中國機械工程學刊第四十六卷第二期第 205~210 頁(民國一百一十四年)
Journal of the Chinese Society of Mechanical Engineers, Vol.46, No.2, pp 205~210 (2025)

Automated Detection of Arc Welding Defects by Using a Convolutional Neural Network Model

Shu-Hsien Huang*, Ting-En Wu** and Chia-Hung Lai ***

Keywords：Arc Welding, Defect Detection, Convolutional Neural Network (CNN), Welding Quality Control

ABSTRACT

Arc welding has low costs and offers high welding efficiency. Visual inspection, the easiest welding inspection method, is prone to human error. This research proposes a convolutional neural network (CNN)-based method for classifying welding defects into the following categories: 'no defect', 'blow hole', 'lack of fusion', 'incompletely filled groove', and 'undercut'. The weld bead was positioned under a camera, and the captured images were transmitted to a computer, which recognised defects in the images. After recognition, the computer saved the images and added them to a defect detection dataset to ensure that the dataset was continuously updated. The proposed defect recognition model achieved an accuracy as high as 97.2%. A comparison was conducted for different numbers of images and training iterations. We recommend collecting a minimum of 800 images when a CNN model is to be trained from scratch to detect welding defects.

INTRODUCTION

Arc welding is a method in which metals are joined by creating a high-temperature arc by using electric current. This method is widely utilised in industries because of its low cost and high welding efficiency. Welding inspection in the form of visual inspection is typically conducted after welding is completed.

Paper Received April, 2024. Revised July, 2024. Accepted August, 2024. Author for Correspondence: Chia-Hung Lai
E-mail(s): chlai@gm.ncut.edu.tw

**Department of Information Management, National Chin-Yi University of Technology, No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan (R.O.C.)*

***Department of Industrial Education and Technology, National Changhua University of Education, No.2, Shi-Da Road, Changhua City50074, Taiwan (R.O.C.)*

****Department of Intelligent Automation Engineering, National Chin-Yi University of Technology, No.57, Sec. 2, Zhongshan Rd., Taiping Dist., Taichung 411030, Taiwan (R.O.C.)*

However, visual inspection is susceptible to human factors, leading to potential flaws in the identification of welding defects. Therefore, designing a defect recognition system to enhance the accuracy of defect identification is essential (Wang & Zhu, 2021).

Image recognition combined with deep learning has proven to be highly effective in defect detection. Luo (2005) showcased the ability of artificial neural networks to identify welding abnormalities during laser welding. The aforementioned study focused on laser welding, which limits the applicability of its findings to other welding methods such as arc welding. Moreover, previous studies have not addressed the real-time processing needs of industrial environments. Kasban (2011) successfully detected welding defects through radiographic and ultrasonic imaging, even in noisy environments. Despite the robustness of this method, the method is infeasible for widespread industrial application because it requires complex and costly imaging equipment. Moreover, interpreting radiographic images is a specialised skill demanded by the aforementioned approach. Related studies have shown the potential of machine learning and deep learning for improving precision and predictive capabilities in various applications. For instance, Peng et al. (2024) conducted nonlinear dynamic analysis and forecasting of symmetric aerostatic cavities bearing systems. Wang et al. (2022) demonstrated the use of an optimized XGBoost model for predicting turning precision. Lin et al. (2020) applied support vector machine based on the artificial fish-swarm algorithm for diagnosing ball-bearing faults. Jian et al. (2019) used a general regression neural network to predict spindle displacement caused by heat. Both Yang (2020) and Yang (2021) have reported high accuracy rates for a deep-learning-based method that was used to perform defect localisation during welding.

