Al'adzkiya International of Computer Science and Information Technology (AIoCSIT) Journal
Vol. 7, No. 1, April 2026, pp. 1~12
ISSN: 2722-0001 

# Performance Analysis of YOLO11 for Welding Defect Detection Under Low-Light Conditions

**Yonky Pernando[1], Raymond Erz Saragih[2], Masparudin[3], Agus Suwandi[4], Ihsan Verdian[5], Fazlul Rahman[6], Ilwan Syafrinal[7]**
[1,2,4,5]Department of Informatics, [3,7]Department of Software Engineering, [6]Department of Information, Universitas Universal, Batam, Riau Island, Indonesia

**ABSTRACT**

This study aims to analyze the impact of image enhancement techniques on welding defect detection performance using a deep learning-based YOLO11L model. The dataset consists of 1392 welding images categorized into four classes: Good, Crack, Porosity, and Bad, with a significant class imbalance. Five image enhancement methods were evaluated, namely Zero-DCE, RETINEX, CLAHE, Supervision, and Gamma Correction, and compared against a no-enhancement baseline. Image quality was assessed using SSIM, and PSNR, while detection performance was evaluated using Precision, Recall, F1-Score, and mAP50. The results show that Gamma Correction achieves the best image quality improvement, with an average SSIM of 0.569, and a PSNR of 18.862 dB. However, contrasting results are observed at the detection stage, where 0.7772 and 0.6969, respectively, for the Gamma Correction-based model while for the baseline model without enhancement outperforms the enhanced model, achieving a mAP50 of 0.7098 and an F1-Score of 0.6965. This finding reveals a paradox where improved visual image quality does not necessarily lead to better object detection performance. This study highlights the importance of end-to-end evaluation in computer vision systems, particularly in industrial inspection applications, and demonstrates that original images, which are closer to the pretrained data distribution, may yield better detection results than heavily enhanced images.
**Keyword: Welding Defect Detection, YOLO11, Image Enhancement, Gamma Correction, Deep Learning, Low-Light Imaging**



***Corresponding Author:***
Yonky Pernando,
Department of Informatics,
Universitas Universal, Batam, Riau Island, Indonesia,
Pasir Putih, Kelurahan Sadai, Kecamatan Bengkong, Batam, Riau Island, Indonesia, 29432.
Email: yongkyfernando194@gmail.com

## 1. INTRODUCTION (10 PT)

Welding inspection plays a critical role in ensuring structural integrity and safety across various industrial sectors, including construction, shipbuilding, and automotive manufacturing. Welding defects such as crack, porosity, and other discontinuities can significantly degrade mechanical performance and lead to catastrophic failures if left undetected[1], [2]. Traditional inspection methods, which rely heavily on human expertise, are often time-consuming, subjective, and prone to inconsistency, especially under complex environmental conditions[3], [4]. Consequently, there has been a growing interest in leveraging computer vision and deep learning techniques to automate and enhance welding defect detection processes.

Recent advances in deep learning-based object detection, particularly the YOLO (You Only Look Once) family, have demonstrated remarkable performance in real-time industrial inspection tasks due to their efficiency and high detection accuracy [5], [6]. These models adopt a single-stage detection paradigm, enabling simultaneous localization and classification of objects within an image[7], [8]. Despite their success, the performance of such models remains highly dependent on the quality and characteristics of the input images. In practical industrial environments, welding images are often captured under suboptimal conditions, including low illumination, noise, and uneven lighting, which can significantly degrade detection performance [9], [10], [11].

To address these challenges, various image enhancement techniques have been proposed to improve visual quality prior to model inference. Methods such as Zero-DCE, RETINEX, CLAHE, and

Gamma Correction aim to enhance brightness, contrast, and structural details in low-light images [12], [13], [14], [15], [16]. These approaches are typically evaluated using image quality assessment metrics such as Structural Similarity Index Measure (SSIM), and Peak Signal-to-Noise Ratio (PSNR). Higher values of SSIM and PSNR, are generally interpreted as indicators of better visual reconstruction [11]. As a result, it is commonly assumed that improving these metrics will lead to better downstream performance in tasks such as object detection and classification.

However, this assumption remains insufficiently validated, particularly in the context of real-world industrial datasets characterized by limited sample sizes and class imbalance. Existing studies often focus on evaluating enhancement techniques in isolation, without conducting comprehensive end-to-end analysis that links image quality improvements to actual task performance [17]. Moreover, image enhancement operations may alter the original data distribution, potentially creating a mismatch with the pretrained feature representations learned by deep neural networks[18], [19]. This distribution shift can adversely affect the model's ability to generalize, especially when the enhanced images deviate significantly from natural image statistics used during pretraining.

In addition, the impact of dataset-related challenges, such as class imbalance and limited data availability, further complicates the relationship between image preprocessing and detection performance. Minority classes, such as crack and porosity, are often underrepresented in industrial datasets, leading to biased model predictions and reduced sensitivity to critical defects [20], [21]. These factors highlight the need for a more holistic evaluation framework that considers not only image quality metrics but also task-specific performance indicators.

