Research Article | Vol. 34, No. 13 / 29 Jun 2026 / Optics Express 24062
Optics EXPRESS

High-speed 3D light field sensing via snapshot compressive acquisition and domain-adaptive deep equilibrium reconstruction

RUIXUE WANG,[1] XUE WANG,[1] ZHAOLIN XIAO,[2] GUOQING ZHOU,[1] AND QING WANG[1,*]

[1]School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China
[2]Department of Software Engineering, School of Computer Science and Engineering, Xi'an University of Technology, Xi'an, Shaanxi 710048, China
[*]qwang@nwpu.edu.cn

Abstract: Constrained by the finite space-bandwidth product of sensors, 3D light field (LF) sensing using consumer LF cameras is often restricted to single-digit frame rates. To bridge this gap, we present a physical snapshot compressive acquisition system that encodes multiple 5D LF frames into a single 4D measurement via a digital micromirror device (DMD). To recover high-fidelity LF data, we propose the domain-adaptive multi-stage deep equilibrium (DAM-DEQ) framework. This model replaces uniform processing strategies with a structure-aware design that first integrates spatial and angular priors, followed by a multi-domain fusion stage where parallel spatial, angular, and epipolar features are weighted adaptively and refined via spatiotemporal convolution. Experiments on synthetic (Sintel), real (Lytro), and custom array-camera datasets demonstrate that our method outperforms existing techniques and DEQ baselines in terms of PSNR and SSIM. Furthermore, our system achieves effective reconstruction fidelity even on physical measurements degraded by optical aberrations and mask misalignment, confirming the practical viability of the proposed hardware-algorithm co-design.

1. Introduction

Three-dimensional (3D) information acquisition is a cornerstone of modern optical systems, enabling applications ranging from autonomous driving and robotics to biomedical microscopy and immersive displays. Among various 3D sensing modalities, Light Field (LF) imaging stands out for its ability to capture the spatial, angular, and radiometric distribution of light rays in a single exposure. Typically represented by the four-dimensional plenoptic function [1], this parameterization allows for versatile computational operations, such as post-capture refocusing [2], aberration correction [3], and view synthesis [4,5]. LF imaging has demonstrated significant success in capturing static 3D scenes for computational photography [6], rendering [7,8], and a wide range of cross-disciplinary applications [9–11]. Consequently, high-fidelity LF acquisition is particularly critical for emerging technologies in 3D visualization, including wearable augmented reality (AR) displays and autostereoscopic screens, where the accurate reproduction of parallax and depth cues is essential for visual comfort and immersion.

However, in dynamic real-world environments, the finite space–bandwidth product (SBP) of image sensors imposes a fundamental trade-off between temporal and spatial–angular resolutions [12]. Capturing high-density angular information for robust 3D depth sensing inherently consumes sensor pixels, sacrificing spatial resolution. When temporal resolution is introduced for dynamic scene recording (i.e., 5D dynamic LF $L(x, y, u, v, t)$), the data throughput requirements exceed the readout speeds of standard sensors. This limitation often restricts conventional LF cameras to single-digit frame rates, rendering them inadequate for capturing transient phenomena or fast-moving objects, which creates a bottleneck in dynamic 3D sensing and content creation.

#599949 https://doi.org/10.1364/OE.599949
Journal © 2026 Received 26 Mar 2026; accepted 11 Jun 2026; published 22 Jun 2026

Optics EXPRESS

To acquire higher-dimensional spatiotemporal LF data and overcome this resolution trade-off, prior efforts have focused on optical design [11,13,14], electronic implementation [12,15], and algorithmic reconstruction [16,17]. However, these approaches often rely on bulky hardware modifications or scanning mechanisms that introduce motion artifacts due to sequential capture. Recently, computational imaging, particularly Snapshot Compressive Imaging (SCI), has emerged as a promising solution. By leveraging the co-design of hardware and algorithms, SCI facilitates an end-to-end pipeline that enables the recovery of high-dimensional data from low-dimensional measurements, ultimately enhancing temporal resolution and data efficiency [10,18].

In this work, we propose a high-speed 3D LF sensing framework based on snapshot compressive acquisition. As illustrated in Fig. 1, our system integrates a Digital Micromirror Device (DMD) into the optical path to perform high-rate temporal modulation within a single exposure time. This effectively multiplexes temporal information into a single coded 4D measurement without sacrificing the crucial spatial–angular sampling pattern required for 3D perception [Fig. 1(a)]. To resolve the ill-posed inverse problem of recovering the 5D LF from the compressed 4D snapshot, we introduce a Domain-Adaptive Multi-stage Deep Equilibrium (DAM-DEQ) decoder [Fig. 1(b)]. Unlike standard deep unfolding networks that are memory-intensive, our approach leverages the implicit layer formulation of DEQ to achieve high-performance reconstruction with constant memory costs. Crucially, this architecture enables a hardware-algorithm co-design strategy [Fig. 1(c)] in which the final multi-domain fusion stage serves a dual purpose, acting not only as a refinement module to fuse diverse priors but also as a learnable domain adapter that compensates for hardware-specific imperfections.

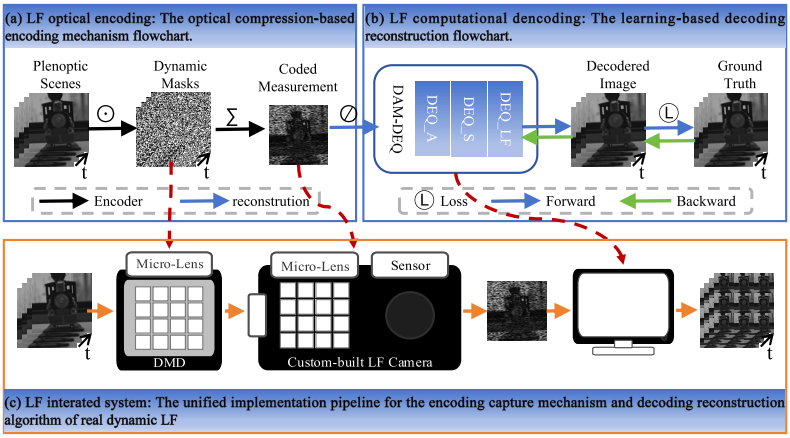

Fig. 1. Overall framework of the proposed dynamic LF snapshot compressive imaging system. (a) Optical encoder: dynamic LF frames are temporally modulated by high-rate random masks and compressed into a single coded measurement. (b) Computational decoder: a domain-adaptive multi-stage DEQ (DAM-DEQ) module reconstructs high-frame-rate LF sequences. (c) Integrated system: unified acquisition–reconstruction pipeline for dynamic LF acquisition and reconstruction.

Building upon our preliminary exploration [19], this paper presents a complete hardware-algorithm co-design tailored for dynamic 3D sensing. Our specific contributions are as follows:

- **Domain-Adaptive Deep Equilibrium Reconstruction.** We propose the DAM-DEQ model, which replaces the uniform processing in [19] with a sequential decomposition into spatial, angular, and multi-domain fusion stages. By employing lightweight DEQ submodules for each stage, this multi-stage design enables efficient 5D LF reconstruction

Optics EXPRESS

from a 4D measurement, ensuring numerical stability and constant memory usage while explicitly capturing structural priors.

- **Prototype Construction and Physical Validation.** Distinct from [19] and other studies relying on simulated encoding, we construct a prototype system integrating a DMD with a custom-built microlens-array camera. Integrated with our reconstruction model, this hardware implementation validates the physical feasibility of the proposed encoding scheme and achieves flexible frame-rate upscaling determined by the compression ratio.
- **Robustness and Sim-to-Real Generalization.** We provide a comprehensive evaluation on synthetic (Sintel), real (Lytro), and custom array-camera datasets. The results demonstrate effective improvements in reconstruction quality. Significantly, we demonstrate that our model, despite being trained primarily on simulated data with uniform masks, generalizes well to real-world prototype measurements. This effectively bridges the "sim-to-real" gap caused by optical aberrations and pixel-to-microlens misalignment that were previously unaccounted for in simulation-based studies.

