Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Compact snapshot meta-optic HDR camera for handheld, edge-deployable imaging

Praneeth Chakravarthula

cpk@cs.unc.edu

University of North Carolina at Chapel Hill https://orcid.org/0000-0002-3092-7435

Zhihan Peng

University of North Carolina at Chapel Hill

Debabrata Mandal

University of North Carolina at Chapel Hill

Yujie Wang

University of North Carolina at Chapel Hill

Zihao Zou

University of North Carolina at Chapel Hill

Myunghoo Lee

University of Washington

Arka Majumdar

University of Washington

Johannes Fr¨och

University of Washington

Article

Keywords:

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11254042/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. Arka Majumdar is a cofounder of Tunoptix Inc. The remaining authors declare no competing interests.

Compact snapshot meta-optic HDR camera for handheld, edge-deployable imaging

Zhihan Peng[1], Debabrata Mandal[1], Yujie Wang[1], Zihao Zou[1], Myunghoo Lee[2], Arka Majumdar[2,3], Johannes E. Fröch[2], and Praneeth Chakravarthula[1]

[1]University of North Carolina at Chapel Hill, Dept. of Computer Science, Chapel Hill, NC, USA
[2]University of Washington, Dept. of Electrical and Computer Engineering, Washington, USA
[3]University of Washington, Dept. of Physics, Washington, USA

ABSTRACT

Real-world scenes can span more than 100 dB of luminance, while conventional image sensors typically capture less than 90 dB in a single exposure. This gap causes bright regions to saturate and dark regions to lose detail. Existing cameras extend the dynamic range by combining multiple sequentially captured exposures, but this increases total acquisition time and produces artifacts from scene or camera motion. In this work, we translate multi-exposure acquisitions from the temporal domain to the optical domain. A 2×3 metalens array with engineered transmittance across the full visible spectrum simultaneously captures six differently attenuated exposures in a single snapshot, which extends the dynamic range by 23 dB over a conventional single-aperture camera. However, each metalens in the array produces different, spatially varying blur and aberrations across the image, making the six measurements difficult to restore, align, and fuse into a high dynamic range (HDR) image. Correcting these degradations and performing HDR fusion typically requires GPU-intensive processing, which undermines the size and efficiency advantages of compact meta-optics. We overcome this computational bottleneck using content- and position-adaptive filters and a lightweight image restoration network. The complete optical-computational pipeline runs directly on an NVIDIA Jetson Orin Nano with 8 GB of memory, with network inference running at 2.7 fps at 256×256 resolution and 8.5 fps at 128×128 resolution, bringing HDR reconstruction within the compute and memory constraints of mobile-class hardware. Experiments on real-world high-contrast, low-light, and fast-moving scenes demonstrate robust reconstruction under conditions that challenge sequential multi-exposure capture. Together, these results move computational metalens imaging from GPU-intensive offline reconstruction toward compact, edge-deployable cameras.

Introduction

Real-world imaging rarely occurs under controlled illumination. Natural scenes can span luminance ratios exceeding 100,000:1 [1], placing deep shadows and bright highlights within the same field of view. A vehicle entering a tunnel on a sunny day, for example, must simultaneously image the dark tunnel interior and the brightly lit road outside. Conventional image sensors capture only a fraction of this range in a single exposure, saturating bright regions and losing information in dark regions to sensor noise. For photography, this produces an undesirable image. But for safety-critical systems such as autonomous vehicles, drones, surveillance systems, and other machine-vision platforms, it can remove precisely the scene information needed for reliable operation. High-dynamic-range (HDR) imaging is therefore not only a problem of image quality, but of capturing information reliably under uncontrolled real-world illumination.

One approach is to increase the dynamic range of the sensor itself through larger full-well capacities [2], stacked sensor architectures [3], and per-pixel exposure coding [4]. These advances have substantially improved sensor performance, but dynamic range remains coupled to the fundamental trade-offs among full-well capacity, read noise, pixel size, resolution, sensitivity, and operating speed. Specialized sensor architectures also require changes to the imaging hardware at the semiconductor level, limiting their flexibility and accessibility.

The most widely adopted alternative is bracketed exposure imaging (Fig. 1a), common in modern smartphone cameras, where a sequence of images is captured at different exposures, aligned, and their complementary information is fused into an HDR image [5]. This works well when the scene is sufficiently static. In dynamic scenes, however, the measurements no longer describe the same instant in time, with scene or camera motion between exposures producing misalignment and ghosting that must be corrected computationally [6]. Longer exposure brackets also increase acquisition time, with latency scaling at least linearly with the number of exposures [7, 8]. These limitations become particularly important in autonomous vehicles, drones, wearables, and other edge platforms, where scenes are often dynamic, especially when sensing must operate under strict latency, power, and compute constraints. The dominant HDR strategy therefore becomes increasingly difficult to deploy

in many of the settings where robust dynamic-range capture is most valuable.

An alternative is to move exposure multiplexing from time into optics. Beam-splitter systems [9], spatially varying filters [10], coded apertures [11, 12], and polarization cameras [13, 14] capture multiple exposures simultaneously, avoiding inter-exposure motion artifacts. However, they require bulky optics or complex alignment, limiting scalability and deployment on compact, weight-constrained platforms. Meta-optics offer a compact alternative by integrating multiple optical functions on a single planar substrate [15, 16, 17], and have recently been explored for HDR imaging [18, 19, 20, 21, 22]. However, they are restricted to laboratory setups and demand heavy computation for image reconstruction. We argue that compact optics alone do not make a compact camera. Broadband metalenses introduce spatially varying blur and aberrations, making the captured images difficult to restore. Correcting these degradations typically requires GPU-intensive reconstruction, which can offset the size and efficiency advantages of meta-optics [23, 24, 25]. Practical metalens HDR imaging, therefore, requires solving the optical and computational problems together, with computation efficient enough to run within the resource constraints of mobile and edge devices.

In this work, we jointly address both the optical and computational sides of this problem. We design a 2×3 metalens array whose six sub-apertures are engineered with distinct optical transmittances, simultaneously capturing a six-exposure bracket in a single snapshot across the visible spectrum (Fig. 1b). This spatial exposure multiplexing extends the captured dynamic range by 23 dB over a conventional single-aperture camera while avoiding sequential acquisition. We then address the heterogeneous blur and aberrations introduced by the metalens array with a lightweight reconstruction network suitable for deployment on standalone edge devices. Specifically, our network uses a Decoupled Dynamic Filter (DDF) [26] front-end that adapts spatial and channel filtering to image content and position, followed by deformable alignment, exposure-aware fusion, and a static reconstruction backbone. To improve generalization to real-world scenes, we pretrain on laboratory data and then fine-tune with a bootstrapped real-scene training pipeline. Coarse reconstructions from the pretrained model facilitate registration with co-captured smartphone HDR images, enabling pseudo-reference construction for real-scene fine-tuning. The trained network jointly restores, aligns, and fuses the six measurements into an HDR image.

The resulting optical-computational system integrates broadband snapshot acquisition, adaptive reconstruction, and edge deployment in a single meta-optical camera (Fig. 1c). The metalens array operates across 400–700 nm, while the complete reconstruction pipeline runs directly on an 8 GB NVIDIA Jetson Orin Nano, whose memory and power constraints approach those of mobile computing platforms (Fig. 1d). More importantly, our lightweight image recovery network infers high fidelity images at rates reaching 2.7 fps at 256×256 resolution and 8.5 fps at 128×128 resolution. We validate the system on real-world high-contrast, low-light, and fast-moving scenes where sequential multi-exposure capture typically struggles. To our knowledge, this is the first metalens HDR camera to combine broadband snapshot acquisition with reconstruction directly on embedded edge hardware. Our results show that the computational burden introduced by compact meta-optics can be addressed within the same resource-constrained system. More broadly, by moving computational metalens imaging from GPU-intensive offline reconstruction into the embedded computing regime, this work provides a path toward compact HDR cameras for smartphones, wearables, drones, and other edge devices.

