Enhancing Welding Training with an AI System for Welding Skill Assessment

Thien Tran Ngoc, Phong Duong The, Tanh Tong Huynh and Thoai Tran Van

August 4, 2024

Enhancing Welding Training with an AI System for Welding Skill Assessment

Thien Tran Ngoc*
Department of Mechatronics
HCMC University of Technology and Education
Ho Chi Minh City, Vietnam
thientn@hcmute.edu.vn

Phong Duong The†
Department of Mechatronics
HCMC University of Technology and Education
Ho Chi Minh City, Vietnam
phongdt@hcmute.edu.vn

Tanh Tong Huynh†
Department of Mechatronics
HCMC University of Technology and Education
Ho Chi Minh City, Vietnam
20134024@student.hcmute.edu.vn

Thoai Tran Van†
Department of Mechatronics
HCMC University of Technology and Education
Ho Chi Minh City, Vietnam
20134025@student.hcmute.edu.vn

***Abstract*— Visual assessment of welds is crucial in welding training, yet traditional instructor-led evaluations are prone to subjectivity and inconsistent feedback. This research introduces an innovative artificial intelligence (AI)-driven system for objective evaluation and analysis of welding skills through visual inspection. The system utilizes deep learning models, incorporating a Vision Transformer for welding skill classification and a YOLO (You Only Look Once) algorithm for precise weld defect detection. Experimental validation demonstrates the effectiveness of the Vision Transformer, achieving an 89% accuracy in classifying welding skills base on visual inspection, while the YOLO model attained an F1-score of 0.8 in weld defect detection. This AI-powered approach has the potential to enhance the objectivity, consistency, and efficiency of welding skill assessment in educational settings.**

Keywords—welding training, weld defect, AI-based welding assessment

I. Introduction

In the field of welding, particularly manual metal arc welding (MMAW), the quality of welds is a critical determinant of the structural integrity and durability of fabricated products. Ensuring high-quality welds requires a combination of skill, experience, and meticulous attention to detail. However, *traditional welding training methods often rely on subjective assessments by instructors*, which can introduce variability and potential bias into the evaluation process. In particular, as the volume of samples to be assessed increases, the likelihood of undetected errors also rises, potentially compromising the accuracy and reliability of the assessment.

Previous study has explored the integration of AI and machine learning in the welding evaluation. For instance, in a recent study [1], an artificial neural network (ANN) based model was developed to predict weld defects. The authors used data from 289 specimens produced by an automated Gas Metal Arc Welding (GMAW) system. The input data consisted of three welding process measurements: welding current, travel speed, and protective gas flow. The corresponding non-destructive test results for four defect types (underfill, lack of penetration, incomplete fusion, and porosity) served as the output data. The authors of another study [2] conducted experiments on tube-to-tube butt joints, varying process parameters to predict good welds and three types of defects. While these research shares the goal of weld defect detection using machine learning, it differs from our project by using welding process parameters as inputs rather than weld images. In addition to direct identification of surface weld defects, we also use a separate model to classify the welding skill level of students.

Since the MMAW images data can be challenging to obtain, the authors of another previous research [3] have developed a system using Total Focusing Method (TFM) imaging, combining Finite Element (FE) simulations, Deep Convolutional Generative Adversarial Network (DCGAN) for rapid image generation, and the addition of real-world noise to enhance dataset realism. However, to enhance the practical applicability of the model, we elected to utilize real-world MMAW images for both training and evaluation process.

Conversely, comprehensive AI-driven systems that integrate both welding skill classification and detailed defect detection within the specific context of arc welding training, particularly for student-produced welds, are scarce in the literature. To address this gap, this paper presents a novel system that concurrently classifies welding skill levels and detects the weld defects, with the overarching goal of enhancing welding training efficacy. The proposed methodology comprises two distinct models: an image classification model designed to classify students' welding skills based on visual assessment, and an object detection model tailored to identify and localize defects on the weld surface itself. Additionally, we also collected a dataset containing raw MMAW images data.

