Prediction of Weld Penetration Using Dynamic Weld Pool Arc Images

An improved approach correlates weld penetration with fused images through a convolutional neural network

BY W. JIAO, Q. WANG, Y. CHENG, R. YU, AND Y. M. ZHANG

ABSTRACT

This work aims to study an improved method to predict weld penetration that is not directly observable during manufacturing but is critical for the integrity of the weld produced. Previous methods used signals acquired at a time, typically a single image or multiple images/signals from the process, to derive the penetration at that given time. Although deep learning appears to extract data well, analyses of weld pool physics, previous studies, and skilled weld operation all suggest that the dynamic welding phenomena give a more solid mechanism to assure the adequacy of the needed information. Therefore, this paper proposes to fuse the present weld pool arc image with two previous images, acquired $\frac{1}{6}$ and $\frac{2}{6}$ s earlier. The fused single image thus reflects the dynamic welding phenomena. Due to the extraordinary complexity, the weld penetration is correlated to the fused image through a convolutional neural network (CNN). Welding experiments have been conducted in a variety of welding conditions to synchronously collect the needed data pairs to train the CNN. Results show that this method improved the prediction accuracy from 92.7 to 94.2%. Due to the critical role of weld penetration and the negligible impact on system/implementation, this method represents major progress in the important field of weld penetration monitoring and is expected to provide more significant improvements during welding using pulsed current, where the process becomes highly dynamic.

KEYWORDS

- Weld Penetration
- Fused Single Image
- Early Fusion
- Convolutional Neural Network (CNN)
- Weld Pool Arc Image
- Dynamic Weld Phenomena

Introduction

Welding has been automated/robotized greatly. However, in typical automated/robotic welding applications, the welding parameters are preset and not adjusted adaptively to overcome the effect from unpredicted disturbances. This imperfection cannot meet the increasing requirements from the welding/manufacturing industry on quality, efficiency, and flexibility. Combining information sensing/processing with traditional welding manufacturing techniques has been a major directive to revolutionize the welding industry (Ref. 1). In practical welding, the weld penetration, as measured by the back-side weld bead width, is a critical factor determining the integrity of the weld produced. However, the back-side bead width is difficult to monitor directly during manufacturing because it occurs underneath the surface of the workpiece being processed. Therefore, predicting the back-side bead width using conveniently sensible information from the welding process becomes a fundamental issue in intelligent welding.

Many studies have been done to predict the weld penetration using different characteristic information from the welding process. They typically 1) sense observable phenomena from the welding process using, or based on, different sensors/phenomena such as infrared, pool oscillation, laser ultrasonic, and active vision methods (Refs. 2–5); 2) define and extract characteristic features from sensed phenomena; and 3) build a model to correlate the extracted characteristic features to the penetration state (Refs. 6, 7). However, the characteristic features are proposed subjectively based on the individual’s understanding of the physics, thus lacking a systematic way to ensure success in leading to a good model. Iteration is often needed such that the development efficiency is low. To address this general challenge, researchers recently started to apply deep-learning-based methods to extract the information automatically. Therefore, the major remaining challenge is reduced to acquiring adequate information from the welding process.

Skilled welders can judge the weld penetration per their observed welding phenomena during the process. The welding community believes that images from the observable welding scene, including the 3D weld pool surface, contain sufficient information to predict the weld penetration (Ref. 8). While earlier efforts followed the aforementioned procedure to first propose characteristic features, the deep learning method has recently been applied, with a concentration on using convolutional neural networks (CNNs), to directly map images to the penetration (Refs. 9–14). The training for the parameters, including the convolutional kernels and

https://doi.org/10.29391/2020.99.027
weights in fully connected layers as well as the feature extraction and reasoning, is done automatically. However, these images of the observable weld scene and the weld penetration state are collected synchronously for use in training CNN models. There is an implied assumption that the current weld pool/scene can fully determine the weld penetration. The temporal dynamic information may have been missed, and the operation of skilled welders may have been interpreted incompletely. In a previous effort, the temporal dynamic weld pool surface has been used to predict the weld penetration using an adaptive neuro-fuzzy inference system model (Ref. 15).

