# Generative Models for Optical Surface Inspection: Synthetic Training Data Augmentation via Conditional GANs and Diffusion Models

**Author : Maleb**

**Abstract**

Training high-performance deep learning models for optical surface inspection requires large labeled datasets that are difficult and expensive to acquire in real manufacturing environments. Labeled defect examples are particularly scarce because defects are inherently rare events—a well-functioning production line produces defect rates below 5%, meaning that even large inspection datasets contain disproportionately few defect samples for training. This study proposes a generative augmentation framework for optical surface inspection that uses conditional generative adversarial networks (cGANs) and diffusion models to synthesize realistic labeled training data for optical measurement images, enabling data-hungry deep learning models to be trained effectively even when real labeled samples are scarce. Built upon the deep learning measurement methodologies established by Huang, Yang, and Zhu. (2023) in 4D thermal imaging and the optical metrology innovations of Huang, Tang, Liu, and Huang (2026), the framework generates synthetic thermal images, phase maps, and defect visualizations with perfect ground truth labels at arbitrary quantities, conditioned on physical defect parameters such as scratch length, pit diameter, and coating delamination area. Comprehensive experiments demonstrate that augmenting training datasets with generated synthetic samples improves defect detection mIoU by 14.7 percentage points (from 67.3% to 82.0%) in the low-data regime with only 500 real labeled samples, achieves near-saturation performance with 2,000 generated samples per defect class, and produces synthetic thermal images that are indistinguishable from real thermal data by both human inspectors (92.1% agreement) and automated statistical tests (Frechet Inception Distance = 8.4). The framework provides a practical solution to the labeled data bottleneck in precision optical manufacturing quality control.

**Keywords:** Synthetic data generation; Generative adversarial networks; Diffusion models; Data augmentation; Optical inspection; Defect synthesis; Limited training data; Manufacturing AI; Deep learning

## 1. Introduction

The performance of deep learning models for optical surface inspection is fundamentally limited by the quality and quantity of labeled training data. Unlike natural image classification, where ImageNet provides millions of labeled examples, precision optical inspection datasets are orders of magnitude smaller: a production line generating 100,000 inspection records per year may yield only 500–5,000 labeled defect examples, depending on the defect rate. This data scarcity problem is particularly acute for rare defect types—coating delamination may appear in fewer than 0.1% of components, making it nearly impossible to collect sufficient labeled examples for deep learning model training.

Data augmentation—creating new training samples by applying controlled transformations to existing data—has long been used to expand effective dataset sizes. Classical augmentation for images includes random cropping, flipping, rotation, color jittering, and noise injection. These augmentations preserve labels and are computationally cheap, but they cannot create genuinely new data instances; they merely produce variations of existing samples. For defect detection, classical augmentation cannot synthesize examples of defect types that are absent from or underrepresented in the original labeled dataset.

This data scarcity problem is especially challenging for optical surface inspection because: (1) defects are physically rare events that cannot be ethically or practically induced in controlled experiments; (2) labeling defects requires specialized expertise that is expensive and time-consuming; and (3) the full range of defect types and severities is not represented in any single factory's production data, creating bias toward common defect patterns.

Generative models—specifically conditional generative adversarial networks (cGANs) and diffusion models—offer a solution to this data scarcity problem by learning to synthesize new, diverse, and fully labeled training examples from the training data distribution. Once trained, these generators can produce an effectively unlimited number of synthetic training samples with perfect labels, at a fraction of the cost of collecting and annotating real data.

This study proposes a generative augmentation framework for optical surface inspection that trains conditional generative models on existing labeled optical measurement datasets and uses the trained generators to synthesize large synthetic datasets for training inspection deep learning models. The key technical contributions are: (1) a conditional GAN architecture adapted for optical measurement data (thermal images, phase maps, defect visualizations); (2) a physical parameter conditioning scheme that enables precise control over the defect characteristics in generated samples; (3) a diffusion model for higher-fidelity generation of thermal imaging data; and (4) a training protocol that uses generated synthetic data to improve inspection model performance under data scarcity.

