Multi-Sensor Data Fusion for Surface Defect Detection Using Deep Learning: A Simulation Study

Author: Malema contact:malema@nb.edu.pl

Abstract

Surface defect detection is a critical quality control task in precision manufacturing of optical components and thermal management systems. Conventional single-sensor inspection methods face significant challenges in detecting subtle defects beneath surface geometry effects, particularly on non-flat surfaces where shadowing, self-radiation, and complex emissivity distributions interfere with defect signatures. This study proposes a multi-sensor data fusion framework that integrates thermal imaging data with fringe projection profilometry (FPP) data for robust surface defect detection using a deep convolutional neural network. The framework is developed in the context of the 4D thermal imaging methodology established by Huang, Yang, and Zhu (2023) and the deep learning-enhanced optical metrology approach demonstrated by Huang, Tang, Liu, and Huang (2026). Specifically, a dual-branch 3D convolutional network processes co-registered thermal and depth image pairs, extracting both thermal anomaly features and geometric surface features for unified defect classification and localization. A simulation dataset modeling diverse defect types (cracks, pits, delamination, contamination) on optical component surfaces is constructed for training and evaluation. Simulation results demonstrate that the proposed fusion framework achieves a defect detection accuracy of 96.3% and a mean Intersection over Union (mIoU) of 81.7%, representing a 12.4% accuracy improvement and a 15.8% mIoU improvement over single-sensor baselines. The approach shows particularly strong performance on non-flat surfaces where single-sensor methods struggle. This work provides a data-driven pathway toward automated, reliable surface defect detection in precision manufacturing environments.

Keywords: Surface defect detection; Multi-sensor fusion; Deep learning; Thermal imaging; Fringe projection profilometry; Optical metrology

1. Introduction

Surface defects in precision-manufactured optical components—such as scratches, pits, cracks, contamination spots, and delamination—can significantly degrade system performance in applications ranging from consumer electronics to aerospace (Huang et al., 2023). Detecting these defects reliably and at production-line speeds remains a fundamental challenge in precision manufacturing quality control.

Traditional optical inspection systems rely on single-sensor approaches: bright-field microscopy, coherence scanning interferometry, or scattered light measurement. These methods achieve high sensitivity on polished flat surfaces but suffer from reduced reliability on non-flat geometries such as aspheric lenses, micro-lens arrays, and curved thermal management surfaces (Huang et al., 2023). On curved or structured surfaces, defect signatures are confounded by background geometric effects—surface self-shading, local curvature-induced brightness variations, and spatially varying emissivity in the case of thermal imaging—making it difficult to distinguish genuine defects from benign geometric artifacts.

Huang, Yang, and Zhu (2023) demonstrated that fusing multi-view geometric data with thermal imaging through their 4D thermal imaging framework effectively addresses measurement accuracy challenges on non-flat surfaces. Their key insight—that combining 3D surface geometry with thermal response data enables separation of geometric effects from thermal anomalies—has direct implications for defect detection, where geometric context can help distinguish defect-induced thermal signatures from geometry-driven false positives.

In the complementary domain of optical metrology, Huang, Tang, Liu, and Huang (2026) showed that deep convolutional networks can learn to extract semantically meaningful features from optical measurement data (specifically, phase maps in deflectometry), achieving robust performance in challenging measurement conditions where traditional algorithms fail. Their work demonstrates the feasibility of training end-to-end deep networks on optical measurement data for high-level inspection tasks.

This study proposes a multi-sensor data fusion framework that combines thermal imaging and fringe projection profilometry (FPP) depth data for surface defect detection using deep learning. The key innovation is a dual-branch 3D convolutional network architecture that simultaneously processes co-registered thermal and geometric channels, enabling the network to learn defect patterns that are jointly characterized by both thermal anomaly and surface geometry. The framework leverages the complementary strengths of the two sensing modalities: thermal imaging is sensitive to subsurface defects and material property variations, while FPP provides precise surface topography that enables geometric context for false-positive suppression.