Results

A compact metalens array enables snapshot multi-exposure imaging

Our compact imaging module converts temporal exposure bracketing into a single spatially multiplexed measurement, with computation running natively on an NVIDIA Jetson Orin Nano, a mobile-class power- and latency-constrained edge device. A 2×3 metalens array is fabricated on a 0.5 mm-thick substrate and mounted directly above a Sony IMX183 CMOS sensor, achieving an effective field of view of 33°. The field-dependent blur that remains across the un-optimized imaging field is addressed computationally by the reconstruction network. Each 1-mm-diameter metalens forms an image of the same scene, while integrated attenuation layers assign a different effective transmittance to each aperture. A baffle isolates the six image fields on the sensor. The complete optical track length is 2.5 mm, achieving a substantially smaller form factor than refractive optics. Unlike conventional HDR cameras, which acquire multiple exposure levels sequentially over time, the array records all six exposure levels in a single shot (Fig. 1).

The metalenses are designed for broadband imaging across the visible spectrum using an established phase design framework [23]; full details are given in the Methods and Supplementary Information. Optical and scanning-electron-microscope images of the fabricated meta-optic (Fig. 1c) show the arrangement of the sub-wavelength nanostructures that locally modulate the phase of incident light and confirm the accuracy of the fabricated design. We characterize the array using the wavelength-resolved point spread function (PSF), modulation-transfer-function (MTF), and per-aperture throughput measurements across various wavelengths and exposure times (Fig. 2). Specifically, we measure the PSF with a broadband laser source and a tunable filter, from which we derive the MTF across wavelengths. Each aperture maintains a compact focal spot across the measured band (Fig. 2a), with MTF contrast cutoff ($\mathrm{MTF} > 0.01$) above 100 lp/mm for nearly all measured

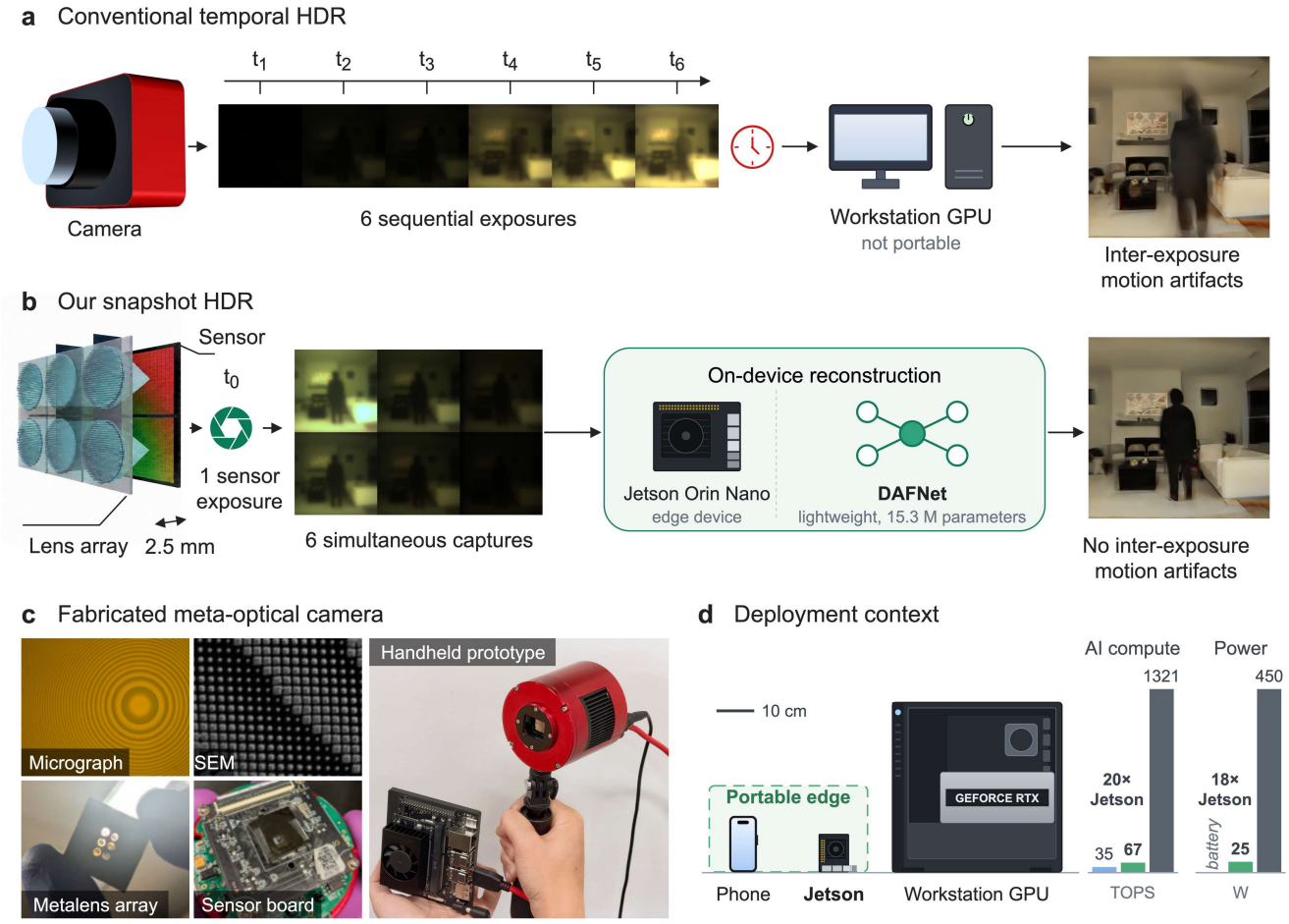


Figure 1. Single-snapshot HDR imaging with a portable meta-optical camera. a) Conventional HDR imaging acquires a sequential exposure burst. The resulting acquisition latency causes motion ghosting when the scene or camera moves between exposures. These methods further rely on computationally demanding reconstruction networks that require workstation-class hardware. **b)** Our 2×3 metalens array images six exposure levels onto the sensor in a single snapshot, eliminating the acquisition latency and the inter-exposure motion artifacts. We develop a lightweight reconstruction network that runs on an 8-GB NVIDIA Jetson Orin Nano, yielding HDR reconstructions on mobile-class edge devices. **c)** Optical and scanning electron micrographs of our fabricated metalens (top row), and the metalens array and sensor board (bottom row). When assembled, our handheld portable prototype achieves an optical track length of just 2.5 mm. **d)** The Jetson Orin Nano used in our system is comparable in compute capability and form factor to current-generation smartphone hardware, while operating far below workstation-class compute and power requirements. The Jetson provides $\sim 20\times$ less compute than an NVIDIA RTX 4090 (67 versus 1321 TOPS) while consuming $18\times$ less power (25 versus 450 W), and is closer in compute to the Apple A18 (35 TOPS).

wavelengths (Fig. 2b). This validates the broadband performance of the designed meta-optic. While residual wavelength- and field-dependent blur remains, we characterize and address it computationally in the reconstruction stage described below.

The attenuation layers create the exposure diversity required for HDR capture. To illustrate the multiplexed exposure bracket, we measured PSFs at varying exposure times to allow imaging of varying intensities. At 550 nm, a short 1.5 ms exposure places the highest-transmittance aperture near full scale while the lowest-transmittance aperture remains well above the noise floor (Fig. 2c). At a 30-fold longer exposure of 46 ms, the highest-transmittance aperture saturates while the lowest-transmittance aperture reaches full scale (Fig. 2d). Thus, regions that saturate in one aperture remain measurable in another.

Across the measured spectrum, the six metalenses form an approximately logarithmic exposure ladder, which we quantified by measuring the integrated signal per unit exposure time across all wavelengths. The exposure spans $14\times$ in band-averaged (490–680 nm) throughput, with approximately 0.76 exposure stops between adjacent metalenses (Fig. 2e). This yields a gap-free exposure bracket realized in space rather than time, with an approximately uniform logarithmic step between adjacent apertures at each wavelength. The relative attenuation varies moderately with wavelength (Fig. 2f), producing both brightness

and spectral differences across apertures, which we handle through the reconstruction network.

