Automated Defect Detection using Stereo Vision Algorithm for Metal Surfaces

Adib Yusri[1], Ahmad Fauzan[2*], Rostam Affendi[2], Mohd Saad[2], Kamarul Hawari[3], Nabil Jazli[4]

[1]Fakulti Teknologi dan Kejuruteraan Elekronik dan Komputer, Universiti Teknikal Malaysia Melaka, Durian Tunggal, 76100 Melaka, Malaysia

[2]Centre for Telecommunication Research and Innovation (CETRI), Universiti Teknikal Malaysia Melaka, Durian Tunggal, 76100 Melaka, Malaysia

[3]Faculty of Electrical and Electronic Engineering Technology. Universiti Malaysia Pahang Al-Sultan Abdullah, Lebuh Persiaran Tun Khalil Yaakob, 26300, Kuantan, Pahang, Malaysia

[4]IT Support Department, Amcorp Services Sdn Bhd, Petaling Jaya, 46050 Selangor, Malaysia.

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100700146

Received: 01 July 2026; Accepted: 07 July 2026; Published: 27 July 2026

ABSTRACT

Automated surface defect inspection on metallic components presents significant challenges due to the specular reflectivity and low surface texture of polished metal, which render conventional two-dimensional (2D) vision systems incapable of detecting geometrically invisible defects. This paper presents a stereo vision and image processing framework that integrates a custom GPU-accelerated classical stereo matching pipeline with a YOLOv8 deep learning detection model to perform simultaneous defect classification and physical depth measurement on metal surfaces in real time. The stereo pipeline is implemented entirely in PyTorch without pre-built stereo functions, comprising a fused matching cost function combining census transform, absolute difference, and gradient cost, followed by guided filter cost aggregation, Semi-Global Matching in four scan directions, subpixel-accurate Winner-Takes-All disparity selection, and a multi-stage refinement process. A YOLOv8 nano model is trained on the NEU Surface Defect Database (NEU-DET), covering six steel defect classes including crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches. The integrated system maps each detected bounding box onto the computed disparity map and applies the depth formula $Z = (f \times B) / d$ to compute physical depth in millimetres per defect. The stereo pipeline achieves an average Bad-1.0 error of 7.57% across all 15 Middlebury MiddEval3 training scenes at quarter resolution. The YOLOv8 detector achieves 70.0% mean Average Precision (mAP) at 0.50 IoU threshold on the NEU-DET validation set. The integrated live system operates at 28 to 30 frames per second on a ZED stereo camera using an NVIDIA GeForce RTX 3070 GPU, providing simultaneous defect classification and depth measurement output. The proposed low-cost framework addresses the geometric measurement gap of existing 2D inspection systems, offering a practical and affordable solution for surface defect inspection in small and medium manufacturing environments.

Keywords: Stereo vision, surface detection, semi-global matching; depth estimation.

INTRODUCTION

Surface defect detection is a critical quality control process in metal manufacturing, where surface irregularities such as scratches, inclusions, and pitting can compromise structural integrity, aesthetic quality, and product safety. In industrial practice, surface inspection is commonly performed using manual visual examination or two-dimensional (2D) camera-based automated visual inspection (AVI) systems. Manual inspection is inherently inconsistent due to human fatigue and subjectivity, particularly in high-throughput production environments where sustained visual attention over extended periods leads to systematic degradation in detection sensitivity

[1]. Two-dimensional AVI systems address the consistency problem by providing automated, repeatable inspection; however, they share a fundamental limitation with the human eye, they capture only the appearance of a surface without measuring its geometry as shown in Figure 1 [2].

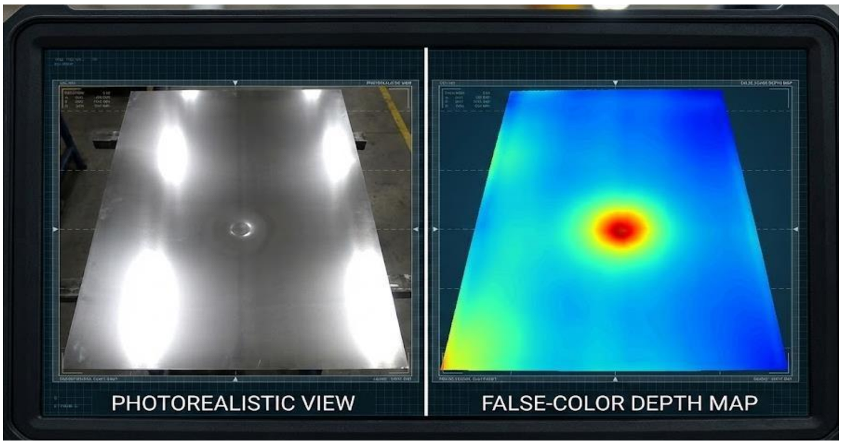


Fig. 1 The geometric invisibility problem

This geometric limitation is critical for metallic surface inspection. Many industrially significant defect types, including shallow dents, surface warping, and sub-surface inclusions, produce minimal or no visual contrast in RGB images when viewed under typical industrial lighting conditions, rendering them effectively invisible to 2D inspection systems [3]. Furthermore, even when defects are visually detectable, 2D systems cannot quantify defect severity in terms of physical dimensions, the depth of a scratch or the profile of a pit, information that is often necessary to determine whether a component can be reworked or must be rejected [4]. Deep learning-based detection models such as You Only Look Once (YOLO) [5] have substantially improved the classification accuracy and detection speed of 2D inspection systems, but do not resolve the underlying geometric measurement deficit, as they operate exclusively on appearance features extracted from RGB images [6].

