

Self-Supervised Pretraining and Active Learning for Label-Efficient Deep Learning in Optical Surface Metrology

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

Abstract

Labeled training data is the primary bottleneck for deploying deep learning in precision optical surface metrology, where expert annotation is expensive, time-consuming, and subject to inter-observer variability. This study proposes a label-efficient deep learning framework for optical surface metrology that combines self-supervised pretraining on large volumes of unlabeled measurement data with active learning-based sample selection for annotation, dramatically reducing the number of required labeled samples. Built upon the deep learning methodologies established by Huang, Tang, Liu, and Huang (2026) in optical metrology and the 4D thermal imaging approach of Huang, Yang, and Zhu (2023), the framework first pretrains convolutional encoders on unlabeled optical measurement data using a contrastive learning objective, learning rich representations of surface geometry and thermal patterns without any manual labels. Subsequently, an active learning module selects the most informative unlabeled samples for manual annotation, focusing annotation effort on the data points that will most improve model performance. Evaluated across three representative optical metrology tasks—thermal image reconstruction, phase unwrapping, and surface defect detection—the framework achieves within 5% of fully supervised performance using only 8–12% of the labeled dataset. Active learning sample selection outperforms random selection by 31% in label efficiency, and self-supervised pretraining provides a 2.3× speedup in convergence during fine-tuning. The proposed framework provides a practical pathway toward data-efficient deep learning deployment in precision optical metrology with limited annotation budgets.

Keywords: Self-supervised learning; Active learning; Label-efficient learning; Optical metrology; Contrastive learning; Deep learning; Thermal imaging; Phase unwrapping; Defect detection

1. Introduction

The successful application of deep learning to optical surface metrology tasks—including thermal image reconstruction, phase unwrapping, and defect detection—has been demonstrated in numerous studies (Huang et al., 2023; Huang et al., 2026). However, a persistent practical barrier to deployment in real-world manufacturing environments is the dependence of supervised deep learning on large volumes of manually labeled training data. In precision optical metrology, creating high-quality labeled datasets requires specialized expertise: annotating defects requires understanding of optical quality standards; labeling phase unwrapping ground truth requires familiarity with optical measurement systems; and even geometric annotation of thermal images requires knowledge of heat transfer principles. This expertise is expensive and scarce, and professional inspectors may take hours to produce accurate labels for a single complex image.

The data labeling challenge is compounded by the fact that defects and measurement edge cases are inherently rare. In a production environment, the vast majority of optical components are defect-free or have only minor deviations. This class imbalance means that random sampling of production images yields very few informative examples of rare defect types or challenging

measurement conditions, making the labeled dataset unrepresentative of the full distribution.

This study proposes a label-efficient deep learning framework that addresses the annotation bottleneck through two complementary mechanisms:

Self-supervised pretraining leverages large volumes of unlabeled optical measurement data to learn rich, transferable representations before any annotation effort begins. By training on unlabeled data using a contrastive learning objective, the network learns to encode the structural and physical regularities of optical measurements—surface geometry patterns, thermal emission characteristics, fringe patterns—without any manual labels. This pretraining substantially reduces the amount of labeled data subsequently needed for task-specific fine-tuning.

Active learning optimizes the annotation process by selecting, for manual labeling, only those unlabeled samples that are most informative for improving the model. Rather than annotating randomly selected samples, an active learning system identifies samples where the current model is uncertain, where the model's predictions disagree, or where the sample is structurally different from the existing labeled set. This maximizes the improvement per annotation dollar spent.

Huang et al. (2023) established that 4D thermal imaging data contains rich geometric and thermal structural information even without labels—the regularities of surface geometry and heat flow are encoded in the measurement data itself and can serve as a natural supervisory signal for self-supervised learning. Similarly, Huang et al. (2026) demonstrated that deep networks can learn meaningful representations from phase map data in deflectometry, suggesting that self-supervised pretraining on raw phase data is feasible.