The remainder of this paper is organized as follows. Section 2 reviews related work on computational imaging for dynamic LFs and compressive reconstruction methods. Section 3 details the proposed method, including the mathematical formulation of the forward model, the architecture of the DAM-DEQ framework, and the construction of the hardware prototype. Section 4 presents the experimental results, encompassing quantitative evaluations on synthetic and real-world datasets, qualitative comparisons, and ablation studies. Finally, Section 5 concludes the paper and discusses future directions.

2. Related work

2.1. Acquisition strategies for dynamic light fields

Enhancing the temporal resolution ofLF imaging typically involves navigating the trade-off between spatial, angular, and temporal dimensions. Approaches generally fall into hardware-centric or algorithm-centric categories.

Hardware-centric strategies primarily utilize parallel acquisition or specialized optics. Camera arrays [16,17,20] offer high performance but are bulky and costly. Microlens array (MLA) cameras [13,21] and LF microscopes [10,11,18] are compact but inherently bound by the sensor's readout frame rate. Other modalities include phase modulation and event-based encoding [15,22], which offer high temporal resolution but often sacrifice spatial fidelity or color information due to the sparsity of event streams or diffraction limits.

Algorithm-centric strategies, particularly coded aperture techniques, attempt to multiplex temporal information into a single exposure. Early methods required multiple coded shots or dictionary learning, significantly increasing reconstruction complexity [23,24]. Sakai et al. [25] utilized alternating patterns but were limited to two distinct time points. Mizuno et al. [26] synchronized aperture and pixel encoding, yet such methods often rely on mode-switching assumptions, making them vulnerable to complex motion artifacts. Additionally, components like polarization modulators introduce significant light loss, degrading the signal-to-noise ratio in high-speed scenarios. Habuchi et al. [27] proposed a computational imaging method that combines encoded apertures and event cameras.

While Deep Learning (DL) has shown promise in optimizing encoding patterns [28], applying it to 5D LF data remains challenging due to the curse of dimensionality. Unlike previous approaches that rely on heavy hardware or restrictive motion models, our work adopts the SCI paradigm. We leverage a DMD to perform high-speed temporal masking, effectively compressing 5D data into 4D measurements without compromising the optical throughput or the MLA's angular sampling structure.

Optics EXPRESS

2.2. Compressive reconstruction and deep equilibrium models

The inverse problem of recovering high-dimensional video or LF data from compressed measurements has evolved from iterative optimization to data-driven approaches.

Iterative and Feed-forward Methods. Early SCI reconstruction relied on model-based priors (e.g., sparsity, Total Variation) solved via iterative algorithms [29,30]. While theoretically sound, they are computationally expensive and slow. End-to-end Convolutional Neural Networks (CNNs) [31] improved speed but often lack physical interpretability. To bridge this gap, deep unfolding networks [32,33] and Plug-and-Play (PnP) priors [34] were introduced, treating iterations as network layers. However, most existing SCI algorithms are tailored for monocular video (3D data: x, y, t), and simply extending them to LFs (5D data: x, y, u, v, t) ignores the rich angular coherence and epipolar geometry essential for depth perception. Furthermore, deep unfolding becomes memory-intensive as the number of stages increases, limiting its depth for high-dimensional LF reconstruction.

Deep Equilibrium (DEQ) Models. The DEQ framework [35] solves for the fixed point of a nonlinear operator, allowing for infinite effective depth with constant memory cost via implicit differentiation. This property is particularly advantageous for ill-posed inverse problems [36,37]. In the context of LF, our previous work (DLFDEQ) [19] applied a densely connected DEQ to dynamic LFs. However, it relied on the assumption of view-consistent encoding (i.e., identical masks across viewpoints) typical of ideal simulations, which allowed for a uniform treatment of the LF structure using non-standard 4D convolution. Consequently, it failed to exploit domain-specific correlations and did not account for the inter-view mask inconsistencies inherent in physical DMD-based systems, where encoding patterns vary across viewpoints due to optical misalignment.

Differentiation of Proposed Work. The primary distinction of this work lies in the transition from simulation to physical realization. While our previous work, DLFDEQ [19], established the theoretical potential of DEQ for dynamic LF reconstruction using simulated compression, it treated the high-dimensional LF tensor uniformly. To confront real-world optical challenges (e.g., aberrations, alignment errors) inherent in a tangible hardware prototype, we introduce the DAM-DEQ model. Unlike generic video SCI methods that treat frames as 2D sequences or prior LF approaches that ignore structural priors, DAM-DEQ explicitly decomposes the inverse problem into spatial, angular, and multi-domain sub-problems. This enables the integration of domain-specific regularizers within the equilibrium framework. Ultimately, this hardware-algorithm co-design moves beyond simulation-based studies, ensuring robust reconstruction performance and superior fidelity on actual physical data.

3. Method

3.1. Problem statement

In dynamic LF reconstruction, we consider a sequence of B consecutive LF frames, denoted as $\{\mathbf{X}_k\}_{k=1}^{B}$, where $\{\mathbf{X}_k\}_{k=1}^{B} \in \mathbb{R}^{n_x \times n_y \times n_u \times n_v}$. Each frame is modulated by a corresponding coding mask $\{\mathbf{M}_k\}_{k=1}^{B} \in \{0,1\}^{n_x \times n_y \times n_u \times n_v}$ during a single exposure. These coded frames are temporally integrated into a single snapshot measurement $\mathbf{Y}$, expressed as:

$$\mathbf{Y} = \sum_{k=1}^{B} \mathbf{X}_k \odot \mathbf{M}_k + \mathbf{G}, \tag{1}$$

where $\odot$ denotes the Hadamard (element-wise) product, and $\mathbf{G}$ represents additive measurement noise after temporal integration. As illustrated in Fig. 1(a), this forward model follows the SCI paradigm, preserving the spatial-angular sampling structure while multiplexing temporal information into a single snapshot.

Optics EXPRESS

By vectorizing each frame into $n = n_x n_y n_u n_v$, the compressive acquisition process is modeled as:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{g}, \tag{2}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the compressive measurement, $\Phi \in \mathbb{R}^{n \times nB}$ denotes the sensing matrix, $\mathbf{x} \in \mathbb{R}^{nB}$ is the dynamic LF signal, and $\mathbf{g} \in \mathbb{R}^n$ is the noise vector. The reconstruction of $\mathbf{x}$ is formulated as an optimization problem:

$$\widehat{\mathbf{x}} = \arg\min_{\mathbf{x}} \frac{1}{2}\|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + r(\mathbf{x}), \tag{3}$$

where the first term ensures data fidelity and $r(\mathbf{x})$ is a regularization term encoding structural priors.

3.2. Overview of the architecture

Deep Equilibrium (DEQ) models represent a paradigm shift from traditional deep unfolding networks. Instead of stacking a finite number of explicit layers, DEQ formulates the forward pass as finding the fixed point $\mathbf{x}^\star$ of a shared non-linear transformation f_θ:

$$\mathbf{x}^\star = f_\theta(\mathbf{x}^\star; \mathbf{y}), \tag{4}$$

where θ denotes the learnable parameters. The solution is obtained via root-finding algorithms (e.g., Anderson acceleration) and trained using implicit differentiation, allowing for infinite effective depth with constant memory cost ($O(1)$).

While our previous work [19] successfully applied a generic densely connected DEQ to dynamic LF reconstruction, it treated the 5D LF tensor as a uniform block, failing to exploit its intrinsic geometric structure. To address this and handle the complex degradations of real-world hardware, we evolve the framework into a Domain-Adaptive Multi-stage Deep Equilibrium (DAM-DEQ) architecture. As illustrated in Fig. 1(b), the proposed DAM-DEQ introduces three critical structural advancements to bridge the gap between mathematical elegance and physical complexity.