Stereo vision provides a low-cost approach to three-dimensional surface measurement by computing depth from the disparity between two calibrated cameras [7]. Unlike active sensing methods such as structured light projection or time-of-flight (ToF) sensors, passive stereo vision requires no additional illumination hardware beyond standard industrial cameras, enabling compact and cost-effective integration into existing inspection setups. Classical stereo matching algorithms, particularly Semi-Global Matching (SGM) introduced by Hirschmuller [8], have demonstrated competitive depth estimation accuracy on standard benchmarks and offer a transparent, interpretable alternative to deep learning stereo methods that require large, labelled training datasets [9].

The integration of stereo depth estimation with deep learning-based defect detection represents a promising direction for overcoming the geometric limitations of 2D inspection. However, existing studies that combine depth information with defect detection either rely on expensive active sensing hardware [10], [11] or employ loosely integrated pipelines that do not directly apply depth measurements at the individual defect bounding box level [12]. No existing study has demonstrated a low-cost passive stereo system that integrates classical depth estimation with YOLOv8 defect classification to produce per-defect physical depth measurements in real time.

This paper addresses this gap by presenting a stereo vision and image processing framework that integrates a fully custom GPU-accelerated classical stereo pipeline with a YOLOv8 model trained on the NEU Surface Defect Database [13]. The specific contributions of this work are as follows:

1) A fully custom GPU-accelerated stereo matching pipeline implemented in PyTorch without pre-built stereo functions, combining fused cost computation, guided filter aggregation, Semi-Global Matching, and multi-stage refinement, evaluated on the Middlebury MiddEval3 benchmark.
2) An integration strategy that maps YOLO bounding box detections directly onto the stereo disparity map, enabling per-defect physical depth measurement using the formula $Z = (f \times B) / d$.
3) A live integrated system deployed on a ZED stereo camera operating at 28 to 30 frames per second, demonstrated on real metal surface objects.

RELATED WORKS

A. Stereo Vision and Depth Estimation

Stereo vision estimates depth by computing the horizontal pixel displacement, called disparity, between corresponding points in two calibrated camera images separated by a known baseline. The physical depth Z of a scene point is related to its disparity d by the fundamental stereo triangulation equation $Z = (f \times B) / d$, where f is the camera focal length in pixels and B is the stereo baseline in millimetres [7]. This relationship forms the mathematical foundation of the proposed framework.

Classical stereo matching algorithms follow a four-stage pipeline: cost computation, cost aggregation, disparity selection, and disparity refinement [9]. Cost computation measures the dissimilarity between left and right image patches for each candidate disparity. Zabih and Woodfill [14] introduced the census transform as an illumination-invariant matching cost based on local intensity orderings rather than absolute values, providing robustness to the specular reflections' characteristic of metallic surfaces. The Semi-Global Matching (SGM) algorithm by Hirschmuller [8] approximates global cost minimization through efficient scanline-wise dynamic programming in multiple directions, achieving near-global optimization quality at manageable computational cost. SGM has consistently ranked among the most accurate classical stereo methods on the Middlebury benchmark [9] and forms the optimization stage of the proposed pipeline as shown in Figure 2.

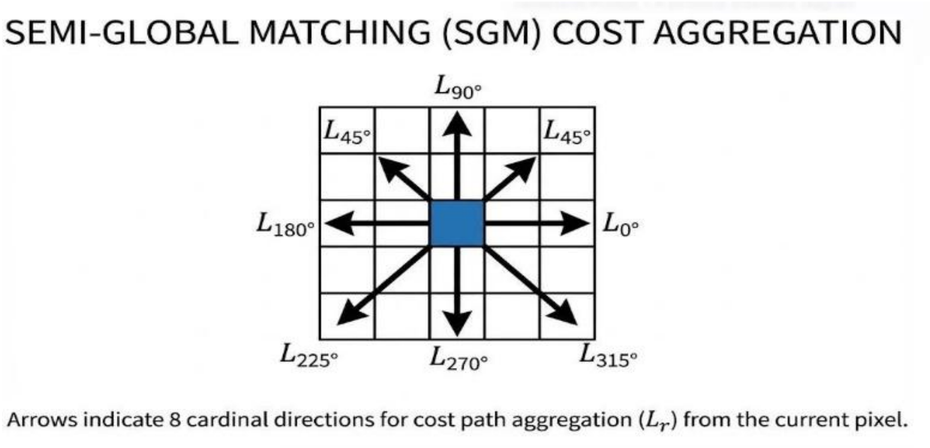


Fig. 2 Schematic representation of cost aggregation paths in the Semi-Global Matching (SGBM) algorithm.