To address these research gaps, the present study developed a convolutional neural network (CNN)-based approach for the automated detection of arc welding defects. Previous studies primarily focused on defect localization rather than defect classification, and the datasets used were limited, potentially not representing the diverse range of welding defects encountered in real-world applications. Our method involves training a CNN model using images of welding defects, designed to recognize defects such as porosity, poor weld formation, insufficient filler metal, weld erosion, and defect-free regions. This approach effectively addresses errors overlooked during visual inspection. The main contributions of this research are as follows:

1. A CNN-based approach was developed to detect arc welding defects to improve welding inspection systems.
2. The developed CNN-based system can classify different types of welding defects.
3. A total of 1250 images were annotated for classifying welding defects.
4. Welding defects were classified with an accuracy as high as 97.2%.

Classification of welding defects

Blow hole

Blow holes form when excessive gases are present during welding, which leads to the formation of cavities on the surface as the metal solidifies. Blow holes can reduce the strength of the weld bead and compromise the integrity of the structure (Choudhury, 2023). Welding fractures typically initiate from blow holes (Ro, 2021). From an engineering perspective, blow holes form when (1) the surface of the welded workpiece is contaminated with rust or oil and when (2) the welding flux on the welding rod is damp. Robotic inspection achieves a 95% accuracy in the detection of blow holes in the weld bead.

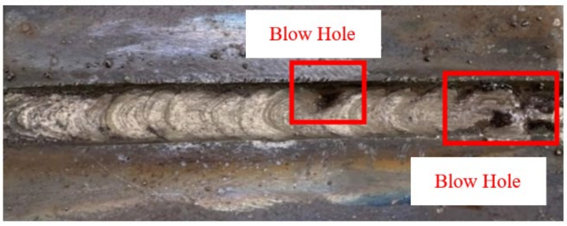

Fig. 1 Schematic of blow holes in the weld bead

Lack of fusion

Lack of fusion occurs when the weld bead and the base material are incompletely fused together. From an engineering perspective, lack of fusion can occur because of (1) low electric current, which results in insufficient temperature for complete metal melting, and (2) welding rod bias towards one side during welding, which results in uneven fusion of the workpiece (Hong, 2018). Robotic inspection offers high accuracy in the detection of lack of fusion during the welding of ultrathin plates. This method can serve as a foundation for intelligent defect detection.

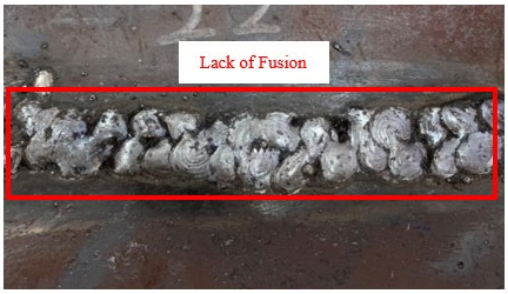

Fig. 2 Schematic of lack of fusion between the weld bead and the base material

Incompletely filled groove

An incompletely filled groove refers to a groove with insufficient molten fill during the welding process. Such grooves can occur on the surface and bottom of the weld bead. Incompletely filled grooves occur because of (1) insufficient welding current and (2) excessive welding speed. For workpieces with poor heat conduction, flame heating can be employed to prevent the occurrence of defects.

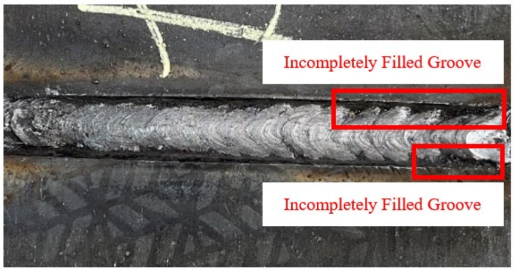

Fig. 3 Schematic of a weld bead with incompletely filled grooves

Undercut

An undercut refers to a groove that appears at the junction between the weld bead and the workpiece; such grooves cause excessive stress concentration within the material. Undercuts can occur because of (1) excessive current, which leads to inadequate molten fill in the weld groove, and (2) failure to clean the plate before welding. As the welding speed decreases, the occurrence of undercuts gradually decreases until they are completely eliminated.

Fig. 4 Schematic of an undercut in the weld bead.

