

DeepWelding: A Deep Learning Enhanced Approach to GTAW Using Multisource Sensing Images

Yunhe Feng , Zongyao Chen , Dali Wang , Jian Chen , and Zhili Feng 

Abstract—Deep learning has great potential to reshape manufacturing industries. In this article, we present DeepWelding, a novel framework that applies deep learning techniques to improve gas tungsten arc welding process monitoring and penetration detection using multisource sensing images. The framework is capable of analyzing multiple types of optical sensing images synchronously and consists of three deep learning enhanced consecutive phases: image preprocessing, image selection, and weld penetration classification. Specifically, we adopted generative adversarial networks (pix2pix) for image denoising and classic convolutional neural networks (AlexNet) for image selection. Both pix2pix and AlexNet delivered satisfactory performance. However, five individual neural networks with heterogeneous architectures demonstrated inconsistent generalization capabilities in the classification phase when holding out multisource images generated with specific experimental settings. Therefore, two ensemble methods combining multiple neural networks are designed to improve the model performance on unseen data collected from different experimental settings. We have also found that the quality of model prediction is heavily influenced by the data stream collection environment. We think these findings are beneficial for the broad intelligent welding community.

Index Terms—Arc welding, deep neural networks, gas tungsten arc welding (GTAW), monitoring and classification, multisource, pix2pix, sensing images.

I. INTRODUCTION

ARC WELDING inspection and quality control are crucial to the safety and structural soundness of a large number

Manuscript received June 30, 2019; accepted August 20, 2019. Date of publication August 27, 2019; date of current version January 4, 2020. This article was supported in part by the US Department of Energy, in part by the Office of Nuclear Energy (Advanced Methods for Manufacturing Program), in part by the Office of Science (Advanced Scientific Computing Research Program), and in part by the AI Initiative at Oak Ridge National Laboratory. Paper no. TII-19-2952. (Corresponding author: Dali Wang.)

Y. Feng and Z. Chen are with the University of Tennessee, Knoxville, TN 37996 USA (e-mail: yfeng14@vols.utk.edu; zchen25@vols.utk.edu).

D. Wang is with the Energy and Environmental Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: wangd@ornl.gov).

J. Chen and Z. Feng are with the Physical Sciences Directorate, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA (e-mail: chenj2@ornl.gov; fengz@ornl.gov).

Color versions of one or more of the figures in this article are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TII.2019.2937563

of metal products, since welding defects such as incomplete penetration, lack of fusion, and porosity left in the structure could become significant threats to the integrity of a welding structure. An experienced welder can use personal sensing ability (e.g., visual observation of the weld pool and sound from the arc) to determine whether a welding defect has occurred [1], but doing so is time-consuming and laborious. Although many welding robots and automation systems have been applied in manufacturing industries to improve productivity and reduce cost, the demand for automatic welding quality detection is still very high [2].

To monitor dynamical arc welding processes, many traditional machine learning algorithms have been developed to extract features from weld pool surface profiles and determine the welding penetration condition [3]. The interference of arc light, one of the major challenges in capturing weld pool images using visual sensors, has usually been reduced by adopting active vision sensing systems equipped with auxiliary light sources [4]. The key features of weld pool three-dimensional (3-D) geometry have then been reconstructed through image segmentation [5] and reconstruction [6]. A recent work [7] proposed a supervised neuro-fuzzy inference model to correlate the 3-D weld pool characteristic parameters to the welder's adjustments. In addition, a passive vision system was developed to monitor the weld pool using an arc light for illumination [8], [9]. To be specific, it allows the weld pool area to be partially visualized with an appropriate camera exposure time. Based on the calculated weld pool geometry, machine learning techniques have been applied to estimate weld penetrations [10]–[12]. Recently, Zhang *et al.* [13] proposed the support vector machine with grid search optimization and cross-validation (SVM-GSCV) model, which combined both weld pool images and arc sound to predict penetrations.

Additionally, there exist multiple studies that take advantage of neural networks to monitor welding processes. In the literature, Andersen *et al.* [14] is one of the earliest works applying artificial neural networks (ANNs) to arc welding processes and monitoring. Lee *et al.* [15] demonstrated that ANNs performed better than multiple regression analyses in back-bead prediction in gas metal arc welding (GMAW). Similarly, another work by Kim *et al.* [16] estimated bead height in robotic arc welding using a neural network. Nagesh and Datta [17] fed numerical variables such as voltage, current, and arc length

into back-propagation neural networks to predict bead geometry and penetration. Ates [18] explored the possibility of predicting GMAW parameters based on ANNs. Wu *et al.* [19] applied the single-hidden-layer feed-forward neural network to extracted keyhole geometry and acoustic signatures to monitor variable polarity plasma arc welding. Ghanty *et al.* [20] used ANNs to estimate weld bead width and depth of penetration from an infrared thermal image of the weld pool.

However, most existing arc weld monitoring methodologies consist of two separate steps (i.e., extracting handcrafted weld pool features explicitly, and correlating extracted features to weld quality) and implies several limitations. First, it is difficult to develop reliable and robust algorithms to extract the desired features from weld pool profiles that present various irregular shapes and uneven surfaces under different welding conditions. Second, useful information, except handcrafted features, may not be found through conventional image processing techniques [21]. In contrast, deep learning algorithms learn complex and implicit features automatically and have demonstrated powerful capabilities in the field of computer vision and pattern recognition. For example, a state-of-the-art neural network [22] became the first to beat humans at image recognition on a large-scale ImageNet data set [23] in 2015. Inspired by deep learning achievements, major manufacturing companies such as GE, Siemens, and Bosch are adopting deep learning to improve all aspects of manufacturing processes, aiming to lower labor costs, reduce product defects, increase production speed, and so on. We also notice that several recent research studies [24]–[26] are exploring the possibility of utilizing deep learning in weld manufacturing.

