Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

U-Net–Based 3D Structured Illumination Microscopy of Transparent Materials

Zili Lei, Liqing Wan, Wei Shen, Da Liu, Zhongsheng Zhai

How to cite: Lei Z, Wan L, Shen W, Liu D, Zhai Z. U-Net–Based 3D Structured Illumination Microscopy of Transparent Materials. Textile & Leather Review. 2026; 9:2629-2654. https://doi.org/10.31881/TLR.2026.2629

How to link: https://doi.org/10.31881/TLR.2026.2629

Published: 25 April 2026

U-Net–Based 3D Structured Illumination Microscopy of Transparent Materials

Zili Lei*, Liqing Wan, Wei Shen, Da Liu, Zhongsheng Zhai

Hubei Key Laboratory of Modern Manufacturing Quantity Engineering, School of Mechanical Engineering, Hubei University of Technology, Wuhan 430068, China

*115394315@qq.com

Article

https://doi.org/10.31881/TLR.2026.2629

Published 25 April 2026

ABSTRACT

Translucent materials are widely used in micro-optical components, and related optically transmissive structures, where surface three-dimensional morphology directly influences optical transmission efficiency and interfacial reliability. Structured Illumination Microscopy (SIM), with its optical sectioning capability, is an important technique for 3D measurement of transparent and translucent materials. However, their inherently low reflectivity often reduces fringe modulation contrast, making weak fringes difficult to resolve. Moreover, refractive-index inhomogeneity can induce uncontrolled fringe phase shifts, which further amplify fringe frequency deviations caused by DMD projection offsets and optical-path misalignment, degrading 3D reconstruction stability and limiting high-throughput, real-time inspection. To address these issues, this study proposes a single-frame U-Net reconstruction framework integrating a Squeeze-and-Excitation (SE) module and a Transformer Bottleneck (TB) for global dependency modeling. A physics-informed Frequency Perturbation Augmentation (FPA) strategy is further introduced to improve weak-fringe parsing and reconstruction robustness. Trained on 8,542 pairs of structured illumination and optically sectioned images, the proposed method significantly improves imaging quality while reducing reconstruction time, achieving about an eightfold acceleration. It provides an efficient, scalable solution for high-throughput 3D surface morphology reconstruction of transparent and translucent materials under weakly modulated imaging conditions.

KEYWORDS

structured illumination microscopy, transparent and translucent materials, 3D surface metrology, U-Net, attention mechanism

INTRODUCTION

Transparent materials, owing to their excellent optical transmittance, are widely used in optoelectronic devices such as integrated circuit packaging, optical lenses, and optical fibers. The surface morphology of

transparent and translucent materialsplays a critical role in determining device performance, coating uniformity, and interfacial quality. However, due to their high transmittance and low reflectance, Optical Sectioning Structured Illumination Microscopy (OS-SIM) often fails to produce stable fringe modulation signals on such surfaces. Moreover, the illumination pattern is prone to fringe phase shift across the field of view, resulting in reduced image contrast, insufficient signal-to-noise ratio, and poor reconstruction robustness, thereby limiting its application in 3D surface morphology reconstruction of transparent samples [1], [2]. In this study, "reconstruction robustness" refers to the resilience of the reconstruction results to fringe phase shifts, fringe period deviations, and noise perturbations, as reflected by the continuity of the optically sectioned structure and the consistency of peak localization in Adaptive Reference Search (ARS).

Conventional 3D inspection methods mainly rely on geometric or physical modeling. The refractive-geometry reconstruction method based on ray tracing proposed by Kutulakos et al. can theoretically recover the surface morphology of transparent objects, but it is computationally demanding and requires stringent experimental conditions [3]. Murase et al. inverted surface normals of transparent media through optical-flow estimation, but the method is only applicable to specific refractive environments [4]. Although the invasive inspection method of Aberman et al. can achieve high accuracy, it compromises sample integrity [5]. Terahertz ghost imaging by Olivieri et al. and the mid-infrared thermographic approach developed by Fraunhofer IOF have shown progress in small-sample inspection; however, both methods involve complex instrumentation and high cost [6]. Meanwhile, existing CNN-based SIM reconstruction methods still struggle to maintain accuracy when applied to transparent samples [7]. Overall, current techniques are either constrained by hardware requirements or exhibit pronounced trade-offs among accuracy, robustness, and high-throughput inspection capability [8].

It is worth noting that although HiLo-based structured illumination microscopy performs well in terms of optical sectioning quality, its reconstruction pipeline inherently depends on multi-frame phase-shift demodulation and axial scanning. Specifically, in a typical HiLo imaging process, at least two images with different illumination conditions—uniform illumination and fringe illumination—must be acquired at each focal plane, and optically sectioned reconstruction is then performed through frequency-domain filtering and contrast computation [9]. This process involves pixel-wise peak fitting and frequency-domain operations, and its computational complexity increases approximately linearly with image size and the number of scanned layers, resulting in a relatively long overall reconstruction time.

In industrial scenarios such as micro-optical component fabrication, reconstruction is often required at millisecond-level speed to enable continuous in-line monitoring and rapid defect identification [10]. Multi-frame acquisition not only increases sampling time but also imposes stricter requirements on system stability. When slight vibration or dynamic variation is present, multi-frame fusion can easily introduce artifacts. Therefore, reducing the number of acquired frames and the computational complexity while maintaining measurement accuracy remains a key challenge in 3D surface morphology reconstruction of transparent materials.

In recent years, deep learning has been introduced into the field of structured illumination microscopy reconstruction. For example, Chai proposed a deep-learning-based one-shot structured illumination reconstruction method that enabled relatively fast surface measurement [11]. However, this method was primarily designed for general reflective samples, and its training data were derived from structured illumination images acquired under ideal modulation conditions, which limits its adaptability to transparent-material scenarios characterized by high transmittance and low modulation depth. In addition, most existing deep-learning-based SIM methods still rely on multi-frame inputs or do not adequately account for physical characteristics specific to transparent materials, such as fringe phase shifts and modulation attenuation. Therefore, the development of a physics-consistent single-frame reconstruction framework tailored to the weakly modulated fringe characteristics of transparent materials remains of considerable research significance.

Motivated by the above analysis, this study proposes a single-frame U-Net reconstruction method that integrates an attention mechanism with global dependency modeling. Specifically, a Squeeze-and-Excitation (SE) module and a lightweight Transformer Bottleneck (TB) are incorporated into the network to enhance fringe modulation representation and capture long-range spatial dependencies, respectively, thereby improving both the expression of structural details and the global consistency of surface morphology during reconstruction [12], [13]. In addition, a Frequency Perturbation Augmentation (FPA) strategy is designed, in which ±5% spatial frequency perturbations are applied to the input structured illumination images during training to simulate the period deviation and angular errors commonly encountered in digital micromirror device (DMD) projection [14], [15]. Although transparent and translucent materials differ in their optical scattering mechanisms, the present work does not classify samples according to material taxonomy itself. Instead, the proposed framework is driven by shared degradation patterns observed in reflective OS-SIM, including weak fringe modulation, phase perturbation, and fringe-frequency mismatch. When transparent surfaces and weakly scattering/translucent samples generate similarly degraded raw fringe observations under low-reflectivity

conditions, they can be reconstructed using a unified degradation-driven model. Experimental results show that the proposed method significantly improves imaging quality, substantially reduces reconstruction time compared with conventional approaches, and exhibits superior robustness and generalization under complex sample conditions. These results demonstrate that the proposed framework provides an efficient single-frame solution for real-time 3D surface morphology reconstruction of transparent materials.

METHODS

Optical-sectioning Structured Illumination Microscopy

As shown in Fig.1, an Optical Sectioning Structured Illumination Microscopy (OS-SIM) system was developed based on a digital micromirror device (DMD; DLP LightCrafter 6500, Texas Instruments). The illumination module adopts a Köhler illumination configuration, in which the condenser aperture diaphragm and field diaphragm are adjusted to provide uniform and controllable illumination, thereby optimizing both illumination intensity and imaging field of view. The system employed an incoherent green LED illumination source with a center wavelength of 518 nm, and the light was directed to the DMD panel by a total internal reflection (TIR) prism. Acting as a fringe pattern generator, the DMD enables high-speed modulation and projection of structured illumination patterns with different spatial frequencies and phases.