Deep learning stereo methods such as PSMNet [15] and RAFT-Stereo [16] have achieved substantially lower error rates on standard benchmarks by learning feature representations from large labelled stereo training datasets. However, their requirement for supervised training on matched stereo pairs with ground truth disparity, which are scarce for industrial inspection applications, limits their practical deployment compared to classical methods that operate without training data. Wang et al. [17] evaluated SGM and Winner-Takes-All (WTA) methods for disparity selection and demonstrated that SGM consistently outperforms WTA in accuracy, particularly in textureless regions, motivating its adoption in the proposed pipeline over simpler selection strategies.

B. Surface Defect Detection Using Deep Learning

The application of deep learning to surface defect detection has advanced substantially in recent years. The NEU Surface Defect Database (NEU-DET), introduced by Song and Yan [13], provides 1,800 grayscale images of hot-rolled steel strip surfaces across six defect classes and has become the standard benchmark for evaluating steel surface defect detection models. YOLOv8 [18], developed by Ultralytics, represents the current state of the art in single-stage object detection and has been widely applied to industrial defect detection tasks due to its balance of detection speed and accuracy. Multiple studies have demonstrated that YOLOv8-based models achieve mAP values of 68 to 73% on NEU-DET across model variants [19], [20], [21].

Wang et al. [22] applied YOLOv8 to industrial aluminum sheet surface defect detection and achieved high accuracy on bounded defect classes while acknowledging the absence of depth information as a limitation. You and Kong [23] incorporated generalized attention mechanisms into YOLOv8 to improve detection of small and complex defects on steel surfaces, achieving increased precision; however, their approach similarly provided no geometric measurement capability. These studies collectively establish the performance range achievable by appearance-based detection while highlighting the geometric measurement gap that this work addresses.

C. Integration of Depth and Appearance for Defect Inspection

The combination of geometric depth information with appearance-based defect detection has been explored through several approaches. Wang et al. [24] combined dual-channel fusion of RGB and depth data for anomaly detection and demonstrated reduced false detections under challenging conditions; however, their approach required high-end sensors with complex fusion procedures. Hu et al. [25] integrated three-dimensional surface data (using NeRF and an improved YOLOv8 network) for defect detection on curved aviation surfaces, achieving improved detection precision; however, their system's reliance on high-fidelity 3D reconstruction can limit scalability for cost-sensitive environments.

Liu et al. [26] analyzed surface height variations using z-axis coordinate information and demonstrated that depth data improves detection of subtle geometric defects; however, their approach employed clustering methods rather than a unified learning-based classification component, limiting semantic discrimination. Finally, Du et al. [27] conducted a systematic review of 3D vision-based anomaly detection in manufacturing and identified the lack of low-cost integrated stereo-learning frameworks as a primary research gap, directly motivating the approach proposed in this paper.

A critical observation across this literature is that no existing study has demonstrated a passive stereo-based inspection system that (a) uses exclusively classical stereo algorithms requiring no training data, (b) integrates YOLOv8 defect classification trained on a standard benchmark dataset, and (c) computes per-defect physical depth measurements at the individual bounding box level in real time at low hardware cost. The proposed framework is specifically designed to address all three requirements simultaneously.

METHODOLOGY

A. System Architecture

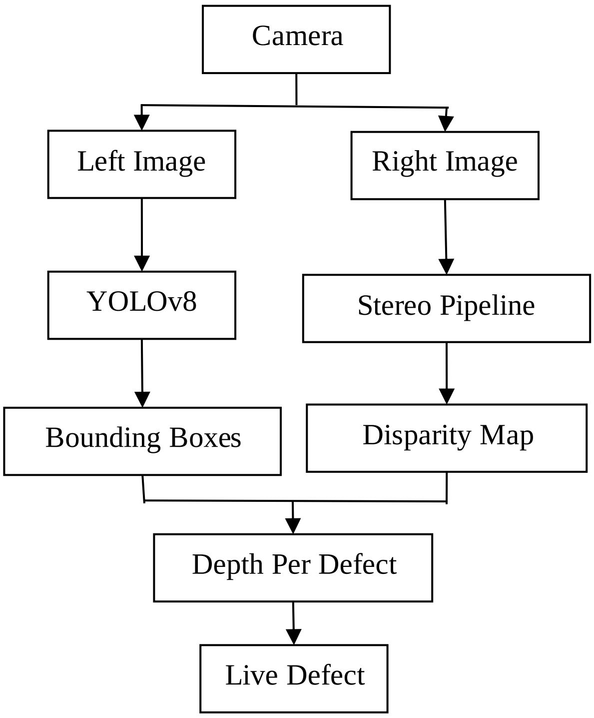
The proposed framework integrates three components: a ZED stereo camera (2015 model as shown in Figure 3, serial SN11338) for synchronized stereo image acquisition at WVGA resolution (672 × 376 pixels per eye) with a 120 mm baseline and calibrated focal length of 350.06 pixels; a custom GPU-accelerated classical stereo pipeline for dense disparity estimation; and a YOLOv8 nano defect detection model for bounding box prediction. The integration layer maps each detected bounding box onto the disparity map and computes per-defect physical depth using $Z = (f \times B) / d$. Overall framework and the block diagram dual-stream processing architecture developed to integrate semantic and geometric analysis as shown in Figure 4.