When deep learning techniques are applied to monitor arc welding processes, the following challenges might arise. First, the domain knowledge is required to determine the quality of weld pool images and annotate weld penetrations during the creation of image data sets. Second, unexpected noises captured in the welding process, such as random light reflection from the dynamic weld pool surface, may cause deep neural networks to make incorrect decisions. Third, different experimental settings could lead to large variations in weld pool images: a challenging situation for image classification via deep neural networks. Finally, it is challenging to choose suitable deep neural networks, as one specific network with distinctive architectures may behave inconsistently over different weld pool images.

In this article, we propose DeepWelding, an end-to-end deep learning enhanced framework for multisource sensing weld pool images, to overcome the challenges mentioned. DeepWelding consists of three steps: weld pool image preprocessing, high-quality image selection, and penetration classification. For each step, we created corresponding data sets, trained neural networks, and proved their effectiveness. Although DeepWelding was applied only to autogenous, bead-on-plate gas tungsten arc welding (GTAW) processes in this article, we think it is flexible enough for other common types of welding procedures (such as metal inert gas welding, GMAW, and flux-cored arc welding) because DeepWelding has demonstrated its capability of handling various types of weld pool images. We deployed five individual

neural networks with heterogeneous architectures to classify the weld penetration and found that they demonstrated inconsistent generalization capabilities when holding out multisource images generated with specific experimental settings. Therefore, we adopted two plurality voting-based ensemble models combining multiple neural networks to improve the classification performance on unseen data collected from different experimental settings. Our contribution can be summarized as follows.

- 1) We designed and developed a DeepWelding framework, which integrates welding image preprocessing, image quality control, and robust classification.
- 2) Unlike most existing solutions that merely handle a single type of weld pool images, DeepWelding incorporates multisource sensing images, namely, the active vision, passive vision, and reverse electrode images (REIs), to monitor welding status simultaneously.
- 3) To the best of our knowledge, we are the first to apply the pix2pix [27], an advanced generative adversarial network (GAN) model, to removing noise in the weld pool images.
- 4) DeepWelding is highly flexible and extensible. For example, the individual components of DeepWelding can be further integrated with other existing welding monitoring solutions to improve their performance.
- 5) We summarized the knowledge we learned when applying DeepWelding to GTAW, e.g., the classification performance of neural networks was heavily influenced by the data stream collection environment, and discussed implications and future directions for DeepWelding.

II. METHODOLOGY

A. Framework Overview

The proposed DeepWelding framework contains three deep learning enhanced phases: 1) crop and remove noise from raw weld pool images; 2) select high-quality preprocessed images for further classification; and 3) classify the weld penetration using deep learning-based ensemble methods. Fig. 1 illustrates the main components of DeepWelding and the associated workflow between the three key phases. First, three types of weld pool images, which are harvested concurrently using different visual sensors, are cropped and further denoised if needed. Then, deep learning based image selectors dichotomize those preprocessed images as high-quality and low-quality images. Only images of high quality are kept and fed to weld classifiers. In the classification phase, for each image type, five independent deep neural networks with heterogeneous architectures are used to categorize images into different penetration classes, such as partial penetration and full penetration. To improve the classification stability and enhance the generalization capability simultaneously, we designed two-level plurality voting ensemble methods (See Ensemble-1 and Ensemble-2 in Fig. 1.) At the first level, three Ensemble-1s combine the five heterogeneous neural networks for active vision images, passive vision images, and REIs, respectively, to produce improved results. At the second level, Ensemble-2 further makes the three Ensemble-1s agree on the final classification of the ongoing weld penetration

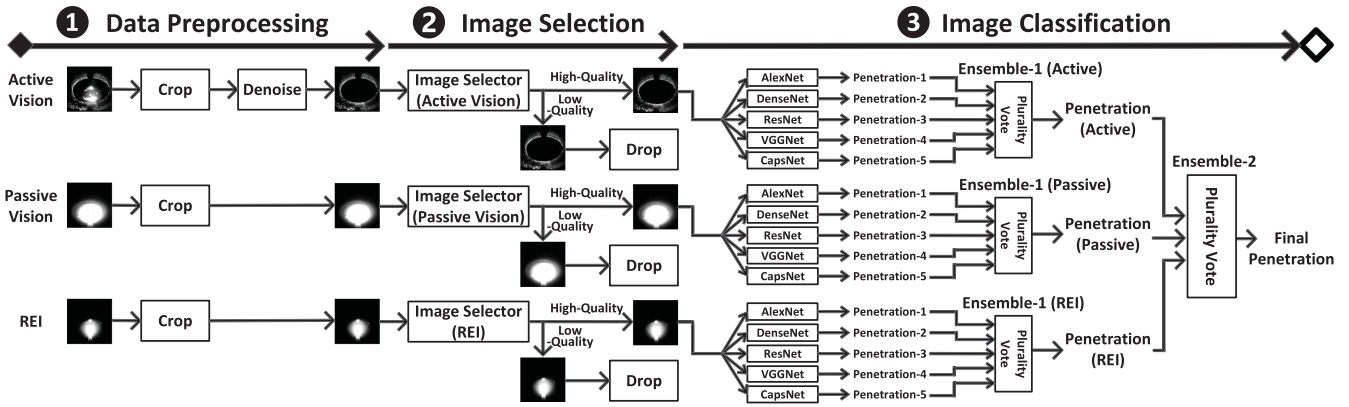


Fig. 1. Framework overview. The framework handles multisource sensing images (active vision images, passive vision images, and REIs) simultaneously using five different neural networks. On each sensing channel, Ensemble-1 concludes a single weld penetration label by aggregating five individual neural networks' results using plurality voting. Considering penetration labels generated by multisource sensing images might be inconsistent, we designed Ensemble-2 to produce the final weld penetration through another plurality voting.

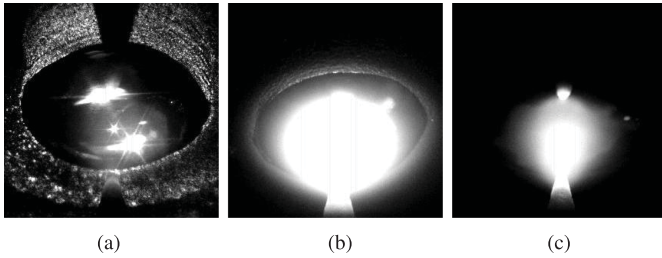


Fig. 2. Examples of collected multisource sensing weld pool images. (a) Active vision. (b) Passive vision. (c) REI.

status at the time when the multisource sensing images were captured.

