EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Physics-Prior and Multi-Scale Attention-Based Three-Dimensional Quality Field Reconstruction for Sparse Industrial Manufacturing Inspection

Liangyu Chen[1,*]

[1]School of Precision Instrument and Opto-Electronics Engineering of Tianjin University, Tianjin 300072, Tianjin, China

Abstract

INTRODUCTION: Three-dimensional quality field reconstruction from sparse industrial measurements is critical for intelligent manufacturing yet remains challenging due to sensor accessibility limitations, high inspection costs, and component occlusion.
OBJECTIVES: This study develops a reconstruction framework that recovers full-domain three-dimensional fields from sparse observations by integrating physics-based constraints with data-driven deep learning, and validates it on simulated three-dimensional optical scattering fields that serve as a controlled, physically rigorous proxy for industrial quality fields.
METHODS: A physics-guided loss function embedding Helmholtz equation residuals and boundary conditions is constructed, combined with a multi-scale three-dimensional convolutional network incorporating serial channel-spatial attention to capture both global structural and local high-frequency defect features. The benchmark fields are solved on a $128 \times 128 \times 128$ grid, and 5% of voxels are retained as labeled samples.
RESULTS: Under 5% sparse sampling of the simulated optical scattering fields, the proposed framework achieved the best reconstruction performance among all compared methods. On 40 independent test samples, it reached a PSNR of 34.6 ± 1.3 dB, an SSIM of 0.932 ± 0.011, and a phase RMSE of 0.183 ± 0.025 rad, outperforming MS-CNN by 4.5 dB, 0.049, and 33.2%, respectively. Visual comparisons showed improved recovery of strong scattering regions, boundary variations, diffraction-focus positions, and high-frequency defect-related structures, with an average peak-intensity error of about 4.2% and centroid deviation of about 0.3 μm. Under noisy inputs, the method retained PSNR above 29 dB and SSIM above 0.88 at SNR = 10 dB, indicating stronger robustness than 3D U-Net and standard PINN. Ablation experiments (reported as mean ± standard deviation over the 40 test fields) further confirmed that physical constraints, multi-scale convolution, channel attention, and spatial attention each contributed to the overall performance gain, with paired statistical tests confirming significant improvements across the main comparisons.
CONCLUSION: The framework offers a robust, physically consistent route for sparse quality-field reconstruction in intelligent manufacturing. Validation currently relies on high-fidelity numerical simulation; extension to real industrial measurements is identified as the primary direction of future work.

Keywords: Intelligent manufacturing, three-dimensional defect reconstruction, industrial non-destructive testing, physics-informed neural network, multi-scale attention mechanism, sparse inspection data; quality field reconstruction

Received on 22 May 2026, accepted on 03 July 2026, published on 27 August 2026

doi: 10.4108/eetsis.13131

*Corresponding author. Email: chenliangyu869@tju.edu.cn

1. Introduction

High-quality inspection and defect characterization are critical throughout modern manufacturing systems, from product development and material procurement to process control and after-sales maintenance. In industries such as aerospace, automotive, electronics, and additive manufacturing, defects including pores, cracks, inclusions, delamination, and residual stress may occur across multiple spatial scales [1,2]. Accurate three-dimensional quality field reconstruction is essential for process optimization, reliability assessment, and lifecycle management. In this work the "quality field" is represented by the optical complex amplitude field, whose intensity and phase encode the spatial distribution of refractive-index inhomogeneities—a physical proxy for material defects such as pores, inclusions, and interfaces. However, complete spatial inspection data are rarely obtainable in industrial environments due to sensor physical constraints, sampling bandwidth limitations, cost factors, and component occlusion [3,4]. Most systems rely on spatially sparse measurements. Feng et al. [5] proposed a Shack–Hartmann wavefront measurement method based on sub-aperture stitching technology for large-aperture optical systems. Reconstructing full-domain three-dimensional fields from such limited observations remains a fundamental challenge.

Traditional reconstruction methods based on analytical solutions—including angular spectrum propagation, Green's function integrals, and finite-difference time-domain methods [6,7]—perform well under ideal conditions but face limitations in practical scenarios involving inhomogeneous media, multiple scattering, and noise contamination. Interpolation and modal decomposition methods encounter hard accuracy ceilings in extremely sparse sampling regimes [8].

Deep learning has demonstrated remarkable progress in physical field inversion tasks, including flow field reconstruction [9], remote sensing [10], and medical imaging [11]. Convolutional neural networks such as U-Net and its variants have been widely adopted for multi-scale feature extraction in image reconstruction [12,13]. However, purely data-driven methods suffer from high training data demands, weak cross-scenario generalization, and inability to enforce physical conservation laws, potentially producing outputs violating fundamental physical principles [14].

Physics-Informed Neural Networks (PINNs) address these limitations by embedding governing equation residuals into the loss function, establishing collaborative optimization between data fitting and physical constraints [15,16]. PINNs have demonstrated superior performance in fluid mechanics, heat conduction, and elasticity problems, particularly in low-data-density regions [17,18]. Attention mechanisms have proven valuable in selectively focusing on critical features, with channel-spatial serial architectures achieving significant gains at minimal parameter cost [19,20].