First, we transition from a single-stage solving paradigm to a sequential refinement strategy. Instead of solving the high-dimensional inverse problem in a single black-box step, we decompose it into three manageable stages: *Spatial*, *Angular*, and *Multi-domain Fusion*. This sequential approach progressively recovers details, using the output of one domain to initialize and guide the equilibrium of the next. Second, we move from generic regularization to domain-specific priors. We replace the unified regularizer with a set of domain-adaptive implicit priors $\{R_{\theta_s}, R_{\theta_a}, R_{\theta_{\text{lf}}}\}$, specifically designed to enforce spatial texture fidelity, angular flux consistency, and epipolar geometric coherence, respectively. Finally, to ensure stability and efficiency within the equilibrium loop, we replace general-purpose convolutional layers with specialized blocks (LipCNN and ISC). These blocks are engineered to maintain Lipschitz continuity while efficiently aggregating multi-scale features across the LF structure.

3.3. DAM-DEQ architecture

For each domain $i \in \{s, a, lf\}$, DAM-DEQ adopts a preconditioned, gradient-based equilibrium formulation. The equilibrium mapping is defined as:

$$\mathbf{x}_i^\star = f_{\theta_i}(\mathbf{x}; \mathbf{y}, \Phi) = \mathbf{x} + \eta\Phi^\top\mathbf{P}(\mathbf{y} - \Phi\mathbf{x}) - \eta R_{\theta_i}(\mathbf{x}), \tag{5}$$

where η is the step size and $\mathbf{P} = (\Phi\Phi^\top)^{-1}$ approximates the inverse Hessian. In practice, $\mathbf{P}$ is approximated by a diagonal or block-diagonal preconditioner to ensure computational efficiency. This formulation links unrolled gradient descent with implicit DEQ inference.

Optics EXPRESS

3.3.1. Core building blocks

Integrated Scale-aware Context (ISC) Module. To efficiently capture multi-scale spatial structures while balancing computational overhead and representational capacity, we design the Integrated Scale-aware Context (ISC) module, as illustrated in Fig. 2. Operating in a split-transform-merge fashion, the module first processes input features through parallel convolution branches with varying kernel sizes to simultaneously capture spatial context across different receptive fields. Subsequently, the aggregated features are refined by a spatial-channel gating mechanism, which employs spatial attention to highlight salient regions and channel-wise modulation to suppress redundant information. A final residual connection fuses these modulated features with the original input, preserving high-frequency details and facilitating gradient flow. By leveraging these lightweight operations, the ISC module effectively enhances feature representation for complex LF data without incurring excessive memory costs.

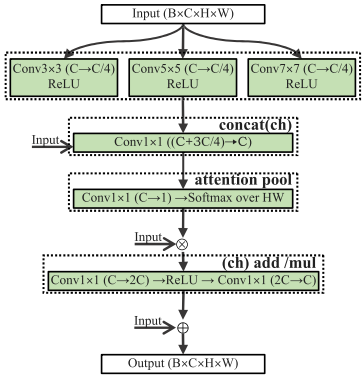

Fig. 2. Illustration of the Integrated Scale-aware Context (ISC) module, which captures multi-scale context through parallel convolutions, spatial attention, and feature modulation for efficient light field data processing.

Spectral-Normalized Convolution (LipCNN). To ensure the numerical stability and convergence of the DEQ fixed-point iteration, we introduce LipCNN, a convolutional module constrained by spectral normalization to enforce near 1-Lipschitz continuity. Regardless of the specific application, the architecture follows a unified expansion-refinement-projection design: an input convolution maps the data from its native dimension to a high-dimensional feature space ($D_{in} \to F$), followed by n sequential convolutions ($F \to F$) with ReLU activations for deep feature extraction, and a final convolution projects the refined features back to the original space ($F \to D_{in}$). We instantiate two variants based on the input dimensionality D_{in}. The domain-specific variants ($\text{LipCNN}_{s,a,e}$) set $D_{in} = 1$ to process individual spatial, angular, or epipolar slices via a bottleneck structure. Conversely, the multi-frame variant (LipCNN_{ms}) sets $D_{in} = C$ (where C is the number of frames) to operate directly on the video tensor $\mathbf{x} \in \mathbb{R}^{B\times C\times H\times W}$, thereby preserving temporal coherence and motion information without reshaping.

3.3.2. Sequential equilibrium stages

The framework sequentially solves three sub-problems, where the output of one stage initializes the next.

Stage 1: Spatial-Domain DEQ (R_{θ_s}). As illustrated in Fig. 3(a), the spatial subnetwork prioritizes the recovery of high-fidelity spatial structures. It first reshapes the input into a spatial-dominant 4D tensor to isolate spatial correlations. The core feature extraction is performed by the ISC module, which aggregates multi-scale spatial context, reinforced by LayerScale and DropPath to ensure training stability and dynamic range adjustment. The extracted features are fused with

Optics EXPRESS

the original input via a residual connection and subsequently refined by the spectral-normalized LipCNN$_s$. Within the DEQ framework, this module iteratively solves for the spatial equilibrium state $\mathbf{x}_s^\star$:

$$\mathbf{x}_s^\star = \mathrm{DEQ}_s(\mathbf{y}, \Phi; R_{\theta_s}, \mathbf{x}_0). \tag{6}$$

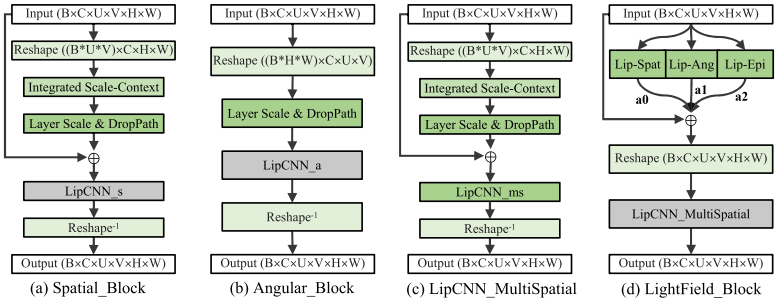

Fig. 3. Spatial, angular, and LF multi-domain equilibrium layers in the DAM-DEQ framework. (a) Spatial_Block (R_{θ_s}) using ISC, LayerScale, DropPath, and LipCNN$_s$ to obtain stable spatial reconstruction $\mathbf{x}_s^\star$. (b) Angular_Block (R_{θ_a}) with LayerScale, DropPath, and LipCNN$_a$ to recover angular low-frequency information, producing the final angular output $\mathbf{x}_a^\star$. (c) LipCNN_MultiSpatial leverages ISC, LayerScale, DropPath, and LipCNN$_{ms}$ to achieve learnable spatial-temporal coherence modeling. (d) LightField_Block ($R_{\theta_{\mathrm{lf}}}$) Three domain-specific submodules, Lip-Spat, Lip-Ang, and Lip-Epi, are hierarchically fused with learnable coefficients, followed by refinement with LipCNN_MultiSpatial to produce the final LF multi-domain reconstruction $\mathbf{x}_{\mathrm{lf}}^\star$.