Experimental setup

This research developed a CNN model for welding defect detection. The welding images used in this study were captured by the researchers at a factory. The captured images were sorted into a training database and test database, which facilitated subsequent deep learning training. Specific types of welding defects could be identified by the developed model.

The welding defect detection system developed in this study consists of a computer, camera, and lighting system. The weld bead is positioned beneath the camera, and the captured images are transmitted to the computer, which performs defect recognition on the images, stores the images, and executes preprocessing. The images are then added to a welding defect dataset to ensure that this dataset is continuously updated.

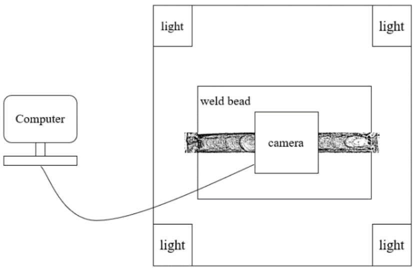

Fig. 5 Schematic of welding defect detection

Image preprocessing

The light source was positioned directly above the weld seam during imaging. System use was minimised to avoid excessive carbon emissions. The captured images were preprocessed by adjusting the brightness level by ±20. The images were subjected to contrast adjustments after they had been classified as ‘blow hole’, ‘lack of fusion’, ‘incompletely filled groove’, ‘undercut’, or ‘no defect’. The contrast was individually increased and decreased, and rotations were applied. Because the captured images were colorised and were resized to 64 during deep learning training, all of them had the standardised dimensions of 64 × 64 × 3. These images were then sorted in a 4:1 ratio into training and testing datasets. The quantity of images under each defect category is detailed in Table 1.

Table 1 Image count in the training and testing datasets

	Good	Blow Hole	Lack of Fusion	Incompletely Filled Groove	Undercut
Training Database	448	96	248	172	44
Test Database	112	24	62	43	11

Convolutional neural network

A CNN is a type of deep learning model composed of convolutional, pooling, flatten, dropout, and dense layers. The convolutional layer is the core of a CNN model; this layer extracts local features from images by using filters. The pooling layers effectively compress images without losing information by dividing them into multiple windows and selecting the maximum value from each window. The flatten layer transforms the output data generated by the convolutional and pooling layers into one-dimensional data, which are required for input into the dense layer. The dropout layer randomly discards some neuron outputs to prevent overfitting, thereby enhancing the model’s generalizability. The dense layer is a fully connected layer that classifies images. The output generated by the convolutional and pooling layers is fed to the fully connected layer, where the extracted features are weighted to improve training accuracy. The complexity of deep learning models increases with the depth of training. However, although increased model complexity improves accuracy, it also increases the likelihood of overfitting.

The CNN model developed in this study primarily consists of four blocks. Each of the first three blocks contains three convolutional layers and one pooling layer. The convolutional layers in the first, second, and third blocks have dimensions of 3 × 3 × 32, 3 × 3 × 64, and 3 × 3 × 128, respectively. Moreover, the pooling layers in the first, second, and third blocks have dimensions of 2 × 2 × 32, 2 × 2 × 64, and 2 × 2 × 128, respectively. The fourth block is composed of three (3 × 3 × 256) convolutional layers, one (2 × 2 × 256) pooling layer, one flatten layer, one dropout layer, and one dense layer (Figure 6).

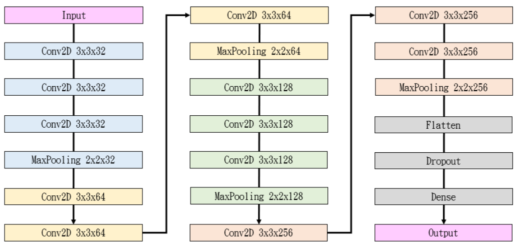

Fig. 6 Architecture of the CNN model developed in this study

Confusion matrix

A confusion matrix is a table used for evaluating the prediction results of a supervised deep learning model. This matrix has dimensions of N × N, where N represents the number of classification categories (Table 2). The effectiveness of a deep learning model can be quickly determined by calculating its accuracy. The formula for calculating accuracy is presented in Equation (1). Moreover, the formulas for calculating F1 score, precision, and recall are presented in Equations (2)–(4), respectively.

