

McLDI-INR: mask-constraint lensless dynamic imaging by dual-domain collaborative implicit neural representation

Wei Li, Weizhi Song, Chenyu Xu, Bo Xiong, You Zhou, Xun Cao and Zhan Ma

Citation: Li W, Song WZ, Xu CY et al. McLDI-INR: mask-constraint lensless dynamic imaging by dual-domain collaborative implicit neural representation. *Intelligent Opto-Electronics* 2, 260002 (2026).

Received: 24 January 2026; Accepted: 27 April 2026; Published online: 27 May 2026

Related articles

In-situ and ex-situ twisted bilayer liquid crystal computing platform for reconfigurable image processing

Kang Zeng, Yougang Ke, Zhangming Hong, et al

Opto-Electronic Advances 2025, 8(12): 250226 doi: [10.29026/oea.2025.250226](https://doi.org/10.29026/oea.2025.250226)

Meta-lens digital image correlation

Zhou Zhao, Xiaoyuan Liu, Yu Ji, et al

Opto-Electronic Advances 2025, 8(9): 250014 doi: [10.29026/oea.2025.250014](https://doi.org/10.29026/oea.2025.250014)

Unsupervised learning enabled label-free single-pixel imaging for resilient information transmission through unknown dynamic scattering media

Fujie Li, Haoyu Zhang, Zhilan Lu, et al

Opto-Electronic Advances 2025, 8(10): 250013 doi: [10.29026/oea.2025.250013](https://doi.org/10.29026/oea.2025.250013)

Dual-frequency angular-multiplexed fringe projection profilometry with deep learning: breaking hardware limits for ultra-high-speed 3D imaging

Wenwu Chen, Yifan Liu, Shijie Feng, et al

Opto-Electronic Advances 2025, 8(9): 250021 doi: [10.29026/oea.2025.250021](https://doi.org/10.29026/oea.2025.250021)

More related articles in Opto-Electronic Journals Group website 



McLDI-INR: mask-constraint lensless dynamic imaging by dual-domain collaborative implicit neural representation

Wei Li^{1,2†}, Weizhi Song^{1†}, Chenyu Xu¹, Bo Xiong^{3*}, You Zhou^{2*}, Xun Cao¹ and Zhan Ma¹

Abstract: Lensless imaging, which replaces conventional optical lenses with computational techniques, enables miniaturized and lightweight imaging systems. However, inherent phase loss, manifested as twin-image artifacts and defocus noise, often necessitates multi-shot reconstruction. This dependency induces spatiotemporal coupling, inevitably compromising imaging resolution in dynamic scenarios. To address this, we propose McLDI-INR, an implicit neural representation (INR) framework that integrates temporal embedding with radial encoding for mask-constrained dynamic lensless imaging, enabling dynamic scene reconstruction from single-shot measurements. A binary encoding mask modulates the incident wavefront to build an interpretable imaging model, while joint spatial- and Fourier-domain regularization establishes a continuous spatiotemporal representation for accurate motion compensation, thereby reducing artifacts in dynamic scenes. Experiments demonstrate that our method achieves higher fidelity and resolution, offering a new paradigm for high-dimensional lensless imaging across space and time.

Keywords: lensless imaging; dynamic imaging; implicit neural representation; self-supervised reconstruction

DOI: [10.67704/ioe.2026.260002](https://doi.org/10.67704/ioe.2026.260002)

Citation: Li W, Song WZ, Xu CY et al. McLDI-INR: mask-constraint lensless dynamic imaging by dual-domain collaborative implicit neural representation. *Intell Opto-Electron* 2, 260002 (2026).

1 Introduction

Lensless imaging^{1–4} has emerged as a promising technology to advance next-generation miniaturized imaging systems^{5,6}. In classical coherent lensless imaging systems, image sensors integrated with wavefront modulation elements^{7–11} directly capture the diffracted light field from the target scene, followed by computational reconstruction to generate high-quality images¹². By eliminating optical lenses, this architecture reduces size, weight, and cost while avoiding optical aberrations and chromatic dispersion inherent to conventional imaging systems. It establishes a solid foundation for miniaturized imaging within the optical field encoding-computational decoding paradigm. The versatility of lensless imaging is evident in its wide applications, including optical microscopy^{3,4,9,13}, biomedical analysis^{14–16}, digital holography^{17,18}, wearable devices¹⁹, and deep-space exploration²⁰. These applications highlight its potential to transcend the constraints of conventional optics.

Traditional lensless imaging is often posed as a regularized least-squares deconvolution problem, solved via iterative reconstruction (e.g., ptychographic iterative engine (PIE)²¹) or optimization with gradient- or frequency-domain sparsity constraints^{22–24}. However, diffraction-induced defocus, spatial disorder, and the loss of phase information make single-frame measurements highly mixed and often unrecognizable, rendering lensless dynamic reconstruction severely ill-posed. To reduce ambiguity, strong priors such as spatiotemporal smoothness²⁵, sparse representations^{26,27}, or physics-based constraints²⁸, have been introduced. But these approaches rely heavily on accurate prior models and are prone to over-regularization, reduced spatiotemporal resolution, and motion artifacts under rapid non-rigid motion. Diversity acquisition strategies, including multi-height scanning^{29–31}, multi-angle illumination^{8,11}, and multi-wavelength illumination^{32,33}, can improve robustness but entangle spatiotemporal information and introduce artifacts in dynamic scenes. Thus, achieving high-quality, temporally

Received: 24 January 2026

Accepted: 27 April 2026

Published online: 27 May 2026

¹School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China; ²Medical School, Nanjing University, Nanjing 210093, China; ³Institute of Medical Technology, Peking University Health Science Center, Peking University, Beijing 100871, China.

[†]These authors contributed equally to this work.

*Correspondence: B Xiong, Email: xiongbo@pku.edu.cn; Y Zhou, Email: zhouyou@nju.edu.cn

resolved imaging requires a novel few-shot approach that effectively integrates physical priors to mitigate measurement deficiencies and suppress motion blur.

Current compensation strategies for lensless imaging include: 1) physical constraint regularization^{25,28}; 2) sparsity regularization²⁶; 3) deep-learning priors^{34,35}; and 4) optical model priors^{29,36,37}. However, these methods often depend on assumptions of static scenes or rigid smoothing constraints, which hinder their capacity to recover spatiotemporal details and ultimately degrade resolution. Deep learning has substantially advanced lensless imaging^{2,34,35,38–40}. Supervised end-to-end approaches, such as deep unfolding^{38,39}, image-domain deconvolution³⁵, and feature-domain mapping⁴⁰, improve perceptual quality but depend heavily on large paired datasets, with limited generalization to unseen scenarios. Implicit neural representations (INRs) or Neural radiance fields (Nerf) parameterize signals as continuous neural functions and, through coordinate encoding^{41–45} embed low-dimensional inputs into high-dimensional feature spaces to learn nonlinear mappings. Widely adopted in 3D vision^{46,47}, INR frameworks have recently been extended to computational imaging tasks, including lensless imaging^{29,37,48,49}, optical microscopy⁴², and light-field imaging⁵⁰. Their differentiable coupling with physical models enables self-supervised learning and reduces dependence on paired data.

Nevertheless, INR-based lensless imaging often relies on multi-diversity measurements and lacks effective priors for dynamic scenes^{29,51}, which limits single-frame recovery and spatiotemporal resolution. While INR has shown promise for dynamic reconstruction in other domains^{52–54}, directly applying these solutions to lensless imaging still faces challenges arising from differences in the imaging model. Specifically, 1) data mismatch: dynamic NeRF relies on dense

multi-view inputs or temporally consistent observations, whereas lensless imaging provides only single-frame encoded measurements without phase information, lacking the viewpoint diversity and temporal correspondence required to disentangle spatiotemporal information; and 2) prior incompatibility: the effective priors used by existing dynamic NeRF methods, such as natural-image-based data priors and sparsity constraints, are difficult to reconcile with the mask-coded diffraction physics of lensless imaging.

In this work, we propose mask-constraint lensless dynamic imaging with implicit neural representation (McLDI-INR), a self-supervised, single-shot framework for mitigating spatiotemporal resolution loss and information degradation in dynamic scenes (Fig. 1). Our method integrates a space-time INR network^{52,54} with radial encoding and a mask-constrained imaging model, using a binary mask as an interpretable prior to modulate the incident wavefront. The reconstruction backend incorporates a spatial-frequency dual-domain collaborative loss, which combines spatial-domain loss with a physics-model-driven frequency-domain (PMFD) loss, aimed at enhancing the fidelity of reconstructed continuous complex optical fields. Experiments show that McLDI-INR effectively suppresses motion artifacts and simultaneously improves both resolution and reconstruction fidelity in lensless dynamic imaging.