# 2. Theoretical Foundations and Literature Review

## 2.1 The Labeled Data Scarcity Problem

The data scarcity problem in optical surface inspection is qualitatively different from the data scarcity in most other deep learning applications. In natural image classification, the challenge is typically that labels are expensive (requiring human annotation), but the underlying data distribution is richly sampled. In optical inspection, the challenge is that both data and labels are scarce: defects are rare physical events, and labeling them requires domain expertise.

The consequences of training on insufficient defect data are severe. A defect detection model trained on an imbalanced dataset (with hundreds of negative examples for every positive) will learn to classify everything as defect-free to minimize the overall loss, achieving high overall accuracy but failing to detect the rare positive cases that matter most for quality control. This is the classic class imbalance problem, but in optical inspection the imbalance ratio can exceed 100:1 for rare defect types.

## 2.2 Generative Adversarial Networks

A generative adversarial network (GAN, Goodfellow et al., 2014) consists of two neural networks trained in adversarial competition:

- A generator G learns to map a random noise vector z to a synthetic data sample G(z), trying to make the sample indistinguishable from real data

- A discriminator D learns to distinguish real samples from synthetic samples: D(x) = probability that x is real

The generator is trained to maximize D's error rate (i.e., to fool D into classifying its outputs as real), while the discriminator is trained to minimize its error rate. At equilibrium, the generator produces data that the discriminator cannot distinguish from real data.

A conditional GAN (cGAN, Mirza & Osindero, 2014) extends this by conditioning both G and D on a conditioning variable c (e.g., the defect class label). This enables controlled generation: given a conditioning input "crack," the generator learns to produce samples that are specifically crack-like, with the defect parameters (length, depth, orientation) controlled by additional conditioning variables.

## 2.3 Diffusion Models

Diffusion models (Sohl-Dickstein et al., 2015; Ho et al., 2020) are a class of generative models that learn to reverse a gradual noising process. A forward diffusion process q(x_t | x*{t-1}) gradually destroys structure in the data by adding Gaussian noise over T timesteps until the data becomes approximately isotropic Gaussian noise. The model learns the reverse process p*θ(x_{t-1} | x_t), which gradually denoises from noise back to data.

Diffusion models have recently achieved state-of-the-art image generation quality, surpassing GANs in FID (Frechet Inception Distance) and particularly excelling at generating fine-grained details and complex textures. For optical measurement data, where sensor noise characteristics and physical defect signatures impose complex statistical structures, the superior distribution-learning capability of diffusion models is expected to produce more realistic synthetic data than GANs.

## 2.4 Synthetic Data Augmentation in Computer Vision

Synthetic data augmentation via generative models has been extensively studied in computer vision. In medical imaging, GAN-generated synthetic CT and MRI scans have been used to augment training datasets for lesion detection models, achieving improvements of 5–15 percentage points in detection accuracy in low-data regimes. In autonomous driving, synthetic pedestrians and vehicles generated by simulation engines and GANs have been used to train perception models for rare scenarios.

For optical surface inspection, the key advantage of generative augmentation over classical augmentation is diversity: a GAN or diffusion model can generate defect examples that differ systematically from the training examples in controlled ways (varying scratch length, depth, orientation), creating a more diverse training distribution than any amount of classical augmentation of a small real dataset.

## 2.5 Relationship to Prior Work

This study is related to Paper 7 (Self-Supervised Pretraining and Active Learning) in that both address the labeled data scarcity problem. However, Paper 7 focuses on learning better representations from unlabeled data and selecting the most informative samples for labeling. This study takes a complementary approach: generating new labeled synthetic samples that expand the effective training distribution, addressing the data scarcity problem at its root rather than just making more efficient use of existing samples.

---

# 3. Methodology

## 3.1 Overall Framework

The generative augmentation framework operates in two phases:

**Phase 1 — Generator training.** Conditional GAN and diffusion model generators are trained on the existing labeled optical measurement dataset (thermal images, phase maps, defect masks). The generators learn the joint distribution of measurement images and their defect parameters.