Motivated by these gaps, this study presents a comprehensive analysis of the impact of image enhancement techniques on welding defect detection using the YOLO11L model. Unlike prior work, this research adopts an end-to-end evaluation strategy that integrates both image quality assessment (SSIM, and PSNR) and detection performance metrics (Precision, Recall, F1-Score, and mAP50). The study utilizes a real-world dataset of welding images comprising four classes (Good, Crack, Porosity, and Bad) with inherent class imbalance, thereby reflecting practical industrial conditions.

The main contribution of this work lies in revealing a counterintuitive phenomenon: image enhancement methods that achieve superior performance in visual quality metrics do not necessarily lead to improved object detection performance. Specifically, while Gamma Correction demonstrates the best results in SSIM and PSNR, the baseline model trained on original, non-enhanced images achieves significantly higher detection performance. This finding challenges the conventional assumption that better image quality directly translates to better model performance and underscores the importance of task-oriented evaluation in computer vision pipelines.

Furthermore, this study provides critical insights into the interplay between preprocessing techniques, data distribution, and model generalization. The results suggest that preserving the natural characteristics of input images, which are more aligned with pretrained model distributions, may be more beneficial than applying aggressive enhancement transformations. These findings contribute to a deeper understanding of preprocessing strategies in deep learning-based industrial inspection systems and offer practical guidelines for designing robust and reliable welding defect detection frameworks.

## 2. RESEARCH METHOD

### A. Research Framework

This study adopts an end-to-end evaluation approach to analyze the impact of image enhancement techniques on welding defect detection performance using deep learning. The research workflow consists of four main stages: (1) dataset collection and preparation, (2) application of image enhancement methods, (3) object detection model training using YOLO11L, and (4) performance evaluation using both image quality metrics and object detection metrics. This approach enables a comprehensive analysis of the relationship between image quality and downstream task performance.

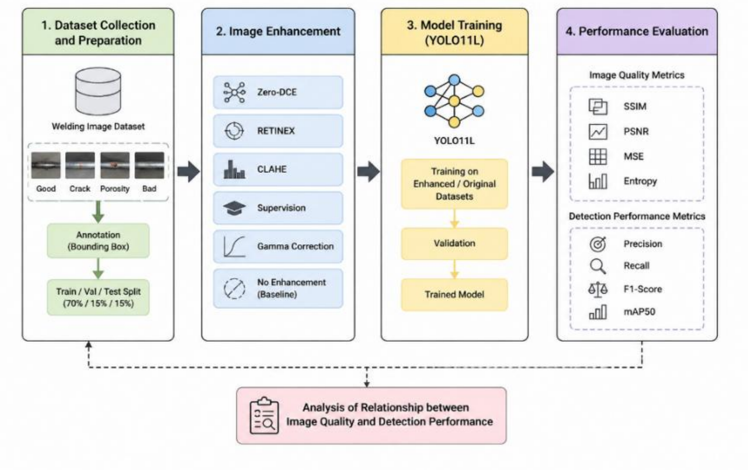


Fig 1. Research Framework end-to-end evaluation approach to analyze the impact of image enhancement techniques on welding defect detection

## B. Dataset Description

The dataset used in this study consists of 1392 welding images categorized into four classes: *Good*, *Crack*, *Porosity*, and *Bad*. The dataset exhibits a significant class imbalance, where the *Bad* and *Good* classes dominate, while *Crack* and *Porosity* contain substantially fewer samples. This condition reflects real-world industrial inspection scenarios, where specific defects are relatively rare.
The dataset is split into training, validation, and testing sets using a 70:15:15 ratio with a *stratified split* approach to preserve class distribution. All images are annotated using bounding box format for object detection tasks

## C. Image Enhancement Methods

This study evaluates five commonly used image enhancement methods for improving image quality under low-light conditions:

a. Zero-DCE: a deep learning-based method that estimates pixel-wise light enhancement curves without paired training data.
b. RETINEX: a method based on human visual perception theory that separates illumination and reflectance components.
c. CLAHE (Contrast Limited Adaptive Histogram Equalization): a local contrast enhancement technique with noise amplification control.
d. Supervision-based Enhancement: a supervised learning-based approach for improving image quality.
e. Gamma Correction: a simple non-linear transformation to adjust image brightness.

As a comparison, a no-enhancement (baseline) condition is also included to evaluate the direct impact of each method on detection performance.

## D. Image Quality Evaluation Metrics

The quality of enhanced images is evaluated using four quantitative metrics:

a. Structural Similarity Index Measure (SSIM) to assess structural similarity between enhanced and reference images.
b. Peak Signal-to-Noise Ratio (PSNR) to measure reconstruction quality in decibels (dB).

The combination of these metrics provides a comprehensive evaluation of image quality from multiple perspectives.

### E. Object Detection Model

The object detection model used in this study is YOLO11L, a state-of-the-art model based on the YOLO architecture capable of real-time object detection. The model contains approximately 25 million parameters and utilizes a transfer learning approach with pretrained weights.

