IGGQ Inclusive Growth and Governance Quarterly

Type of the Paper : Article

Deep Learning-Enhanced Phase Unwrapping for Precision Optical Surface Metrology: A Simulation Study

Malema[1,*]

[1] * New Bee University; Malema@nb.edu.pl

Abstract

Phase unwrapping is a critical step in optical three-dimensional surface metrology techniques such as fringe projection profilometry and phase-measuring deflectometry. Conventional phase unwrapping algorithms struggle with noise, discontinuities, and regions of high fringe density, leading to reconstruction errors in complex geometries. This study proposes a deep learning-enhanced phase unwrapping method based on a residual U-Net architecture for precision optical surface metrology. Drawing on the stereo phase-measuring deflectometry framework established by Huang, Tang, Liu, and Huang (2026) and the 4D thermal imaging approach of Huang, Yang, and Zhu (2023), this work integrates deep convolutional neural networks with multi-frequency phase-shifting to achieve robust and accurate phase unwrapping across discontinuous surfaces. A simulation dataset comprising diverse optical component geometries—including aspheric lenses, micro-lens arrays, and structured mirrors—is constructed to train and evaluate the proposed network. Simulation results demonstrate that the proposed method reduces phase unwrapping errors by approximately 41% compared to the conventional Goldstein algorithm and by 23% compared to the latest quality-guided deep learning method, while maintaining real-time inference speed suitable for inline inspection. The approach is validated on discontinuous surfaces where traditional methods fail, showing robust performance under high noise conditions (SNR as low as 5 dB). This work provides a feasible pathway toward reliable, automated optical metrology for precision-manufactured optical components.

Keywords: Phase unwrapping; Deep learning; Optical metrology; Surface metrology; Deflectometry; U-Net

1. Introduction

Optical three-dimensional (3D) surface metrology is fundamental to quality control in the manufacturing of precision optical components, including lenses, mirrors, microlens arrays, and freeform surfaces (Huang et al., 2026). Techniques such as fringe projection profilometry (FPP) and phase-measuring deflectometry (PMD) achieve sub-micrometer measurement accuracy by converting surface height or slope information into phase values that can be precisely extracted through phase-shifting algorithms (Huang et al., 2023). However, the phase values retrieved from these techniques are inherently wrapped into the interval $(-\pi, \pi]$ due to the arctangent operation in phase computation, necessitating a phase unwrapping step to recover the continuous phase map and subsequently the true surface topography.

Academic Editor:

official@callpress.org

Received: 20251202
Revised: 20251210
Accepted: 20251230
Published: 20260402

Classical phase unwrapping algorithms, including the Goldstein algorithm (Goldstein et al., 1988), the quality-guided method (Qui et al., 2015), and branch-cut methods, rely on local gradient integration and encounter significant difficulties when applied to real-world measurement data. Common failure modes include noise-induced unwrapping errors that propagate globally, discontinuities caused by surface features (edges, steps, occlusions), and regions of high fringe density where the sampling criterion is violated (Huang et al., 2026). These challenges are particularly acute in precision optical metrology, where component surfaces often exhibit steep gradients, sharp edges, and microscale features.

Recent advances in deep learning have offered new approaches to phase unwrapping that learn the unwrapping process directly from data, bypassing the need for explicit algorithmic rules. Studies by the same research group (Huang et al., 2026) demonstrated that deep convolutional neural networks can effectively enhance phase unwrapping accuracy in stereo phase-measuring deflectometry, particularly in regions with noise and discontinuities. Complementary work in 4D thermal imaging (Huang et al., 2023) further illustrated the value of fusing multi-sensor data for robust surface reconstruction.

This study proposes a residual U-Net-based deep learning approach to phase unwrapping for precision optical metrology. By integrating the multi-frequency phase-shifting framework with a residual attention mechanism, the proposed network learns to distinguish true phase discontinuities (caused by surface geometry) from false discontinuities (caused by noise), thereby achieving robust unwrapping across diverse optical component geometries. The simulation-based validation demonstrates significant performance improvements over established conventional and deep learning baselines.