In summary, the major contributions of this work are as follows:

- We collect a dataset consisting of MMAW product images from students.
- We analyze and categorize the welding skill of the students by doing a visual analysis of the weld bead surface characteristics.
- We identify and visualize defects present on the weld surface.

II. Methodology

A. Dataset description

In Manual Metal Arc Welding (MMAW), five primary types of welded joints exist: butt, corner (angle), edge, fillet (or tee), and lap joints. This research specifically focuses on the butt weld joint configuration, wherein two metal pieces are positioned end-to-end without overlap and subsequently welded along the joint interface. More specifically, the dataset comprises approximately 1200 MMAW images acquired from student welding exercises.

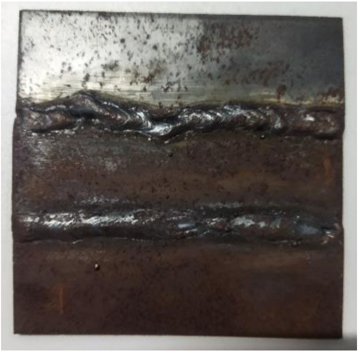

Fig. 1. Real weld sample with two weld beads

1) Welding skill classification data

Based on an analysis of the collected student welding exercise data and consultation with welding experts, MMAW products from students have been categorized into four distinct level classes as summarized in Table I.

TABLE I. WELDING SKILL LEVEL DEFINITION

Level	Definition
A	Perfect weld bead with small or no surface defects.
B	The weld bead has been disrupted by appearing of several surface defects than class A.
C	Weld bead created but the welding line has been disrupted continuously by appearing of more surface defects than class B.
D	Not creating a weld or burn through.

Fig. 2 through Fig. 5 provide illustrative examples for each distinct welding skill level.

Fig. 2. Level A weld joint.

Fig. 3. Level B weld joint.

Welds classified as level B closely approximate the quality of level A welds; however, minor imperfections such as uneven or non-uniform metal ripples, spatter, or slag inclusions may be present on the weld bead surface, resulting in a less aesthetically pleasing appearance.

Fig. 4. Level C weld joint.

Level C welds exhibit non-linearity and discontinuity in the weld bead, characterized by inconsistent shape and dimensions. Additionally, multiple surface defects are typically present.

Fig. 5. Level D weld joint.

Lastly, Level D welds indicate a lack of mastery of the welding technique and/or inappropriate welding parameters. This manifests as a weld bead that is interrupted, discontinuous, or even exhibits burn-through.

2) Weld defect detection data

Following a collaborative analysis of numerous student-produced weld images with welding experts, six recurring defects were identified as the most prevalent among students among 25 common defects listed in a reference literature[4]. Consequently, the system was designed to focus on the detection of these six specific defects, which will be annotated in the dataset if present: burn-through, incomplete fusion (infusion), porosity, surface roughness, slag inclusions, and spatter. Fig. 6 and 7 present representative examples of defect on the weld image dataset.

Fig. 6. Sample with burn-through defect.

Fig. 7. Sample with porosity defect.

B. Proposed workflow

We use a basic workflow for developing and evaluating an AI-driven system for welding skill classification and defect detection is structured into three main stages: Data Preparing, Data Processing, and Model Evaluation as depicted in Fig. 8.

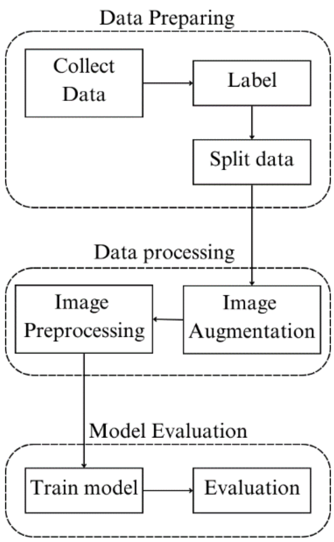

Fig. 8. Models workflow.

