Optics & Laser Technology 198 (2026) 114856

Contents lists available at ScienceDirect

Optics and Laser Technology

journal homepage: www.elsevier.com/locate/jolt

Full length article

Structure-guided lensless reconstruction via physics-aware decomposition in low-light conditions

Ziyang Liu [a,c], Tianjiao Zeng [b,c,*], Xu Zhan [a], Xiaoling Zhang [a], Yunqi Wang [d], Edmund Y. Lam [e]

[a] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China
[b] School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China
[c] National Laboratory on Adaptive Optics, Chengdu, 610209, China
[d] Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02215, United States of America
[e] Department of Electrical and Electronic Engineering, The University of Hong Kong, Pokfulam, Hong Kong, China

ARTICLE INFO

Keywords:
Low-light
Lensless imaging
Diffusion enhancement
Component extraction

ABSTRACT

Lensless imaging is promising for miniature applications owing to its compact, lightweight, and low-cost design. However, reconstruction quality severely degrades under low-light conditions, where the pervasive interference between structural information and measurement residuals (e.g., noise and brightness variations) poses a critical challenge for existing methods to recover clean, high-fidelity details. To address this, our work fundamentally revisits the lensless measurement paradigm, introducing a novel structure-guided model for low-light lensless imaging that embeds an explicit structure–residual decomposition within the forward process. This formulation breaks with the convention of holistic scene treatment by redefining the scene as a combination of intrinsic structural and residual components, enabling precise and targeted component-wise processing to achieve enhanced fidelity and noise reduction. Building on this model, we develop a two-stage physics-aware reconstruction approach: (1) a multilevel, multiscale extraction module embedded with the forward model initially extracts the components from the measurements, reducing noise impact on the structure through feature extraction across multiple scales and levels; (2) conditional diffusion modules, trained bidirectionally for stability, refine structures and optimize residuals generatively to boost detail recovery before fusion. Experiments on low-light datasets from our custom lensless camera (22,000 images with phase/amplitude masks) show that our approach outperforms state-of-the-art methods in both objective assessments and visual inspection, validating its advantages in denoising, structural fidelity, and overall image quality.

1. Introduction

Optical imaging systems are widely used in photography and microscopy, but traditional lens-based systems are constrained by physical limitations such as diffraction, making miniaturization, weight reduction, and cost optimization difficult [1]. Lensless imaging overcomes these limitations by replacing traditional lenses with lightweight masks or coded apertures to modulate light, then reconstructing images using computational algorithms [2–4]. This approach frees the system from point-to-point mapping constraints, enabling thinner, more compact designs. It shows significant promise in fields such as biosensing, 3D imaging, computational photography, and visual diagnostics [5–9].

With the development of lensless imaging technology, high-quality image reconstruction has become one of the core challenges in this field [10–12]. This is mainly because its inverse problem is inherently highly ill-posed: different scenes may correspond to similar observation signals (non-uniqueness of solutions), while small noise or measurement errors can be amplified (instability of solutions), leading to a reconstruction process that lacks both stability and robustness. Early approaches primarily relied on classical optimization frameworks, such as Tikhonov regularization, Iterative Shrinkage-Thresholding Algorithm (ISTA), and Alternating Direction Method of Multipliers (ADMM), introducing manually designed priors to regularize the solution process [13–15]. While these model-driven methods offer physical interpretability, their imaging quality heavily depends on parameter tuning, and reconstruction degrades significantly in challenging scenarios like low-light conditions. Deep learning has advanced data-driven methods in

* Corresponding author at: School of Aeronautics and Astronautics, University of Electronic Science and Technology of China, Chengdu, 611731, Sichuan, China.
Email address: tzeng@uestc.edu.cn (T. Zeng).

https://doi.org/10.1016/j.optlastec.2026.114856
Received 1 September 2025; Received in revised form 20 December 2025; Accepted 28 January 2026
Available online 6 February 2026

various reconstruction tasks [16–19], improving quality through end-to-end mapping from measurements to images. However, these require large training datasets and significantly degrade on unseen distributions. Hybrid strategies, such as unrolled networks (e.g., Le-ADMM-U, FlatNet [20–23]), combine both, but still face issues like model mismatches. Compensation mechanisms, including multi-stage loss recovery, learnable primal-dual frameworks, multi-scale deconvolution, and generative integration [24–28], address some limitations. Although these methods succeed under adequate lighting, their performance declines in low-light conditions.

Under low-light conditions, the signal-to-noise ratio (SNR) drops sharply, significantly degrading image quality. This challenge is particularly severe for lensless imaging systems relying on optical encoding and computational reconstruction. Without optical focusing, such systems are more susceptible to SNR fluctuations. Furthermore, the sparse energy distribution of the encoded light field, combined with limited sensor sensitivity and background noise, creates observed entanglement between the scene's inherent content (e.g., textures, edges, colors) and external factors easily affected by the external environment (e.g., brightness, noise). We define the former as structural information and the latter as residual information.

This entanglement poses significant challenges to existing lensless reconstruction methods. First, these methods do not incorporate structural-residual decomposition in modeling the forward imaging process. Consequently, they lack the capability for targeted component-wise processing, often resulting in the loss of structural details during noise suppression. Second, residual optimization (e.g., brightness recovery) also remains inadequate, limiting overall quality. Unfortunately, research on extracting high-fidelity structure from severely noise-contaminated lensless measurements is very limited. Our prior work based on Wiener filtering and diffusion-based methods [29,30], while achieving noise suppression and brightness enhancement to some extent, falls short in fully extracting and utilizing intrinsic structure due to its holistic recovery strategy.

In this study, we address these issues by exploring high-fidelity structural reconstruction techniques for low-light lensless imaging, encompassing both the measurement modeling and solution strategies. To extract core structural information effectively, we develop a new lensless measurement model that decomposes the scene into two key components, structural (inherent textures, edges, colors) and residual (ambient brightness, noise) components. This establishes a clearer link between raw measurements and the scene's inherent content, effectively reframing the reconstruction task as a structure-centered decomposition problem.

To solve this model, we propose a physics-aware two-stage reconstruction method designed to simultaneously achieve high-fidelity structure extraction, noise suppression, and optimized brightness recovery. In this approach, the extracted structural and residual components are processed independently before recombination into the final high-quality image. The first stage employs a multi-level, multi-scale extraction module designed to decompose the measurements into distinct structural and residual components. Through multi-scale feature analysis and by fusing the forward physical model into various feature-learning levels, the module introduces the necessary constraints for effective separation. The second stage employs a generative conditional diffusion model to refine both components. It specifically enriches structural details while simultaneously suppressing noise and enhancing brightness within the residuals. Finally, the separately enhanced structural and residual components are fused to produce the final reconstruction. To ensure consistency between diffusion generation and physical constraints, we employ a bidirectional training strategy that integrates the forward and reverse processes for improved stability.

In summary, the contributions of this work are as follows.

1. We introduce a novel conceptual perspective on low-light lensless reconstruction, shifting from traditional holistic scene recovery to a dedicated processing paradigm of "structure guidance + residual optimization". This reformulation provides a new pathway to address structural degradation caused by noise and brightness variations in low-light measurements.
2. To realize this perspective, we propose a structure-guided lensless measurement model that explicitly decomposes the scene into intrinsic structural and external residual components within the forward process, enabling more precise extraction and alignment with physical imaging constraints.
3. We develop a physics-aware two-stage reconstruction framework based on this model, achieving high-fidelity structural recovery alongside noise suppression and brightness enhancement in residual components, resulting in superior overall image quality under low-light conditions.
4. We construct a lensless imaging prototype system and collect two real-world low-light datasets (phase and amplitude masks, each with 11,000 measurements and ground truths). Both qualitative and quantitative evaluations demonstrate that our method achieves state-of-the-art performance, with a PSNR gain of 0.83 dB, a 4.4 percentage-point improvement in SSIM, and leading performance on all remaining metrics.

2. Methodology

2.1. Structure-guided measurement model

The lensless imaging process can be modeled as a linear shift-invariant system, where the forward measurement is represented by the two-dimensional discrete convolution between the scene intensity distribution and the system's point spread function (PSF):

$$\mathbf{y} = \mathbf{H} * \mathbf{x} + \mathbf{n} \tag{1}$$

where $\mathbf{y}$ denotes the measurement acquired by the lensless imaging system, $*$ represents the 2D discrete convolution operator, $\mathbf{H}$ is the system's forward model (i.e., the PSF), $\mathbf{x}$ is the scene intensity distribution, and $\mathbf{n}$ denotes the additive measurement noise.