The modulated beam is then coupled through a tube lens and a beam-splitting prism into the objective lens (20×, NA 0.45, Nikon), which projects the structured light field onto the sample surface with high fidelity. Axial displacement of the objective is realized using a high-precision objective scanning stage (P-721, Physik Instrumente), allowing the acquisition of a stack of structured-illumination wide-field images at different focal planes. All images are captured by a high-speed CMOS camera (acA1920-155 µm, Basler), providing a balance between high spatial resolution and a sufficiently large field of view [16].

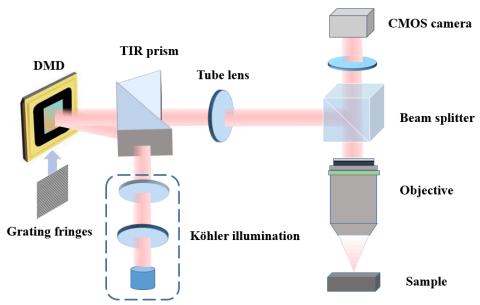

Figure 1. Schematic of the custom-built optical sectioning structured illumination microscopy (OS SIM) system

A digital micromirror device (DMD) is used to generate structured illumination patterns, which are projected onto the sample through a microscope objective. Axial scanning enables the acquisition of structured illumination image stacks for three-dimensional surface reconstruction

As shown in Fig.2, which presents a photograph of the OS-SIM system, the acquired images in this setup have a resolution of 512 × 512 pixels, a pixel size of 0.325 μm/pixel, and a bit depth of 12 bits. The spatial frequency of the projected fringes is set to 0.85 times the system cutoff frequency, which was estimated from the objective numerical aperture (NA = 0.45) and the illumination wavelength (518 nm), corresponding to a fringe period of approximately 7.2 μm on the sample surface. For a high-reflectivity standard grating sample, the fringe modulation depth is approximately 18%–22%. In contrast, for transparent samples, most of the incident light is transmitted, and only weak Fresnel reflection contributes to image formation. As a result, the modulation depth decreases significantly to 6%–10%, accompanied by a strong scattering background.

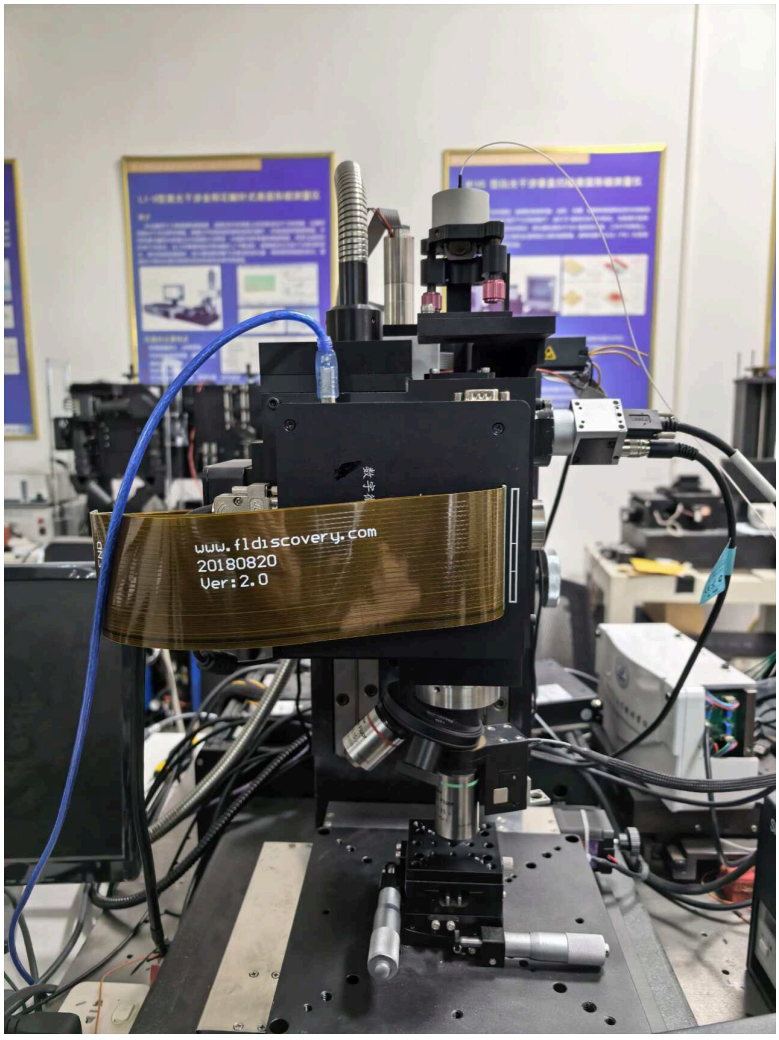

Figure 2. Photograph of the Optical Sectioning Structured Illumination Microscopy (OS-SIM) system

To illustrate the characteristics of the network input, Fig.3 presents representative imaging results acquired by the OS-SIM system. As shown in Fig.3(a), the raw single-frame fringe image of a microlens array sample was captured under structured illumination. Because the surface reflectivity of the transparent material is low, only weak Fresnel reflection contributes to image formation, resulting in low fringe modulation depth and generally weak image contrast. Fig.3(b) shows a magnified view of the region marked by the red box in Fig.3(a), in which the structured illumination fringes can be observed more clearly. Owing to the influence of surface microstructures and local refractive-index variations, the fringes exhibit slight fringe phase shifts and contrast fluctuations in some regions.

Fig.3(c) presents the optically sectioned image reconstructed using the HiLo method, which serves as the reference labelfor network training. Compared with the raw fringe image, the optically sectioned result effectively suppresses background scattering and enhances the contrast of surface structures. However, its reconstruction process requires multi-frame demodulation and frequency-domain computation, leading to substantial computational overhead. As shown in Fig.3(d), refractive-index gradients and microscopic surface undulations in the transparent material introduce regional fringe phase shifts and fringe frequency deviation, causing variations in the fringe period across the field of view. These effects reduce reconstruction robustness in 3D reconstruction and ultimately degrade the reconstruction performance.

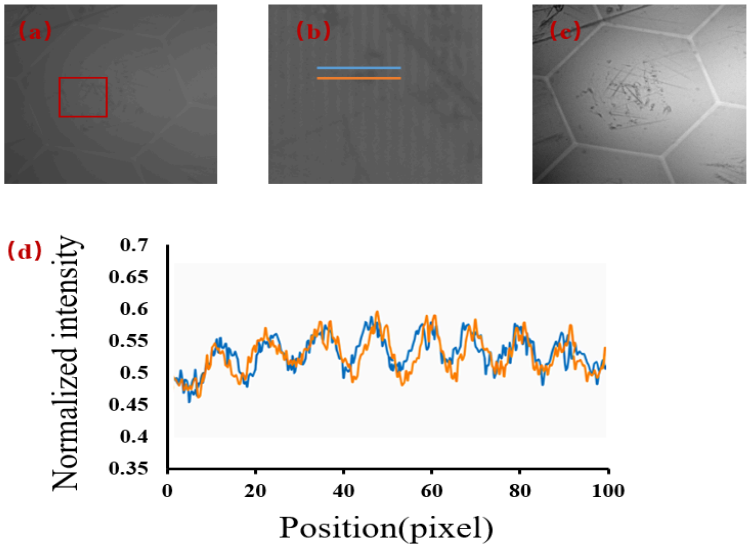

Figure 3. Representative images acquired by the OS-SIM system .(a) Raw single-frame structured illumination image of a microlens array sample.(b) Enlarged view of the region marked in (a), showing weak modulation fringes.(c) Optical section reconstructed using the HiLo method, used as the reference labels for network training. (d)Grayscale intensity profile extracted along the line indicated in Fig. (b), showing regional fringe phase shifts and local variations in peak spacing

The core idea of OS-SIM is to enhance image contrast using spatial fringe modulation, perform optical sectioning via multi-frame phase-shift demodulation, and then obtain a tomographic stack through axial scanning [1,2]. Under ideal conditions, a single-frame fringe image can be expressed as:

$$I(x,y,z) = I_0(x,y,z)[1 + m(x,y,z)\cos(2\pi f(x\cos\theta + y\sin\theta) + \phi(x,y,z))] \quad (1)$$

where $I_0(x,y,z)$ is the background intensity, $m(x,y,z)$ denotes the local fringe modulation depth, f is the spatial frequency, θ is the fringe orientation angle, and $\phi(x,y,z)$ is the phase term. For transparent materials, weak reflective signals and local optical-path variations may degrade demodulation stability.