2 Methods

2.1 Physical modeling of McLDI-INR

To enhance the dynamic imaging capability, we developed a mask-constraint lensless imaging system (see the schematic in Fig. 2). Using a coherent laser source and positioning the dynamic object at a sufficient distance to achieve normal

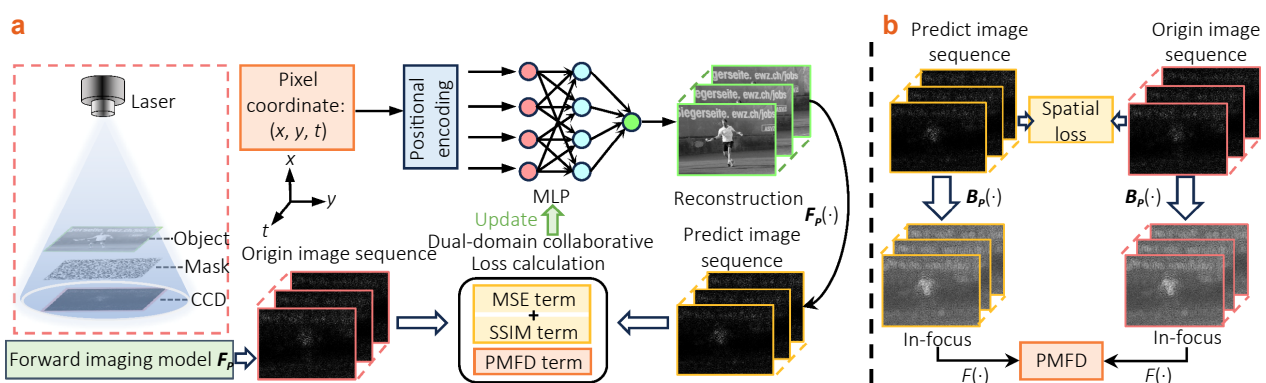


Fig. 1 | Method overview. (a) Pipeline of McLDI-INR. The lensless system first captures a raw diffraction image sequence of the dynamic target. A forward physical model of the lensless imaging process (Eq. (4)) is then established. McLDI-INR encodes spatiotemporal coordinates (x, y, t) to drive a multi-layer perceptron (MLP) that fits the implicit complex amplitude field of the scene. Using the forward model ($F_p(\cdot)$) predicted diffraction sequences are generated and pixel-wise compared with the raw data for self-supervised optimization. (b) Dual-domain collaborative loss. Based on the spatial domain loss, to better preserve high-frequency details, the physics-model-driven frequency-domain (PMFD) loss module back-propagates both raw and predicted sequences to the object's in-focus plane via angular spectrum diffraction theory, obtaining high-frequency representations. Fast Fourier Transforms (FFT) are then applied to compute the frequency-domain loss. $B_p(\cdot)$ denotes the backpropagation computation operator. $F(\cdot)$ denotes the 2D Fourier transform.

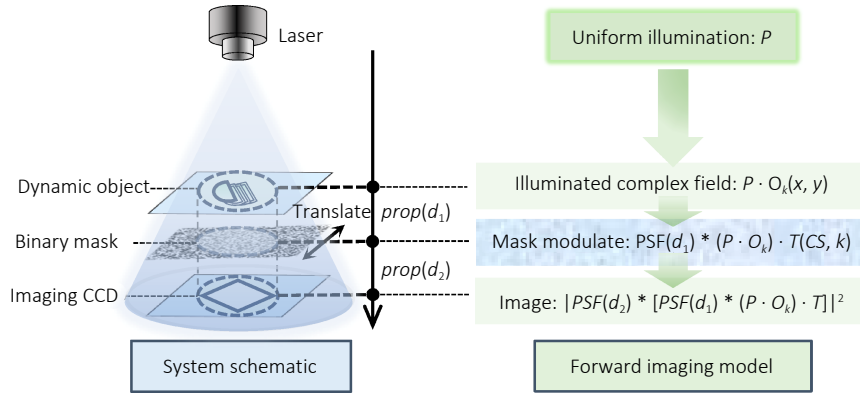


Fig. 2 | Schematic of the optical system and forward imaging model. The system uses a coherent laser source to illuminate the dynamic target at normal incidence. The light propagates a short distance of approximately $2mm$, sequentially reaching the modulation mask and sensor planes. The forward imaging process is modeled as follows: the coherent illumination field first propagates a distance d_1 under Fresnel diffraction, is then modulated by the mask, and finally propagates an additional distance d_2 to the camera surface, where the diffraction image is captured.

incidence, the illumination can be modeled as a constant plane wave $P \in \mathbb{C}^{m \times n}$, meaning that every 2D coordinate (x, y) in the object plane receives uniform intensity. The complex-amplitude distribution of the object is denoted as $O(x, y)$. Consequently, the illumination complex field at the object plane is computed as $P \cdot O(x, y)$, where $\cdot$ denotes element-wise multiplication. According to Fresnel diffraction theory, after propagating a distance of d_1 , the object's exit wavefront at the mask plane $W_k(x, y)$ is given by

$$W_k(x, y) = PSF(d_1) * [P \cdot O_k(x, y)], \quad (1)$$

where $PSF(d_1)$ denotes the point spread function for Fresnel diffraction over a propagation distance of d_1 , $k = 1, 2, \dots, K$ denotes the indexes of each dynamic measurement, K is the total number of frames, and $*$ represents the convolution operator. We modulate the optical forward model using a mask, denoted by $CS(x, y)$, to introduce a known physical prior for subsequent reconstruction.

Moreover, by applying a unidirectional dynamic translation to the mask, each frame experiences a distinct modulation, which captures inter-frame differential information and enhances single-frame reconstruction quality. After the wavefront $W_k(x, y)$ propagates over a distance d_2 to the sensor, the resulting complex amplitude field CF_k is given by

$$CF_k = PSF(d_2) * [W_k(x, y) \cdot T(CS(x, y), k)], \quad (2)$$

where $PSF(d_2)$ represents the point spread function when light wave propagates from the mask plane to the sensor plane with a distance of d_2 , and T denotes the translation operator.

Since the sensor can only capture the intensity of the light wave, the raw diffraction image acquired at the k -th frame is

$$I_k(x, y) = |CF_k(x, y)|^2 = |PSF(d_2) * [W_k(x, y) \cdot T(CS(x, y), k)]|^2, \quad (3)$$

where $|\cdot|$ denotes the magnitude of complex value. Thus, the forward imaging model $F_p(\cdot)$ can be concisely formulated as

$$F_p(k) = |PSF(d_2) * [PSF(d_1) * (P \cdot O_k) \cdot T(k)]|^2. \quad (4)$$

Through the forward imaging model in Eq. (4), the high-resolution target at the k -th frame O_k can be mapped to the lensless diffraction image I_k . This forward computation process is differentiable and suitable for the implicit neural representation described in Section 2.2.

2.2 McLDI-INR architecture

As shown in Fig. 1(a), McLDI-INR aims to achieve high-resolution reconstruction of dynamic targets from raw lensless diffraction image sequences. Considering the imaging system's anisotropic resolution in spatial and temporal dimensions, McLDI-INR applies positional encoding in the temporal domain and designs a differentiated spatiotemporal encoding strategy. Specifically, the input coordinate $r(x, y, t)$ is decomposed into $v(x, y)$ and t , using distinct approaches to expand the spatial coordinates (x, y) and the temporal coordinate t . In particular, McLDI-INR employs radial encoding⁴³ for the (x, y) coordinates:

$$\gamma(v) = \begin{pmatrix} \sin(2^0 \pi R_\theta v), \cos(2^0 \pi R_\theta v), \dots, \\ \sin(2^{L_{xy}-1} \pi R_\theta v), \cos(2^{L_{xy}-1} \pi R_\theta v) \end{pmatrix}, \quad (5)$$

where L_{xy} controls the number of extended frequencies in the (x, y) dimensions, and R_θ represents the set of rotation matrices corresponding to coordinate translation angles θ :

$$R_\theta = \left\{ \begin{bmatrix} \cos(\theta_j) & -\sin(\theta_j) \\ \sin(\theta_j) & \cos(\theta_j) \end{bmatrix} \right\}_{j=1}^J, \quad (6)$$

where J denotes the number of translational angle frequencies.