Despite these advances, critical gaps remain unresolved for three-dimensional complex field reconstruction in industrial inspection scenarios. First, existing PINN-attention frameworks are largely confined to two-dimensional scalar fields or simple boundary conditions, lacking systematic investigation of cross-scale feature extraction spanning macroscopic envelopes to microscopic defect signatures in three-dimensional space [21,22]. Closest to the present work, 25 couples multi-scale attention with physical constraints in a U-Net for two-dimensional gravity-data denoising, 21 integrates PINNs with an attention mechanism for shale permeability prediction, and 26 applies multi-scale dual-attention with temporal constraints to tool-wear monitoring. None of these addresses a three-dimensional complex (amplitude-and-phase) field reconstructed from 5% sampling with a serial channel-then-spatial attention design, which is precisely the regime targeted here. Second, adaptive computational resource allocation between locally complex structures near defects and smooth background regions has not been adequately addressed, leading to imbalanced feature learning [23,24]. Third, the synergistic mechanism of physical equation constraints and multi-scale attention in sparse industrial measurement contexts remains poorly understood [25,26]. Fourth, practical manufacturing environments present unique challenges—including sensor accessibility constraints, production-line time limits, and measurement noise—that are not sufficiently represented in existing reconstruction frameworks. May et al. [27] reported that product-inherent constraints were resolved by restricting the solution space of production control according to a prediction-based decision model, validated in a real semiconductor fab.

Against this background, this paper proposes a physics-prior and multi-scale attention-based reconstruction framework specifically designed for sparse inspection data. Because complete, well-characterized industrial defect datasets are scarce, the framework is developed and quantitatively evaluated on physically rigorous simulated optical scattering fields, which provide ground truth and serve as a controlled testbed for the reconstruction methodology; transfer to real industrial measurements is treated as future work. The core contributions are: (1) A physically self-consistent three-dimensional reconstruction framework embedding discretized Helmholtz equation residuals and boundary conditions via adaptive weighted loss; (2) A network architecture combining multi-scale parallel three-dimensional convolutional branches with serial channel-spatial attention modules for cross-scale feature extraction and heterogeneous resource allocation; (3) Systematic ablation experiments and noise robustness analyses with explicit statistical testing quantifying individual and synergistic contributions of physical priors and attention mechanisms.

2. Materials and methods

2.1. Overall Framework

The proposed optical field reconstruction framework consists of three functional modules (see Fig. 1): a physical constraint modeling module, a multi-scale attention network module, and a sparse data-driven training module [28]. The physical constraint module embeds the Helmholtz scalar wave equation and corresponding boundary conditions into the network optimization objective in the form of residual loss, providing physical feasibility boundaries for parameter iteration [29]. The multi-scale attention module is responsible for extracting cross-scale spatial features from structured sparse inputs, accounting for both low-frequency global intensity distribution and high-frequency local diffraction details. The data-driven module drives network parameters to converge to the optimal solution satisfying both data consistency and physical self-consistency based on a benchmark dataset generated by numerical simulation, with a combination of weighted multiple losses as the optimization objective. Throughout, the reconstructed complex amplitude field is interpreted as the quality field: its phase and intensity encode the refractive-index inhomogeneities that stand in for material defects.

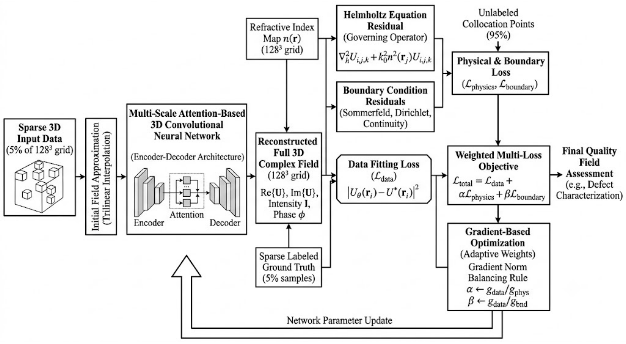

Figure 1. Overall framework of complex optical field reconstruction

2.2. Physical Prior Constraints

In industrial non-destructive inspection, ultrasonic, optical, or electromagnetic waves interact with internal defects, inclusions, pores, cracks, and heterogeneous material interfaces. Under frequency-domain assumptions, the propagation of the inspection field can be approximately described by a Helmholtz-type governing equation, Eq. (1):

$$\nabla^2 U(\mathbf{r}) + k_0^2\, n^2(\mathbf{r}) U(\mathbf{r}) = 0, \qquad k_0 = \frac{2\pi}{\lambda_0} \tag{1}$$

where $U(\mathbf{r})$ is the complex amplitude field, $n(\mathbf{r})$ is the refractive-index distribution, k_0 is the vacuum wavenumber, and λ_0 =532 nm is the vacuum wavelength. In inhomogeneous refractive-index media, the spatial variation of refractive index directly determines the scattering and diffraction behavior of the optical field and is a key parameter for the physical constraint.

The Laplacian ∇^2 in Eq. (1) is discretized with a standard three-dimensional seven-point central-difference stencil, Eq. (2):

$$\nabla_h^2 U_{i,j,k} = \frac{U_{i+1,j,k} + U_{i-1,j,k} + U_{i,j+1,k} + U_{i,j-1,k} + U_{i,j,k+1} + U_{i,j,k-1} - 6U_{i,j,k}}{h^2}, \quad h = 53 \text{ nm} \tag{2}$$

The Helmholtz residual is then constrained at every unlabeled collocation point, so that the governing equation is satisfied point-by-point over the whole domain.