In OS-SIM 3D reconstruction, the axial response signal (ARS) of each pixel can be approximated by a Gaussian distribution:

$$m(z) = r\exp\left(-\left(\frac{z - z_a}{\mathrm{FWHM}}\right)^2\right) \quad (2)$$

whre is the axial peak position, represents the axial resolution of the system, and r is the peak modulation amplitude coefficient. The ARS algorithm recovers the sample surface height by localizing this peak. When fringe modulation decreases, the ARS peak becomes less distinguishable and can be easily buried in noise, thereby degrading 3D surface reconstruction accuracy.[17][18][19]

Therefore, three key challenges arise in the imaging of transparent materials:

(1) Weak fringe modulation under low-reflectivity conditions makes accurate ARS peak localization difficult. [20].

(2) Local optical-path variation introduces fringe phase shifts across the field of view..

(3) Refractive-index inhomogeneity may aggravate fringe-frequency mismatch induced by projection and optical-path errors [19], [21], [22].

Network Architecture for Single-frame Reconstruction

To address the three key challenges described above, this study introduces both a Squeeze-and-Excitation (SE) module and a Transformer Bottleneck (TB) into the baseline U-Net architecture for targeted enhancement. The proposed SE-Transformer U-Net for single-frame reconstruction is illustrated in Fig.4. The network takes a single structured illumination image $I(x,y)$ as input and outputs the corresponding optically sectioned

image $S(x, y)$ at the focal plane. The overall architecture consists of components: an encoder, a Transformer bottleneck layer, and a decoder. The encoder is used to extract local fringe and edge features, the TB module captures global spatial dependencies and enables cross-region information interaction, and the decoder progressively restores spatial resolution to generate a high-quality optically sectioned result.

The network design is motivated by two key considerations:

(1) SE channel attention module: By exploiting global channel-wise statistical information, the SE module adaptively recalibrates feature channels, enhancing the responses of channels closely related to structured illumination fringe modulation while suppressing channels dominated by background noise. In this way, the effective signal-to-noise ratio of weakly modulated fringes is improved in the feature domain, which enhances the ability of the subsequent decoding stages to recover low-contrast structures.

(2) Transformer Bottleneck (TB) module: The TB module is introduced at the compressed bottleneck layer to enable cross-region information interaction, thereby improving the network's ability to characterize global phase consistency and periodic structural continuity, while mitigating spatial distortion caused by local fringe phase shifts.

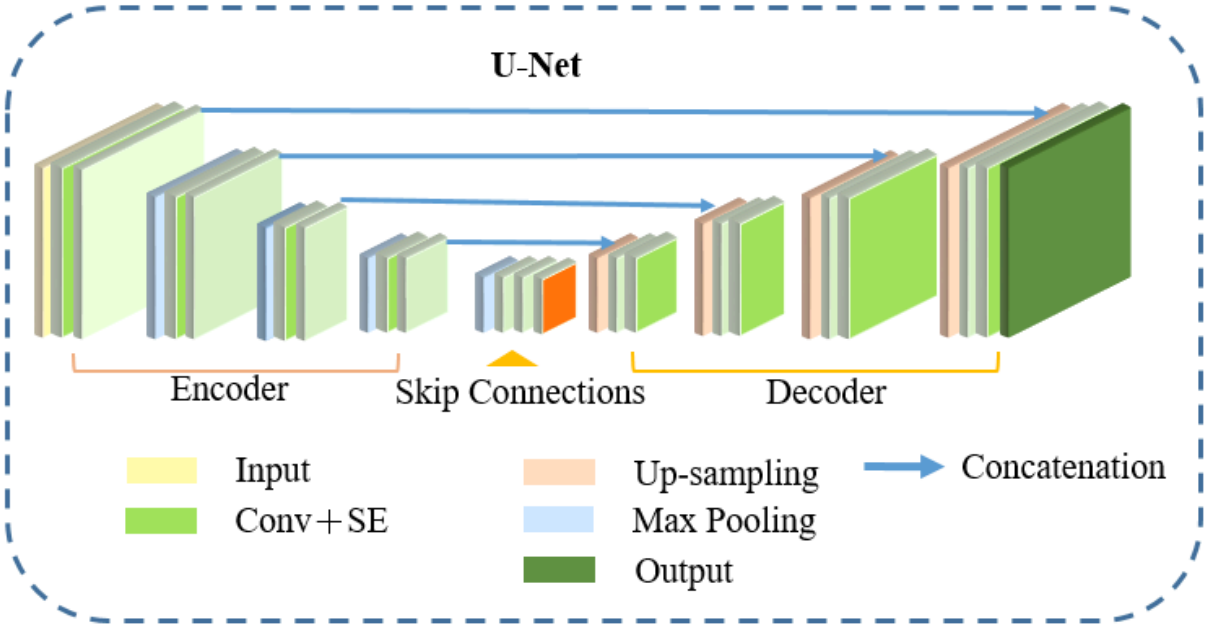

Figure 4. Architecture of the proposed SE-Transformer U-Net network. The network consists of an encoder–Transformer bottleneck–decoder structure with skip connections. Squeeze-and-Excitation (SE) modules and a lightweight Transformer bottleneck are integrated to enhance weak modulation features and capture global spatial dependencies

Weak-modulation Enhancement Using the SE Module

To enhance weakly modulated fringe features, an SE module is introduced after each convolutional block to enhance the responses of feature channels closely associated with structured illumination fringe modulation. Specifically, the module first extracts channel-wise statistical information through global average pooling and then learns a channel weight vector for feature recalibration [12]. As shown in Fig.5, the Transformer bottleneck is integrated with the CNN backbone through a CNN–Transformer fusion scheme, enabling effective interaction between local convolutional features and global contextual representations.

Let the input feature map be $X \in R^{C\times H\times W}$. The SE module first compresses spatial information through global average pooling to compute channel-wise statistics:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i,j), c = 1,2,...,C \tag{3}$$

The resulting vector is then fed into a bottleneck structure consisting of two fully connected layers to capture nonlinear inter-channel relationships:

$$s = \sigma(W_2 \cdot \delta(W_1 \cdot z)) \tag{4}$$

where $W_1 \in R^{C/r\times C}$ and $W_2 \in R^{C\times C/r}$ are the weight matrices for dimensionality reduction and expansion, respectively; $\delta(\cdot)$ denotes the ReLU activation function; $\sigma(\cdot)$ denotes the Sigmoid activation function; and r is the channel reduction ratio. Finally, the excitation vector $s \in R^C$ is used as attention weights to rescale the original feature channels:

$$\tilde{X}_c = s_c \cdot X_c, \forall c = 1,2,...,C \tag{5}$$

Through this mechanism, the SE module adaptively enhances key channels related to fringe modulation while suppressing noise-dominant channels, thereby improving the effective signal-to-noise ratio of fringe features in the learned feature space.

Global phase-consistency modeling with the Transformer Bottleneck

To mitigate fringe phase shifts caused by spatial displacement arising from refractive-index gradients and microscopic surface undulations, a lightweight Transformer Bottleneck (TB) module is introduced into the bottleneck layer of the U-Net. By leveraging self-attention to model long-range spatial dependencies, the module enhances the network's ability to capture global phase consistency and periodic structural continuity, while reducing spatial distortion induced by local fringe phase shifts and pattern deformation [13].

Specifically, let the bottleneck convolutional feature map be $F \in R^{H\times W\times C}$. We first reshape it into a sequence $X \in R^{N\times C}$ with ($N = H \times W$), and then compute global correlations among features using Multi-Head Self-Attention (MHSA):

$$\text{Attention } (Q, K, V) = \text{Softmax } \left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{6}$$

where $Q = XW_Q$, $K = XW_K$, $V = XW_V$; W_Q , W_K , W_V are learnable projection matrices; and d_k is the scaling factor. This mechanism models global dependencies by evaluating correlations among all spatial positions.

To maintain feature-scale consistency, we adopt the standard residual design with layer normalization (LN) and a feed-forward network (MLP). The Transformer Bottleneck can be written as:

$$Z_1 = X + \text{MHSA } (\text{LN } (X)) \tag{7}$$

$$Z_2 = Z_1 + \text{MLP } (\text{LN } (Z_1)) \tag{8}$$

This structure integrates global information while preserving convolutional representations. Considering that CNNs remain advantageous in extracting local textures and high-frequency details, we do not fully replace the convolutional bottleneck with the Transformer. Instead, a complementary fusion strategy is adopted by concatenating the Transformer output feature F_{TB} with the convolutional bottleneck feature F_{CNN} along the channel dimension:

$$F_{\text{fusion}} = \text{Concat } (F_{\text{CNN}}, F_{\text{TB}}) \tag{9}$$

The fused features are then integrated by a convolution layer and fed into the decoder. This design enables the network to balance local detail recovery and global structural consistency under weak modulation and fringe drift. Because self-attention is only applied at the spatially downsampled bottleneck, the additional computational overhead is limited and does not substantially increase inference time. The TB module models cross-region dependencies through global self-attention, thereby strengthening the global constraint on fringe periodicity under fringe phase shifts and pattern deformation, and ultimately improving demodulation consistency.

