

ARTICLE

Open Access

# Super-resolution image projection over an extended depth of field using a diffractive decoder

Hanlong Chen<sup>1,2,3</sup>, Çağatay Işıl<sup>1,2,3</sup>, Che-Yung Shen<sup>1,2,3</sup> , Shiqi Chen<sup>1,2,3</sup> , Tianyi Gan<sup>1</sup>, Mona Jarrahi<sup>1</sup>  and Aydogan Ozcan<sup>1,2,3</sup> ✉

## Abstract

Image projection systems must be efficient in data storage, computation and transmission while maintaining a large space-bandwidth-product (SBP) at their output. Here, we introduce a hybrid image projection system that achieves extended depth-of-field (DOF) with improved resolution, combining a convolutional neural network (CNN)-based digital encoder with an all-optical diffractive decoder. A CNN-based encoder compresses input images into compact phase representations, which are subsequently displayed by a low-resolution (LR) projector and processed by an analog diffractive decoder for all-optical image reconstruction. This optical decoder is completely passive, designed to synthesize pixel super-resolved image projections that feature an extended DOF while eliminating the need for additional power consumption for super-resolved image reconstruction. Our pixel super-resolution (PSR) image projection system demonstrates high-fidelity image synthesis over an extended DOF of  $\geq 250\lambda$ , where  $\lambda$  is the illumination wavelength, concurrently offering up to  $\sim 16$ -fold SBP improvement at each lateral plane. The feasibility of this approach is validated through proof-of-concept experiments conducted in both the terahertz and visible spectral regions, demonstrating the scalability of our hybrid system across different wavelengths. This image projection architecture can reduce data storage and transmission requirements for display systems without imposing additional power constraints on the optical decoder. Beyond extended DOF PSR image projection, the underlying principles of this approach can be extended to various applications, including optical metrology and microscopy.

## Introduction

Augmented and virtual reality (AR/VR) systems hold immense promise across diverse fields, including education, entertainment, and healthcare<sup>1,2</sup>. These systems aim to deliver immersive user experiences. Near-eye displays, while crucial for achieving this objective, have presented a hurdle in realizing the full potential of AR/VR technology. A critical issue plaguing most near-eye displays, including stereoscopic, autostereoscopic, and multi-view systems, is the well-documented discomfort and fatigue experienced by users<sup>3,4</sup>. This primarily stems from the mismatch between vergence and accommodation cues of the eye. Holographic displays, in contrast to these widely adopted

approaches, possess the ability to offer all depth cues of the human visual system. This capability can potentially alleviate eye strain and visual discomfort associated with vergence-accommodation conflict<sup>4–9</sup>, which is important for AR/VR systems. However, the space-bandwidth product (SBP) of holographic displays is constrained by the limited pixel count of spatial light modulators (SLMs). In addition to this, further research into advanced computer-generated holography (CGH) techniques and supporting hardware is crucial for enabling real-time, full-parallax holographic displays with low power consumption<sup>7,8,10,11</sup>. Advances in machine learning techniques have also been applied to improve hologram calculation in terms of accuracy and speed<sup>12–21</sup>, as well as to enhance the SBP of SLMs<sup>22,23</sup> through passive surfaces. An alternative strategy involves computing holograms on a remote server and transmitting them to local receivers with limited, low-cost processors<sup>4</sup>. However, hologram compression presents challenges, as conventional image compression

Correspondence: Aydogan Ozcan ([ozcan@ucla.edu](mailto:ozcan@ucla.edu))

<sup>1</sup>Electrical and Computer Engineering Department, University of California, Los Angeles, CA, USA

<sup>2</sup>Bioengineering Department, University of California, Los Angeles, CA, USA

Full list of author information is available at the end of the article

These authors contributed equally: Hanlong Chen, Çağatay Işıl, Che-Yung Shen

© The Author(s) 2026



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.



techniques like JPEG (Joint Photographic Experts Group) inevitably sacrifice the fine details critical for synthesizing high-quality holographic images. These limitations highlight the need for innovative image projection approaches that can simultaneously address the current challenges associated with data transmission, processing power, and restricted SBP in display technologies.

Here, we present a deep learning-enabled image projection approach that synergistically integrates a digital encoder with a passive all-optical decoder, achieving both extended depth of field (DOF) and pixel super-resolution (PSR). This hybrid architecture employs a convolutional neural network (CNN) to efficiently encode high-resolution image information into compact phase-only representations, which are subsequently displayed by a low-resolution (LR) projector and processed by a passive diffractive decoder to synthesize super-resolved images over an extended DOF. We validated the scalability of this framework through comprehensive numerical simulations and proof-of-concept experiments conducted in both the terahertz and visible parts of the spectrum, demonstrating its ability to project images over a DOF of  $\geq 250\lambda$  with approximately 16-fold SBP improvement compared to the SBP of the input projector, where  $\lambda$  is the illumination wavelength. Furthermore, we rigorously evaluated the system's external generalization capabilities on structurally distinct unseen object classes and its robustness against non-idealities, including component misalignment, phase bit depth, and quantization. This digital encoder–analog decoder architecture provides a viable path to address data storage and transmission requirements while reducing computational costs, as it incurs no additional power consumption or significant latency at the user end, leveraging the intrinsic advantages of its all-optical passive decoder. This hybrid super-resolution image projection system, with its extended DOF and adaptability across different wavelengths, might find various applications in optical metrology and microscopy, as well as in next-generation holographic displays.

## Results

The hybrid architecture of our extended DOF PSR image projection system integrates a digital encoder with an all-optical passive diffractive decoder, as depicted in Fig. 1. The electronic encoder, shown in Fig. 1a, employs a lightweight CNN-based encoder to process the input image through a sequence of convolutional layers (Fig. 1a). Each layer is equipped with instance normalization<sup>24</sup> and Parametric Rectified Linear Unit (PReLU) activation<sup>25</sup>. In our proof of concept demonstration, this CNN-based encoder is designed to compress grayscale images with  $24 \times 24$  pixels into compact phase

representations with  $\frac{24}{k} \times \frac{24}{k}$  pixels where  $k = 6$  in Fig. 1a – i.e., a reduction of 36-fold in the total number of pixels per image. The all-optical analog decoder is composed of three diffractive layers ( $L = 3$ ), illustrated in Fig. 1b, and it is jointly designed/optimized to process the encoded compact phase patterns to synthesize high-resolution images over an extended DOF (EDOF), denoted by  $z_e$ . Stated differently, our projection framework performs a continuous EDOF projection while improving the effective SBP of the reconstructed images, which is illustrated using 2D projection slices at the selected discrete positions represented by  $z_{3,1}, z_{3,2}, \dots, z_{3,i}, \dots, z_{3,n}$ , with the corresponding cross-sectional images denoted as  $I_1, I_2, \dots, I_i, \dots, I_n$ , respectively. To train our hybrid models, we utilized the EMNIST<sup>26</sup> handwritten letter dataset, comprising 80,000 training, 8800 validation, and 14,800 testing images, each with  $24 \times 24$  pixels, and employed the AdamW optimizer, as detailed in the **Methods** section.