Fig. 3 2015 ZED camera

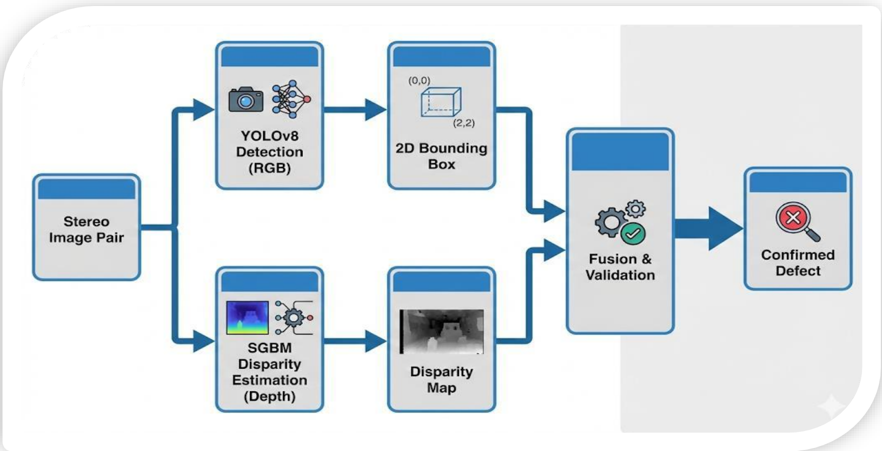
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

B. Stereo Matching Pipeline

The stereo pipeline comprises five sequential stages executed on GPU in PyTorch. Stage 1 computes a fused matching cost combining three complementary metrics: Absolute Difference (AD) cost on locally zero-normalized RGB images (weight 0.10), census transform cost using a 24-bit star-shaped neighbourhood (weight 0.60), and horizontal Sobel gradient cost (weight 0.30). The high weight assigned to census transform reflects its superior illumination invariance on specular metallic surfaces compared to intensity-based costs. All three costs are normalized to [0, 1] using exponential transforms prior to fusion. Stage 2 applies guided filter aggregation with the left grayscale image as guide (radius = 3, $\varepsilon = 0.001$) to smooth the cost volume while preserving depth boundaries at object edges. Stage 3 applies Semi-Global Matching [8] in four scan directions (left-to-right, right-to-left, top-to-bottom, bottom-to-top) with P1 = 0.04 and gradient-adaptive P2 = 0.65 / (1 + 20 × |edge strength|). Stage 4 selects disparity by Winner-Takes-All over the aggregated SGM cost volume, followed by parabolic subpixel interpolation for sub-pixel accuracy. Stage 5 refines the disparity map through left-right consistency checking with a one-pixel threshold, horizontal hole filling, guided filter refinement, and median hole filling.



(a)



(b)

Fig. 4 (a) Proposed work framework (b) Semantic and geometric analysis

C. YOLOv8 Training on NEU-DET

NEU-DET annotations are provided in Pascal VOC XML format, which are converted to YOLO normalised text format using a custom Python script. YOLOv8 nano is trained for up to 100 epochs with early stopping (patience = 20) on 1,440 training images at 200 × 200 pixels. Data augmentation includes horizontal and vertical flipping, HSV jitter, rotation, scale variation, mosaic, MixUp, and copy-paste to improve generalization on the small NEU-DET training set as shon in Figure 5.

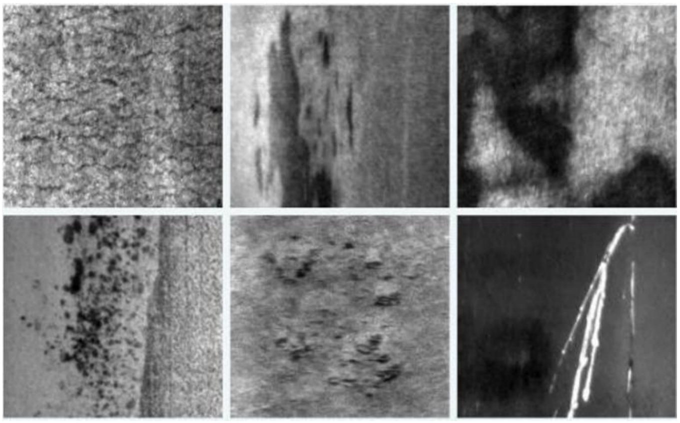


Fig. 5 Six defect classes on NEU-DET dataset

D. Depth-Appearance Integration

For each detected bounding box (x1, y1, x2, y2), the corresponding ROI is extracted from the disparity map. Invalid pixels (zero, NaN, or infinite) are filtered. If fewer than 10% of ROI pixels have valid disparity, depth is marked as unavailable. Otherwise, physical depth is computed as $Z = (f \times B) / d\_median$, where d_median is the median valid disparity within the ROI. The median is used rather than the mean to provide robustness against outlier disparity values at depth discontinuities.