Deep learning facilitates an appreciated automatic process to directly map images as original process information to the process outputs without tedious, trial-and-error, hand-crafted feature extraction. However, the information must be adequate; although, it may be redundant and complex. For the prediction of weld penetration, a possible inadequacy in temporal dynamic information needs to be addressed. Because weld pool arc images are easier to measure than 3D weld pool surfaces, this paper proposes to fuse three consecutive images, each taken \( \frac{1}{6} \) s apart, into a composite image as the input of a CNN to predict weld penetration.

### System Configuration

As shown in Fig. 1, a weld torch was set vertically in the experimental platform with a rectangular hole directly below to facilitate the camera and observe the workpiece backside information. A passive vision system (camera 1) observed the weld scene and acquired the images to be fused to form the input of the CNN. A second camera observed the back side of the workpiece to obtain the actual weld penetration to be predicted. The welding torch and two cameras were set stationary, and the workpiece was moved by a stepper motor. This design ensured that the cameras could always capture high-resolution paired images from the top side and back side of the workpiece. To eliminate the strong interference from the arc radiation, camera 1 (Point Grey FL3FW03S1C) used a 685-nm, center-wavelength bandpass optical filter. The same filter was also used by camera 2 for the same reason. The configurations for both cameras are shown in Table 1. The operations of the system and sensors were controlled by the same computer, in particular, con-

---

**Fig. 1 — Welding experiment platform.**

**Fig. 2 — Weld pool evolution: A — Incomplete joint penetration; B — complete joint penetration; C — excessive penetration.**

**Table 1 — Camera Configuration Applied**

<table>
<thead>
<tr>
<th>Configuration</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Filter Center (nm)</td>
<td>685 ± 2</td>
</tr>
<tr>
<td>Filter FWHM (nm)</td>
<td>10 ± 2</td>
</tr>
<tr>
<td>Image size (pixel)</td>
<td>480 × 640</td>
</tr>
<tr>
<td>Format</td>
<td>Mono8</td>
</tr>
<tr>
<td>Frame rates (FPS)</td>
<td>30</td>
</tr>
<tr>
<td>Shutter time (s)</td>
<td>0.03</td>
</tr>
<tr>
<td>Sharpness</td>
<td>3000</td>
</tr>
<tr>
<td>Gain</td>
<td>0</td>
</tr>
<tr>
<td>Gamma</td>
<td>2.5</td>
</tr>
</tbody>
</table>
trolling the cameras, collecting the images with a peripheral component interconnect (PCI) expansion card (IEEE 1394), and controlling the welding power supply and stepper motor using a data acquisition card (PCI 6229).

Weld Penetration Analysis

Welding Process Analysis

When welding started, the solid metal melted and the liquid metal expanded due to thermal expansion, such that the surface started to rise as shown in Fig. 2A. Under the continuous application of the heat from the arc, the volume of the liquid weld pool increased. As a result, the surface area of the weld pool increased, and the penetration depth also deepened until the liquid weld pool completely penetrated the solid metal. The width of the back-side weld bead (Wd) increased after the solid metal was fully melted as shown in Fig. 2B, and Wd was the key parameter to quantify the penetration state. The weld integrity requirement was considered satisfied when Wd reached a certain width in actual production. If the liquid weld pool continuously increased, the back-side weld pool surface would become more convex due to gravity. When the convex volume was greater than the thermal expansion volume, the top-side surface became concave as shown in Fig. 2C. It is thus clearly suggested that the top- and back-side weld pool must follow this physical relationship, and the use of the observable weld scene from the top side to predict the back-side bead width is justifiable. In addition, the dynamic evolution of the top-side weld scene is even more informative and can better correlate with what occurs on the back side.

We note that welders can see not only the surface of the weld pool but also the welding arc. Until recently, the arc was regarded as a strong light interference affecting the effective processing of images to extract more critical information, such as the weld pool boundary. Efforts were thus taken to filter the arc out. Possible information from the arc relevant to the development of the weld pool was ignored. In fact, because the heat is directly applied by the arc, the arc cone directly affects the energy density distribution and the heated area, which is related to the direction and speed of the weld pool extension.

As shown in Fig. 3, as the arc voltage increased, the arc length became longer, causing the arc cone to widen and the arc heating area to become broader. Conversely, when the arc voltage decreased, the arc length shortened, the arc cone became narrower, and the heating area was more focused (Ref. 16). The surface of the weld pool will constantly change with the development of the weld pool, and the deformed surface also affects the arc cone. As a result, we collected the top-side image, which included both the arc and weld pool as the raw information of the relevant welding phenomena.