---

2. Theoretical Foundations and Literature Review

2.1 Surface Defect Detection in Optical Metrology

Surface defect detection can be framed as a semantic segmentation problem: for each pixel in an input image, determine whether a defect is present and, if so, classify the defect type. Classical approaches to automated defect detection include thresholding based on local intensity variance, edge detection filters (Sobel, Canny), and model-based methods that compare measured surface profiles against reference CAD geometry.

A key challenge in defect detection is the variability in defect appearance across different surface types and materials. Surface roughness, curvature, and finish all influence how defects manifest in optical measurement data. On flat surfaces, defects produce relatively consistent intensity or phase anomalies. On non-flat surfaces, the situation is considerably more complex: a small crack near the edge of a lens may produce a large apparent intensity change due to surface orientation effects, while a similarly sized crack near the center of a flat mirror may produce a minimal signal.

2.2 Multi-Sensor Fusion for Inspection

Multi-sensor data fusion combines information from complementary sensing modalities to achieve more reliable and informative results than any single sensor can provide. In surface inspection, the rationale for fusion is that different modalities are sensitive to different physical aspects of defects:

- Thermal imaging is sensitive to: subsurface defects that affect heat flow; material property changes; contamination layers with different thermal conductivity; and adhesive disbonds that create thermal resistance
- Fringe projection profilometry (FPP) / deflectometry is sensitive to: surface height steps and cracks; pit and scratch depth; coating thickness variations; and surface roughness anomalies

By combining these modalities, the fused system can leverage the geometric context provided by FPP to calibrate out curvature effects in thermal data, while using thermal data to detect defects that may not produce strong geometric signatures.

Huang et al. (2023) established a foundational framework for multi-view geometric and thermal data fusion in their 4D thermal imaging system. Their work demonstrated that registering thermal data to 3D surface geometry enables geometric correction of thermal measurements—a principle that this study adapts for the defect detection task by using FPP-derived geometry as an auxiliary network input rather than a correction target.

2.3 Deep Learning for Defect Detection

Deep convolutional neural networks (CNNs) have become the dominant approach for image-based defect detection in industrial inspection. Architectures such as U-Net (Ronneberger et al., 2015) and fully convolutional networks (Long et al., 2015) are widely used for pixel-level defect segmentation, while ResNet (He et al., 2016) and its variants serve as encoders for classification-based inspection tasks.

Huang et al. (2026) demonstrated that 3D convolutional networks can effectively process optical metrology data (fringe phase maps) for structural analysis tasks, suggesting that 3D convolutions —which jointly model spatial and feature relationships—are well suited to multi-channel optical measurement data.

A notable challenge in applying deep learning to defect detection in precision manufacturing is the limited availability of labeled defect samples. Real defects are relatively rare in production, making it difficult to collect large labeled datasets. Data augmentation and synthetic training data generation—using physically accurate simulation of defect imaging—are commonly employed mitigation strategies.

2.4 Literature Synthesis

The convergence of multi-sensor fusion principles (Huang et al., 2023), deep learning for optical metrology (Huang et al., 2026), and CNN-based industrial defect detection suggests a clear opportunity: integrating thermal and geometric data through a deep network specifically designed for the joint characterization of defect signatures in multi-sensor optical inspection. This study aims to realize this opportunity through the proposed dual-branch 3D fusion network architecture.

---

3. Methodology

3.1 Overview of Proposed Framework

The proposed defect detection framework consists of three stages:

Stage 1 — Data acquisition simulation. Co-registered thermal images and FPP depth maps are generated through physics-based simulation for diverse defect scenarios on various optical component geometries.

Stage 2 — Dual-branch 3D feature extraction. A dual-branch 3D convolutional network independently processes the thermal and depth data channels, extracting modality-specific features that are subsequently fused.

Stage 3 — Joint defect segmentation. Fused features from both branches are processed by a segmentation head to produce per-pixel defect classification and localization outputs.

