Weld Defect Detection Model

Zamam Moeez

Alina Iunusova

Abstract

In the welding industry, radiographic inspection is commonly used to detect, classify, and assess various welding defects. However, this technique requires significant human resource allocation. Therefore, in this study, a convolutional neural network (CNN) was trained on radiographic images from the RIAWELC dataset to automate detection and classification of 4 different defect classes: Porosity (PO), Crack (CR), and Lack of Penetration (LP), and No Defect (ND). When presented with an input image from a radiograph patch in the RIAWELC data set, it categorizes the patch into one of 4 weld defect types. The model was assessed on the final test set and achieved an impressive accuracy score of 0.9562 and a low loss score of 0.1229. These results indicate robust generalization and minimal overfitting. Consequently, our proposed model is expected to perform effectively on new, unseen data, leading to cost savings related to labor and improved efficiency in weld defect detection. Additionally, it can help mitigate potential labor shortage issues.

Keywords: Radiographic inspection • Welding Defects • Convolutional Neural Networks (CNNs)

Table of Contents

Introduction

In the modern industry, welding is a critical manufacturing procedure that is widely used in many engineering disciplines, from civil to mechanical to chemical engineering. [1], [2] Welding possesses numerous advantages, which include flexibility and high stiffness of joints. [2] However, strength and service life of welds can be reduced by welding defects, which occur during the production process and can include porosity, slag inclusions, cracks, etc. [1], [2]

To ensure the quality of welds meets industry standards, various non-destructive examination (NDE or NDT) techniques are used to detect welding defects, which can include radiographic, ultrasonic, and electromagnetic testing. [1], [3] Non-destructive testing ensures that welding defects are located, assessed, and repaired in accordance with existing industry standards to prevent safety risks to the public and environment.

Background and Motivation

In traditional radiographic inspection (RT), welding defects are detected and sized manually, which poses a variety of challenges and introduces risk of human error. Radiographic images produced using X-ray or gamma-ray imaging often exhibit poor quality and low contrast, which can make detection and assessment of smaller and thinner defects, such as porosity, challenging. [1], [4] Moreover, the NDT process is costly and time-consuming, as NDT technicians need multiple levels of certification and training, and the assessment and reporting process can often take several weeks due to backlog. [5]

Therefore, there is high demand for automation of welding defect detection, sizing, and assessment. Use of deep learning-based approaches will significantly simplify, cheapen, and speed up the assessment process, as well as increase the accuracy of defect detection and sizing. Moreover, use of artificial intelligence will allow certified NDT technicians to focus on more complicated and ambiguous cases that may not be easily resolved using computer vision software.

Considering the background information presented above, the objective of this project is to employ convolutional neural networks (CNNs) to automatically detect and classify three common welding defect types: porosity, lack of penetration, and crack.

Related Work

Traditional Inspection Methods

NDT/NDE or Non-Destructive Testing are a set of technologies used in the industry to inspect and examine materials to identify imperfections and defects in their structure without affecting their structural integrity. In other words, while in other engineering testing methods, particularly the tensile stress test, the materials are deformed until failure, non-destructive testing methods preserve the structure of the material during identification and measurement of defects or imperfections. Materials maintain their original properties and are thus suitable for service after the testing is complete. [3]

A variety of NDT techniques exist, the most common ones used on ferrous materials are Radiographic Testing (RT), Ultrasonic Testing (UT), Magnetic Particle Inspection (MPI), Eddy Current or Electromagnetic Testing (ET), and Penetrant Testing (PT). [3] The description of each method, as well as their advantages and disadvantages are included in Table 1.

As other NDT techniques do not provide a pictorial result, this model and paper focus on the radiographic (RT) inspection technique, aiming to optimize the testing process and shorten the assessment time by employing imaging techniques to automatically detect and classify the material discontinuities shown on the RT films.