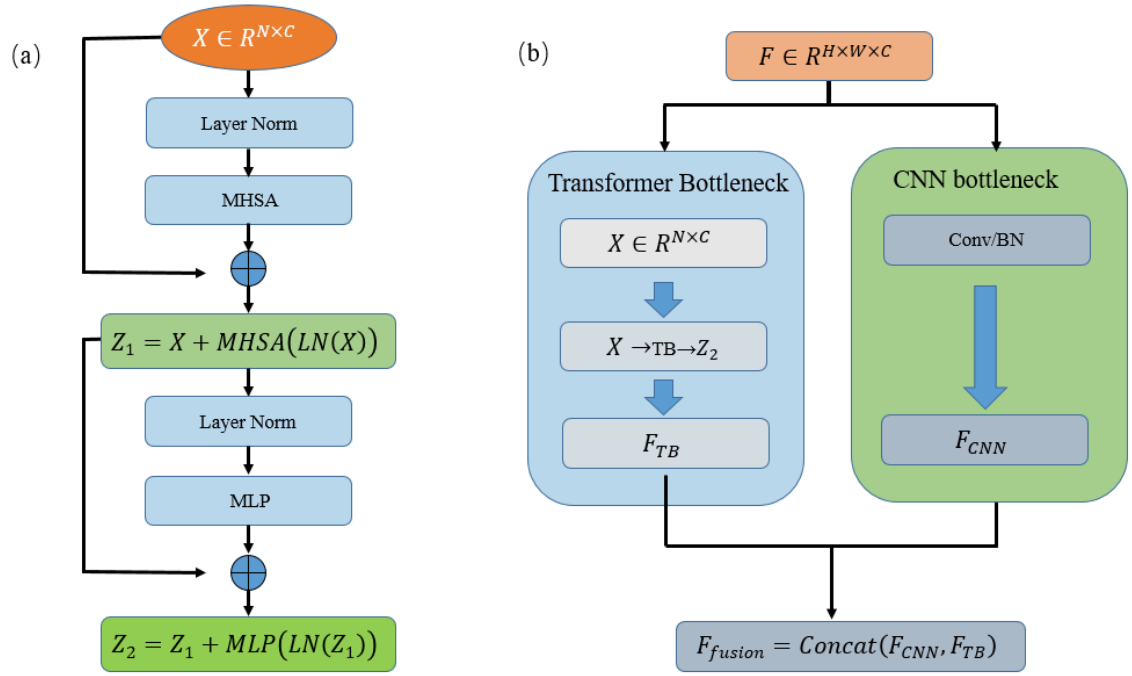

Figure 5. Structure of the Transformer bottleneck and CNN-Transformer fusion scheme: (a) Structure of the Transformer bottleneck; (b) CNN-Transformer fusion scheme

Frequency Perturbation Augmentation Strategy

To improve robustness against fringe-frequency mismatch, we propose a Frequency Perturbation Augmentation (FPA) strategy.

Specifically, during training, the input structured illumination images are subjected to spatial frequency perturbations within ±5% to emulate the period deviation and angular errors frequently observed in practice. The underlying rationale is to expose the network to samples with realistic perturbation patterns at the data level, thereby encouraging the learning of feature representations that are less sensitive to fringe frequency

variation. We note that real DMD tilt and optical-path misalignment may also introduce spatially varying distortions such as chirped fringes or small rotational deviations. In the present work, FPA is used as a first-order approximation of the dominant fringe-period mismatch at the patch level, where slowly varying chirp or small angular deviation can be effectively represented as a local frequency offset.

In this work, simulated perturbation samples are used instead of real perturbed measurements for two reasons. First, simulated perturbations provide better controllability and repeatability, whereas real perturbations are influenced by multiple experimental factors, such as system conditions, temperature variation, and mechanical noise, making the perturbation magnitude and distribution difficult to control precisely. Second, this choice avoids additional hardware complexity. Generating real perturbed samples would require repeated tuning of DMD projection parameters or the intentional introduction of projection offsets, which would increase the experimental workload and may further compromise optical stability. By contrast, FPA can be implemented entirely in software without modifying the hardware setup, making it more portable and easier to generalize.

In the image domain, the fringe pattern can be modeled as a sinusoidally modulated signal:

$$I(x,y) = A(x,y) + B(x,y) \cdot \cos(2\pi x/P) \tag{10}$$

where P denotes the fringe period, $f = 1/P$ is the spatial frequency, $A(x,y)$ is the background intensity, and $B(x,y)$ is the fringe modulation amplitude. In FPA, a perturbation factor $\epsilon \in [-0.05, 0.05]$ is sampled to obtain a perturbed fringe image:

$$I'(x,y) = A(x,y) + B(x,y) \cdot \cos(2\pi x/(P(1+\epsilon))) \tag{11}$$

These perturbed samples expose the network to different frequency offsets during training, thereby improving robustness to fringe drift and phase distortion.

To systematically evaluate the effectiveness of FPA, we perform five-fold cross-validation (K = 5). Let the dataset contain N samples, which are randomly divided into $K = 5$ non-overlapping subsets $\{D_1, D_2, ..., D_5\}$ with comparable sizes, while ensuring that different material types are distributed evenly across subsets. In the k -th experiment, D_k is used as the validation set and the remaining four subsets are used for training. This procedure is repeated five times so that each subset serves once as the validation set.

By comparing model performance with and without FPA, we adopt image entropy and standard deviation as quantitative indicators to characterize information content and reconstruction stability, and report the averaged results across folds. Image entropy is defined as

$$H = -\sum_{i=1}^{L} p_i \log p_i \tag{12}$$

where p_i is the probability of gray level i and L is the number of gray levels. The standard deviation is defined as

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2} \tag{13}$$

where x_i is the intensity of the i -th pixel, μ is the mean image intensity, and N is the total number of pixels. As shown in Fig.6, introducing physics-driven frequency perturbations results in smoother convergence in the early training stage and markedly reduces oscillations. Although the entropy and standard deviation converge to similar final levels in both cases, the model trained with FPA exhibits smaller fluctuations on the validation set, indicating improved stability and generalization. Nevertheless, spatially varying chirp and rotational distortion are not explicitly modeled in the current FPA design and will be considered in future extensions.

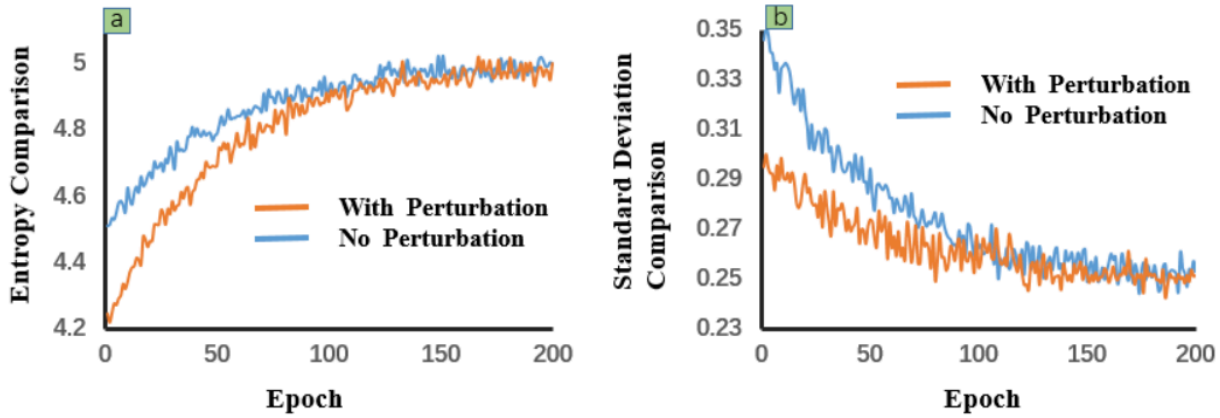

Figure 6. Training curves of image entropy and standard deviation with and without frequency perturbation augmentation. The curves represent the average results of five-fold cross-validation (K=5)

EXPERIMENTS AND RESULTS

Training and Reconstruction Strategy

To train the proposed single-frame reconstruction network, structured illumination image data were acquired using a self-built Optical Sectioning Structured Illumination Microscopy (OS-SIM) system from a variety of transparent and translucent samples exhibiting weak reflected fringe modulation, as illustrated in Fig.7,including transparent glass, compound-eye lenses, microlens arrays, micro/nano chips, large-scale chips, freeform lenses, and cylindrical lenses. These chip-related samples were used to broaden the diversity of weakly modulated fringe patterns. Structured illumination image stacks were obtained through axial scanning, and the corresponding optically sectioned images were generated using an improved HiLo algorithm as reference labels [23], [24].

