Development of Image Sensing Technology for Automatic Welding (Image Recognition by Deep Learning)

Tsuyoshi ASHIDA*1, Akira OKAMOTO*2, Keita OZAKI*2, Masatoshi HIDA*2, Dr. Takayoshi YAMASHITA*3

*1 AI Promotion Project Department, Technical Development Group
*2 Production Systems Research Laboratory, Technical Development Group
*3 Department of Computer Science, College of Engineering, Chubu University

A system has been developed for automatic MAG welding with ceramic backing. This system comprises a camera to capture the images of the molten pool for recognizing feature points to control the torch. A regression-based deep convolutional neural network (DCNN), which outputs continuous values from image inputs, was used to recognize feature points such as arc center and the leading end of the molten pool. This has enabled the accurate recognition of the distance from the arc center to the leading end of the molten pool, as well as the width of the molten pool, with an average error of 0.44 mm or less. The formation of a proper back bead has been confirmed in a welding experiment on a test piece with a tapered gap (from 3 to 10 mm).

Introduction

In the field of production, automatic welding technology using robots has become indispensable in recent years to improve productivity to strengthen cost competitiveness and to compensate for the shortage of welders due to the aging of skilled welding technicians. However, there are many welding tasks that are difficult to automate due to the restrictions of apparatus and skills; these tasks still rely on skilled welding technicians. MAG penetration welding with ceramic backing material is one such task.

MAG welding is a type of gas shield arc-welding method and involves high electric current density to allow a large electric current to pass through thin wires. This makes it possible to deeply melt the welded portion to increase the strength. The ceramic backing material, if used, shields the back side and holds the weld bead, which results in the formation of a proper back bead to stably ensure high strength of the welded portion. To form a proper back bead is referred to as "To form a penetration bead." This requires controlling welding while watching the status of the molten pool and the arc, which can only be dealt with by skilled welding technicians and has not hitherto been automated.

Recently, in the field of general image recognition, a technique called "deep learning" has enabled the recognition of humans and objects, as well as the estimation of their positions, with high accuracy. This deep learning technology was applied to MAG penetration welding with ceramic backing material for automation, as described in this paper.

1. MAG penetration welding with ceramic backing material

This paper focuses on butt welding with a V groove that has a gap varying from 3 to 10 mm, to which a ceramic backing plate is attached (**Fig. 1**). During the welding of the first layer, the state of the molten pool changes, depending on−among other factors−fluctuations in the gap width and in the groove angle, and the mounting condition of the backing material. Hence, the existing robots of playback systems encounter arc interruptions and/or joint penetration failures, and they often are unable to yield proper back beads. In order to form a sound back bead, it is necessary to properly maintain the formation of the molten pool and the state of the arc. To this end, parameters such as the electric current, voltage, wire feed rate, and torch/electrode manipulation must be sensed and controlled in real time. In this development, the following control policy has been laid down for stable welding:

(1) The torch speed is controlled so as to keep the arc near the leading end of the molten pool.
(2) Left-right control is performed to hold the torch in the center of the gap width to prevent the weld line from deviating.
(3) When the gap width exceeds a certain value, weaving is initiated in accordance with the gap width.

In order to realize the above control, the coordinates of the leading end of the molten pool, arc center, etc., (feature points) were extracted from

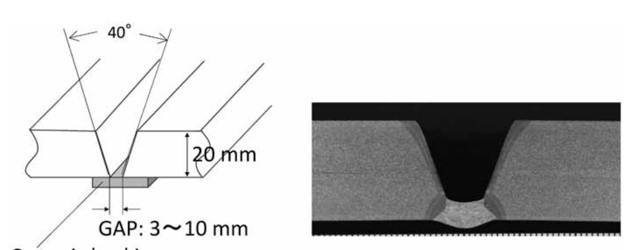


Fig. 1 MAG penetration welding with ceramic backing material

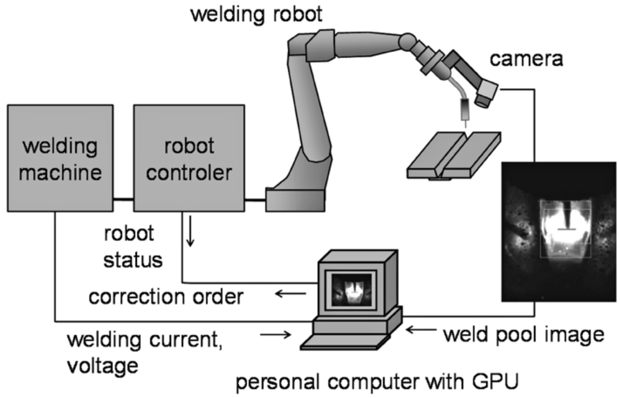


Fig. 2 System Configuration

the camera images capturing the state of welding.