In actual lensless imaging, we often observe that measurement signals do not come from a single source, but are jointly constituted by two types of components: one is the structural information inherent to the scene, such as textures, edges, and colors, while the other primarily manifests as residual interference information including environmental light changes and noise. If these two are modeled together in a mixed manner, noise suppression often accompanies the weakening of structural details, making it difficult to balance noise reduction and fidelity in reconstruction results. Especially under low-light conditions, signal attenuation and noise enhancement further exacerbate this strong coupling between structural and residual components, causing the overall modeling to face a dilemma between "noise suppression and distortion." It is evident that reasonably separating measurement signals into structural and residual components not only helps maintain structural stability while suppressing external interference but also enhances the accuracy of subsequent extraction and reconstruction. However, existing lensless reconstruction methods generally lack explicit modeling of this difference, often treating the two as a single entity, resulting in limitations in noise suppression, brightness optimization, and structure recovery. Particularly under complex noise conditions in low light, this deficiency is further intensified, posing significant challenges for high-fidelity structural reconstruction. This also highlights the necessity of constructing new measurement models that can simultaneously account for structure preservation and noise suppression.

To overcome the aforementioned limitations, we propose a measurement model for lensless imaging that integrates a system convolution degradation model with the joint representation of structural and residual components, as illustrated in Fig. 1. The proposed formulation characterizes the imaging process as:

$$\mathbf{y} = \mathbf{H} * [\mathbf{S}(\mathbf{x}) \odot \mathbf{R}(\mathbf{x})] + \mathbf{n} \tag{2}$$

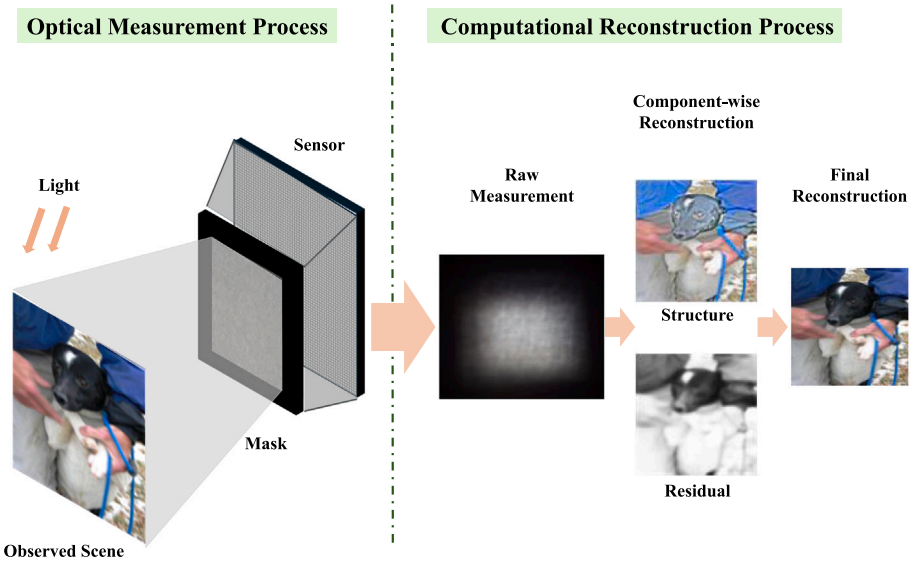

Fig. 1. Illustration of the structure residual decomposition-based lensless measurement model. Left: Optical measurement process, where the inherent structural information and external residual information from the scene are modulated by the mask and captured by the sensor, generating raw measurements. Right: Computational reconstruction process, where the structure and residual components are extracted separately and recombined to obtain the final high-fidelity reconstruction.

where $\mathbf{S}(\mathbf{x})$ represents the structural component, containing inherent information such as scene textures, edges, and colors, $\mathbf{R}(\mathbf{x})$ represents the residual component, containing external information such as ambient brightness and noise, and $\odot$ denotes the Hadamard product, i.e., element-wise multiplication of these components. Here we adopt a multiplicative interaction between the structural and residual components, because illumination and noise variations modulate the underlying structure rather than simply adding an independent component on top. In contrast, an additive form would treat the two components as loosely coupled and easily separable, which is inconsistent with their strong coupling in practice and leads to a less faithful description of the imaging process.

Unlike existing methods, our approach no longer views the reconstructed scene as a single entity, but instead establishes a clearer correspondence between the measurements and the scene's inherent information through modeling, thus transforming the reconstruction task into a decomposition problem centered on structure extraction. Specifically, we directly extract $\mathbf{S}(\mathbf{x})$ and $\mathbf{R}(\mathbf{x})$ from the measurement $\mathbf{y}$, perform targeted enhancement optimization on them, and finally combine the two components to obtain the reconstruction result of the scene information. This approach better aligns with the physical imaging process and can more effectively separate and enhance the structural and residual components, thereby improving the quality and stability of image reconstruction.

2.2. Two-stage physic-aware approach

Based on our established lensless measurement model, we design a physics-aware two-stage approach for reconstruction, including initial extraction and fine enhancement (Fig. 2). This method not only achieves high-fidelity structural component reconstruction, but also simultaneously performs noise suppression and brightness optimization for the residual component. Our approach effectively extracts structural information and improves component quality through physics-aware mechanisms. In the first stage, we design a multi-level multi-scale structure-residual dual-branch extraction module to separately extract structure and residual components from measurements. This module extracts global structural information through conditional transformer

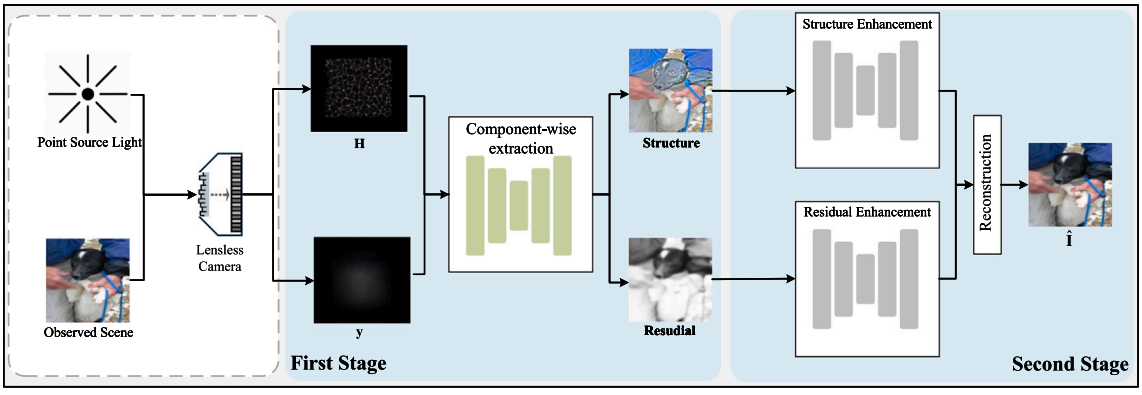

Fig. 2. Overview of the proposed physics-aware two-stage reconstruction method. In the first stage, structure and residual components are extracted from raw measurements through multi-scale multi-level extraction modules, guided by the physical model fusion module. In the second stage, these two components are enhanced and optimized through enhancement modules based on conditional diffusion models, and then combined to obtain the final reconstruction result, achieving high-quality imaging under low-light conditions.

layers in the upper branch, while gradually extracting residual information using convolutional layers in the lower branch. This design strengthens structural components through long-range dependencies, while residual components are effectively extracted through local convolution processing. Meanwhile, the physical model fusion block integrates physical prior information into features at different levels, enhancing the network's physical constraints.

In the second stage, we introduce enhancement modules based on conditional diffusion models. These modules adopt identical diffusion network structures for both structure and residual components, but are independently trained for their respective components: the structure enhancement model focuses on recovering missing or blurred structural information, while the residual enhancement model primarily suppresses noise and optimizes brightness levels. Through this component-specific enhancement strategy, the two diffusion models fully function within their respective subspaces, ultimately collaborating to generate high-quality reconstruction results. During training, we simultaneously optimize forward diffusion (learning the conditional distribution of structure and residual representations transitioning to noise states) and introduce reverse diffusion as a constraint. This ensures that generated results maintain consistency with real measurements during the denoising process, improving reconstruction stability and reducing random variations. In the following sections, we will introduce the modules of the first and second stages separately.

2.2.1. Stage 1: multi-level multi-scale component extraction

This section is divided into two parts: the extraction network structure overview and the physical model fusion mechanism.

(1) Extraction Network Structure Overview.

We construct a multi-level, multi-scale parallel dual-branch network to separately extract structure and residual components from the original measurement **y**, with its overall architecture shown in Fig. 3. The structure component contains inherent geometric textures, edge contours, and color distributions of the scene, requiring higher demands for global consistency and multi-scale context modeling. Therefore, the structure branch adopts a structure stacked with conditional transformer blocks to capture long-range dependencies and effectively represent key structural features in a multi-level, multi-scale manner. In contrast, the residual component is mainly composed of external factors, including environmental brightness fluctuations and noise interference, with information that is more localized and high-frequency, not requiring complex global modeling. Thus, the residual branch adopts a lightweight convolutional structure, more suitable for characterizing local details and noise patterns. The two branches differ in complexity design based on their target component characteristics, but both incorporate a physical model fusion mechanism at their core positions to ensure the network output remains consistent with the measurement model, avoiding artifacts that don't conform to the actual physical process.