For the temporal dimension t , McLDI-INR employs standard positional encoding for expansion:

$$\gamma(t) = \begin{pmatrix} \sin(2^0 \pi t), \cos(2^0 \pi t), \dots, \\ \sin(2^{L_t-1} \pi t), \cos(2^{L_t-1} \pi t) \end{pmatrix}, \quad (7)$$

where L_t denotes the number of expansion frequencies along the temporal dimension; optimal performance for different objects can be achieved by tuning L_{xy} , L_t , and θ . By applying spatiotemporal encoding, McLDI-INR establishes a continuous functional representation in the spatiotemporal domain (x, y, t) . This approach not only captures higher spatial frequency details but also reduces temporal motion blur. An MLP then fits the high-definition dynamic object, with the coordinate-to-image mapping defined as $g_\theta(r): (x, y, t) \rightarrow amp, phs \rightarrow O_k = amp \cdot e^{j \cdot phs}$. Consequently, the forward model in Eq. (4) produces the predicted diffraction image sequence, $I_k' = F(g_\theta(r), k)$. The loss between the predicted sequence I_k' and the original sequence I_k is then computed via a dual-domain collaborative loss function:

$$L = [\alpha \cdot L_{MSE}(CS(S, D)) + (1 - \alpha) \cdot L_{SSIM}(CS(S, D))] + [\beta \cdot L_{PMFD}] , \quad (8)$$

where α and β are hyperparameters that balance the spatial and frequency-domain loss terms, respectively. $CS(S, D)$ denotes the binary mask with feature size S and duty cycle D , which models the physical modulation process in the lensless imaging system. Based on Eq. (8), the mask constraint is incorporated into the spatial-domain reconstruction loss, thereby enforcing physically consistent modulation during optimization and improving reconstruction fidelity.

For the spatial-domain loss calculation, McLDI-INR simultaneously enforces mean square error (MSE) and SSIM constraints at the measurement plane. The initial optimization is performed by computing a weighted MSE-SSIM loss between the measured data I_k and the predicted reconstruction I_k' .

However, according to Parseval's theorem⁵⁵, spatial MSE is equivalent to an L_2 loss uniformly weighted across all frequency components. This isotropic constraint may suppress high-frequency details during gradient descent due to the dominance of low-frequency components. To address this, we introduce the PMFD loss in the frequency domain to mitigate high-frequency information deficiencies in single-frame reconstruction. By optimizing loss functions in both the spatial and frequency domains, McLDI-INR achieves better alignment with the lensless forward model during back-propagation, ultimately enabling high-fidelity reconstruction of dynamic objects.

2.3 PMFD loss

Since recovering the original image from a single lensless diffraction image is a highly ill-posed problem, existing self-supervised methods often suffer from issues such as information loss and motion artifacts, and constraining solely in the spatial domain is insufficient to enhance high-frequency representation. To address these challenges, we propose a frequency-domain feature compensation mechanism guided by a physical model to regularize the solution space.

As shown in Fig. 1(b), the measured image, being in a defocused plane, exhibits significant attenuation of high-frequency information, which makes direct frequency-domain constraints ineffective for improving reconstruction quality. We find that mapping the measured data onto the sample's in-focus plane via inverse diffraction propagation effectively recovers the spatial frequency characteristics of the target scene, thereby providing a basis for constructing a loss function that is sensitive to high-frequency features.

Therefore, based on the forward model in Eq. (4), we design the corresponding inverse model $B_p(\cdot)$ as

$$B_p(k) = I(k) * PSF(-d_2) * PSF(-d_1) , \quad (9)$$

where $B_p(\cdot)$ is a back-propagation operator that follows the Fresnel diffraction model with a known propagation distance. This operator back-propagates the optical field carrying the structural information of the mask to the in-focus sample plane. After Fourier transformation, it contributes to construct a frequency-domain regularization constraint in the PMFD loss that is sensitive to high-frequency information. The PMFD loss function based on the $B_p(\cdot)$ operator is calculated as follows

$$L_{PMFD} = \frac{1}{K} \sum_{k=1}^K \left| F(h_n(B_p(I_k))) - F(h_n(B_p(I_k')) \right| , \quad (10)$$

where h_n denotes the Hanning window function and F denotes the 2D Fourier transform. The Hanning window is used to suppress spectral leakage, i.e., reduce spurious high-frequency components caused by boundary discontinuities, allowing the frequency-domain loss to focus on true spectral structures. It also improves inter-frame consistency in dynamic or multi-frame scenarios, enhancing training stability. By introducing the PMFD loss in Eq. (10), we impose a regularization constraint in the frequency domain, thereby effectively compensating for motion blur and enabling the reconstruction of finer texture details in dynamic objects.

3 Results and discussions

3.1 Simulation results

3.1.1 Simulation of dynamic scenes

We evaluated McLDI-INR on a custom dataset for dynamic lensless imaging. Both qualitative and quantitative comparisons demonstrate its superior performance over state-of-the-art methods. Under identical parameter settings for two scenarios involving linear and complex nonlinear motion (Fig. 3), PIE²¹ suffers from severe distortion due to its single-frame reconstruction without iterative refinement. Spatiotemporally regularized inversion (STRIVER²⁵) achieves single-shot reconstruction using spatiotemporal priors,

but strong smoothing reduces resolution. The original Nerf⁴¹ shows motion artifacts, and although temporally embedded Nerf (Nerf-T⁵⁴) mitigates some motion blur, it fails to capture high-frequency details. In contrast, McLDI-INR closely approximates the ground truth (GT) by effectively compensating for motion and preserving high-frequency components, achieving higher resolution and fidelity for both linear and nonlinear motions. Its self-supervised training further enables robust adaptation to scene transitions, including natural and microscopic scenes.

Quantitative results (Table 1) confirm these observations: McLDI-INR consistently achieves the best SSIM (Structural Similarity Index Measure), PSNR (Peak Signal-to-Noise Ratio), and LPIPS (Learned Perceptual Image Patch Similarity) across three scenarios.

We also investigated the effect of the encoding mask (Fig. 4). Reconstruction quality is strongly influenced by both feature size and duty cycle: reducing the feature size from 9 μm to 3 μm markedly improves SNR and fidelity, highlighting the superior wavefront modulation of smaller masks. Images reconstructed without a mask exhibit pronounced artifacts, and a duty cycle of 25% results in significant high-frequency loss. Therefore, all experiments employ a mask with a 3 μm feature size and a 50% duty cycle to optimize imaging performance.

3.1.2 Simulation implementation details

Data source. As the integration of INR with dynamic lensless imaging in coherent microscopy systems is still relatively unexplored, there currently exists no standardized

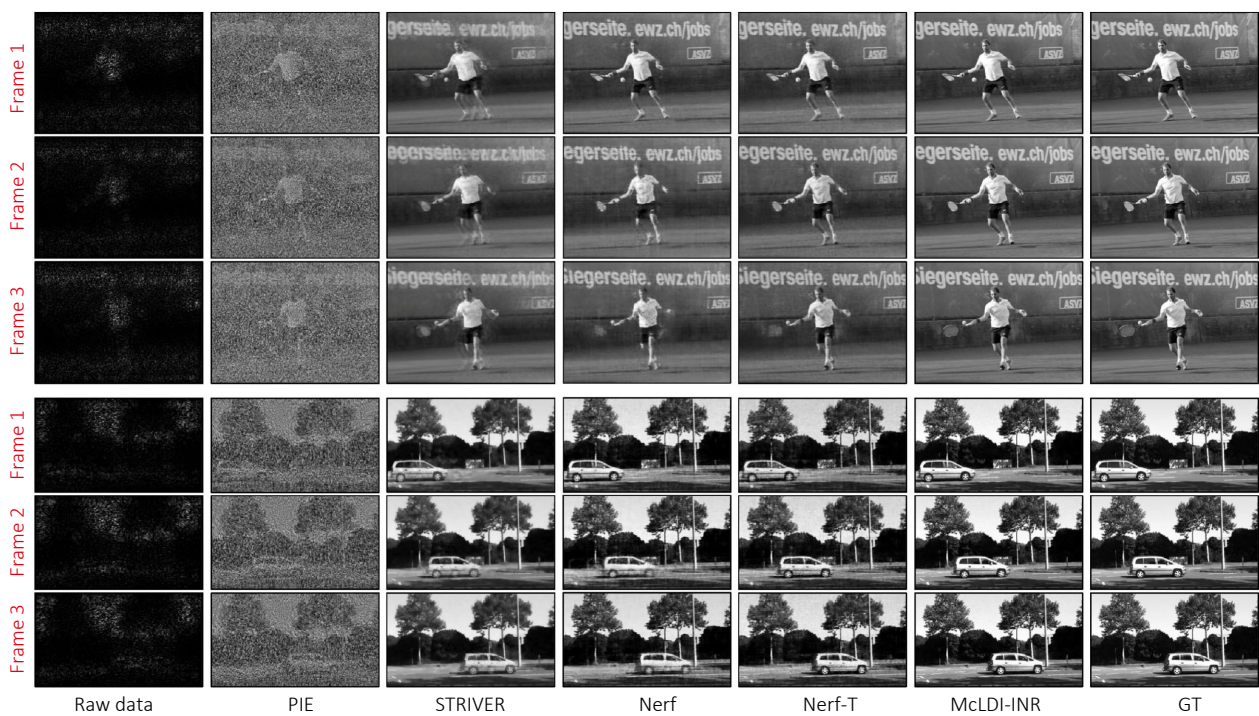


Fig. 3 | visual comparison of McLDI-INR and baselines. We evaluate reconstruction methods in two scenarios. The leftmost column shows raw lensless captures, while the others present three consecutive frames. McLDI-INR effectively suppresses motion blur under both fast linear and complex non-linear motion, achieving superior spatiotemporal resolution. It also preserves static background fidelity, closely matching the ground truth (GT).