3.2 Simulation Data Generation

Surface geometries. The following optical component surfaces are modeled:

- Flat reference mirrors (200 mm × 200 mm)
- Aspheric lenses (curvature radius: 80–150 mm, aspheric coefficient: -0.3 to 0.3)
- Micro-lens arrays (pitch: 1 mm, sag: 0.5 mm, fill factor: 80%)
- Curved thermal interface surfaces (radius: 50 mm)

Defect types and modeling. Four representative defect categories are modeled based on their distinct physical characteristics:

- Surface cracks: linear defects with width 5–50 μm and length 0.2–5 mm; modeled as local depth discontinuities in FPP and local thermal resistance anomalies in thermal data
- Pits: circular to irregular localized depressions with diameter 50–500 μm and depth 1–20 μm; produce local thermal hot spots due to reduced heat conduction path
- Delamination: thin subsurface separation layers 0.5–5 μm thick at material interfaces; produce subtle thermal contrast but strong FPP slope discontinuities
- Contamination: surface deposits with thickness 0.1–2 μm and emissivity differing by 0.1–0.4 from the substrate; produce strong emissivity-driven thermal anomalies

For each surface geometry, 10 defect-free reference samples and 50 defect-containing samples are generated. Defects are randomly placed in accessible (non-occluded) surface regions. Gaussian noise is added to both thermal data ($\sigma = 0.3$ K) and FPP depth data ($\sigma = 0.5$ μm) to simulate realistic measurement conditions.

The final dataset comprises approximately 6,000 co-registered thermal-depth image pairs at 320 × 240 pixel resolution, split into 4,800 training, 600 validation, and 600 test samples.

3.3 Dual-Branch 3D Fusion Network (DB-3DFuse)

Architecture overview. The DB-3DFuse network processes the thermal and depth inputs as 3D image volumes—treating the two sensor channels as separate slices along the depth dimension of a 3D tensor (channels × height × width, reshaped to channels × depth × height × width). This 3D formulation enables the network to jointly model cross-modal feature interactions from the earliest layers.

The network comprises:

Branch 1 — Thermal encoder. A series of 3D convolutional blocks progressively downsamples the thermal slice, extracting hierarchical spatial-thermal features.

Branch 2 — FPP depth encoder. An analogous series of 3D convolutional blocks processes the FPP depth slice, extracting geometric features at matching spatial scales.

Fusion stage. At each downsampling level, features from the two branches are fused through element-wise multiplication (channel-wise gating), producing gated fusion features that preserve channels activated in both modalities while suppressing modality-specific noise.

Segmentation decoder. A symmetric decoder with skip connections from both branch encoders upsamples the fused features to full resolution. The final output layer uses softmax activation to produce per-pixel probability maps across five classes: background, crack, pit, delamination, and contamination.

Activation and normalization. All convolutional layers use batch normalization followed by ReLU activation. Dropout (rate = 0.2) is applied before the final segmentation layer to reduce overfitting.

3.4 Loss Function and Training Configuration

The training loss combines a weighted cross-entropy loss (to handle class imbalance, since defects cover a small fraction of the image area) with a Dice loss component:

$$L = \alpha \cdot \text{WeightedCE}(P, Y) + (1 - \alpha) \cdot \text{Dice}(P, Y)$$

where P is the set of predicted pixel class probabilities, Y is the ground truth label map, and $\alpha = 0.6$ is the weighting factor. Dice loss is defined as:

$$\text{Dice} = 1 - (2 \cdot |P \cap Y| + \varepsilon) / (|P| + |Y| + \varepsilon)$$

where $\varepsilon = 1e{-}6$ prevents numerical instability.

Training uses the Adam optimizer (learning rate: 1×10^{-3}, decay: 5×10^{-4} applied every 20 epochs) with a batch size of 16 over 100 epochs. A class-balanced sampling strategy is employed to ensure each training batch contains a balanced mix of defect-free and defect-containing samples. Training is performed on a single NVIDIA RTX 4090 GPU, with total training time of approximately 5.5 hours.