While the complex raw information increases the redundant information, it increases the difficulty in extracting the characteristic information and computational complexity. Fortunately, the deep-learning-based, data-driven approach may thus be a right choice. In this paper, we further designed an image sequence dataset to increase the temporal information by a CNN directly without extracting image features first. Unfortunately, the harsh imaging conditions may thus be a right choice. In this paper, we further designed an image sequence dataset to increase the temporal information by a CNN directly without extracting image features first. In addition, camera noises may corrupt individual images, but such noises can be suppressed using an image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17). In addition, camera noises may corrupt individual images, but such noises can be suppressed using an image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17).

An image sequence contains more information than any of its constituent images. For this reason, image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17). In addition, camera noises may corrupt individual images, but such noises can be suppressed using an image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17).

Image Sequence

An image sequence contains more information than any of its constituent images. For this reason, image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17). In addition, camera noises may corrupt individual images, but such noises can be suppressed using an image sequence analysis has been used in computer vision for quite a long time. Especially for relatively moving objects, analyzing image sequences is very effective for object detection (Ref. 17).

As shown in Table 2, the characteristics of the different welding processes are: dark chroma, low contrast, and relatively slow development of the weld pool and arc. The top-side images in the same process thus had a high degree of similarity. What's more, in the welding process, the characteristic information was dynamic while the surrounding environment was static, as can be seen from the sequence of the top-side images in Fig. 4. The development of the weld pool and arc cone can be observed from this sequence, but the harsh imaging conditions made extracting image features a challenge. Processing the image by a CNN directly without extracting image features first may thus be a right choice. In this paper, we further designed an image sequence dataset to increase the temporal information in the raw data, which enabled the CNN to extract dynamic characteristics from the process.
The main connectivity pattern category for fusing temporal information through the CNN includes early fusion, late fusion, and slow fusion (Ref. 19). The early fusion is to fuse multiple features that are extracted from early layers in the CNN and then use the fused features to train a predictor (later layers). The late fusion uses separate networks to predict and then fuse the prediction results from these networks together. The slow fusion model needs a much larger dataset and computation time, which makes it difficult to predict the penetration states in real time. The goal of our design was to find the hidden dynamic features in the feature space so the neural network could use the dynamic features to predict the weld penetration status. The early fusion appeared to be more appropriate than the other two choices for our application.

**Dataset and Model**

A series of experiments were designed to perform spot welding using the gas tungsten arc welding (GTAW) process. Each 1.85-mm-thick stainless steel sheet was welded in 12 spots with a 2-cm distance between them. For completeness and reasonableness of the experiments, the range of welding currents and time were designed to simulate a wide range of conditions. The detailed parameters are shown in Table 2. During each experiment, cameras captured the images of the entire welding process from incomplete joint penetration to excessive penetration, with 28,494 image pairs collected as raw data from more than 120 experiments. Image pairs were segmented into training, validation, and test datasets with sizes of 22,794, 2849, and 2851, respectively.

**Image Sequence Design**

In the typical CNN model, the input images are passed through the network one by one. The image generator yields N, W, H, C data, where N is the batch size, W and H are the width and height of the images, and C is the number of channels (three for RGB images and one for grayscale images). The...
image sequence dataset needs several frames in a sequence, and the data dimension becomes $N, W, H, C, F$, where $F$ is the number of frames in a sequence. In our system, the frame rate of the camera was 30 frames per second, so we set every sequence to be $\frac{1}{3}$ s, which had ten frames (grayscale images), and the stride was set to one to prevent missing critical information. In our application, every frame in the sequence was not equally important as it was in other applications, such as motion analysis and video classification (Refs. 20, 21). We picked the last image (image 10), the image in the middle of the sequence ($\frac{1}{6}$ s earlier or image 5), and the first image (image 1) in the sequence as the samples from the sequence, which effectively reduced the computation costs of redundant information. The sampled sequence was used as the input of the CNN.

The image sequence design is shown in Fig. 5. This design can effectively control the size of the input data, and the neural network can control the stride size to gradually fuse dynamic information from the whole image sequence.