B. Data Preprocessing

1) Data Description: We collected three types of weld pool images (the active vision, passive vision, and REIs) to describe the weld pool at the same time from different perspectives, as shown in Fig. 2. The principle of image acquisition can be found in our previous work [1], [28]. The active vision image was obtained under laser light illumination that was almost suppressed during data collection. The passive vision image was obtained without any auxiliary light source. The weld pool was partially visualized via arc light by controlling the camera exposure time appropriately. The REI was obtained under a very short exposure time. Compared with the passive vision image, the arc light intensity was reduced. Because of weld pool surface reflections, only the electrode and the REI can be visualized in this type of image.

2) Cropping: Original captured images were usually very large and contained useless pixels (see dark areas around the weld pool). To improve the learning efficiency and focus on the meaningful parts of images, we preprocessed all three types of images. First, we determined the pixel location of the electrode, which was unchanged during the whole experiment and was in

the same location in all three types of captured images when the calibration was done. Then, we used a square bounding box to extract useful information, as shown in Fig. 2.

3) Denoising: When streaming images are collected, unpredicted light reflections may introduce random image noise. For example, the bright zones inside the weld pool in Fig. 2(a) are light reflection noise, which may fool neural networks into learning trivial patterns and making incorrect decisions. Since shapes of weld pools are not always regular circles or ellipses, it would be difficult to locate exact weld pool edges. Moreover, light reflections across weld pool edges make it hard to remove noise based only on the brightness.

We adopted pix2pix [27], a conditional GAN (CGAN) based image-to-image translation model, to remove the light reflection noise. This image-to-image translation model consists mainly of two components, the generator and the discriminator. The former generates synthesized images to fool the discriminator, whereas the latter learns to distinguish between fake and real images. In our experiments, original streaming images were used as inputs and manually denoised images were used as targets to train the discriminator and the generator. Examples of denoised images are shown in Table IV.

C. Image Selection

The extreme environment during the arc welding process, such as the intense arc and surface oxidation, may greatly affect the quality of welding images. Not all captured images are qualified for classification because low-quality images may trick deep neural networks into making inaccurate decisions. The low-quality images obtained from different sources are shown in Fig. 3. In Fig. 3(a), the active vision weld pool image was overwhelmed by the laser light reflection. This condition occasionally occurs as a result of dynamic change in the weld pool surface and the material surface condition of the unmelted area. Note that although meaningless pixels inside active vision images of weld pools are filtered out during data preprocessing, the laser light reflections outside the weld pool may also

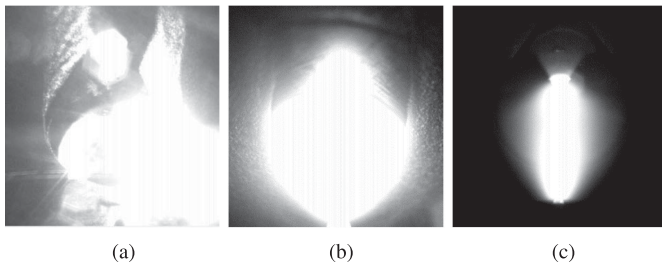


Fig. 3. Low-quality multisource sensing images. (a) Active vision. (b) Passive vision. (c) REI.

TABLE I
COMPARISON OF INVOLVED NEURAL NETWORKS

Name	# Layers	# Parameters	Release Years
AlexNet	8	~60M	2012
VGGNet	16	~138M	2014
ResNet	50	~26M	2015
DenseNet	121	~8M	2016
CapsNet	3	~7M	2017

demonstrate different patterns that cause incorrect classifications. For passive vision weld pool images, the image quality is associated with the amount of light reaching the optical sensors, which can be controlled by the camera exposure time and aperture. The passive vision weld pool image shown in Fig. 3(b) was overwhelmed by the arc light because of the long exposure time and large aperture. The REI shown in Fig. 3(c) was unqualified because the center area was saturated and connected with the REI area.

Therefore, for different types of weld pool images, we designed corresponding image selectors to justify their qualities using the convolutional neural network [29]. We trained a biclassification neural network to distinguish between good and bad images. The bad images were directly dropped, while the good images were fed forward.

D. Classification

The classification phase in Fig. 1 mainly consists of two components, i.e., individual classifiers and ensemble models. Individual classifiers are responsible for weld penetration predictions. Ensemble models summarize prediction results of individual classifiers and determine the final prediction.

1) *Individual Classifiers*: Similar to image selectors, classifiers are also image-type-wise. Therefore, we trained classifiers for each type of weld pool image independently. Considering the distinct patterns of active vision, passive vision, and REIs of weld pools, multiple convolutional neural networks in different architectures, which are capable of learning different features, were trained and evaluated on collected data sets. To be specific, we selected the candidate neural networks as illustrated in Table I according to the number of layers, the number of parameters, and the novelty.

a) *AlexNet*: AlexNet [29] aroused interest in applying deep learning to machine vision when it won the ImageNet challenge of 2012 [23]. Although AlexNet contained only

five convolutional layers and three fully connected layers, it is a breakthrough development for the feasibility of running it on GPUs and the techniques it introduced, such as data augmentation and dropout to reduce overfitting.

b) *VGGNet*: VGGNet [30] is famous for its simplicity and depth. To keep its architecture simple, VGGNet strictly used a 3×3 convolutional filter with a stride of 1, along with a 2×2 max pooling layer with a stride of 2. VGGNet pushed the depth to 16–19 weight layers that were very deep in 2014. The performance of VGGNet reinforced the notion that deeper layers promise more effective hierarchical representations of visual data typically [31].

c) *ResNet*: ResNet [32] demonstrated the possibility of training thousands of layers and still achieving compelling performance. It introduced the identity shortcut connection between layers (a middle layer receives outputs from the preceding second or third layer) to gain maximal representations and alleviate the vanishing gradient problem.

d) *DenseNet*: DenseNet [33] connected each layer to every other layer in a feed-forward fashion, i.e., for each layer, outputs of all preceding layers are used as inputs, and its own output feeds all subsequent layers. Compared with ResNet, DenseNet demonstrated more efficiency in terms of parameters and computation for the same level of accuracy because of its further strengthened feature propagation and feature reuse.

e) *CapsNet*: CapsNet [34], [35] introduced a new groundbreaking concept of capsules, or groups of neurons, to capture properties of one entity inside an image. Unlike traditional convolutional neural networks, CapsNet is robust against geometric distortions and transformations because a single capsule contains viewpoints from multiple layers.