Table 1: Descriptions, advantages, and disadvantages of different NDT techniques. [3], [6]

Inspection Method	Description	Advantages	Disadvantages
RT	X-rays or Gamma rays are directed through the tested material and onto a film. Internal voids, which show up as darkened areas on the RT film, can then be detected, classified, and sized.	• Suitable for ferrous and non-ferrous materials • Provides a permanent pictorial record of the test results • Most applicable for thin-walled materials and emptied vessels	• Cannot detect surface defects • Not suitable for thick-walled materials or filled liquid tanks • Film processing and interpretation by a trained professional is required, adding to assessment time
UT	Ultrasonic waves are passed through the material. Upon encountering a defect, the sound wave will reflect against it back towards the material surface, transforming into an electrical signal which can then be displayed on a screen.	• Can detect both surface and subsurface defects • Suitable for thick materials and filled liquid vessels, such as odorant tanks, dump tanks, separators, etc. • Instantaneous test results	• Higher skill and training requirements compared to other methods. • Defects parallel to the sound wave can be missed • Thin-walled materials can pose challenges
MPI	Magnetic current is applied to the tested material. Flux lines will be interrupted when encountering a discontinuity, creating a secondary magnetic field and making it possible to pinpoint the defect using magnetic particles.	• Simple to apply and operate • Test results are visible on the material surface with a naked eye • Can detect small defects	• Application limited to ferromagnetic materials • Can only detect surface or non-deep subsurface defects • Improper application can lead to defects being missed
ET	Alternating current is applied to two coils, generating a magnetic field and inducing eddy currents in the material. A change in the material properties or geometry disrupts the flow of eddy currents and is detected as a change in coil impedance.	• Can detect small defects • Instantaneous test results • Applicable to complex shapes and sizes of materials	• Restricted to conductive materials • Limited to surface and near-surface defects • Change in material properties can interfere with defect detection
PT	Material is coated in surface penetrant, which gets drawn into surface defects using capillary action. Excess penetrant is then removed, and a developer is applied, drawing penetrant from the defect and forming a colored or fluorescent indication of the defect.	• Simple to use • Applicable to ferrous and non-ferrous materials • Instantaneous results are visible with a naked eye	• Extensive surface cleaning and preparation is required • Limited to surface defects • Limited to non-porous materials

Dataset

Dataset Description

The Weld Defect Detection Model (WDDM) was designed to identify and classify weld defects based on the RIAWELC dataset [7]. The model is a convolutional neural network (CNN), which is well-suited for image-based classification due to its ability to extract hierarchical features from images, making it ideal to detect welding defects.

The RIAWELC dataset features X-ray weld images taken during an RT inspection in an industrial setting, with each image initially sized at 2000×8640 pixels and saved in JPEG format. The images were first processed to isolate the weld bead by removing the background. A software tool then extracted tiles from these cropped images to focus on regions where defects could appear. Tiles were varied in size from 32×32 to 150×150 pixels to capture different defect types.

Smaller tiles (e.g. 32×32) were effective for detecting small or closely spaced defects like porosity (PO), but larger tiles (e.g. 150×150) were needed for larger defects such as cracks (CR) and lack of penetration (LP). An 80×80-pixel tile size was chosen as a balanced option, providing good detail for a range of defects.

To fit popular Convolutional Neural Networks (CNNs) like SqueezeNet, AlexNet, GoogleNet, ResNet, and VGG16, the tiles were resized to 224×224 pixels while maintaining their aspect ratio. To enhance defect visibility and edge clarity, contrast-limited adaptive histogram equalization was applied to the images.

The organization of the RIAWELC dataset is presented in Table 2. The dataset is split into four defect classes: Crack (CR), Porosity (PO), Lack of Penetration (LP), and No Defect (ND). It contains a total of 24,407 samples, distributed into training, validation, and test sets.

Table 2: The RIAWELC dataset organization. [7]