Stage 2: Angular-Domain DEQ (R_{θ_a}). Building upon the spatial estimation, this stage takes $\mathbf{x}_s^\star$ as its initialization to enforce angular coherence (Fig. 3(b)). The subnetwork reshapes the tensor to prioritize angular dimensions and employs a streamlined architecture utilizing LayerScale and DropPath for regularization. Specifically, it leverages LipCNN$_a$ to capture low-frequency angular components, thereby maintaining flux consistency across viewpoints. The angular equilibrium state $\mathbf{x}_a^\star$ is iteratively solved via:

$$\mathbf{x}_a^\star = \mathrm{DEQ}_a(\mathbf{y}, \Phi; R_{\theta_a}, \mathbf{x}_s^\star). \tag{7}$$

Stage 3: Multi-Domain Fusion DEQ ($R_{\theta_{\mathrm{lf}}}$). Initialized by the angular estimate $\mathbf{x}_a^\star$, the final stage functions as a comprehensive prior integrator and domain adapter (Fig. 3(d)). This subnetwork operates through three parallel branches tailored to specific LF properties. The Lip-Spat branch mirrors the sophisticated architecture of the spatial layer by integrating the ISC module, LayerScale, and DropPath to resolve fine-grained spatial details. In parallel, the architecture employs two streamlined branches, Lip-Ang and Lip-Epi, utilizing standalone LipCNN modules to specifically capture low-frequency angular flux consistency and long-range epipolar geometric dependencies, respectively. To synthesize these diverse priors, the branch outputs are aggregated via learnable adaptive weights, allowing the network to dynamically balance feature contributions. This fused representation is subsequently refined by the LipCNN_MultiSpatial module (Fig. 3(c)), which combines the ISC module with the multi-frame LipCNN$_{ms}$ to enforce spatiotemporal coherence. The final 5D LF reconstruction is obtained as the equilibrium state:

$$\mathbf{x}^\star = \mathbf{x}_{\mathrm{lf}}^\star = \mathrm{DEQ}_{\mathrm{lf}}(\mathbf{y}, \Phi; R_{\theta_{\mathrm{lf}}}, \mathbf{x}_a^\star). \tag{8}$$

Convergence Analysis. To promote the existence and uniqueness of the equilibrium state, our framework relies on the Banach Fixed-Point Theorem. By decomposing the intractable high-dimensional inverse problem into three sequential, lower-dimensional subproblems, we

Optics EXPRESS

effectively reduce the optimization complexity. Within each stage, spectral normalization is applied to the weight matrices f_{θ_t} to enforce a Lipschitz constraint (weak contractivity). In practice, this theoretical stability is further augmented by residual scaling and Anderson acceleration, which ensure rapid and robust convergence to the fixed point.

3.4. Prototype construction

To validate the proposed hardware-algorithm co-design strategy, we constructed a physical prototype of the snapshot compressive LF imaging system. The detailed experimental setup is illustrated in Fig. 4. The flexibility of our system stems from the decoupling of sampling from sensor readout. The temporal upscaling factor B is determined solely by the number of modulation patterns loaded onto the DMD within a single exposure. Since the DMD is fully programmable, the compression ratio can be adjusted without modifying the optical hardware. The proposed DAM-DEQ framework is designed to be agnostic to the fixed temporal length, enabling adaptable frame-rate upscaling across different compression settings.

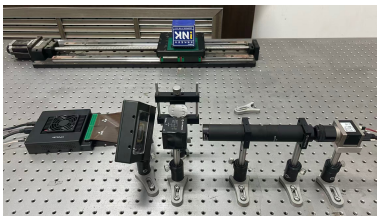

Fig. 4. Experimental setup of the customized LF-SCI prototype. Top: Motion control platform comprising a screw-driven linear translation stage. Bottom: Detailed optical path comprising an imaging lens (f = 100 mm), a 50:50 beam splitter, a high-speed DMD for temporal modulation, and a 4f relay system. The modulated scene is projected onto a custom LF camera, enabling the multiplexed acquisition of 4 temporal frames within a single snapshot.

Optical Encoding and Detection. The core encoding unit is depicted in the bottom panel of Fig. 4. The incident LF is collected by a plano-convex imaging lens (f = 100 mm, Hanyang Optics) and directed through a 50:50 cube beam splitter (Daheng Optics GCC-M403102) onto a high-speed Digital Micromirror Device (DMD, UPOLabs HDSLM108D95-Smart). The DMD operates at a refresh rate exceeding 20 kHz with a resolution of 1920 × 1080 (10.8 μm pitch). To ensure precise spatial modulation, we construct a custom 4f relay system (f_1 = 125 mm, f_2 = 60 mm), which optically conjugates the DMD plane to the sensor plane of the backend LF camera. The detection module consists of an industrial CMOS sensor (Basler acA4096-30uc, 4096 × 2168, 3.45 μm pitch) coupled with a square microlens array. This configuration provides a 9 × 9 angular sampling rate (with a central 7 × 7 subset used for reconstruction) and a spatial resolution of 456 × 241 pixels per view.

Data Acquisition Configuration. For dynamic scene experiments, the object is mounted on a linear translation stage, as shown in the top panel of Fig. 4, to generate controlled motion, while illumination is provided by an external laser source. During acquisition, the DMD is synchronized with the camera to multiplex N_t temporal frames into a single snapshot, where $N_t \in 2, 4, 6, 8, 16$. This enables flexible temporal upscaling from a physical frame rate of 2 fps to an effective rate of $2N_t$ fps in reconstruction.

To ensure precise spatial modulation in our prototype, we implement a three-step calibration protocol. (1) *Rotational alignment*: we calculate the rotational transformation between the DMD and the sensor planes to compensate for the DMD's inclined installation. (2) *Geometric mapping*: we determine the optical parameters of the 4f relay system to establish a coarse

Optics EXPRESS

geometric correspondence between the DMD micromirrors and the sensor's macro-pixels. (3) *Sub-aperture calibration*: by leveraging our "Dual-Mode Data Acquisition" strategy, we capture static calibration patterns to derive the precise point-to-point mapping from the mask to the sensor plane for each angular view. This enables the DAM-DEQ network to learn residual non-linear warping and adapt to hardware-specific aberrations and pixel-to-microlens misalignments.

4. Experimental results

4.1. Datasets and implementation details

Standard Datasets and Simulation Compression. To simulate temporal compression and minimize the workload of physical acquisition, we utilize preprocessed LF video datasets following the procedure established in our prior work (which considers $B = 4$). In this work, each snapshot measurement encodes B temporal frames, where $B \in 2, 4, 6, 8, 16$. Consistent with the video SCI paradigm, we simulate the encoding process by applying identical binary coding patterns to each sub-aperture view across the temporal sequence. This approach efficiently emulates optical modulation while preserving temporal and angular consistency.

For model training, we use the Raytrix LF dataset [38]. To verify performance, we adopt the Sintel synthetic LF dataset [39], the Illum (Lytro) real-world LF dataset [40], and custom array-camera prototype data, all of which are uniformly preprocessed to a spatial resolution of 128×128. These test scenes are meticulously curated to encompass a broad range of geometric complexities and motion dynamics. Specifically, we select five Sintel sequences that cover a disparity range of 0.191 to 1.005 pixels to assess depth-sensitive reconstruction. For the Illum dataset, we prioritize four sequences with significant motion captured by the hybrid Lytro-DSLR system. To further ensure data diversity, we explicitly supplement the evaluation with the *tree_branch* outdoor scene, which presents significant challenges due to its complex edges and depth discontinuities. To further avoid selection bias, we additionally include all remaining Sintel LF scenes (excluding those used for validation) and all available test scenes from the Illum dataset. This comprehensive selection ensures a rigorous evaluation of the system's ability to handle the spatiotemporal trade-offs inherent in real-world 3D sensing.

Prototype Data for Sim-to-Real Adaptation. To effectively bridge the domain gap between ideal simulations and the physical system, we implement a dual-mode data acquisition strategy for model fine-tuning. This strategy isolates hardware imperfections from temporal dynamics:

(1) **Instantaneous Static Acquisition:** We capture static scenes while dynamically switching masks to obtain instantaneous measurements. This mode isolates hardware-specific optical characteristics (e.g., aberrations and misalignment) without the interference of motion blur, providing a clean supervision signal for spatial hardware adaptation.

(2) **Pseudo-Cumulative Dynamic Acquisition:** To model realistic temporal integration, we capture sequential frames of moving objects (acquiring both real coded measurements and uncoded ground truth). We then digitally integrate the sequential ground truth to generate "simulated" cumulative measurements. This allows the model to learn the actual time-accumulation process and the temporal statistics of real-world motion.