Table 2 Structure of a 5 × 5 confusion matrix

	A	B	C	D	E
A	M_{11}	M_{12}	M_{13}	M_{14}	M_{15}
B	M_{21}	M_{22}	M_{23}	M_{24}	M_{25}
C	M_{31}	M_{32}	M_{33}	M_{34}	M_{35}
D	M_{41}	M_{42}	M_{43}	M_{44}	M_{45}
E	M_{51}	M_{52}	M_{53}	M_{54}	M_{55}

$$Accuracy = \frac{\sum_{i=1}^{5} M_{ii}}{\sum_{i=1}^{5}\sum_{j=1}^{5} M_{ij}} \quad (1)$$

$$\text{F1} - \text{score} = \frac{2\times\text{precision}\times\text{recall}}{\text{precision}+\text{recall}} \quad (2)$$

$$\text{Precision} = \frac{M_{kk}}{\sum_{i=1}^{5} M_{kj}} \quad (3)$$

$$\text{Recall} = \frac{M_{kk}}{\sum_{j=1}^{5} M_{kj}} \quad (4)$$

Results

The objective of the present study was to classify welding defects. The collected images were compiled into a dataset, and image recognition was conducted after deep learning training to categorise welding defects. The accuracy of our CNN model was evaluated at epochs 30, 50, 80, 100, and 150 by using 400, 600, 800, and 1000 training images (Figure 7).

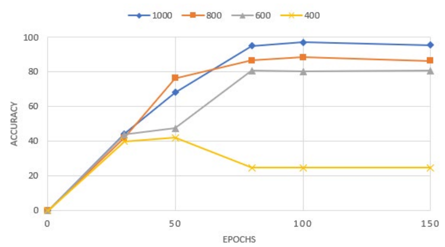

Fig. 7 Accuracy of the developed CNN model under different numbers of epochs and images

Table 3 summarises the training times exhibited by the proposed CNN model at epochs 30, 50, 80, 100, and 150 when 400, 600, 800, and 1000 training images were used. The training time increased with the number of images. Table 3 and Figure 7 reveal that the model accuracy did not vary considerably with the number of images at epoch 30. At epoch 50, the use of 800 and 1000 images resulted in a considerable enhancement in the model accuracy, whereas the use of 400 and 600 images did not produce a notable increase in accuracy. At this epoch, the best results were obtained with 800 images, followed by 1000 images. At epoch 80, the best accuracy was achieved when 1000 images were used for training. At this epoch, training with 600 and 800 images also resulted in a considerable accuracy improvement, whereas training with 400 images led to a noticeable decline in accuracy. At epoch 100, training with 1000 images yielded the highest accuracy. Training with 400, 600, and 1000 images resulted in only marginal variations in accuracy. At epoch 150, training with 400, 600, 800, and 1000 images resulted in a decrease in accuracy, which indicated the occurrence of overfitting.

Table 3 Training times of the proposed CNN model at different epochs when different numbers of training images were used

Counts / Epochs	400	600	800	1000
	Time(s)			
30	230.38	380.67	516.64	590.84
50	383.53	642.54	844.43	948.38
80	631.46	1029.91	1375.76	1546.25
100	820.39	1182.17	1701.25	1974.03
150	1176.03	1951.98	2606.38	3261.75

The results indicated that the highest model accuracy was achieved when 1000 training images were used. Therefore, the confusion matrix was determined for training with 1000 images. At epoch 30, the model failed to recognise welding defects, yielding an accuracy of only 44.4% (Figure 8). At epoch 50, the model's welding defect recognition improved, with its accuracy being 68.2% (Figure 9). At epoch 80, the model exhibited good accuracy in welding defect recognition, with its accuracy in detecting incompletely filled grooves being 95.2% (Figure 10). At epoch 100, the model achieved an accuracy of 97.2% (Figure 11). Finally, at epoch 150, the model accuracy marginally reduced to 95.6% (Figure 12), which indicated the occurrence of overfitting.