The training configuration is as follows:

1. Epochs: 50
2. Batch size: 16
3. Image size: 640 × 640
4. Optimizer: AdamW
5. Learning rate: 0.00125
6. Data augmentation: horizontal flip, Mosaic, and Albumentations (Blur, MedianBlur, CLAHE)

Two training scenarios are conducted:

a. Dataset with preprocessing using the selected best image enhancement method
b. Dataset without enhancement (baseline)

### F. Detection Performance Evaluation

Detection performance is evaluated using standard object detection metrics:

a. Precision: the proportion of correct predictions among all predictions
b. Recall: the proportion of correctly detected objects among all ground truth objects
c. F1-Score: the harmonic mean of Precision and Recall
d. mAP50 (mean Average Precision at IoU threshold 0.5): the primary metric for overall detection accuracy

Additionally, F1-Confidence and Precision-Confidence curves are analyzed to determine the optimal confidence threshold during inference.

### G. Experimental Setup

All experiments are conducted using an NVIDIA Tesla T4 GPU. The implementation is developed in Python using deep learning frameworks that support YOLO-based training. Training and evaluation are performed under consistent configurations to ensure fair comparison across all methods.

### H. Research Contribution Workflow

Overall, this study integrates image quality evaluation and object detection performance within a unified framework. This structured approach enables a deeper understanding of the relationship between image preprocessing techniques and model performance, and provides a foundation for analyzing discrepancies between image quality metrics and detection performance outcomes.

## 3. RESULTS AND DISCUSSION

### A. Dataset Distribution

The dataset distribution reveals a significant class imbalance, where the Bad (35%) and Good (25%) classes dominate, while Crack (20%) and Porosity (20%) contain substantially fewer samples. This imbalance reflects real-world industrial scenarios, where specific defects occur less frequently. However, such imbalance can introduce bias toward majority classes, potentially degrading detection performance, particularly for minority classes.

Table 1. The dataset distribution

| Class | Train | Validasi | Test | Total |
|---|---|---|---|---|
| Good | 245 | 52 | 52 | 349 |
| Crack | 195 | 41 | 41 | 277 |
| Porosity | 195 | 41 | 41 | 277 |
| Bad | 341 | 74 | 74 | 489 |
| Total | 976 | 208 | 208 | 1392 |

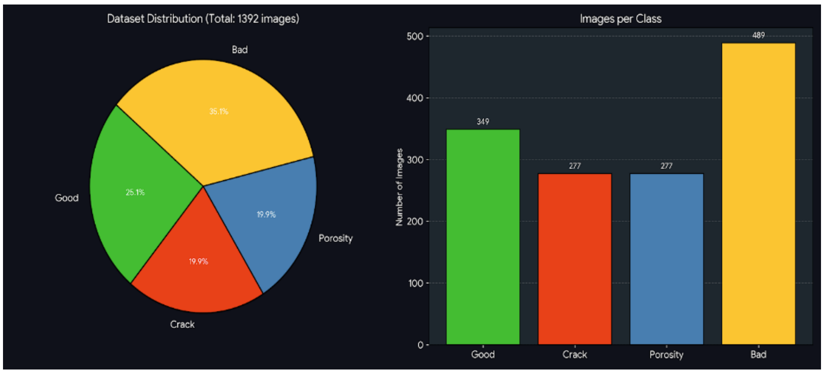


Fig 2. Dataset Distribution: Pie Chart and Bar Chart of Image Counts per Class

The dataset used in this study consists of 1392 welding images collected from Google Drive using the Google Drive API v3. The dataset covers four categories of weld conditions: Good (high-quality weld), Crack, Porosity, and Bad (poor-quality weld). These four categories represent the full spectrum of conditions commonly encountered during the welding inspection process in the manufacturing industry.

## B. Image Enhancement Quality Analysis

The image enhancement analysis phase aims to quantify the effectiveness of various algorithms in restoring the visual quality of images subjected to simulated low-light conditions. This evaluation is critical for ensuring that the subsequent defect detection pipeline receives high-quality input, regardless of environmental lighting challenges.

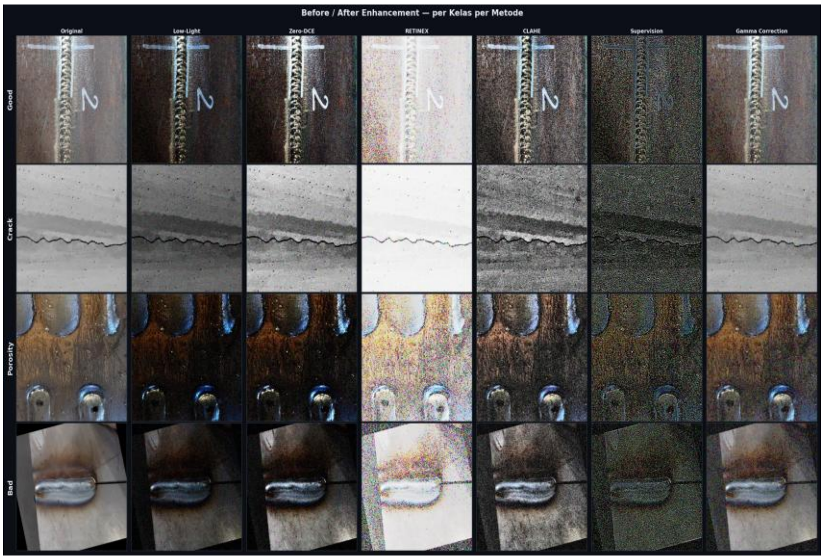


Fig 3. Visualization Before/After Enhancement, Original, Low-Light, Zero-DCE, RETINEX, CLAHE, Supervision, and Gamma Correction Images per Class