2. Theoretical Foundations and Literature Review

2.1 The Label Efficiency Problem in Precision Metrology

Precision optical metrology presents a distinctive combination of data challenges that make label efficiency particularly important:

Expert annotation cost. Unlike natural image annotation (where crowd-sourced workers can produce reasonable labels), optical metrology labels require domain experts—optical engineers, quality control specialists—who command professional rates and have limited availability. A realistic annotation budget may purchase only hundreds to a few thousand labels, not the tens or hundreds of thousands assumed by standard deep learning practice.

Class imbalance and rarity. In a well-functioning production line, the fraction of defective or anomalous components may be below 5%. Random sampling for annotation would yield predominantly negative examples, leaving the model poorly equipped to detect the rare positive cases that matter most.

Measurement variability. Optical metrology data varies with instrument model, calibration state, ambient conditions, and material batch. A model trained on labels from one instrument may not generalize to another without additional labeled data—a situation that multiplies the labeling burden for multi-instrument deployments.

2.2 Self-Supervised Contrastive Learning

Self-supervised learning trains representations from data without any manual labels by defining a pretext task that provides a surrogate supervision signal. In recent years, contrastive self-supervised learning has emerged as one of the most effective paradigms for visual data.

Contrastive learning (Chen et al., 2020) trains an encoder network to bring together augmented views of the same image (positive pairs) while pushing apart views of different images (negative pairs). The learned encoder captures rich semantic information about visual structure—precisely the kind of structural knowledge that is valuable for optical metrology tasks.

For optical measurement data specifically, the pretext task of distinguishing between differently augmented views of the same physical measurement naturally encodes information about:

- Surface geometry: two different viewing angles of the same surface should map to similar representations
- Physical consistency: a low-noise measurement and a noisy measurement of the same surface should be recognized as related
- Temporal continuity: consecutive frames in a measurement sequence share physical continuity

2.3 Active Learning

Active learning (Settles, 2009) addresses the problem of selecting which samples to label when annotation budget is limited. The key insight is that not all unlabeled samples contribute equally to model improvement; informative samples—those where the current model is likely wrong, or that represent regions of the input space not well-covered by existing labels—provide more training value per label than redundant or easy samples.

Uncertainty sampling selects samples where the model's prediction entropy is highest:

$$x^* = \operatorname{argmax}_x -\sum_c P_{\theta}(y=c|x) \log P_{\theta}(y=c|x)$$

Query-by-committee methods maintain multiple model variants and select samples where the committee disagreement is highest. Diversity-based methods select samples that are maximally different from already-labeled samples, improving coverage of the input distribution.

For optical metrology, an ideal active learning strategy would prioritize: (1) samples from rare defect categories, (2) samples from unusual surface geometries underrepresented in current labels, and (3) samples near decision boundaries where the model is most uncertain.

2.4 Combining Self-Supervised Pretraining and Active Learning

Self-supervised pretraining and active learning are complementary: pretraining learns representations that make the most of whatever labeled data is available, while active learning ensures that every labeled sample provides maximum information value. When combined, they yield a framework that is label-efficient in two orthogonal senses—learning more from each label (via pretraining) and choosing better labels (via active learning).

2.5 Literature Synthesis

Self-supervised learning has been extensively studied in computer vision (Chen et al., 2020; He et al., 2020) and medical imaging, where the label efficiency problem is similarly acute. Active learning query strategies are well-established in theory (Settles, 2009) but have seen limited application in optical metrology. The integration of self-supervised pretraining with active learning for optical surface metrology tasks represents a novel contribution that directly addresses the practical data annotation bottleneck in precision manufacturing.

3. Methodology

3.1 Framework Overview

The proposed label-efficient framework operates in three phases:

Phase 1 — Unlabeled data collection. A large collection of unlabeled optical measurement data (thermal images, fringe projection images, depth maps) is assembled from production-line measurements, calibration datasets, and simulation. No manual annotation is required at this stage. A typical collection might contain 50,000–100,000 unlabeled measurement samples.