Figure 2 compares the image projection performances of diffractive projection systems with different numbers of passive diffractive layers ( $L = 1, 3$ ) trained for different PSR factors ( $k = 4, 6, 8$ ). Each configuration, represented by a distinct row in Fig. 2, is separately trained (see Supplementary Figure S1 for the decoders' phase profiles corresponding to each design), and each design is blindly evaluated using the test dataset not seen by the models during the training phase. As an example, a test image of handwritten “P” is processed through the diffractive image projection systems and the resulting EDOF image projections are presented using lateral cross-sections at selected propagation distances, which showcase the system's ability to maintain high-fidelity projections throughout an extended DOF, axially spanning  $z_e \approx 266.85\lambda$ . For instance, the three-layer-based decoder configurations ( $L = 3$ ) effectively maintain a similar image quality across the entire axial depth range at PSR factors of  $k = 4, 6$ , and 8. On the other hand, single-layer decoder configurations ( $L = 1$ ) show a more pronounced trade-off between resolution and DOF performance (see Fig. 2). Notably, configurations with a free-space decoder ( $L = 0$ , i.e., no diffractive decoder) present significantly deteriorated images synthesized across the same DOF, confirming the essential role of the diffractive decoder layers jointly optimized with the digital encoder at the front-end.

We further explored our hybrid projection system by demonstrating its *external generalization* capabilities in Fig. 3. Horizontal and vertical gratings, which differ from the training and test datasets used in the previous results, were employed to test the external generalization of our extended DOF PSR image projection system. As illustrated in Fig. 3, despite being trained using solely the EMNIST dataset, the three-layer-based decoder models accurately synthesized images of gratings with a linewidth



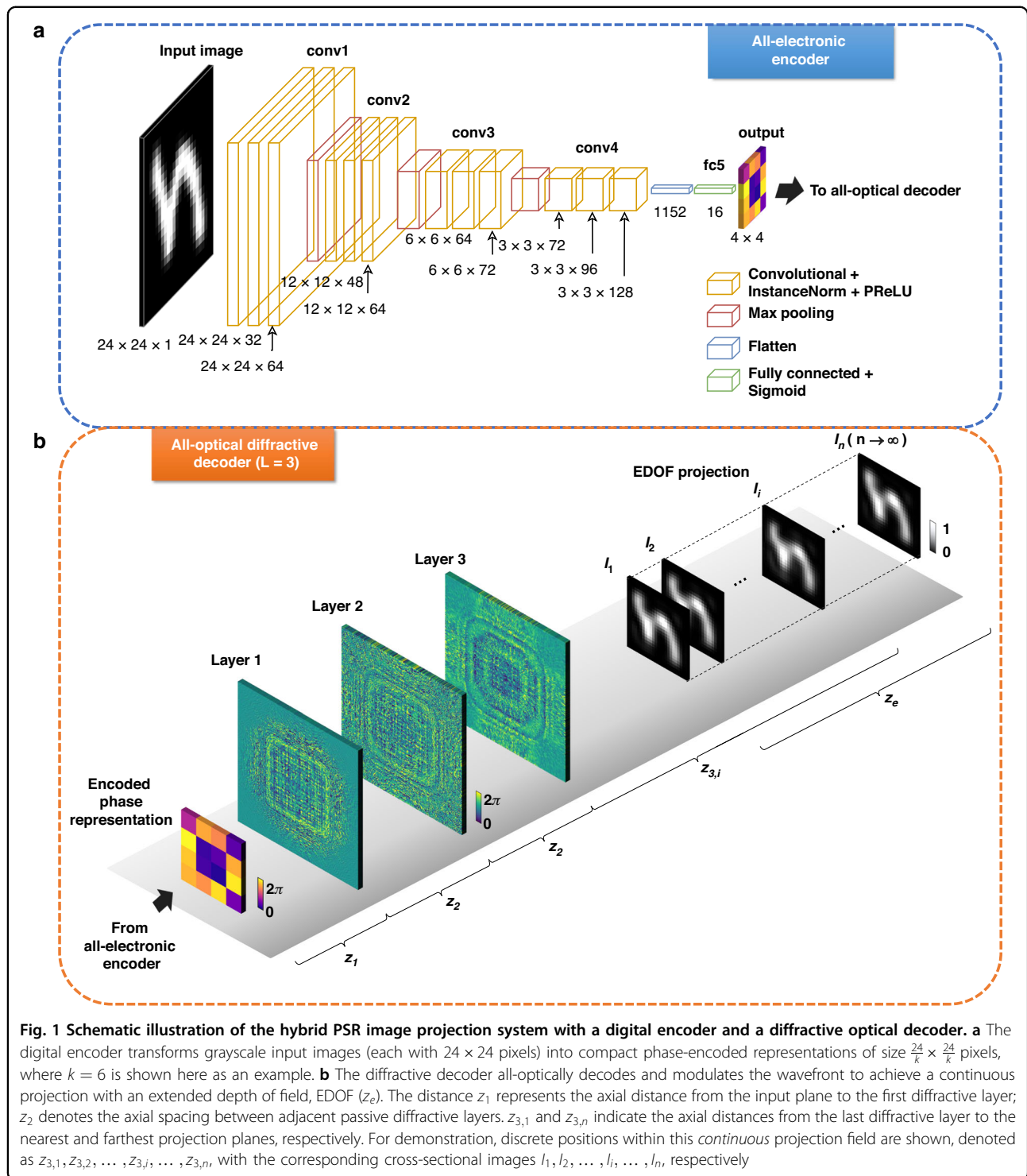


**Fig. 1** Schematic illustration of the hybrid PSR image projection system with a digital encoder and a diffractive optical decoder. **a** The digital encoder transforms grayscale input images (each with  $24 \times 24$  pixels) into compact phase-encoded representations of size  $\frac{24}{k} \times \frac{24}{k}$  pixels, where  $k = 6$  is shown here as an example. **b** The diffractive decoder all-optically decodes and modulates the wavefront to achieve a continuous projection with an extended depth of field, EDOF ( $z_e$ ). The distance  $z_1$  represents the axial distance from the input plane to the first diffractive layer;  $z_2$  denotes the axial spacing between adjacent passive diffractive layers.  $z_{3,1}$  and  $z_{3,n}$  indicate the axial distances from the last diffractive layer to the nearest and farthest projection planes, respectively. For demonstration, discrete positions within this *continuous* projection field are shown, denoted as  $z_{3,1}, z_{3,2}, \dots, z_{3,i}, \dots, z_{3,n}$ , with the corresponding cross-sectional images  $I_1, I_2, \dots, I_i, \dots, I_n$ , respectively

of  $10.6\lambda$  ( $\sim 7.94$  mm) over an extended DOF  $\{z_3 \in \mathbb{R} \mid 150 \leq z_3 \leq 350\}$  for different PSR factors of  $k = 4, 6, 8$ . Notably, the single-layer designs for  $k = 4$  and  $6$  also performed image projections over an extended DOF with some minor degradation in image quality compared to three-layer decoder results. Figure 3 further reveals that

our extended DOF method achieves up to  $\sim 16$ -fold SBP improvement compared to the input/encoding plane, in addition to proving external generalization of the presented framework. Specifically, the model trained for  $k = 8$  successfully projected gratings with a linewidth of  $10.6\lambda$  ( $\sim 7.94$  mm), translating to an effective pixel size of