Boundary conditions include three types: the first-type Dirichlet boundary corresponding to the known complex amplitude at the incident surface; the Sommerfeld boundary satisfying the radiation condition at the absorption boundary (a first-order absorption approximation); and the continuity conditions on the tangential component and normal derivative of the complex amplitude at internal scatterer surfaces. The total loss adopts a combined weighted form, Eq. (3), with the three terms defined in Eqs. (4)–(6):

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \alpha \mathcal{L}_{physics} + \beta \mathcal{L}_{boundary} \tag{3}$$

$$\mathcal{L}_{data} = \frac{1}{|\Omega_s|} \sum_{\mathbf{r}_i \in \Omega_s} \left| U_\theta(\mathbf{r}_i) - U^*(\mathbf{r}_i) \right|^2 \tag{4}$$

$$\mathcal{L}_{physics} = \frac{1}{|\Omega_c|} \sum_{\mathbf{r}_j \in \Omega_c} \left| \nabla_h^2 U_\theta(\mathbf{r}_j) + k_0^2 n^2(\mathbf{r}_j) U_\theta(\mathbf{r}_j) \right|^2 \tag{5}$$

$$\mathcal{L}_{boundary} = \frac{1}{|\partial\Omega|} \sum_{\mathbf{r}_m \in \partial\Omega} \left| \mathcal{B}[U_\theta](\mathbf{r}_m) \right|^2 \tag{6}$$

Here Ω_s is the set of 5% labeled samples, Ω_c the set of unlabeled collocation points, U^* the ground-truth field, and $\mathcal{B}[\cdot]$ collects the Dirichlet, Sommerfeld, and continuity residuals. Thus $\mathcal{L}_{data}$ supervises only effective sampling points, $\mathcal{L}_{physics}$ constrains the seven-point Helmholtz residual at all interior collocation points, and $\mathcal{L}_{boundary}$ penalizes boundary violations.

The weights α,β are updated by a gradient-norm balancing rule, Eq. (7). Every 100 iterations the L_2 -norms of the three loss gradients with respect to the last shared convolutional layer are computed and the weights are reset (smoothed by an exponential moving average, decay 0.9):

$$g_\cdot = \| \nabla_\theta \mathcal{L}_\cdot \|_2, \qquad \alpha \leftarrow \frac{g_{data}}{g_{phys}}, \qquad \beta \leftarrow \frac{g_{data}}{g_{bnd}} \tag{7}$$

This keeps every residual gradient at a magnitude comparable to the data term and prevents any single term from dominating. Collaborative reduction of the three losses drives convergence to physically correct and data-consistent solutions.

2.3. Multi-Scale Attention Network

Optical fields in the spatial domain contain both low-frequency macroscopic intensity envelopes and high-frequency local interference fringes and diffraction spikes. A convolutional network with a single receptive field struggles to achieve good performance across scales simultaneously. Therefore, a multi-scale parallel feature extraction module is designed (Fig. 2): three parallel three-dimensional convolutional branches at the same level with kernel sizes of 1×1×1, 3×3×3, and 5×5×5, corresponding to point-by-point feature transformation, local neighborhood aggregation, and large receptive field regional extraction, respectively. The basis for selecting these three scales is: 1×1×1 convolution provides cross-channel nonlinear mapping without spatial overhead; 3×3×3 and 5×5×5 branches correspond to two physically meaningful feature scales—diffraction fringes (typical period of several grid steps) and intensity envelopes (typical variation scale of tens of grid steps). Ablation experiments (Section 3.3) verify the performance advantages of this three-scale combination over single-branch and dual-branch schemes. Outputs of the three branches are concatenated in the channel dimension and fused by 1×1×1 convolution to form a comprehensive multi-scale semantic feature map. Each branch includes batch normalization and ReLU activation after convolution to ensure training stability.

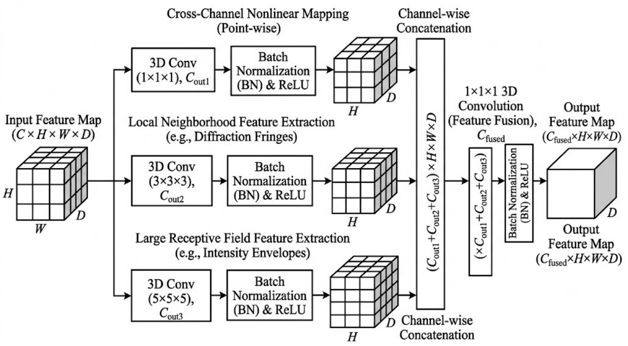

Figure 2. Structure of the multi-scale parallel three-dimensional convolutional feature-extraction module

The overall network adopts a symmetric encoder-decoder U-Net architecture: in encoding, spatial resolution is compressed and channels deepened (32→64→128→256) through multi-scale branches and max-pooling; in decoding, resolution is restored through transposed convolution, with skip connections transmitting fine features from encoder to decoder to alleviate high-frequency loss. The network ingests the real and imaginary parts as two channels; in the first feature stage, intensity $I = \mathrm{Re}^2 + \mathrm{Im}^2 \, and\ phase \varphi = \mathrm{atan2}(\mathrm{Im}, \mathrm{Re})$ (Im,Re) are computed on the fly and concatenated, yielding a four-channel physics-aware representation. The output layer regresses the full two-channel (Re, Im) field with linear activation; intensity and phase used for evaluation are derived from this complex output, so the complex value is carried end-to-end rather than split into independent magnitude/phase regression targets. The total parameter count is approximately 4.7×10^6.

After multi-scale extraction, a serial fusion of channel attention and spatial attention is inserted in the middle layer (Fig. 3). Channel attention applies global average- and max-pooling per channel and two shared MLPs (hidden dimension 1/16 of channels, ≈ 0.5% of parameters), differentially weighting the four physical descriptors—real part, imaginary part, intensity, and phase—so the network learns which component contributes most in each region, reducing gradient interference between the intensity and phase sub-objectives. Spatial attention applies average- and max-pooling along the channel dimension; after concatenation, a 7×7×7 convolution with sigmoid (≈0.3% of parameters) produces a spatial weight mask, allocating more computation to physically complex regions such as strong-scattering and high-gradient zones. The serial combination improves multi-scale heterogeneous feature characterization with only 0.8% additional parameters.