2) *Ensemble Models*: Considering individual neural networks' potential unstable classification performance on different types of images and overfitting on unseen data, we designed two-level plurality-voting-based ensemble models, Ensemble-1 and Ensemble-2 in Fig. 1, to predict the final weld penetrations in a more robust way. At the first level, the image-type-wise Ensemble-1 concluded classification results for each type of image (i.e., active vision, passive vision, and REI) using the plurality voting scheme. Suppose two labels, partial penetration and full penetration, are available. If three or more than three of the five individual neural networks (AlexNet, VGGNet, ResNet, DenseNet, and CapsNet) predict partial penetration, Ensemble-1 will interpret the overall individual classifier decisions as partial penetration. Otherwise, full penetration is concluded by Ensemble-1.

As mentioned earlier, three Ensemble-1s (Ensemble-1 (Active Vision), Ensemble-1 (Passive Vision), and Ensemble-1 (REI) in Fig. 1 act on three types of weld images. When they disagree with one another, i.e., the labels predicted based on active vision, passive vision, and REI are inconsistent, the second-level Ensemble-2 combines their results using the plurality voting scheme. For example, if both active Ensemble-1 and REI Ensemble-1 produce a partial penetration label, while passive Ensemble-1 gives a full penetration label, Ensemble-2 will select partial penetration as the final label.

Algorithm 1: Two-level Ensemble Models.

Data: $S = \{Active, Passive, REI\}$: the set of sensing image types; $N = \{AlexNet, VGGNet, ResNet, DenseNet, CapsNet\}$: the set of neural networks; P : the set of weld penetration labels

Result: $p^* \in P$: the final weld penetration label

```

// Ensemble-1s
for each  $s$  in  $S$  do
  for each  $p$  in  $P$  do
     $C_p \leftarrow 0$ ; // count of penetration  $p$ 
  end
  for each  $n$  in  $N$  do
     $p \leftarrow P_s^n$ ; // penetration by neural network  $n$  on sensing channel  $s$ 
     $C_p \leftarrow C_p + 1$ ; // update the count of  $p$ 
  end
   $C_{max} \leftarrow 0$ ; // the max count
   $P_s^* \leftarrow None$ ; // Ensemble-1 label on sensing channel  $s$ 
  for each  $p$  in  $P$  do
    if  $C_p > C_{max}$  then
       $C_{max} \leftarrow C_p$ ; // update max count
       $P_s^* \leftarrow p$ ; // Ensemble-1 label on  $s$ 
    end
  end
end
// Ensemble-2
for each  $p$  in  $P$  do
   $C_p \leftarrow 0$ ; // count per penetration
end
for each  $s$  in  $S$  do
   $p \leftarrow P_s^*$ ; // Ensemble-1 label on  $s$ 
   $C_p \leftarrow C_p + 1$ ; // update the count of  $p$ 
end
 $C_{max} \leftarrow 0$ ; // the max count
for each  $p$  in  $P$  do
  if  $C_p > C_{max}$  then
     $C_{max} \leftarrow C_p$ ; // update max count
     $p^* \leftarrow p$ ; // update final penetration
  end
end
return  $p^*$ 

```

III. EXPERIMENTS

We first describe the GTAW platform and the computational platform used in our article, then we list the data sets used in each phase of DeepWelding. Finally, we present the parameter settings for deep learning models involved in the DeepWelding framework.

A. Welding Experiments

1) *GTAW Platform and Weld Bead:* Fig. 4 shows the configuration of the experimental platform used in our article. The welding torch traveled from the left toward the right to perform

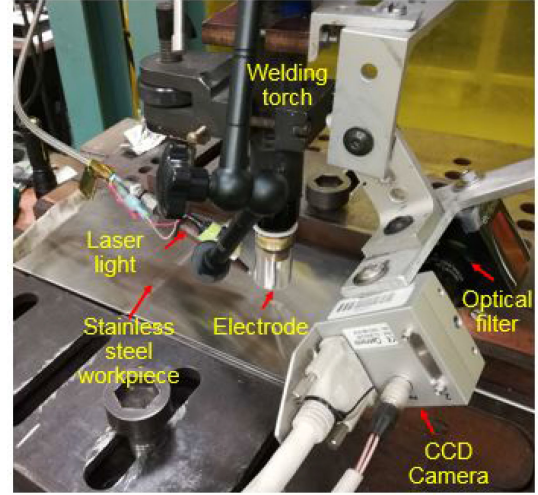


Fig. 4. Experimental setup of GTAW.

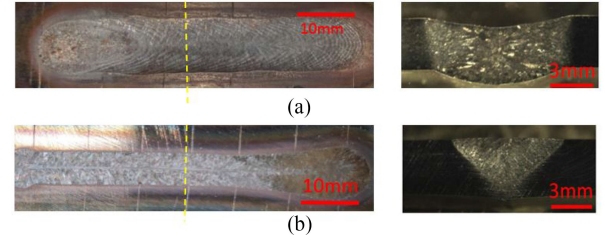


Fig. 5. Weld bead examples. (a) Fully penetrated weld. (b) Partially penetrated weld.

a weld. The optical sensing system was mounted on the welding torch. A pulsed diode laser source (905 nm wavelength) was fixed on the left of the torch with 45° incident angle. A CCD camera was mounted on the right-hand side of the torch with a 45° incline angle. A bandpass filter (905 nm) was placed in front of the camera lens to suppress the arc light. The camera shutter and the laser light were synchronized and controlled through the computer, which allowed us to obtain three types of weld pool images sequentially within 2 ms. Specifically, 1) an active vision image of the weld pool surface illuminated by the pulsed laser and captured with a relatively short camera exposure (approximately 1 s); 2) a passive vision image of the weld pool surface illuminated by the welding arc without a laser and captured with a relatively long exposure time (100 to 500 s); and 3) an REI also illuminated by the welding arc only and captured with camera exposure duration at proximate 5 to 10 s. The REI image contains the position information of both the electrode and its reflection from the weld pool. Such position information could be used to determine the height of weld pool surface [2].