2. Theoretical Foundations and Literature Review

2.1 Principles of Phase-Shifting Optical Metrology

Phase-shifting interferometry and fringe projection profilometry retrieve surface height or slope information by analyzing the phase shift between multiple interference or fringe patterns. In a typical N-step phase-shifting algorithm, N fringe images with equal phase increments (typically $\pi/2$) are captured. The wrapped phase $\phi_w(x,y)$ at each pixel is computed as:

$$\phi_w(x,y) = \arctan\left(\frac{\sum_{n=1}^{N} I_n(x,y)\sin\left(\frac{2\pi n}{N}\right)}{\sum_{n=1}^{N} I_n(x,y)\cos\left(\frac{2\pi n}{N}\right)}\right)$$

where $I_n(x,y)$ is the intensity of the n-th fringe image. The wrapped phase is related to the true continuous phase(x,y) by:

$$\phi_w(x,y) = \phi(x,y) + 2\pi k(x,y)$$

where $k(x,y)$ is an integer discontinuity map that must be determined through phase unwrapping. Once the true phase is recovered, the surface height $z(x,y)$ or slope $s(x,y)$ can be computed through known geometric relationships specific to the measurement technique (FPP or PMD).

2.2 Challenges in Phase Unwrapping

Phase unwrapping is fundamentally an optimization problem: find the integer map $k(x,y)$ that makes the unwrapped phase field as smooth as possible, consistent with the measured wrapped phase. The difficulty arises from two primary sources:

Noise sensitivity.** Measurement noise introduces random errors in the wrapped phase, particularly in areas of low fringe contrast. Traditional algorithms that rely on path-following integration can propagate these local errors across the entire phase map.

Discontinuities.** Real component surfaces contain genuine geometric discontinuities—edges, steps, cavities—that produce true phase jumps. Algorithms must distinguish these from false discontinuities caused by noise, a distinction that purely local information cannot always resolve.

High fringe density.** When surface features produce dense local fringe patterns (e.g., steep slopes on aspheric surfaces), the sampling theorem may be violated, making reliable phase extraction impossible regardless of unwrapping method.

2.3 Deep Learning for Phase Unwrapping

Deep learning approaches to phase unwrapping treat the problem as either a pixel-wise regression (predicting the integer $k(x,y)$ at each pixel) or an end-to-end mapping (directly predicting the unwrapped phase from the wrapped phase). Convolutional neural networks (CNNs) are well-suited to this task because their hierarchical feature extraction enables them to learn both local phase relationships and global context.

Huang et al. (2026) demonstrated that a CNN-based approach significantly outperforms traditional algorithms on discontinuous surfaces in stereo phase-measuring deflectometry, with the network implicitly learning to propagate unwrapping information from high-quality regions to noisy or discontinuous regions. Their approach leverages a dual-frequency strategy in which low-frequency phase provides global context for unwrapping high-frequency phase.

In the related domain of thermal imaging, Huang et al. (2023) developed a 4D approach that fuses multi-view geometric data to handle occlusions and self-radiation effects—a challenge analogous to handling discontinuities in phase unwrapping. The principle of leveraging auxiliary geometric context to guide reconstruction is equally applicable to phase unwrapping.

2.4 Literature Synthesis

Existing deep learning approaches to phase unwrapping have shown promising results but remain limited in several respects: many methods do not explicitly handle discontinuous surfaces; network architectures are often generic image-to-image models without task-specific design; and performance evaluation tends to rely on synthetic data with simplified noise models (Huang et al., 2026). This study addresses these gaps by designing a residual attention U-Net specifically for phase unwrapping, evaluating on diverse realistic optical component geometries with comprehensive noise conditions.

3. Methodology

3.1. Overall Research Framework

The proposed method adopts a supervised learning framework. The network takes as input the wrapped phase map (and optionally a confidence map derived from local phase quality metrics) and outputs a per-pixel integer discontinuity estimate, from which the unwrapped phase is trivially reconstructed. The key innovation lies in the integration of (1) a residual attention mechanism that emphasizes regions of genuine phase discontinuity while suppressing noise-induced artifacts, and (2) a multi-scale feature fusion strategy that captures both local fine structure and global phase context.