Structure branch. To reconstruct the inherent structural information in the scene that does not change with external conditions (such as texture, edges, and color details), the structure component extraction branch uses Conditional Transformer Blocks (CTB) as core units, capturing key structural details at different feature resolutions through multi-scale attention mechanisms. Each CTB consists of Multi-Dconv Head Transposed Attention (MDTA), Gated-Dconv Feed-Forward Network (GDFN) [31], and Physical Model Fusion Mechanism (PMFM). MDTA performs depth convolution and transposed attention in parallel on multi-scale feature maps, effectively combining local spatial context with global dependencies, enhancing response to inherent structures such as textures and edges; GDFN introduces gating mechanisms and depth convolution paths in the feed-forward process, enhancing non-linear modeling capabilities of features and reducing cross-layer information decay; PMFM introduces prior constraints based on imaging physics in feature updates, ensuring the extracted structure component maintains high-resolution details under physical consistency. The structure branch, through cross-layer skip connections, transmits low-level texture and edge information to high-level semantic features, ultimately forming a structural component representation that reflects the scene's inherent properties, providing a reliable foundation for subsequent enhancement and fusion.

Residual branch. To capture residual information introduced by the external environment and imaging conditions (such as environmental brightness changes, noise interference, etc.), the residual component extraction branch adopts a lightweight convolutional network structure, focusing on high-frequency and local detail features with lower computational complexity. Through shallow convolutional units, this branch can effectively characterize unstable brightness variations and random noise distributions in the image. Similar to the structure branch, a physical model fusion mechanism is also introduced at the core position to ensure residual information is effectively recovered under physical model constraints. Ultimately, the residual component output by this branch provides a reliable foundation for subsequent denoising and brightness enhancement.

(2) Physical Model Fusion Mechanism.

As mentioned in the previous section, the structure component extraction branch and residual component extraction branch both incorporate a Physical Model Fusion Mechanism (PMFM) at their core. This mechanism serves the same basic function in

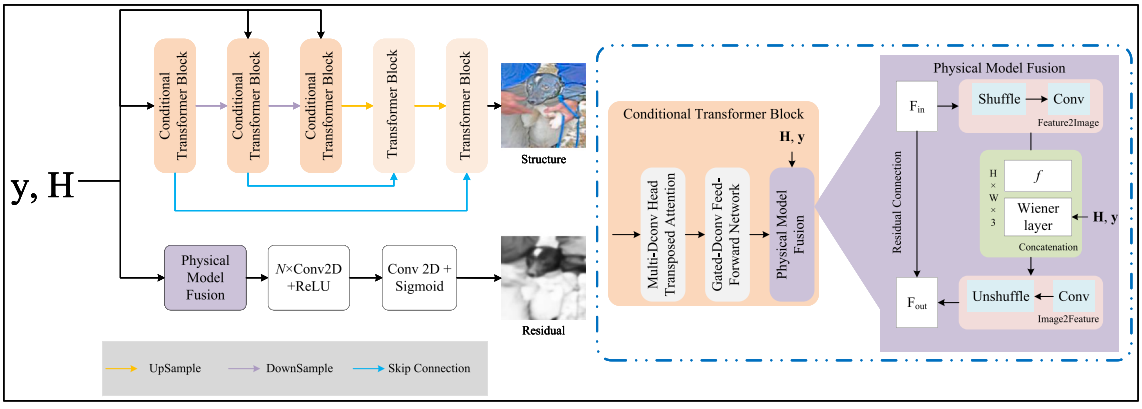

Fig. 3. Architecture of the proposed multi-level multi-scale component extraction module. The left side shows the main extraction process for structural and residual components, while the right side details the physical model fusion module, which integrates physical priors and learned features through a Wiener filter-based combination mechanism.

both branches: utilizing physical measurement information (measurement matrix $\mathbf{H}$ and original measurement $\mathbf{y}$) to generate a physically-driven preliminary estimate, which is then mapped back to the feature domain for subsequent network utilization.

In implementation, PMFM first maps features to the image domain through a combination of pixel shuffle and convolution operations (i.e., Feature2Image), then obtains preliminary inversion results through Wiener filtering by combining $\mathbf{H}$ and $\mathbf{y}$:

$$\hat{\mathbf{S}}_{\text{init}} = \mathcal{F}^{-1}\left\{\mathcal{F}(\mathbf{y}) \odot \left(\frac{\mathcal{F}(\mathbf{H})^*}{\gamma + |\mathcal{F}(\mathbf{H})|^2}\right)\right\} \tag{3}$$

where $\hat{\mathbf{S}}_{\text{init}}$ represents the physically-driven preliminary estimate, $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ are the Fourier transform and inverse Fourier transform respectively, $(\cdot)^*$ denotes the complex conjugate, and γ is a regularization coefficient. This preliminary estimate can quickly generate imaging results in the presence of noise, containing both structural information of the scene (such as edges and textures) and inevitably carrying certain residual information (such as noise interference and brightness fluctuations). Therefore, it provides a physically-driven reference for subsequent feature extraction. To fully leverage its reference value, we design a fusion mechanism: concatenating the preliminary estimate $\hat{\mathbf{S}}_{\text{init}}$ with the image-domain features f generated by Feature2Image along the channel dimension. The concatenated result is mapped back to the feature domain through convolution and pixel unshuffle (i.e., Image2Feature) operations, ensuring that the fused features align with the residual connection branch in both spatial dimensions and feature dimensions. Finally, the network outputs enhanced features $\mathbf{F}_{\text{out}}$, achieving effective interaction between the physical model prior and deep features. This design enables the model to not only preserve the structural outline information provided by the preliminary estimate but also further recover details and suppress noise, significantly improving overall reconstruction quality.

In the structure component extraction branch, the network uses $\hat{\mathbf{S}}_{\text{init}}$ as a reference input, enhancing the expressive capability of detailed structures while ensuring overall coherence through multi-scale feature fusion, thereby achieving robust and comprehensive extraction of structural information. In the residual component extraction branch, the network similarly receives $\hat{\mathbf{S}}_{\text{init}}$ as a reference, but adaptively focuses on unstable high-frequency components (such as noise disturbances and lighting variations) through convolutional mapping, effectively separating and characterizing these components under physical prior constraints. PMFM provides unified physical prior input for both branches while maintaining differentiated feature utilization strategies for each branch.

2.2.2. Stage 2: conditional diffusion-based component enhancement

This section is divided into three parts: the forward/reverse diffusion process, the conditional enhancement of components, and the bidirectional diffusion training strategy.

(1) Diffusion Forward/Reverse Process Overview.

The diffusion model consists of two key processes: the forward diffusion process and the reverse diffusion process [32].

The forward diffusion process is a Markov chain that gradually adds Gaussian noise to images, starting from the original image $\mathbf{x}_0$ and proceeding through multiple timesteps t until reaching an approximately pure noise distribution. This process can be represented as:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\,\mathbf{x}_{t-1}, \beta_t\mathbf{I}) \tag{4}$$

where $\mathbf{x}_t$ represents the image state at timestep t, β_t is a predefined noise schedule parameter, and $\beta_t \in (0,1)$. Through reparameterization, we can directly sample from any arbitrary timestep:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\,\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\,\boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}) \tag{5}$$

where $\bar{\alpha}_t = \prod_{i=1}^{t}(1-\beta_i)$.

The reverse diffusion process is a denoising process that aims to learn the conditional transition probability $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$ to gradually recover the original image:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1};\, \boldsymbol{\mu}_\theta(\mathbf{x}_t, t),\, \sigma_t^2\mathbf{I}\right), \tag{6}$$

where the variance $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$, and the mean can be expressed as:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{1-\beta_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) \tag{7}$$

where, $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$ is the noise vector predicted by the neural network $\boldsymbol{\epsilon}_\theta$. This process can be understood as progressively removing noise and restoring image details, enabling the model to generate high-quality images.

(2) Conditional Enhancement of Components.

In the second stage, two parallel enhancement modules based on conditional diffusion models further enhance and optimize the structural component and residual component obtained from the first stage—the structure diffusion model is used to recover lost structural information, while the residual diffusion model suppresses noise and optimizes brightness distribution. The two diffusion models maintain consistent architecture but set different optimization objectives according to the characteristics of different components, as shown in Fig. 4.