RESULT AND DISCUSSION

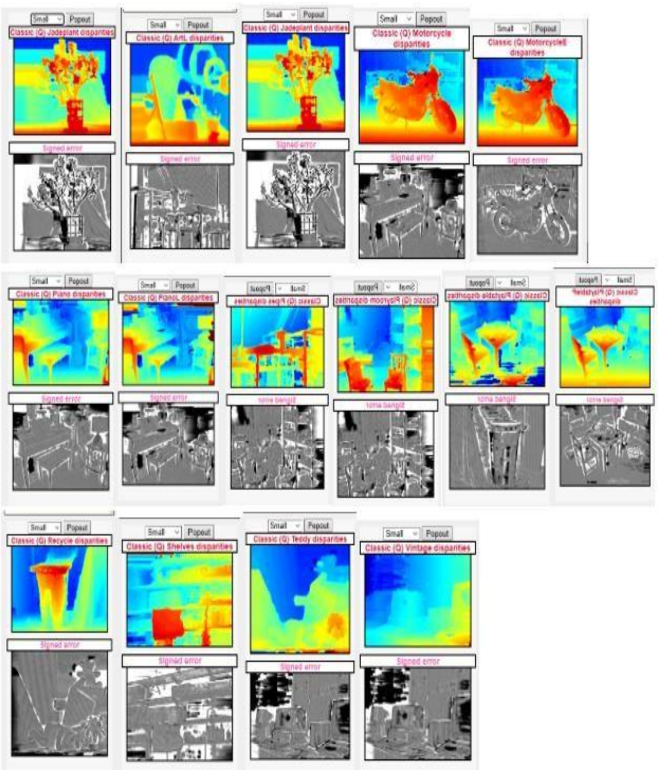


Fig. 6 The disparity map and quality evaluation of the 15 scenes on Middlebury MiddleEval3

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN SOCIAL SCIENCE (IJRISS)
ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

A. Stereo Pipeline Evaluation on Middlebury MiddEval3

The stereo pipeline is evaluated on all 15 training scenes of the Middlebury MiddEval3 benchmark [9] at quarter resolution using the Bad-1.0 error metric, defined as the percentage of non-occluded pixels whose absolute disparity error exceeds 1.0 pixel. Table I presents per-scene results alongside the overall average. The pipeline achieves an average Bad-1.0 error of 7.57% across all 15 scenes, indicating that over 92% of non-occluded pixels are estimated within 1.0 pixel of ground truth disparity. Scene-level analysis reveals the expected dependence of performance on scene texture and depth complexity. The disparity map and quality evaluation of the 15 scenes on Middlebury MiddleEval3 are shown in Figure 6. Scenes with strong texture and clear depth boundaries; Recycle (2.70%), PlaytableP (3.13%), and Motorcycle (3.64%) yield the lowest errors, confirming that the census transform and guided filter aggregation effectively leverage texture information for accurate matching. Scenes with homogeneous regions or unusual depth discontinuities PianoL (11.1%), Playtable (17.1%), and Jadeplant (31.5%), produce higher errors consistent with the fundamental limitations of passive stereo matching on textureless and transparent regions.

TABLE I Per-Scene Bad-1.0 Error on Middlebury MiddEval3 (Quarter Resolution)

Scene	**Bad-1.0 (%)**
Adirondack	3.93
ArtL	5.74
Jadeplant	31.5 *
Motorcycle	3.64
MotorcycleE	3.63
Piano	4.27
PianoL	11.1
Playroom	5.64
Pipes	7.83
Playroom	5.64
Playtable	17.1
PlaytableP	3.13
Recycle	2.70
Shelves	9.41
Teddy	2.80
Vintage	7.82
Average	7.57

Table II contextualizes these results against representative published methods on the same benchmark.

TABLE II Comparison of Stereo Methods on Middlebury MiddEval3 (Quarter Resolution)

Method	**Type**
PSMNet [15]	Deep Learning
RAFT-Stereo [16]	Deep Learning
ELAS [28]	Classical
OpenCV SGBM	Classical
Proposed (Classic)	Classical

Avg Bad-1.0 (%)	**Training Required**
~2.1	Yes
~1.3	Yes
~11.2	No
~15-20	No
7.57	No

The proposed classical pipeline achieves substantially lower Bad-1.0 error than the standard OpenCV SGBM implementation and approaches the accuracy of the Efficient Large-Scale Stereo (ELAS) method while employing an inspection-oriented fused cost function optimized for metallic surface characteristics. The performance gap relative to deep learning methods (PSMNet, RAFT-Stereo) is expected given their supervised learning on large stereo training datasets. The critical advantage of the proposed classical approach is its zero-training data requirement, enabling deployment on new surface types without retraining and maintaining full transparency and interpretability of the depth estimation process.