### Labels

The back-side images were used as the penetration state labels for training the CNN model. The penetration state is defined by the range of the real back-bead width. However, it is not possible to be measured simultaneously with the top-side images in real time during the welding process. Therefore, we calculated the back-bead width value through a model that correlates the back-side image to the real back-bead width. To this end, we first defined a bright area using a threshold and then calculated it from the back-side image as shown in Fig. 6. In this paper, we set the threshold at 170. Then, the relationship between the area and the actual back-bead width was established through calibration experiments. Since the square of the actual back-bead width was proportional to the area (number of pixels), we proposed the following formula:

$$\text{Width} = k \times \sqrt{\text{Area} - \text{Error}} \quad (1)$$

where $k$ and $\text{Error}$ are unknown parameters determined by least squares approximation, and the fitting curve is shown in Fig. 7. The coefficient of determination is 0.9879, which indicates that the regression predictions well approximate the real data.

With Equation 1, the back-bead width was calculated from the back-side images, and the penetration state was classified by the back-bead width into three categories: under penetration, desirable penetration, and excessive penetration. Table 3 shows the detailed classification criteria used in this paper. Such penetration states from the back-side images were used as labels for CNN model training.

### CNN Model

Early fusion can immediately expand the entire time win-
dow to fuse information at the pixel level. In the first convolutional layer of an early fusion, the kernel size will extend through the full depth of the input volume. This is achieved by modifying the filters of the first convolutional layer in the single-frame model by expanding the size to be $W \times H \times (C \times F)$. The designed image sequence can thus be merged into three channels, and the neural network can easily find the pixel changes at the same location by the same receptive field. This matches with our design sequence data. The network can then accurately detect local motion directions and speeds (Ref. 19). With the early fusion method, the proposed CNN model is shown in Table 4, which includes three convolutional layers (Convs), three max-pooling layers (Pools), and three fully connected layers (FCs). The CNN early fusion takes $480 \times 640$ and three-channel images as input, and its Conv1 has 32 kernels of $5 \times 5$ to fuse the temporal information at the pixel level by convoluting over the three-channel raw images. A $3 \times 3$ max-pool layer follows the Conv1 to reduce the computation cost of the whole neural network. Three sequential grayscale inputs are then due to the inputs. The feature information has brighter pixels in the grayscale images, making max pool useful, as it is when the background of the images is dark and the target information has lighter pixels. On Pool1, the max pooling is used to extract the extreme features and give the neural network robustness to position variation. The similar processes with different kernel size and kernel number are applied on Conv2, Pool2, Conv3, and Pool3, respectively, to further extract the features. The CNN model is a tradeoff between speed and accuracy; the $5 \times 5$ kernel is used in Conv1 to better fuse information from three-channel images, and the $3 \times 3$ kernels are used in Conv2 and Conv3 to reduce memory usage and compute faster (Ref. 22). Then, the data is flattened and connected with three FC layers to output the classification of the weld penetration state.

A batch normalization (BN) and rectified linear unit (ReLU) function follow each Conv to improve the performance of the CNN model. BN has the effect of stabilizing the learning process and dramatically accelerating the neural network training speed by reducing internal covariate shift and reducing generalization error (Ref. 23). The ReLU is used as an activation function; compared with other activation functions such as tanh and sigmoid, ReLU can overcome vanishing gradient problems and allows the model to learn faster and perform better (Ref. 24). In addition, the CNN model is trained using a mini-batch back-propagation algorithm. The loss function is set as the cross entropy for this multiclass classification problem, and the Adam optimization is used to reduce the computation cost and improve the convergence rate in the training.

Table 4 — The Architecture of the CNN (Early Fusion)