Table 1 | Quantitative analysis. Frame-averaged metrics were compared between McLDI-INR and baselines. (Upward arrows ($\uparrow$) indicate that higher values are better, downward arrows ($\downarrow$) indicate that lower values are preferable; boldface highlights the best performance, and underlined values denote second-best results).

Method	Hela cell			Fast car			Tennis sport		
	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$	SSIM $\uparrow$	PSNR $\uparrow$	LPIPS $\downarrow$
PIE ²¹	0.371	30.065	0.604	0.148	27.899	0.725	0.115	28.020	0.777
STRIVER ²⁵	<u>0.907</u>	32.231	<u>0.225</u>	<u>0.795</u>	<u>29.121</u>	<u>0.329</u>	0.791	29.001	0.419
Nerf ⁴¹	0.876	35.593	0.271	0.743	29.034	0.366	0.799	<u>30.963</u>	0.427
Nerf-T ⁵⁴	0.897	<u>36.287</u>	0.249	0.733	28.983	0.380	<u>0.818</u>	30.296	<u>0.392</u>
McLDI-INR	0.964	40.266	0.135	0.920	33.140	0.149	0.941	38.173	0.198

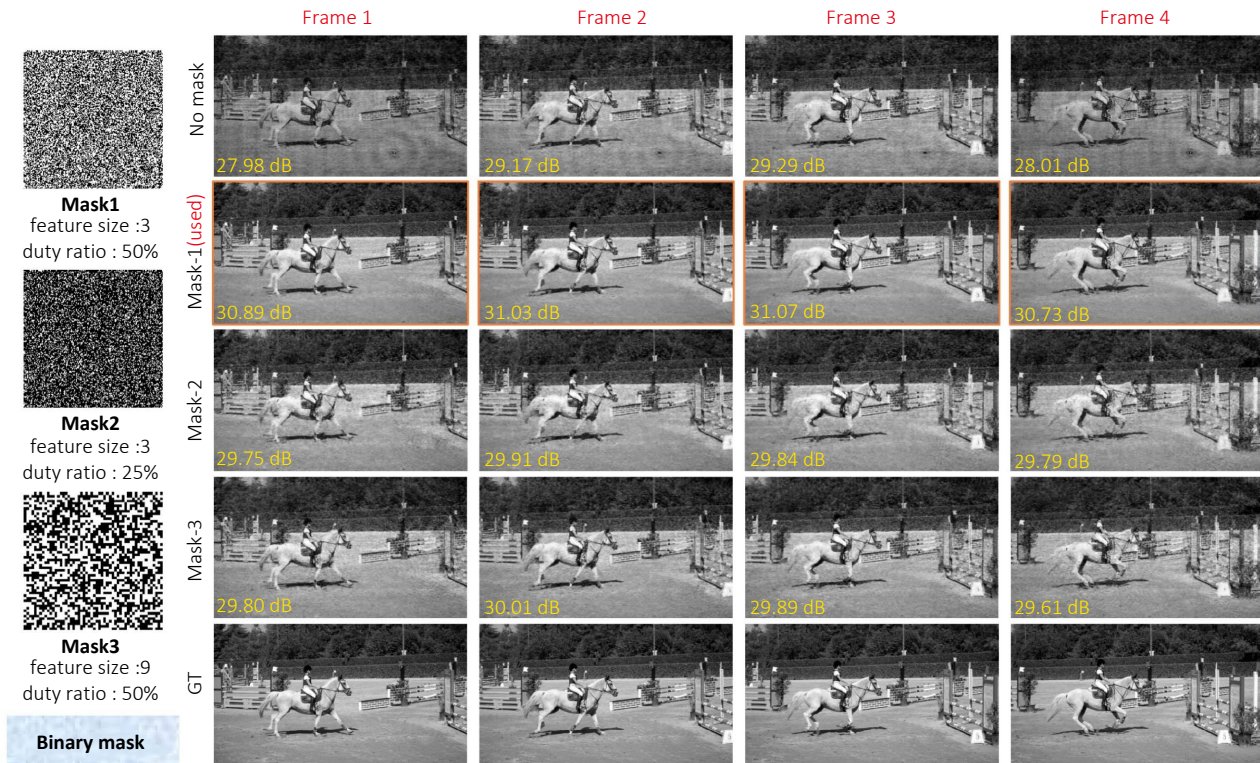


Fig. 4 | Analysis of different mask modulation effects. We compared the reconstruction capability of McLDI-INR for complex dynamic scenes when modulated by masks with varying feature sizes and duty cycles. The lower-left corner of each reconstructed frame is annotated with its peak signal-to-noise ratio (PSNR) calculated against the corresponding ground truth (GT) image. In our experiments, we employed a mask with a feature size of 3 and a duty cycle of 50% to enhance the system's imaging performance.

benchmark for evaluation. To address this, we selected a diverse range of data types to enhance scene variability, including simulations involving complex natural motion scenarios, in order to comprehensively validate the effectiveness of the proposed McLDI-INR method. Specifically, we integrated natural scene data extracted from computer vision conference literature⁵⁶, along with selected subsets from the KITTI dataset⁵⁷, as well as microscopic image samples. Dynamic objects were processed using the forward imaging model Eq. (4) to generate measurements.

Network. Our MLP mapping network g_θ comprises eight fully connected layers with 256 neurons each, using ReLU activations and a sigmoid output. The (x, y) pixel coordinates were normalized to $[-1, 1]$, and the time coordinate t was normalized to $[0, 1]$. In our experiments, the radial coding angle θ was set to 6, the spatial frequency depth L_{xy} was set to 7, and the temporal frequency depth L_t was set to 6.

Imaging system. To generate the simulation data, we followed the mask-constraint lensless dynamic imaging system described above. We used a coherent laser with a wavelength of 473 nm. The mask layer is constructed using randomly generated binary patterns in simulation, with a default feature size of 3 set during the simulation process. A monochrome camera with a pixel size of $1.85 \mu\text{m}$ served as the image sensor. The light source-to-object distance was set

sufficiently large, with an object-to-mask distance of 2.5 mm and a mask-to-sensor distance of 1.5 mm. The mask layer was panned unidirectionally at a rate of 3 pixels per frame.

Training. We trained our network using the Adam optimizer⁵⁸ with an initial learning rate of $1e-3$. An equally spaced learning rate schedule was used, halving the rate every 500 epochs, with a batch size of 1 over 2,000 epochs. In the loss function Eq. (8), we set $\alpha = 0.9$ and $\beta = 0.1$ to balance the spatial-domain and frequency-domain loss. All experiments were conducted on an NVIDIA RTX A100 GPU. For inference-based reconstruction of a 10-frame measurement sequence with a resolution of 300×300 pixels, the GPU memory consumption is approximately 36 GB and the reconstruction time is about 15 minutes.

Baselines and metrics. We evaluated the performance of our method by comparing McLDI-INR with baselines, including traditional lensless imaging methods as well as INR-based methods, as shown below:

1) PIE²¹ is widely used in coherent diffraction imaging, where an iterative algorithm combines diffraction intensity data from multiple local scanning positions to reconstruct the sample's high-resolution complex amplitude distribution. We introduce this method into the single-frame reconstruction scenario of lensless dynamic imaging.

2) STRIVER²⁵ leverages spatiotemporal signal priors to

eliminate phase ambiguities and motion artifacts in lensless dynamic imaging. We employ the official STRIVER implementation with default parameters as a benchmark for comparison.

3) Nerf⁴¹ is commonly employed in static lensless imaging. We adopt a basic spatial coordinate encoding version of Nerf as a baseline for comparison.

4) Nerf-T⁵⁴ is an INR-based dynamic scene reconstruction method. We refer to the time encoding method in D-Nerf⁵⁴ and introduce it into lensless dynamic imaging as a baseline for comparison.

Details and baseline settings. All baseline methods adopt a single-frame reconstruction strategy, where each measured frame corresponds to an independent reconstruction time point. For the conventional method PIE²¹, iterative reconstruction is performed independently on each frame with the number of iterations uniformly set to 100. For STRIVER²⁵, joint reconstruction is performed on the entire video sequence, with 100 outer iterations and 5 inner iterations, while all other parameters are kept at their default settings. For the INR-based baseline methods (NeRF⁴¹ and NeRF-T⁵⁴), the MLP architecture, training epochs, and computational resources are consistent with those of the proposed McLDI-INR method, differing only in the encoding scheme and the use of the PMFD loss, thereby ensuring a fair comparison.