The spatial exposure ladder from the metalens array directly extends the captured dynamic range. The highest-transmittance channels preserve weak shadow signals, while the lowest-transmittance channels preserve highlights that saturate brighter apertures. Relative to the 72 dB single-aperture dynamic range of the IMX183 sensor at zero gain, the measured exposure ladder adds 23 dB, yielding an estimated 95 dB captured dynamic range in a single snapshot. The dynamic range extension is achieved optically rather than through modifications to the sensor, allowing the same spatial exposure-multiplexing principle to be combined with sensors of different native dynamic ranges.

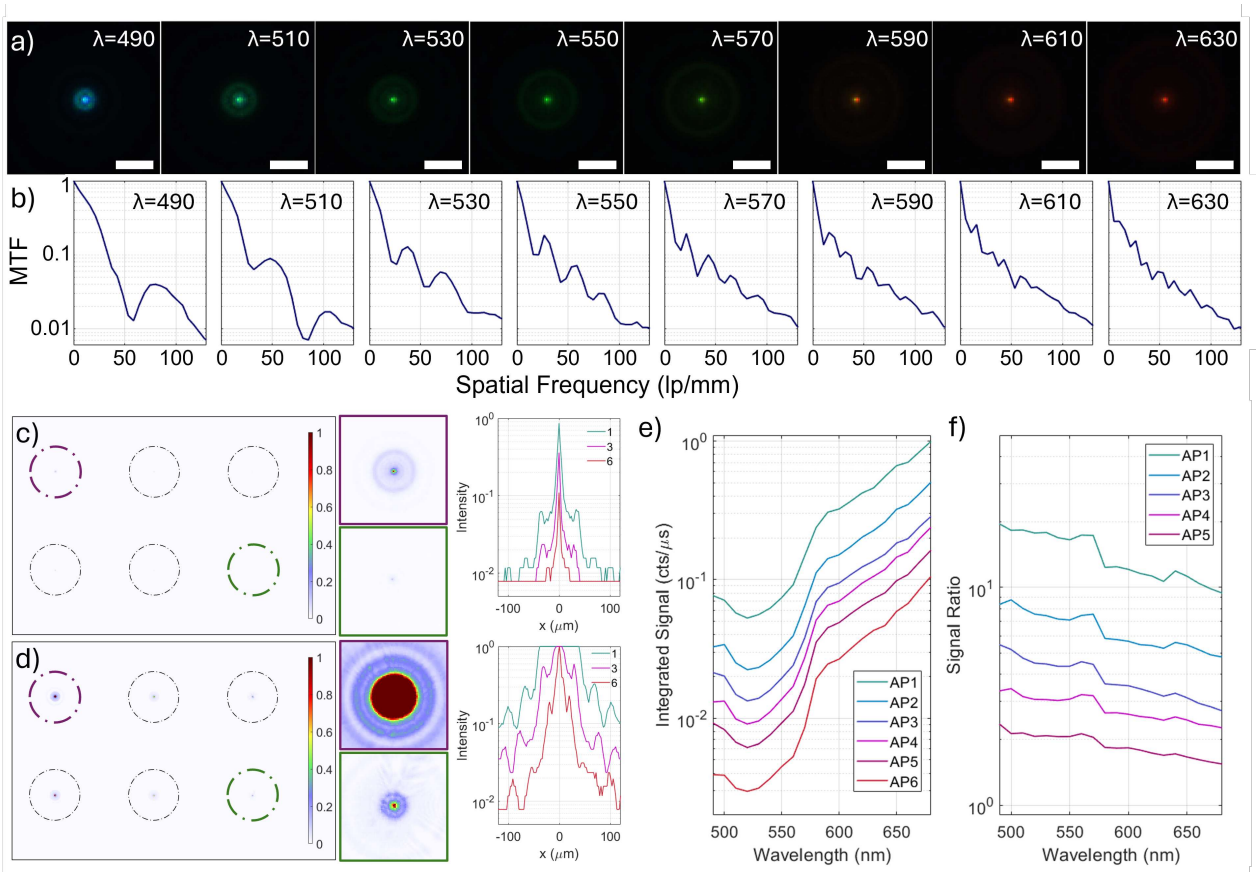


Figure 2. Meta-optic array characterization. a) Measured wavelength-dependent PSFs for a single lens across the visible wavelength range from 490 nm to 630 nm. Scale bars correspond to 50 μm. **b)** Corresponding modulation transfer function at each wavelength. **c,d)** Capture of a point source at short (1.5 ms) and long (46 ms) exposure, respectively. The dashed circles indicate the center of the sub-aperture fields in the 2×3 array, with apertures 1 (purple) and 6 (green) highlighted; the correspondingly framed insets show their PSFs magnified. Line plots show the horizontal intensity profiles through the peaks for apertures 1, 3, and 6. At short exposure, the max intensity in aperture 1 is close to unity, while aperture 6 is well above the noise floor. At long exposure aperture 1 is saturated, illustrating the complementary exposure coverage of the array. **e)** Integrated signal relative to exposure time per aperture (AP1–AP6) versus wavelength. **f)** Per-aperture signal normalized to the most attenuated aperture (AP6).

Adaptive reconstruction recovers HDR images from heterogeneous metalens measurements

Snapshot exposure multiplexing solves the temporal acquisition problem, but creates a challenging reconstruction problem. The six metalens measurements contain complementary exposure information across the scene's dynamic range. However, in addition to differences in brightness, each aperture also shows a different spectral response, spatially varying blur due to field-dependent aberrations, and viewpoint-dependent parallax due to the spatial separation between apertures. Directly combining these measurements therefore produces blurred, misaligned, and color-inconsistent results. Recovering an HDR image requires restoring each measurement, bringing the six views into correspondence, and then fusing the exposure information that remains reliable in each region.