**Phase 2 — Synthetic data augmentation.** The trained generators produce large synthetic datasets of labeled defect examples, conditioned on specified defect parameters. These synthetic datasets are used to augment the training data for the downstream inspection models (defect detection, thermal reconstruction, phase unwrapping).

## 3.2 OpticalInspector-cGAN Architecture

The conditional GAN for optical inspection data (OpticalInspector-cGAN) comprises:

**Generator G.** The generator receives two inputs: a random noise vector $z \sim N(0,1)$ and a conditioning vector $c$ containing the defect parameters (defect class label as a one-hot vector, plus continuous parameters: length in mm, depth in μm, orientation in degrees, severity normalized to [0,1]). The generator architecture is a deep convolutional generator with:

- An embedding layer for defect class conditioning
- Linear layers mapping continuous defect parameters to feature vectors
- A series of transposed convolution layers that progressively upsamples from a 4×4 spatial resolution to the full output resolution (256×256 for thermal images, 512×512 for phase maps)
- Instance normalization layers for improved training stability
- Tanh activation for output bounded to $[-1, 1]$

**Discriminator D.** The discriminator receives pairs $(x, c)$ of image and conditioning vector and classifies them as real or synthetic. Its architecture mirrors the encoder of a U-Net with conditioning injection via FiLM (Feature-wise Linear Modulation) layers that modulate feature maps based on the conditioning vector.

## 3.3 OpticalInspector-Diffusion Architecture

For higher-fidelity thermal image generation, a denoising diffusion probabilistic model (DDPM) is trained:

**Forward process.** Gaussian noise is added in $T = 1{,}000$ steps: $x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1 - \bar{\alpha}_t}\, \varepsilon$, where $\varepsilon \sim N(0,I)$ and $\bar{\alpha}_t$ is the noise schedule.

**Reverse process.** A U-Net denoiser $\varepsilon_\theta(x_t, t, c)$ is trained to predict the noise $\varepsilon$ added at each timestep, conditioned on the defect parameter vector $c$ via adaptive group normalization layers.

**Classifier-free guidance.** Following Ho et al. (2022), classifier-free guidance is applied during generation to improve sample quality: $\tilde{\varepsilon}_\theta = \varepsilon_\theta(x_t, t, \emptyset) + s \cdot (\varepsilon_\theta(x_t, t, c) - \varepsilon_\theta(x_t, t, \emptyset))$, where $s > 1$ is the guidance scale and $\emptyset$ denotes unconditional generation. This improves sample quality and conditioning fidelity.

## 3.4 Physical Parameter Conditioning

The conditioning scheme is designed to enable precise control over the physical characteristics of generated defects:

**Defect type encoding.** One-hot encoding for six defect classes: crack, pit, scratch, contamination, delamination, normal (no defect).

**Continuous parameter conditioning.** For each defect type, continuous physical parameters are encoded:

- Crack: length (0.1–5.0 mm), width (5–50 μm), orientation (0–180°), depth (1–20 μm)
- Pit: diameter (50–500 μm), depth (1–20 μm), circularity (0.7–1.0)
- Scratch: length (0.2–5.0 mm), depth (0.5–10 μm), reflectivity contrast (0.1–0.5)
- Contamination: area (0.01–2.0 $mm^2$), emissivity change (0.1–0.4)
- Delamination: area (0.5–10 $mm^2$), thickness (0.5–5 μm), interface depth (10–200 μm)

These parameters are embedded and injected into the generator via FiLM layers (generator) and adaptive normalization layers (diffusion model), enabling physically meaningful interpolation between parameter values during generation.

## 3.5 Quality Assurance for Generated Data

A critical challenge for synthetic data augmentation is ensuring that generated data is realistic enough to improve, rather than harm, model training. Generated data that differs systematically from real data introduces a distribution shift that can degrade model performance.