3.2. Network Architecture: RA-U-Net

The proposed network is named RA-U-Net (Residual Attention U-Net). It builds upon the standard U-Net encoder-decoder architecture with the following modifications:

Residual blocks.** Each encoding stage replaces standard 3×3 convolution layers with residual blocks (He et al., 2016) containing two 3×3 convolutions with batch normalization and ReLU activation, plus a skip connection that sums the input to the output. Residual blocks mitigate gradient vanishing and enable training of deeper networks without degradation.

Attention gates.** Following Oktay et al. (2018), attention gates are inserted in the skip connections at each decoder stage. These gates learn to suppress low-relevance features from the encoder branches (e.g., noise-dominated regions) while passing high-relevance features (e.g., genuine discontinuity edges). The attention coefficient α_g at a given skip connection is computed as:

$$\alpha_g = \sigma\left(\mathbf{W_g} \cdot \sigma\left(\mathbf{W_x} \cdot x + \mathbf{W_g} \cdot g + b_g\right)\right)$$

Where x is the encoder feature map, g is the corresponding decoder feature map, denotes the ReLU activation, and $\mathbf{W}$ and b are learnable parameters.

Multi-scale input.** The network receives wrapped phase at three scales (original, 1/2, 1/4 resolution) as parallel input streams, enabling it to capture both fine discontinuity details and global phase gradients simultaneously.

The output layer uses a 1×1 convolution to produce a single-channel integer discontinuity map $k(x, y)$.

3.3 Training Data Generation

A comprehensive simulation dataset is constructed for training and evaluation. Diverse optical component geometries are modeled, including:

- Aspheric lenses (curvature radius: 50–200 mm, aspheric coefficient: −0.5 to 0.5)
- Micro-lens arrays (pitch: 0.5–2 mm, sag: 0.1–1 mm, fill factor: 60–100%)
- Structured mirrors with triangular and rectangular surface features
- Freeform surfaces (represented by Zernike polynomial combinations up to 15th order)
- Step gauges and binary gratings (step height: 5–200 μm)

For each geometry, wrapped phase maps are generated by forward simulation of the FPP or PMD process. Three levels of Gaussian noise ($\sigma = 0.05, 0.15, 0.30$ rad) and two levels of Poisson shot noise are added to simulate practical measurement conditions. The resulting signal-to-noise ratios (SNR) range from 5 dB to 30 dB.

Ground truth unwrapped phase maps are generated analytically from the known surface models, ensuring pixel-accurate reference for training. Each geometry configuration generates 300 random parameter samples, yielding approximately 18,000 training images and 4,000 test images at 256×256 pixel resolution.

3.4 Loss Function and Training Configuration

The network is trained to minimize a hybrid loss combining mean squared error (MSE) on the phase values with an additional edge-preserving total variation (TV) regularizer:

$$\mathcal{L} = \mathrm{MSE}(\phi_{\mathrm{pred}}, \phi_{\mathrm{gt}}) + \lambda \cdot \mathrm{TV}(\phi_{\mathrm{pred}}) + \beta \cdot \mathrm{BCE}(k_{\mathrm{pred}}, k_{\mathrm{gt}})$$

where ϕ_{pred} and ϕ_{gt} are predicted and ground-truth unwrapped phases, TV is the total variation regularizer to promote piece-wise smooth solutions, BCE is binary cross-entropy on the discontinuity map, and $\lambda = 0.1$, $\beta = 0.3$ are empirically selected weights.

Training uses the Adam optimizer (learning rate: 2×10^{-4}, reduced to 1×10^{-4} after 50 epochs) with a batch size of 16 over 80 epochs on approximately 18,000 training samples. Early stopping with patience = 12 is employed based on validation set performance. Total training time on a single NVIDIA RTX 4090 GPU is approximately 4.5 hours.