2. Overall system configuration

The overall configuration of the system is shown in **Fig. 2**. The camera is disposed so as to capture the arc and molten pool from a position diagonally forward of the torch. The images are imported into a PC at the rate of 50 fps, and the feature points of the image are calculated every 20 ms. The image feature quantity, however, varies randomly for each image; hence, the past 20 images are averaged to calculate the average feature quantity of the images to determine the amount of control correction (for the speed command, torch left-right command, and weaving width command). The amount of control correction is sent to the robot controller every 200 ms for the welding motion of the robot. Since the calculation of the image feature quantity is based on the average image feature quantity of the last 20 images captured every 20 ms, the amount of control correction for every 200 ms comes to a moving average value of 400 ms.

3. Feature point detection of molten pool based on deep learning

The conventional approach to extracting the leading end of the molten pool, and other features, by image processing generally involved human engagement in combining image feature quantities such as edge and labeling to construct the extraction logic. The shapes and appearances, however, change with alterations in the gap width and welding conditions, even for joints having an identical shape; and each time, human engagement has been necessary to make additional modifications to the image processing logic. Hence a decision has been made to use the regression-based Deep Convolutional Neural Network (hereinafter referred to as "DCNN") in extracting feature points such as the positions of the leading end of the molten pool and of the arc center.[1), 2)] The DCNN has been used for the estimation of coordinates for, e.g., facial organ detection and human posture estimation and has achieved high recognition accuracy.

3.1 Regression-based DCNN

The process flow is shown in **Fig. 3**. The network consists of a grayscale-image entry, four each of convolution layers, batch normalization layers and pooling layers, followed by three fully-connected layers. As shown in **Fig. 4**, the final fully-connected layer outputs values with a total of 10 points, including the coordinate values of the arc center, wire tip, left & right end points of the molten pool (indicated as "edge of pool lead (X,Y)" in the figure), and left & right ends of the molten pool (indicated as "edge of pool (X)", etc., in the figure). In the convolution layer, convolution calculation is carried out with the weight for a filter size of N×N to generate the convoluted value, u. The size of

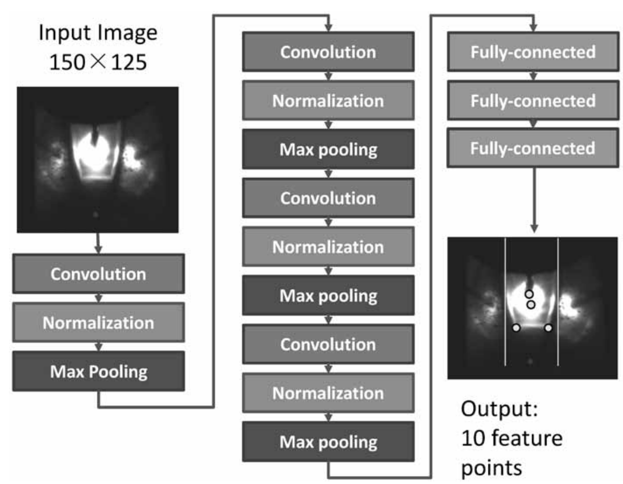


Fig. 3 Regression-based Deep Convolutional Neural Network

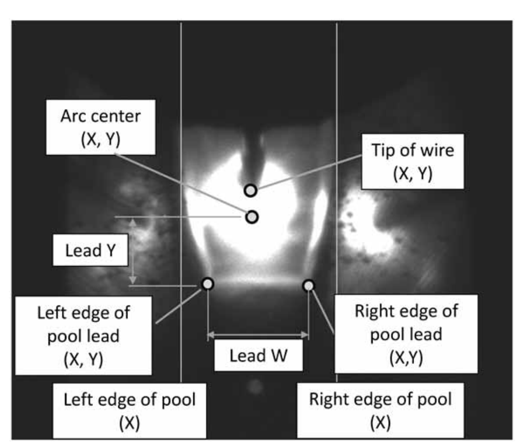


Fig. 4 Feature points of molten pool

the filter used this time is 3×3. Next, an activation function is applied to the value, u.

A sigmoid function, hyperbolic tangent function, or a rectified linear unit (ReLU) is generally used for the activation function. The present system employs the ReLU for the activation function. The ReLU is a function that returns 0 when the input value of u is negative and returns the value of u as-is when it is positive. In the convolution layer and the batch normalization layer which follows the activation function, normalization is performed so that each element of the feature map of a mini batch unit, entered upon learning, has an average of 0 and distribution of 1.[3] This reduces the influence of the spots with high brightness values, in particular the arc light, so that the output values can be obtained taking into account the molten pool and welding target plates with relatively low brightness values. In the pooling layer, processing is performed to shrink the feature map that has passed through the batch normalization layer. The pooling techniques include an average value pooling and a maximum value pooling. The present development adopts the maximum value pooling of 2×2 size, which is commonly used for general object recognition. In the fully-connected layer, the input feature map is transformed to one dimension to generate the input value to the next fully-connected layer. In the fully-connected layers, a technique for suppressing overlearning, called dropout,[4] is adopted, in which the probability of dropout is set to $p=0.5$. The ten output values of this fully-connected layer correspond to the ten feature points to be forecasted.