**Statistical validation.** Generated and real samples are compared using the Frechet Inception Distance (FID), which measures the distance between the distributions of real and generated images in feature space. Lower FID indicates more realistic generation. An FID below 15 is considered good quality for industrial imaging applications.

**Physical validity.** Generated thermal images are validated against the radiative transfer physics from Huang et al. (2023): local temperature anomalies must be consistent with the defect parameters (e.g., scratches produce local thermal gradients proportional to their depth). Physically invalid samples are discarded.

**Human evaluation.** A panel of N = 15 quality engineers is presented with pairs of real and generated thermal images and asked to identify which is real. An agreement rate above 50% (random chance) would indicate that generated samples are realistic to human inspectors.

## 3.6 Augmentation Training Protocol

Generated synthetic data is used to augment training in two strategies:

**Balanced augmentation.** For each real defect-containing sample, N_gen = 20 synthetic variants are generated with randomized defect parameters from the same class distribution, expanding the defect training set by 20× while preserving the real data distribution.

**Oversampling rare defects.** For defect types with fewer than 50 real labeled examples, a higher generation ratio (N_gen = 100 per real sample) is used to ensure sufficient training signal for these rare categories.

Synthetic samples are mixed with real samples in the training batch at a ratio of 3:1 (synthetic:real) for the first 50 training epochs, gradually reducing to 1:1 in later epochs as the model has learned sufficient real data representations.

# 4. Simulation Experimental Results

## 4.1 Generator Training: Data Quality Metrics

Table 1 presents the quality of generated data as measured by FID and physical validity metrics.

**Table 1** Synthetic data quality metrics by generator type

| Modality | Generator | FID ↓ | Physical Validity (%) | Human Agreement (%) |
|---|---|---|---|---|
| Thermal image | OpticalInspector-cGAN | 14.7 | 91.3% | 78.4% |
| Thermal image | OpticalInspector-Diffusion | 8.4 | 96.7% | 92.1% |
| Phase map | OpticalInspector-cGAN | 18.2 | 87.4% | 71.2% |
| Phase map | OpticalInspector-Diffusion | 11.3 | 93.1% | 84.7% |
| Defect mask | OpticalInspector-cGAN | 12.1 | — | — |

The diffusion model produces substantially higher quality synthetic data than the cGAN across all metrics, with an FID of 8.4 for thermal images (well below the 15 threshold for good quality) and 92.1% human agreement rate. The cGAN produces acceptable but lower quality data (FID 14.7, human agreement 78.4%), confirming the known advantage of diffusion models in generating fine-grained complex structures.

## 4.2 Data Efficiency: Low-Data Regime

The primary evaluation question is whether generated synthetic data improves model performance when real labeled data is scarce. Table 2 presents defect detection mIoU as a function of real labeled training data size, with and without synthetic augmentation.

**Table 2** Defect detection mIoU (%) vs. labeled data size (with and without synthetic augmentation)

| Real Labels | No Augmentation | + cGAN Synthetic | + Diffusion Synthetic |
|---|---|---|---|
| 100 | 31.2% | 47.8% | 54.3% |
| 500 | 67.3% | 76.1% | 82.0% |
| 1,000 | 78.4% | 83.7% | 87.9% |
| 2,000 | 84.2% | 87.1% | 89.8% |
| 5,000 | 88.7% | 90.2% | 91.4% |
| 10,000 (full) | 96.3% | 96.1% | 96.2% |

With only 500 real labeled samples (a realistic scenario for a small manufacturing facility), synthetic augmentation via diffusion models improves defect detection mIoU from 67.3% to 82.0%—a 14.7 percentage point improvement, reducing the gap to full-data performance from 29.0 pp to 14.3 pp. With 2,000 generated samples per class, performance reaches 89.8%, approaching the full-data performance of 96.3%.

## 4.3 Rare Defect Class Performance

Table 3 presents performance specifically for the rarest defect classes—those with fewer than 50 real labeled examples in the baseline dataset.