For data processing, all captures are cropped to the central 5×5 angular views. The fine-tuning dataset consists exclusively of linear motion scenes, whereas the evaluation dataset includes both linear and rotational motions to test generalization. Additionally, we acquire raw single-exposure snapshot measurements directly from the prototype for qualitative analysis.

Implementation Details. All models are implemented in PyTorch and trained on a single NVIDIA RTX 3090 GPU. We utilize the Adam optimizer with an initial learning rate of 1×10^{-3} and a cosine annealing schedule. To ensure robust generalization, we adopt a decoupled optimization paradigm that separates intrinsic prior learning from hardware adaptation. In the

Optics EXPRESS

initial phase, the spatial and angular subnetworks are pre-trained on synthetic data (20 epochs) to establish a hardware-agnostic feature manifold representing ideal LF structures. Subsequently, the multi-domain fusion stage functions as a domain-adaptive interface. For synthetic evaluation, it is optimized to refine feature fidelity; crucially, for the physical prototype, this stage is fine-tuned directly on real measurements, transforming it into a correction operator that absorbs hardware-specific aberrations. Regarding the DEQ solver, we employ Anderson acceleration to expedite convergence during forward iteration and utilize implicit differentiation for backward gradient computation, ensuring constant memory efficiency.

Baselines and Metrics. We compare DAM-DEQ against three representative methods: (1) Mizuno et al. [26] (hardware-based dynamic LF acquisition), (2) Zhao et al. [36] (DEQ for video SCI), and (3) our previous DLFDEQ [19] (Dense DEQ for dynamic LF). To ensure a fair comparison, all models are evaluated on identical data splits and mask configurations with a fixed compression ratio of $B = 4$. For methods not originally tailored to our specific setup, we retrained them using the authors' official configurations. We report PSNR and SSIM to quantify structural fidelity, alongside MSE, MAE, and MaxErr to evaluate pixel-wise reconstruction accuracy and consistency.

4.2. Performance evaluation and analysis

We evaluate on representative LF scenes from synthetic [39] and real [40] datasets, and extend to autonomously collected dynamically compressed LF data. All methods target a 4× effective frame-rate increase; DAM-DEQ consistently outperforms the baselines.

Quantitative Evaluation on Synthetic Scenes. Table 1 compares the reconstruction quality on the Sintel synthetic LF dataset. DAM-DEQ achieves the state-of-the-art performance with an Overall PSNR of 29.59 dB and SSIM of 0.907, significantly outperforming Mizuno et al. [26] by 7.27 dB in PSNR. Compared to our previous DLFDEQ [19], the proposed domain-adaptive multi-stage mechanism further pushes the average PSNR from 29.39 dB to 29.63 dB. Notably, in challenging sequences like *chickenrun_1* and *thebigfight_2*, DAM-DEQ exhibits superior robustness, surpassing Zhao et al. [36] by 0.29 dB and 0.40 dB, respectively. These results across both *Avg.* (selected scenes) and *Overall* (full dataset) metrics validate the superior generalization of our framework in complex synthetic environments.

Quantitative Evaluation on Real-World Scenes. Table 2 summarizes the performance on the Lytro real-world dataset. DAM-DEQ consistently outperforms all baseline methods, achieving a leading Overall PSNR of 34.13 dB and SSIM of 0.940. Compared to Mizuno et al. [26], our framework provides a substantial margin of 8.29 dB in overall PSNR. In particular, in the high-texture *toy tiger* scene, DAM-DEQ reaches an impressive 41.02 dB PSNR, significantly exceeding Zhao et al. [36] and our previous DLFDEQ [19]. These results across diverse real-world sequences, including the complex *tree branch*, demonstrate the superior robustness and texture-preserving capability of our domain-adaptive multi-stage mechanism in practical LF reconstruction.

Quantitative Evaluation on Custom Array-Camera Scenes. Table 3 presents results on our custom LF dataset captured via a camera array. DAM-DEQ achieves the highest average PSNR of 29.63 dB and SSIM of 0.930. It shows a massive 10.14 dB PSNR improvement over Mizuno et al. [26], and consistently outperforms both Zhao et al. [36] and our baseline DLFDEQ [19] by margins of 0.49 dB and 1.23 dB in average PSNR, respectively. Particularly in the challenging *lego* sequence, which features low textures and complex illumination, DAM-DEQ achieves 28.30 dB PSNR, surpassing all baselines and demonstrating strong robustness in real-world, wide-baseline array captures.

Cross-Dataset Analysis and Discussion. Comparing results across datasets highlights a key advantage of DAM-DEQ. While the monocular method by Zhao et al. [36] performs well on the synthetic Sintel dataset (only 0.02 dB behind ours in overall PSNR), its performance degrades

Optics EXPRESS

Table 1. Quantitative evaluation on the Sintel synthetic dataset. Avg. denotes the average over the selected five scenes, while Overall denotes the average over all scenes in this dataset.

Method	Metric	*ambushfight_5*	*chickenrun_1*	*foggyrock_1*	*questbegin_1*	*thebigfight_2*	Avg.	Overall
Mizuno et al. [26]	PSNR ↑	24.03	20.30	24.33	27.70	22.80	23.83	22.32
	SSIM ↑	0.719	0.565	0.731	0.671	0.594	0.702	0.565
Zhao et al. [36]	PSNR ↑	29.73	25.97	26.21	37.62	28.17	29.54	29.57
	SSIM ↑	0.933	0.864	0.875	**0.971**	0.868	0.902	0.907
DLFDEQ [19]	PSNR ↑	29.35	25.75	26.12	**37.63**	28.10	29.39	29.27
	SSIM ↑	0.926	0.857	0.874	0.970	0.865	0.898	0.901
DAM-DEQ	PSNR ↑	**30.00**	**26.26**	**26.50**	36.83	**28.57**	**29.63**	**29.59**
	SSIM ↑	**0.935**	**0.871**	**0.885**	0.964	**0.878**	**0.907**	**0.907**

Table 2. Quantitative evaluation on the Lytro real-world dataset. Avg. denotes the average over the selected five scenes, while Overall denotes the average over all scenes in this dataset.

Method	Metric	*toy engine*	*toy train*	*toy tiger*	*toy cat*	*tree branch*	Avg.	Overall
Mizuno et al. [26]	PSNR ↑	26.48	27.73	30.11	32.95	28.36	29.13	25.84
	SSIM ↑	0.829	0.740	0.758	0.905	0.775	0.801	0.669
Zhao et al. [36]	PSNR ↑	31.47	33.45	40.02	36.32	33.43	34.94	33.53
	SSIM ↑	0.952	**0.955**	0.982	0.981	0.948	0.963	0.935
DLFDEQ [19]	PSNR ↑	30.72	33.39	39.94	35.81	33.19	34.61	33.48
	SSIM ↑	0.947	0.948	0.981	0.980	0.945	0.960	0.934
DAM-DEQ	PSNR ↑	**32.31**	**33.87**	**41.02**	**36.39**	**33.95**	**35.51**	**34.13**
	SSIM ↑	**0.958**	0.951	**0.984**	**0.983**	**0.954**	**0.966**	**0.940**

Table 3. Quantitative evaluation on our custom array-camera dataset.

Method	Metric	*lego*	*bicycle*	*trash*	*hydrant*	*trolley*	Avg.
Mizuno et al. [26]	PSNR ↑	12.74	16.47	26.21	22.72	19.31	19.49
	SSIM ↑	0.288	0.409	0.777	0.690	0.487	0.530
Zhao et al. [36]	PSNR ↑	27.71	27.55	31.45	30.68	28.29	29.14
	SSIM ↑	0.938	0.898	0.950	0.927	0.910	0.925
DLFDEQ [19]	PSNR ↑	26.31	27.24	30.58	30.28	27.59	28.40
	SSIM ↑	0.903	0.891	0.945	0.929	0.901	0.914
DAM-DEQ	PSNR ↑	**28.30**	**27.97**	**32.09**	**31.15**	**28.62**	**29.63**
	SSIM ↑	**0.940**	**0.906**	**0.954**	**0.935**	**0.915**	**0.930**

on real-world data. Specifically, DAM-DEQ outperforms Zhao et al. by larger margins of 0.60 dB and 0.49 dB on the Lytro and Custom Array datasets, respectively. This gap arises from how multi-view information is handled. Repeatedly applying a monocular network works adequately on simulated data because the underlying geometries are perfectly idealized. However, simulated environments cannot fully replicate real-world physical complexities, such as intricate object occlusions and viewpoint-dependent illumination changes. Treating views independently struggles to resolve these ambiguities in practical scenes. By being explicitly designed for LF data, DAM-DEQ effectively models authentic cross-view dependencies, leading to superior robustness and reconstruction quality in the real world.