A dataset comprising 8,542 pairs of structured illumination images and corresponding optically sectioned labels was ultimately established. The original images were cropped using a sliding-window strategy to generate training samples of 512 × 512 pixels, covering 30 different fields of view and tilt angles to improve data diversity and enhance the generalization capability of the model [25], [26]. The dataset was divided into training, validation, and test sets at a ratio of 70% / 15% / 15% for network optimization and performance evaluation. Before training, all input images were normalized to ensure consistency in numerical distribution and to improve training stability [27], [28].

Model training was conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB memory), an Intel Core i9 CPU, and 64 GB RAM. The network was implemented in PyTorch 1.12 with CUDA 11.6. We used the Adam optimizer with an initial learning rate of 1×10^{-4} , a batch size of 8, and 200 training epochs.

The network was optimized using the L1 loss function:

$$L_{L1} = \frac{1}{N} \sum_{i=1}^{N} | \, S_i - \hat{S}_i \, | \tag{14}$$

where S_i is the pixel value of the reference optical-section image, $\hat{S}_i$ is the network prediction, and N is the number of pixels.

During training, as illustrated in Fig.8 , the structured illumination images were used as inputs and the HiLo-reconstructed optical sections served as supervision for end-to-end learning. Since these HiLo-generated optical sections are used as reference supervision, the goal of the network is to approximate HiLo-quality

optical sectioning from a single structured-illumination image. The encoder extracts multi-scale features and enhances weakly modulated fringes via SE modules; the bottleneck incorporates the Transformer module to model global spatial dependencies; and the decoder progressively restores spatial resolution with skip connections to predict high-quality optical sections[27] [29][30].

During reconstruction, an optical section is obtained directly from a single structured illumination image. The normalized input image is fed into the network, which produces the corresponding optical section through encoder feature extraction, Transformer-based global modeling, and decoder reconstruction. Subsequently, the axial peak position at each pixel is determined from the optical-section stack acquired during axial scanning using the ARS algorithm, thereby recovering the 3D surface topography. This approach substantially reduces the multi-frame demodulation and computational burden required by conventional HiLo reconstruction. The reconstruction accuracy and computational efficiency of the proposed method are quantitatively evaluated in the following experimental section[28][31][32].

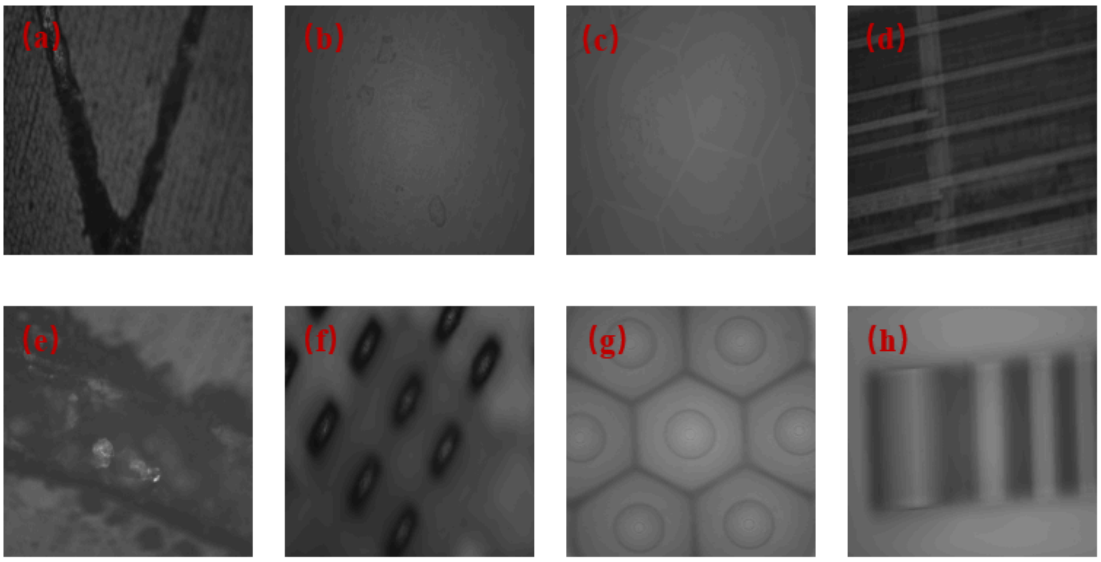

Figure 7. (a)–(h) show structured illumination images of (a) transparent glass, (b) a compound-eye lens, (c) a microlens array, (d) a micro-/nano-patterned chip, (e) a large-area chip, (f) a free-form lens, and (g) a cylindrical lens, respectively

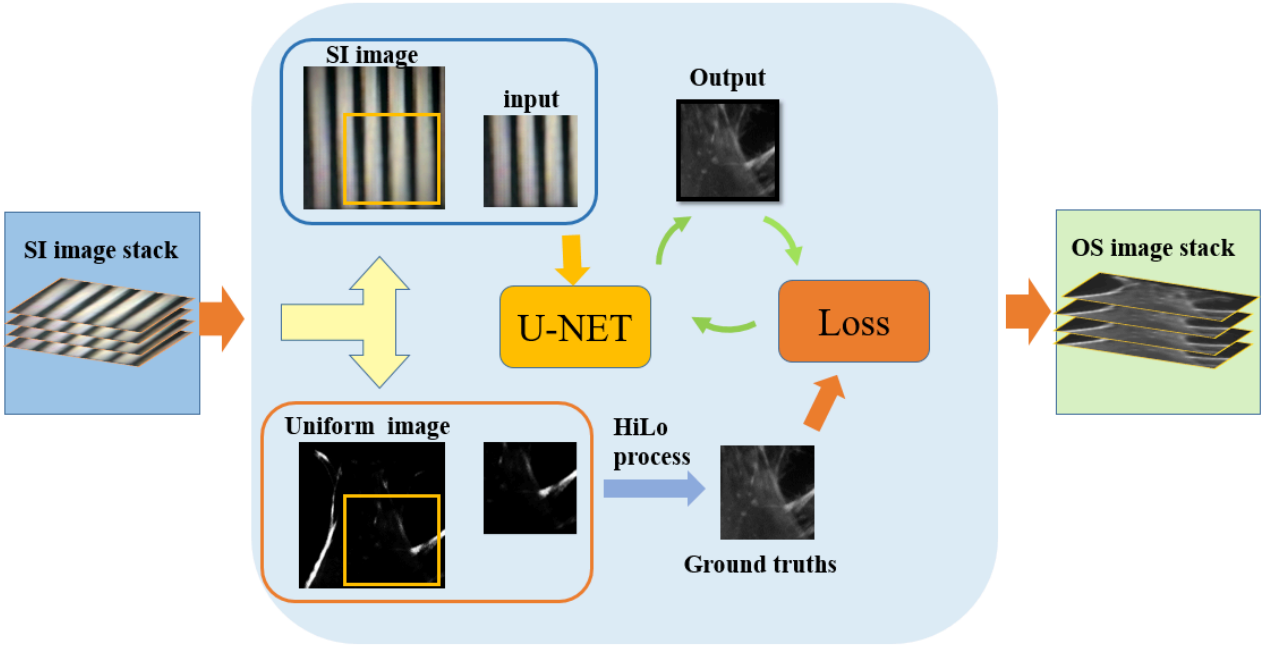

Figure 8. Training pipeline of the proposed OS-SIOS method. Structured illumination image stacks are paired with optical section images generated by the HiLo algorithm as reference labels for end-to-end network training

Quantitative Evaluation of Network Performance and Computational Efficiency

To validate the effectiveness of the proposed single-frame reconstruction method for 3D surface topography inspection of transparent and semi-transparent materials, we trained and compared four networks under the same training data and hyperparameter settings: the baseline SIOS, SE-SIOS (SIOS with only the SE module), Transformer-SIOS (SIOS with only the Transformer Bottleneck), and the full model OS-SIOS that integrates both SE and the Transformer Bottleneck and further employs Frequency Perturbation Augmentation (FPA). Specifically, SIOS contains no structural enhancement modules; SE-SIOS and Transformer-SIOS are used to evaluate the contribution of each individual module; and OS-SIOS represents the complete model[25][26][29][33][30].