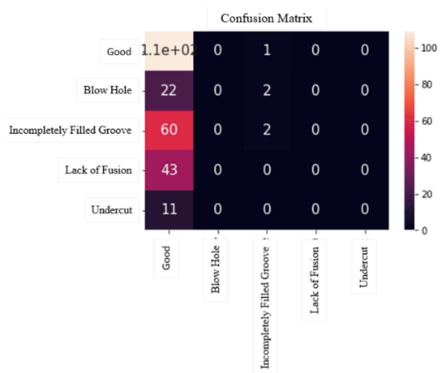

Fig. 8 Confusion matrix at epoch 30

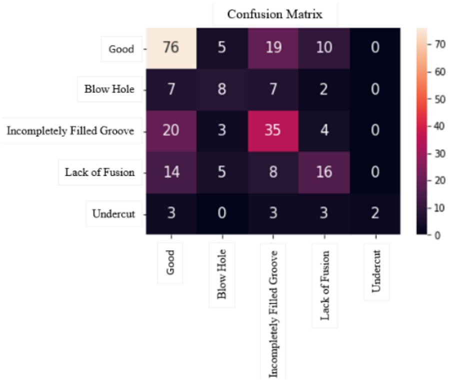

Fig. 9 Confusion matrix at epoch 50

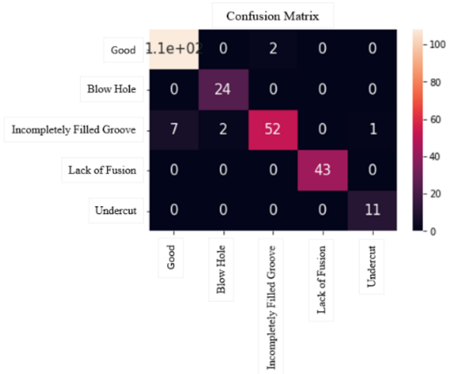

Fig. 10 Confusion matrix at epoch 80

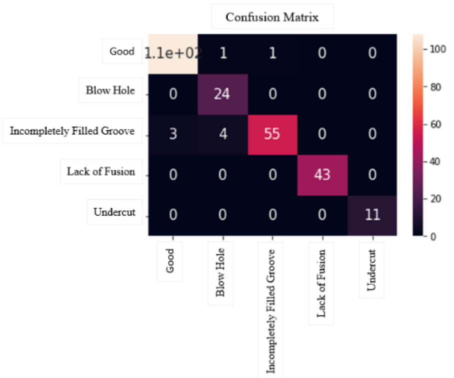

Fig. 11 Confusion matrix at epoch 100

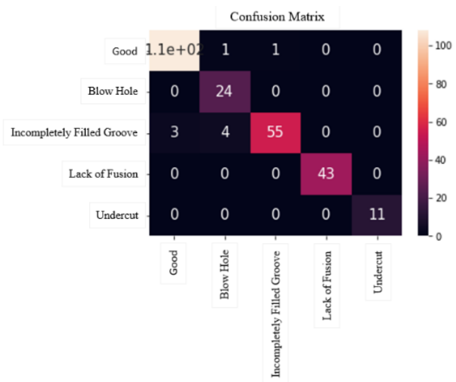

Fig. 12 Confusion matrix at epoch 150

The highest model accuracy was achieved at epoch 100 (Figure 7). Therefore, precision, recall, and F1 score were calculated at epoch 100. Extremely high precision, recall, and F1 scores were obtained for 'lack of fusion' and undercuts. However, relatively low recall and precision were achieved for incompletely filled grooves and blow holes, respectively.

Table 4 Precision, recall, and F1 scores at epoch 100.