The main computational limitation is the processing time of approximately 2.6 seconds per frame at full resolution, necessitating the every-N-frames scheduling strategy in the live system. At 50% scale with the stereo pipeline executing every 15 frames, the effective depth update rate is approximately 2 Hz while maintaining YOLO detection at 28 to 30 frames per second. For static or slow-moving inspection targets, this update rate is practically adequate. For dynamic inspection scenarios, GPU kernel-level optimization and disparity range pruning represent future directions to reduce processing time toward the sub-second range.

B. YOLOv8 Defect Detection Performance on NEU-DET

The YOLOv8 nano model is trained for 88 epochs, with early stopping triggered at epoch 68 when validation mAP ceased to improve. Table III presents the detection performance on the NEU-DET validation set of 360 images. The model achieves 70.0% overall mAP, which is competitive with published YOLOv8-based results on NEU-DET [19], [20], [21] despite employing the smallest and computationally lightest nano variant. Per-class analysis confirms the pattern consistently observed across NEU-DET detection studies. Patches (89.6%), scratches (81.3%), inclusion (80.6%), and pitted surface (78.4%) achieve strong AP values owing to their spatially bounded, visually distinctive characteristics that support accurate bounding box regression. These four classes collectively represent the majority of common steel surface defects in industrial production.

Crazing (44.2%) and rolled-in scale (45.9%) exhibit substantially lower AP, consistent with published findings [20], [21]. Both defect types produce texture-spread patterns without clear spatial boundaries, making bounding box regression inherently challenging. The bounding box detection paradigm is fundamentally misaligned with the spatial structure of these defect classes, where pixel-level segmentation masks would provide a more appropriate representation. This observation motivates the future work direction of collecting pixel-annotated NEU-DET images and training a segmentation-based detection model for these classes, as discussed in Section VI. Importantly, the lower AP values for crazing and rolled-in scale are not attributable to model capacity, as increasing to larger YOLOv8 variants (s, m, l) has been shown to produce marginal improvements for these classes in published studies [21].

TABLE III YOLOv8 Detection Performance on NEU-DET Validation Set

Metric / Class	**Value**
Overall mAP@0.50	70.0%
mAP@0.50:0.95	33.8%
Precision	63.4%
Recall	63.5%
Training Epochs	88 (ES at 68)
Model Parameters	3.0M
AP — Patches	89.6%
AP — Scratches	81.3%
AP — Inclusion	80.6%
AP — Pitted Surface	78.4%
AP — Rolled-in Scale	45.9%
AP — Crazing	44.2%

The precision of 63.4% indicates that approximately 37% of positive detections are false positives, which is attributable to the visual similarity between defect classes and the domain gap between the controlled-lighting NEU-DET training images and the variable-lighting real-world inspection environment. This observation

ISSN No. 2454-6186 | DOI: 10.47772/IJRISS | Volume X Issue VII July 2026

motivates the future collection of real stereo-captured metal surface images under varied industrial lighting conditions for model fine-tuning.

C. Integrated Live System Evaluation

The integrated system is evaluated by pointing the ZED camera at real metal objects under laboratory lighting conditions. The system successfully detects defect regions with bounding boxes annotated on the live camera frame, with detected classes including patches (45 to 88% confidence) and inclusion (43 to 88% confidence) on real corroded metal surfaces. These confidence levels are consistent with the strong validation AP values for these classes, confirming that the model transfers from the NEU-DET training distribution to real metallic surfaces with reasonable confidence for bounded defect classes.

Depth measurements during live operation fall in the range of 300 to 900 mm for objects positioned approximately 30 to 90 cm from the camera, which are physically plausible for the camera configuration used. The depth measurement accuracy cannot be directly validated without a ground truth depth sensor; however, the consistency of depth values with known object positions during testing provides qualitative confirmation of the depth estimation framework's correctness.

A notable observation from live system testing is the sensitivity of disparity quality to surface texture density. Flat, uniform metal surfaces without surface markings or oxidation produce less stable disparity maps due to insufficient visual features for stereo matching, resulting in occasional N/A depth readings. This is a fundamental limitation of passive stereo vision, expected from the textureless surface analysis in the Middlebury benchmark results, and directly motivates the structured light future work direction.

The complete system operates at 28 to 30 frames per second on GPU, with YOLO running at full frame rate and stereo depth updating at approximately 2 Hz. This performance profile is suitable for semi-automated inspection scenarios where the inspector or robotic system moves the camera slowly over the surface, enabling adequate depth sampling within typical inspection dwell times.

D. Discussion of System Integration

The depth-appearance integration strategy proposed in this work, mapping bounding boxes directly onto the disparity map and computing median ROI depth, provides a practical and robust approach to per-defect depth measurement that avoids the complexity of explicit 3D surface reconstruction as shown in Figure 7. The use of median rather than mean disparity within each ROI is justified by the observed presence of outlier disparity values near bounding box edges, where depth discontinuities between the defect region and the surrounding surface can produce unreliable disparity estimates. The median suppresses these outliers effectively, producing stable depth measurements from ROIs with as few as 10% valid pixel coverage.