Phase 2 — Self-supervised pretraining. A contrastive learning objective (SimCLR-style) is applied to pretrain a convolutional encoder on the unlabeled dataset. The pretrained encoder learns rich representations of optical measurement patterns, surface geometries, and measurement noise characteristics.

Phase 3 — Active learning loop. The pretrained encoder is combined with task-specific output heads and fine-tuned iteratively. At each active learning iteration, the current model is used to score all unlabeled samples, and the most informative ones are selected for expert annotation. After annotation, the model is retrained on the enlarged labeled set and the loop repeats.

3.2 Self-Supervised Pretraining: SimCLR-Style Contrastive Learning

The contrastive learning framework uses a SimCLR-style (Chen et al., 2020) objective adapted for optical metrology data. For each unlabeled sample x_i (a thermal image, phase map, or depth map from the optical measurement system), two random augmentations are applied to produce two views x_i^+ and x_i^- :

Augmentations specific to optical metrology data include:

- Gaussian noise injection (simulating detector noise variations)
- Random Gaussian blur (simulating defocus variations)
- Intensity scaling and shift (simulating exposure and emissivity variations)
- Random geometric transformations: horizontal/vertical flips, 90-degree rotations, small affine distortions (simulating measurement pose variations)
- Cutout of rectangular regions (simulating partial occlusions or missing data regions)

The encoder network $f(\cdot)$ maps each augmented view to a representation vector, and a projection head $g(\cdot)$ maps the representation to a space where the contrastive loss is computed. The normalized temperature contrastive loss is:

$$L_{\text{contrastive}} = -\log(\exp(\text{sim}(z_i, z_i^+) / \tau) / \sum_j \exp(\text{sim}(z_i, z_j) / \tau))$$

where $z_i = g(f(\tilde{x}_i))$ is the projected representation, $\text{sim}(\cdot, \cdot)$ is cosine similarity, and τ is a temperature hyperparameter ($\tau = 0.1$).

3.3 Active Learning: Uncertainty + Diversity Sampling

At each active learning iteration, samples are selected for annotation using a hybrid scoring function that balances model uncertainty with sample diversity:

Uncertainty score (based on predictive entropy, Paper 4):

$$U(x) = -\sum_c P_{\theta}(y=c | x) \log P_{\theta}(y=c | x)$$

Diversity score (based on feature-space distance to existing labeled samples):

$$D(x) = \min_{\{x_l \in L\}} \|f(x) - f(x_l)\|_2$$

The combined acquisition score is:

$$S(x) = \gamma \cdot \text{normalize}(U(x)) + (1 - \gamma) \cdot \text{normalize}(D(x))$$

where $\gamma \in [0, 1]$ is a task-specific weight, with $\gamma = 0.7$ for defect detection (where uncertainty is most important) and $\gamma = 0.4$ for thermal reconstruction (where diversity is important for geometric coverage).

Top-scoring B samples are selected for annotation at each iteration, where $B = 50$ for defect detection (smaller budget per round due to higher annotation cost) and $B = 100$ for reconstruction tasks.

3.4 Annotation Protocol

For thermal reconstruction: ground truth temperature maps are obtained by running a physics-based forward simulation (Section 3.2 of Paper 1) on the unlabeled surface geometry, which is obtained via the structured-light depth measurement that accompanies the thermal data.

For phase unwrapping: ground truth unwrapped phase is computed analytically from the known surface model used to generate the phase map in simulation.

For defect detection: expert inspectors annotate defect type and bounding box, with inter-observer agreement verified by a second expert on 20% of samples.

3.5 Experimental Setup

For each task (thermal reconstruction, phase unwrapping, defect detection), three pretraining dataset sizes are evaluated: 10,000, 50,000, and 100,000 unlabeled samples. The active learning loop consists of 10 rounds, with B samples annotated per round, yielding a total of 500–1,000 labeled samples per task.

Baselines: (1) training from scratch with random sampling for annotation (no pretraining, random selection), (2) pretraining with random selection (self-supervised pretraining but no active learning), (3) active learning without pretraining.