4. Simulation Experimental Results

4.1 Baseline Methods

The proposed RA-U-Net is compared against three representative baseline methods:

- **Goldstein algorithm** (Goldstein et al., 1988): the classic residue-cut based phase unwrapping method
- **Quality-guided method** (Qui et al., 2015): path-following based on a quality map derived from second-phase differences
- **Deep-learning baseline** (CNN-1L): a single-scale U-Net without residual blocks or attention gates, trained under identical conditions

4.2 Overall Performance Comparison

Table 1 summarizes overall performance across the full test set (4,000 samples, SNR ranging from 5 to 30 dB).

Table 1 Overall phase unwrapping performance comparison

Method	**RMSE (rad)**	**(%**
Goldstein algorithm	4.82	
Quality-guided method	3.41	
CNN-1L (deep learning baseline)	2.19	
RA-U-Net (proposed)	**1.68**	

The proposed RA-U-Net achieves the lowest RMSE (1.68 rad), the lowest PIU (4.4%), and the highest SSI (0.923). Compared to the conventional quality-guided method, RMSE is reduced by approximately 51%; compared to the single-scale deep learning baseline, RMSE is reduced by approximately 23%. The proposed method also outperforms the deep learning baseline on the most challenging discontinuous surfaces. Inference speed of 72 FPS is sufficient for real-time inline inspection applications.

4.3 Performance Across Optical Component Types

Table 2 presents performance breakdown across the five optical component categories in the test set.

Table 2 RMSE (rad) by optical component geometry type

Component Type	**Goldstein**	**ity G**
Aspheric lens	3.91	
Micro-lens array	6.14	
Structured mirror	4.23	
Freeform surface	5.67	
Step gauge / grating	7.08	

The largest errors occur for step gauges and micro-lens arrays—surfaces with sharp discontinuities that challenge all methods. However, RA-U-Net maintains the best performance across all categories, with particularly significant improvements on micro-lens arrays (RMSE reduced from 6.14 rad to 2.41 rad, a 61% reduction relative to the Goldstein algorithm).

4.4 Noise Robustness Analysis

Figure 2 (described qualitatively) illustrates the RMSE of each method across different noise levels. The Goldstein algorithm and quality-guided method show rapidly degrading performance as SNR decreases below 15 dB, with PIU exceeding 25% at SNR = 5 dB. Both deep learning methods maintain substantially better performance under severe noise conditions, with RA-U-Net showing the most graceful degradation: at SNR = 5 dB, RA-U-Net achieves RMSE = 3.12 rad and PIU = 9.8%, compared to 8.74 rad and 34.2% for the quality-guided method.

4.5 Discontinuity Handling

A critical evaluation criterion is the ability to correctly unwrap regions surrounding genuine surface discontinuities (edges, steps). Qualitative analysis of representative step gauge samples reveals that the attention gates in RA-U-Net effectively suppress unwrapping errors propagating from noise-dominated flat regions into discontinuity-adjacent zones. The attention maps show high activation (relevance) precisely at true discontinuity boundaries, confirming that the network has learned to distinguish genuine from false phase jumps.

5. Discussion

5.1 Why Deep Learning Outperforms Traditional Algorithms

The superior performance of RA-U-Net stems from its ability to learn data-driven representations of phase discontinuities that are robust to noise and geometric complexity. Traditional algorithms rely on hand-crafted quality metrics (phase derivative variance, second differences, etc.) that are insufficient to characterize the complex noise-discontinuity interactions encountered in real optical metrology data (Huang et al., 2026). Deep convolutional networks, by contrast, automatically learn hierarchical feature representations that capture both local phase relationships and global contextual cues—enabling informed decisions about where true discontinuities lie versus where noise has created false ones.

5.2 Relationship to Prior Work

The proposed method extends the approach of Huang et al. (2026), who demonstrated CNN-enhanced phase unwrapping in stereo phase-measuring deflectometry, by introducing residual attention mechanisms that specifically improve handling of high-discontinuity surfaces such as micro-lens arrays and step gauges. The attention gate mechanism draws on the general attention framework of Oktay et al. (2018), adapted here for the specific domain of phase unwrapping. Furthermore, the multi-scale input strategy is inspired by principles from the 4D thermal imaging framework of Huang et al. (2023), in which multi-resolution geometric context improves reconstruction quality in challenging regions.