Evaluation metrics. We report the reconstruction quality of each approach with three standard error metrics: SSIM, PSNR, and perceptual similarity through LPIPS⁵⁹, on both the entire scene. The reported values are obtained by averaging the metrics across all reconstructed frames.

3.2 Experiments results

3.2.1 Dynamic imaging experiments

Using a customized mask-constrained lensless imaging system, we conducted imaging experiments on a moving USAF resolution target and dynamic rotifer samples. For the USAF resolution target experiment, object motion was simulated with a motorized translation stage to evaluate the spatiotemporal resolution (see Fig. 5(a)). In the experiments, adjacent frames were captured at intervals of 0.1 s. Reconstruction results (Fig. 5(b, c)) show that STRIVER suffers increasing blur across frames due to large inter-frame displacements and reliance on spatiotemporal continuity priors. In contrast, McLDI-INR encodes temporal information to more accurately represent spatial displacements and applies the PMFD loss to enforce temporal consistency across frames, effectively reducing motion blur. Amplitude profiles (Fig. 5(d)) further demonstrate higher edge contrast for McLDI-INR. Some digits remain missing in the reconstructed chart of McLDI-INR, primarily due to optical throughput loss from the binary mask and physical model deviations, limiting single-frame information capture. To

address this issue, we further propose a multi-frame overlapping reconstruction strategy in the following section to recover missing details and improve overall fidelity. For the swimming rotifer samples, live specimens were placed on a glass slide to enable dynamic imaging. The reconstructed amplitude and phase results from single-frame measurements are shown in Fig. 5(e). The proposed McLDI-INR method successfully resolves the motion details of the rotifers, such as tail contraction and displacement in water, enabling effective reconstruction even for samples undergoing non-rigid motion.

3.2.2 Multi-frame overlapping reconstruction strategy

To mitigate information loss caused by blocked regions of the binary mask and mismatches between the physical and theoretical models in real-world experiments, we introduce a parameter self-adaptation mechanism combined with a Multi-frame Overlapping Reconstruction (MfOR) strategy. By leveraging inter-frame correlations, MfOR effectively compensates for missing information, while the self-adaptation mechanism reduces the impact of modeling errors, enhancing imaging fidelity and robustness. As illustrated in Fig. 6(a), MfOR leverages motion continuity and inter-frame redundancy by integrating multiple frames into a single reconstruction. The step size between consecutive reconstructions is set to 1 to preserve sufficient correlation among the input frames. Figure 6(b) and 6(c) compare MfOR with single-frame reconstruction, demonstrating that MfOR effectively recovers missing USAF digit details and enhances spatial resolution. However, since the inter-frame motion is approximately 8 pixels ($\sim 14.8 \mu\text{m}$), excessively increasing the number of overlapping frames (e.g., $j = 4$ in Fig. 6(c)) introduces significant motion blur, where the total displacement across overlapping frames exceeds 32 pixels ($\sim 59.2 \mu\text{m}$). This degradation primarily arises because object motion reduces inter-frame correlation. Consequently, selecting the number of overlapping frames requires a trade-off governed by the motion characteristics of the scene: faster object motion necessitates fewer overlapping frames. In summary, the MfOR strategy effectively compensates for information loss by exploiting the overlap information between consecutive frames, thereby improving the reconstruction quality.

3.2.3 Experimental implementation details

In real-world experiments, our MLP mapping network g_θ adopts a 6-layer MLP with 128 neurons per layer to reduce memory consumption. We also observed that using a larger-scale MLP yields negligible performance gains in real experiments due to calibration errors in system parameters. To better capture high-frequency spatial variations, we increased the radial encoding frequency depth for the (x, y) axes to $L_{xy} = 8$. Additionally, the weight coefficient of the PMFD loss was set to $\beta = 0.01$ to achieve improved reconstruction quality. All other parameters remain consistent

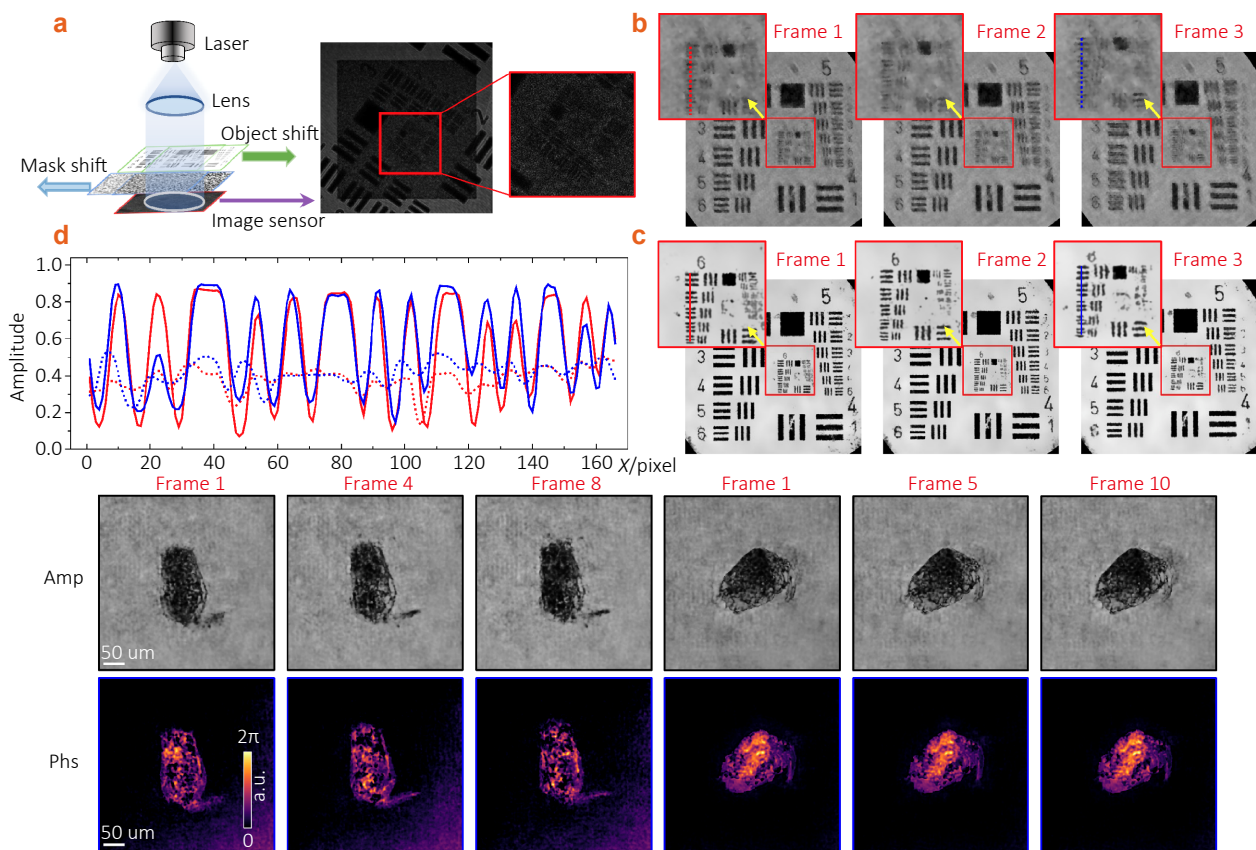


Fig. 5 | Real-world system design and experimental results. (a) System design and measurements: the real system closely follows the simulation setup. Wide-field measurements are captured using an image sensor, and a certain region under the mask (700×700 pixels) is selected for reconstruction. (b, c) Reconstruction results of STRIVER and McLDI-INR. Yellow arrows indicate the motion direction of the USAF target. Red boxes highlight the 6th and 7th element groups. McLDI-INR achieves higher resolution, markedly reduces motion blur, and produces a clearer background. (d) Amplitude distributions along the red and blue lines in (b) and (c) show that McLDI-INR produces noticeably sharper profiles than STRIVER across three individual frames. (e) Amplitude and phase results of two swimming rotifer samples using McLDI-INR. Our method successfully recovers the motion details of rotifers, such as tail contraction and displacement in water.

with those used in the simulation. For inference-based reconstruction of a 10-frame measurement sequence with a resolution of 400×400 pixels, the GPU memory consumption is approximately 39 GB and the reconstruction time is about 20 minutes.