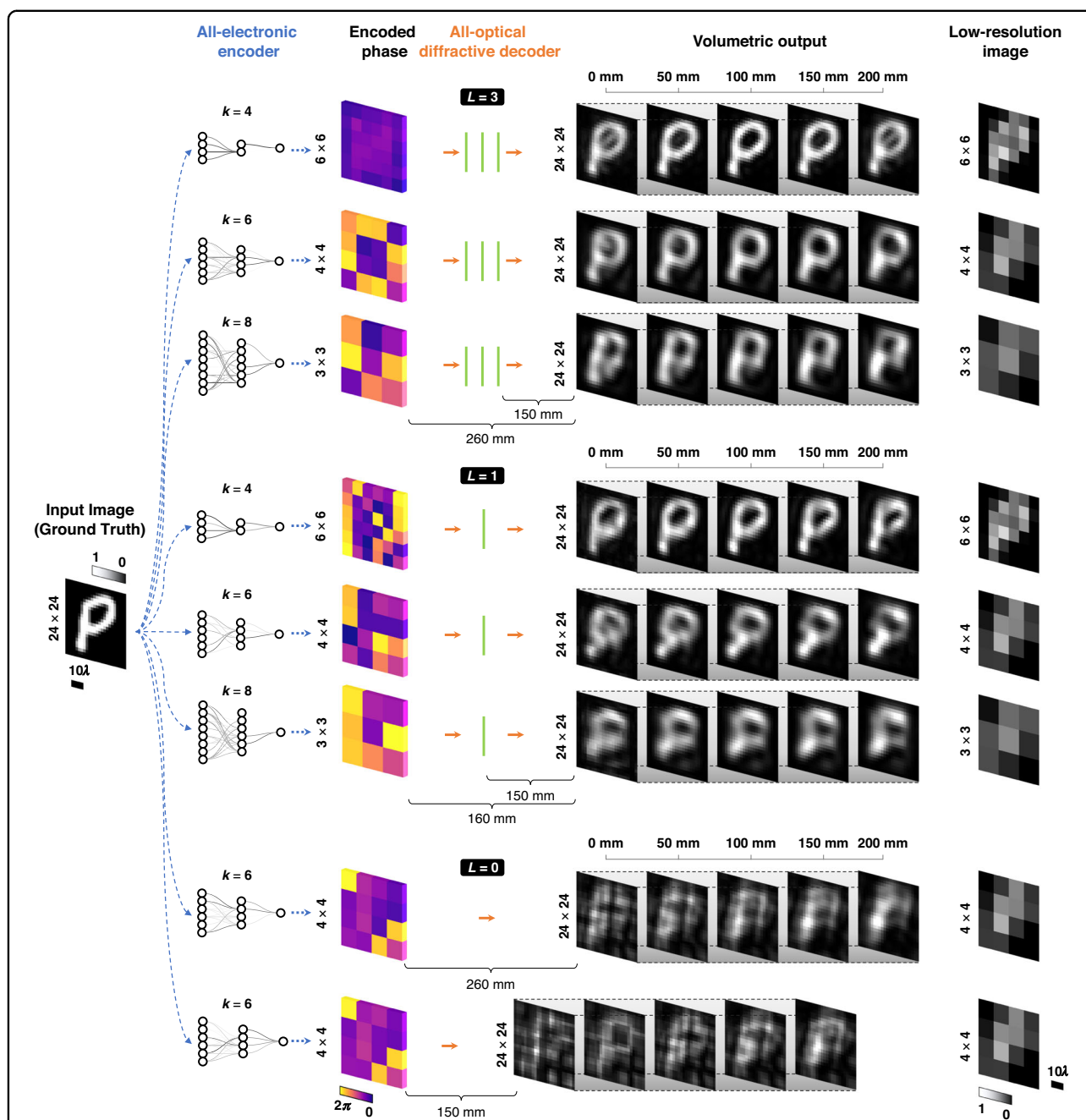


**Fig. 2 Performance comparison of hybrid PSR image projection systems.** The input images (each  $24 \times 24$  pixels) are processed through our PSR image projection systems trained for varying PSR factors ( $k = 4, 6, 8$ ) and different numbers of passive diffractive layers ( $L = 1, 3$ ) for the diffractive decoder. Each row represents a distinct system configuration with jointly trained encoder-decoder pairs, optimized independently from other configurations. The outputs are continuous projections with an extended DOF, i.e.,  $z_e = 200$  mm ( $\sim 266.85\lambda$ ). For illustration purposes, several propagation distances, i.e., 0 mm, 50 mm ( $\sim 66.71\lambda$ ), 100 mm ( $\sim 133.43\lambda$ ), 150 mm ( $\sim 200.14\lambda$ ), 200 mm ( $\sim 266.85\lambda$ ), are selected to showcase cross-sectional views. The rightmost column displays the low-resolution images corresponding to each PSR factor,  $k$ . The bottom two rows illustrate system configurations with *no* diffractive layers (the encoder-only cases, where  $L = 0$ ), serving as a comparison, highlighting the significance of the diffractive decoder

$5.3\lambda$  ( $\sim 3.97$  mm) at the output plane. These results are obtained using an effective pixel size of  $16$  mm ( $\sim 21.35\lambda$ ) at the encoded phase plane, corresponding to a  $\sim 16$ -fold improvement in the effective SBP<sup>22,27</sup>. Even smaller

linewidths can be imaged through this diffractive platform by specifically training it, end-to-end, for the projection of densely-packed fringe patterns as illustrated in Supplementary Figure S2.