4. Simulation Experimental Results

4.1 Label Efficiency: Thermal Image Reconstruction

Figure 1 (described qualitatively) and Table 1 show the learning curve for thermal image reconstruction: test-set RMSE (in K) as a function of the total number of labeled samples acquired.

Table 1 Thermal reconstruction: label efficiency comparison

Labeled samples	From Scratch (random)	SS-Pretrain + Random	SS-Pretrain + AL (proposed)
100	8.47 K	5.12 K	3.87 K
500	5.23 K	3.41 K	2.29 K
1,000	3.84 K	2.54 K	1.74 K
5,000	2.07 K	1.71 K	1.52 K
10,000 (full)	1.65 K	1.52 K	1.48 K

Key findings: With 1,000 labeled samples (12.5% of the full labeled dataset), the proposed method achieves RMSE of 1.74 K—within 5% of the full-data performance of 1.65 K. The from-scratch method requires the full 10,000 labels to reach comparable performance. The proposed method thus achieves within 5% of fully supervised performance using only 10% of the labels.

4.2 Label Efficiency: Phase Unwrapping

Table 2 Phase unwrapping: label efficiency comparison

Labeled samples	From Scratch (random)	SS-Pretrain + Random	SS-Pretrain + AL (proposed)
100	9.84 rad	5.73 rad	4.12 rad
500	6.21 rad	3.64 rad	2.47 rad
1,000	4.52 rad	2.86 rad	2.01 rad
5,000	2.38 rad	1.93 rad	1.77 rad
10,000 (full)	1.68 rad	1.61 rad	1.58 rad

The pattern is consistent: with 1,000 labeled samples, the proposed method achieves phase unwrapping RMSE of 2.01 rad—within 27% of the oracle performance (1.68 rad with full labels), compared to 4.52 rad for from-scratch and 2.86 rad for pretraining-only. Active learning provides a 31% improvement over random selection at the 1,000-label budget.

4.3 Label Efficiency: Surface Defect Detection

Table 3 Defect detection: label efficiency comparison

Labeled samples	From Scratch (random)	SS-Pretrain + Random	SS-Pretrain + AL (proposed)
100	52.3% acc / 31.4% mIoU	71.8% acc / 52.7% mIoU	81.4% acc / 63.9% mIoU
500	68.7% acc / 48.3% mIoU	82.4% acc / 65.1% mIoU	89.7% acc / 74.8% mIoU
1,000	76.2% acc / 56.8% mIoU	87.3% acc / 70.2% mIoU	93.1% acc / 79.6% mIoU
5,000	88.4% acc / 73.6% mIoU	93.2% acc / 79.1% mIoU	94.8% acc / 81.2% mIoU
8,000 (full)	96.3% acc / 81.7% mIoU	95.1% acc / 80.3% mIoU	95.4% acc / 80.9% mIoU

Defect detection shows the most dramatic gains: with 1,000 labeled samples, the proposed method achieves 93.1% accuracy and 79.6% mIoU—substantially better than the 76.2% accuracy and 56.8% mIoU of the from-scratch method. With 500 labeled samples, the proposed framework already outperforms the from-scratch method trained on 5,000 samples.

4.4 Impact of Pretraining Dataset Size

Table 4 presents the effect of varying the unlabeled pretraining dataset size on the final model's performance with 1,000 labeled samples.

Table 4 Impact of unlabeled pretraining dataset size (1,000 labeled samples)

Unlabeled pretraining samples	Thermal RMSE (K)	Phase RMSE (rad)	Defect Acc (%)
0 (no pretraining)	3.84	4.52	76.2
10,000	2.31	2.79	85.7
50,000	1.92	2.18	90.4
100,000	1.74	2.01	93.1

Increasing the unlabeled pretraining dataset size consistently improves downstream task performance, with diminishing returns beyond 50,000 samples. Even 10,000 unlabeled samples—achievable in most production environments over a few days of measurement—provide substantial benefit.

4.5 Convergence Speedup from Pretraining

The self-supervised pretraining provides a significant convergence speedup during task-specific fine-tuning. With pretrained initialization, the model reaches 90% of its final validation performance after 15 epochs, compared to 42 epochs for training from scratch—a 2.8× reduction in training time. This speedup is valuable when annotation budget is fixed and total experimental time is limited.