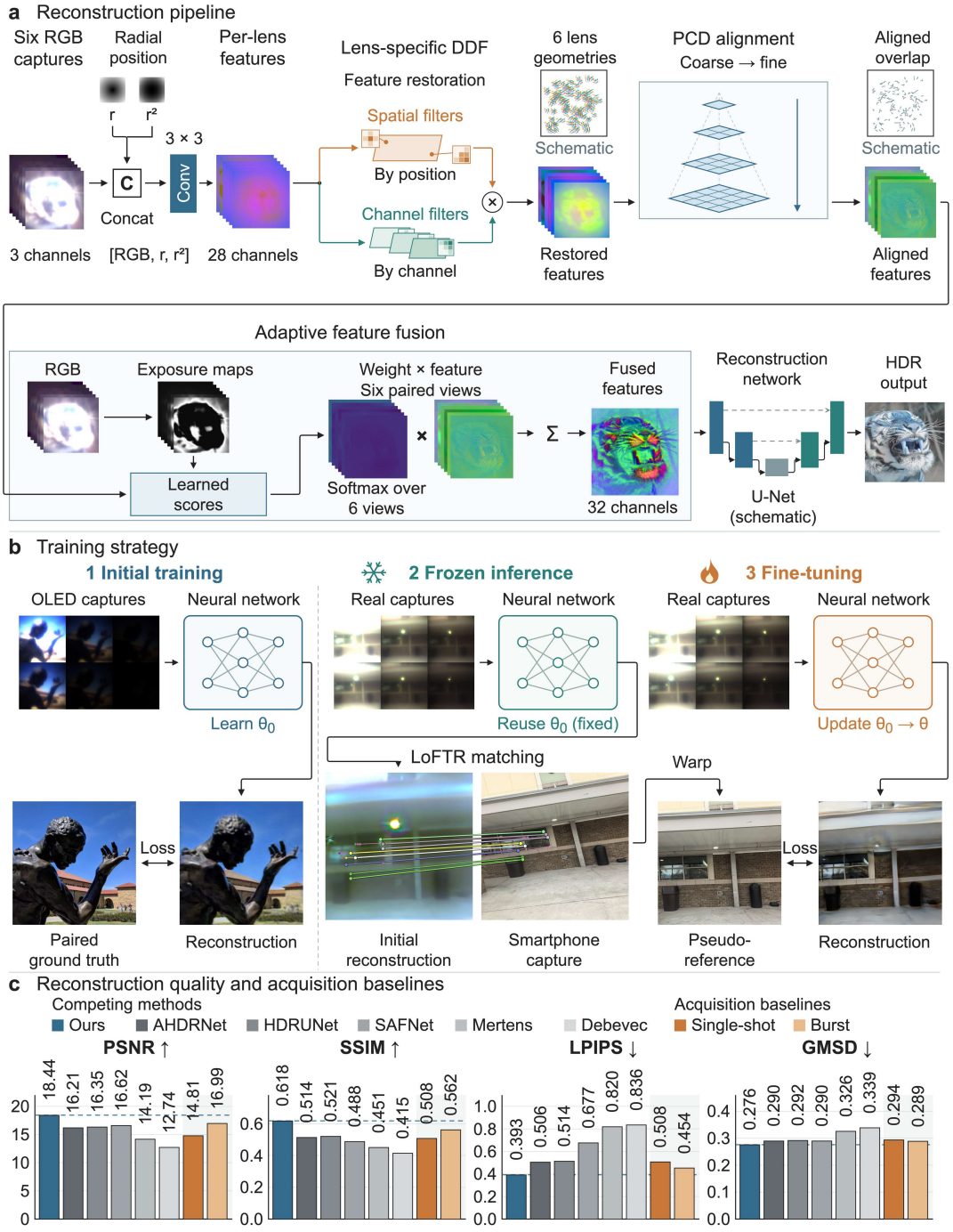


Figure 3. Computational reconstruction of exposure-coded metalens measurements. a) *Snapshot HDR reconstruction pipeline.* We devised DAFNet to restore, align, and fuse the six degraded sub-aperture measurements. We augment each input with radial field coordinates and process them with Decoupled Dynamic Filters (DDFs) to enable content- and field-adaptive restoration. A multiscale PCD (pyramid, cascading, and deformable convolution) alignment and confidence-weighted fusion combine the six views, and a residual U-Net produces the HDR reconstruction. Here, features before and after DDF are visualized using a shared PCA projection. **b)** *Bootstrapped training for real scenes.* The neural network is first pretrained on paired OLED measurements to obtain the parameters θ_0, which are later fine-tuned to θ. For real captures, the pretrained model produces a coarse reconstruction that enables registration to a co-captured smartphone HDR image using local feature transformer (LoFTR). This aligned HDR image provides a pseudo-reference for fine-tuning; this step is required only during training. **c)** *Quantitative comparison on held-out real scenes.* Our method achieves the best performance across pixel fidelity (PSNR), structural (SSIM), perceptual (LPIPS), and image quality (GMSD) metrics on in-the-wild measurements.

We address these coupled degradation challenges with DAFNet (DDF-Align-Fuse), a lightweight reconstruction network that jointly restores, aligns, and fuses the six sub-aperture measurements to produce a single tone-mapped HDR image (Fig. 3a). The first stage uses Decoupled Dynamic Filters (DDFs) [26] to compensate for field-dependent metalens blur. To provide explicit field-position cues, we concatenate each RGB input with a two-channel radial position code $[r, r^2]$, where $r = \sqrt{x^2+y^2}/\sqrt{2}$ and $x, y \in [-1, 1]$ are normalized pixel coordinates centered on the input image. Conditioned on both image content and these position cues, DDFs *dynamically* adapt their filtering to spatially varying optical degradation. Predicting a full dynamic filter at every location is expensive. We therefore *decouple* the filter into spatial and channel components, retaining local adaptivity at a substantially lower computational cost.

The restored sub-aperture features must also be aligned. Because the six metalenses occupy different positions on the array, they observe the scene from slightly different viewpoints, producing inter-aperture parallax; fabrication and mounting tolerances introduce additional fixed geometric offsets between the sub-images. Fusion on a fixed pixel grid therefore combines features corresponding to different scene locations, which can produce displaced or duplicated edges. Indeed, disabling the learned alignment offsets at inference produces visible edge distortion and ghosting in real captures (Supplementary Fig. S3). To address these spatial discrepancies, we use coarse-to-fine deformable alignment that locally adapts the sampling locations to register features across apertures. Once aligned, the six measurements are then fused using spatially varying confidence weights that account for their complementary exposure information. High-transmittance apertures provide signals in dark regions, whereas low-transmittance apertures preserve structure in highlights that saturate brighter measurements. In summary, DAFNet jointly addresses three degradations introduced by the optical system: DDF restores field-dependent blur, deformable alignment compensates for inter-aperture spatial misalignment, and confidence-weighted fusion combines complementary exposure information across apertures.

Training the reconstruction pipeline on real scenes introduces a second bottleneck: aligned HDR ground truth pairs are difficult to acquire. To achieve accurate training of the HDR recovery network, paired supervision is ultimately required, where the same scene is captured with the metalens array and a camera with high-quality HDR imaging capability. The metalens and reference cameras observe the scene from different viewpoints, producing depth-dependent parallax and occlusions. Moreover, the raw metalens measurements are too heavily blurred and too spectrally different from conventional HDR photographs for reliable direct registration. As a result, real scenes do not provide naturally pixel-aligned references for supervised training. We overcome this problem using a bootstrapped training strategy (Fig. 3b). We first pretrain DAFNet using controlled OLED-display measurements, where the approximately planar scene allows reliable alignment between the metalens measurements and known displayed targets. Although the OLED measurements do not reproduce the full radiance distribution of real HDR scenes, they allow for initial supervised training. In particular, this stage teaches the network an initial inverse mapping from degraded metalens measurements to visually meaningful coarse multi-aperture fusion reconstructions.

For real scenes, we use a smartphone to capture HDR references via the standard multi-exposure burst imaging pipeline. The pretrained DAFNet model first produces a coarse reconstruction of each metalens capture, recovering scene structure that facilitates matching to the co-captured smartphone HDR image despite residual blur and appearance differences. We establish correspondences using a pre-trained local feature transformer, LoFTR [27], whose coarse-to-fine, attention-based matching uses broader scene context to resolve locally ambiguous features. We then use RANSAC [28] to reject outlier matches and robustly estimate a homography, imposing global geometric consistency on the alignment. The smartphone image is warped into the metalens coordinates and used as a pseudo-reference for fine-tuning. The pretrained reconstruction is used only to establish correspondences; the pseudo-reference content comes from the independently captured smartphone HDR image. At inference, the system requires only the metalens snapshot measurements. On held-out real scenes, we compare against five existing HDR reconstruction methods applied to the same metalens captures: three learned methods (AHDRNet [29], HDRUNet [30], and SAFNet [31]) retrained on the same paired data (Supplementary Note 4), and two classical multi-exposure fusion methods (Mertens [32] and Debevec [1]). We also compare against two acquisition baselines using the same optics: a single-aperture single shot and a six-frame temporal burst. Our model achieves the best performance across pixel fidelity (PSNR), structural similarity (SSIM [33]), perceptual similarity (LPIPS [34]), and gradient-based image quality (GMSD [35]) (Fig. 3c). We attribute the advantage over the temporal burst, even for static scenes, to aperture diversity: all burst frames share the same aperture and therefore the same field-dependent aberrations, whereas the six apertures with varying transmittance exhibit different aberrations and provide complementary information for reconstructions.

Snapshot HDR imaging recovers detail across challenging real-world scenes

We next evaluate the complete system on real scenes with large brightness variations. Figure 4a shows three representative held-out examples comparing a conventional snapshot low dynamic range (LDR) capture and our snapshot metalens array reconstruction against a reference HDR measurement. A conventional single exposure loses information when bright regions saturate. The metalens array instead records these regions at several transmittance levels within the same exposure. DAFNet combines this complementary information to recover detail that is missing from the conventional LDR capture.