Each GTAW welding experiment was performed on 304 L stainless steel plates with a fixed thickness and constant welding speed. The welding current was purposely adjusted to produce different states of weld penetration.

TABLE II
DATA SET FOR DENOISING AND IMAGE SELECTION

Thickness (mm)	Speed (mm/s)	Current (A)	# of images
2	1	50	357
2	1	60	276
3	1	70	414
3	2	100	181
6	2	120	734
6	2	150	540

TABLE III
DATA SET FOR PENETRATION CLASSIFICATIONS

Thickness (mm)	Speed (mm/s)	Current (A)	# of images
2	1	45~70	1275
3	1	70~110	1227
3	2	100~125	1011

Fig. 5 illustrates the topside view and cross section of welded parts from a GTAW platform with two different penetration states.

B. Data Set

We collected active vision images, passive vision images, and REIs of weld pools under different experimental settings in terms of the thickness of the material, the welding speed, and the electrical current for data preprocessing, image selection, and penetration classification. For simplicity, other welding parameters, such as voltage, shielding gas rate, material, and surface finish, were also considered fixed.

We used the same data set with diverse experimental settings to perform both denoising and image selection, as shown in Table II. The data set included thicknesses of materials, varying from 2 to 6 mm; welding speeds from 1 to 2 mm/s; and electrical currents from 50 to 150 A.

The data set for penetration classification (both training and testing) is illustrated in Table III. We grouped experiments with the same material thicknesses and moving speeds together. For each group, we changed electrical current values to generate images with labels of partial penetrations and full penetrations. Note that experiments with material thicknesses of 6 mm listed in Table II were not included in this data set (see Table III) because we have only one label for penetration under this setting. It is meaningless to train and evaluate classifiers on a data set with only one label. But we still conducted data preprocessing and image selection on these images, so their representative examples are included in Table II.

C. Computational Experiments

The deep learning studies were performed on an Nvidia DGX server with four cutting-edge Nvidia Tesla V100 GPUs. Each Tesla V100 was equipped with 640 Tensor Cores and 16 GB memory. We implemented all involved neural networks using TensorFlow 1.7.1 and Python 3.5.2.¹ Furthermore, we listed

technical details (such as network parameter setting) on each of the computational experiment phases: data preprocessing, image selection, and image classification.

1) *Data Preprocessing*: The pix2pix [27] image translation model requires pairs of images (A, B) as input to learn the mapping between image A and image B . In our experiments, we set the originally captured active vision images as the image A , and the images denoised by experts as the image B . When training the pix2pix model, we used the default values of all parameters, except that we set the max_epochs as 200.

2) *Image Selection*: As mentioned earlier, we adopted AlexNet to distinguish between the low- and high-quality images. We used a batch size of 5, learning rate of 10^{-5} , dropout probability of 0.5, and weight decay of 5×10^{-4} to train the AlexNet image selector for 2000 epochs.

3) *Image Classification*. *AlexNet*: All the parameters were the same for AlexNet, which was used for image selection, except that we set the epoch as 400. *VGGNet*: To shorten the training time, we fine-tuned our model based on the pre-trained VGGNet on ImageNet. Specifically, we first loaded the pretrained VGGNet and trained the last fully connected layer with a learning rate of 10^{-3} for 40 epochs. Then, we trained the full network with a learning rate of 10^{-5} for 400 epochs. The batch size, dropout probability, and weight decay were the same as for AlexNet. *ResNet*: We tuned the ResNet pretrained on ImageNet using the Adam optimizer with a learning rate of 10^{-4} for 400 epochs. We also set the patience as 10 for early training, stopping if the performance did not increase within 10 epoch. *DenseNet*: The parameters were exactly the same as for ResNet except that we tuned it on the pretrained DenseNet. *CapsNet*: We used all default parameters in [34] and [35].

IV. RESULTS AND ANALYSIS

The performance of three components in the proposed framework is presented in the following three sections. To be specific, we illustrated and analyzed the denoising performance on active vision images, the selection of high-quality multisource images, and the weld penetration classification under different experimental settings.

A. Illustration of Image Denoising Using pix2pix

We ran the deep learning based pix2pix model to remove the light reflection noise in active vision images of weld pools; the results are shown in Table IV. Each column of the table is a specific experimental setting corresponding to the images in each line in Table II. For example, the last column titled $t = 6, s = 2, c = 150$ contains active vision images generated with a material thickness of 6 mm, welding speed of 2 mm/s, and current of 150 A. The first row except the header in Table IV is the original input image for the network, and the second row lists the output image after denoising. The last row lists the associated ground-truth images after manual generation by experts. Generally speaking, Table IV shows satisfactory performance achieved by the pix2pix model under diverse experimental settings, as the denoised images are very close to the images generated by human experts.

¹The project code is available online. [Online]. Available: <https://github.com/YunheFeng/deepwelding>

TABLE IV
DENOISE PERFORMANCE UNDER DIFFERENT EXPERIMENTAL SETTINGS OF THICKNESS (t), SPEED (s), AND CURRENT (c)

Labels	$t=2, s=1, c=50$	$t=2, s=1, c=60$	$t=3, s=1, c=70$	$t=3, s=2, c=100$	$t=6, s=2, c=120$	$t=6, s=2, c=150$
Input						
Output						
Ground Truth						

TABLE V
PERFORMANCE OF IMAGE SELECTORS ON MULTISOURCE IMAGES

Thickness	Speed	Current	Denoise Acc.	Passive Acc.	REI Acc.
2	1	50	91.8%	90.1%	95%
2	1	60	90.2%	96.2%	98.2%
3	1	70	87.4%	92.5%	100%
3	2	100	81.8%	98.3%	97.3%
6	2	120	92.9%	100%	100%
6	2	150	90.9%	98.5%	98.5%

B. Performance of Image Selectors

For denoised active vision images, passive vision images, and REIs in each line in [Table II](#), we selected randomly 80% images to train three image-type-wise AlexNets to select high-quality images. The fourth to sixth columns in [Table V](#) demonstrate the corresponding test accuracy on the other 20% images. We can see that the average selection performance on passive vision images and REIs is slightly better than that on denoised active vision images. But note that even “low-quality” denoised active vision images actually have a higher quality than the corresponding raw active vision images, since most noise inside weld pools has been removed by pix2pix during preprocessing. Overall, the trained AlexNets can achieve satisfactory accuracy to filter out unnecessary, low-quality images for each of the sensory input data streams under different experimental settings.