<table>
<thead>
<tr>
<th>Layer Name</th>
<th>Kernel Size</th>
<th>No. of Filters</th>
<th>Output Size</th>
<th>CNN (Early Fusion)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conv1</td>
<td>$5 \times 5$</td>
<td>32</td>
<td>$476 \times 636$</td>
<td>Batch normalization, ReLU</td>
</tr>
<tr>
<td>Pool1</td>
<td>$3 \times 3$</td>
<td>-</td>
<td>$158 \times 212$</td>
<td>Max pool</td>
</tr>
<tr>
<td>Conv2</td>
<td>$3 \times 3$</td>
<td>64</td>
<td>$156 \times 210$</td>
<td>Batch normalization, ReLU</td>
</tr>
<tr>
<td>Pool2</td>
<td>$3 \times 3$</td>
<td>-</td>
<td>$52 \times 70$</td>
<td>Max pool</td>
</tr>
<tr>
<td>Conv3</td>
<td>$3 \times 3$</td>
<td>128</td>
<td>$50 \times 68$</td>
<td>Batch normalization, ReLU</td>
</tr>
<tr>
<td>Pool3</td>
<td>$3 \times 3$</td>
<td>-</td>
<td>$16 \times 22$</td>
<td>Max pool</td>
</tr>
<tr>
<td>FC1</td>
<td>-</td>
<td>-</td>
<td>1080</td>
<td>-</td>
</tr>
<tr>
<td>FC2</td>
<td>-</td>
<td>-</td>
<td>64</td>
<td>-</td>
</tr>
<tr>
<td>FC3</td>
<td>-</td>
<td>-</td>
<td>3</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 5 — Confusion Matrix

<table>
<thead>
<tr>
<th>Actual Class</th>
<th>Predicted Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1073</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 7 — Fitting curve between the pixel area and back-bead width.

Fig. 8 — Single image data and image sequence data in the validating process.
process (Refs. 25, 26). All training was done on a computer with an Intel® Core™ i7-6700K central processing unit and an NVIDIA® GeForce GTX 1080 graphics processing unit.

Results and Analysis

Figure 8 shows the validating results after each CNN parameter was updated at each epoch during the training. The validation accuracy for the model using the image sequence data fluctuated significantly during the first ten epochs and then slowly stabilized and reached the best value in the 14th epoch, which was 95.8%. However, using a single image, the highest validating accuracy was 93.6%. We note that the corresponding cross-entropy loss did not exactly match: The loss using single image data was less than that using image sequence. The loss value is not a necessary nor a sufficient condition for accuracy in the neural network. The image sequence data increases the size of the input while increasing the temporal information. To prevent running out of memory, we set the batch size at 10, and the batch size of single image data was 40. The large batch size can reduce the loss value due to the relationship between batch size and speed and stability of the learning process (Ref. 27). However, the accuracy achieved using image sequence data was higher than that using single image data, proving that the designed image sequence was effective and met the design requirements. The model producing the highest validating accuracy was saved to predict the test data: The model using image sequence data had an accuracy of 94.2%, while the model using single image data had an accuracy of 92.7%. The testing results are shown in Table 5. Data examination shows that most incorrect predictions occurred in the neighboring classes.

The prediction using the trained model is considered real time. It requires 0.0254 s for the computer to transfer the image sequence to the CNN and for the CNN to output the classification of the weld penetration. It is less than the time for acquiring one image, which is 0.033 s.

Conclusion and Future Work

In this paper, the image sequence was used to include the dynamic welding phenomena to improve the accuracy for weld penetration prediction. The proposed network used the early fusion method to fuse the dynamic information. In comparison with the model trained using the single top-side image data, using the image sequence data had a higher prediction accuracy. We found 1) the designed sequence of top-side images can effectively add temporal information and improve the accuracy of the prediction; 2) the early fusion approach can enable the CNN to find the dynamic information from the sequence data quickly and accurately; and 3) the data-driven and end-to-end prediction is an efficient and simple method to predict the weld penetration state from top-side images in the complex welding process.

The work was performed in laboratory conditions using bead-on-plate experiments without a gap and filler metal for stainless steel. Future work should extend experiments to butt joints with filler metal as well as to other materials to further verify the effectiveness of the proposed method that is based on dynamic weld pool arc images in predicting weld penetration.

References

Dear members of the welding research community,

Note that in each issue of the Welding Journal Research Supplement, we are including Direct Object Identifier (DOI) numbers with each of the papers published in print and online. A direct object identifier is a unique alphanumeric string assigned by a registration agency (we are using Crossref.org) to identify content and provide a persistent link to its location on the Internet. Our decision to begin assigning a DOI for each paper comes directly from a request by the research community.

As part of our obligation to Crossref.org, we are asked to provide DOI numbers, when available, in the references section of papers. So, if you have submitted a paper to the Welding Journal or are planning on submitting a paper, we ask that you update your references to include DOI numbers whenever possible.

Thank you.

Annette Alonso
Publisher, Welding Journal