In the real experimental system, as shown in Fig. 4(a), we place a lens directly beneath the laser to collimate the beam, thereby achieving more uniform illumination. The MBL-FN-473 laser light source from changchun new industries optoelectronics technology Co., Ltd. was used, providing coherent light with a wavelength of 473nm . Two motorized translation stages move the sample and the mask in opposite directions to enhance the diversity of measurements. The mask is fabricated on a silica substrate via direct laser writing: a thin chromium coating on the substrate is patterned with a random binary motif of feature size $5\text{mm} \times 5\text{mm}$. The image sensor used was the Image Source DMM DMK 33UX226 USB 3.0 monochrome camera, featuring a pixel size of $1.85\text{ }\mu\text{m}$.

3.2.4 Parameter calibration

To enable accurate algorithmic reconstruction, the acquired experimental images first undergo a calibration process to extract the relevant physical parameters. Specifically, we employ the sparsity of the gradient (SoG⁶⁰) method to estimate the axial distances in the experimental setup, namely the sample-to-mask distance d_1 and the mask-to-sensor distance d_2 . In addition, we adopt a displacement calibration technique⁶¹ based on discrete Fourier transform (DFT) and cross-correlation to compute the positional offset $(\Delta x_i, \Delta y_j)$ between the captured and the designed mask patterns for each acquisition.

However, the parameters obtained through the above calibration procedures inevitably contain errors due to limitations in measurement accuracy. In practical imaging systems, however, due to the non-ideal uniform planar light source and the inability to maintain perfectly horizontal sample and encoding planes, the parameters obtained through the aforementioned calibration procedure inevitably contain errors. To address this, we treat d_1 and d_2 as

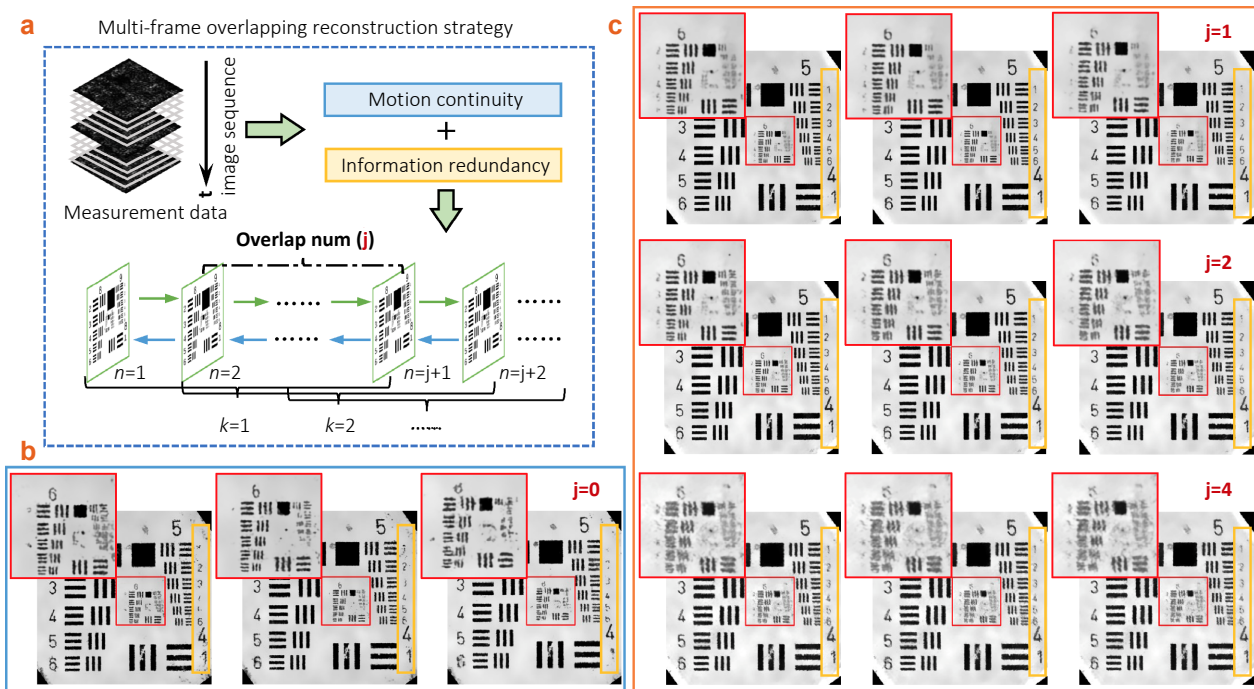


Fig. 6 | Multi-frame overlapping reconstruction (MfOR) strategy and USAF results. (a) MfOR principle: Given the continuity of object motion and high inter-frame redundancy across N raw measurements, we adopt the overlapping reconstruction. Multiple frames are merged into a single reconstruction to enhance information recovery. The number of overlapping frames between adjacent reconstructions is denoted as j . (b) Single-frame reconstruction ($j = 0$): Each frame is reconstructed independently. The yellow rectangle highlights a region of the USAF resolution target containing digits. (c) MfOR results with varying overlap ($j = 1, 2, 4$): in these cases, digits within the yellow rectangle are completely recovered. Moderate overlap improves information compensation and reconstruction fidelity, while excessive overlap can introduce motion blur in high-frequency regions.

learnable parameters during the neural network training phase and optimize them using gradient descent, thereby enabling the recovery of more accurate physical parameters for improved reconstruction fidelity

3.3 Ablation experiments

To validate the effectiveness of McLDI-INR, we conducted ablation experiments under five conditions: without mask-constraint (w/o Mask), without temporal embedding (w/o T Embed), without PMFD loss (w/o L_{PMFD}), without both temporal embedding and PMFD loss (w/o T Embed, L_{PMFD}), and without both mask-constraint and PMFD loss (w/o Mask, L_{PMFD}). Experimental results are presented in Fig. 7(a). The quantitative analysis is shown in Table 2.

Excluding the mask constraint alone produces reconstructions with high SSIM but markedly lower PSNR. Omitting the PMFD loss leads to a loss of high-frequency details and reduced cell resolution. While removing temporal embedding slightly improves reconstruction, likely due to the high inter-frame similarity from slow cell motion, its performance still falls short of the complete McLDI-INR system. Notably, simultaneously removing temporal embedding, the mask constraint, and the PMFD loss severely degrades image quality, highlighting the PMFD loss's critical role in preserving high-frequency details. Error maps between the

reconstructed third frame and the ground truth clearly illustrate the impact of each component: the PMFD component contributes the most, and the full strategy exhibits the smallest error. Additional quantitative analyses across various scenarios (Table 2) further confirm the relative importance of each component, with the full strategy achieving the best overall performance.

To further validate the necessity of the PMFD loss, we compared the spectral characteristics of the reconstructed images with and without this loss function, as shown in Fig. 7(b). It can be observed that, after introducing the PMFD loss based on the $B_p(\cdot)$ operator, the range of high-frequency details in the reconstructed images is significantly expanded. In particular, the magnified high-frequency texture details within the red and green boxed regions exhibit notably improved reconstruction quality.

4 Conclusions

We propose McLDI-INR, a mask-constraint lensless dynamic imaging framework based on implicit neural representations. By integrating a physical model into a spatial-frequency dual-domain collaborative loss, our method effectively regularizes the solution space and achieves high spatiotemporal resolution. Quantitative and qualitative experiments show that McLDI-INR outperforms