1) Data preparing

Following the acquisition of MMAW weld images from student samples, each image was meticulously labeled in accordance with the aforementioned criteria. To mitigate overfitting and model selection bias, both datasets were partitioned into three distinct subsets: 75% for training, 15% for validation, and 10% for testing.

2) Data processing

Due to the diverse capture environments and varying camera settings of the collected MMAW images, data augmentation techniques were employed during the training phase to enhance data diversity and improve model robustness. These techniques included:

- Vertical and horizontal flipping.
- Adding random noise.
- Adding random Gaussian Blur.
- Brightness adjustments.

After augmentation, the train set comprised approximately 2700 MMAW images.

To ensure compatibility with the Vision Transformer [5] model, the following preprocessing steps were applied to the skill classification dataset: resizing to a standard dimension of 224x224 pixels, rescaling by a factor of 1/255, normalizing using a mean of [0.5, 0.5, 0.5], resampling and formatting into tensors.

Conversely, for the defect detection model, we employ fundamental data processing techniques tailored for YOLO [6], specifically leveraging YOLOv8[7] for this work. The preprocessing steps include resizing the images to a fixed dimension of 640x640 pixels to maintain a consistent input size, normalizing pixel values by scaling them to the range [0, 1], and applying padding as necessary to preserve the aspect ratio. Finally, the images are converted into tensors, facilitating efficient training and inference of the YOLOv8 model.

3) Model evaluation metric

For the welding skill classification model, accuracy was defined as the ratio of correct predictions to the total number of predictions.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{1}$$

For the defect detection model, performance was assessed using F1-score metrics[8].

$$Precision = \frac{TP}{TP+FP} \tag{2}$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

$$F1\ score = 2 * \frac{Precision*Recall}{Precision+Recall} \tag{4}$$

Where:

- TP is True positives
- TN is True negatives
- FP is False positives
- FN is False negatives

Furthermore, we also trained the welding skill classification dataset using stable established convolutional neural network (CNN) architectures, including InceptionsV3[9] , Resnet-50[10] and VGG-16[11], to provide a comparative performance assessment against the Vision Transformer model.

III. RESULT AND DISCUSSION

A. Welding skill classification

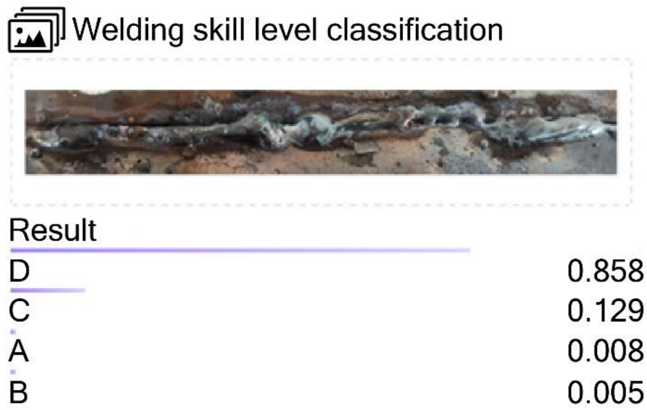

Fig. 9. Example of welding skill classification result.

Fig. 9 presents the results of a visual inspection-based welding skill classification model, analyzing a weld sample that appears rough and uneven. The classification strongly indicates a Level D weld with 85.8% confidence, suggesting low-quality work by an unskilled student.

Table II shows that the Vision Transformer model achieves the best accuracy (89%), outperforming traditional CNN architectures on test dataset. InceptionV3 (82.60%) demonstrates strong performance among CNNs, while Resnet-50 (72.1%) and VGG-16 (70.2%) show moderate effectiveness.