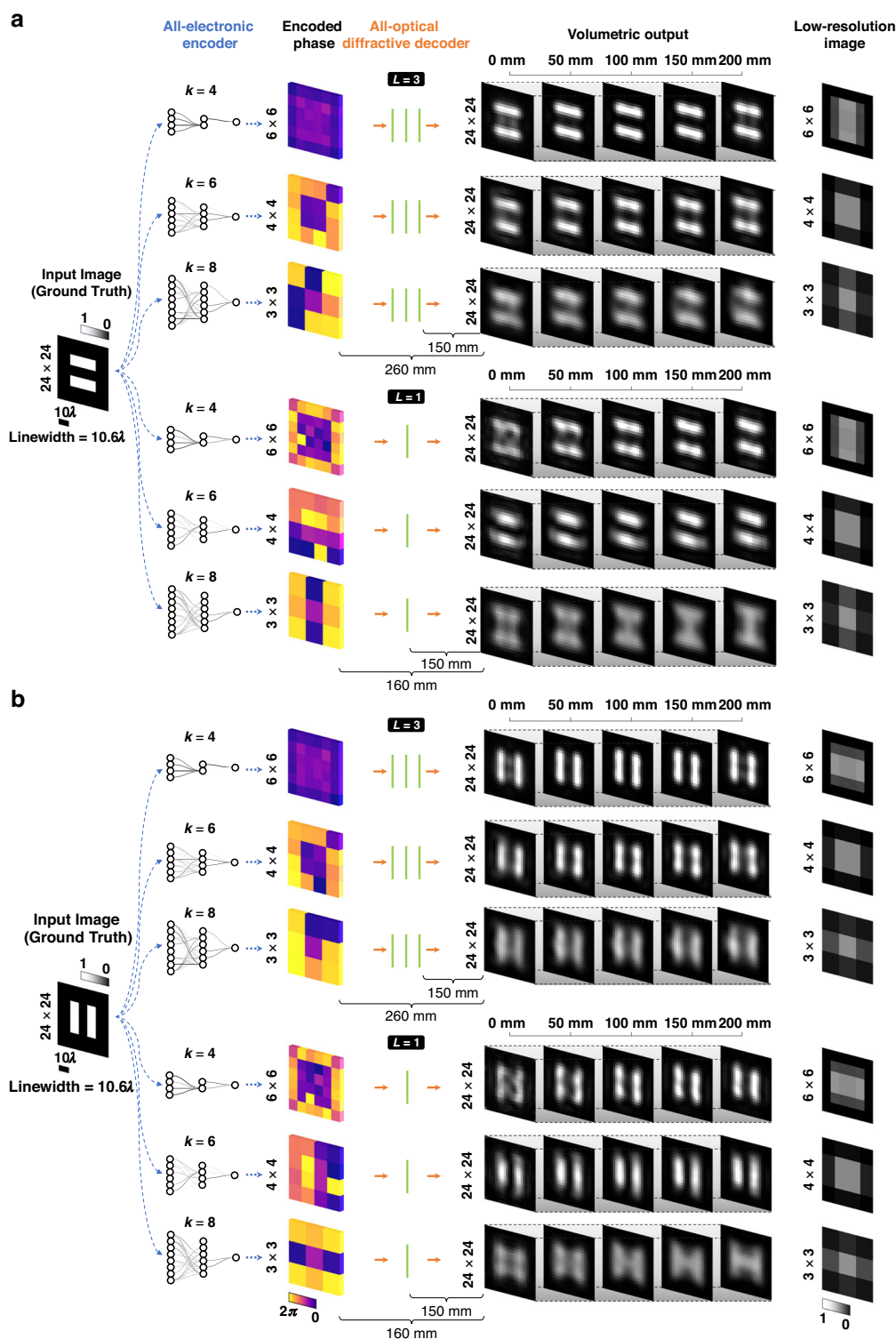


**Fig. 3 External generalization performance of hybrid PSR image projection systems on grating patterns—never seen before.** This figure demonstrates the efficacy of the system in generalizing to unseen data types, utilizing encoder-decoder models trained with the EMNIST dataset for each combination of  $k$  and  $L$ . **a** Horizontal and **b** vertical grating inputs ( $24 \times 24$  pixels) are processed through each configuration. Encoded phase patterns and EDOF outputs, axially sampled at 0 mm, 50 mm ( $\sim 66.71\lambda$ ), 100 mm ( $\sim 133.43\lambda$ ), 150 mm ( $\sim 200.14\lambda$ ), 200 mm ( $\sim 266.85\lambda$ ) showcase continuous projection over a DOF of 200 mm ( $\sim 266.85\lambda$ ). The rightmost column presents the corresponding low-resolution images for each PSR factor,  $k$



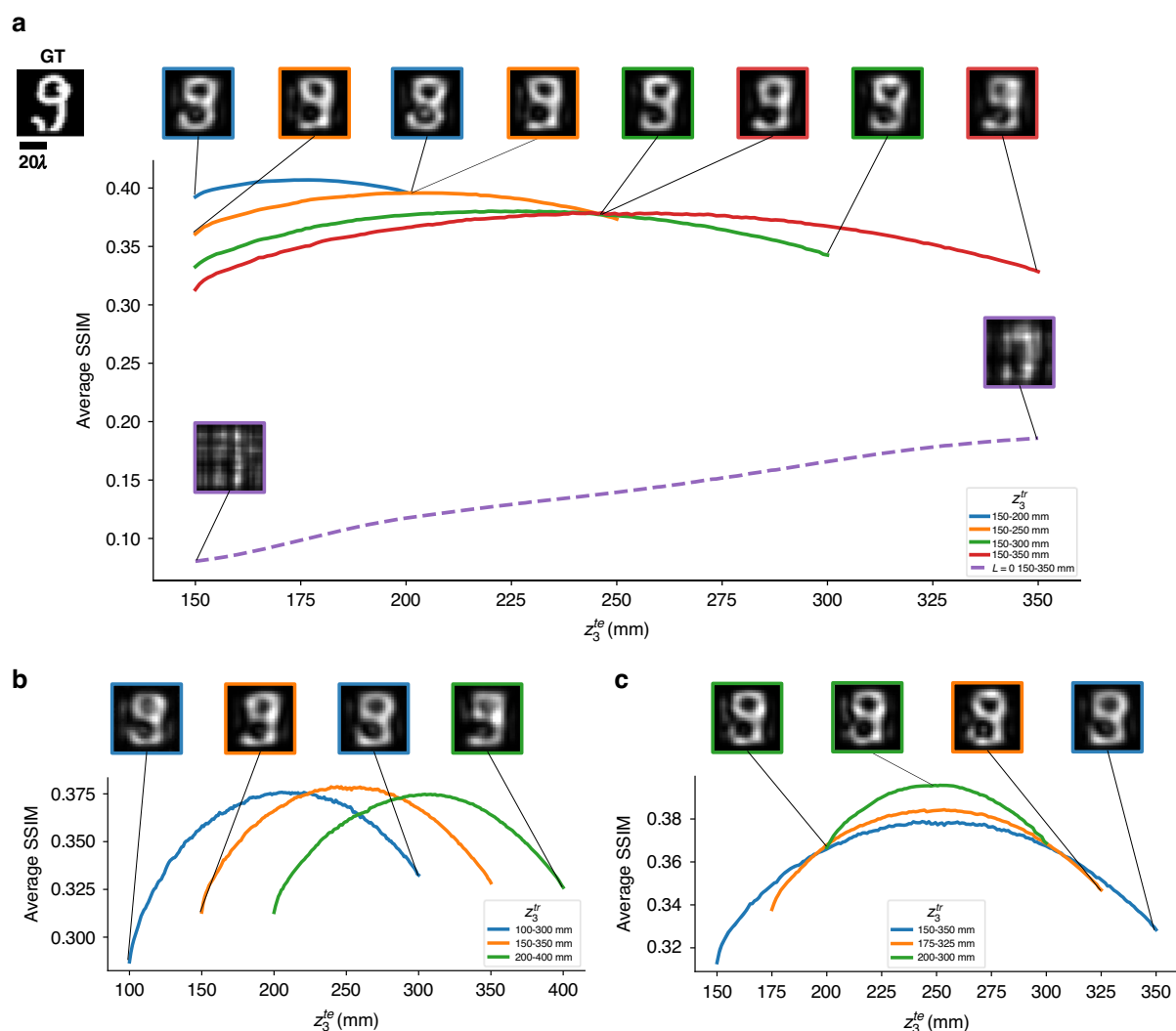


**Fig. 4** Quantitative analysis of the hybrid PSR image projection systems across varying  $z_3$  ranges. Each solid curve represents a hybrid system ( $L = 3, k = 6$ ) trained over a specific  $z_3^{tr}$  training range and evaluated over the corresponding  $z_3^{te}$  testing range, as indicated in the legend. The dashed purple curve depicts the performance of a projection system without a diffractive decoder (i.e.,  $L = 0$ ), trained and evaluated over an equivalent axial depth range. Representative reconstructed images at selected  $z_3^{te}$  positions are displayed above the corresponding curves. **a** Impact of increasing the depth range on the image projection fidelity, with all ranges ( $z_3^{te}$ ) starting from 150 mm ( $\sim 200.14\lambda$ ). **b** Performance comparison for depth ranges of a constant width  $z_e$  of 200 mm ( $\sim 266.85\lambda$ ) but varying central positions, elucidating the influence of range placement on the PSR image projection quality. **c** Average SSIM as a function of  $z_3^{te}$  for axial ranges sharing a common midpoint at 250 mm ( $\sim 333.56\lambda$ ) but differing in width,  $z_e$