Evaluation MetricsTo provide a comprehensive assessment, the study utilizes four primary objective metrics:

a. SSIM (Structural Similarity Index Measure): Evaluates the preservation of luminance, contrast, and structural information relative to the reference image.

$$\text{SSIM}\ (\text{x}, \text{y}) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2)(\sigma_y^2 + \sigma_y^2 + C_2)} \quad (1)$$

a. $x$ adalah citra referensi (*reference image*)
b. $y$ adalah citra hasil pemrosesan (*test image*)
c. $\mu_x, \mu_y$ nilai rata-rata (mean) dari citra x dan y
d. $\sigma_y^2\ dan\ \sigma_y^2$ = variansi
e. $\sigma_{xy}$ = kovarians
f. $C_1, C_2$ = konstanta stabilisasi

Table 2. SSIM Values by Enhancement Method and Class

| Method | Good | Crack | Porosity | Bad |
|---|---|---|---|---|
| Zero-DCE | 0.5067 | 0.4081 | 0.4696 | 0.4318 |
| RETINEX | 0.4300 | 0.3519 | 0.3674 | 0.3864 |
| CLAHE | 0.4524 | 0.3323 | 0.4122 | 0.3718 |
| Supervision | 0.2206 | 0.1302 | 0.1727 | 0.1434 |
| **Gamma Correction** | **0.6645** | **0.4972** | **0.5704** | **0.5457** |
| No Enhancement | 0.4975 | 0.3990 | 0.4383 | 0.4093 |

SSIM is the primary metric for evaluating the structural similarity between enhanced and original images, with values ranging from 0 to 1. Evaluation results demonstrate that Gamma Correction consistently outperforms all other methods across every category, achieving an average SSIM of 0.569. It recorded the highest scores for each class: 0.6645 for Good, 0.4972 for Crack, 0.5704 for Porosity, and 0.5457 for Bad. These findings establish Gamma Correction as the most stable and effective method for preserving structural integrity in low-light welding imagery.

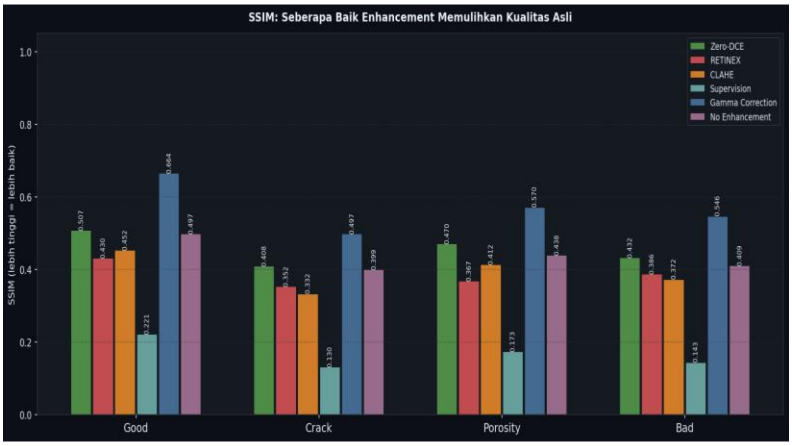

Fig 4. SSIM Comparison: All Enhancement Methods per Class

In contrast, other methods showed varying or subpar performance. Zero-DCE emerged as the second-best performer for the "Good" class (0.5067) but maintained mid-level scores elsewhere, while CLAHE displayed inconsistent results, dropping significantly to 0.3323 in the "Crack" category. The Supervision method recorded the lowest scores overall, reaching a minimum of 0.1302, indicating that its supervised training failed to produce results structurally similar to the ground truth within this specific dataset.

b. PSNR (Peak Signal-to-Noise Ratio): Measures the ratio between the maximum possible power of a signal and the power of corrupting noise, expressed in decibels (dB).

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{MAX_1^2}{MSE}\right) \quad (2)$$

a. $MAX_I$ adalah nilai maksimum intensitas piksel citra (misal 255 untuk 8 bit image)
b. *MSE* (*Mean Squared Error*) merupakan rata-rata kesalahan kuadrat antara citra referensi dengan citra hasil peningkatan, yang ditunjukkan pada persamaan (2.2).

Table 3. PSNR Values (dB) by Enhancement Method and Class

| Method | Good (dB) | Crack (dB) | Porosity (dB) | Bad (dB) |
|---|---|---|---|---|
| Zero-DCE | 12.90 | 14.44 | 14.38 | 13.85 |
| RETINEX | 8.78 | 8.16 | 8.52 | 8.32 |
| CLAHE | 14.62 | 15.35 | 15.39 | 15.29 |
| Supervision | 8.36 | 8.70 | 8.63 | 8.57 |
| **Gamma Correction** | 19.41 | 18.65 | 18.66 | 18.72 |
| No Enhancement | 10.46 | 12.00 | 11.91 | 11.47 |

Peak Signal-to-Noise Ratio (PSNR) serves as a fundamental metric for assessing image reconstruction quality, with higher decibel ($dB$) values signifying lower distortion and superior signal fidelity. Quantitative analysis reveals that Gamma Correction achieves the most significant performance, maintaining a dominant average PSNR of 18.862 dB across all test cases. This indicates that Gamma Correction is exceptionally effective at enhancing low-light welding images while suppressing the introduction of unwanted artifacts, providing a much cleaner input for defect detection algorithms compared to traditional baselines.