Qualitative Analysis. In our qualitative evaluation, we present a comprehensive visual analysis by combining full reconstructed images, zoomed-in crops for fine local details, and pixel-wise

Optics EXPRESS

error maps (MSE/MAE/MaxErr) for global consistency. This combined approach explicitly reveals both the recovery of intricate textures and the overall distribution of reconstruction errors across the entire field of view. Visual comparisons in Fig. 5 substantiate the quantitative gains. Note that the metrics shown in Fig. 5 refer to the specific frames displayed, while tables report the averages of the datasets.

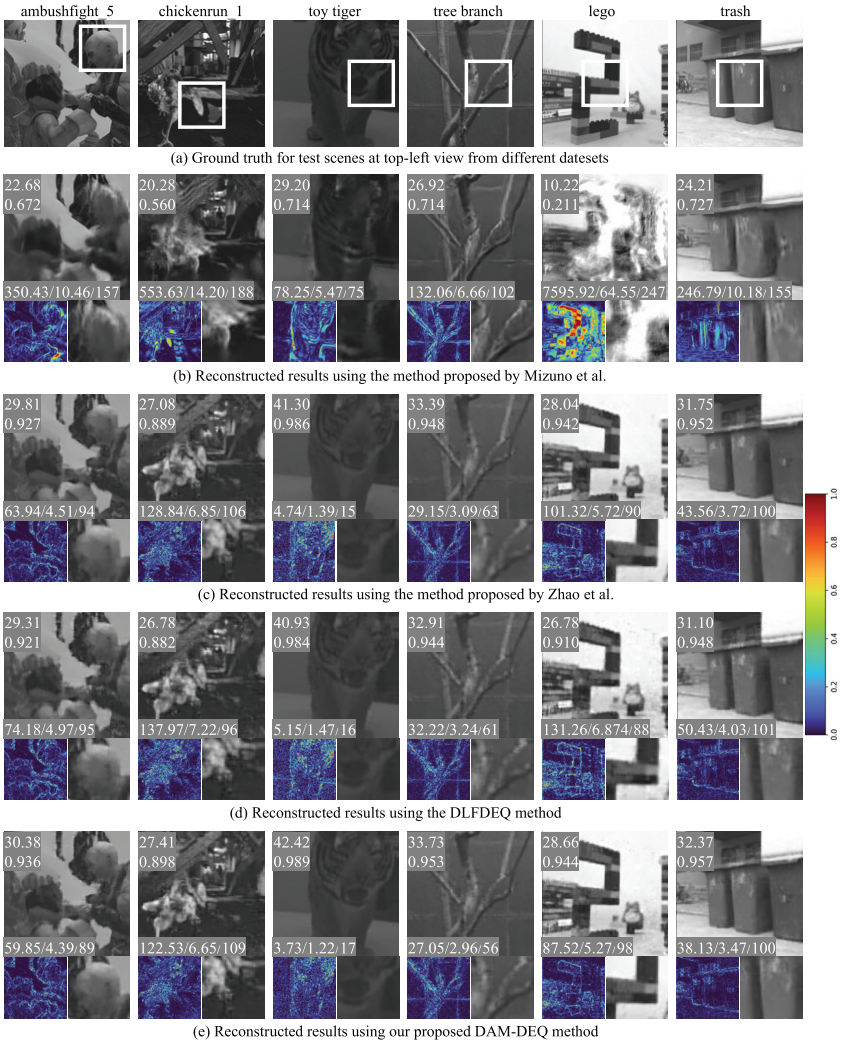

Fig. 5. Qualitative comparison across six test scenes. (a) Ground truth. (b) Mizuno et al. [26]. (c) Zhao et al. [36]. (d) DLFDEQ [19]. (e) Ours (DAM-DEQ). Numbers show PSNR/SSIM (top) and MSE/MAE/MaxErr (bottom) for the specific displayed frames. Our method achieves superior reconstruction quality across synthetic, real-world, and array-camera scenes.

In synthetic scenes, DAM-DEQ demonstrates superior reconstruction quality: for *ambushfight_5*, our method achieves 30.38 dB PSNR and 0.936 SSIM with significantly reduced MSE (59.85) compared to competing methods, effectively suppressing aliasing artifacts in complex motion scenes while maintaining sharp boundaries. Similarly, for *chickenrun_1*, DAM-DEQ attains 27.41 dB PSNR and 0.898 SSIM, with the lowest error metrics (MSE: 122.53, MAE: 6.65), preserving fine textural details in challenging scenes.

Optics EXPRESS

In real-world scenes, DAM-DEQ consistently outperforms baseline methods: the *toy tiger* scene shows exceptional performance with 42.42 dB PSNR and 0.969 SSIM, achieving remarkably low reconstruction errors (MSE: 3.73, MAE: 1.22), indicating robust handling of high-texture regions and natural lighting variations. For the *tree branch* scene, our method maintains superior fidelity with 33.73 dB PSNR and 0.953 SSIM while effectively managing occlusions and depth discontinuities, as evidenced by the controlled error metrics (MSE: 27.05, MaxErr: 56).

In array camera scenes, DAM-DEQ demonstrates enhanced robustness under challenging conditions: the *lego* scene achieves 28.66 dB PSNR and 0.944 SSIM with effective noise suppression, while the *trash* scene shows 32.37 dB PSNR and 0.957 SSIM, providing enhanced edge sharpness despite low-texture characteristics and variable illumination. These comprehensive quantitative improvements across all error metrics (MSE, MAE, MaxErr) confirm the framework's superior generalization capability and robustness across diverse scene domains. Collectively, these results validate that our DAM-DEQ design effectively harmonizes learnable priors with iterative data fidelity, yielding robust reconstruction even under complex illumination and geometric variations.

Performance on Prototype Measurements (Sim-to-Real Validation). To validate the practical deployability of our framework, we evaluated DAM-DEQ on raw compressed measurements acquired directly from the custom DMD-microlens prototype. Leveraging the decoupled optimization paradigm detailed above, we fine-tuned the multi-domain fusion subnetwork using the proposed dual-mode acquisition strategy. This process specifically aligned the model with hardware-specific optical characteristics (via instantaneous static measurements) and realistic temporal dynamics (via pseudo-cumulative dynamic data).

Visual results presented in Fig. 6 demonstrate that the fine-tuned model exhibits remarkable robustness to physical non-idealities, effectively compensating for optical aberrations and mask-to-sensor misalignment. By transforming the fusion stage into a learnable correction operator, DAM-DEQ successfully bridges the sim-to-real gap, recovering high-frequency spatial details and preserving temporal continuity despite the complex constraints of the physical hardware.

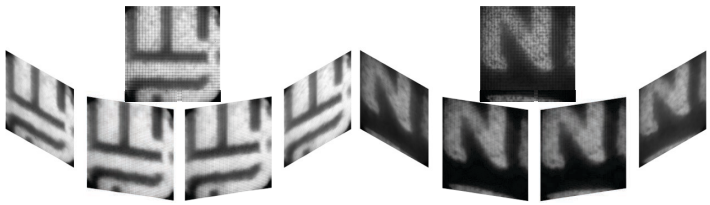

Fig. 6. Real-prototype demonstration. Left: pseudo-cumulative compressed measurements and reconstructions. Right: real compressed measurements and reconstructions.