Figure 4 further presents quantitative and visual analysis of our hybrid image projection approach across varying projection distances and depths. For this purpose, different three-layer designs ( $L = 3$ ) targeting a PSR factor of  $k = 6$  were trained for various  $z_3^{tr}$  ranges and tested using different  $z_3^{te}$  ranges. Here,  $k = 6$  was chosen as it strikes a balance between high compression and reconstruction fidelity, thereby providing a more stringent and representative evaluation of the hybrid diffractive decoder's performance. Figure 4a elucidates the impact of expanding the depth range  $z_e$  to 50 mm ( $\sim 66.71\lambda$ ), 100 mm ( $\sim 133.43\lambda$ ), 150 mm ( $\sim 200.14\lambda$ ), and 200 mm ( $\sim 266.85\lambda$ ), while maintaining a

fixed starting point  $z_{3,1} = 150$  mm ( $\sim 200.14\lambda$ ). As the DOF broadens, the peak average structural similarity index measure (SSIM) gradually declines, indicating a natural trade-off between the DOF and image projection fidelity. In contrast, an image projection system with a free-space-based decoder (i.e.,  $L = 0$ , trained and evaluated over an equivalent axial range), shown with the dashed purple curve in Fig. 4a, exhibits significantly degraded performance, once again highlighting the crucial role of the passive diffractive decoder in extending the DOF while maintaining high image projection fidelity. Similarly, Fig. 4b, c illustrate the influence of depth range placement and its axial width on



the image projection performance. While narrower DOF ranges yield higher SSIM values at the output projection volume, the system, in general, maintains high image quality and discernible features across all the tested DOF ranges reported in Fig. 4.

To quantitatively characterize the effective DOF enabled by the presented hybrid image projection architecture in the terahertz domain, we further evaluated our design using resolution test targets and image-quality metrics, including Michelson contrast and peak signal-to-noise ratio (PSNR). DOF is commonly defined based on the preservation of image contrast under defocus<sup>28,29</sup>. Accordingly, defining the DOF as the axial range over which an image contrast metric remains above a specified fraction of its peak value (e.g., contrast > 50%) is a widely adopted practical criterion in imaging and projection systems<sup>30,31</sup>. In this work, we adopt a more stringent threshold, requiring the Michelson contrast to remain above 90% of its peak value, which corresponds to minimal perceptual degradation and reliable preservation of fine spatial features. In addition, PSNR is used as a complementary metric to quantify overall image reconstruction fidelity. As shown in Supplementary Fig. S3a, these metrics were analyzed both within and beyond the training axial depth range, enabling a direct assessment of the system's external generalization capability along the axial direction. Consistent with the SSIM trends discussed above, the results indicate that image contrast and PSNR values remain stable across the trained DOF range and degrade gradually outside this range. Based on these quantitative metrics, using a PSNR threshold of 1 dB drop below the maximum and a Michelson contrast threshold of 90% of the peak value, the effective DOF was determined to span from  $z_3 = 170.5$  mm to  $z_3 = 357.7$  mm, spanning an axial width of  $\sim 250\lambda$ , within which the system consistently delivers high-contrast, high-fidelity image projections.

Complementing this quantitative analysis, Supplementary Fig. S3b shows representative projected images at multiple axial locations ( $z_3$ ), further demonstrating the preserved contrast and resolvable features across a wide span of image projection distances. This consistent visual performance corroborates the quantitative findings and highlights the decoder's ability to mitigate diffraction-induced degradation over extended depths.

In addition to these numerical analyses, our EDOF PSR image projection approach was also validated in proof-of-concept experiments spanning the terahertz and visible parts of the spectrum, demonstrating the scalability and generality of the hybrid system across different wavelengths. Specifically, for the experimental demonstration in the terahertz band ( $\lambda \approx 0.75$  mm), we designed a model with a single-layer diffractive decoder trained for  $z_3^{tr} = 135 - 185$  mm ( $180.12\lambda - 246.84\lambda$ ), corresponding to a

$z_e = 50$  mm ( $\sim 66.71\lambda$ ) and a PSR factor of  $k = 6$ . A single diffractive layer was chosen to minimize performance degradation from potential inter-layer misalignments, which can be an important error source in experimental implementations, particularly for cascaded diffractive systems. To further enhance robustness and practical feasibility of our design, an additional power efficiency-related loss term was used, and random displacements (see the Methods section for details) were applied to the diffractive decoder surface during the training phase. As depicted in Fig. 5, we used a THz detector and a continuous-wave THz illumination source in our experiments after the 3D fabrication of the single-layer diffractive surface and the LR encoded-phase profiles, all of which were fabricated using a 3D printer (see the "Methods" section). Figure 6 illustrates the experimental and the corresponding numerical results for different positions ( $z_3^{te}$ ) of the output plane, including 135 mm ( $\sim 180.12\lambda$ ), 160 mm ( $\sim 213.48\lambda$ ), and 185 mm ( $\sim 246.84\lambda$ ). Collectively, these results confirm the system's capability to deliver consistent, pixel super-resolved projections over an extended axial range in the terahertz domain.

To further validate the scalability and versatility of our hybrid framework, we conducted an additional proof-of-concept experiment in the visible range ( $\lambda = 635$  nm). As illustrated in Fig. 7a, this implementation employed a CNN-based electronic encoder with a super-resolution factor of  $k = 4$  and a single-layer passive diffractive decoder ( $L = 1$ ). In this visible-light setup, we employed two spatial light modulators (SLMs) to display the encoded phase patterns and the static diffractive decoder layer, respectively, axially separated by 17.5 mm (see the Methods section). While an SLM was used to implement the diffractive decoder in this proof-of-concept experiment to facilitate testing of different designs, it functioned strictly as a static phase mask, emulating a passive optical layer that can be fabricated using, e.g., lithography. Similar to the terahertz configuration, the system was jointly optimized to project super-resolved images across an axial range of  $z_3 \in [130, 160]$  mm. During the training phase, a random displacement strategy was applied to ensure robustness against potential misalignments (see the Methods section). Figure 7b presents the blind testing results for six randomly selected EMNIST characters, comparing numerical diffractive outputs with experimentally measured images across seven axial planes spanning the target volume. The experimental measurements exhibit good agreement with the simulations, maintaining high-fidelity image features over an extended DOF of 30 mm. These results, as shown in Figs. 6, 7, confirm the system's capability to synthesize super-resolved projections over an extended DOF significantly larger than the illumination wavelength. Given the