Each diffusion branch uses the corresponding output from the first stage as conditional input, combining these conditions with the current noise state as joint input to the diffusion network during both forward and reverse processes. This way, the network is continuously constrained by physical priors when performing gradual noise injection and denoising restoration: for the structure component branch, the conditional input guides the diffusion process to better complete missing structural information; for the residual component branch, the conditional input helps the diffusion process more effectively suppress noise and optimize brightness distribution.

The conditional reverse diffusion process can be expressed as:

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{s}_k) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu}_\theta(\mathbf{x}_t, \mathbf{s}_k, t), \sigma_t^2\mathbf{I}\right) \tag{8}$$

where, $k \in \{\mathrm{s}, \mathrm{r}\}$ refers to the structure component and residual component output from the first stage, respectively, and $\mathbf{s}_k$ represents the conditional information. The mean term is typically parameterized as:

$$\boldsymbol{\mu}_\theta(\mathbf{x}_t, \mathbf{s}_k, t) = \frac{1}{\sqrt{1-\beta_t}}\left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\,\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \mathbf{s}_k, t)\right) \tag{9}$$

where, by optimizing the parameters θ of the noise prediction network $\boldsymbol{\epsilon}_\theta$, we predict the noise vector $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, \mathbf{s}_k, t)$, which receives the current noisy image $\mathbf{x}_t$, conditional signal $\mathbf{s}_k$, and timestep t as inputs.

(3) Bidirectional Diffusion Training Strategy.

Traditional conditional diffusion models have inevitable randomness in their sampling process when generating from random noise, which causes content inconsistency issues in reconstruction tasks: even with the same overall structure, different samplings may produce variations at the detail level. Inspired by the limitations addressed in [33], we designed a Bidirectional Diffusion

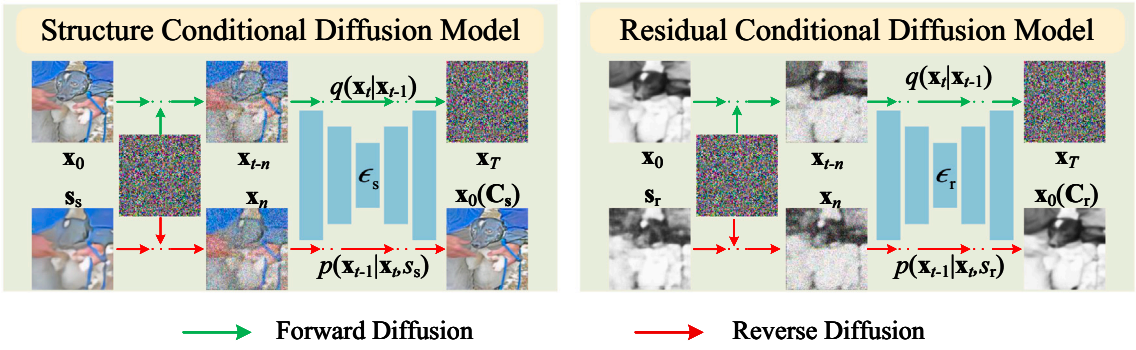

Fig. 4. The figure shows the architecture of the conditional diffusion enhancement module in the second stage. On the left is the structure component conditional diffusion model (with the structure component $\mathbf{s}_s$ output from the first stage as conditional input), used to further recover missing structural information; on the right is the residual component conditional diffusion model (with the residual component $\mathbf{s}_r$ output from the first stage as conditional input), used to suppress noise and optimize brightness distribution.

Training Process based on this foundation. Unlike simply optimizing forward diffusion, this method incorporates the reverse diffusion process as a supervision target during training, enabling the model to maintain a stable approximation to the real distribution during gradual denoising, thereby effectively reducing detail differences during sampling.

In the reverse diffusion process, training iteratively restores clean components step by step: at each time step t, the noise estimation network ϵ_θ receives the current state $\mathbf{x}_t$ and conditional inputs (component outputs from the first stage), predicts and removes residual noise, generating results closer to the target component. By continuously constraining generated results to converge toward the target results, reverse diffusion not only suppresses randomness during sampling but also makes the generation process more consistent with the real distribution at the detail level, ensuring the stability and reliability of output results.

Meanwhile, forward diffusion serves as a data perturbation mechanism during training, gradually transitioning real components to noisy states, enabling the network to learn how to effectively utilize conditional inputs at different noise levels. The joint optimization of forward and reverse diffusion processes enhances model adaptability under multiple noise conditions while ensuring physical consistency constraints on generated results, significantly improving structural components in detail restoration and residual components in noise suppression and brightness enhancement.

At the end of the second stage, after optimization by the conditional diffusion model enhancement module guided by the bidirectional diffusion training strategy, the enhanced structural component $\mathbf{C}_s$ and residual component $\mathbf{C}_r$ are recombined to obtain the final reconstruction result, with the combination process expressed as:

$$\hat{\mathbf{I}} = \mathbf{C}_s \odot \mathbf{C}_r \tag{10}$$

This stage significantly improves the overall quality of the image, ensuring that both structural and residual components are fully enhanced, ultimately achieving high-quality reconstruction results.

2.3. Training objectives for two-stage approach

2.3.1. Training loss design

To effectively train the proposed two-stage physics-aware reconstruction framework, we optimize the structure–residual extraction module (Stage 1) and the conditional diffusion-based component enhancement module (Stage 2) separately, with each module being guided by a task-specific training objective that is tailored to its role in the reconstruction pipeline.

(1) Stage 1: Multi-level Multi-scale Component Extraction Module.

Let $\mathbf{S}(\mathbf{x})$ and $\mathbf{R}(\mathbf{x})$ denote the structure and residual components predicted by the Stage 1 network, and let $\mathbf{S}_{gt}$, $\mathbf{R}_{gt}$, and $\mathbf{I}_{gt}$ denote the corresponding structural target, residual target, and full ground-truth image, respectively. The targets $\mathbf{S}_{gt}$ and $\mathbf{R}_{gt}$ are derived from $\mathbf{I}_{gt}$ using a pre-trained generic component extraction network that shares the backbone of the Stage 1 extractor but excludes the PMFM, and is trained offline with simple reconstruction and regularization losses. Based on these variables, the loss for this stage consists of the following three terms:

Component Fidelity Loss: This term constrains each extracted component to be close to its corresponding reference, ensuring accurate and independent extraction. The formulation is as follows:

$$\mathcal{L}_{\text{comp}} = \text{MAE}\left(\mathbf{S}(\mathbf{x}), \mathbf{S}_{gt}\right) + \text{MAE}\left(\mathbf{R}(\mathbf{x}), \mathbf{R}_{gt}\right) \tag{11}$$

Reconstruction Consistency Loss: This term enforces that the recombined output from the structure and residual components matches the ground-truth image $\mathbf{I}_{gt}$, thereby ensuring the global fidelity of the reconstruction. The formulation is as follows:

$$\mathcal{L}_{\text{recon}} = \text{MSE}\left(\mathbf{R}(\mathbf{x}) \odot \mathbf{S}(\mathbf{x}), \mathbf{I}_{gt}\right) \tag{12}$$

Spatial Smoothness Regularization: This term applies a Total Variation (TV) penalty to both components to suppress high-frequency noise while maintaining edge continuity, which is particularly beneficial in low-light conditions. The formulation is as follows:

$$\mathcal{L}_{\text{smooth}} = \text{TV}(\mathbf{R}(\mathbf{x})) + \text{TV}(\mathbf{S}(\mathbf{x})) \tag{13}$$

Total Loss: Based on the three loss components above, the total loss for the stage 1 is defined as their weighted sum, where λ_1, λ_2, λ_3 are the respective weight coefficients for balancing the importance of different loss terms, formulated as follows:

$$\mathcal{L}_{\text{stage1}} = \lambda_1 \cdot \mathcal{L}_{\text{comp}} + \lambda_2 \cdot \mathcal{L}_{\text{recon}} + \lambda_3 \cdot \mathcal{L}_{\text{smooth}} \tag{14}$$

By simultaneously optimizing these three terms, the network can achieve under low-light noise interference: (i) high-fidelity structure and residual extraction; (ii) overall consistency of the reconstructed results (iii) noise suppression and structural detail preservation.

(2) Stage 2: Conditional Diffusion-based Component Enhancement Module.

To enhance the reconstruction quality of structural and residual components, while avoiding content inconsistencies in the conditional diffusion model generation process (i.e., "diversity-induced inconsistencies"), the loss design for this stage comprises two parts.