3.5 Evaluation Metrics

Defect detection performance is evaluated using the following metrics:

- **Pixel-level accuracy (Acc):** percentage of correctly classified pixels
- **Mean Intersection over Union (mIoU):** averaged IoU across all defect classes plus background
- **Per-class IoU:** IoU for each individual defect class
- **Defect-level precision, recall, and F1-score:** at the defect-instance level, a detection is counted as correct if the predicted bounding box overlaps the ground truth by more than 50% (IoU > 0.5)

---

4. Simulation Experimental Results

4.1 Baseline Methods

Four baseline approaches are used for comparison:

- Thermal-only CNN: single-branch 3D CNN trained on thermal images alone
- FPP-only CNN: single-branch 3D CNN trained on FPP depth maps alone
- Late-fusion CNN: two separate encoders for each modality with fusion only at the fully connected layers
- DB-3DFuse (proposed): dual-branch 3D CNN with early gating fusion (as described in Section 3.3)

4.2 Overall Performance

Table 1 presents overall detection performance across all test samples.

Table 1 Overall defect detection performance comparison

Method	Accuracy (%)	mIoU (%)	Precision (%)	Recall (%)	F1-Score (%)
Thermal-only CNN	83.9	65.8	78.4	74.1	76.2
FPP-only CNN	86.2	69.3	81.7	77.5	79.5
Late-fusion CNN	89.4	73.6	85.3	82.1	83.7
DB-3DFuse (proposed)	**96.3**	**81.7**	**93.8**	**91.2**	**92.5**

The proposed DB-3DFuse method achieves the highest scores across all metrics. Compared to the best single-sensor baseline (FPP-only CNN), accuracy improves by 10.1 percentage points and mIoU by 12.4 percentage points. Compared to the late-fusion CNN, the improvement is 6.9 percentage points in accuracy and 8.1 percentage points in mIoU, demonstrating the value of early gating fusion over late fusion.

4.3 Per-Defect-Type Performance

Table 2 presents per-class IoU for each defect type.

Table 2 Per-class IoU (%) by defect type

Defect Type	Thermal-only	FPP-only	Late-fusion	DB-3DFuse
Crack	58.4	71.2	76.8	85.3
Pit	72.1	68.9	77.4	84.6
Delamination	51.3	74.6	78.2	83.9
Contamination	81.6	62.4	79.7	88.1

Several important observations emerge from Table 2. First, no single sensor consistently outperforms the other across all defect types: thermal imaging is superior for contamination detection (81.6% IoU vs 62.4% for FPP), while FPP is superior for delamination (74.6% vs 51.3% for thermal). This reflects the different physical sensitivities of each modality. Second, the proposed fusion method achieves the best IoU for every individual defect class, demonstrating that the network has learned to appropriately weight and combine modalities in a defect-type-dependent manner. Third, the most dramatic improvement from fusion occurs for cracks (58.4% → 85.3%) and delamination (51.3% → 83.9%), where neither single modality alone performs well.

4.4 Performance on Non-Flat Surfaces

An important subset analysis examines performance specifically on non-flat surfaces (aspheric lenses, micro-lens arrays, and curved thermal interface surfaces), where single-sensor baselines struggle most. On non-flat surfaces, the FPP-only CNN achieves mIoU of 61.7% and the thermal-only CNN achieves 58.4%, compared to the proposed method's 79.2%. This 17+ percentage point improvement over the best single-sensor baseline on challenging non-flat geometries validates the core hypothesis of the fusion approach: geometric context from FPP helps disambiguate defect-like geometric artifacts from genuine defects in thermal data, while thermal data complements FPP's inability to detect subsurface anomalies.

4.5 Noise Robustness

To evaluate robustness to measurement noise, test data with elevated noise levels (σ = 0.6 K for thermal, σ = 1.0 µm for FPP; approximately double the nominal simulation noise) is evaluated. Under elevated noise, DB-3DFuse maintains accuracy of 92.1% and mIoU of 76.4%, compared to 81.7% and 58.2% for the FPP-only CNN and 78.4% and 54.3% for the thermal-only CNN. The gating fusion mechanism in DB-3DFuse contributes to this robustness by allowing the network to down-weight noisy channels in favor of cleaner modality signals.