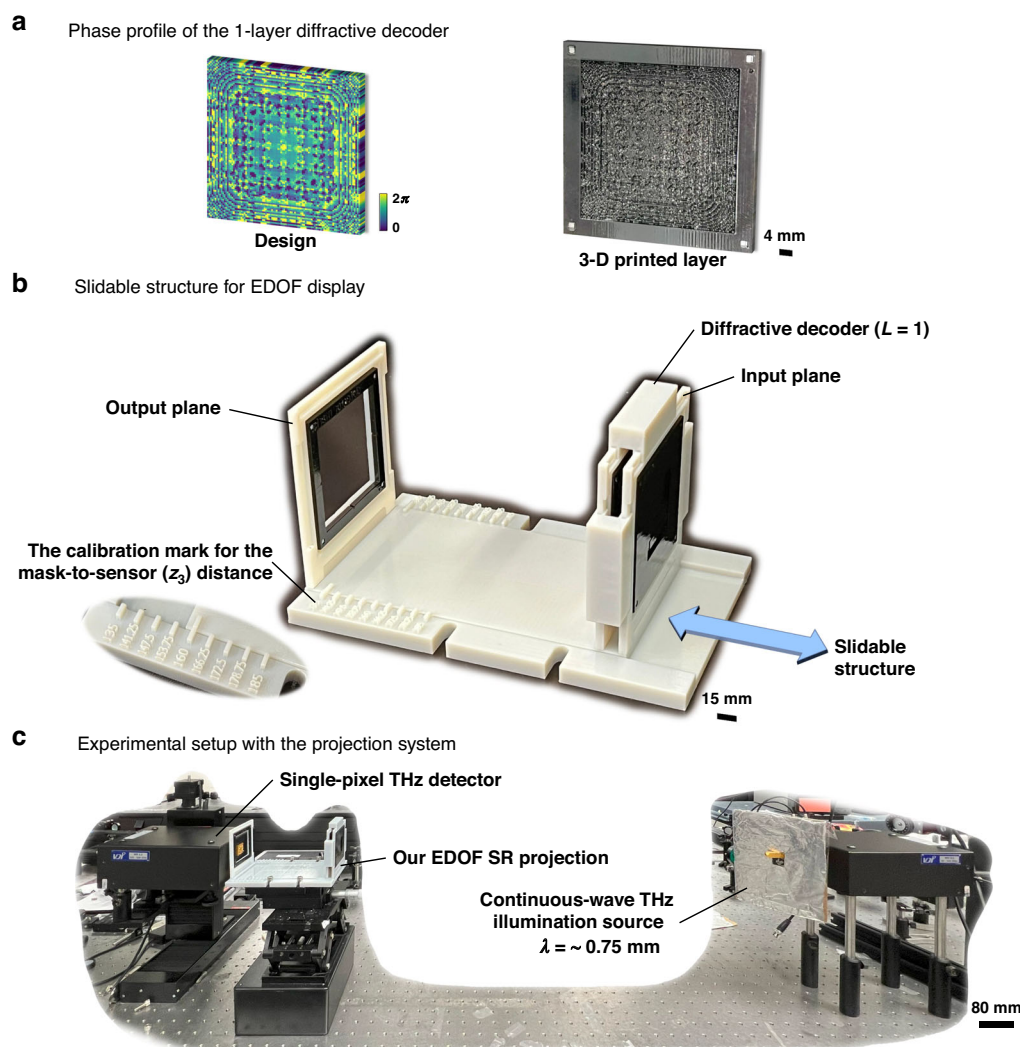
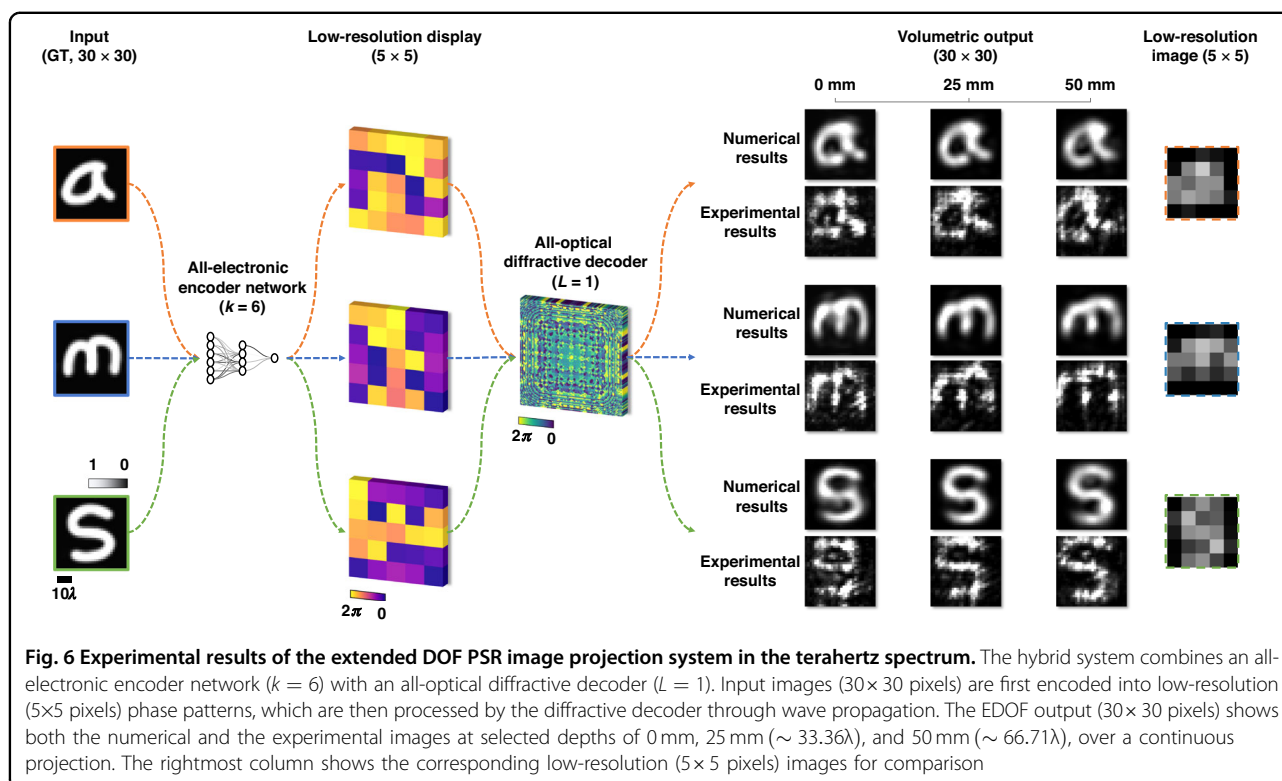


**Fig. 5** Experimental setup for EDOF PSR image projection system with  $L=1$ . **a** The phase profile of the designed diffractive decoder and its 3D fabricated version. **b** This 3D printed structure (white color) facilitates the relative axial translation of the diffractive decoder with respect to the terahertz detector at the output plane, enabling a variable layer-to-sensor distance to measure the resulting image projections at different axial positions. A calibration mark is integrated into the structure to ensure accurate measurement of the  $z_3$  distance. **c** Experimental setup employing a continuous wave terahertz illumination and a THz detector

structural simplicity of the single-layer diffractive decoder, the results further highlight the framework's efficiency in delivering substantial improvements in resolution and image projection depth with minimal optical hardware complexity.

Notably, our hybrid system is not tailored to a specific type of features or simple patterns to be projected, but functions as a general-purpose image projection platform. To demonstrate the versatility of our hybrid approach, extending beyond the testing on unseen EMNIST characters (Fig. 8a), we further numerically tested the external generalization capability of our hybrid system using image types/classes absent from training: resolution test targets (Fig. 8b) and Quickdraw

doodles (Fig. 8c). For this study, we optimized a five-layer ( $L = 5$ ) diffractive design based on visible illumination ( $\lambda = 635 \text{ nm}$ ) and compared it with the performance of a single-layer ( $L = 1$ ) diffractive design that was used in Fig. 7. Both models share an identical CNN encoder architecture and were trained end-to-end independently using exclusively the EMNIST dataset to achieve  $4\times$  PSR over an extended DOF of 130–160 mm. As shown in Fig. 8, both models demonstrate robust, consistent performance across diverse datasets. Quantitatively, the  $L = 5$  model achieved high fidelity on the internal generalization task (unseen EMNIST images), with an average test PSNR of  $23.89 \pm 2.72 \text{ dB}$  and an SSIM of  $0.793 \pm 0.105$ . Regarding