4.3. Ablation study

Ablation on Domain-Adaptive Fusion. To validate the efficacy of the domain-adaptive multi-branch architecture, we conducted a comprehensive ablation study using the fine-tuning dataset derived from our prototype. Consistent with the dual-mode acquisition strategy detailed in Section 4.1, the evaluation is stratified into two subsets: Static (Instantaneous) measurements, which isolate spatial hardware characteristics in the absence of motion blur, and Dynamic (Pseudo-Cumulative) measurements, which incorporate simulated temporal integration to represent realistic physical conditions. Although the Static subset generally yields higher metrics due to its lower complexity, the Dynamic subset serves as a more representative benchmark for assessing robust sim-to-real transfer under hardware constraints.

We designated the LipCNN_ms module as the *Space-Only* backbone and progressively integrated complementary domain branches. As evidenced in Table 4, relying exclusively on

Optics EXPRESS

auxiliary domains (*Angle-Only* or *Epi-Only*) proves insufficient, confirming the necessity of a strong spatial foundation. While the simple addition of angular and epipolar cues (+*Angle*, +*Angle+Epi*) offers marginal benefits on the simpler Static data, the advantages of our approach become pronounced on the rigorous Dynamic subset. By employing learnable residual coefficients to dynamically weight branch contributions, our *Fusion* strategy (DAM-DEQ) effectively harmonizes conflicting priors. This adaptive mechanism achieves the highest performance on the Dynamic dataset (26.11 dB PSNR, 0.910 SSIM), demonstrating that adaptive multi-domain synergy is essential for mitigating the complex degradation found in real-world accumulated measurements.

Table 4. Ablation study of multi-domain branches and fusion strategies on prototype data.

Dataset	Metric	Angle-Only	Epi-Only	Space-Only	+Angle	+Angle+Epi	Fusion (Ours)
Static (Instantaneous)	PSNR ↑	19.53	22.47	28.44	**28.86**	**28.86**	28.66
	SSIM ↑	0.630	0.712	0.888	**0.897**	**0.897**	0.889
Dynamic (Pseudo-Cumul.)	PSNR ↑	20.06	22.06	26.04	25.34	25.78	**26.11**
	SSIM ↑	0.723	0.816	0.908	0.894	0.910	**0.910**

Ablation on Spatiotemporal Scalability. To rigorously test the limits of our model under restricted angular data and extreme temporal entanglement, we subsequently introduce a sparse 3×3 sampling strategy (extracting views from the 1st, 3rd, and 5th rows and columns). Compared to the dense 5×5 setup at $B = 4$, this sparse configuration understandably reduces performance (e.g., dropping to 29.17 dB on Sintel and 34.79 dB on Lytro) due to the significantly enlarged view-to-view disparities and diminished angular redundancy.

We further push this challenging 3×3 setup across an extended range of compression ratios $B \in \{2, 4, 6, 8, 16\}$, as summarized in Table 5. At a mild compression ratio ($B = 2$), DAM-DEQ effectively resolves the large spatial disparities, yielding a clear performance margin over baselines. However, for $B \geq 4$, the performance of all evaluated methods noticeably converges, with our model maintaining a marginal lead. We attribute this plateau to the compounding physical constraints of the inverse problem: the extreme temporal entanglement at high compression ratios, coupled with the large spatial disparities of our artificially sparse sampling, creates a severe information bottleneck that bounds theoretical performance. Nevertheless, despite these extreme physical limitations up to $B = 16$, DAM-DEQ consistently demonstrates resilient structural fidelity, proving its robustness under highly degraded spatiotemporal conditions.

These experiments highlight our framework's temporal flexibility: adaptable frame-rate upscaling is achieved simply by adjusting DMD patterns and retraining for a specific B. Crucially, DAM-DEQ maintains an $O(1)$ memory complexity relative to the solver's iterations. Unlike traditional deep unfolding networks whose memory scales linearly with depth, our approach allows increasing iterations for higher accuracy without additional memory overhead. This decoupling of memory cost from network depth ensures high-throughput, memory-efficient reconstruction of 5D dynamic LF data.

Optics EXPRESS

Table 5. Ablation study of reconstruction performance across varying compression ratios *B* and datasets.

Method	Dataset	Metric	$B = 2$	$B = 4$	$B = 6$	$B = 8$	$B = 16$
Zhao et al. [36]	Sintel (Synthetic)	PSNR	31.61	**29.53**	**25.33**	23.46	22.76
		SSIM	0.938	**0.901**	**0.816**	0.773	0.738
	Lytro (Real-world)	PSNR	37.03	34.80	31.37	29.61	**28.82**
		SSIM	0.980	0.962	0.927	0.907	**0.890**
DLFDEQ [19]	Sintel (Synthetic)	PSNR	31.63	28.87	25.28	23.44	22.73
		SSIM	0.939	0.890	0.815	0.773	0.737
	Lytro (Real-world)	PSNR	37.31	34.05	31.22	29.53	28.71
		SSIM	0.979	0.955	0.925	0.904	0.887
DAM-DEQ	Sintel (Synthetic)	PSNR	**31.98**	29.17	**25.33**	**23.47**	**22.79**
		SSIM	**0.944**	0.898	**0.816**	**0.776**	**0.740**
	Lytro (Real-world)	PSNR	**37.91**	**34.79**	31.34	**29.70**	28.78
		SSIM	**0.982**	**0.962**	0.927	**0.908**	0.899

5. Conclusion

In this work, we present a complete hardware–algorithm co-design for dynamic LF sensing, bridging theoretical simulation and physical realization through a snapshot compressive acquisition system. By integrating a high-speed DMD with a microlens-array camera, the proposed prototype enables efficient spatio-angular–temporal multiplexing within a single exposure. To address the resulting ill-posed inverse problem, we propose DAM-DEQ, a domain-adaptive deep equilibrium reconstruction framework. By explicitly modeling spatial, angular, and multi-domain fusion priors within an implicit equilibrium formulation, DAM-DEQ reconstructs high-fidelity 5D LF data from a single coded 4D measurement. Experimental results demonstrate a fourfold improvement in effective frame rate and consistent performance gains over state-of-the-art methods.

Several practical insights can be drawn from this study. First, replacing uniform processing with domain-specific modeling that enforces spatial texture, angular flux conservation, and epipolar geometric coherence significantly improves reconstruction quality. Second, the deep equilibrium formulation enables parameter-efficient reconstruction with constant memory usage, which is critical for high-dimensional 5D LF processing. Notably, DAM-DEQ exhibits strong sim-to-real generalization: despite being trained primarily on simulated data with ideal masks, the model remains robust to domain shifts caused by optical aberrations and pixel-to-microlens misalignment, validating its effectiveness under real-world imaging conditions.

We explicitly acknowledge the physical boundaries of the current system. While the DMD enables high compression ratios, the primary limitation is the photon budget: increasing B reduces the integration time per sub-frame, inherently lowering the signal-to-noise ratio (SNR). Furthermore, the cost-effective prototype configuration introduces spherical aberrations and mask-sensor misalignment. While our domain-adaptive algorithm mitigates these non-idealities, future work will aim to address these constraints by developing a specialized hybrid imaging system integrating advanced optical correctors. We aim to leverage multi-view learning and LLMs to improve robust 3D reconstruction. Additionally, we plan to extend our hardware-algorithm co-design to support emerging visualization technologies, such as AR and 3D displays.

Funding. National Key Research and Development Program of China (2024YFB2809002); National Natural Science Foundation of China (62571449, 62031023).

Disclosures. The authors declare no conflicts of interest.