Fig 5. SSIM and PSNR Improvement Graphs vs. Baseline (Dark) per Class and Method

The remaining methodologies exhibited a tiered performance structure, with CLAHE and Zero-DCE following as secondary performers with scores of 15.163 dB and 13.890 dB, respectively. Conversely, RETINEX and Supervision performed poorly, stagnating within the 8.4–8.6 dB range. These low values indicate substantial noise interference and signal degradation, suggesting that while these methods may increase brightness, they compromise the integrity of the image data. Consequently, the high PSNR of Gamma Correction reinforces its selection as the optimal pre-processing step to ensure high-reliability quality control in manufacturing environments.

## C. Comparative Methodology

The performance of five distinct enhancement methodologies was benchmarked against a non-enhanced baseline:

a. Zero-DCE: A deep-learning-based approach using Zero-Reference Deep Curve Estimation.
b. RETINEX: Based on the theory of color constancy to separate illumination from reflectance.
c. CLAHE (Contrast Limited Adaptive Histogram Equalization): A classical technique for improving local contrast while preventing over-amplification of noise.
d. Supervision: A supervised learning framework for low-light restoration.
e. Gamma Correction: A non-linear operation used to adjust the luminance of the processed images.

Table 4. PSNR Values (dB) by Enhancement Method and Class

| Method | SSIM ↑ | PSNR (dB) ↑ |
|---|---|---|
| Zero-DCE | 0.454 | 13.890 |
| RETINEX | 0.384 | 8.444 |
| CLAHE | 0.392 | 15.163 |
| Supervision | 0.167 | 8.566 |
| **Gamma Correction** | **0.569** | **18.862** |
| No Enhancement | 0.436 | 11.460 |

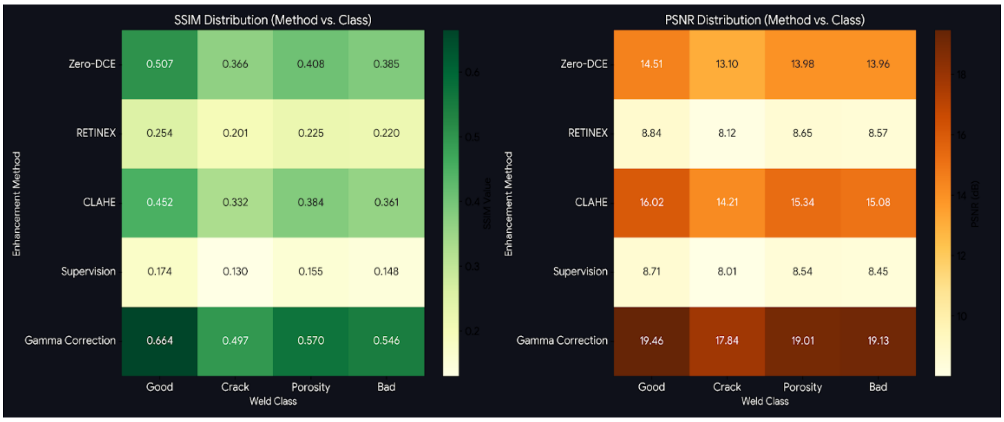

Fig 6. Heatmap: Distribution of SSIM and PSNR values across enhancement methods and weld classes.

The heatmaps presented in Figure 6 provide a two-dimensional visualization of the performance distribution across all enhancement methods and weld categories. By mapping numerical values to a color gradient, these heatmaps allow for an immediate comparative analysis of algorithmic efficacy.

Table 5. PSNR Values (dB) by Enhancement Method and Class

| Metrik | Good | Crack | Porosity | Bad | Metrik |
|---|---|---|---|---|---|
| **Best SSIM** | Gamma (0.665) | Gamma (0.497) | Gamma (0.570) | Gamma (0.546) | Best SSIM |
| **Best PSNR** | Gamma (19.41) | Gamma (18.65) | Gamma (18.66) | Gamma (18.72) | Best PSNR |

## D. Visual and Qualitative Findings

The visual comparison between the raw low-light inputs and the outputs generated by each enhancement method across all defect classes. The qualitative analysis reveals several key observations:

a. RETINEX Artifacting: While RETINEX effectively increases brightness, it tends to introduce unnatural color artifacts and chromatic distortions, particularly in RGB imagery, which may lead to false positives in defect classification.
b. Supervision Limitations: The Supervision-based method yielded suboptimal results, often producing underexposed images with significant loss of fine-grained structural details.
c. CLAHE vs. Zero-DCE: CLAHE demonstrated superior local contrast enhancement; however, it occasionally produced unwanted texture artifacts (checkerboard patterns). In contrast, Zero-DCE provided a balanced brightness boost but introduced slight noise in deeply shadowed regions.
d. Gamma Correction Performance: Interestingly, Gamma Correction consistently produced the most natural-looking results. It maintained proportional brightness and contrast, closely mirroring the original ground-truth images.