C. Performance of Penetration Classification

We used the data set in [Table III](#) to evaluate the penetration classification performance. There are four possible ways to design the experiments, regarding both the policy for grouping data and the strategy for splitting data sets into training and testing. For the data grouping policy, we can first combine images in all three lines in [Table III](#) and then train the neural networks on them. Or we can treat each line in [Table III](#) as an independent data set and train neural networks on them

respectively. For the strategy to split data, we can split the data with the same thickness (t), welding speed (s), and electrical current (c) into 80% training and 20% testing. Thus, the data with the same welding experimental settings can be seen in both training and testing data sets. Recall that we changed c to generate different welding penetration labels under a given t and s . Or we can hold out specific data collected with certain currents as testing data sets and treat the rest as training data sets; that approach enables us to see the generalization capability of our methods, since the data generated with certain currents are unseen in training neural networks.

When the mentioned 80%–20% training–testing splitting is conducted, no matter which grouping policy is taken, individual neural networks perform very well. Thus, these neural networks are capable of learning useful features of welding images automatically during training. After our two-level ensembles were applied, the accuracy achieved was near 100%, which is a clear index of overfitting. Therefore, this kind of data splitting was shown not to be a good choice.

For the grouping policy, if data with different experimental settings are not aggregated, individual neural networks have to be trained and applied for every combination of thickness t and moving speed s . However it seems impractical in industry because 1) the thickness and the speed are continuous values, meaning a large number of models have to be trained, and 2) the experimental setting must be explicitly given to switch between different pretrained neural networks.

Therefore, we decided to group data generated under multiple thicknesses t and moving speeds s together, and held out unseen data for testing. The performance of individual neural networks and the two assemble methods on three different holdout sets are listed in [Table VI](#). Individual neural networks behaved inconsistently regarding different types of sensing images and different holdout sets. For example, AlexNet performed well on Passive images, but failed to achieve a comparable accuracy on REIs. Similarly, AlexNet demonstrated unstable performance

TABLE VI

GENERALIZATION CAPABILITY OF GROUPING ALL DATA AND KEEPING SPECIFIC EXPERIMENTAL SETTINGS OF THICKNESS (t), SPEED (s), AND CURRENT (c) UNSEEN DURING TRAINING

Holdout Set	Classifier	Active	Denoise	Passive	REI
t=2,s=1,c=50	Alexnet	71.8%	88.4%	91.4%	70.7%
	VGGNet	94.6%	94.2%	76.9%	92.2%
	ResNet	75.0%	54.7%	70.9%	78.9%
	DenseNet	82.5%	93.5%	77.8%	99.4%
	CapsNet	75.7%	90.2%	82.6%	97.6%
t=3,s=1,c=110	Ensemble-1	84.9%	93.1%	77.2%	92.9%
t=3,s=2,c=105	Ensemble-2 (4-way)	96.1%			
	Ensemble-2 (3-way)	91.8%			
t=2,s=1,c=45	Alexnet	81.4%	88.0%	99.7%	83.0%
	VGGNet	80.9%	80.4%	98.7%	84.5%
	ResNet	91.6%	88.0%	99.5%	82.6%
	DenseNet	85.5%	85.9%	84.1%	87.4%
	CapsNet	86.9%	81.8%	64.8%	80.0%
t=3,s=2,c=120	Ensemble-1	87.0%	90.1%	99.7%	85.0%
t=3,s=2,c=120	Ensemble-2 (4-way)	87.5%			
	Ensemble-2 (3-way)	93.1%			
t=2,s=1,c=50	Alexnet	74.3%	89.0%	83.6%	62.9%
	VGGNet	85.1%	94.2%	68.7%	75.4%
	ResNet	73.0%	88.1%	68.4%	59.7%
	DenseNet	85.8%	80.2%	71.8%	64.9%
	CapsNet	75.6%	88.1%	79.8%	91.7%
t=3,s=2,c=125	Ensemble-1	88.4%	91.3%	74.5%	67.1%
t=3,s=2,c=125	Ensemble-2 (4-way)	94.4%			
	Ensemble-2 (3-way)	88.1%			

on different holdout sets (see an accuracy of 99.7% on the second holdout set but 83.6% on the third one).

In Table VI, we also included the raw active vision images without denoising (see the Active column). Denoising procedure over the raw active vision images leads to a better result using Ensemble-1s, which indicates the effectiveness of using pix2pix to remove noise. On each type of sensing images, although the Ensemble-1 was not always the one achieving the highest accuracy, it kept the classification more robust and stable because it outperformed the worst-case results produced by individual neural networks significantly. We showed the performance of two Ensemble-2s: the 4-way Ensemble-2 and the 3-way Ensemble-2. In addition to denoised active vision images, passive vision images, and REIs, 4-way Ensemble-2s took into account the raw active vision image results. In contrast, 3-way Ensemble-2s excluded the raw active vision image results. On each holdout data set, Ensemble-2s aggregated the results of Ensemble-1s into final penetration labels, and demonstrated acceptable performance that was much better than the low accuracy from Ensemble-1s. Even the 87.5% accuracy from the 4-way Ensemble-2 in the second holdout data set is better the 87.0% and 85.0% by Ensemble-1s. Furthermore, the generalization capability results infer a good possibility that we may be able to establish a unified neural network training procedure with a single comprehensive data set that contains diverse welding settings in terms of thickness t and moving speed s .