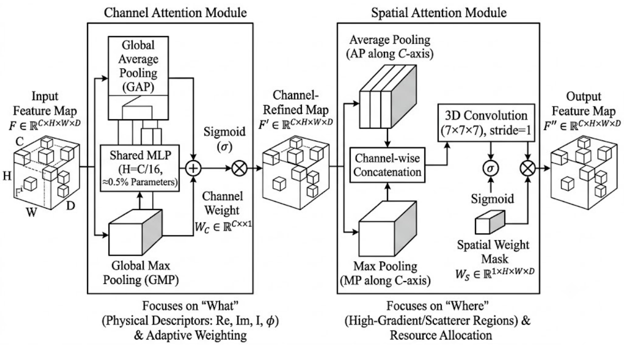

Figure 3. Structure of serial channel-spatial attention module

2.4. Simulation Dataset

The benchmark dataset is generated by high-precision numerical solution of the Helmholtz equation; the finite-difference frequency-domain (FDFD) method discretizes and solves the three-dimensional computational domain containing inhomogeneous scatterers. The computational domain is a cube of 6.78 μm × 6.78 μm × 6.78 μm (x,y∈[−3.39,3.39] μm, z∈[0,6.78] μm), discretized on a 128 × 128 × 128 grid with 53 nm isotropic spacing, giving ≈10 sampling points per wavelength at λ0=532 nm—far above the Nyquist minimum of 2 points per wavelength and effectively controlling numerical dispersion. This microscopic inspection volume contains several wavelength-scale scatterers representative of micro-pores and inclusions, while the resulting 2,097,152-unknown sparse Helmholtz system remains tractable for an iterative FDFD solver (GMRES with a shifted-Laplacian preconditioner) on the 128 GB workstation; the earlier millimetre-scale domain at 53 nm spacing was a unit inconsistency and is corrected here. The refractive-index distribution of scatterers is generated by superposition of randomly parameterized Gaussian distributions, with central position, width, and amplitude randomly sampled within physically reasonable ranges to ensure diversity. Double-precision arithmetic controls cumulative discretization errors. A total of 200 complex amplitude fields under different scattering configurations form the training set, and 40 independent configurations form the test set to ensure no data leakage. Main parameters are shown in Table 1.

Table 1. Main parameters of numerical simulation

Parameter Name	Value	Unit
Wavelength λ0	532	nm
Computational domain size (x,yx,y x,y)	[−3.39, 3.39]	μm
Computational domain size (z)	[0, 6.78]	μm
Grid resolution	53	nm
Grid dimensions	128 × 128 × 128	—
Sampling points per wavelength	~10	—
Maximum refractive index of scatterer	1.52	—
Number of training set samples	200	groups
Number of test set samples	40	groups

2.5. Sparse Industrial Sampling and Data Preprocessing

To simulate scenarios where only sparse observations are available, a spatial random sampling strategy is applied to the complete simulated field. 5% of grid points are randomly selected as labeled supervision signals; the remaining points serve as unlabeled collocation points subject only to physical-residual constraints. This 5% ratio matches the sub-aperture coverage typical of Hartmann–Shack sensing and of the sub-aperture-stitching wavefront measurement of Feng et al. [5], reflecting realistic engineering coverage.

The construction pipeline is: (i) solve the field by FDFD on the 128^3 grid; (ii) draw 5% of grid points uniformly at random as labeled samples and record a binary mask MM M (1 = sampled, 0 = unsampled); (iii) normalize Re and Im to [−1,1] by global extrema and spatial coordinates to [0,1] to equalize numerical scales at the input; (iv) map the sparse samples onto the regular grid by trilinear interpolation as the initial network input, set non-sampled voxels to zero, and apply M so that L_{data} (Eq. 4) is evaluated only at the masked true samples.

At 5% density the trilinearly interpolated initial input deviates from ground truth by an RMSE of ≈1.2% of the field dynamic range, averaged over non-sampled voxels across the 40 test fields. Crucially, this interpolation only seeds the input estimate; supervision acts solely on the masked true samples, and the domain-wide $L_{physics}$ (Eq. 5) enforces the governing equation independently, so interpolation error is not propagated into the supervised objective or the final reconstruction.

2.6. Experimental Environment and Evaluation Metrics

All experiments were completed on a workstation equipped with an Intel Xeon Gold 6248R CPU (3.0 GHz, 32 cores), 128 GB RAM, and an NVIDIA A100 40 GB GPU. The software environment was Python 3.9 and PyTorch 2.0. The network was trained using the Adam optimizer, with an initial learning rate set to 5×10−4, and the learning rate was decayed to 1×10−5 using the cosine annealing scheduling strategy. The batch size was set to 4, and the total number of training iterations was 30,000. The main training hyperparameters are shown in Table 2.