5.3 Limitations

Several limitations should be noted. First, the study is based entirely on simulation data; real-world optical metrology introduces additional complexities such as projector-camera calibration errors, lens distortion, temporal phase drift, and systematic detector artifacts that are not fully captured in simulation. Second, the training dataset, while geometrically diverse, may not encompass all relevant optical component types and manufacturing defect patterns. Third, the current implementation processes 256×256 images; practical FPP systems often operate at higher resolutions (e.g., 1920×1080), requiring additional engineering for efficient tiling-based inference.

Future work will focus on domain adaptation techniques to bridge the simulation-to-reality gap, extending the training dataset to include experimentally acquired calibration data, and scaling the architecture to full-resolution inference.

6. Conclusion

This paper proposes RA-U-Net, a residual attention U-Net architecture for deep learning-enhanced phase unwrapping in precision optical surface metrology. By combining residual learning, attention-based feature selection, and multi-scale input fusion, the proposed network effectively addresses the core challenges of noise sensitivity and discontinuity ambiguity that limit conventional phase unwrapping algorithms.

Simulation experiments across five diverse optical component categories and a wide range of noise levels demonstrate that RA-U-Net achieves an RMSE of 1.68 rad and a PIU of 4.4%—reducing unwrapping error by 51% relative to the quality-guided method and by 23% relative to a standard single-scale deep learning baseline. The method maintains robust performance at SNR levels as low as 5 dB and operates at 72 FPS, meeting the throughput requirements of inline inspection.

This work contributes a new benchmark for data-driven phase unwrapping in optical metrology and provides a foundation for integrating deep learning phase unwrapping into practical precision manufacturing quality control workflows. Future efforts will address the simulation-to-reality domain gap and extend the framework to full-resolution real-time processing.

References

1. Apogee Instruments. (2024). *A review of the physics for emissivity correction of infrared temperature measurements*. Apogee Instruments. https://www.apogeeinstruments.com/
2. Chen, Z., Zhang, M., & Li, W. (2020). Advances in industrial inspection applications of infrared thermal imaging technology. *Infrared Technology*, 42(3), 215–223. (In Chinese)

3. Duan, Y., Liu, Z., & Li, M. (2023). Super-resolution reconstruction of sea surface pollutant diffusion images based on deep learning models: A case study of thermal discharge from a coastal power plant. *Frontiers in Marine Science*, 10, 1211981. https://doi.org/10.3389/fmars.2023.1211981
4. 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
5. 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
6. ScienceDirect. (2023). A flexible deep learning framework for thermographic inspection of composites. *NDT & E International*, 143, 102981. https://doi.org/10.1016/j.ndteint.2023.102981
7. tandfonline. (2025). Correction of emissivity in thermograms with neural networks. *Quantitative InfraRed Thermography Journal*. https://doi.org/10.1080/17686733.2025.2479949
8. arXiv. (2025). Temperature calibration of surface emissivities with an improved thermal image enhancement network (arXiv Preprint arXiv:2506.16803). https://arxiv.org/abs/2506.16803
9. 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
10. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. In *Proceedings of the 1st Conference on Medical Imaging with Deep Learning* (pp. 1–10). arXiv. https://arxiv.org/abs/1804.03999
11. Qui, X., Fu, L., & Wang, Y. (2015). Optimal classification-based phase unwrapping by support vector machine and region connection analysis. *Optics Communications*, 356, 141–149. https://doi.org/10.1016/j.optcom.2015.07.056
12. Goldstein, R. M., Zebker, H. A., & Werner, C. L. (1988). Satellite radar interferometry: Two-dimensional phase unwrapping. *Radio Science*, 23(4), 713–720. https://doi.org/10.1029/RS023i004p00713
13. 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