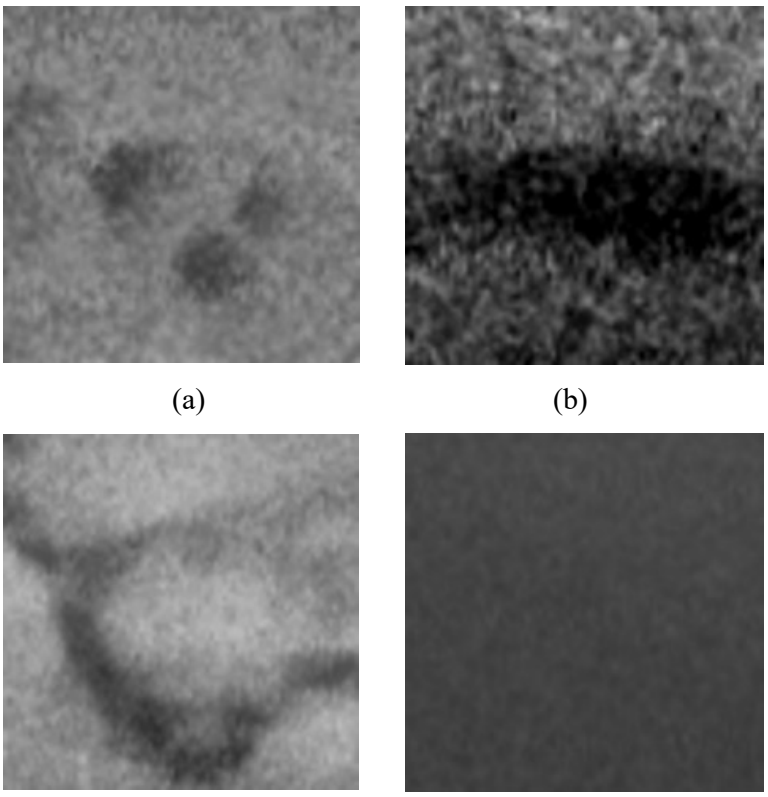
Defect	Training	Validation	Test	Total
Crack (CR)	4,962	1,908	765	7,635
Porosity (PO)	4,108	1,580	632	6,320
Lack of Penetration (LP)	2,893	1,113	446	4,452
No Defect (ND)	3,900	1,500	600	6,000
Total	15,863	6,101	2,443	24,407

The RIAWELC dataset is well-suited for training the Weld Defect Detection Model (WDDM) due to its comprehensive and realistic setup. Collected from real-world industrial environments, the dataset provides a diverse range of welding defect images, ensuring a robust representation of various scenarios encountered in practice. Its large size and balanced distribution of defect types offer a strong foundation for training deep learning models. The dataset's adherence to standard CNN input sizes and preprocessing techniques ensures smooth integration with advanced models, enhancing the accuracy of defect classification.

Sample images from the RIAWELC dataset are shown in Figure 1. These images represent 4 defect classes as follows [8]:

(a) porosity, which happens when gas bubbles get trapped inside the weld
(b) lack of penetration, which forms when the weld material doesn't extend to full thickness of the joint
(c) crack, when the material gets separated inside the weld or on the surface
(d) no defect

Figure 1: Sample images from the RIAWELC dataset. [7]



Data Preprocessing

The datasets for training, validation, and testing were loaded using TensorFlow's **image_dataset_from_directory**, which is well-suited for organizing data in a folder structure, where each subdirectory represents a different class. This method simplified the dataset loading process and automatically inferred labels from directory names. The labels were one-hot encoded using the categorical label mode, necessary for multi-class classification tasks. The training and validation datasets were shuffled to ensure that the model did not learn any order bias, while the test dataset was not shuffled as this was unnecessary for evaluation.

The images were then resized to (28, 28) pixels to standardize the input dimensions, which is essential for neural networks that require uniform input shapes. The smaller input size also helped reduce computational complexity and sped up training while retaining features needed for effective classification.

Finally, the datasets were pre-fetched using TensorFlow's **AUTOTUNE** to optimize performance by ensuring efficient data loading during training, preventing bottlenecks between data fetching and GPU computations.

Methodology

Model Architecture

The weld defect detection model architecture consists of seven layers, structured as a sequential Convolutional Neural Network (CNN) specifically designed for the classification of weld defects. The model begins with a Conv2D layer featuring 16 filters, each with a 3x3 kernel and employing the ReLU activation function. This layer is essential for extracting local spatial features from the input images. It is succeeded by a MaxPooling2D layer with a 2x2 pooling size, which effectively reduces the spatial dimensions of the feature maps, thus enhancing computational efficiency and mitigating the risk of overfitting.

Subsequently, a second Conv2D layer is introduced, consisting of 32 filters with a 3x3 kernel and ReLU activation. This layer facilitates detection of more intricate patterns within the feature maps, further deepening the model's representational capacity. Another MaxPooling2D layer follows, maintaining the 2x2 pooling size to continue the down sampling process.

The model architecture proceeds with a Flatten layer that converts the 2D feature maps into a one-dimensional vector, preparing the data for the subsequent fully connected layers. This is followed by a Dense layer consisting of 64 neurons, utilizing ReLU activation to capture complex feature representations. The final Dense layer comprises 4 neurons corresponding to the four distinct defect classes: Crack, Lack of Penetration, No Defect, and Porosity, applying a softmax activation function to enable probabilistic classification across these categories.