3.2 Learning data and data expansion

Learning data was inputted for the molten pool image of welding, taken in advance, while focusing on the coordinate values, visually determined, of the arc center, wire tip, left & right end points of the molten pool, and left & right ends of the molten pool (these coordinate values are hereinafter referred to as "feature points"). These learning data were inputted from 2,400 images of welding performed 12 times under various conditions.

In deep learning, high forecasting accuracy can be obtained for learning data; however, the forecast accuracy may decline for non-learning test data. One of the factors that cause this decline in accuracy is the difference in appearance between learning data and test data. For example, there may be a case where the misalignment of the camera or the setting position of the recognition target causes a difference between the learning data and test data in the position of the recognition target in an image. There may be another case where the difference in the distance between the camera and the recognition target causes a difference in the size of the recognition target in an image.

In order to suppress the decline of forecast accuracy due to this factor, there is a method of collecting a sufficient amount of data with different appearance and adding it to the learning data. However, this requires a number of experiments and human designation of the positions of feature points for all the images captured. Hence, data expansion was performed to increase the accuracy of recognition based on limited learning data.[5] The data expansion is a technique of providing learning data with the image changes, such as parallel shift, horizontal/vertical inversion, and scaling, and adding the changes to the learning data. This time, the data expansion was carried out by randomly adding a mirror surface inversion to the horizontal direction, parallel shifts by 5, 10, 15, 20 pixels in 8 directions of every 45° angle, and scaling from 0.8 to 1.2 times in each longitudinal/lateral direction at an increment of 0.1. On the basis of the above, learning was carried out on 12,000 data, including the original 2,400 learning data and 9,600 expanded data.

4. Evaluation experiment

In order to verify the validity of the present system, the errors in the values forecasted by the regression-based DCNN were evaluated while carrying out welding experiments.

4.1 Error evaluation of forecast values

In order to evaluate the errors of the values forecasted by the regression-based DCNN, a comparison was made with feature points extracted by the conventional image processing technique. As shown in **Fig. 5**, on the basis of the conventional image processing technique, the left & right lower edges, which constitute the contour of the molten pool, were detected and fitted into curves to identify their intersections with the leading edge, the intersections referred to as the left & right end points (indicated as "edge of pool lead") of the molten pool.

For test data, apart from learning data, the distance, Lead Y, from the arc center to the leading end of the molten pool, and the distance, Lead W, between the left & right ends of the molten pool (**Fig. 6**) were calculated to compare the techniques. These, Lead Y and Lead W, are important values used to control welding.

Fig. 7 compares the average errors, with respect to the coordinates visually selected from the images