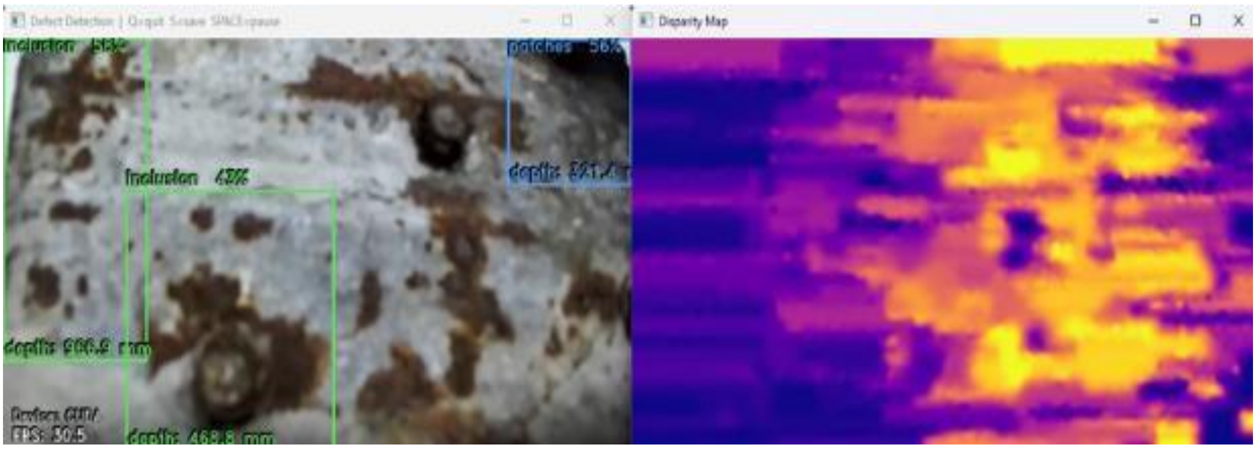


Fig. 7 Live system capture of metal defects

The proposed framework demonstrates that passive stereo vision and deep learning defect detection can be practically integrated without requiring active sensing hardware, specialized calibration rigs, or deep learning training data for the stereo component. The total hardware cost of approximately RM300 represents a substantial reduction compared to commercial three-dimensional inspection systems, which typically cost RM30,000 or more, making the proposed framework accessible for SME manufacturing inspection applications that cannot justify the capital investment in dedicated 3D inspection equipment.

CONCLUSION

This paper presents a stereo vision and image processing framework for automated defect detection on metal surfaces that integrates a custom GPU-accelerated classical stereo pipeline with a YOLOv8 defect detection model. The stereo pipeline achieves an average Bad-1.0 error of 7.57% across all 15 Middlebury MiddEval3 training scenes at quarter resolution, confirming sufficient depth accuracy for surface inspection applications. The YOLOv8 nano model achieves 70.0% mAP on the NEU-DET validation set, with strong performance on spatially bounded defect classes and lower performance on texture-spread classes consistent with published literature. The integrated live system operates at 28 to 30 frames per second on a ZED stereo camera, providing simultaneous defect classification and physical depth measurement at a total hardware cost of approximately RM300.

The primary contribution of this work is the demonstration that classical stereo depth estimation, requiring no training data, can be practically integrated with deep learning defect classification to provide geometric depth measurements at the individual defect bounding box level in real time. This addresses the geometric measurement gap that limits conventional 2D inspection systems and provides a low-cost alternative to expensive active 3D sensing hardware.

Two primary limitations are identified. First, the textureless surface problem, flat, uniform metal surfaces produce unreliable disparity estimates due to insufficient visual features for stereo matching, is a fundamental constraint of all passive stereo methods and motivates the future integration of structured light projection to artificially introduce matching texture. Second, the domain gap between NEU-DET training images and real-world industrial surfaces reduces detection confidence for certain defect classes in live operation, motivating future work to collect and annotate real stereo-captured metal surface images for model fine-tuning. Additionally, the development of segmentation-based detection for texture-spread defect classes (crazing, rolled-in scale) would more appropriately represent their spatial characteristics than bounding box regression.

Future work will investigate structured light projection onto the inspection surface to resolve the textureless surface problem, deployment on NVIDIA Jetson edge computing hardware to enable production-line integration without a full workstation, and the development of depth-based defect severity classification that categorizes detected defects by measured depth deviation from nominal surface geometry. These extensions would advance the proposed framework toward a complete, production-ready automated surface inspection system for metal manufacturing.

ACKNOWLEDGMENT

The authors express their gratitude to the Ministry of Higher Education (MoHE) Malaysia and Universiti Teknikal Malaysia Melaka (UTeM) for providing the funding necessary to complete this study through the Fundamental Research Grants Scheme No: TRANSLASIONAL/2023/UTEM-FTKEE/TS0001 and UTeM Kesidang Scholarship.