Hyperparameter optimization for the model was conducted using **Keras Tuner**, which facilitated determination of the number of filters and kernel sizes. The **Adam** optimizer was employed due to its adaptive learning rate capabilities, and the model's loss function was defined as categorical cross entropy, suitable for the multi-class classification task.

Model Training and Evaluation

The training of the weld defect detection model involved a systematic approach to ensure effective learning and validation. The model was trained over a specified number of epochs (11), using a combination of the training and validation datasets. During each epoch, the model adjusted its weights based on the training data while simultaneously evaluating its performance on the validation dataset to monitor for overfitting. The training process employed the **Adam** optimizer, chosen for its adaptive learning rate, which helps achieve faster convergence. Categorical cross entropy was utilized as the loss function, which is appropriate for multi-class classification problems, ensuring that the model effectively distinguishes between different weld defect classes.

To evaluate the model’s performance, the model was tested on a separate test dataset after training completion. The evaluation metrics focused primarily on accuracy to provide insights into model’s performance on unseen data. The evaluate method was employed to compute the final loss and accuracy, allowing for a quantitative assessment of the model's effectiveness in detecting weld defects. Additionally, training and validation accuracy were plotted over the epochs to visualize the learning progress and identify potential overfitting trends. This comprehensive evaluation framework ensured that the model is not only trained effectively but is also robust and reliable for practical applications in weld defect detection.

Experimental Results

Performance on Training, Validation, and Test Data

Training Results: During training, the model's loss decreased steadily from 1.3135 in Epoch 1 to 0.1460 by Epoch 11, while accuracy increased from 56.14% to 94.89%. These trends demonstrate effective model learning and optimization over the course of training.

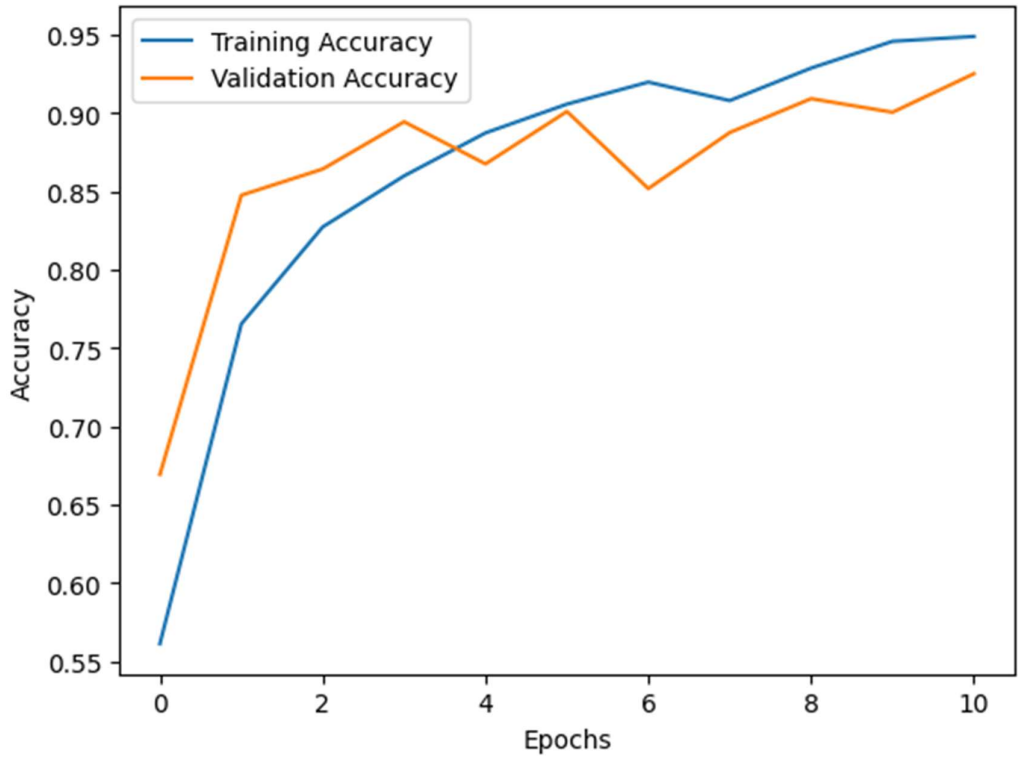
Validation Results: For the validation set, loss initially dropped from 0.7964 in Epoch 1 to 0.3249 in Epoch 4, but then increased to 0.5378 by Epoch 7 before stabilizing at 0.2719 by Epoch 11. Validation accuracy improved from 66.94% to 92.51% over the same period. The fluctuations in loss and accuracy between Epochs 5 and 8 indicate potential overfitting or challenges with certain validation data, though the overall performance improved by the end of training.