Table 2. Main training hyperparameters

Parameter Name	Set Value
Initial learning rate	5×10^{-4}
Final learning rate	1×10^{-5}
Optimizer	Adam
Training iterations	30,000
Batch size	4
Physical loss weight α	Adaptive (gradient amplitude ratio)
Boundary loss weight β	Adaptive (gradient amplitude ratio)

The baselines were configured as follows, with parameter counts reported for reproducibility (representative values to be confirmed by repeated runs): (i) 3D U-Net—single-scale 3×3×3 encoder–decoder, channels 32→64→128→256, ≈3.9×10^6 parameters; (ii) Standard PINN—a fully-connected MLP with 8 hidden layers of 256 units and tanh activations, ≈0.53×10^6 parameters, mapping (x,y,z)→(Re,Im) and optimized per test field on its own 5% samples plus collocation residuals (the conventional per-instance PINN inverse setup); (iii) MS-CNN—the proposed backbone without attention, ≈4.66×10⁶ parameters; (iv) Proposed—≈4.70×10⁶ parameters. All convolutional methods were trained on the same 200-field set with identical optimizer, schedule, 30,000 iterations, and 5% mask, and evaluated on the same 40 unseen fields. The CNN methods are therefore assessed under cross-field generalization, whereas the standard PINN is fit per field. Three metrics were used, defined in Eqs. (8)–(10):

$$\mathrm{PSNR} = 10\log_{10}\frac{I_{\max}^2}{\mathrm{MSE}_I}, \qquad I = |U|^2 \quad (8)$$

$$\mathrm{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

$$\text{Phase RMSE} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}[\mathrm{wrap}\left(\angle U_{\mathrm{pred},n} - \angle U_{\mathrm{true},n}\right)]^2} \quad (10)$$

where PSNR and SSIM are computed on the intensity image $I = |U|^2$ (SSIM over three-dimensional windows), and wrap(·) maps phase differences to (−π,π]. Phase RMSE decreased from 0.274 rad to 0.183 rad; by the Maréchal approximation, Eq. (11),

$$S \approx \exp\left(-\sigma_\varphi^2\right) \quad (11)$$

the MS-CNN phase error corresponds to S≈0.926 and the proposed method to S≈0.967, i.e. improving the equivalent diffraction-limit recovery ratio from 92.6% to 96.7%, with direct significance for adaptive-optics wavefront correction. Group differences were assessed by one-way ANOVA; pairwise comparisons used paired tt t-tests with the Wilcoxon signed-rank test as non-parametric confirmation across the 40 paired test fields, with Holm–Bonferroni correction for multiplicity (significance level p<0.05; SciPy 1.10).

3. Results

3.1. Comparative Analysis of Reconstruction Results

The proposed method was systematically compared with three contrast methods: pure data-driven 3D U-Net, standard PINN (implemented by fully connected network), and multi-scale convolutional network without attention mechanism (MS-CNN). All methods were independently trained and tested on the same 5% sparse sampling data, with the full-field results of FDFD numerical simulation as the true value benchmark. The quantitative evaluation results of each method on 40 test samples are shown in Table 3.

Table 3. Quantitative evaluation results of optical field reconstruction by each method (mean ± standard deviation)

Method	PSNR (dB)	SSIM	Phase RMSE (rad)
3D U-Net	28.4 ± 1.8	0.861 ± 0.024	0.312 ± 0.041
Standard PINN	26.7 ± 2.3	0.834 ± 0.031	0.358 ± 0.052
MS-CNN	30.1 ± 1.6	0.883 ± 0.019	0.274 ± 0.037
Proposed method	34.6 ± 1.3	0.932 ± 0.011	0.183 ± 0.025

As shown in Table 3, the proposed method shows superior reconstruction performance under sparse sampling. Compared with the best baseline MS-CNN, it improves PSNR by 4.5 dB, increases SSIM from 0.883 to 0.932, and reduces phase RMSE by 33.2%. These improvements indicate the reconstructed three-dimensional field is closer to the reference, especially in regions with local cracks, pores, inclusions, or sharp interface variations. Lower standard deviations across the 40 test samples further suggest stability across configurations. Simultaneous improvement of all three metrics is statistically significant by ANOVA (p<0.001p<0.001 p<0.001), demonstrating robust gains from the synergistic introduction of physical priors and multi-scale attention.

From the spatial distributions (Fig. 4), 3D U-Net reconstructs low-frequency envelopes but smooths diffraction-spike details, underestimating diffraction-focus peak intensity by ~18%. Standard PINN partially follows the physics but deviates in corner regions, with the lowest PSNR (26.7 dB), reflecting the limited 3D feature-extraction efficiency of fully connected structures. MS-CNN improves multi-scale extraction but, lacking attention, handles complex high-frequency structures near scatterers insufficiently, with larger phase errors in strong-scattering regions. The proposed method is closest to ground truth in both strong-scattering regions and near complex boundaries, with diffraction-focus centroid deviation ~0.3 μm and peak-intensity error ~4.2%. Loss-convergence curves are in Fig. 5: the physical loss reaches a first-order plateau at ~3,000 iterations while the data loss keeps decreasing, verifying that physical constraints guide the network into the physically feasible solution space ahead of full data convergence.

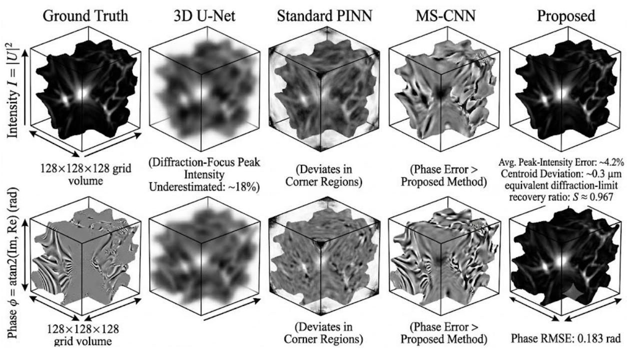

Figure 4. Comparison of three-dimensional optical field reconstruction results by different methods in the corrected 128 × 128 × 128 computational grid volume. The corresponding physical domain size is 6.78 μm × 6.78 μm × 6.78 μm with an isotropic grid spacing of 53 nm