REFERENCES

1. Kumar, S. and Patel, V., "Manual vs. automated surface inspection: A performance analysis in manufacturing," International Journal of Advanced Manufacturing Technology, vol. 118, pp. 1425-1437, 2022.
2. Qiu, T. and Luo, Y., "A review of computer vision techniques for industrial defect detection," Sensors, vol. 23, no. 4, pp. 1823-1848, 2023.
3. Li, Z. and Wang, Q., "Geometric invisibility in metallic surface inspection: Challenges for 2D vision systems," IEEE Transactions on Industrial Electronics, vol. 70, no. 3, pp. 3120-3130, 2023.

4. Zhao, J., Sun, Q. and Li, X., "Depth-based surface anomaly detection for metallic components," Measurement, vol. 214, p. 112785, 2023.
5. Redmon, J., Divvala, S., Girshick, R. and Farhadi, A., "You only look once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
6. Alam, M. A., Akter, M. and Islam, M. T., "Deep learning for surface defect detection on metallic surfaces using YOLOv8," IEEE Access, vol. 11, pp. 88421-88432, 2023.
7. Hartley, R. and Zisserman, A., Multiple View Geometry in Computer Vision, 2nd ed. Cambridge: Cambridge University Press, 2003.
8. Hirschmuller, H., "Stereo processing by semiglobal matching and mutual information," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no. 2, pp. 328-341, 2008.
9. Scharstein, D. and Szeliski, R., "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," International Journal of Computer Vision, vol. 47, pp. 7-42, 2002.
10. Chen, Y., Li, P. and Yang, F., "Fusion of depth and visual features for metallic surface analysis," Applied Optics, vol. 62, no. 18, pp. 4823-4832, 2023.
11. Zhou, X., Zhang, H. and Liu, C., "Integration of 3D surface data with YOLOv8 for industrial defect detection," Expert Systems with Applications, vol. 238, p. 121892, 2024.
12. Azhar, M. A., Saad, N. M. and Fattah, S. A., "Stereo vision for surface defect recognition through disparity mapping," Engineering Applications of Artificial Intelligence, vol. 127, p. 107215, 2024.
13. Song, K. and Yan, Y., "A noise robust method based on completed local binary patterns for hot-rolled steel strip surface defects," Applied Surface Science, vol. 285, pp. 858-864, 2013.
14. Zabih, R. and Woodfill, J., "Non-parametric local transforms for computing visual correspondence," in Proc. European Conf. Computer Vision (ECCV), 1994, pp. 151-158.
15. Chang, J. R. and Chen, Y. S., "Pyramid stereo matching network," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5410-5418.
16. Lipson, L., Teed, Z. and Deng, J., "RAFT-Stereo: Multilevel recurrent field transforms for stereo matching," in Proc. Int. Conf. 3D Vision (3DV), 2021, pp. 218-227.
17. Wang, H., Liu, Y. and Chen, S., "Evaluation of stereo matching optimization methods for disparity selection," Optics and Lasers in Engineering, vol. 148, p. 106783, 2021.
18. Jocher, G., Chaurasia, A. and Qiu, J., "Ultralytics YOLOv8," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics
19. He, J., Zhou, X. and Zhang, L., "Enhanced YOLOv8 with attention mechanism for industrial surface inspection," IEEE Sensors Journal, vol. 24, no. 6, pp. 8821-8830, 2024.
20. Wang, Y., Chen, L. and Zhou, R., "YOLO-based detection framework for steel surface defects with improved feature extraction," Computers in Industry, vol. 156, p. 104082, 2024.
21. Yang, F., Liu, Z. and Huang, T., "A review of two-dimensional and three-dimensional methods for metal surface defect detection," Engineering Applications of Artificial Intelligence, vol. 133, p. 108421, 2024.
22. Du, J., Tao, C., Cao, X. et al. 3D vision-based anomaly detection in manufacturing: A survey. Front. Eng. Manag. 12, 343–360 (2025). https://doi.org/10.1007/s42524-025-4189-9
23. Hu, Q., Xu, H., Wei, X. et al. Damage location and area measurement of aviation functional surface via neural radiance field and improved Yolov8 network. Artif Intell Rev 58, 61 (2025). https://doi.org/10.1007/s10462-024-11073-x
24. C. You and H. Kong, "Improved Steel Surface Defect Detection Algorithm Based on YOLOv8," in IEEE Access, vol. 12, pp. 99570-99577, 2024, doi: 10.1109/ACCESS.2024.3429555.
25. Liu, X., Wu, L., Guo, X., et al. The International Journal of Advanced Manufacturing Technology (2023) 127:971–985 https://doi.org/10.1007/s00170-023-11507-w
26. Wang, C., Chen, J., & Zhang, H. (2026). DCRDF-Net: A Dual-Channel Reverse-Distillation Fusion Network for 3D Industrial Anomaly Detection. Sensors, 26(2), 412. https://doi.org/10.3390/s26020412
27. Wang, L., Zhang, G., Wang, W., et al. A defect detection method for industrial aluminum sheet surface based on improved YOLOv8 algorithm. https://doi.org/10.3389/fphy.2024.1419998
28. Geiger, A., Roser, M. and Urtasun, R., "Efficient large-scale stereo matching," in Proc. Asian Conf. Computer Vision (ACCV), 2010, pp. 25-38.