To quantitatively assess reconstruction quality, we use mean absolute error (MAE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR). MAE measures the pixel-wise absolute difference between the predicted image and the reference reference labels:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}| \hat{Y}_i - Y_i | \tag{15}$$

where $\hat{Y}_i$ is the predicted value at the i -th pixel, Y_i is the corresponding reference labels value, and N is the total number of pixels. A smaller MAE indicates a reconstruction closer to the reference labels.

PSNR is used to evaluate reconstruction fidelity and is defined as

$$\mathrm{PSNR} = 10 \log_{10} \left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}} \right) \tag{16}$$

where

$$\mathrm{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{Y}_i - Y_i)^2 \tag{17}$$

and MAX denotes the maximum possible pixel value (set to 1 for normalized images). A higher PSNR indicates better reconstruction quality.

SSIM evaluates the similarity between two images in terms of luminance, contrast, and structural information:

$$\mathrm{SSIM}\ (x, y) = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{18}$$

where μ_x and μ_y are the mean intensities of the predicted and ground-truth images, σ_x^2 and σ_y^2 are their variances, σ_{xy} is the covariance, and C_1 and C_2 are stabilizing constants to avoid division by zero. SSIM ranges from 0 to 1, with values closer to 1 indicating higher structural similarity.

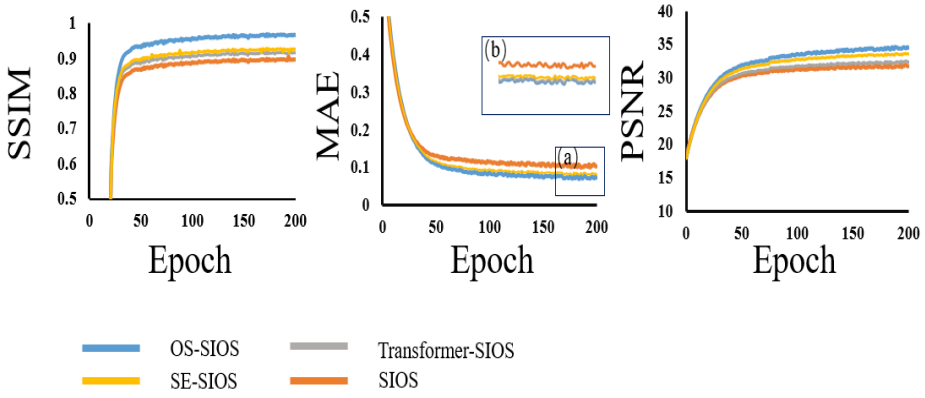

Figure 9. Training curves of reconstruction quality metrics for four networks: SIOS, SE-SIOS, Transformer-SIOS, and OS-SIOS. The inset in the MAE plot provides a magnified view of the highlighted region for clearer comparison of the converged errors

The MAE, SSIM, and PSNR curves of the four networks during training are shown in Fig.9. Compared with the baseline SIOS (SSIM = 0.897, MAE = 0.108 μm, PSNR = 31.8 dB), introducing the SE module improves SE-SIOS to SSIM = 0.927, MAE = 0.082 μm, and PSNR = 34.1 dB, indicating that channel attention strengthens weak-fringe representations and improves the effective signal-to-noise ratio. With only the Transformer Bottleneck, Transformer-SIOS achieves SSIM = 0.913, MAE = 0.074 μm, and PSNR = 32.4 dB; notably, the larger reduction in MAE suggests that global dependency modeling helps reduce overall prediction errors. By integrating both structural enhancements, OS-SIOS achieves the best performance (SSIM = 0.968, MAE = 0.069 μm, PSNR = 35.1 dB), demonstrating the complementary benefits of channel enhancement and global structural constraints and leading to substantially improved reconstruction accuracy and stability.

To evaluate computational efficiency, we measured the single-frame reconstruction time of different methods at the same input resolution (512×512). The conventional HiLo method requires approximately 135 ms per frame on CPU for optical section reconstruction. In contrast, deep learning models eliminate iterative computation during inference, resulting in markedly reduced computational complexity. The baseline SIOS achieves an average inference time of 14.5 ms, while SE-SIOS, Transformer-SIOS, and OS-SIOS require 15.8 ms, 15.6 ms, and 16.2 ms, respectively (Tab 1). The added overhead relative to SIOS is below 12% for all enhanced models, whereas OS-SIOS simultaneously improves SSIM to 0.968. These results indicate that the proposed method significantly improves reconstruction quality with only an additional ~1.7 ms inference cost. Although OS-SIOS is slightly slower than the baseline SIOS (16.2 ms vs 14.5 ms), it still supports a processing rate of approximately 61.7 frames per second. Given the substantial gain in reconstruction quality (SSIM: 0.897 to 0.968), this trade-off is considered acceptable for real-time or near-real-time online inspection scenarios in which reconstruction fidelity is a priority [27,33].

Table 1. Comparison of inference time and SSIM for HiLo, SIOS, SE-SIOS, Transformer-SIOS, and OS-SIOS

Method	Inference Time (ms)	SSIM
HiLo	135	0.870
SIOS	14.5	0.897
SE-SIOS	15.8	0.927
Transformer-SIOS	15.6	0.913
OS-SIOS	16.2	0.968

Table 1 Inference time denotes the average runtime per 512×512 input frame, and SSIM denotes the structural similarity index. The HiLo runtime corresponds to optical-section reconstruction on CPU, whereas the inference times of the deep learning models (SIOS, SE-SIOS, Transformer-SIOS, and OS-SIOS) are averaged measurements obtained on the same workstation during inference.

Validation Experiments

To further validate the effectiveness and accuracy of the proposed method for surface topography reconstruction of transparent and semi-transparent materials, we selected a standard high-precision grating sample with a line density of 100 lines/mm and a nominal height of 7μm as the test specimen. As shown in Fig.10, a single-frame structured illumination image of the standard high-precision grating was acquired using the OS-SIM system, and the corresponding optical section was reconstructed using an improved HiLo algorithm. In the experiments, the proposed OS-SIOS was compared with the baseline SIOS and the conventional HiLo reconstruction method [25,17]. In this study, the conventional HiLo-based 3D reconstruction method was adopted as the reference for comparison, primarily because of its widespread use and its strong capability for high-quality 3D surface morphology reconstruction. The conventional HiLo 3D reconstruction pipeline consists of two main steps. First, uniform-illumination and fringe-illumination images are acquired at each axial position to perform optically sectioned demodulation. Then, ARS-based peak localization is carried out pixel by pixel along the axial direction on the reconstructed optically sectioned image stack, thereby yielding the 3D height morphology.

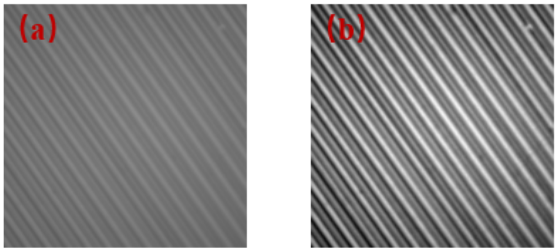

Figure 10. Representative images of the standard high-precision grating sample. (a) Single-frame structured illumination fringe image acquired by the OS-SIM system (network input). (b) Corresponding optical section reconstructed using an improved HiLo algorithm (supervision label)

As shown in Fig.10, HiLo preserves the fringe structure well but suffers from reduced contrast in low-reflectance regions. The SIOS results exhibit blurred textures and local distortion of periodic structures. In contrast, OS-SIOS more stably restores fringe continuity and edge details, producing visual results that are highly consistent with those of HiLo.

Furthermore, the reconstructed 3D surface maps (Fig.10) indicate that both HiLo and OS-SIOS accurately recover the periodic groove structure and overall surface flatness of the grating, whereas SIOS shows pronounced height fluctuations due to noise. For quantitative evaluation, Fig.11 provides a comparison of height profiles extracted from the reconstructed surfaces. The OS-SIOS profile closely matches the HiLo profile, with only minor deviations (< 0.1 μm) in local high-frequency regions. By contrast, the SIOS profile shows noticeable over-smoothing and groove misidentification, reflecting loss of structural information.