external generalization to new classes of test objects, evaluation on 5000 Quickdraw samples (50 randomly selected images from each of 100 test categories) yielded an average test PSNR of  $19.13 \pm 1.69$  dB and an SSIM of  $0.724 \pm 0.070$ . We observed that the  $L = 5$  configuration achieves lower background noise and enhanced contrast compared to the  $L = 1$  model, indicating that additional passive diffractive layers effectively increase the accessible optical degrees of freedom and further improve image projection quality across various types of test objects. To quantitatively substantiate the super-resolution capability of our designs, further analyses, reported in Supplementary Figs. S7 and S8, show the spatial-frequency spectra and the cross-sections of the projected images. The radially averaged power spectra demonstrate that both the  $L = 1$  and  $L = 5$  decoders successfully recover high-frequency components extending beyond the Nyquist sampling limit of the low-resolution input display. While the single-layer ( $L = 1$ ) projections successfully retrieve the spatial-frequency spectra (Supplementary Fig. S7), their limited optical degrees of freedom result in residual noise and localized distortions. In contrast, the  $L = 5$  architecture (Supplementary Fig. S8) refines this spectral recovery, yielding cleaner 2D spatial spectra and sharper spatial cross-sections without the visual artifacts of the  $L = 1$  setup. Collectively, these results confirm that the encoder–decoder architecture learns a broadly

applicable super-resolution mapping rather than memorizing dataset-specific features, thereby establishing our hybrid system as a generalized PSR image projection platform.

## Discussion

We presented a deep learning-enabled hybrid image projection system with enhanced DOF and resolution. This diffractive image projection system, integrating a CNN-based digital encoder and an all-optical passive decoder, reduces computational and hardware costs across various projection distances for a single wavefront modulator while also enhancing both the DOF and the spatial resolution. Moreover, this framework enhances data storage and transmission efficiency by encoding image information into compact phase patterns with significantly compressed dimensions. These patterns are subsequently processed by an all-optical decoder, resulting in the projection of high-resolution images with extended DOF. Specifically, the projected content can be dynamically updated via a compact display engine at the input encoded phase plane, such as an SLM, while the all-optical decoder part can be fabricated using passive optical materials. Notably, the overall image projection performance is not constrained by the nominal pixel count or the native SBP of the SLM itself. Instead, by introducing additional diffractive layers at the optical decoder, our hybrid system effectively



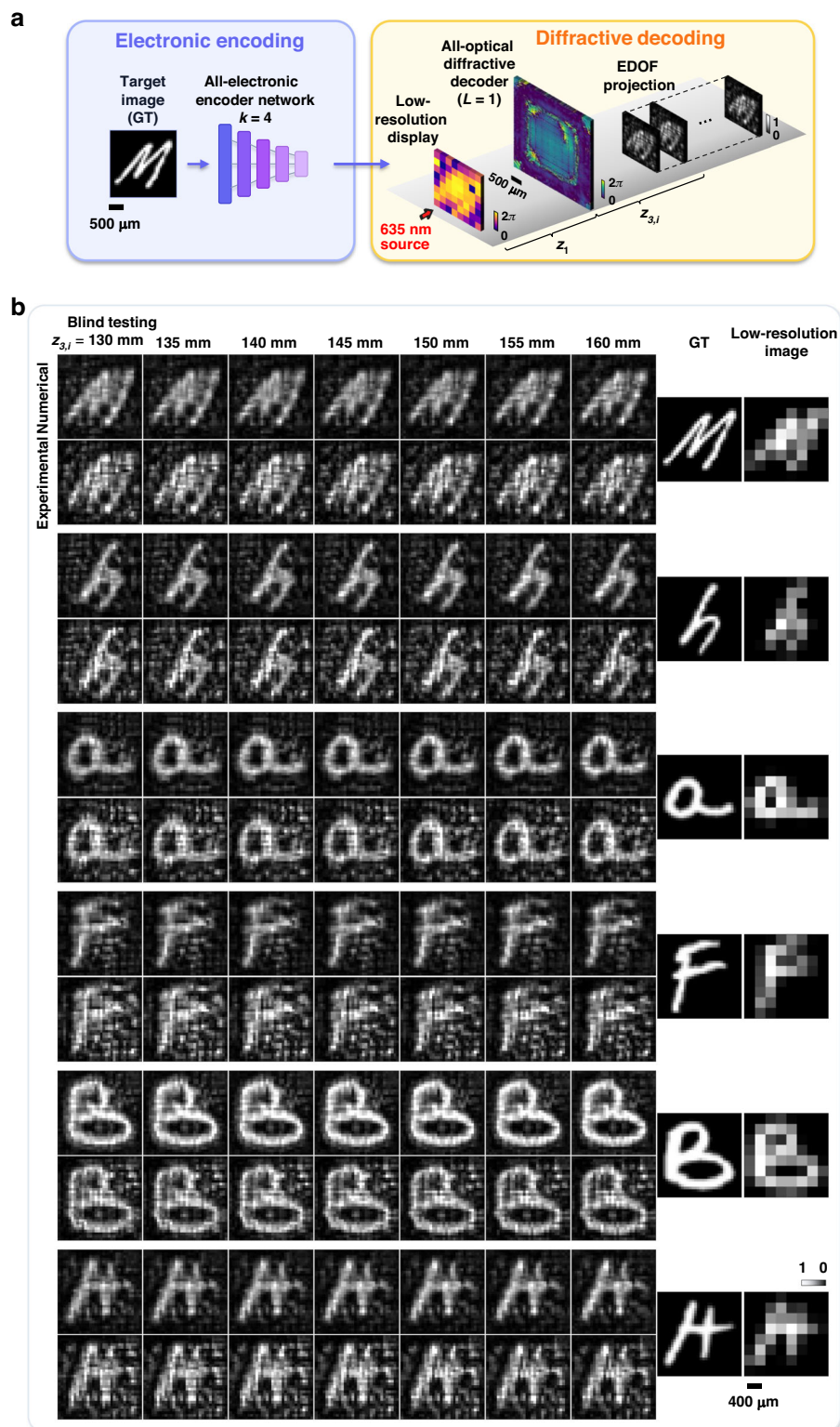


**Fig. 7** Experimental results of extended DOF PSR image projection in the visible spectrum. **a** Schematic of the hybrid system operating at 635 nm, comprising a CNN-based electronic encoder ( $k = 4$ ) and a single-layer ( $L = 1$ ) passive diffractive decoder. **b** Blind testing results for EMNIST characters never seen during the training. For each character, numerical (upper) and experimental (lower) output images are shown at seven axial planes spanning  $z_{3,l} = 130 - 160 \text{ mm}$ , demonstrating an extended DOF of  $\sim 30 \text{ mm}$ . Ground truth (GT) images and the corresponding low-resolution inputs are displayed on the right side for comparison