V. CONCLUSION

In this article, we presented a DeepWelding framework that uses multisource sensing images and deep learning technology to classify weld penetration status during a demonstrative GTAW process. The main challenges we encountered in the

article were data quality control among the multiple experimental settings (a combination of material thickness, head speed, and electrical current changes). Traditional methods of weld penetration prediction involve interactive measurement and empirical knowledge. When we adopted deep learning in our framework, image calibration, denoising, and image selection became necessary and time-consuming processes. Neural networks performed well in image denoising via CGAN-based pix2pix and image selection via AlexNet. Five different neural networks (AlexNet, VGGNet, ResNet, DenseNet, and CapsNet) behaved inconsistently in the classification phase over three individual sensing images. Therefore, we adopted two ensemble methods to improve the model generalization capability by collecting all the images from different experimental settings into a single training data set. We found that the quality of model prediction was heavily influenced by the data stream collection environment. Compared with conventional approaches, deep learning augmented approaches require more restricted and standardized methods of data harvesting and preprocessing.

There exist several interesting and important directions for future work. First, we will collect and annotate a larger number of welding images generated under different welding parameter settings (e.g., welding speed and thickness of the material) and consistent experimental conditions to further fine-tune the individual neural networks involved in DeepWelding. Second, we will integrate the CGAN-based pix2pix, which is used in DeepWelding to remove noise in active vision images, into existing arc welding solutions to enhance the quality of welding images. In addition, deploying DeepWelding on other types of other common types of welding procedures (such as metal inert gas welding, GMAW, and flux-cored arc welding) would also be a future direction.

ACKNOWLEDGMENT

The research used computing resources within the Computing and Data Environment for Science at Oak Ridge National Laboratory, which is managed by UT-Battelle LLC for the Department of Energy.

REFERENCES

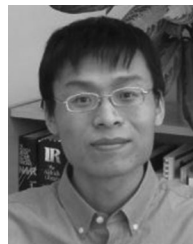
- [1] Z. Feng, J. Chen, and Z. Chen, "Monitoring weld pool surface and penetration using reversed electrode images," *Weld. J.*, vol. 96, no. 10, pp. 367S–375S, 2017.
- [2] S. A. David, J. Chen, B. T. Gibson, and Z. Feng, "Intelligent weld manufacturing: Role of integrated computational welding engineering," in *Transactions on Intelligent Welding Manufacturing*. New York, NY, USA: Springer, 2018, pp. 3–30.
- [3] S.-B. Chen and J. Wu, *Intelligentized Methodology for Arc Welding Dynamical Processes*. New York, NY, USA: Springer, 2009.
- [4] R. Kovacevic, Y. Zhang, and L. Li, "Monitoring of weld joint penetrations based on weld pool geometrical appearance," *Weld. J.-Including Weld. Res. Suppl.*, vol. 75, no. 10, pp. 317–329, 1996.
- [5] G. Saeed and Y. Zhang, "Weld pool surface depth measurement using a calibrated camera and structured light," *Meas. Sci. Technol.*, vol. 18, no. 8, pp. 2570–2578, 2007.
- [6] H. S. Song and Y. M. Zhang, "Three-dimensional reconstruction of specular surface for a gas tungsten arc weld pool," *Meas. Sci. Technol.*, vol. 18, no. 12, pp. 3751–3767, 2007.
- [7] Y.-K. Liu and Y.-M. Zhang, "Supervised learning of human welder behaviors for intelligent robotic welding," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 3, pp. 1532–1541, Jul. 2017.