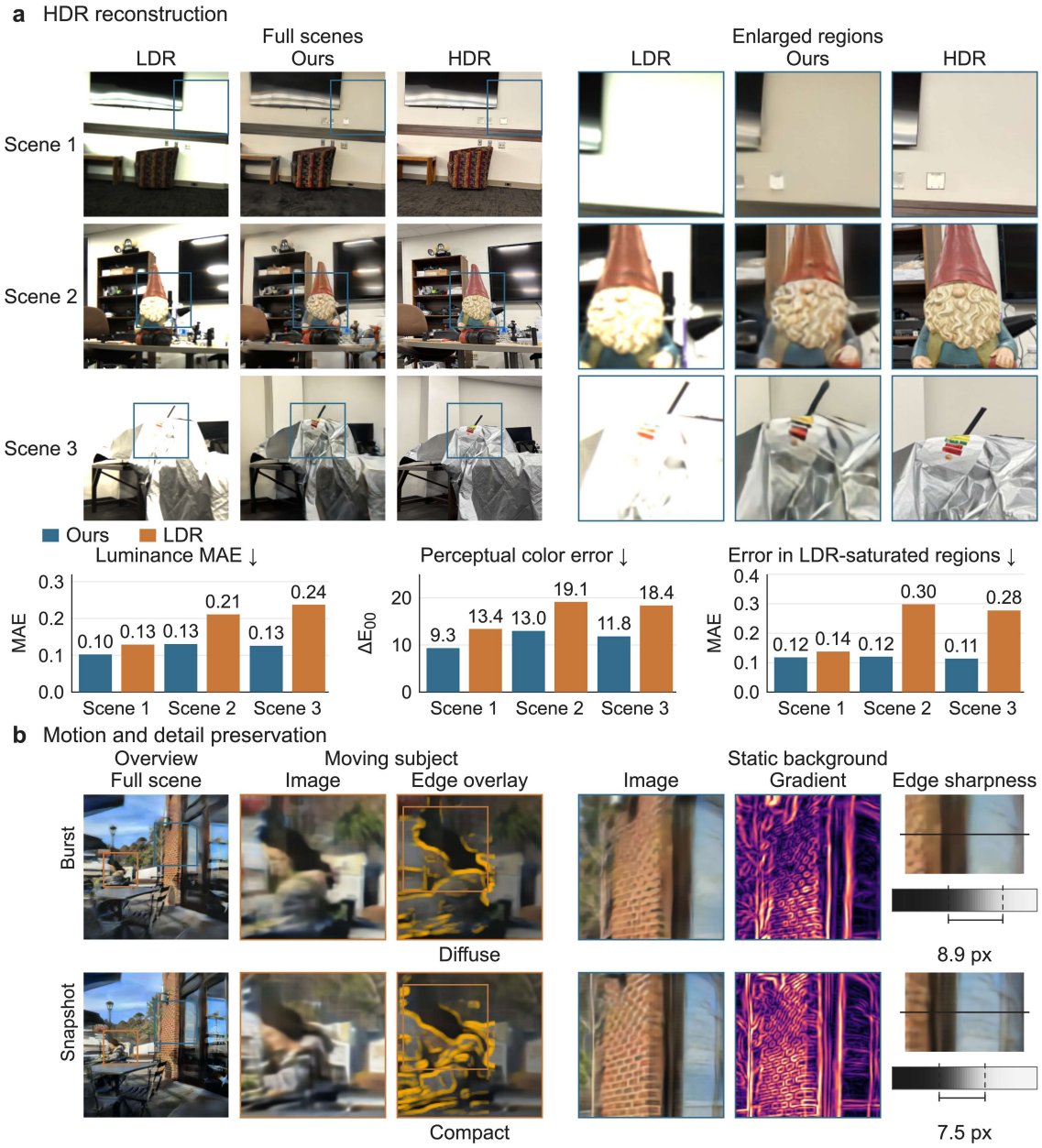


Figure 4. Snapshot HDR reconstruction fidelity and motion robustness in real scenes. a) *Reconstruction of three static HDR scenes with different forms of saturation.* In Scene 1, the LDR capture clips the bright wall and loses low-contrast structures; our reconstruction recovers the wall tone, outlet, and boundary above the baseboard. In Scene 2, saturation removes facial color and beard texture; our reconstruction recovers more of both. In Scene 3, saturation washes out the reflective surface; our reconstruction restores the colored strip and surrounding folds. The plots quantify reconstruction fidelity against the same HDR reference. Luminance mean absolute error (MAE) measures the mean display-space luminance error over the full image. Perceptual color error is the mean CIEDE2000 difference (ΔE_{00}) [36]. Saturated-region error is the RGB MAE computed only over pixels saturated in the LDR input, using the same mask for LDR and Ours. **b)** *Comparison of temporal-burst and snapshot HDR reconstruction during handheld capture of a moving subject.* Blue insets show a static region affected primarily by camera motion, while orange insets show the moving subject region affected by both camera and scene motion. For each inset, the yellow edge overlay and the gradient-magnitude map highlight edge sharpness and ghosting. The burst measurement produces diffuse and partially merged contours, whereas our snapshot method preserves more localized subject boundaries and background gradients. We measure the edge sharpness using a 10–90% intensity-transition width. For a static edge, this width decreases from 8.9 pixels for burst to 7.5 pixels for snapshot, an approximately 16% narrower transition and hence reduced motion blur.

The three scenes highlight different forms of information recovery and span three common failure modes of LDR imaging. In Scene 1, saturation makes the bright wall nearly uniform and removes low-contrast boundaries. Our reconstruction recovers the wall tone, electrical outlet, and boundary above the baseboard. In Scene 2, the LDR image loses texture and color in the saturated face and beard of the gnome. Our reconstruction recovers more of the beard structure and better separates the face, beard, and clothing. Scene 3 contains a strongly reflective surface. Saturation removes both the colored strip and folds in the surrounding material in the LDR image. These structures become visible in our snapshot HDR reconstruction.

We quantify these effects against the co-captured HDR references (Fig. 4a). Compared with the LDR capture, our reconstruction reduces global luminance error by 23–46% and perceptual color error by 31–36% across the three scenes. We also measure reconstruction error only within regions that saturate in the LDR image. Within these regions, our method reduces the error by 14–61%. The largest improvements occur in Scenes 2 and 3, where clipping removes substantial color and structure from the LDR input. These results show that the exposure-coded measurements of our metalens array recover scene information that is lost to clipping in a single exposure. These experiments directly evaluate the image regions where additional dynamic range is needed and complement the reconstruction metrics in Fig. 3c.

Snapshot acquisition becomes particularly important when the scene or camera moves. A temporal HDR burst records different exposure levels at different times. Motion between these measurements changes the position of scene content and can produce blur or ghosting after fusion. Our metalens array instead records the full exposure bracket simultaneously, removing this inter-exposure motion at acquisition.

Figure 4b compares snapshot and multi-exposure temporal-burst reconstruction during handheld capture of a moving subject. The region marked in orange contains both subject and camera motion. The burst reconstruction shows diffuse and partially merged contours around the face, upper body, and hand. In contrast, the snapshot reconstruction preserves a more compact subject profile and more localized boundaries. Edge overlays computed using the same gradient criterion across both measurements show the same trend, with responses more concentrated around the head, shoulder, and hand in the snapshot reconstruction. This indicates that simultaneous exposure capture reduces the spatial mixing caused by subject motion across a temporal exposure bracket.