Test Result: The model achieved a final loss of 0.1229 and an accuracy of 95.62% on the test set, indicative of its high performance. This accuracy underscores the model's strong generalization capabilities, demonstrating its effectiveness in accurately classifying weld defects in previously unseen data. The consistent results reinforce the model's robustness and reliability, highlighting its potential for practical applications in the field of RT inspection.

Table 3: Model loss and accuracy on training and validation datasets per epoch.

Epoch	Training Loss	Training Accuracy	Validation Loss	Validation Accuracy
1	1.3135	0.5614	0.7964	0.6694
2	0.6371	0.7654	0.4716	0.8476
3	0.4648	0.8273	0.3884	0.8643
4	0.3836	0.8599	0.3249	0.8944
5	0.3170	0.8873	0.3717	0.8676
6	0.2662	0.9057	0.3024	0.9010
7	0.2292	0.9198	0.5378	0.8518
8	0.2757	0.9080	0.3502	0.8876
9	0.2162	0.9287	0.3087	0.9092
10	0.1589	0.9458	0.3517	0.9005
11	0.1460	0.9489	0.2719	0.9251

Figure 2: Model accuracy on training and validation datasets per epoch.



Discussion

Analysis of Results

The training results suggest effective learning, with a significant reduction in loss from 1.3135 to 0.1460 and an increase in accuracy from 56.14% to 94.89% over 11 epochs. This indicates that the model successfully adapted to the training data.

However, the validation results displayed fluctuations, with the loss initially decreasing from 0.7964 to 0.3249, then increasing to 0.5378 before stabilizing at 0.2719. Despite an overall increase in validation accuracy from 66.94% to 92.51%, the instability during Epochs 5 to 8 may suggest overfitting or challenges with specific validation data.

The test results revealed a final loss of 0.1229 and an accuracy of 95.62%, demonstrating strong generalization. This performance confirms the model's robustness for practical applications in weld defect detection. Addressing the validation fluctuations could further enhance its efficacy.

Challenges and Limitations

Training the WDDM model presented numerous challenges and constraints due to limited resources. As students, we utilized Google Colab, which offered restricted access to GPU resources, particularly the T4 GPU. This impacted the pace of training and limited our capacity to experiment with more intricate models or conduct extensive hyperparameter tuning. The dataset was relatively small and the likelihood of accessing larger supplementary datasets was low, as companies typically keep RT inspection results confidential. This constraint hindered the model's learning potential and heightened the risk of overfitting.

Due to limitations of the RIAWELC dataset, the WDDM model was trained to detect three major defect types: porosity, crack, and lack of penetration. The model will not be able to distinguish these defects from other common welding defects, such as lack of fusion, slag, or undercut, and may thus misclassify these defects as one of the three defects it is trained to detect. Moreover, the model is not trained to distinguish different types of porosity and is thus only suitable to detect and label single or uniformly scattered porosity. It may misclassify cluster or wormhole porosity as crack or lack of penetration.

As other NDT techniques do not produce a hard-copy pictorial result, WDDM is not suitable for use with other common industrial NDT methods. To employ automatic defect detection in other visual NDT methods, such as Penetrant Testing (PT) and Magnetic Particle Testing (MPI), separate field-applicable AI models must be developed.

Conclusion and Future Work

Summary of Findings

The sequential Convolutional Neural Network (CNN) weld defect detection model showed promising results during the training, validation, and testing phases. The training process demonstrated effective learning, with a substantial decrease in loss from 1.3135 to 0.1460 and an increase in accuracy from 56.14% to 94.89% over 11 epochs. Although the validation results initially improved, fluctuations in loss and accuracy indicated potential overfitting in certain epochs. However, the model exhibited strong generalization on the test set, achieving a final loss of 0.1229 and an impressive accuracy of 95.62%. These results affirm the model's robustness and reliability for practical applications in weld defect detection. Addressing the validation inconsistencies could further enhance its performance in

real-world scenarios. Overall, this study underscores the model's potential to improve quality assurance processes within the welding industry.