TABLE II. COMPARISON OF MODEL ACCURACIES FOR WELDING SKILL CLASSIFICATION

Model	Accuracy (%)
Vision Transformer	**89**
InceptionV3	82.6
Resnet-50	72.1
VGG-16	70.2

B. Weld defect detection

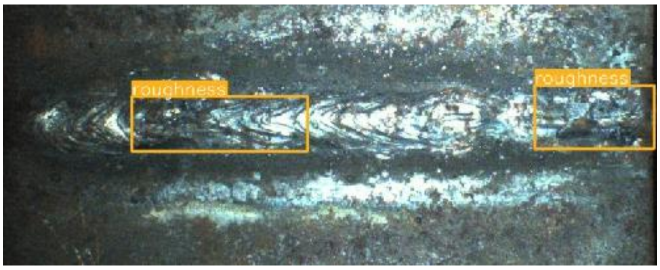

Fig. 10. Testing defect detection model on real weld sample.

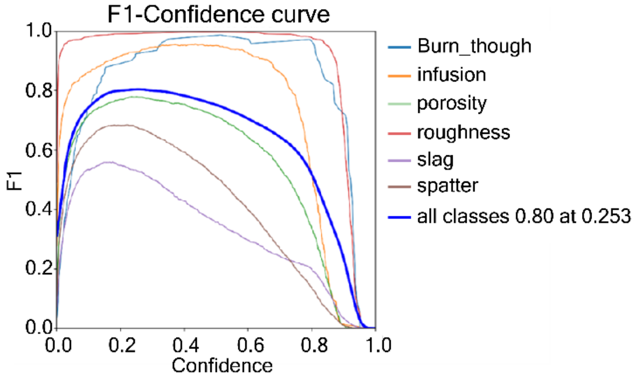

Fig. 11. F1-Confidence curve of defect detection model.

Fig. 11 depicts the performance of the defect detection model. The overall model performance (blue line) reveals a substantial F1 score of 0.80 at a confidence threshold of 0.253. While the model appears effective for most defect types, spatter detection presents the most challenging defect type for the model.

IV. CONCLUTION

This study aimed to develop a system for classification of welding skill levels and detection of weld defects. The system utilizes deep learning models, specifically a Vision Transformer for image classification and YOLO for object detection, trained on a dataset of images captured directly from student-produced welds. Various data processing techniques were employed to enhance model performance. Evaluation metrics demonstrated the effectiveness of the proposed system in assessing weld quality. Future work may include the development of integrated hardware for real-time student interaction and feedback during welding exercises, applied in welding training.

ACKNOWLEDGMENT

We would like to thank Ho Chi Minh City University of Technology and Education (HCMUTE) for the support of this study. We would also like to thank the students who contributed to the manual metal arc welding data collecting process.

REFERENCES

[1] Ho, M. P., Ngai, W. K., Chan, T. W., & Wai, H. W. (2022). An artificial neural network approach for parametric study on welding defect classification. The International Journal of Advanced Manufacturing Technology, 120(1), 527-535.

[2] Moinuddin, S. Q., Hameed, S. S., Dewangan, A. K., Kumar, K. R., & Kumari, A. S. (2021). A study on weld defects classification in gas metal arc welding process using machine learning techniques. Materials Today: Proceedings, 43, 623-628.

[3] Gantala, T., & Balasubramaniam, K. (2021). Automated defect recognition for welds using simulation assisted TFM imaging with artificial intelligence. Journal of Nondestructive Evaluation, 40(1), 28.

[4] Hadzihafizovic, Dzevad. (2013). WELDER'S Visual Inspection HANDBOOK, unpublished.

[5] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

[6] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).

[7] Jocher, G., Chaurasia, A., & Qiu, J. (2023). Ultralytics YOLO (Version 8.0.0) [Computer software]. https://github.com/ultralytics/ultralytics

[8] Wu, T., & Dong, Y. (2023). YOLO-SE: Improved YOLOv8 for remote sensing object detection and recognition. Applied Sciences, 13(24), 12977.

[9] Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2818-2826).

[10] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[11] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.