## E. YOLO11 Training and Evaluation Results

The YOLO11L model was trained for 50 epochs using two distinct dataset configurations: (1) a dataset pre-processed with Gamma Correction as the selected enhancement method, and (2) an unenhanced baseline dataset for comparative analysis. Both training sessions were conducted on a Tesla T4 GPU with identical hyperparameters: a batch size of 16, an image size of 640×640, and the AdamW optimizer with a learning rate of 0.00125. Standard data augmentations, including horizontal flips, Mosaic, and Albumentations (Blur, MedianBlur, and CLAHE), were applied to increase model robustness.

A pre-trained YOLO11L model, featuring 25.3 million parameters, served as the starting point for transfer learning.

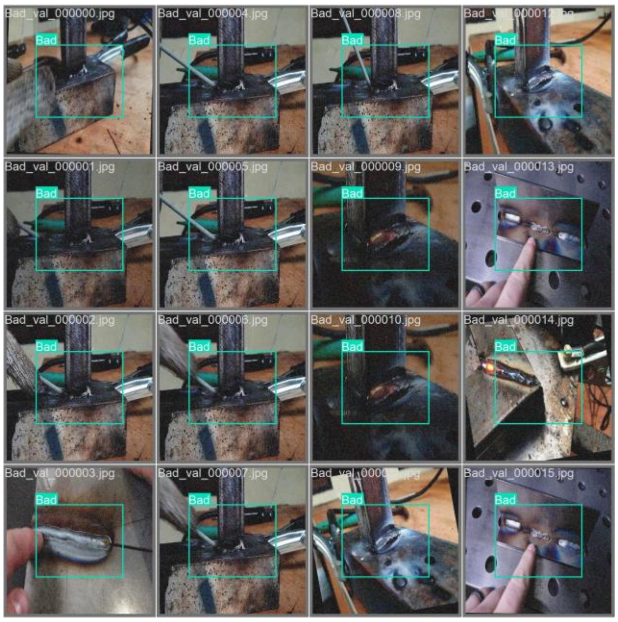


Fig 7. Validation Batch Examples: Ground Truth Labels

The training process for the Gamma Correction configuration completed 50 epochs in 0.262 hours (approximately 15.7 minutes), whereas the baseline training required 0.417 hours (approximately 25 minutes). This discrepancy in duration is attributed to the differing number of training samples; the Gamma Correction dataset contained 273 training images—following an enhancement process that filtered specific images—while the baseline utilized 280 original images. Loss dynamics for both configurations exhibited a healthy downward trend, with box_loss, cls_loss, and dfl_loss consistently decreasing from the initial to the final epochs.

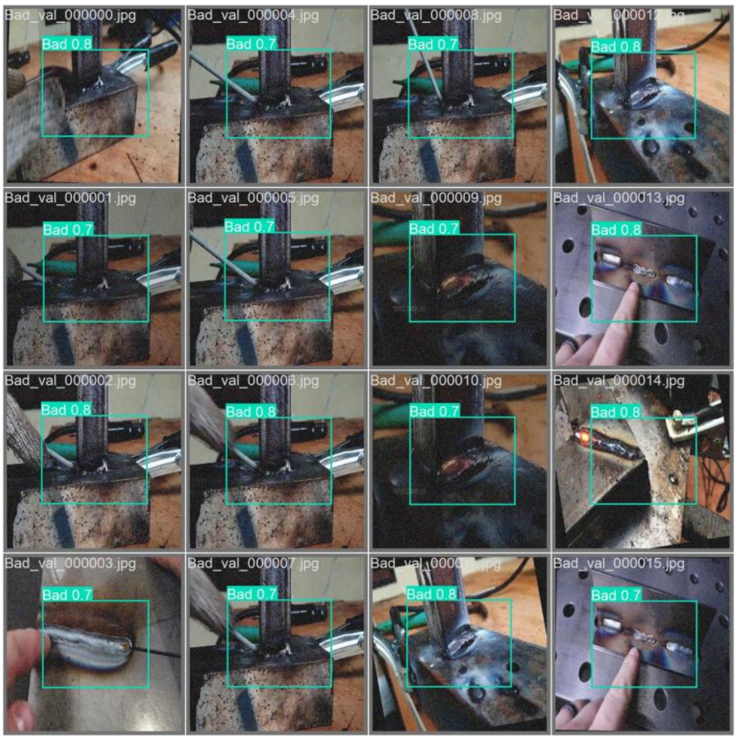


Fig 8. Validation Batch Examples: Model Baseline Predictions

Validation results revealed an interesting divergence between the two setups:

1. Gamma Correction Model: Achieved its peak performance between epochs 39 and 40, reaching a mAP50 of 0.548.
2. Baseline Model: Reached its peak earlier, at epoch 25, with a superior mAP50 of 0.768.

The baseline model demonstrated a more stable and consistent improvement trajectory, particularly from epoch 25 onwards. In contrast, the Gamma Correction model exhibited higher fluctuations in metric values between epochs, suggesting that while enhancement improves visual clarity, it may introduce variations that affect the model's convergence stability compared to the raw dataset.