Overall, the SE module enhances feature representations associated with weakly modulated fringes through channel recalibration, thereby effectively improving the effective signal-to-noise ratio under low-contrast conditions and increasing the stability of ARS-based peak localization. Meanwhile, the TB bottleneck models cross-region dependencies through global self-attention, strengthens global phase consistency, and significantly suppresses spatial distortion caused by local fringe phase shifts. As a result, OS-SIOS achieves reconstruction quality close to that of HiLo-based reconstruction, while retaining the efficiency and robustness of deep learning. These results indicate that the proposed method provides a practical, high-throughput solution for real-time 3D surface morphology reconstruction of transparent and translucent materials.

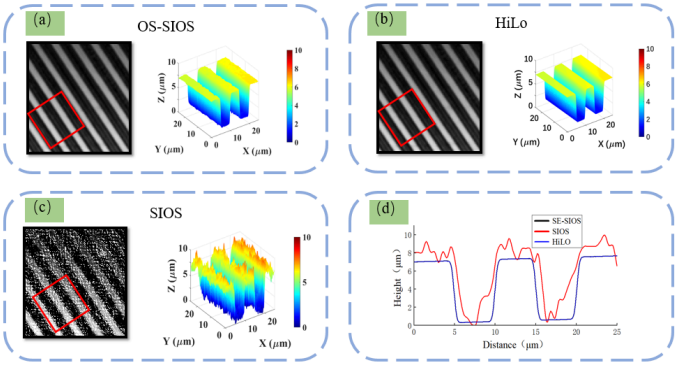

Figure 11. Reconstruction results of a standard grating sample. Maximum intensity projection images, three-dimensional surface reconstructions, and height profiles obtained using the HiLo method, SIOS, and OS-SIOS

To evaluate the robustness of the modelunder more challenging weakly scattering conditions, we further tested a ground-glass sample with typical random roughness, As shown in Fig.12, a single-frame structured illumination image of the ground-glass sample was acquired using the OS-SIM system, and the corresponding optical section was reconstructed using an improved HiLo algorithm. Compared with a standard grating or smooth glass, the random microstructures on ground glass cause phase randomization and local spatial-frequency drift of the incident fringes during propagation. Specifically, small fluctuations in surface normals lead to variations in fringe period, accompanied by reduced modulation depth and increased random scattering noise. These effects markedly weaken fringe modulation in SIM, providing a representative non-ideal imaging scenario for assessing model robustness and generalization.

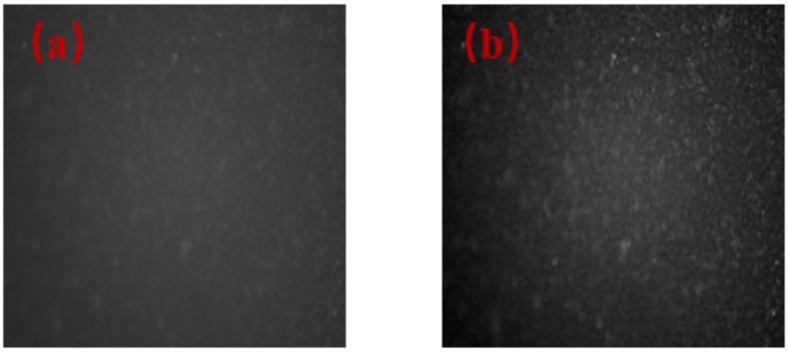

Figure 12. Representative images of the ground-glass sample. (a) Single-frame structured illumination image acquired by the OS-SIM system (network input). (b) Corresponding optical section reconstructed using an improved HiLo algorithm

As shown in Fig.13(a), the MIP results of SIOS exhibit pronounced texture loss and overall over-smoothing. HiLo preserves the overall distribution of random textures, although local blurring remains in high-frequency detail regions. OS-SIOS shows clearly improved texture preservation and detail clarity compared with the baseline, and its results are more consistent with those of HiLo. Further inspection of the top-view reconstruction and 3D surface maps (Fig.13(b,c)) indicates that OS-SIOS captures the random height fluctuations of the ground-glass surface more accurately, whereas SIOS produces noticeable distortions and tends to underestimate surface roughness.

For quantitative evaluation, we computed the areal surface roughness parameters—arithmetical mean height Sa and root-mean-square height Sq—according to ISO 25178 [34,35,36]. As summarized in Tab 2, HiLo yields Sa=0.482 μm and Sq=0.617 μm. The reconstruction results of SE-SIOS are Sa=0.469 μm and Sq=0.598 μm, corresponding to relative errors of 2.7% and 3.1%, respectively. By contrast, SIOS yields Sa=0.438 μm and Sq=0.551 μm, with relative errors of 9.1% and 10.7%, which are markedly higher than those of OS-SIOS. These results indicate that OS-SIOS can achieve roughness estimation accuracy close to HiLo for randomly rough surfaces.

Table 2. Comparison of roughness estimation accuracy for the ground-glass surface reconstructed by HiLo, OS-SIOS, and SIOS

Method	Sa (μm)	Sq (μm)	Relative Error of Sa	Relative Error of Sq
HiLo(reference)	0.482	0.617	—	—
OS-SIOS	0.469	0.598	2.7%	3.1%
SIOS	0.438	0.551	9.1%	10.7%

Table 2 Surface Roughness Parameters

Overall, the results indicate that OS-SIOS is capable of maintaining both global structural consistency and local texture accuracy in the reconstruction of randomly rough surfaces. The SE-based channel enhancement mechanism effectively reinforces weakly modulated signal features, whereas the Frequency Perturbation Augmentation (FPA) strategy improves robustness to fringe period variation. Consequently, the proposed network can still deliver high-fidelity surface morphology reconstruction under low-SNR and non-ideal imaging conditions, demonstrating the robustness and generalization advantage of OS-SIOS for complex surface measurement in both transparent and weakly scattering/translucent samples under similar low-modulation conditions [37].

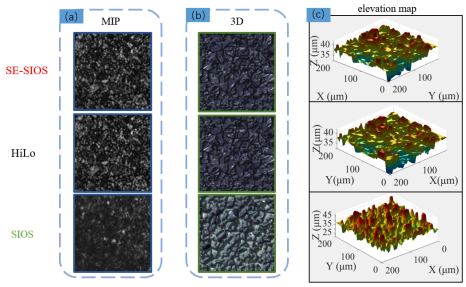

Figure 13. Reconstruction results of a frosted glass sample with random surface roughness. (a) Maximum intensity projection images, (b) top-view height maps, and (c) three-dimensional surface reconstructions obtained using different methods

CONCLUSION

In conclusion, we propose OS-SIOS, a single-frame structured illumination microscopy framework for 3D surface morphology reconstruction of transparent and translucent materials. By incorporating an SE channel attention module, a lightweight TB module, and the Frequency Perturbation Augmentation (FPA) strategy into a U-Net backbone, the proposed method improves the recovery of weakly modulated fringes, enhances optically sectioned reconstruction, and provides bettergeneralization for complex surface measurement. Because the framework is designed to address shared degradation patterns in reflective OS-SIM, including weak fringe modulation, phase perturbation, and fringe-frequency mismatch, it can serve as a unified reconstruction model for transparent and weakly scattering/translucent samples that exhibit similar degraded fringe observations. Experimental validation shows that, compared with the baseline SIOS network, OS-SIOS significantly improves reconstruction performance, increasing SSIM from 0.897 to 0.968 and reducing MAE from 0.108 μm to 0.069 μm, while maintaining a single-frame inference time of 16.2 ms. Moreover, it offers substantially higher computational efficiency than the HiLo-based 3D reconstruction pipeline, which requires 135 ms to achieve a comparable reconstruction quality. Overall, the proposed method achieves a practical balance between reconstruction fidelity and inference efficiency for real-time or near-real-time OS-SIM applications. These results demonstrate that OS-SIOS is a promising solution for high-throughput, non-contact 3D surface morphology reconstruction of transparent materials under low-modulation imaging conditions, and provides a feasible pathway for real-time OS-SIM applications. A limitation of the present study is that the experimental validation does not explicitly cover multilayer IC-packaging structures with strong internal reflections or ghost fringes. Although chip-related samples were included in the training set, dedicated validation under such multilayer interference conditions remains for future investigation.