We separately examine the static background marked in blue to isolate the effect of handheld camera motion. The burst reconstruction spreads the brick pattern and building boundary over a wider region, whereas the snapshot reconstruction preserves more localized edges. Gradient-magnitude maps computed from the same brick-wall region show broader responses for the burst reconstruction and more concentrated responses for our snapshot reconstruction. To quantify the visibility of motion blur, we measure the 10–90% intensity-transition width along an edge; a smaller width indicates a sharper transition and less visible motion blur. Along a representative static edge, the intensity-transition width decreases from 8.9 pixels for the burst to 7.5 pixels for the snapshot, corresponding to an approximately 16% narrower transition. A pixel-aligned HDR reference cannot be reliably obtained for a moving scene. We therefore use the dynamic experiment to only evaluate motion-induced blur and not the full-reference HDR fidelity.

Together, the static and dynamic experiments demonstrate two complementary benefits of spatial exposure multiplexing. It recovers information across a wider brightness range while avoiding the inter-exposure motion inherent to sequential HDR capture. Additional results across high-contrast, low-light, and dynamic scenes are provided in the Supplementary Information.

Efficient reconstruction enables deployment on embedded mobile-class hardware

Compact meta-optics provide little system-level advantage if their images require workstation-scale computation to reconstruct. We therefore deploy the complete reconstruction network on an NVIDIA Jetson Orin Nano with 8 GB of memory, representative of mobile-class hardware. Figure 5a shows consecutive snapshot measurements acquired during a dynamic scene (top row) and the corresponding video reconstructions (bottom row). Each frame is reconstructed independently from a single sensor exposure, and no temporal fusion or multi-frame buffering is used. These results demonstrate that our snapshot approach produces sharp, temporally consistent reconstructions across consecutive frames.

We next measure network inference across input resolutions and report results in Fig. 5b. The deployed model contains 15.33 M parameters. At 256×256 pixel resolution, inference on the Jetson Orin Nano takes 365 ms in eager mode, corresponding to 2.7 fps, with 0.32 GB peak memory, while the network reaches 8.5 fps at 128×128 pixels. These measurements report network-only inference and exclude sensor readout, preprocessing, and display overhead. We demonstrate inference latency reduction by a further 13–16% at 256–512 px through CUDA-graph execution; details are in the Supplementary Information.

The low memory footprint of our pipeline allows the complete reconstruction to run directly on the embedded device without workstation processing. Combined with the 2.5 mm optical track length, this brings both acquisition and reconstruction into a compact edge-computing regime. These results move computational metalens HDR imaging from offline reconstruction toward self-contained cameras for portable and embedded imaging on mobile-class hardware.

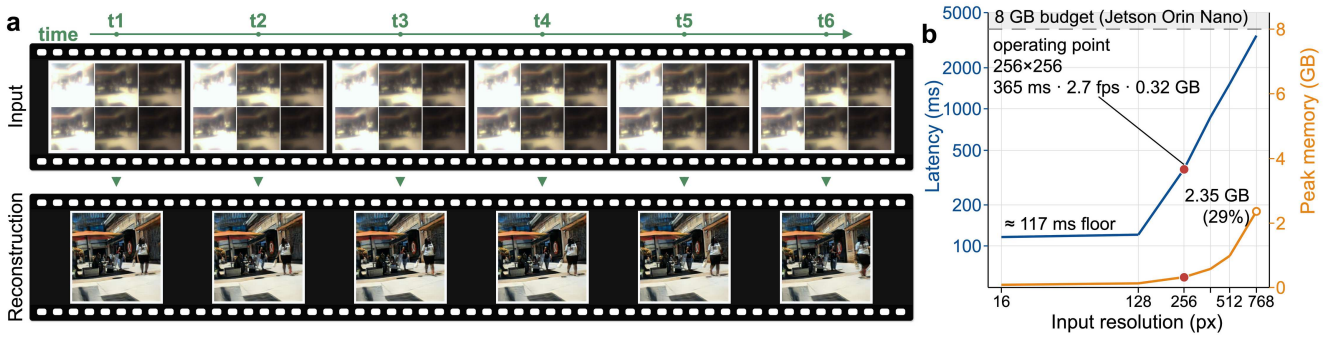


Figure 5. Continuous reconstruction and embedded hardware deployment. a) *Consecutive metalens snapshots and corresponding reconstructions during a dynamic scene.* From each snapshot sensor exposure (top), we independently reconstruct individual frames (bottom). Our results demonstrate temporally consistent continuous frame-by-frame reconstruction, without temporal fusion or multi-frame buffering. **b)** *Network-only inference latency and peak memory on an NVIDIA Jetson Orin Nano (8 GB) across input resolutions.* At 256×256 pixel resolution, inference takes 365 ms (2.7 fps) with 0.32 GB peak memory, while at 128×128 pixel resolution, the inference speed is around 8.5 fps. At low resolution, latency is dominated by a fixed kernel-launch and device overhead of about 117 ms (Supplementary Fig. S6). Additional details are reported in the Supplementary Information.

Discussion and Conclusion

We have demonstrated a snapshot HDR camera that moves exposure bracketing from time to optics. A 2×3 metalens array with engineered per-aperture transmittance captures six differently attenuated images within a single sensor exposure, and a lightweight network reconstructs the measurements directly on an embedded edge device. We engineered the metalens array to span a $\approx 14\times$ attenuation range, extending the captured dynamic range by ≈ 23 dB (≈ 3.8 stops). Because the exposure bracket is acquired simultaneously, the system avoids the inter-exposure motion that causes ghosting and motion blur in sequential HDR capture. Our experiments show this benefit during handheld imaging of moving scenes, while also recovering information lost to saturation in static high-contrast scenes.

The dynamic-range extension is introduced by the optical front-end rather than by modifying the detector. It is therefore complementary to improvements in sensor dynamic range and could be combined with detectors that provide higher native performance. In this sense, our metalens array acts as a passive front-end for snapshot dynamic-range extension. The attenuation itself does not require meta-optics and could, in principle, be implemented with conventional optics and neutral-density filters. The advantage of the metalens array is that exposure multiplexing and image formation are integrated into a thin planar element, making the optics lightweight and compatible with wafer-scale fabrication. Each aperture can independently control its phase profile while the transmittance ladder controls the exposure level. This combination enables six usable images within a 2.5-mm optical track length.

The current prototype also has limitations. Spatial multiplexing trades sensor area for exposure diversity because the six images share the same detector. The reconstructed spatial resolution is therefore lower than that of a conventional camera using the full sensor area. Resolution is further limited by the optical transfer characteristics of the current single-layer metalenses. The spatial separation between apertures also introduces parallax, which becomes harder to correct at short working distances and near depth discontinuities. Strong point sources can produce residual halo artifacts that are not fully removed by reconstruction. In addition, our real-scene HDR references require cross-camera registration and should therefore be interpreted as pseudo-references for relative comparison rather than pixel-exact radiometric ground truth. Because the reconstruction is display-referred and fine-tuned on smartphone HDR renderings, the real-scene metrics measure fidelity to these renderings rather than scene radiance. Finally, embedded reconstruction currently reaches 2.7 fps at 256×256 pixels and 8.5 fps at 128×128, demonstrating continuous on-device reconstruction. Interactive video-rate operation at high resolutions remains a future engineering target.

These limitations expose opportunities for joint optical and computational design. The number and arrangement of apertures, their transmittances, and the reconstruction network were not jointly optimized in the present system. Co-designing these parameters could better balance spatial resolution, dynamic range, parallax, and computational cost. Higher-resolution metalenses and higher native dynamic range sensors could improve spatial fidelity, while more efficient implementations of the reconstruction network could increase on-device frame rate. The planar architecture of the metalens array camera also provides a path toward scalable fabrication using high-throughput lithographic processes.

More broadly, our results show that exposure diversity need not be acquired sequentially. Information that is normally

distributed across multiple exposures can instead be encoded across a compact optical array and recovered computationally from a single sensor integration. Here, this principle provides ≈ 23 dB of snapshot dynamic-range extension while avoiding inter-exposure motion and supporting reconstruction on embedded hardware. This combination of optical multiplexing and efficient computation provides a path toward compact HDR cameras for dynamic real-world scenes. The same principle may extend to other forms of optical diversity that are conventionally acquired sequentially.