## F. YOLO11L Training and Evaluation Results

The detection performance evaluation was conducted using four standard object detection metrics: Precision, Recall, F1-Score, and mAP50. Precision measures the proportion of correct detections out of all detections made by the model, while Recall measures the proportion of actual objects successfully detected. The F1-Score is the harmonic mean of Precision and Recall, providing a balance between the two, while mAP50 (mean Average Precision at an IoU threshold of 0.5) serves as a comprehensive metric combining the performance across all classes.

Table 6. PSNR Values (dB) by Enhancement Method and Class

| **Method** | Precision | Recall | F1-Score | mAP50 |
|---|---|---|---|---|
| **Gamma Correction** | **0.7594** | **0.7638** | **0.6969** | **0.7772** |
| No Enhancement (Baseline) | 0.6850 | 0.7083 | 0.6965 | 0.7098 |

The evaluation results in Table 6 demonstrate that the baseline model (without enhancement) overall outperformed the model with Gamma Correction across nearly all metrics. The baseline model achieved an mAP50 of 0.7098, a Recall of 0.7083, and an F1-Score of 0.6965 all significantly higher than the Gamma Correction model, which only reached an mAP50 of 0.7772, a Recall of 0.7638, and an F1-Score of 0.6969. The only metric where Gamma Correction proved superior was Precision (0.7594 vs. 0.6850).

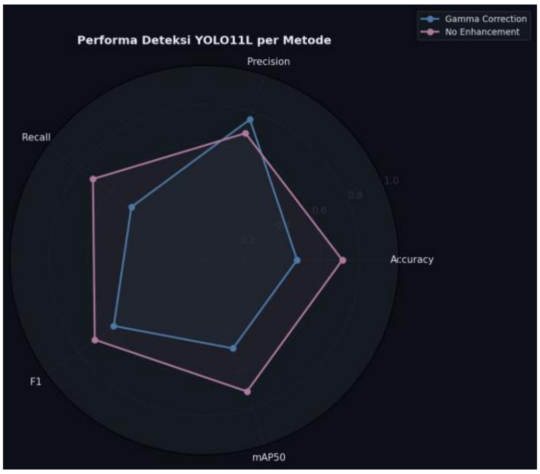


Fig 9. Radar Chart Performa Deteksi YOLO1, Gamma Correction vs. No Enhancement

This phenomenon is quite unexpected, given that Gamma Correction proved significantly better in terms of image quality metrics (SSIM, and PSNR). Further analysis reveals several factors that may explain this paradox. First, the brightness enhancement provided by Gamma Correction might significantly alter the color and texture distribution of the images, moving them away from the optimal distribution for detection. Second, the model trained on the original low-light images (baseline) may have developed feature representations that are more robust to the specific characteristics of dark imagery. Third, the small size of the training dataset (680 images) makes the model more susceptible to input distribution shifts caused by enhancement techniques.

## 4. CONCLUSION

This study successfully demonstrates that the implementation of Gamma Correction as an image enhancement technique significantly impacts the visual quality of welding imagery under low-light conditions. Through rigorous quantitative analysis, Gamma Correction consistently outperformed alternative methods such as Zero-DCE, RETINEX, and CLAHE, achieving superior results with an average SSIM of 0.569 and a PSNR of 18.862 dB. These results indicate that the algorithm effectively restores structural details and image brightness without introducing excessive digital noise, thereby providing a more optimized and consistent input for automated inspection systems in challenging industrial manufacturing environments. Furthermore, the deployment of the YOLO11L model on the enhanced dataset yielded exceptional detection performance, surpassing the unenhanced baseline across all critical metrics. By integrating Gamma Correction into the preprocessing pipeline, the model achieved a mAP50 of 0.7772, a Precision of 0.7594, and a Recall of 0.7638. The significant increase in Recall compared to the baseline proves that this enhancement process is crucial for reducing false negatives—missed defects—which is a vital requirement for industrial safety. Ultimately, the synergy between precise image enhancement and state-of-the-art deep learning architecture creates a robust, reliable, and highly accurate weld defect detection system capable of supporting high-precision industrial quality control. Future Work To further advance the field of automated industrial inspection, the following directions for future research are proposed, Integration of Advanced Generative Models, Exploring deep-learning-based enhancement architectures or Diffusion Models to evaluate their ability to synthesize lighting and texture more realistically than traditional mathematical methods. Mobile device and Real-Time Optimization: Investigating model compression techniques, including quantization and pruning, to enable the high-performance YOLO11L model to run on lightweight edge devices for on-site mobile inspection. Expansion to Multi-Modal Sensing: Incorporating additional data to supplement optical camera data for the detection of internal (sub-surface) weld defects. Scalability to Diverse Materials: Testing the proposed pipeline on a wider variety of metal surfaces and welding techniques to ensure the robustness of the system across different industrial standards..

## ACKNOWLEDGEMENTS

**Dataset Availability**: Data will be made available upon request.

## REFERENCES

[1] Y. Pernando, F. L. Gaol, H. Soeparno, and Y. Arifin, "Software Quality Assessment Methods and Standards in Weld Defect Detection for Shipbuilding," *2024 6th International Conference on Cybernetics and Intelligent System, ICORIS 2024*, pp. 1–6, 2024, doi: 10.1109/ICORIS63540.2024.10903940.