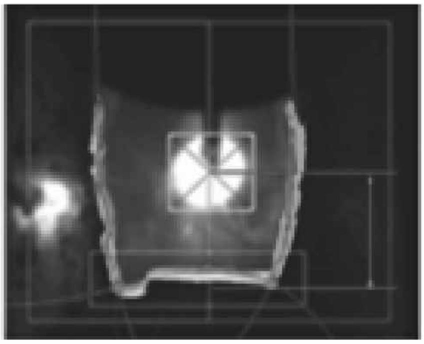


Fig. 5 Edge detection by curve fitting

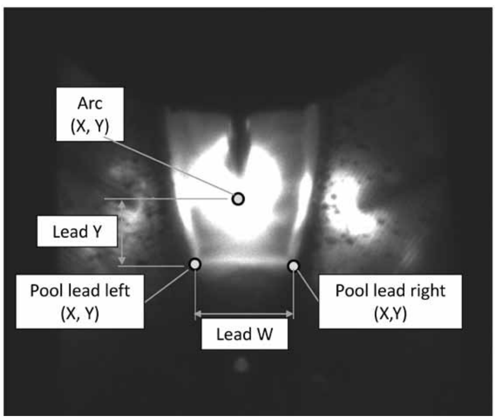


Fig. 6 Feature points Lead Y and Lead W

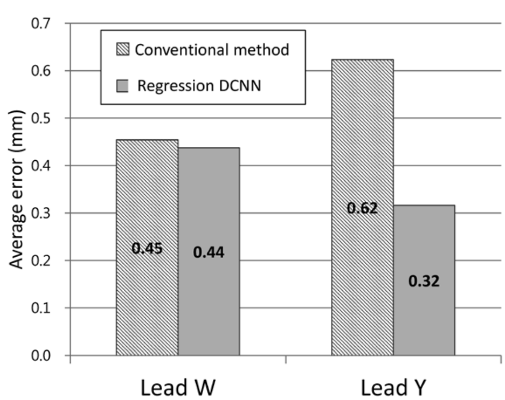


Fig. 7 Average error in coordinates of feature points

of the molten pool, between the coordinates of the feature points extracted by regression-based DCNN, as well as with those extracted by conventional image processing. For Lead W, the same level of errors was confirmed between the conventional image processing and regression-based DCNN, while, for Lead Y, an improvement of 0.3 mm was confirmed in comparison with the conventional image processing. This improvement is attributable to the fact that the conventional image processing erroneously recognizes spatter as the leading end of the molten pool, while regression-based DCNN was able to accurately recognize the leading end of the molten pool.

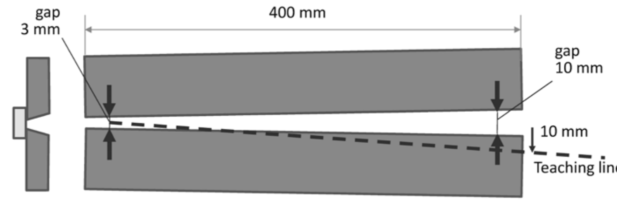
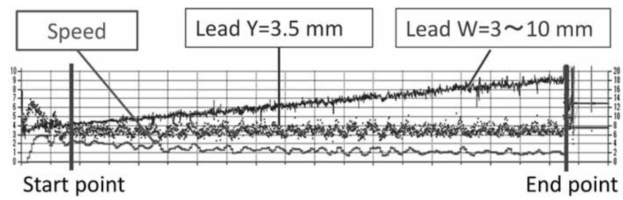


Fig. 8 Test piece



Movie

Fig. 9 Changes of control parameters

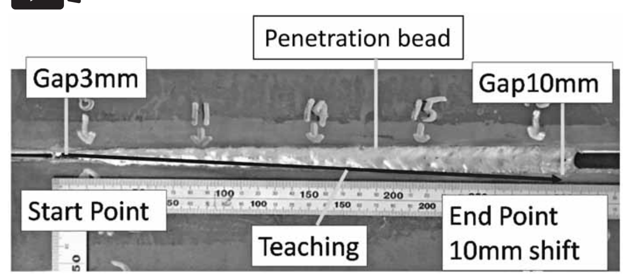


Fig.10 Experimental result (appearance of back bead)

4.2 Welding experiment

To verify the validity of the present system, a welding experiment was conducted, in which a test piece with a tapered gap having a varying groove width (3→10 mm) was disposed at an offset of 10 mm to the right with respect to the final teaching point of welding. The welding was carried out using flux-cored wire (FCW) under the conditions of electric current 200 A, voltage 25 V, and CO_2 gas. The shape of the test piece is shown in **Fig. 8**, and the system configuration is shown in Fig. 2. The images of the state of welding were captured by the camera attached to the torch tip, while the feature points were extracted by PC processing to calculate Lead W and Lead Y, and these values were used as the basis for determining the control variables such as welding speed. **Fig. 9** shows the change in the welding speed, Lead W, and Lead Y from welding start to finish, and **Fig.10** shows the back bead. As shown in Fig. 8, the distance, Lead Y, from the arc center to the leading end of the molten pool is kept at a constant value. The distance, Lead W, between the left & right end points of the molten pool has the same value as the gap width of the workpiece, showing that this gradually increasing width is recognized successfully. As the gap width increases, the pooling of molten metal slows down and the welding speed must be slowed down accordingly.

It is shown also that the welding speed is gradually slowed down to keep Lead Y constant. Fig.10 verifies that a proper back bead is formed regardless of whether there is a misalignment to the right or left, or a gap fluctuation.

Conclusions

An elemental technology for automation was developed for "MAG penetration welding with ceramic backing material." Automatic welding was conducted to control the torch on the basis of the state recognition achieved by deep learning of the molten pool images captured by a camera. As a result, it has been confirmed that a proper back bead can be formed, although still in a limited environment. In the future, we will expand the target joints for deep learning and improve the versatility of molten pool recognition.

References

1) T. Ashida et al. *Preprints of the National Meeting of JWS*, Vol.101. 2017, pp.450-451.
2) T. Yamashita et al. "Facial Point Detection Using Convolutional Neural Network Transferred from a Heterogeneous Task," *International Conference on Image Processing*. 2015.
3) S. Ioeff et al. "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *ICML*.
4) T. Okatani. Deep learning. KODANSHA LTD. 2015, p.30.
5) T. Okatani. Deep learning. KODANSHA LTD. 2015, p.35.