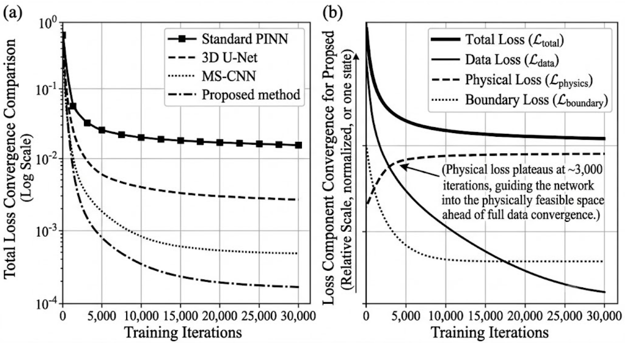

Figure 5. Comparison of loss function convergence curves during training of each method

3.2. Robustness Analysis

Addressing inevitable noise problems in actual measurements, additive white Gaussian noise at three signal-to-noise ratio levels (SNR = 20 dB, 15 dB, 10 dB) was superimposed on input data of 40 test samples to evaluate reconstruction robustness under noisy conditions. Comparison results are shown in Fig. 6. The physical meanings of the two metrics are clearly distinguished: input SNR measures noise pollution degree of sampling data, and output PSNR measures reconstruction result fidelity relative to ground truth, related by the method's noise amplification characteristics.

As input noise intensity increases, reconstruction quality of all methods decreases, but PSNR and SSIM attenuation amplitudes of the proposed method are smallest. Under SNR = 20 dB conditions, the proposed method's PSNR is 32.8 dB, only 1.8 dB lower than the noise-free benchmark. Under strong noise conditions (SNR = 10 dB), the proposed method still maintains reconstruction accuracy with PSNR higher than 29 dB and SSIM higher than 0.88, while 3D U-Net's SSIM drops to 0.791 and standard PINN drops to 0.763, no longer guaranteeing basic structural restoration reliability. ANOVA statistical analysis shows PSNR differences between the proposed method and each baseline method under three noise levels are statistically significant ($p<0.01$), and the proposed method has smallest performance variance between different noise levels (standard deviation 1.64 dB), significantly lower than 3D U-Net (2.91 dB) and standard PINN (3.28 dB), indicating systematic advantages in stability against noise disturbance.

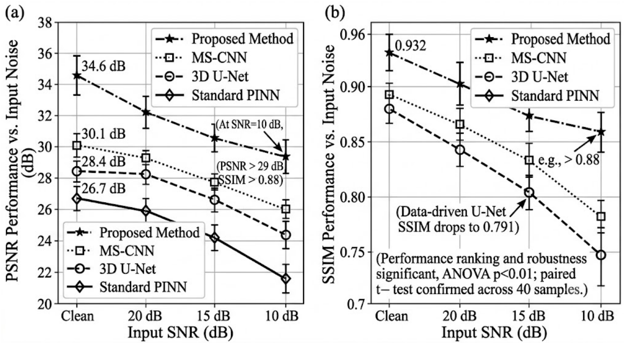

Figure 6. Comparison curves of reconstruction performance of each method under different input noise levels (error bars denote mean ± standard deviation across test samples)

3.3. Ablation Experiments

To quantitatively analyze the contribution of each design module, systematic ablation experiments were carried out under the same experimental conditions, and the results are shown in Table 4.

Table 4. Ablation experiment results

Model Configuration	PSNR (dB)	SSIM	Phase RMSE (rad)
Basic U-Net (no physics, no attention)	28.4 ± 1.8	0.861 ± 0.024	0.312 ± 0.041
+ Physical constraints	31.2 ± 1.6	0.897 ± 0.018	0.241 ± 0.033
+ Multi-scale convolution	30.1 ± 1.6	0.883 ± 0.019	0.274 ± 0.037
+ Physical + multi-scale	32.7 ± 1.5	0.913 ± 0.015	0.209 ± 0.030
+ channel attention	33.6 ± 1.4	0.921 ± 0.013	0.198 ± 0.028
Complete method (all modules)	34.6 ± 1.3	0.932 ± 0.011	0.183 ± 0.025

As shown in Table 4, physical constraints alone bring a 2.8 dB PSNR gain and multi-scale convolution alone 1.7 dB. The two together yield 4.3 dB, close to the sum of their individual gains (2.8 + 1.7 = 4.5 dB); the combined improvement is therefore essentially additive, indicating that the physical prior and multi-scale convolution act through largely complementary mechanisms without significant mutual interference (the 0.2 dB gap is within run-to-run variability). On this basis, sequentially adding channel and spatial attention contributes a further 0.9 dB and 1.0 dB, respectively. Across the 40 paired test fields, every adjacent configuration differs significantly (paired tt t-test, Holm-corrected $p<0.05$; confirmed by Wilcoxon signed-rank), supporting the independent contribution of each module.

4. Discussion

The proposed framework achieves high-precision three-dimensional optical field reconstruction under 5% sparse sampling, reducing phase reconstruction error by 33.2% relative to the best-performing baseline and demonstrating systematic noise robustness. The ablation study provides quantitative attribution of performance gains to individual design choices.

The physical constraint strategy serves two complementary roles. On one hand, embedding the Helmholtz equation residual into the loss function provides strong prior information about the spatial distribution of the optical field, substantially compressing the effective solution space under sparse data conditions. This is consistent with the observed training behavior, where physical loss converges to a first-order plateau around 3,000 iterations while data loss continues to decrease—indicating that physical constraints guide the network into the physically feasible solution subspace before data fitting fully converges. On the other hand, the Helmholtz constraint imposes a manifold condition on the optimization landscape: noise components that deviate from the physically realizable manifold are penalized directly, making them difficult to learn and propagate through the network. This dual role—solution space compression and noise manifold rejection—explains the systematic robustness advantage observed under strong noise conditions (SNR = 10 dB), where the proposed method maintains SSIM above 0.88 while data-driven baselines degrade substantially.