Methods

Hardware architecture

Our snapshot camera consists of six metalenses with different optical transmittances. The system integrates three primary components: a metalens array, an aperture array with baffles, and a CMOS sensor, as illustrated in Supplementary Fig. S1a. Each metalens is 1 mm in diameter with a 2 mm focal length, corresponding to a numerical aperture of 0.24. All six apertures share the same freeform phase profile and differ only in their optical transmittance. Optical coatings set the transmittance of each aperture, creating six distinct exposure levels. A 3D-printed baffle separates the six image fields on the sensor and suppresses inter-aperture crosstalk. The baffle and metalens array are mounted directly on the sensor cover glass. We use a Sony IMX183 1-inch CMOS sensor.

End-to-end metalens optimization

Aspherical phase surfaces We parameterize the metalens using an aspheric phase profile to reduce the spherical aberrations and coma inherent in standard lenses. Unlike a spherical surface, an aspheric profile varies its radius of curvature with radial position, providing additional degrees of freedom for aberration correction [37]. The rotationally symmetric central lens profile $\phi_c(r)$ is defined as:

$$\phi_c(r) = \frac{r^2\rho}{1+\sqrt{1-(1+\kappa)(r\rho)^2}} \cdot \frac{2\pi}{\lambda_{\text{nominal}}} \tag{1}$$

where $r = \sqrt{x^2+y^2}$ is the radial coordinate, ρ is the inverse radius of curvature, and κ is the conic constant. We set the nominal wavelength to $\lambda_{\text{nominal}} = 550$ nm, near the center of the visible spectrum. The parameters ρ and κ are selected through a grid search instead of being directly optimized to ensure that the resulting profiles remain within the space of feasible aspheric phase surfaces.

We further augment the phase profile with a differentiable, radially symmetric residual phase $\phi_{\text{res}}(r)$ to maximize focusing efficiency. This residual is parameterized by a vector of size $N = R/p$, where R is the metalens radius and $p = 350$ nm is the metalens pitch. The resulting lens phase is computed as:

$$\Phi(r) = \phi_c(r) + \phi_{\text{res}}(r) \tag{2}$$

Fabrication The metalens array is fabricated in silicon nitride on a quartz substrate using electron-beam lithography and reactive-ion etching. The six-level transmittance ladder is formed by sequential masked Au/Pd depositions on the backside of the chip. The completed array is integrated with a 3D-printed baffle and mounted directly on the sensor cover glass. Full fabrication parameters and process details are provided in Supplementary Information.

Neural electromagnetic proxy We devise a Multi-Layer Perceptron (MLP) as a bidirectional surrogate model for electromagnetic simulation. This neural proxy model maps between continuous phase profiles and discrete physical meta-atom geometries, represented as their duty cycles. This proxy circumvents the need for repeated full-wave electromagnetic simulations during the metalens design, accelerating the optimization process and reducing the computational overhead.

Optimization We optimize the uncoated reference metalens across the visible spectrum from 410 nm to 690 nm and over incidence angles $\Theta = \{0^\circ, \pm 5^\circ, \pm 7.5^\circ\}$, corresponding to a 15° field of view per lens in the array. The optimization objective combines three figures of merit: focal concentration measured by the spot intensity at each wavelength; modulation transfer; and the spot radius over an extended depth of focus:

$$Loss = \sum_{\lambda\in\Lambda}\sum_{\theta\in\Theta}\sum_{|r|<r_f(\theta)} -PSF_{\lambda,\theta}(r) + \texttt{ReLU}[C - MTF_{\lambda,\theta}(r)] + \sum_{\delta\in[-D,D]} w_{\lambda,\delta}(F+\delta). \tag{3}$$

Here, $PSF_{\lambda,\theta}$ denotes the point spread function at wavelength λ, angle of incidence θ at the focal plane (F). $MTF_{\lambda,\theta}$ is the corresponding modulation transfer function, and `ReLU` denotes the rectified linear unit, which clips negative values to zero. The parameter C defines the target MTF threshold, while $r_f(\theta)$ (<R; the metalens radius) defines the angle of incidence dependent radial region over which the PSF and MTF terms are aggregated. The final term $w_{\lambda,\delta}(F+\delta)$ evaluates the spot radius over

defocus distances $\delta \in [-D, D]$ around the focal plane at F. We use the one-dimensional Rayleigh–Sommerfeld propagation operator to obtain the radially symmetric spot profile (at incidence angle 0°) over the broadband wavelength range. This term encourages the metalens to maintain its focusing behavior over an extended depth range. The optimized metalens and its meta-atom distribution are shown in Fig. 1 and Supplementary Fig. S2.

Reconstruction network

The six sub-aperture images are demultiplexed and processed in a fixed read-out order. We use aperture 3 (AP3) as the geometric reference (Supplementary Fig. S1). The lightweight DAFNet (DDF-Align-Fuse) maps the six raw measurements to a single tone-mapped HDR RGB reconstruction in three stages: restoration and alignment, exposure-aware fusion, and residual reconstruction. In the first stage, a Decoupled Dynamic Filter (DDF) restores each sub-aperture image independently. The filter adapts to both local image content and field position to compensate for the spatially varying metalens PSF. Normalized radial coordinates $[r, r^2]$ are concatenated with each input to provide explicit field information. DDF factorizes the adaptive filtering into spatial and channel components to reduce computational cost. A coarse-to-fine deformable alignment module then registers the five neighboring sub-apertures to the reference in feature space. This step compensates for inter-aperture misregistration, comprising fixed geometric offsets between the sub-images and depth-dependent parallax from the spatial separation of the six apertures. In the second stage, an exposure- and confidence-adaptive fusion module combines the aligned features. A well-exposedness measure derived from the raw inputs identifies which apertures contain reliable information at each image location. The network uses this information to fuse the complementary exposures. In the third stage, a residual U-Net predicts a correction that is added to the raw reference image to produce the final reconstruction. Layer configurations, parameter counts, and component-level ablations are provided in the Supplementary Information.

Training and real-scene supervision

We construct paired supervision in two stages. We first pretrain DAFNet using controlled OLED-display measurements. Images are displayed on a planar OLED panel and captured by the metalens camera. Because the target is approximately planar, a homography can reliably register the displayed image to the reference sub-aperture, giving pixel-registered targets while the six sub-aperture inputs remain unwarped, as at inference. These pairs provide controlled supervision for learning an initial inversion of the metalens degradation and multi-aperture fusion. Direct supervision is more difficult for real three-dimensional scenes. The metalens camera and conventional reference camera observe the scene from different viewpoints, producing depth-dependent parallax and occlusions. In addition, the strong blur and spectral differences in the raw metalens measurements make direct registration to conventional photographs unreliable. We therefore construct real-scene supervision in the reconstruction domain.

For each real capture, the OLED-pretrained model first produces a coarse reconstruction. This image is sufficiently similar to a co-captured smartphone HDR image to establish correspondences using a local feature transformer, LoFTR [27]. We reject inconsistent matches and estimate the geometric transformation using RANSAC [28]. The smartphone HDR image is then warped into the metalens reconstruction coordinates and used as a pseudo-reference for fine-tuning. The pretrained reconstruction is used only to establish correspondence; the supervision itself comes from the independently captured smartphone HDR image.

The deployed model is initialized with the OLED-pretrained weights and fine-tuned on these real-scene pseudo-references. We use a combined perceptual-gradient objective consisting of a Charbonnier loss, SSIM, LPIPS, and gradient losses, together with an output-range penalty and a radially weighted L1 term (Supplementary Note 2), and optimize the network using AdamW [38]. The 1101 OLED captures used for pretraining are split 80/20 into training and validation sets, and the 628 real-world captures used for fine-tuning are likewise split 80/20, with the held-out real captures used for the real-scene evaluation. For controlled component-level ablations, models are instead trained from scratch to provide clean single-variable comparisons. Full acquisition, registration, training, and optimization details are provided in the Supplementary Information.

References

1. Debevec, P. E. & Malik, J. Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 369–378, 10.1145/258734.258884 (1997).