Noise Matching Loss: This term constrains the predicted noise $\epsilon_\theta(\cdot)$ in the diffusion process to match the added noise ϵ_t, following the formulation in [32]. For a given component index $k \in \{\mathrm{s}, \mathrm{r}\}$, with corresponding conditioning $\mathbf{s}_k$, the formulation is as follows:

$$\mathcal{L}_{\text{diff}}^{(k)} = \mathbb{E}_{x_t, x_0, t}\left[\left\|\epsilon_t - \epsilon_\theta(x_t, \mathbf{s}_k, t)\right\|^2\right] \tag{15}$$

Content Consistency Loss: For a generic component $\mathbf{C}_k$ and its reference $\mathbf{C}_{k,\mathrm{gt}}$ (where $k \in \{\mathrm{s}, \mathrm{r}\}$ denotes structure or residual components). This loss ensures the content quality of generated results through multi-level constraints. At the pixel intensity level, it uses Mean Absolute Error (MAE); at the structural level, it incorporates the Structural Similarity Index (SSIM); and at the perceptual level, it employs Learned Perceptual Image Patch Similarity (LPIPS). The combination of these three metrics maintains consistency between the generated results and the reference in terms of intensity, structure, and perception dimensions, effectively reducing differences caused by random sampling. The formulation is as follows:

$$\begin{aligned}\mathcal{L}_{\text{con}}^{(k)} = {} & \mathrm{MAE}\left(\mathbf{C}_k, \mathbf{C}_{k,\mathrm{gt}}\right) + \left[1 - \mathrm{SSIM}\left(\mathbf{C}_k, \mathbf{C}_{k,\mathrm{gt}}\right)\right] \\ & + \lambda_4 \cdot \mathrm{LPIPS}\left(\mathbf{C}_k, \mathbf{C}_{k,\mathrm{gt}}\right)\end{aligned} \tag{16}$$

This formulation jointly constrains intensity, structural, and perceptual consistency, thereby mitigating discrepancies introduced by random sampling.

Branch Training Loss: During training, the same loss form is applied to different component branches, only changing the index k:

$$\begin{aligned}&\text{Structure branch:} \quad \mathcal{L}_{\text{stage2}}^{(\mathrm{s})} = \mathcal{L}_{\text{diff}}^{(\mathrm{s})} + \lambda_5 \cdot \mathcal{L}_{\text{con}}^{(\mathrm{s})} \\ &\text{Residual branch :} \quad \mathcal{L}_{\text{stage2}}^{(\mathrm{r})} = \mathcal{L}_{\text{diff}}^{(\mathrm{r})} + \lambda_5 \cdot \mathcal{L}_{\text{con}}^{(\mathrm{r})}\end{aligned} \tag{17}$$

where λ_5 is used to balance the relative importance of noise matching loss and content consistency loss.

2.3.2. Training protocol

During training, Stage 1 is trained independently until convergence to obtain stable extractions of the structural component $\mathbf{S}(\mathbf{x})$ and the residual component $\mathbf{R}(\mathbf{x})$. This stage integrates physical-model-based constraints, such as physical model fusion, with deep feature representations to achieve accurate extraction of corresponding components from the input measurements.

After Stage 1 is completed, its network parameters are fixed, and the extracted $\mathbf{S}(\mathbf{x})$ and $\mathbf{R}(\mathbf{x})$ are used as the conditional inputs $\mathbf{s}_\mathrm{s}$ and $\mathbf{s}_\mathrm{r}$ for Stage 2, respectively. Stage 2 consists of a structure enhancement branch and a residual enhancement branch, associated with the loss functions $\mathcal{L}_{\text{stage2}}^{(\mathrm{s})}$ and $\mathcal{L}_{\text{stage2}}^{(\mathrm{r})}$, which are computed and updated separately during training.

This staged training strategy avoids interference from cross-stage gradient propagation, allows each stage to focus on its specific reconstruction objective, and ensures that Stage 2 further enhances detail restoration and noise suppression based on the extracted components, while maintaining physical consistency with the measurement model.

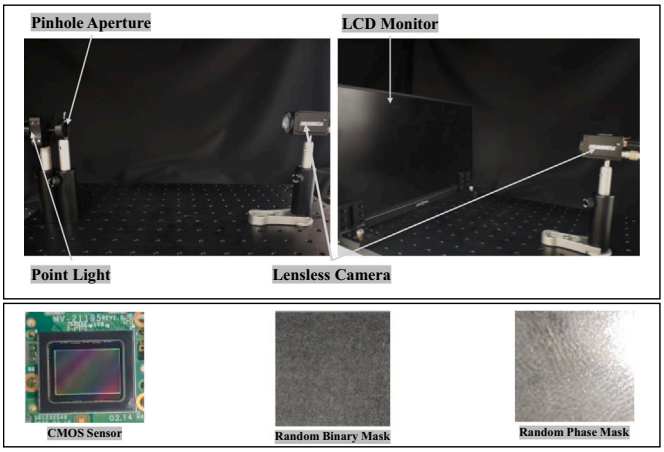

Fig. 5. Self-built lensless imaging system. Top: PSF and dataset capture setup. Bottom: Core hardware components including a Sony IMX183 CMOS sensor, a diffractive phase mask, a random binary amplitude mask, and a pinhole-filtered LED point light source.

3. Experiments

3.1. Implementation details

To verify the effectiveness of the imaging algorithm, we constructed a lensless imaging system for collecting lensless datasets under low-light conditions, as shown in Fig. 5. The system uses a HIKROBOT color camera equipped with a Sony IMX183 sensor, with a pixel size of $2.4\ \mu\mathrm{m} \times 2.4\ \mu\mathrm{m}$. The system employs two types of masks: a Luminit 0.5° diffuser phase mask with a transparent phase having pseudo-random gradient heights, and a random binary amplitude mask consisting of non-transparent amplitude formed by pseudo-random 0–1 sequences. Both masks are placed 3 mm in front of the sensor and equipped with a $2.0\,\mathrm{mm} \times 2.5\,\mathrm{mm}$ aperture to obtain an appropriate area for encoding modulation. The light source used for imaging system calibration is located 20 cm from the mask, maintained on the same horizontal plane as the aperture and sensor. As this distance gradually increases, the PSF hardly changes. We use a white LED light with brightness adjustable in the range of 0–255 as the light source, which is filtered through a pinhole aperture to serve as a point source.

We built a training environment based on the Pytorch framework on Windows, with a hardware configuration of 256 GB memory and an NVIDIA RTX A6000 GPU. In the first stage of training, the loss function weight coefficients were set to $\lambda_1 = 0.6$, $\lambda_2 = 0.5$, and $\lambda_3 = 0.01$. We used a learning rate of 1×10^{-4} with no weight decay, training for 250 epochs using the AdamW optimizer with a batchsize of 22. In the second stage of training, the loss function weight coefficients were set to $\lambda_4 = 0.2$ and $\lambda_5 = 0.001$, maintaining the same learning rate but switching to the Adam optimizer for 400 epochs, with the batchsize adjusted to 8. The diffusion time step T was 200, with 10 implicit sampling steps. To ensure training stability, we applied exponential moving average (EMA) to the model parameters at a rate of 0.9999, and set the dropout value to 0.3 in ResNet blocks.

3.2. Datasets and evaluation metrics

Since there is currently no low-light lensless dataset available, we collected two low-light lensless datasets, each containing 11,000 sample pairs, based on different mask types. Specifically, we selected 11,000 images from the Cats and Dogs [34] as ground truth for our dataset, and placed an LCD display 20 centimeters away from the camera, with an image size of $5\,\mathrm{cm} \times 5\,\mathrm{cm}$ and a black background in the area outside the image. To achieve continuous acquisition, images were displayed sequentially at 8-second intervals, and measurement data were

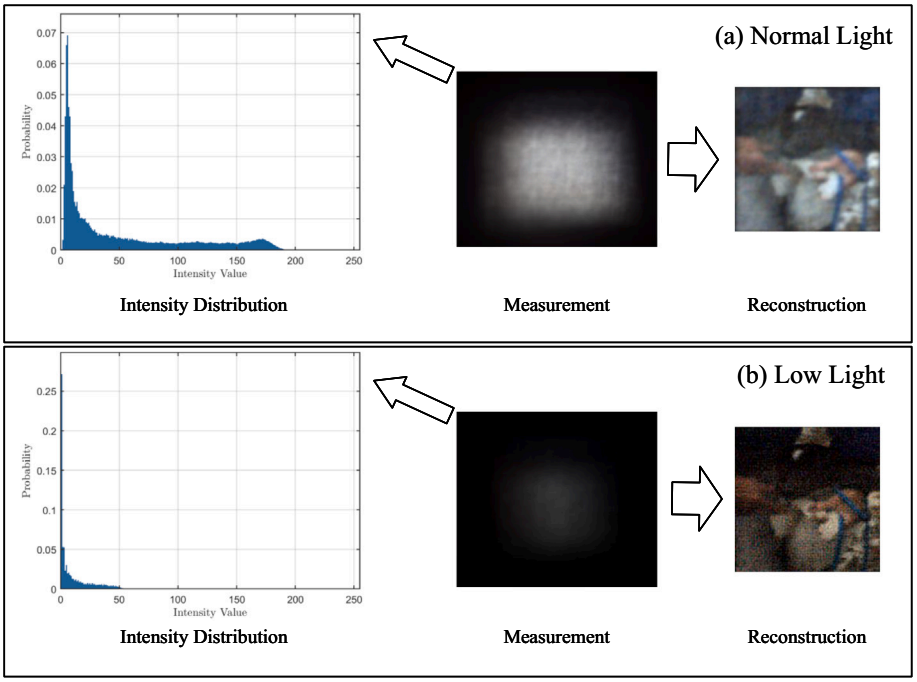

Fig. 6. Comparison of measurements and coarse reconstructions under (a) normal-light and (b) low-light conditions. Each subfigure presents: intensity distribution histogram, raw measurement, and corresponding coarse reconstruction.

