follows:

$$X = f(\Sigma w_i x_i + b) \quad (1)$$

$$W_{out} = (W_{in} - F + 2P)/S + 1 \quad (2)$$

where W_{out} and W_{in} are the width of the output and input volume, respectively; F is the filter size (kernel size); P is the zero padding; and S is the stride. The formula in Equation 1 calculates the weight and height of the image; the convolutional layer output is X. For this paper, padding type was set as the "valid," in which no zero padding in two convolution layers was used such that the output had the length and weight as shown in Fig. 8. Hence, the outputs of the two convolutional layers calculated using Equation 2 are a 44×32 grid of nodes with 128 channels and a 20×14 grid of nodes with 128 channels, respectively.

The pooling layer, which classifies as max pooling, mean pooling, and stochastic pooling, was periodically inserted after the convolution layer and performed a down-sampling operation along spatial dimensions with width and height. The max-pooling layer was selected in this structure to progressively reduce the spatial size and overfitting. Meanwhile, it reduced the number of parameters and computations in the network. The operation of max pooling is selecting a max value among the filter area and doesn't change the size of the depth dimension. In this work, filters of size 2×2 were applied to every depth slice of the input with a stride of two steps along both width and height using the max-pooling operation. The input volume of sizes $W_{in} \times H_{in} \times D_{in}$ in Equation 3 were generated as the output volume of sizes $W_{out} \times H_{out} \times D_{out}$ that is defined in Equa-

Table 3 — The Kernel Size in Two Convolutional Layers

Layer No.	1-Convolutional Layer	2-Convolutional Layer
#1	3×3	3×3
#2	3×3	5×5
#3	5×5	3×3
#4	5×5	5×5

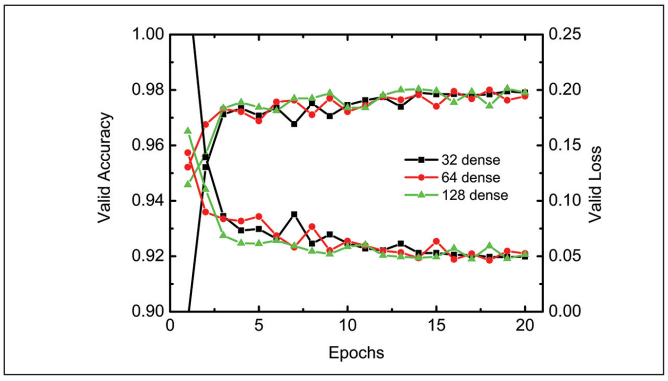

Fig. 12 — CNN model optimization of dense sizes.

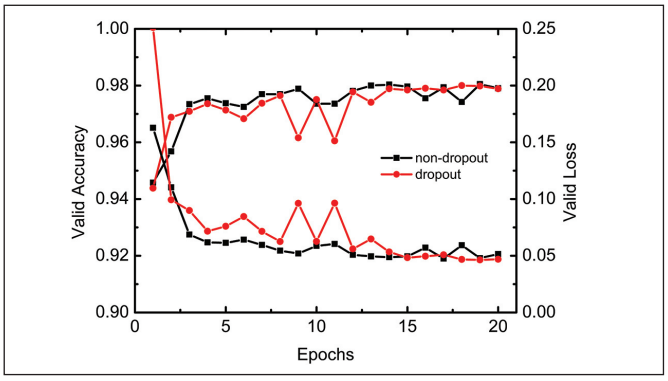

Fig. 13 — CNN model optimization of dropout.

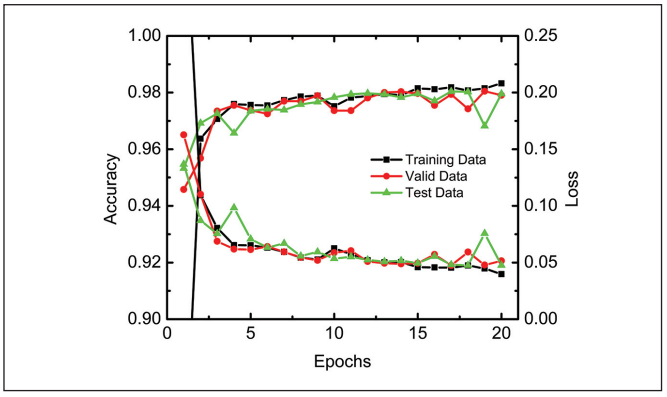

Fig. 14 — The training, validation, and test performance in the CNN model.

tion 4, respectively, where two hyperparameters are spatial extent F and stride S. Hence, the outputs of two pooling layers were a 22×16 grid of nodes with 128 channels and a 10×7 grid of nodes with 128 channels, respectively.