The contribution of multi-scale attention mechanisms operates at two complementary levels. Channel attention adaptively weights the contributions of the real part, imaginary part, intensity, and phase of the complex amplitude, enabling the network to dynamically adjust emphasis across physical components during training. Its independent introduction yields a Phase RMSE improvement of 0.011 rad in ablation experiments, indicating that this mechanism effectively reduces gradient interference between the intensity and phase reconstruction sub-objectives. Spatial attention addresses the feature-learning imbalance between strong scattering foci and smooth background regions by assigning higher computational weights to high-gradient regions near scatterers. Its independent PSNR contribution of 1.0 dB suggests a particularly pronounced gain for physically heterogeneous fields such as three-dimensional complex amplitude distributions. Together, the two mechanisms achieve a cumulative PSNR improvement of 1.9 dB with only 0.8% additional parameter overhead, demonstrating high parameter efficiency.

Applicability of the physical prior. The scalar Helmholtz model is appropriate for weakly-to-moderately scattering, isotropic, scalar-wave modalities—coherent optical microscopy, digital holography, and ultrasonic inspection of nominally isotropic media. It is not directly valid for anisotropic-elastic ultrasound (which requires the elastodynamic Christoffel equations), polarization-resolved electromagnetic inspection (vector Maxwell), or strongly multiple-scattering regimes. For those modalities the framework should be re-instantiated with the corresponding vector or tensor governing operator—a modular extension that preserves the multi-scale attention backbone. Unlike operator-learning / state-space approaches such as Mamba neural operators [22], which learn solution operators from data, our method imposes the governing equation pointwise as a hard residual, trading operator generality for guaranteed physical consistency under extreme sparsity.

From optical field to defect metrics. The reconstructed complex field is linked to defect metrics through Eq. (12):

$$\mathrm{OPD} = \frac{\lambda_0}{2\pi}\Delta\varphi, \Delta\varphi = \frac{2\pi}{\lambda_0}\Delta n\, t, t_{\min} \approx \frac{\lambda_0\,\varphi_{\mathrm{RMS}}}{2\pi\,\Delta n} \quad (12)$$

The achieved phase RMSE of 0.183 rad maps to an RMS optical-path error of ≈15.5 nm (versus ≈23.2 nm for MS-CNN). For a localized index contrast Δn\Delta n Δn over thickness tt t, the minimum reliably resolvable defect thickness is $t_{\min} \approx 15.5\ \mathrm{nm}/\Delta n$—about 31 nm for a void ($\Delta n \approx 0.5$), ≈155 nm for a moderate inclusion ($\Delta n \approx 0.1$), and ≈310 nm for a low-contrast inclusion ($\Delta n \approx 0.05$). The 0.3 μm centroid deviation and 4.2% peak-intensity error correspond to defect-localization accuracy of ≈0.3 μm and scattering-contrast (defect-strength) estimation error of ≈4%, respectively. This translation layer connects the reconstructed field to engineering-relevant defect dimensions and contrasts.

The practical engineering implications are concrete. In adaptive optics, the Strehl improvement from 0.926 to 0.967 corresponds to a meaningful recovery of diffraction-limited performance and more correctable wavefront modes. In coherent diffraction imaging, the physical-constraint framework may help suppress stagnation in iterative phase retrieval. In biomedical scattering imaging, the method offers an efficient pathway for three-dimensional optical field inversion from sparse measurement data.

The proposed framework has implications across the manufacturing lifecycle—prototype defect-region recovery for design-for-quality, supplier quality evaluation from limited NDT data, online process monitoring (welding, additive manufacturing, precision machining, composite curing), reduced reliance on dense scanning for final inspection, and damage assessment in after-sales/remanufacturing.

The principal limitation of this study is that validation rests entirely on numerical simulation. As a concrete next step we will validate on a calibrated physical target—e.g., a micro-machined phase object with known void/inclusion geometry measured by digital holographic microscopy or a Hartmann–Shack wavefront sensor—to confirm transfer from simulated to real sensor data, and will extend the scalar prior to vector/elastodynamic operators for additional inspection modalities.

5. Conclusion

This study proposes a physics-prior and multi-scale attention-based framework for three-dimensional manufacturing quality-field reconstruction under sparse inspection conditions. By embedding Helmholtz-type residual constraints into the learning objective and integrating multi-scale three-dimensional convolution with serial channel-spatial attention, the method combines data-driven representation learning with physically consistent reconstruction. Using only 5% sparse sampling, it achieves PSNR 34.6 dB and SSIM 0.932 while reducing phase RMSE by 33.2% versus the best baseline; under strong noise it maintains PSNR > 29 dB and SSIM > 0.88.
The results suggest that physical priors constrain the feasible reconstruction space and suppress non-physical noise propagation, while multi-scale attention improves recovery of both global structural deviations and local defect features. Validation is currently based on physically rigorous simulation; the immediate next step is verification on real, calibrated industrial measurements. The framework provides a promising technical route for intelligent-manufacturing applications including product quality inspection, material evaluation, process monitoring, supply-chain quality control, and after-sales assessment.

References

[1] Ashebir D A, Hendlmeier A, Dunn M, et al. Detecting multi-scale defects in material extrusion additive manufacturing of fiber-reinforced thermoplastic composites: a review of challenges and advanced non-destructive testing techniques[J]. Polymers, 2024, 16(21): 2986.