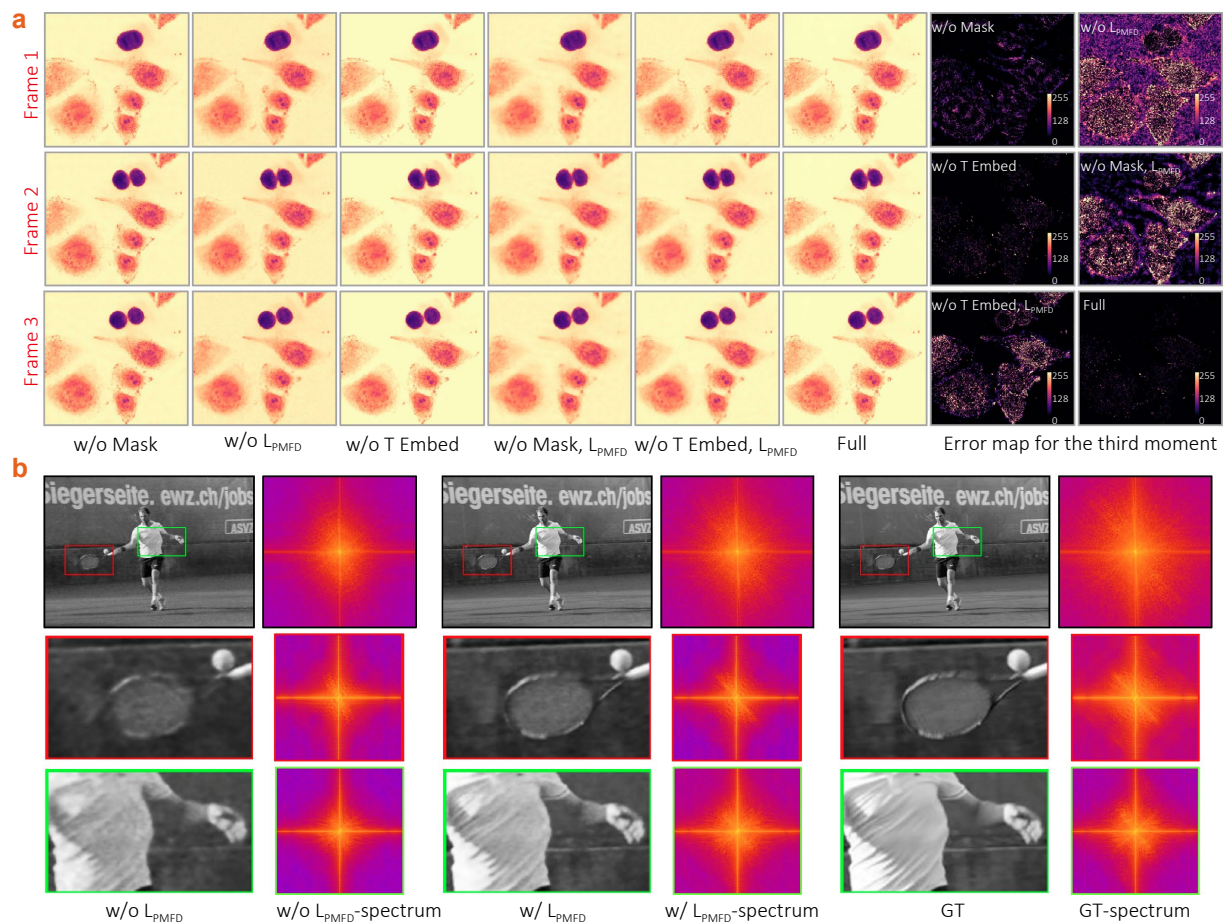


Fig. 7 | Ablation experiments. (a) In the lensless imaging of HELA cells, we compared the results (shown in pseudocolor) obtained with the full McLDI-INR method against variants with individual components removed, using a consistent error map at the third time point relative to the ground truth. The results indicate that each component affects image quality in a distinct manner, with the frequency-domain loss component, L_{PMFD} , having the most pronounced impact on reconstruction. (b) In the spectral analysis experiment, we compared the impact of L_{PMFD} on the spectrum of the reconstructed images. The results show that incorporating the L_{PMFD} significantly expands the range of high-frequency details in the reconstructed images.

Table 2 | Ablation experiments. Quantitative metrics for each component removal.

Method	Hela cell			Fast car			Tennis sport		
	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓	SSIM↑	PSNR↑	LPIPS↓
w/o Mask	0.968	35.361	0.197	0.869	30.702	0.206	0.871	31.494	0.294
w/o L_{PMFD}	0.897	30.507	0.323	0.880	31.437	0.218	0.868	31.428	0.328
w/o T Embed	0.952	<u>39.632</u>	<u>0.147</u>	<u>0.899</u>	<u>32.262</u>	<u>0.194</u>	<u>0.901</u>	<u>36.576</u>	<u>0.262</u>
w/o Mask, L_{PMFD}	0.875	31.975	0.349	0.772	30.137	0.338	0.812	31.795	0.387
w/o T Embed, L_{PMFD}	0.898	33.712	0.303	0.831	30.405	0.280	0.815	32.679	0.401
Full	<u>0.964</u>	40.266	0.135	0.920	33.140	0.149	0.941	38.173	0.198

existing computational frameworks, recovering high-quality images in complex scenarios while suppressing motion blur. Simulation studies further demonstrate its robustness across diverse imaging conditions, and real-world experiments validate its practicality and effectiveness. The framework is flexible and generalizable, with potential extensions to 3D dynamic lensless imaging and light-field microscopy.

Like conventional INR methods, McLDI-INR uses an MLP to map spatiotemporal coordinates to pixel values, which increases computational demands at higher resolutions. It also does not explicitly address occlusions, which can degrade reconstruction when inter-frame information is inconsistent. The PMFD loss may be compromised by inaccuracies in the physical model, and irreversible mask

modulation inherently limits available information for its computation.

Future work may include optimizing the encoding scheme to reduce the computational cost of ultra-wide field-of-view imaging, exploring reversible mask designs to enhance the PMFD loss and further improve reconstruction performance, and addressing the degradation of the MfOR strategy in scenarios involving rapid or highly non-rigid deformations by introducing enhanced approaches with explicit motion compensation mechanisms. Specifically, optical flow estimation⁶² or a deformable registration module⁶³ could be incorporated prior to reconstruction to perform pixel-level pre-alignment of the input frames, thereby restoring effective inter-frame correlation in high-speed dynamic scenes.

References

- Boominathan V, Robinson JT, Waller L et al. Recent advances in lensless imaging. *Optica* 9, 1–16 (2022).
- Sinha A, Lee J, Li S et al. Lensless computational imaging through deep learning. *Optica* 4, 1117–1125 (2017).
- Jiang SW, Song PM, Wang TB et al. Spatial- and Fourier-domain ptychography for high-throughput bio-imaging. *Nat Protoc* 18, 2051–2083 (2023).
- Zheng GA, Shen C, Jiang SW et al. Concept, implementations and applications of Fourier ptychography. *Nat Rev Phys* 3, 207–223 (2021).
- Gurkan UA, Moon S, Geckil H et al. Miniaturized lensless imaging systems for cell and microorganism visualization in point-of-care testing. *Biotechnol J* 6, 138–149 (2011).
- Adams JK, Boominathan V, Avants BW et al. Single-frame 3D fluorescence microscopy with ultraminiature lensless FlatScope. *Sci Adv* 3, e1701548 (2017).
- Lu CC, Zhou Y, Guo YX et al. Mask-modulated lensless imaging via translated structured illumination. *Opt Express* 29, 12491–12501 (2021).
- Zhang ZB, Zhou Y, Jiang SW et al. Invited article: mask-modulated lensless imaging with multi-angle illuminations. *APL Photon* 3, 060803 (2018).
- Adams JK, Yan D, Wu JM et al. In vivo lensless microscopy via a phase mask generating diffraction patterns with high-contrast contours. *Nat Biomed Eng* 6, 617–628 (2022).
- Wu JC, Zhang H, Zhang WH et al. Single-shot lensless imaging with fresnel zone aperture and incoherent illumination. *Light Sci Appl* 9, 53 (2020).
- Zhou Y, Wu JM, Suo JL et al. Single-shot lensless imaging via simultaneous multi-angle LED illumination. *Opt Express* 26, 21418–21432 (2018).
- Zhou Y, Hua X, Zhang ZB et al. Wirtinger gradient descent optimization for reducing Gaussian noise in lensless microscopy. *Opt Lasers Eng* 134, 106131 (2020).
- Shin J, Tran DN, Stroud JR et al. A minimally invasive lens-free computational microendoscope. *Sci Adv* 5, eaaw5595 (2019).
- Liu TR, Li YZ, Koydemir HC et al. Rapid and stain-free quantification of viral plaque via lens-free holography and deep learning. *Nat Biomed Eng* 7, 1040–1052 (2023).
- Jiang SW, Guo CF, Bian ZC et al. Ptychographic sensor for large-scale lensless microbial monitoring with high spatiotemporal resolution. *Biosens Bioelectron* 196, 113699 (2022).
- Song PM, Guo CF, Jiang SW et al. Optofluidic ptychography on a chip. *Lab Chip* 21, 4549–4556 (2021).
- Zhang JL, Sun JS, Chen Q et al. Resolution analysis in a lens-free on-chip digital holographic microscope. *IEEE Trans Comput Imaging* 6, 697–710 (2020).
- Luo ZX, Yurt A, Stahl R et al. Pixel super-resolution for lens-free holographic microscopy using deep learning neural networks. *Opt Express* 27, 13581–13595 (2019).
- Jain P, Nazar AM, Khan SS et al. FlatTrack: eye-tracking with ultra-thin lensless cameras. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops* 396–404 (IEEE, 2025). [10.1109/WACVW65960.2025.00050](https://doi.org/10.1109/WACVW65960.2025.00050).
- Edwards HB. Lensless scanning telescope. *Appl Opt* 19, 504–507 (1980).
- Maiden A, Johnson D, Li P. Further improvements to the ptychographical iterative engine. *Optica* 4, 736–745 (2017).
- Antipa N, Kuo G, Heckel R et al. DiffuserCam: lensless single-exposure 3D imaging. *Optica* 5, 1–9 (2018).
- Boominathan V, Adams JK, Robinson JT et al. Phlatcam: designed phase-mask based thin lensless camera. *IEEE Trans Pattern Anal Mach Intell* 42, 1618–1629 (2020).
- Reddy D, Veeraraghavan A, Chellappa R. P2C2: programmable pixel compressive camera for high speed imaging. In *CVPR 2011* 329–336 (IEEE, 2011). [10.1109/CVPR.2011.5995542](https://doi.org/10.1109/CVPR.2011.5995542).
- Gao YH, Cao LC. Motion-resolved, reference-free holographic imaging via spatiotemporally regularized inversion. *Optica* 11, 32–41 (2024).
- Zhang WH, Cao LC, Brady DJ et al. Twin-image-free holography: a compressive sensing approach. *Phys Rev Lett* 121, 093902 (2018).
- Denis L, Lorenz D, Thiébaud E et al. Inline hologram reconstruction with sparsity constraints. *Opt Lett* 34, 3475–3477 (2009).
- Oh J, Hugonnet H, Park Y. Non-interferometric stand-alone single-shot holographic camera using reciprocal diffractive imaging. *Nat Commun* 14, 4870 (2023).
- Zhu H, Liu Z, Zhou Y et al. DNF: diffractive neural field for lensless microscopic imaging. *Opt Express* 30, 18168–18178 (2022).
- Jiang SW, Guo CF, Hu P et al. High-throughput lensless whole slide imaging via continuous height-varying modulation of a tilted sensor. *Opt Lett* 46, 5212–5215 (2021).
- Zhou Y, Song WZ, Wu LY et al. Large-scale lensless microscopy with fast acquisition and region-wise focusing. *Laser Photon Rev* 18, 2400498 (2024).
- Song PM, Wang RH, Zhu JK et al. Super-resolved multispectral lensless microscopy via angle-tilted, wavelength-multiplexed ptychographic modulation. *Opt Lett* 45, 3486–3489 (2020).
- Wu XJ, Zhou N, Chen Y et al. Lens-free on-chip 3D microscopy based on wavelength-scanning Fourier ptychographic diffraction tomography. *Light Sci Appl* 13, 237 (2024).
- Cai X, You ZY, Zhang HL et al. PhoCoLens: photorealistic and consistent reconstruction in lensless imaging. In *Proceedings of the 38th International Conference on Neural Information Processing Systems* 391 (Curran Associates Inc., 2024).
- Khan SS, Sundar V, Boominathan V et al. Flatnet: towards photorealistic scene reconstruction from lensless measurements. *IEEE Trans Pattern Anal Mach Intell* 44, 1934–1948 (2022).
- Huang LZ, Chen HL, Liu TR et al. Self-supervised learning of hologram reconstruction using physics consistency. *Nat Mach Intell* 5, 895–907 (2023).
- Cao RM, Divekar NS, Nuñez JK et al. Neural space-time model for dynamic multi-shot imaging. *Nat Methods* 21, 2336–2341 (2024).
- Zeng TJ, Lam EY. Robust reconstruction with deep learning to handle model mismatch in lensless imaging. *IEEE Trans Comput Imaging* 7, 1080–1092 (2021).
- Kingshott O, Antipa N, Bostan E et al. Unrolled primal-dual networks for lensless cameras. *Opt Express* 30, 46324–46335 (2022).