$$W_{out} = (W_{in} - F)/S + 1 \quad (3)$$

$$H_{out} = (H_{in} - F)/S + 1 \quad (4)$$

$$D_{out} = D_{in} \quad (5)$$

Neurons have full connections to the activations of the previous layer in a full-connected layer. In this research, the output from the final pooling layer was a 10×7 grid of nodes with 128 channels, which executed the flatten operation, and the output was one dimension with 8960 nodes. When the first full-connected layer was operated with 128 channels, the output was also one dimension with 128 nodes, and the whole parameters in this layer were 1,147,008 per training sample. For the second full-connected layer, the value was set as 1 node to implement the binary classification between complete joint penetration (designed penetration) and incomplete joint penetration.

The activation function is ReLU in Equation 6, which defined the positive part of its argument. The sigmoid function is a special case of the logical function, which is defined in Equation 7. A flatten operation reshaped the data to have a one-dimensional shape that was equal to the number of elements contented in the two-dimensional shape. As demonstrated by Kingma and Ba (Ref. 21), the optimizer was set to Adam, which is an adaptive learning rate optimization algorithm that has been designed for the deep learning network.

$$f(x) = max(0, x) \quad (6)$$

$$S(x) = 1/1 - e^{-x} \quad (7)$$

Data Augmentation and Image Resizing

To increase robustness and decrease the risk of overfitting, the training dataset was augmented by using a horizontal flip and adding noise, such as Gaussian and salt-and-pepper noise, to boost the original images. For the horizontal flip, the original images were flipped in the horizontal direction as shown in Fig. 9B. For the Gaussian and salt-and-pepper noise, the percentage of the image noise area was set to 0.05 for the original images as shown in Fig. 9C and D. The size of the original and augmentation images was 640×480, which is large and increased the computational cost for training, so the input images were resized to 48×36 at the same ratio. The total images through augmentation of these three methods using the original number of 15,627 images was expended to 62,508 to form the database. The training, validation, and test database were 50,633, 5625, and 6250, respectively, using a random allocation method. Finally, the resized data transmitted to the convolutional neural network to train the model.

Model Optimization

The model training is based on the Keras library with Python programming language, which is a popular library for deep learning and ran on a personal computer with an Intel® Core™ i7-6700 processor and 16G memory for 20 epochs, which was the number of trained times with all samples in the training dataset. In the training process, the parameter optimization was the key issue to improve the performance of the designed CNN model. Due to so many parameters in the model, some core parameters were designed and conducted to verify the stability and improve accuracy. The parameters, including kernel number and size, dense size, and dropout, were selected. The plot in Fig. 10 depicts the validation accuracy and loss, in which kernel numbers were set to 32, 64, and 128. The results show that 128 had a better performance than other values. The plot in Fig. 11 shows the vali-

dation accuracy and loss, in which kernel sizes were set in Table 3. The results show that two convolutional layers had a better performance when kernel sizes were set as 5×5 and 3×3 in two convolutional layers, respectively. The plot in Fig. 12 depicts that validation accuracy and loss performed well when the dense size was 128. The plot in Fig. 13 shows that validation accuracy and loss performed well when the dropout was not added in the model. The plot in Fig. 14 shows that accuracy and loss of training, validation, and test performed well, in which all the prediction accuracies were approximately 98%. Hence, the six-layer CNN gives us about 98% prediction accuracy to identify the penetration statuses of the weld pool in the GTAW process.

Conclusion

This paper presents a novel active vision-sensing system based on convolutional neural networks to identify the weld penetration state in the GTAW process.

1) An active laser vision-sensing system was designed to capture the feature information of the weld pool surface. The two reflection modes, which are the convex and concave reflection modes, correspond to different penetration states. These were verified as the fundamentals of the proposed machine learning method.

2) A six-layer CNN model that included two convolutional layers, two pooling layers, and two full-connected layers was designed. Images of the reflected laser stripe from the weld pool were collected and trained to identify the weld joint penetration state. A data argument method was proposed to tremendously expand the dataset for more accurate training.

3) The verification and test results show that the optimized CNN achieved an accuracy of about 98% for identifying the weld penetration state. The proposed CNN model provides a potential active-image-based approach for online penetration monitoring of the GTAW process.

Acknowledgments

This work was supported by the International Research Cooperation Seed Fund of Beijing University of Technology, National Natural Science Foundation of China (grant no. 51975014), and Beijing Municipal Natural Science Foundation (grant no. 3192004).