Author Contributions

Conceptualization – Liqing Wan, Zili Lei and Wei Shen Z; methodology – Liqing Wan, Zili Lei and Wei Shen Z; formal analysis – Liqing Wan, Zili Lei and Wei Shen Z; investigation – Liqing Wan, Zili Lei and Wei Shen Z; resources – Liqing Wan, Zili Lei and Wei Shen Z; writing-original draft preparation – Liqing Wan; writing-review and editing – Liqing Wan, Zili Lei and Wei Shen Z; visualization – Liqing Wan; supervision – Liqing Wan, Zili Lei and Wei Shen Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

Funding

The National Natural Science Foundation of China (No.12302239). Automatic Inspection and Quality Evaluation of Aerospace Structural Sealing Rubber Components (JCZRLH202500460).

Acknowledgements

Not applicable.

REFERENCES

[1] Gustafsson MG. Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. J Microsc. 2000;198(Pt 2):82-87. doi: 10.1046/j.1365-2818.2000.00710.x

[2] Saxena M, Eluru G, Gorthi SS. Structured illumination microscopy. Adv Opt Photon. 2015;7:241-275. doi: 10.1364/AOP.7.000241

[3] Kutulakos KN, Steger E. A theory of refractive and specular 3D shape by light-path triangulation. Int J Comput Vis. 2008;76:13-29. doi: 10.1007/s11263-007-0049-9

[4] Murase H. Surface shape reconstruction of a nonrigid transport object using refraction and motion. IEEE Trans Pattern Anal Mach Intell. 1992;14(10):1045-1052. doi: 10.1109/34.159906

[5] Qian Y, Gong M, Yang YH. 3D reconstruction of transparent objects with position-normal consistency. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016. p. 4369-4377. doi: 10.1109/CVPR.2016.473

[6] Wetzstein G, Roodnick D, Heidrich W, Raskar R. Refractive shape from light field distortion. In: 2011 International Conference on Computer Vision; 2011. p. 1180-1186. doi: 10.1109/ICCV.2011.6126367

[7] Olivieri L, Totero Gongora JS, Peters L, Cecconi V, Cutrona A, Tunesi J, et al. Hyperspectral terahertz microscopy via nonlinear ghost imaging. Optica. 2020;7:186-191. doi: 10.1364/OPTICA.381035

[8] Chen X, Zhong S, Hou Y, et al. Superresolution structured illumination microscopy reconstruction algorithms: a review. Light Sci Appl. 2023;12:172. doi: 10.1038/s41377-023-01204-4

[9] Mertz J, Kim J. Scanning light-sheet microscopy in the whole mouse brain with HiLo background rejection. J Biomed Opt. 2010;15(1):016027. doi: 10.1117/1.3324890

[10] Catalucci S, Thompson A, Moroni G, et al. Optical metrology for digital manufacturing: a review. Int J Adv Manuf Technol. 2022;120:4271-4290. doi: 10.1007/s00170-022-09084-5

[11] Chai C, Chen C, Liu X, Lei Z. Deep learning based one-shot optically-sectioned structured illumination microscopy for surface measurement. Opt Express. 2021;29(3):4010-4021. doi: 10.1364/OE.415210

[12] Hu J, Shen L, Sun G. Squeeze-and-Excitation Networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018. p. 7132-7141. doi: 10.1109/CVPR.2018.00745

[13] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale. arXiv. 2020. Available from: https://arxiv.org/abs/2010.11929

[14] Hu G, Greene J, Zhu J, Yang Q, Zheng S, Li Y, et al. HiLo microscopy with caustic illumination. Biomed Opt Express. 2024;15:4101-4110. doi: 10.1364/BOE.527264

[15] Rudolf B, Du Y, Turtaev S, Leite IT, Čižmár T. Thermal stability of wavefront shaping using a DMD as a spatial light modulator. Opt Express. 2021;29:41808-41818. doi: 10.1364/OE.442284

[16] Zhao T, Wang Z, Chen T, Lei M, Yao B, Bianco PR. Advances in high-speed structured illumination microscopy. Front Phys. 2021;9:672555. doi: 10.3389/fphy.2021.672555

[17] Zhou X, Lei M, Dan D, Yao B, Qian J, et al. Double-exposure optical sectioning structured illumination microscopy based on Hilbert transform reconstruction. PLoS One. 2015;10(3):e0120892. doi: 10.1371/journal.pone.0120892

[18] Qian J, Lei M, Dan D, et al. Full-color structured illumination optical sectioning microscopy. Sci Rep. 2015;5:14513. doi: 10.1038/srep14513

[19] Wicker K. Non-iterative determination of pattern phase in structured illumination microscopy using auto-correlations in Fourier space. Opt Express. 2013;21:24692-24701. doi: 10.1364/OE.21.024692

[20] Stokseth PA. Properties of a defocused optical system. J Opt Soc Am. 1969;59(10):1314-1321. doi:10.1364/JOSA.59.001314

[21] Smith CS, Slotman JA, Schermelleh L, et al. Structured illumination microscopy with noise-controlled image reconstructions. Nat Methods. 2021;18:821–828. doi:10.1038/s41592-021-01167-7

[22] Feng L, Wang X, Sun X, et al. Efficient multifocal structured illumination microscopy utilizing a spatial light modulator. Appl Sci. 2020;10(12):4396. doi:10.3390/app10124396

[23] Lauterbach MA, Ronzitti E, Sternberg JR, Wyart C, Emiliani V. Fast calcium imaging with optical sectioning via HiLo microscopy. PLoS One. 2015;10(12):e0143681. doi: 10.1371/journal.pone.0143681

[24] Li Z, et al. Fast widefield imaging of neuronal structure and function with optical sectioning in vivo. Sci Adv. 2020;6:eaaz3870. doi: 10.1126/sciadv.aaz3870

[25] Lim D, Ford TN, Chu KK, Mertz J. Optically sectioned in vivo imaging with speckle illumination HiLo microscopy. J Biomed Opt. 2011;16(1):016014. doi: 10.1117/1.3528656

[26] Waller L, Tian L. Machine learning for 3D microscopy. Nature. 2015;523:416-417. doi:10.1038/523416a

[27] Rivenson Y, Göröcs Z, Günaydin H, et al. Deep learning microscopy. Optica. 2017;4(11):1437-1443. doi:10.1364/OPTICA.4.001437

[28] Ling C, Zhang C, Wang M, Meng F, Du L, Yuan X. Fast structured illumination microscopy via deep learning. Photon Res. 2020;8:1350-1359. doi: 10.1364/PRJ.396122

[29] Qiao C, Chen X, Zhang S, Li D, Guo Y, Dai Q, Li D. 3D structured illumination microscopy via channel attention generative adversarial network. IEEE Journal of Selected Topics in Quantum Electronics. 2021;27(4):1-11. doi: 10.1109/JSTQE.2021.3060762

[30] Wang J, Fan J, Zhou B, Huang X, Chen L. Hybrid reconstruction of the physical model with the deep learning that improves structured illumination microscopy. Adv Photon Nexus. 2023;2(1):016012. doi: 10.1117/1.APN.2.1.016012

[31] Wu Y, Rivenson Y, Wang H, et al. Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning. Nat Methods. 2019;16:1323-1331. doi:10.1038/s41592-019-0622-5

[32] Christensen CN, Ward EN, Lu M, Liò P, Kaminski CF, et al. ML-SIM: universal reconstruction of structured illumination microscopy images using transfer learning. Biomed Opt Express. 2021;12(5):2720-2733. doi: 10.1364/BOE.414680

[33] Zhuge H, Summa B, Hamm J, Brown JQ. Deep learning 2D and 3D optical sectioning microscopy using cross-modality Pix2Pix cGAN image translation. Biomed Opt Express. 2021;12(12):7526-7543. doi: 10.1364/BOE.439894

[34] ISO 25178-2:2021. Geometrical product specifications (GPS) - Surface texture: Areal - Part 2: Terms, definitions and surface texture parameters. Available from: https://www.iso.org/standard/74591.html

[35] Pawlus P, Reizer R, Wieczorowski M. Functional importance of surface texture parameters. Materials. 2021;14:5326. doi: 10.3390/ma14185326

[36] Buchenau T, Mertens T, Lohner H, Bruening H, Amkreutz M. Comparison of optical and stylus methods for surface texture characterisation in industrial quality assurance of post-processed laser metal additive Ti-6Al-4V. Materials. 2023;16:4815. doi: 10.3390/ma16134815

[37] Maurya AK, Chatterjee K, Jha R. Ultra-wide range non-contact surface profilometry based on reconfigurable fiber interferometry. Opt Lett. 2024;49:3588-3591. doi: 10.1364/OL.531327