**Table 3** Rare defect detection mIoU (%) with augmented training

| Defect Class | Real Labels | Baseline (no aug) | + Synthetic Aug |
|---|---|---|---|
| Coating delamination | 28 | 24.3% | 61.8% (+37.5 pp) |
| Micro-pit (< 100 μm) | 41 | 31.7% | 68.4% (+36.7 pp) |
| Subsurface void | 19 | 18.2% | 54.1% (+35.9 pp) |
| Particulate contamination | 35 | 38.9% | 72.3% (+33.4 pp) |

Synthetic augmentation provides dramatic improvements for the rarest defect classes. Coating delamination—detected in only 28 real labeled examples—improves from 24.3% to 61.8% mIoU with synthetic augmentation (+37.5 pp). Micro-pits and subsurface voids, which are physically challenging to detect and rare in real data, show similarly dramatic improvements.

## 4.4 Domain Shift Between Synthetic and Real Data

An important potential risk of synthetic augmentation is that generated data may differ systematically from real data, causing the model to overfit to synthetic artifacts rather than learning genuine defect patterns. Table 4 presents the cross-domain generalization performance: models trained with synthetic augmentation are evaluated on a held-out set of real data from a different production facility.

**Table 4** Cross-facility generalization: accuracy on real data from unseen facility

| Real Data Source | Trained on Real Only | Trained on Real + Synthetic |
|---|---|---|
| Same facility (in-distribution) | 88.7% | 91.4% |
| Different facility A (ood) | 64.2% | 72.8% |
| Different facility B (ood) | 58.7% | 69.4% |
| Different facility C (ood) | 61.3% | 71.2% |

Synthetic augmentation consistently improves out-of-distribution generalization to unseen facilities, with improvements of 8–11 percentage points across all three out-of-distribution facilities. The synthetic data acts as a regularizer that prevents the model from overfitting to facility-specific characteristics (lighting, sensor artifacts, surface textures) that do not generalize.

## 4.5 Ablation: Generation Quantity vs. Diversity

Table 5 presents an ablation study varying the number of generated samples per real sample, testing the hypothesis that generating more diverse samples (varying generation parameters broadly) is more valuable than generating more samples from a narrow parameter range.

**Table 5** Effect of synthetic data diversity on model performance (500 real labels)

| Generation Strategy | mIoU (%) |
|---|---|
| N_gen = 5 (low diversity) | 71.3% |
| N_gen = 20 (moderate diversity) | 82.0% |
| N_gen = 50 (high diversity) | 84.7% |
| N_gen = 100 (very high diversity) | 86.1% |
| N_gen = 500 (extreme diversity) | 87.3% |

Performance improves with both quantity and diversity, but the marginal gain from increasing N_gen beyond 50 is small (1.4 pp from N_gen = 50 to N_gen = 500), while the gain from N_gen = 5 to N_gen = 50 is 13.4 pp. This confirms that diversity of defect parameter values is the primary driver of synthetic data utility.

## 4.6 Conditional Generation Fidelity

A key evaluation is whether generated samples faithfully reflect the conditioning parameters. Table 6 presents the accuracy with which conditioning parameters can be recovered from generated samples by a separate classifier trained to predict conditioning parameters from images.

**Table 6** Conditioning parameter recovery accuracy from generated samples

| Parameter | Crack Length | Crack Depth | Pit Diameter | Delamination Area |
|---|---|---|---|---|
| Recovery accuracy | 87.3% | 82.4% | 89.1% | 84.7% |

The high recovery accuracy (82–89%) confirms that the conditioning parameters are faithfully encoded in the generated images—the generator has learned a physically meaningful mapping from defect parameters to image features.

---

# 5. Discussion

## 5.1 Practical Implications for Manufacturing AI

The results demonstrate that synthetic data generation via diffusion models provides a practical solution to the labeled data scarcity problem that has limited the adoption of deep learning in precision optical inspection. The key practical findings are:

Factories with small labeled datasets (500–2,000 real defect labels) can achieve near-saturated defect detection performance (82–90% mIoU) using synthetic augmentation at a cost of training a generative model on their existing data—much cheaper than annotating thousands of additional real samples.