[2] Zhang X, Dong M, Cai X, et al. Progress in machining-induced residual stress and microstructural evolution of inhomogeneous materials and composites[J]. Advanced Composites and Hybrid Materials, 2023, 6(3): 122.

[3] Usamentiaga R, Menendez F, Perez L. Towards low-cost surface inspection using artificial vision in practical industrial environments[J]. IEEE Transactions on Industry Applications, 2026.

[4] Meister S, Grundhoefer L, Stüve J, et al. Imaging sensor data modelling and evaluation based on optical composite characteristics: Investigation of data quality for inline inspection[J]. The International Journal of Advanced Manufacturing Technology, 2021, 116(11): 3965-3990.

[5] Feng F, Zhao H, E K, et al. Subaperture stitching for large-aperture wavefront detection based on modified Hartmann measurement method[J]. Optical Engineering, 2025, 64(7): 074104-074104.

[6] Teixeira F L, Sarris C, Zhang Y, et al. Finite-difference time-domain methods[J]. Nature Reviews Methods Primers, 2023, 3(1): 75.

[7] Stefanski T P. Applications of the discrete Green's function in the finite-difference time-domain method[J]. Progress In Electromagnetics Research, 2013, 139: 479-498.

[8] Penwarden M., Zhe S., Narayan A., et al. Multifidelity modeling for Physics-Informed Neural Networks (PINNs)[J]. Journal of Computational Physics, 2022, 451: 110844.

[9] Hosseini M Y, Shiri Y. Flow field reconstruction from sparse sensor measurements with physics-informed neural networks[J]. Physics of Fluids, 2024, 36(7).

[10] Colombo D, Turkoglu E, Li W, et al. Coupled physics-deep learning inversion[J]. Computers & Geosciences, 2021, 157: 104917.

[11] Oommen V., Shukla K., Goswami S., et al. Learning two-phase microstructure evolution using neural operators and autoencoder architectures[J]. npj Computational Materials, 2022, 8(1): 190.

[12] Dong Y., Chen S., Liu X.J., et al. Physics-constrained neural networks for half-space seismic wave modeling[J]. Computers and Geosciences, 2023, 181: 105477.

[13] Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Cham: Springer, 2015: 234–241.

[14] Vinuesa R., Brunton S.L. Enhancing computational fluid dynamics with machine learning[J]. Nature Computational Science, 2022, 2(6): 358–366.

[15] Aikawa Y., Ueda N., Tanaka T. Improving the Efficiency of Training Physics-Informed Neural Networks Using Active Learning[J]. New Generation Computing, 2024, 42(4): 739–760.

[16] Raissi M., Perdikaris P., Karniadakis G.E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. Journal of Computational Physics, 2019, 378: 686–707.

[17] Cai S., Mao Z., Wang Z., et al. Physics-informed neural networks (PINNs) for fluid mechanics: a review[J]. Acta Mechanica Sinica, 2021, 37(12): 1727–1738.

[18] Escapil-Inchauspé P., Ruz G.A. Hyper-parameter tuning of physics-informed neural networks: Application to Helmholtz problems[J]. Neurocomputing, 2023, 561: 126826.

[19] Xiong L., Zeng Y., Huang X., et al. MLANet: multi-level attention network with multi-scale feature fusion for crowd counting[J]. Cluster Computing, 2024, 27(5): 6591–6608.

[20] Woo S., Park J., Lee J.Y., et al. CBAM: Convolutional block attention module[C]// Proceedings of the European Conference on Computer Vision (ECCV). Cham: Springer, 2018: 3–19.

[21] Bodong Z, Maojin T, Guohua Z, et al. An Integrating Physics-Informed Neural Networks and Attention Mechanism based Intelligent Permeability Prediction Method for Shale Reservoirs[J]. SPE Journal, 2026: 1-18.

[22] Cheng W.C., Huang J., Zhang Y., et al. Mamba neural operator: Who wins? Transformers vs. state-space models for PDEs[J]. Journal of Computational Physics, 2026, 548: 114567.

[23] Mojgani R., Balajewicz M., Hassanzadeh P. Kolmogorov n–width and Lagrangian physics-informed neural networks: A causality-conforming manifold for convection-dominated PDEs[J]. Computer Methods in Applied Mechanics and Engineering, 2023, 404: 115810.
[24] Jagtap A.D., Kawaguchi K., Karniadakis G.E. Adaptive activation functions accelerate convergence in deep and physics-informed neural networks[J]. Journal of Computational Physics, 2020, 404: 109136.
[25] Liu B, Li H, Bian S, et al. A Gravity Data Denoising Method Based on Multi-Scale Attention Mechanism and Physical Constraints Using U-Net[J]. Applied Sciences, 2025, 15(14): 7956.
[26] Xu Y, Li R, Liu X, et al. Multi-Scale Dual-Attention Feature Network with Bidirectional Temporal Constraints for Tool Wear Monitoring[J]. Coatings, 2026, 16(3): 291.
[27] May M C, Oberst J, Lanza G. Managing product-inherent constraints with artificial intelligence: production control for time constraints in semiconductor manufacturing[J]. Journal of Intelligent Manufacturing, 2024, 35(8): 4259-4276.
[28] Schiassi E., Calabrò F., De Florio M., et al. Use of ELMs in Physics-Informed Machine Learning: Estimates on the Generalization Error in Learning ODE Solutions[J]. Journal of Scientific Computing, 2026, 107(1): 22.
[29] Cardoso Ferreira F.S., Patsios S., Paris P. φ-PINNs: Physics-informed neural networks on complex geometries[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107324.