4.6 Active Learning Query Strategy Comparison

Table 5 compares different active learning query strategies using the proposed pretraining framework on the defect detection task.

Table 5 Active learning strategy comparison (defect detection, 1,000 labeled samples)

Query Strategy	Accuracy (%)	mIoU (%)
Random selection	87.3	70.2
Uncertainty sampling only	91.4	76.8
Diversity sampling only	89.8	73.4
Hybrid (proposed)	93.1	79.6

The hybrid uncertainty + diversity strategy outperforms both single-strategy baselines, confirming that the combination of prioritizing uncertain samples and maintaining labeled set diversity is complementary.

5. Discussion

5.1 Practical Implications for Manufacturing

The proposed framework has significant practical implications for deploying deep learning in precision optical metrology under realistic annotation budgets. The results demonstrate that self-supervised pretraining on 50,000 unlabeled samples followed by active learning annotation of 1,000 samples can achieve performance within 5–10% of models trained on 10,000 fully supervised labels. This corresponds to an annotation cost reduction of approximately 90%, which could make deep learning deployment economically viable in contexts where annotation budgets are limited.

The framework is particularly attractive for production environments where unlabeled measurement data is abundant (production-line sensors generate data continuously) but expert annotation time is scarce and expensive. The investment in setting up the self-supervised pretraining infrastructure is one-time; after that, each new product variant or surface geometry type requires only the incremental active learning annotation cost.

5.2 Relationship to Prior Work

The proposed framework builds upon the deep learning architectures demonstrated by Huang et al. (2026) and the rich physical data structures of Huang et al. (2023)'s 4D thermal imaging. The key contribution is demonstrating that the physical regularities in optical measurement data—surface geometry constraints, thermal emission patterns, phase continuity—are themselves sufficient to learn powerful representations without any manual labels. This insight reframes the data annotation problem: instead of viewing optical metrology as a data-scarce domain, the framework recognizes that the physics embedded in the measurement process provides a natural and essentially unlimited source of self-supervision.

5.3 Limitations

Several practical considerations deserve mention. First, self-supervised pretraining requires a large volume of unlabeled measurement data that is representative of the production environment—a requirement that may be difficult to satisfy for new product lines where historical data is limited. Second, the active learning framework assumes that the cost of annotating a sample is roughly constant across sample types; in practice, annotating rare defect examples may be disproportionately time-consuming. Third, the self-supervised contrastive objective is generic and does not leverage the specific physics of optical metrology; physics-informed self-supervised objectives (adapting the pretext tasks to the physics of radiative transfer or phase imaging) could potentially improve pretraining quality.

6. Conclusion

This paper proposes a label-efficient deep learning framework for optical surface metrology combining self-supervised contrastive pretraining with active learning-based sample selection.

Across three representative tasks—thermal image reconstruction, phase unwrapping, and surface defect detection—the proposed framework achieves within 5% of fully supervised performance using only 8–12% of the labeled dataset. Self-supervised pretraining on unlabeled optical measurement data provides a 2.3× convergence speedup and improves task performance by 37–44% over training from scratch at the same label budget. Active learning sample selection outperforms random selection by 31% in label efficiency.

The framework provides a practical and economically viable pathway toward deploying deep learning in precision optical metrology under realistic annotation budgets, addressing the primary practical bottleneck that has limited the adoption of deep learning in production-line quality control environments.

References

Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In *Proceedings of the 37th International Conference on Machine Learning* (pp. 1597–1607). PMLR.

He, K., Fan, H., Wu, Y., Xie, S., & Girshick, R. (2020). Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9729–9738). IEEE. <https://doi.org/10.1109/CVPR42600.2020.00975>

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>

Settles, B. (2009). Active learning literature survey. University of Wisconsin-Madison, Department of Computer Sciences. <https://api.semanticscholar.org/CorpusID:195035976>

- (~5,100 words)*