	Good	Blow Hole	Incompletely Filled Groove	Lack of Fusion	Undercut
Precision	0.9735	0.8276	0.9821	1.0000	1.0000
Recall	0.9821	1.0000	0.8871	1.0000	1.0000
F1-Score	0.9778	0.9057	0.9321	1.0000	1.0000

CONCLUSIONS

This study developed an image recognition model for welding defect classification. The developed CNN model not only detects welding defects but also distinguishes between different types of defects. Images of different types of welding defects were compiled into a dataset for training and testing. This study analysed the training time and accuracy of the proposed model at different epochs when different numbers of training images were used. The following conclusions were drawn on the basis of the results of this study:

1. The proposed CNN model achieved an accuracy of up to 97.2% in welding defect classification.
2. The highest model accuracy was achieved for 100 training epochs. Further enhancement in model performance can be achieved by optimising the training parameters.
3. The use of 400 images resulted in poor model accuracy. Overall, the results indicate that a minimum of 800 images should be used to train the proposed CNN model for welding defect detection and classification.

Future research can improve the system proposed in this paper for welding defect detection. A weld may exhibit multiple defects, which can affect the accuracy of defect identification. An expert system capable of detecting and locating multiple defects can be designed, thereby reducing material wastage during training and ensuring environmental sustainability.

REFERENCES

Choudhury, S. D., Khan, W. N., Lyu, Z., & Li, L. (2023). Failure analysis of blowholes in welded boiler water walls. *Engineering Failure Analysis*, 153, 107560.

Hong, Y., Chang, B., Peng, G., Yuan, Z., Hou, X., Xue, B., & Du, D. (2018). In-process monitoring of lack of fusion in ultra-thin sheets edge welding using machine vision. *Sensors*, 18(8), 2411.

Jian, B. L., Wang, C. C., Hsieh, C. T., Kuo, Y. P., Houng, M. C., & Yau, H. T. (2019). Predicting spindle displacement caused by heat using the general regression neural network. *The International Journal of Advanced Manufacturing Technology*, *104*, 4665-4674.

Kasban, H., Zahran, O., Arafa, H., El-Kordy, M., Elaraby, S. M., & Abd El-Samie, F. E. (2011). Welding defect detection from radiography images with a cepstral approach. *Ndt & E International*, *44*(2), 226-231.

Lin, C. J., Chu, W. L., Wang, C. C., Chen, C. K., & Chen, I. T. (2020). Diagnosis of ball-bearing faults using support vector machine based on the artificial fish-swarm algorithm. *Journal of Low Frequency Noise, Vibration and Active Control*, *39*(4), 954-967.

Luo, H., Zeng, H., Hu, L., Hu, X., & Zhou, Z. (2005). Application of artificial neural network in laser welding defect diagnosis. *Journal of Materials Processing Technology*, *170*(1-2), 403-411.

Peng, T. J., Kuo, P. H., Huang, W. C., & Wang, C. C. (2024). Nonlinear Dynamic Analysis and Forecasting of Symmetric Aerostatic Cavities Bearing Systems. *International Journal of Bifurcation and Chaos*, *34*(04), 2430008.

Ro, C. S., Kim, K. H., Bang, H. S., & Yoon, H. S. (2021). Joint properties of aluminum alloy and galvanized steel by AC Pulse MIG braze welding. *Applied Sciences*, 11(11), 5105.

Wang, C. C., Kuo, P. H., & Chen, G. Y. (2022). Machine learning prediction of turning precision using optimized xgboost model. *Applied Sciences*, *12*(15), 7739.

Yang, L., Fan, J., Liu, Y., Li, E., Peng, J., & Liang, Z. (2020). Automatic detection and location of weld beads with deep convolutional neural networks. *IEEE Transactions on Instrumentation and Measurement*, 70, 1-12.

Yang, L., Wang, H., Huo, B., Li, F., & Liu, Y. (2021). An automatic welding defect location algorithm based on deep learning. Ndt & E International, 120, 102435.