2. Akahane, N. *et al.* A sensitivity and linearity improvement of a 100-db dynamic range CMOS image sensor using a lateral overflow integration capacitor. *IEEE J. Solid-State Circuits* **41**, 851–858, 10.1109/JSSC.2006.870753 (2006).

3. Miyauchi, K. *et al.* A stacked back side-illuminated voltage domain global shutter CMOS image sensor with a 4.0 μm multiple gain readout pixel. *Sensors* **20**, 486, 10.3390/s20020486 (2020).

4. So, H., Martel, J. & Wetzstein, G. MantissaCam: Learning snapshot high-dynamic-range imaging with perceptually-based in-pixel irradiance encoding. In *IEEE International Conference on Computational Photography (ICCP)* (2022).

5. Hasinoff, S. W. *et al.* Burst photography for high dynamic range and low-light imaging on mobile cameras. *ACM Transactions on Graph.* **35**, 192, 10.1145/2980179.2980254 (2016).

6. Wu, S., Xu, J., Tai, Y.-W. & Tang, C.-K. Deep high dynamic range imaging with large foreground motions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 117–132, 10.1007/978-3-030-01216-8_8 (2018).

7. Meuel, H., Ackermann, H., Rosenhahn, B. & Ostermann, J. Physical high dynamic range imaging with conventional sensors. In *2018 Picture Coding Symposium (PCS)*, 209–213, 10.1109/PCS.2018.8456268 (2018).

8. Woo, T. *et al.* High-throughput high-dynamic range imaging by spatiotemporally structured illumination. *APL Photonics* **7**, 106106, 10.1063/5.0099780 (2022).

9. Tocci, M. D., Kiser, C., Tocci, N. & Sen, P. A versatile HDR video production system. *ACM Transactions on Graph.* **30**, 41:1–41:10, 10.1145/2010324.1964936 (2011).

10. Nayar, S. K. & Mitsunaga, T. High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 472–479, 10.1109/CVPR.2000.855857 (2000).

11. Alghamdi, M. M., Fu, Q., Thabet, A. K. & Heidrich, W. Reconfigurable snapshot HDR imaging using coded masks and inception network. In *Vision, Modeling and Visualization (VMV)*, 37–45, 10.2312/vmv.20191316 (The Eurographics Association, 2019).

12. Sun, Q., Tseng, E., Fu, Q., Heidrich, W. & Heide, F. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1383–1393, 10.1109/CVPR42600.2020.00146 (2020).

13. Xie, M., Chan, M. & Metzler, C. Snapshot high dynamic range imaging with a polarization camera (2023). 2308.08094.

14. Pistellato, M., Fatima, T. & Wimmer, M. Exploiting light polarization for deep HDR imaging from a single exposure. *Sensors (Basel, Switzerland)* **23** (2023).

15. Yu, N. & Capasso, F. Flat optics with designer metasurfaces. *Nat. Mater.* **13**, 139–150, 10.1038/nmat3839 (2014).

16. Chen, W. T., Zhu, A. Y. & Capasso, F. Flat optics with dispersion-engineered metasurfaces. *Nat. Rev. Mater.* **5**, 604–620, 10.1038/s41578-020-0203-3 (2020).

17. Pan, M. *et al.* Dielectric metalens for miniaturized imaging systems: progress and challenges. *Light. Sci. & Appl.* **11**, 195, 10.1038/s41377-022-00885-7 (2022).

18. Brookshire, C., Liu, Y., Chen, Y., Chen, W. T. & Guo, Q. MetaHDR: single shot high-dynamic range imaging and sensing using a multifunctional metasurface. *Opt. Express* **32**, 26690–26707, 10.1364/OE.528270 (2024).

19. Bang, J., Kim, C. & Lee, B. Single-shot metasurface-based HDR imaging system with deep neural network. In *Frontiers in Optics + Laser Science*, JTu5A.59, 10.1364/FIO.2022.JTu5A.59 (Optica Publishing Group, 2022).

20. Zhang, N. *et al.* Single-layer metasurface for snapshot high-dynamic-range imaging. *Photonics Res.* **13**, 2270–2280, 10.1364/PRJ.563285 (2025).

21. Liu, Y. & Guo, Q. METAH2: A snapshot metasurface HDR hyperspectral camera. In *2025 IEEE International Conference on Image Processing (ICIP)*, 1918–1923, 10.1109/ICIP55913.2025.11084531 (2025).

22. Mandal, D., Peng, Z., Wang, Y. & Chakravarthula, P. Enabling high-quality in-the-wild imaging from severely aberrated metalens bursts. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 849–859, 10.1109/WACV61042.2026.00089 (2026).

23. Chakravarthula, P. *et al.* Thin on-sensor nanophotonic array cameras. *ACM Transactions on Graph.* **42**, 1–18, 10.1145/3618398 (2023).

24. Yeo, J. *et al.* Neural network enabled wide field-of-view imaging with hyperbolic metalenses. *Nanophotonics* **14**, 3329–3337, 10.1515/nanoph-2025-0354 (2025).

25. Fröch, J. E. *et al.* Beating spectral bandwidth limits for large aperture broadband nano-optics. *Nat. communications* **16**, 3025 (2025).

26. Zhou, J., Jampani, V., Pi, Z., Liu, Q. & Yang, M.-H. Decoupled dynamic filter networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 6647–6656, 10.1109/CVPR46437.2021.00658 (2021).

27. Sun, J., Shen, Z., Wang, Y., Bao, H. & Zhou, X. LoFTR: Detector-free local feature matching with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 8922–8931 (2021).

28. Fischler, M. A. & Bolles, R. C. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24**, 381–395, 10.1145/358669.358692 (1981).

29. Yan, Q. *et al.* Attention-guided network for ghost-free high dynamic range imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 1751–1760, 10.1109/CVPR.2019.00185 (2019).

30. Chen, X., Liu, Y., Zhang, Z., Qiao, Y. & Dong, C. HDRUNet: Single image HDR reconstruction with denoising and dequantization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 354–363, 10.1109/CVPRW53098.2021.00045 (2021).

31. Kong, L. *et al.* SAFNet: Selective alignment fusion network for efficient HDR imaging. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 256–273, 10.1007/978-3-031-73347-5_15 (2024).

32. Mertens, T., Kautz, J. & Van Reeth, F. Exposure fusion. In *15th Pacific Conference on Computer Graphics and Applications (PG)*, 382–390, 10.1109/PG.2007.17 (IEEE Computer Society, 2007).

33. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Process.* **13**, 600–612 (2004).

34. Zhang, R., Isola, P., Efros, A. A., Shechtman, E. & Wang, O. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018).

35. Xue, W., Zhang, L., Mou, X. & Bovik, A. C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE Transactions on Image Process.* **23**, 684–695, 10.1109/TIP.2013.2293423 (2014).

36. Sharma, G., Wu, W. & Dalal, E. N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color. Res. & Appl.* **30**, 21–30, https://doi.org/10.1002/col.20070 (2005).

37. Schuhmann, R. Description of aspheric surfaces. *Adv. Opt. Technol.* **8**, 267–278, 10.1515/aot-2019-0011 (2019).

38. Loshchilov, I. & Hutter, F. Decoupled Weight Decay Regularization. In *International Conference on Learning Representations* (2019).

Author contributions statement

P.C. and J.F. conceived the idea and supervised the overall project. P.C. and D.M. designed the metasurfaces, and J.F. and M.L. fabricated them. Z.P. and D.M. developed the image restoration approach with inputs from Y.W., Z.Z, and P.C. Z.P. performed the experiments with help from D.M., Y.W., and Z.Z. P.C., J.F., and A.M. interpreted the results together with all the authors. P.C., Z.P., J.F., and D.M. wrote the manuscript with input from all the authors.

Competing interests

The authors declare no competing interests.

Supplementary Information

Supplementary Information accompanies this manuscript as part of the submission files.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- snapshotHDRsuppl.pdf