References

1. Li, P. J., and Zhang, Y. M. 2001. Precision sensing of arc length in GTAW based on arc light spectrum. *Journal of Manufacturing Science and Engineering* 123(1): 62–65. DOI: 10.1115/ 1.1349719

2. Alfaro, S. C. A., and Franco, F. D. 2010. Exploring infrared sensoring for real time welding defects monitoring in GTAW. *Sensors* 10(6): 5962–5974. DOI: 10.3390/s100605962

3. Lv, N., Xu, Y., Zhong, J., Chen, H., Wang, J., and Chen, S. 2013. Research on detection of welding penetration state during robotic GTAW process based on audible arc sound. *Industrial Robot* 40(5): 474–493. DOI: 10.1108/IR-09-2012-417

4. Zhang, S., Hu, S., and Wang, Z. 2016. Weld penetration sensing in pulsed gas tungsten arc welding based on arc voltage. *Journal of Materials Processing Technology* 229: 520–527. DOI: 10.1016/ j.jmatprotec.2015.09.034

5. Feng, Z., Chen, J., and Chen, Z. 2017. Monitoring weld pool surface and penetration using reversed electrode images. *Welding Journal* 96(10): 367-s to 375-s.

6. Huang, Y., Wu, D., Lv, N., Chen, H., and Chen, S. 2017. Investigation of porosity in pulsed GTAW of aluminum alloys based on spectral and x-ray image analyses. *Journal of Materials Processing Technology* 243: 365–373. DOI: 10.1016/j.jmatprotec.2016.12.026

7. Li, C., Shi, Y., Du, L., Yufen, G., and Zhu, M. 2017. Real-time measurement of weld pool oscillation frequency in GTAW-P process. *Journal of Manufacturing Processes* 29: 419–426. DOI: 10.1016/ j.jmapro.2017.08.011

9. Zhang, W., Wang, X., and Zhang, Y. 2013. Analytical real-time measurement of a three-dimensional weld pool surface. *Measurement Science and Technology* 24(11): 1–18. DOI: 10.1088/0957-0233/24/ 11/115011

10. Shi, Y., Zhang, G., Ma, X. J., Gu, Y. F., Huang, J. K., and Fan, D. 2015. Laser-vision-based measurement and analysis of weld pool oscillation frequency in GTAW-P. *Welding Journal* 94(5): 176-s to 187-s.

11. Liang, Z., Chang, H., Wang, Q., Wang, D., and Zhang, Y. M. 2019. 3D reconstruction of weld pool surface in pulsed GMAW by passive biprism stereo vision. *IEEE Robotics and Automation Letters* 4(3): 3091–3097. DOI: 10.1109/LRA.2019.2924844

12. Zhang, K., Zhang, Y. M., Chen, J. S., and Wu, J. S. 2017. Observation and analysis of three-dimensional weld pool oscillation dynamic behaviors. *Welding Journal* 96(5): 144-s to 153-s.

13. Lecun, Y., Bengio, Y., and Hinton, G. 2015. Deep learning. *Nature* 521(7553): 436–444. DOI: 10.1038/nature14539

14. LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. 1998. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86(11): 2278–2323. DOI: 10.1109/5.726791

15. Krizhevsky, A., Sutskever, I., and Hinton, G. E. 2017. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* 60(6): 84–90. DOI: 10.1145/3065386

16. Zeiler, M. D., and Fergus, R. 2014. Visualizing and understanding convolutional networks. *European Conference on Computer Vision ECCV 2014* 8689: 818–833. DOI: 10.1007/978-3-319-10590-1_53

17. Simonyan, K., and Zisserman, A. 2015. Very deep convolutional networks for large-scale image recognition. *3rd International Conference on Learning Representations ICLR 2015 — Conference Track Proceedings*: 1–14.

18. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. 2015. Going deeper with convolutions. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition:* 1–9 . DOI: 10.1109/CVPR. 2015.7298594

19. He, K., Zhang, X., Ren, S., and Sun, J. 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*: 770–778. DOI: 10.1109/CVPR.2016.90

20. Li, X. R., Shao, Z., Zhang, Y. M., and Kvidahl, L. 2013. Monitoring and control of penetration in GTAW and pipe welding. *Welding Journal* 92(6): 190-s to 196-s.

21. Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *International Conference on Learning Representations*: 1–15.

YONGCHAO CHENG, JUN XIAO (jun.xiao@bjut.edu.cn), *and SHUJUN CHEN are with the Engineering Research Center of Advanced Manufacturing Technology for Automotive Components, Ministry of Education, College of Mechanical Engineering and Applied Electronics, Beijing University of Technology, Beijing, China. QIYUE WANG, WENHUA JIAO, and YUMING ZHANG are with the Department of Electrical and Computer Engineering, University of Kentucky, Lexington, Ky.*