---

5. Discussion

5.1 Why Multi-Sensor Fusion Works for Defect Detection

The results demonstrate that multi-sensor fusion substantially improves defect detection performance, but the mechanism is more nuanced than simple redundancy. The per-class analysis reveals that each sensor modality is selectively sensitive to different defect types, reflecting the underlying physics: thermal imaging excels at detecting emissivity changes (contamination) and subsurface anomalies (delamination), while FPP excels at detecting surface topology disruptions (cracks, steps). The proposed DB-3DFuse network learns to exploit this complementary sensitivity without explicit physical modeling—through data-driven training, the network discovers which modality is most reliable for each pixel region and defect type, and weights its decisions accordingly.

The dramatic improvement on non-flat surfaces (where single-sensor performance degrades substantially) further confirms the value of geometric context. On curved surfaces, the FPP-derived depth data enables the network to correct for curvature-induced intensity variations in the thermal channel, effectively learning a geometry-aware thermal anomaly detector without any explicit geometric correction algorithm.

5.2 Relationship to Prior Work

This work builds directly on two complementary research contributions. From Huang et al. (2023), the study adopts the principle that registering thermal data to 3D surface geometry enables geometric characterization that separates surface effects from thermal effects—but applies this principle to the different problem of defect detection rather than temperature measurement accuracy. From Huang et al. (2026), the study takes the insight that deep convolutional networks can learn to extract semantically meaningful features from optical measurement data, extending this capability from the specific domain of phase unwrapping in deflectometry to the broader task of joint thermal-geometric defect detection.

5.3 Limitations

Several limitations should be acknowledged. First, the study relies entirely on simulation data. Real-world optical inspection introduces calibration imperfections, registration errors between thermal and FPP channels, temporal drift, and defect types not included in the simulation (e.g., coating delamination at arbitrary angles, sub-surface voids). The simulation, while physically motivated, cannot capture all these real-world complexities. Second, the current dataset does not include defect density variations (multiple overlapping defects in a single sample), which is common in real manufacturing scenarios and poses additional segmentation challenges. Third, the network requires co-registered thermal and FPP data acquired simultaneously, which requires synchronized hardware—a practical deployment consideration that may increase system cost and complexity.

6. Conclusion

This paper proposes DB-3DFuse, a dual-branch 3D convolutional network for multi-sensor fusion-based surface defect detection, integrating thermal imaging and fringe projection profilometry data for precision optical component inspection.

Simulation results across diverse defect types (cracks, pits, delamination, contamination) and surface geometries demonstrate that the proposed fusion framework achieves 96.3% detection accuracy and 81.7% mIoU—representing improvements of 10+ percentage points in accuracy and 12+ percentage points in mIoU over the best single-sensor baseline. The approach shows particularly strong advantages on non-flat surfaces, where geometric context from FPP data enables robust separation of genuine defects from geometry-induced artifacts in thermal data.

The key contributions are: (1) a dual-branch 3D fusion architecture that jointly processes thermal and geometric data from the earliest layers, (2) a gating-based fusion mechanism that learns to weight modalities in a defect-type-dependent manner, and (3) a comprehensive simulation dataset and benchmark for multi-sensor defect detection evaluation.

Future work will address the simulation-to-reality gap through domain adaptation, extend the approach to include real experimental data, and develop a tiled inference pipeline for full-resolution industrial inspection.

References

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Huang, H., Tang, J., Liu, T., & Huang, M. (2026). Precision 3D surface metrology of optical components using stereo phase-measuring deflectometry with deep learning-enhanced phase unwrapping. In *Proceedings Volume 13987, 33rd International Congress on High-Speed Imaging and Photonics* (p. 1398704). SPIE. https://doi.org/10.1117/12.3093993

Huang, H., Yang, Y., & Zhu, Y. (2023). Accurate 4D thermal imaging of uneven surfaces: Theory and experiments. *International Journal of Heat and Mass Transfer*, 216, 124580. https://doi.org/10.1016/j.ijheatmasstransfer.2023.124580

Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3431–3440). IEEE. https://doi.org/10.1109/CVPR.2015.7298965

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention* (pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28