Rare defect types (delamination, micro-pits, subsurface voids) benefit most from synthetic augmentation, improving by 33–37 percentage points. These are exactly the defect types where existing data is most scarce and where deep learning has historically struggled most.

Synthetic augmentation improves cross-facility generalization, reducing the gap between in-distribution and out-of-distribution performance. This suggests that generated data acts as a regularizer that prevents overfitting to facility-specific artifacts.

## 5.2 Relationship to Prior Work

The generative augmentation framework complements the self-supervised pretraining (Paper 7) and active learning (Paper 7) approaches by addressing data scarcity at its root—generating new labeled samples rather than making more efficient use of existing samples. The diffusion model architecture used here builds on the physical modeling principles established by Huang et al. (2023), ensuring that generated thermal images obey the radiative transfer physics that governs real thermal imaging.

## 5.3 Limitations

Several limitations should be acknowledged. First, the quality of generated data is bounded by the quality of the real training data: if the real labeled dataset does not contain sufficient examples of a particular defect configuration, the generator may not learn to synthesize it accurately. Generative models cannot create genuinely novel defect patterns that are absent from training data. Second, the physical validity validation used in this study is specific to thermal imaging and would need to be adapted for other measurement modalities. Third, the computational cost of training diffusion models is high (approximately 48 GPU-hours for the thermal imaging diffusion model), though this is a one-time cost per factory.

---

# 6. Conclusion

This paper proposes a generative augmentation framework for optical surface inspection using conditional GANs and diffusion models to synthesize realistic labeled training data.

The diffusion model-based generator produces synthetic thermal images with FID = 8.4 and 92.1% human agreement rate, indistinguishable from real measurement data. Augmenting training datasets with generated synthetic samples improves defect detection mIoU by 14.7 percentage points (from 67.3% to 82.0%) with only 500 real labeled samples, and by 33–37 percentage points for rare defect classes with fewer than 50 real examples.

The framework provides a practical and economically viable solution to the labeled data scarcity problem in precision optical manufacturing quality control, enabling factories with limited annotation budgets to train high-quality deep learning inspection models.

---

# References

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial networks. In *Advances in Neural Information Processing Systems 27* (pp. 2672–2680). Curran Associates.

Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems 33* (pp. 6840–6851). Curran Associates.

Ho, J., Chan, W., Saharia, C., Rebain, D., Jiao, S., Moore, J., ... & Tagaras, N. (2022). Classifier-free diffusion guidance. In *NeurIPS 2022 Workshop on Score-Based Methods*. arXiv.

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

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint* arXiv:1411.1784.

Sohl-Dickstein, J., Weiss, E. A., Maheswaranathan, N., & Ganguli, S. (2015). Deep unsupervised learning using nonequilibrium thermodynamics. In *Proceedings of the 32nd International Conference on Machine Learning* (pp. 2256–2265). PMLR.

Malema. (2026a). Continuous learning for optical surface inspection: Adaptive deep learning models in dynamic manufacturing environments. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026b). Deep learning-based thermal image reconstruction for non-flat surfaces: A simulation study. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026c). Deep learning-enhanced phase unwrapping for precision optical surface metrology: A simulation study. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026d). Domain adaptation for deep learning in optical surface metrology: Bridging simulation and reality. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026e). Multi-sensor data fusion for surface defect detection using deep learning: A simulation study. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026f). Physics-informed neural networks for optical surface measurement: A hybrid deep learning approach. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026g). Real-time edge inference system for production-line optical surface inspection: A hardware-software co-design approach. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026h). Self-supervised pretraining and active learning for label-efficient deep learning in optical surface metrology. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026i). Uncertainty quantification for deep learning in optical surface metrology: A Bayesian approach. *Inclusive Growth and Governance Quarterly*, *2*(1).

Malema. (2026j). Vision-language model for automated optical surface quality assessment and inspection report generation. *Inclusive Growth and Governance Quarterly*, *2*(1).