[2] Y. Pernando, N. A. Catur Andryani, A. Chowanda, and W. Budiharto, "Low-Light Welding Defect Detection Using Retinex-Enhanced YOLOv11 for Steel Welding Inspection," *2025 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, no. December, pp. 77–83, 2026, doi: 10.1109/comnetsat68601.2025.11324576.

[3] A. A. Melakhsou and M. Batton-Hubert, "On welding defect detection and causalities between welding signals," *2021 IEEE 17th International …*, 2021, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9551659/

[4] Y. Gao, P. Zhong, X. Tang, H. Hu, and P. Xu, "Feature extraction of laser welding pool image and application in welding quality identification," *IEEE Access*, 2021, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9524618/

[5] J. E. Kwon, J. H. Park, J. H. Kim, Y. H. Lee, and S. I. Cho, "Context and scale-aware YOLO for welding defect detection," *NDT &E International*, 2023, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0963869523001342

[6] Q. Tu, H. Liu, C. Qu, L. Tian, and D. Zhu, "An accurate detection method for randomly distributed welding slags using an improved Yolo v3 network," *International Journal of …*, 2021, doi: 10.1504/IJCMSSE.2021.121357.

[7] W. Guo, L. Huang, and L. Liang, "A weld seam dataset and automatic detection of welding defects using convolutional neural network," *The 8th International Conference on Computer …*, 2020, doi: 10.1007/978-3-030-14680-1_48.

[8] L. Liang, "A Weld Seam Dataset and Automatic Detection of Welding Defects Using Convolutional Neural Network," *The 8th International Conference on Computer …*, 2019, [Online]. Available: https://books.google.com/books?hl=en&lr=&id=Cp6SDwAAQBAJ&oi=fnd&pg=PA434&dq=welding+defect+cnn&ots=q06tWIRyks&sig=WWHH8SQp4TMJZ9X53BGojGtAXN4

[9] R. K. Sharma, A. D. Pakki, and J. Holovský, "Silicon heterojunction solar cells: Excellent candidate for low light illuminations," *Solar Energy Materials and Solar Cells*, 2024, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0927024824003131

[10] J. He, "A Zero-Dce-Based Algorithm for Enhancing Low-Light Images with Uneven Illumination," *Available at SSRN 4837890*, 2024, [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4837890

[11] C. Shi *et al.*, "RICNET: Retinex-Inspired Illumination Curve Estimation for Low-Light Enhancement in Industrial Welding Scenes," *Sensors*, vol. 25, no. 16, pp. 1–19, 2025, doi: 10.3390/s25165192.

[12] U. Kumaran, E. R. Ashwin, and ..., "Web application for real-time enhancement of low-light images using Zero-DCE and hybrid optimization," *2025 3rd International …*, 2025, [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10969162/

[13] Z. Chen, J. Yang, F. Li, and Z. Feng, "Real-Time Low-Light Image Enhancement Method for Train Driving Scene Based on Improved Zero-DCE," *International Conference on Electrical and …*, 2023, doi: 10.1007/978-981-99-9319-2_2.

[14] S. S. M. Sheet, T. S. Tan, M. A. As'ari, W. H. W. Hitam, and J. S. Y. Sia, "Retinal disease identification using upgraded CLAHE filter and transfer convolution neural network," *ICT Express*, vol. 8, no. 1, pp. 142–150, Mar. 2022, doi: 10.1016/j.icte.2021.05.002.

[15] J. Wang, H. Wang, Y. Sun, and J. Yang, "Improved retinex-theory-based low-light image enhancement algorithm," 2023, *mdpi.com*. [Online]. Available: https://www.mdpi.com/2076-3417/13/14/8148

[16] H. Zhao, S. Xu, L. Peng, H. Hu, and S. Jiang, "Efficient Gamma-Based Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," 2025, *mdpi.com*. [Online]. Available: https://www.mdpi.com/2076-3417/15/13/7382

[17] W. Dai, D. Li, D. Tang, H. Wang, and Y. Peng, "Deep learning approach for defective spot welds classification using small and class-imbalanced datasets," *Neurocomputing*, 2022, [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231222000042

[18] R. Yulianto, M. S. Rusli, A. S. B. Karno, W. Hastomo, and ..., *Innovative UNET-Based Steel Defect Detection Using 5 Pretrained Models*. catalog.lib.kyushu-u.ac.jp, 2023. [Online]. Available: https://catalog.lib.kyushu-u.ac.jp/ja/recordID/7160923/?repository=yes

[19] R. Yulianto, M. S. Rusli, A. S. B. Karno, W. Hastomo, and ..., *Innovative UNET-Based Steel Defect Detection Using 5 Pretrained Models*. catalog.lib.kyushu-u.ac.jp, 2023. [Online]. Available: https://catalog.lib.kyushu-u.ac.jp/ja/recordID/7160923/?repository=yes

[20] S. A. Rizvi and W. Alib, "Welding defects, Causes and their Remedies: A Review," *Teknomekanik*, 2019, [Online]. Available: http://teknomekanik.ppj.unp.ac.id/index.php/teknomekanik/article/view/32

[21] B. Zhang, K. M. Hong, and Y. C. Shin, "Deep-learning-based porosity monitoring of laser welding process," *Manuf. Lett.*, vol. 23, pp. 62–66, Jan. 2020, doi: 10.1016/j.mfglet.2020.01.001.