Optics EXPRESS

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. J. R. Bergen and E. H. Adelson, "The plenoptic function and the elements of early vision," Computational models of visual processing **1**, 3 (1991).
2. C. Zhao and B. Jeon, "Compact representation of light field data for refocusing and focal stack reconstruction using depth adaptive multi-cnn," IEEE Trans. Comput. Imaging **10**, 170–180 (2024).
3. K. Cheng, L. Pan, Z. Lai, *et al.*, "Blind aberration correction for light field photography," Opt. Lett. **50**(1), 209–212 (2025).
4. Y. Li, X. Wang, G. Zhou, *et al.*, "Sheared epipolar focus spectrum for dense light field reconstruction," IEEE Trans. Pattern Anal. Mach. Intell. **46**(5), 3108–3122 (2024).
5. C. Jin, Y. Chen, T. Luo, *et al.*, "Dual attention-assisted hdr light field imaging using state space model and implicit neural representation," Optics & Laser Technology **192**, 113693 (2025).
6. X. Wang, Y. Lin, and S. Zhang, "Multi-stream progressive restoration for low-light light field enhancement and denoising," IEEE Trans. Comput. Imaging **9**, 70–82 (2023).
7. X. Huang, Q. Zhang, Y. Feng, *et al.*, "Hdr-nerf: High dynamic range neural radiance fields," in *35th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), pp. 18398–18408.
8. Z. Zhang, X. Yu, X. Gao, *et al.*, "High-fidelity light-field display with enhanced information utilization by modulating chrominance and luminance separately," Light:Sci. Appl. **14**(1), 78 (2025).
9. Z. Lu, Y. Cai, Y. Nie, *et al.*, "A practical guide to scanning light-field microscopy with digital adaptive optics," Nat. Protoc. **17**(9), 1953–1979 (2022).
10. R. Guo, Q. Yang, A. S. Chang, *et al.*, "Eventlfm: event camera integrated fourier light field microscopy for ultrafast 3d imaging," Light:Sci. Appl. **13**(1), 144 (2024).
11. Z. Lu, S. Zuo, M. Shi, *et al.*, "Long-term intravital subcellular imaging with confocal scanning light-field microscopy," Nat. Biotechnol. **43**(4), 569–580 (2025).
12. E. Miandji, H.-N. Nguyen, S. Hajisharif, *et al.*, "Compressive hdr light field imaging using a single multi-iso sensor," IEEE Trans. Comput. Imaging **7**, 1369–1384 (2021).
13. Y. Huang, M. M. Hossain, Y. Liu, *et al.*, "Data rectification and decoding of a microlens array-based multi-spectral light field imaging system," Optics and Lasers in Engineering **180**, 108327 (2024).
14. A. Higaki, K. Kitano, Y. Fujimura, *et al.*, "Refractive epitrochoids sampling for light field measurement using wedge prisms," Opt. Express **33**(2), 2604–2619 (2025).
15. F. Linda Liu, G. Kuo, N. Antipa, *et al.*, "Fourier diffuserscope: single-shot 3d fourier light field microscopy with a diffuser," Opt. Express **28**(20), 28969–28986 (2020).
16. X. Cao, Y. Liu, X. Ji, *et al.*, "Vision field capture for advanced 3dtv applications," in *26th Visual Communications and Image Processing* (VCIP) (IEEE, 2011), pp. 1–4.
17. X. Lin, J. Wu, G. Zheng, *et al.*, "Camera array based light field microscopy," Biomed. Opt. Express **6**(9), 3179–3189 (2015).
18. Y. Zhang, Y. Wang, M. Wang, *et al.*, "Multi-focus light-field microscopy for high-speed large-volume imaging," PhotoniX **3**(1), 30 (2022).
19. R. Wang, X. Wang, Z. Xiao, *et al.*, "Dynamic light field reconstruction via densely connected deep equilibrium model," Opt. Express **32**(26), 46829–46848 (2024).
20. B. Wilburn, N. Joshi, V. Vaish, *et al.*, "High performance imaging using large camera arrays," in *SIGGRAPH* (ACM, 2005), pp. 765–776.
21. K. He, X. Wang, Z. W. Wang, *et al.*, "Snapshot multifocal light field microscopy," Opt. Express **28**(8), 12108–12120 (2020).
22. F. Xing, D. Wang, H. Tan, *et al.*, "High-resolution light-field particle imaging velocimetry with color-and-depth encoded illumination," Optics and Lasers in Engineering **173**, 107921 (2024).
23. S. Tambe, A. Veeraraghavan, and A. Agrawal, "Towards motion aware light field video for dynamic scenes," in *14th International Conference on Computer Vision* (ICCV) (IEEE, 2013), pp. 1009–1016.
24. S. Hajisharif, E. Miandji, C. Guillemot, *et al.*, "Single sensor compressive light field video camera," Computer Graphics Forum **39**(2), 463–474 (2020).
25. K. Sakai, K. Takahashi, T. Fujii, *et al.*, "Acquiring dynamic light fields through coded aperture camera," in *16th European Conference on Computer Vision (ECCV)*, (2020), pp. 368–385.
26. R. Mizuno, K. Takahashi, M. Yoshida, *et al.*, "Acquiring a dynamic light field through a single-shot coded image," in *35th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), pp. 19830–19840.
27. S. Habuchi, K. Takahashi, C. Tsutake, *et al.*, "Time-efficient light-field acquisition using coded aperture and events," in *37th IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2024), pp. 24923–24933.
28. Z. Wang, Y. Peng, L. Fang, *et al.*, "Computational optical imaging: on the convergence of physical and digital layers," Optica **12**(1), 113–130 (2025).
29. H. Tang, T. Men, X. Liu, *et al.*, "Single-shot compressed optical field topography," Light:Sci. Appl. **11**(1), 244 (2022).

Optics EXPRESS

30. Y. Chen, Y. Wang, and H. Zhang, "Prior image guided snapshot compressive spectral imaging," IEEE Trans. Pattern Anal. Mach. Intell. **45**(9), 11096–11107 (2023).
31. Y.-C. Miao, X.-L. Zhao, J.-L. Wang, *et al.*, "Snapshot compressive imaging using domain-factorized deep video prior," IEEE Trans. Comput. Imaging **10**, 93–102 (2024).
32. Z. Meng, X. Yuan, and S. Jalali, "Deep unfolding for snapshot compressive imaging," International Journal of Computer Vision **131**(11), 2933–2958 (2023).
33. J. Zhang, H. Zeng, J. Cao, *et al.*, "Dual prior unfolding for snapshot compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2024), pp. 25742–25752.
34. Y. Chen, X. Gui, J. Zeng, *et al.*, "Combining low-rank and deep plug-and-play priors for snapshot compressive imaging," IEEE Trans. Neural Netw. Learning Syst. **35**(11), 16396–16408 (2024).
35. S. Bai, J. Z. Kolter, and V. Koltun, "Deep equilibrium models," Advances in Neural Information Processing Systems **32**, 690–701 (2019).
36. Y. Zhao, S. Zheng, and X. Yuan, "Deep equilibrium models for snapshot compressive imaging," in *Proceedings of the AAAI Conference on Artificial Intelligence*, (2023), pp. 3642–3650.
37. Z. Zou, J. Liu, B. Wohlberg, *et al.*, "Deep equilibrium learning of explicit regularization functionals for imaging inverse problems," IEEE Open J. Signal Process. **4**, 390–398 (2023).
38. L. Guillo, X. Jiang, G. Lafruit, *et al.*, "Light field video dataset captured by a r8 raytrix camera (with disparity maps)," clim (2018) [retrieved 9 April 2018], http://clim.inria.fr/Datasets/RaytrixR8Dataset-5x5/index.html.
39. T. Kinoshita and S. Ono, "Depth estimation from 4d light field videos," ieee-dataport (2021) [retrieved 26 March 2021], https://ieee-dataport.org/open-access/sintel-4d-light-field-video-dataset.
40. T.-C. Wang, J.-Y. Zhu, N. K. Kalantari, *et al.*, "Light field video capture using a learning-based hybrid imaging system," ACM Transactions on Graphics (TOG) **36**, 1–13 (2017).