collected using a diffraction phase mask-based lensless imaging system with a 0.3-second exposure time. This exposure time was chosen based on a practical trade-off between photon budget and sensor saturation, enabling effective low-light simulation while avoiding severe signal degradation and ensuring fair comparison among imaging methods. We constructed a complete dataset comprising 11,000 pairs of measurement data and corresponding ground truth, with 10,000 pairs used for training and 1000 pairs for testing. To visually demonstrate the degradation effect of measurement data under low-light conditions, we compared the intensity distribution, raw measurement data, and reconstruction results under normal and low-light conditions, as shown in Fig. 6. It can be observed that under low-light conditions, the measurement data intensity is significantly reduced, the signal-to-noise ratio decreases, and the reconstructed images exhibit more severe detail blurring and structure loss. Additionally, to verify the robustness of the imaging method, we also replaced the mask with a random binary amplitude mask and collected another dataset following the same procedure, to further evaluate the model's generalization performance under different modulation mechanisms.

To evaluate the quality of reconstruction results, we adopted two types of evaluation metrics: full-reference metrics and no-reference metrics. Full-reference metrics are used to assess the consistency between the reconstructed image and the ground truth, reflecting reconstruction fidelity, including MSE, PSNR, SSIM, and LPIPS [35,36]. No-reference metrics are used to evaluate the visual realism of reconstructed images without requiring reference to the original image, directly assessing the visual quality and naturalness of the image, for which we selected ManIQA [37] as this type of metric.

3.3. Evaluation on phase mask dataset

We evaluated the proposed method both qualitatively and quantitatively in low-light lensless imaging conditions, based on the phase-coded mask model. The comparison methods include four representative approaches: Wiener filtering [38], general-purpose low-light enhancement method Retinexformer [39], iterative unfolding reconstruction method DeepLIR [22], multi-scale convolutional network MWDN [24], and the diffusion model-based WiDiff [29]. All methods were evaluated using the same test dataset and experimental settings, with ground-truth images as references for performance and perceptual quality comparison.

Fig. 7 shows the overall reconstruction effects and their locally magnified detail comparison. As can be seen from the figure, the Wiener filtering results are generally blurry with obvious noise interference and overall low brightness. Retinexformer improves overall brightness and suppresses noise, but it fails to recover high-quality reconstructions—structural details remain blurred and obvious artifacts persist. Although DeepLIR and MWDN suppress noise to some extent, the locally magnified areas show that both still lack sharp edge contours, and MWDN has noticeable color reproduction deviations. WiDiff performs well in maintaining overall structure, but high-frequency textures in locally magnified areas remain insufficient. In comparison, our proposed method demonstrates clear advantages in both the full image and local details: sharper edge contours, richer texture information, and more natural color distribution. For example, in columns 1 and 2, the dog and cat eye details are sharper with more accurate color reproduction; in columns 5 and 6, the background stripe textures are clearer with better continuity.

Table 1 provides the quantitative evaluation results. The proposed method achieves the best performance on all image quality metrics, including full-reference metrics (MSE, LPIPS, PSNR, SSIM) and no-reference metrics (ManIQA). Notably, it excels in perceptual quality (LPIPS, ManIQA) and structural consistency (SSIM), indicating improvements in both reconstruction consistency and visual realism. Compared with the second-best method, DeepLIR, it produces reconstructions that are more faithful and visually realistic, consistent with the qualitative observations in Fig. 7.

3.4. Evaluation on amplitude masks dataset

To verify the adaptability and robustness of the proposed method under different mask types, we conducted additional experiments on the amplitude mask low-light imaging dataset, comparing it with Wiener, Retinexformer, DeepLIR, MWDN, and WiDiff methods.

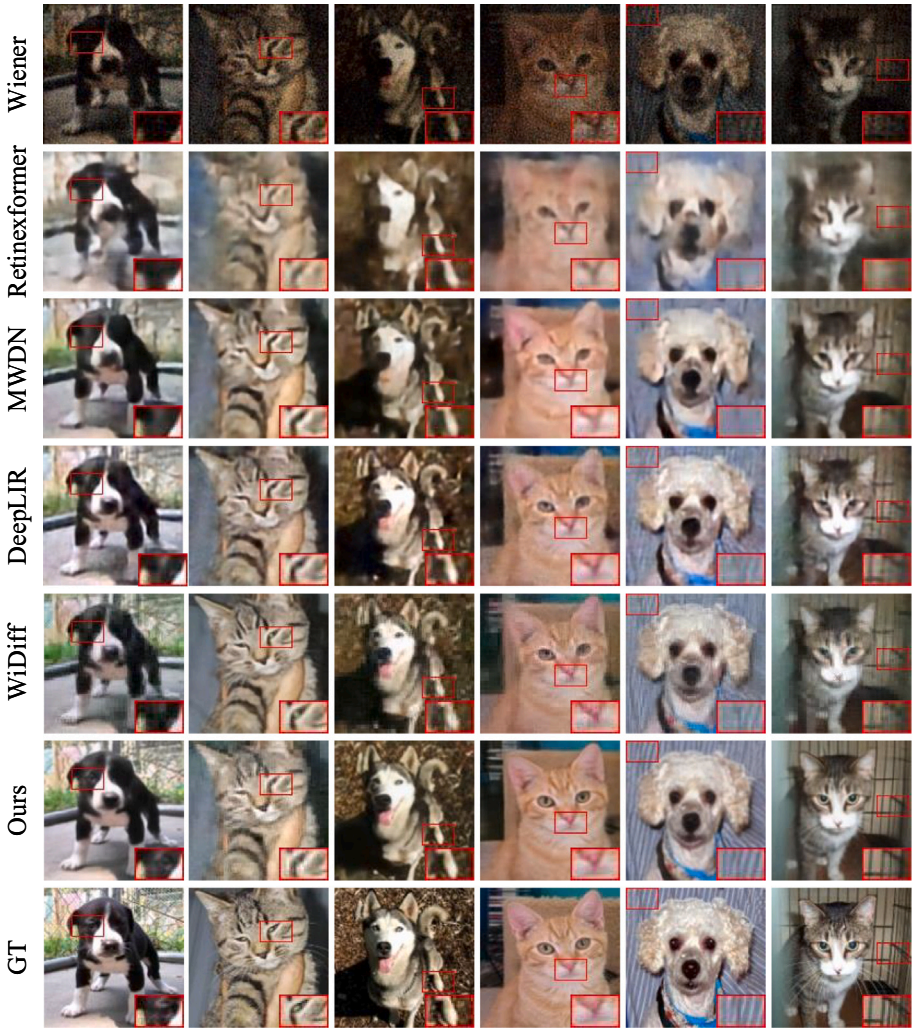

Fig. 7. Comparison of reconstruction results for low-light lensless imaging based on phase-coded masks. From top to bottom: measurement image, Wiener, Retinexformer, MWDN, DeepLIR, WiDiff, our method, and ground truth(GT). Red boxes indicate locally magnified areas to highlight detail comparison.

Table 1
Quantitative reconstruction performance comparison based on phase-coded masks. The table shows the average MSE, LPIPS, PSNR, SSIM, and the no-reference metric ManIQA for each method on the test set.