- [8] H.-Y. Shen, J. Wu, T. Lin, and S.-B. Chen, "Arc welding robot system with seam tracking and weld pool control based on passive vision," *Int. J. Adv. Manuf. Technol.*, vol. 39, nos. 7/8, pp. 669–678, 2008.
- [9] S. Chen and J. Wu, "A survey on intelligentized technologies for visual information acquirement, modeling and control of arc welding pool dynamics," in *Proc. 33rd Annu. Conf. Ind. Electron. Soc.*, 2007, pp. 781–785.
- [10] D. Wu, H. Chen, Y. Huang, Y. He, M. Hu, and S. Chen, "Monitoring of weld joint penetration during variable polarity plasma arc welding based on the keyhole characteristics and PSO-ANFIs," *J. Mater. Process. Technol.*, vol. 239, pp. 113–124, 2017.
- [11] R. Liang, R. Yu, Y. Luo, and Y. Zhang, "Machine learning of weld joint penetration from weld pool surface using support vector regression," *J. Manuf. Processes*, vol. 41, pp. 23–28, 2019.
- [12] L. Liu, H. Chen, and S. Chen, "Online monitoring of variable polarity tig welding penetration state based on fusion of welding characteristic parameters and SVM," in *Transactions on Intelligent Welding Manufacturing*. New York, NY, USA: Springer, 2019, pp. 87–104.
- [13] Z. Zhang, G. Wen, and S. Chen, "Audible sound-based intelligent evaluation for aluminum alloy in robotic pulsed GTAW: Mechanism, feature selection, and defect detection," *IEEE Trans. Ind. Informat.*, vol. 14, no. 7, pp. 2973–2983, Jul. 2018.
- [14] K. Andersen, G. E. Cook, G. Karsai, and K. Ramaswamy, "Artificial neural networks applied to arc welding process modeling and control," *IEEE Trans. Ind. Appl.*, vol. 26, no. 5, pp. 824–830, Sep./Oct. 1990.
- [15] J. Lee and K. Um, "A comparison in a back-bead prediction of gas metal arc welding using multiple regression analysis and artificial neural network," *Opt. Lasers Eng.*, vol. 34, no. 3, pp. 149–158, 2000.
- [16] I. Kim, J. Son, C. Park, C. Lee, and Y. K. Prasad, "A study on prediction of bead height in robotic arc welding using a neural network," *J. Mater. Process. Technol.*, vol. 130, pp. 229–234, 2002.
- [17] D. Nagesh and G. Datta, "Prediction of weld bead geometry and penetration in shielded metal-arc welding using artificial neural networks," *J. Mater. Process. Technol.*, vol. 123, no. 2, pp. 303–312, 2002.
- [18] H. Ates, "Prediction of gas metal arc welding parameters based on artificial neural networks," *Mater. Des.*, vol. 28, no. 7, pp. 2015–2023, 2007.
- [19] D. Wu, H. Chen, Y. He, S. Song, T. Lin, and S. Chen, "A prediction model for keyhole geometry and acoustic signatures during variable polarity plasma arc welding based on extreme learning machine," *Sensor Rev.*, vol. 36, no. 3, pp. 257–266, 2016.
- [20] P. Ghanty *et al.*, "Artificial neural network approach for estimating weld bead width and depth of penetration from infrared thermal image of weld pool," *Sci. Technol. Welding Joining*, vol. 13, no. 4, pp. 395–401, 2008.
- [21] W. Li, Y. Ji, J. Wu, and J. Wang, "A modified welding image feature extraction algorithm for rotating arc narrow gap mag welding," *Ind. Robot: An Int. J.*, vol. 42, no. 3, pp. 222–227, 2015.
- [22] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1026–1034.
- [23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 248–255.
- [24] Y. Zhang, D. You, X. Gao, N. Zhang, and P. P. Gao, "Welding defects detection based on deep learning with multiple optical sensors during disk laser welding of thick plates," *J. Manuf. Syst.*, vol. 51, pp. 87–94, 2019.
- [25] W. Hou, Y. Wei, J. Guo, Y. Jin, and C. Zhu, "Automatic detection of welding defects using deep neural network," *J. Phys.: Conf. Ser.*, vol. 933, no. 1, 2018, Art. no. 012006.
- [26] P. Sassi, P. Tripicchio, and C. A. Avizzano, "A smart monitoring system for automatic welding defect detection," *IEEE Trans. Ind. Electron.*, vol. 66, no. 12, pp. 9641–9650, Dec. 2019.
- [27] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 1125–1134.
- [28] Z. Chen, J. Chen, and Z. Feng, "Welding penetration prediction with passive vision system," *J. Manuf. Processes*, vol. 36, pp. 224–230, 2018.
- [29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015. [Online]. Available: arXiv:1409.1556.
- [31] Y. Aytar, C. Vondrick, and A. Torralba, "Soundnet: Learning sound representations from unlabeled video," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 892–900.
- [32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 770–778.
- [33] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2017, pp. 4700–4708.
- [34] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," in *Proc. Adv. Neural Inform. Process. Syst.*, 2017, pp. 3856–3866.
- [35] G. E. Hinton, S. Sabour, and N. Frosst, "Matrix capsules with em routing," in *Proc. Int. Conf. Learn. Representations*, 2018. [Online]. Available: <https://openreview.net/pdf?id=HJWLfGWRb>



Yunhe Feng received the B.E. degree in computer science and technology and the M.E. degree in computer technology from the Beijing University of Technology, Beijing, China, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree majoring in computer science with the University of Tennessee, Knoxville, TN, USA.

His research interests include data mining and knowledge discovery, mobile security and privacy, and machine learning.



Zongyao Chen received the B.S. and M.S. degrees in electrical engineering from Shanghai Jiaotong University, Shanghai, China, in 2009 and 2012, respectively, and the doctoral degree in mechanical engineering from the University of Tennessee, Knoxville, TN, USA, in 2018.

He is a Research Scientist with the R&D Centre of American Air Liquide, Newark, DE, USA. During his Ph.D. study, he worked with the Welding and Joining Research Group at Oak Ridge National Lab and focused on applying artificial

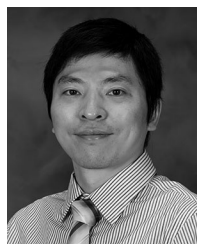
intelligent and computer vision technology to welding manufacturing. His current research interests include robotics, welding automation, computer vision, machine learning, and other artificial intelligence (AI) applications in industry.



Dali Wang received the M.Sc. degree in computational science and engineering, the M.S. degree in computer science, and the Ph.D. degree in environmental engineering (computational science focus) from Rensselaer Polytechnic Institute, Troy, NY, USA, in 2001, 2001, and 2002, respectively.

He is a Senior R&D Staff and a Member of Artificial Intelligent (AI) initiative at Oak Ridge National Laboratory, Oak Ridge, TN, USA. He is also a Joint Associate Professor of computer

science with the University of Tennessee, Knoxville, TN, USA. He is the author (coauthor) of more than 95 peer-reviewed publications (papers, data, and software products) and the principal investigator (PI) (co-PI) of over 30 funded projects. His research interests include AI, high-performance computing, machine learning for data science, Earth and environmental system modeling, advanced manufacturing, developmental biology, and geographic information system.



Jian Chen received the B.S. degree in materials science and engineering from Jiangsu University, Zhenjiang, China, in 2002, the M.S. degree in solid mechanics from the University of Science and Technology of China, Hefei, China, in 2005, and the Ph.D. degree in industrial engineering from The Ohio State University, Columbus, OH, USA, in 2010.

He is an R&D Staff, Materials Joining Group, Oak Ridge National Laboratory, Oak Ridge, TN, USA. He has significant experimental and analytical experiences in developing advanced materials joining technologies and the associated control and monitoring technologies. His research interests include advanced sensing techniques to monitor and control welding processes and high-performance welding simulation techniques.



Zhili Feng received the B.S. degree in mechanical engineering and the M.S. degree in material science from Tsinghua University, Beijing, China, in 1983 and 1989, respectively, and the Ph.D. degree in welding engineering from The Ohio State University, Columbus, OH, USA, in 1994.

He is the Group Leader of the Materials Joining Group and a Distinguished R&D staff member with Oak Ridge National Laboratory, Oak Ridge, TN, USA. His current research includes both fundamental and applied R&D and technology innovations for diverse interdisciplinary subjects related to materials joining and allied materials manufacturing processes. He has published more than 200 peer-reviewed papers.

Dr. Feng is a Fellow of the American Welding Society and also Fellow of the International Institute of Welding.