- (2022).
40. Li Y, Li ZD, Chen KY et al. MWDNs: reconstruction in multi-scale feature spaces for lensless imaging. *Opt Express* **31**, 39088–39101 (2023).
 41. Mildenhall B, Srinivasan PP, Tancik M et al. NeRF: representing scenes as neural radiance fields for view synthesis. *Commun ACM* **65**, 99–106 (2022).
 42. Zhou Y, Xu CY, Jin ZY et al. Physics-informed ellipsoidal coordinate encoding implicit neural representation for high-resolution volumetric wide-field microscopy. Preprint at <https://doi.org/10.1101/2024.10.17.618813> (2024).
 43. Liu RH, Sun Y, Zhu JB et al. Recovery of continuous 3D refractive index maps from discrete intensity-only measurements using neural fields. *Nat Mach Intell* **4**, 781–791 (2022).
 44. Zheng JQ, Ramasinghe S, Lucey S. Rethinking positional encoding. In *The 10th International Conference on Learning Representations* (OpenReview. net, 2022).
 45. Tancik M, Srinivasan P, Mildenhall B et al. Fourier features let networks learn high frequency functions in low dimensional domains. In *Proceedings of the 34th International Conference on Neural Information Processing Systems* 632 (Curran Associates Inc. , 2020).
 46. Xie YH, Takikawa T, Saito S et al. Neural fields in visual computing and beyond. *Comput Graph Forum* **41**, 641–676 (2022).
 47. Gao K, Gao Y, He H, et al. Neural radiance fields in 3D vision: A comprehensive review. *Computational Visual Media*, 10.26599/CVM.2026.9450535, 1-55 (2026).
 48. Madavan RR, Kaimal A, V BK et al. GANESH: generalizable NeRF for lensless imaging. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* 9499–9508 (IEEE, 2025). [10.1109/WACV61041.2025.00920](https://doi.org/10.1109/WACV61041.2025.00920).
 49. Banerjee A, Singh S. Towards lensless image deblurring with prior-embedded implicit neural representations in the low-data regime. *Eng Appl Artif Intell* **162**, 112488 (2025).
 50. Zhao JY, Zhao ZF, Wu JM et al. PNR: physics-informed Neural Representation for high-resolution LFM reconstruction. Preprint at <https://arxiv.org/abs/2409.18223> (2024).
 51. Jin YB, Lu LP, Zhou S et al. Neural-field-assisted transport-of-intensity phase microscopy: partially coherent quantitative phase imaging under unknown defocus distance. *Photon Res* **12**, 1494–1501 (2024).
 52. Chen H, He B, Wang HY et al. NeRV: neural representations for videos. In *Proceedings of the 35th International Conference on Neural Information Processing Systems* 1649 (Curran Associates Inc. , 2021).
 53. Li ZQ, Niklaus S, Snavely N et al. Neural scene flow fields for space-time view synthesis of dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 6494–6504 (IEEE, 2021). [10.1109/CVPR46437.2021.00643](https://doi.org/10.1109/CVPR46437.2021.00643)
 54. Pumarola A, Corona E, Pons-Moll G et al. D-NeRF: neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 10313–10322 (IEEE, 2021). [10.1109/CVPR46437.2021.01018](https://doi.org/10.1109/CVPR46437.2021.01018).
 55. Roberts RA, Mullis CT. *Digital Signal Processing* (Addison-Wesley Longman Publishing, Boston, 1987).
 56. Wang QQ, Chang YY, Cai RJ et al. Tracking everything everywhere all at once. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* 19738–19749 (IEEE, 2023). [10.1109/ICCV51070.2023.01813](https://doi.org/10.1109/ICCV51070.2023.01813).
 57. Geiger A, Lenz P, Stiller C et al. Vision meets robotics: the KITTI dataset. *Int J Robot Res* **32**, 1231–1237 (2013).
 58. Kingma DP, Ba J. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations* (OpenReview. net, 2015).
 59. Zhang R, Isola P, Efros AA et al. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 586–595 (IEEE, 2018). [10.1109/CVPR.2018.00068](https://doi.org/10.1109/CVPR.2018.00068).
 60. Zhang YB, Wang HD, Wu YC et al. Edge sparsity criterion for robust holographic autofocusing. *Opt Lett* **42**, 3824–3827 (2017).
 61. Guizar-Sicairos M, Thurman ST, Fienup JR. Efficient subpixel image registration algorithms. *Opt Lett* **33**, 156–158 (2008).
 62. Teed Z, Deng J. RAFT: recurrent all-pairs field transforms for optical flow. In *European Conference on Computer Vision* 402–419 (Springer, 2020). [10.1007/978-3-030-58536-5_24](https://doi.org/10.1007/978-3-030-58536-5_24).
 63. Zhuo YT, Shen YQ. DiffuseReg: denoising diffusion model for obtaining deformation fields in unsupervised deformable image registration. In *International Conference on Medical Image Computing and Computer Assisted Intervention* 597–607 (Springer, 2024). [10.1007/978-3-031-72069-7_56](https://doi.org/10.1007/978-3-031-72069-7_56).

Acknowledgements

This work was supported by the Jiangsu Association for Science and Technology (JAST) Young Elite Scientists Sponsorship Program (No. JSTJ-2025-730), National Natural Science Foundation of China (No. 62371006), and the Beijing Natural Science Foundation (No. 3242008).

Author contributions

W.L., W.-Z.S., and Y.Z. conceived the idea and developed the theory. W.L. performed simulations of the method. W.L. and W.-Z.S. designed the imaging system and conducted the experiments. W.L. and C.-Y.X. wrote the manuscript. Y.Z. and B.X. supervised the overall project. All authors discussed the results, did the data analysis, commented on the paper, and contributed to the writing of manuscript.

Competing interests

The authors declare no competing financial interests.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> ©The Author(s) 2026.