Method	Metric				
	MSE	LPIPS	PSNR	SSIM	ManIQA
GT	–	–	–	–	0.3091
Wiener [38]	0.0954	0.3467	10.5173	0.2340	0.1292
Retinexformer [39]	0.0065	0.1591	22.3777	0.6144	0.1770
DeepLIR [22]	0.0031	0.0998	25.4578	0.7473	0.1783
MWDN [24]	0.0063	0.1167	22.8098	0.7202	0.1916
WiDiff [29]	0.0051	0.0921	23.5512	0.7140	0.1615
Proposed	**0.0026**	**0.0626**	**26.2919**	**0.7908**	**0.2491**

Fig. 8 shows the overall reconstruction effects and locally magnified details. Wiener filtering exhibits severe noise interference in both the full image and local areas (such as fingers in columns 1 and 6), with overall low brightness. In contrast, while Retinexformer improves overall brightness and suppresses noise, the reconstructed structure and details remain severely blurred (for example, the eyes in the second and fourth columns). Although DeepLIR and MWDN reduce noise and improve brightness, the magnified areas (such as cat and dog eyes in columns 2 and 5) still show issues with insufficiently sharp edges and blurred details, while MWDN has color reproduction deviations (such as inaccurate coloring of the dog's tongue in column 3). WiDiff performs well in maintaining overall structure, but the recovery of detailed textures (such as the cat's eyes in column 5) remains insufficient. In comparison, the proposed method maintains clearer edges, richer textures, and more accurate color reproduction in both the full image and local areas, closely approaching the real image in reproducing key details such as animal eyes, fur, and background textures.

Table 2 provides the corresponding quantitative evaluation results, with the proposed method taking the lead in both full-reference metrics (MSE, LPIPS, PSNR) and no-reference metrics (ManIQA), with particularly significant improvements in perception-related metrics LPIPS and ManIQA. Although the reconstruction quality of all methods decreased compared to the phase mask scenario due to the lower light throughput under the amplitude mask configuration, the overall performance trend remained consistent, which further demonstrates that our method has good robustness and adaptability across different mask types.

In summary, the proposed method achieves reconstruction performance superior to existing methods on both types of mask data, maintaining a stable lead in both subjective visual quality and objective metrics. This advantage stems from the organic combination of structure-residual component synergistic extraction and conditional diffusion enhancement modules, enabling the method to simultaneously address structure recovery, noise suppression, and brightness improvement in low-light lensless imaging tasks, while exhibiting good generalization capability. These results not only verify the method's robustness

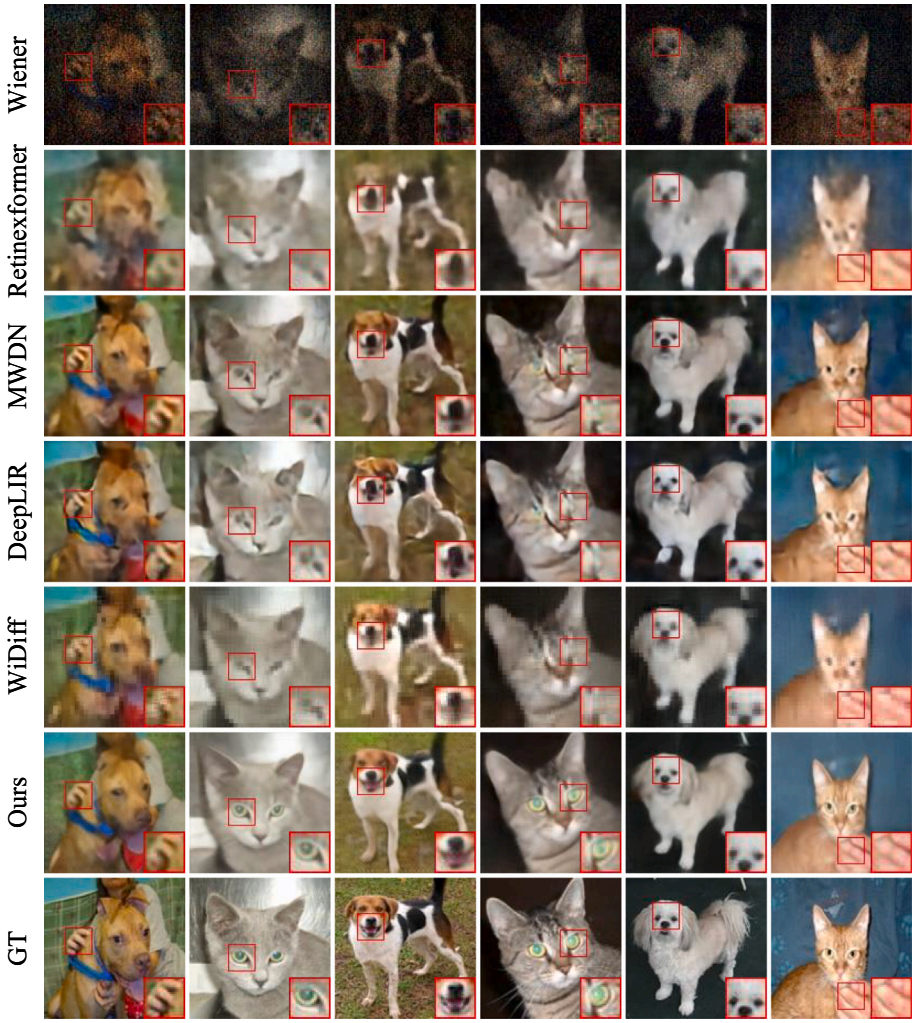

Fig. 8. Comparison of low-light lensless imaging reconstruction results based on amplitude-coded masks. From top to bottom: measurement image, Wiener, Retinexformer, MWDN, DeepLIR, WiDiff, our method, and GT. Regions highlighted by red boxes are magnified to facilitate comparison of fine details.

Table 2
Quantitative reconstruction performance comparison based on amplitude-coded masks. The table shows the average MSE, LPIPS, PSNR, SSIM, and the no-reference metric ManIQA for each method on the test set.

Method	Metric				
	MSE	LPIPS	PSNR	SSIM	ManIQA
GT	–	–	–	–	0.3091
Wiener [38]	0.1222	0.5181	9.4505	0.1380	0.1107
Retinexformer [39]	0.0072	0.1728	21.8715	0.5986	0.1765
DeepLIR [22]	0.0052	0.1230	23.1019	0.6566	0.1728
MWDN [24]	0.0072	0.1142	22.1366	**0.6968**	0.1827
WiDiff [29]	0.0064	0.1291	22.4612	0.6558	0.1411
Proposed	**0.0048**	**0.1056**	**23.5911**	0.6768	**0.2279**

under different modulation mechanisms but also provide a new solution approach for high-fidelity imaging under low-light conditions.

3.5. Ablation study

To evaluate the contribution of each component in the proposed two-stage reconstruction framework, we perform an ablation study on three key factors: (1) the bidirectional diffusion training strategy, (2) the physical model fusion mechanism, and (3) separate versus joint processing of the structure and residual components.

Table 3
Ablation results of different module combinations.

Bidirectional Diffusion	Physical Model Fusion Mechanism	Separate Process	PSNR	SSIM	LPIPS
×	×	×	21.6727	0.5665	0.1393
✓	×	×	22.7746	0.6292	0.1418
✓	✓	×	23.0454	0.6647	0.1370
✓	✓	✓	**23.5911**	**0.6768**	**0.1056**

Table 3 reports the quantitative results for different module combinations. Adding the bidirectional diffusion training strategy to the baseline markedly improves reconstruction fidelity, with only a minor, acceptable drop in perceptual quality. Introducing the physical model fusion mechanism further enhances reconstruction consistency and structural detail. When this mechanism is combined with separate processing of structure and residual components, the model attains its best overall performance, confirming that the full framework delivers superior perceptual quality and structural consistency.

3.6. Discussion

In terms of computational cost, experiments on our dataset show that the proposed method requires about 1.29 s per image with 46.8 million

parameters, whereas WiDiff takes 0.81 s with 22.1 million parameters and DeepLIR 0.05 s with 18 million parameters. The extra cost mainly comes from the diffusion-based dual-branch refinement stage, which in turn improves structural fidelity and perceptual quality. Thus, the current model is better suited to offline reconstruction or scenarios with moderate latency, and designing lighter or distilled variants for real-time or edge deployment is an important direction for future work.

In this work, we used a fixed exposure time to simulate low-light conditions when constructing the dataset, thereby demonstrating the potential of lensless imaging in low-light environments. The sensitivity of the model's performance to changes in exposure time (i.e., lighting intensity) and its ability to adapt to different low-light levels remain open questions; future work will investigate adaptive strategies to improve the robustness of lensless imaging systems under varying illumination.

Although practical lensless systems are still at an early stage, their inherent advantages—miniaturization, low cost, and suitability for constrained environments—make them strong candidates for these application domains. Extending this work to real-world use cases, such as developing adaptive enhancement algorithms for low-light, high-noise conditions on Internet-of-Things devices, is therefore of substantial practical interest.

4. Conclusion

In this work, we address the challenge of high-fidelity structural reconstruction in low-light lensless imaging by introducing a novel decomposition perspective. This approach separates the scene into structural components (intrinsic features like textures, edges, and colors) and residual components (external factors such as brightness variations and noise). We further develop a structure-guided measurement model that decouples these elements within the forward process, establishing a direct link between measurements and structural content for efficient extraction.