Implications

The WDDM has a net positive impact on the NDT inspection industry. The model can be employed in an industrial setting to simplify and speed up the defect classification and evaluation process, as well as lower the workload of NDT technicians and overall backlog of NDT inspections. In turn, this will help streamline many inspection, installation, and improvement projects all over Canada and abroad and allow for earlier in-service dates for new or existing installations.

Future Research Directions

To elevate the performance of the weld defect detection model, several strategies can be implemented. Firstly, the application of data augmentation techniques such as rotation, scaling, and brightness adjustment can heighten dataset variability, thereby helping to address overfitting. Additionally, the exploration of more sophisticated architectures such as ResNet or DenseNet may enhance feature extraction and classification accuracy. Introducing regularization techniques, such as dropout layers or L2 regularization, can further mitigate overfitting by promoting the development of more generalizable features. Lastly, conducting comprehensive hyperparameter optimization could fine-tune the model’s performance and improve training efficiency.

In addition to refining the existing model, alternative approaches such as segmentation could be explored for improved defect detection. Implementing segmentation models like U-Net or Mask R-CNN could enable pixel-level classification of weld defects, offering detailed insights into defect boundaries and sizes. However, this approach presents challenges, including the requirement for substantial computational resources and the generation of segmentation masks for each weld image. This process can be both time-consuming and resource-intensive, necessitating significant human effort to label the entire dataset of 24,000 images. Object detection frameworks, such as YOLO (You Only Look Once) or SSD (Single Shot Multibox Detector), could enable real-time detection of multiple defect types within images. Leveraging transfer learning through pre-trained models on similar tasks may enhance performance, particularly in scenarios with limited training data, leading to faster convergence and improved accuracy. Furthermore, exploring unsupervised learning methods for anomaly detection can identify defects without extensive labeled

datasets, making the approach adaptable to various defect types. Finally, integrating image analysis with other data sources, such as sensor data from welding machines, could enhance overall defect detection capabilities by providing additional contextual information.

References

[1] B. Zhang, X. Wang, J. Cui and X. Yu, "Automated Welding Defect Detection using Point-Rend ResUNet," *Journal of Nondestructive Evaluation,* vol. 43, 2024.

[2] Z. Shu, A. Wu, Y. Si, H. Dong, D. Wang and Y. Li, "Automated identification of steel weld defects, a convolutional neural network improved machine learning approach," *Frontiers of Structural and Civil Engineering,* vol. 18, no. 2, pp. 294-308, 2024.

[3] G. D. Mark Willcox, "A Brief Description of NDT Techniques," 2003. [Online]. Available: https://www.insightndt.com/papers/technical/t001.pdf. [Accessed 31 August 2024].

[4] C. Ajmi, J. Zapata, S. Elferchichi and K. Laabidi, "Advanced Faster-RCNN Model for Automated Recognition and Detection of Weld Defects on Limited X-Ray Image Dataset," *Journal of Nondestructive Evaluation,* vol. 43, 2023.

[5] "Initial NDT Certification Examinations," Government of Canada, 12 July 2024. [Online]. Available: https://natural-resources.canada.ca/science-and-data/non-destructive-testing/about-certification-ndt/learn-about-becoming-certified-ndt/initial-ndt-examinations/19519. [Accessed 19 August 2024].

[6] "Nondestructive Evaluation Techniques," Iowa State University Center for Nondestructive Evaluation, [Online]. Available: https://www.nde-ed.org/NDETechniques/index.xhtml. [Accessed 31 August 2024].

[7] B. Totino, F. Spagnolo and S. Perri, "RIAWELC: A Novel Dataset of Radiographic Images for Automatic Weld Defects Classification," *International Journal of Electrical and Computer Engineering Research,* vol. 3, no. 1, pp. 13-17, 2023.

[8] "The Seven Most Common Welding Defects, Causes and Remedies," Australian Welding Institute, 27 November 2019. [Online]. Available: https://welding.org.au/articles/the-seven-most-common-welding-defects-causes-and-remedies/. [Accessed 21 September 2024].