Based on this model, we propose a physics-aware two-stage reconstruction framework: the first stage employs multi-level, multi-scale extraction modules fused with physical constraints to isolate the components; the second stage uses generative conditional diffusion modules to refine them, enhancing details and optimizing residuals. Using our custom lensless imaging prototype, we curated two large-scale low-light datasets with phase- and amplitude-coded masks. Experimental results demonstrate that our method surpasses state-of-the-art techniques in structural fidelity, detail preservation, and noise suppression across both datasets, confirming its robustness. This framework provides a valuable reference for resource-constrained applications like biological imaging, surveillance, and edge-device sensing.

CRediT authorship contribution statement

Ziyang Liu: Writing – original draft, Visualization, Software, Methodology, Investigation. **Tianjiao Zeng:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Xu Zhan:** Writing – review & editing, Software, Methodology, Conceptualization. **Xiaoling Zhang:** Writing – review & editing. **Yunqi Wang:** Writing – review & editing. **Edmund Y. Lam:** Writing – review & editing, Conceptualization.

Funding

This work is supported by the Fund of National Natural Science Foundation of China (62305049); National Laboratory on Adaptive Optics, China (FNLAO-24-MS-S06); Natural Science Foundation of Sichuan Province, China (2024NSFSC1439); and Aeronautical Science Foundation of China (2024Z071080005).

Declaration of competing interest

The authors declare the following financial interests/personal relationships that may be considered as potential competing interests:

Tianjiao Zeng reports that financial support was provided by the National Natural Science Foundation of China. Tianjiao Zeng reports that financial support was provided by the National Laboratory on Adaptive Optics, China. Tianjiao Zeng reports that financial support was provided by the Natural Science Foundation of Sichuan Province, China. Tianjiao Zeng reports that financial support was provided by the Aeronautical Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] F. Heide, M. Rouf, M.B. Hullin, B. Labitzke, W. Heidrich, A. Kolb, High-quality computational imaging through simple lenses, ACM Trans. Graph. 32 (5) (2013) 1–14.
[2] Z. Yu, Z. Tian, J. Ma, Q. Zhou, P. Su, DCFZA: a high-quality lensless imaging technique, Opt. Laser Technol. 190 (2025) 113236.
[3] A. Ozcan, E. McLeod, Lensless imaging and sensing, Annu. Rev. Biomed. Eng. 18 (1) (2016) 77–102.
[4] Q. You, Y. Gao, F. Zhang, Q. Liao, W. Cao, P. Lu, Broadband multi-distance lensless imaging, Opt. Laser Technol. 184 (2025) 112374.
[5] S.B. Kim, H. Bae, K.-I. Koo, M.R. Dokmeci, A. Ozcan, A. Khademhosseini, Lens-free imaging for biological applications, J. Lab. Autom. 17 (1) (2012) 43–49.
[6] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, L. Waller, DiffuserCam: lensless single-exposure 3D imaging, Optica 5 (1) (2017) 1–9.
[7] E. Bezzam, S. Kashani, M. Vetterli, M. Simeoni, LenslessPiCam: a hardware and software platform for lensless computational imaging with a raspberry Pi, J. Open Source Softw. 8 (86) (2023) 4747.
[8] J. Wu, H. Zhang, W. Zhang, G. Jin, L. Cao, G. Barbastathis, Single-shot lensless imaging with fresnel zone aperture and incoherent illumination, Light: Sci. Appl. 9 (1) (2020) 53.
[9] N. Chen, E. Lam y, Differentiable pixel-super-resolution lensless imaging, Opt. Lett. 50 (4) (2025) 1180–1183.
[10] V. Boominathan, J.T. Robinson, L. Waller, A. Veeraraghavan, Recent advances in lensless imaging, Optica 9 (1) (2021) 1–16.
[11] S. Li, Y. Gao, J. Wu, M. Wang, Z. Huang, S. Chen, L. Cao, Lensless camera: unraveling the breakthroughs and prospects, Fundam. Res. 5 (4) (2025) 1725–1736.
[12] T. Zeng, Y. Zhu, E. Lam y, Deep learning for digital holography: a review, Opt. Express 29 (24) (2021) 40572–40593.
[13] A.N. Tikhonov, V. Arsenin y, Solutions of Ill-posed Problems, Winston and Sons, Washington, DC, 1977.
[14] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. 57 (11) (2004) 1413–1457.
[15] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.
[16] X. Pan, X. Chen, S. Takeyama, M. Yamaguchi, Image reconstruction with transformer for mask-based lensless imaging, Opt. Lett. 47 (7) (2022) 1843–1846.
[17] D. Bae, J. Jung, N. Baek, S.A. Lee, Lensless imaging with an end-to-end deep neural network, in: 2020 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), IEEE, 2020, pp. 1–5.
[18] K. Wang, E. Lam y, Deep learning phase recovery: data-driven, physics-driven, or a combination of both? Adv. Photonics Nexus 3 (5) (2024) 056006.
[19] Y. Wang, T. Zeng, F. Liu, Q. Dou, P. Cao, H.-C. Chang, Q. Deng, E.S. Hui, Illuminating the unseen: advancing MRI domain generalization through causality, Med. Image Anal. 101 (2025) 103459.
[20] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, L. Waller, Learned reconstructions for practical mask-based lensless imaging, Opt. Express 27 (20) (2019) 28075–28090.
[21] S.S. Khan, V. Sundar, V. Boominathan, A. Veeraraghavan, K. Mitra, FlatNet: towards photorealistic scene reconstruction from lensless measurements, IEEE Trans. Pattern Anal. Mach. Intell. 44 (4) (2020) 1934–1948.
[22] A. Poudel, U. Nakarmi, DeepLIR: attention-based approach for mask-based lensless image reconstruction, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2024, pp. 431–439.
[23] H. Qian, H. Ling, X. Lu, Robust unrolled network for lensless imaging with enhanced resistance to model mismatch and noise, Opt. Express 32 (17) (2024) 30267–30283.
[24] Y. Li, Z. Li, K. Chen, Y. Guo, C. Rao, MWDNs: reconstruction in multi-scale feature spaces for lensless imaging, Opt. Express 31 (23) (2023) 39088–39101.
[25] K. Chen, Y. Li, Z. Li, J. Hu, Y. Guo, Enhancing object recognition for lensless cameras through PSF correction and feature loss, Opt. Laser Technol. 190 (2025) 113077.
[26] T. Zeng, E. Lam y, Robust reconstruction with deep learning to handle model mismatch in lensless imaging, IEEE Trans. Comput. Imaging 7 (2021) 1080–1092.
[27] O. Kingshott, N. Antipa, E. Bostan, K. Akşit, Unrolled primal-dual networks for lensless cameras, Opt. Express 30 (26) (2022) 46324–46335.

[28] X. Cai, Z. You, H. Zhang, J. Gu, W. Liu, T. Xue, PhoCoLens: photorealistic and consistent reconstruction in lensless imaging, Adv. Neural Inf. Process. Syst. 37 (2024) 12219–12242.
[29] Z. Liu, T. Zeng, X. Zhan, X. Zhang, E. Lam y, Generative approach for lensless imaging in low-light conditions, Opt. Express 33 (2) (2025) 3021–3039.
[30] Z. Liu, T. Zeng, Q. He, X. Zhan, W. Zhang, X. Zhang, Low-light lensless image enhancement via diffusion model, in: Holography, Diffractive Optics, and Applications XIV, vol. 13240, SPIE, 2024, pp. 130–135.
[31] S.W. Zamir, A. Arora, S. Khan, M. Hayat, F.S. Khan, M.-H. Yang, Restormer: efficient transformer for high-resolution image restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5728–5739.
[32] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, Adv. Neural Inf. Process. Syst. 33 (2020) 6840–6851.
[33] H. Jiang, A. Luo, H. Fan, S. Han, S. Liu, Low-light image enhancement with wavelet-based diffusion models, ACM Trans. Graph. 42 (6) (2023) 1–14.
[34] O.M. Parkhi, A. Vedaldi, A. Zisserman, C.V. Jawahar, Cats and dogs, in: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2012, pp. 3498–3505.
[35] U. Sara, M. Akter, M.S. Uddin, Image quality assessment through FSIM, SSIM, MSE and PSNR—a comparative study, J. Comput. Commun. 7 (3) (2019) 8–18.
[36] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 586–595.
[37] S. Yang, T. Wu, S. Shi, S. Lao, Y. Gong, M. Cao, J. Wang, Y. Yang, MANIQA: multi-dimension attention network for no-reference image quality assessment, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 1191–1200.
[38] N. Wiener, Extrapolation, Interpolation, and Smoothing of Stationary Time Series, The MIT Press, 1964.
[39] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, Y. Zhang, Retinexformer: one-stage retinex-based transformer for low-light image enhancement, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 12504–12513.