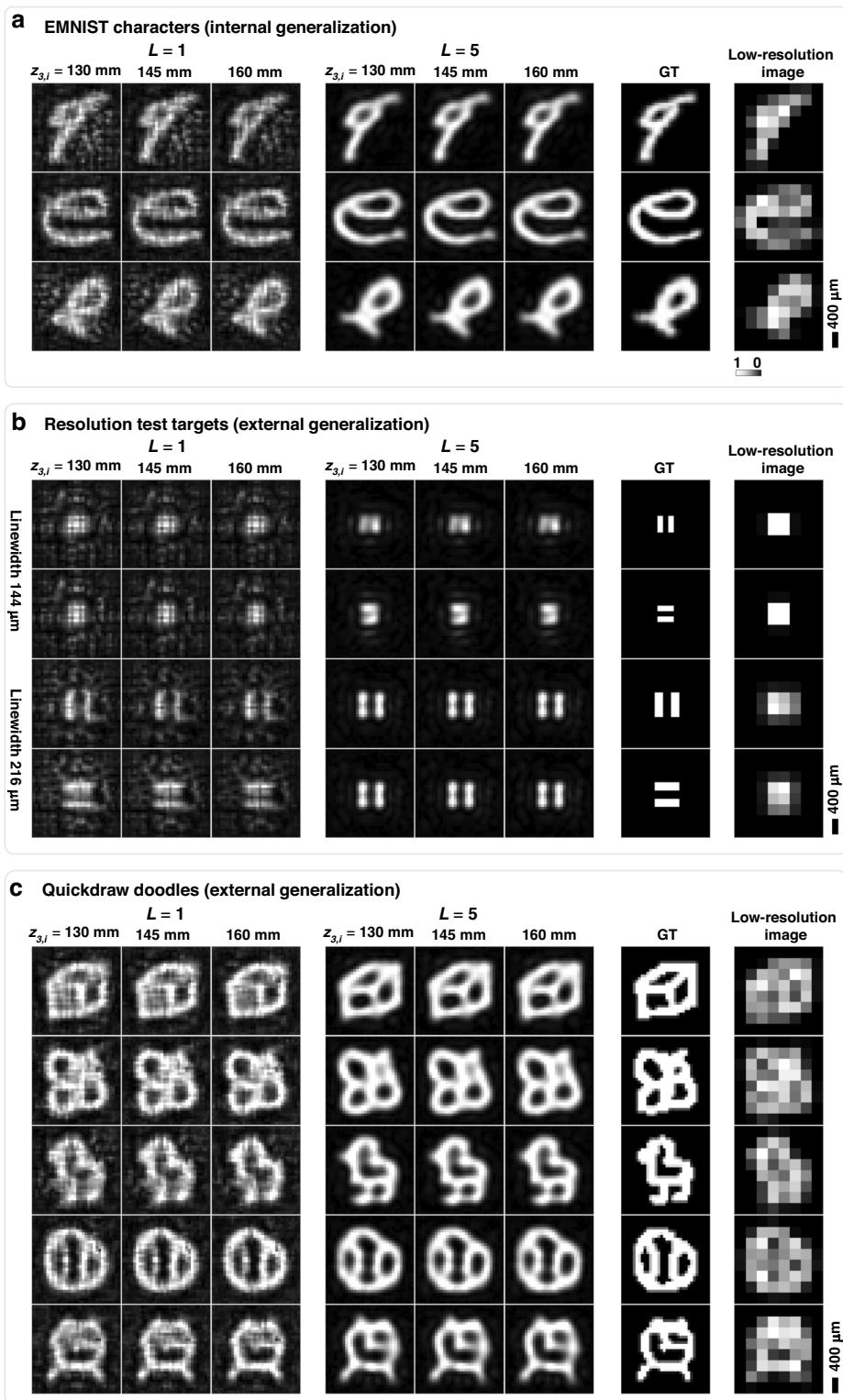


**Fig. 8 Generalization performance analysis of the extended DOF PSR image projection system in the visible spectrum.** Numerical simulations of PSR ( $k = 4$ ) image projection operating at 635 nm, comparing single-layer ( $L = 1$ ) and multi-layer ( $L = 5$ ) diffractive decoders over an extended DOF ( $z_{3,i} = 130 - 160 \text{ mm}$ ). Both models were trained independently using exclusively the EMNIST dataset. **a** Internal generalization results on unseen EMNIST characters. **b** External generalization on resolution test targets. **c** External generalization on Quickdraw doodles. Ground truth (GT) images and the corresponding low-resolution inputs are displayed on the right side for comparison

surpasses the native SBP of the input display. This architecture enables the synthesis of high-fidelity projections with enhanced spatial resolution over an extended DOF.

One key strength of the proposed diffractive PSR framework lies in its inherent scalability, enabled by the passive nature of the optical decoder. The effective output resolution and the corresponding SBP can be systematically increased within the diffraction limit of light by scaling the number of diffractive features or layers, with or without increasing the number of pixels on the encoded pattern. As shown in the analysis in Fig. 8, increasing the number of diffractive features or layers consistently improves the output image fidelity. Accordingly, the number of diffractive layers or features can be tailored to the target image scales and/or performance requirements. Beyond a single device, the proposed framework naturally supports massive parallelization. Arrays of diffractive processors can be fabricated and operated concurrently, potentially enabling hundreds of passive optical decoders to function in parallel. Each diffractive unit can be designed to handle a distinct spatial sub-aperture, depth segment, or projection task, thereby further extending the effective SBP and throughput of the overall system.

Our hybrid approach also synergizes seamlessly with recent advances in deep learning-based CGH for large-scale display and image projection systems. CNN-based hologram generation methods have demonstrated the ability to map target images directly to phase-only holograms in a fast, non-iterative manner<sup>12,14,32,33</sup>, significantly reducing computational overhead compared to traditional CGH algorithms. Importantly, state-of-the-art CNN-based CGH models can be readily integrated into our architecture as digital encoders, enabling learned hologram synthesis to directly benefit from the physical SBP and DOF expansion provided by a jointly trained diffractive decoder. Due to computational resource limitations, the experimental results presented in this work use reduced-scale datasets and system sizes to demonstrate a proof of concept. Nevertheless, the underlying framework and training methodology are inherently scalable and can be designed for larger resolutions, deeper projection volumes, and higher SBP regimes.

Furthermore, precise alignment of the diffractive layers of the decoder can pose additional challenges for operation at shorter wavelengths. Some of these challenges can be mitigated by optimizing misalignment-resilient designs, implemented through a numerical “vaccination” design strategy<sup>34,35</sup>, by introducing potential errors and variations in, e.g., size/shape/orientation of the diffractive layers during the training process. The same approach can also be applied to reduce sensitivity to potential fabrication errors in the phase modulation features of the resulting diffractive layers. To further assess the feasibility

of fabricating the designed multi-layer architectures, we investigated the resilience of our hybrid system against mechanical misalignments between the diffractive layers and the input plane; see Supplementary Figs. S4 and S5. Based on a three-layer decoder configuration ( $L = 3$ ) in Fig. 2, which was designed for a PSR factor of  $k = 6$  trained for an extended DOF of  $z_e = 200$  mm, we evaluated the image projection fidelity under varying degrees of positional errors. Additionally, we employed a vaccination strategy, introducing random lateral ( $D_x \sim U(-\Delta_{xy,tr}, \Delta_{xy,tr})$ ,  $D_y \sim U(-\Delta_{xy,tr}, \Delta_{xy,tr})$ ) and axial ( $D_z \sim U(-\Delta_{z,tr}, \Delta_{z,tr})$ ) displacements to the diffractive layers during the training phase (see the Methods section for details). As shown in Supplementary Fig. S4a, the unvaccinated baseline diffractive model ( $\Delta_{xy,tr} = 0$ ) exhibits a decline in the output PSNR values from  $\sim 15.6$  dB to  $\sim 11.5$  dB as the testing misalignment increases, resulting in image degradation. Conversely, the models trained with vaccinated lateral shifts ( $\Delta_{xy,tr} = 0.2\lambda$  and  $0.4\lambda$ ) demonstrate remarkable robustness, maintaining PSNR values above 15 dB even under significant lateral displacements up to  $\sim \pm 0.5\lambda$ . As visualized in Supplementary Fig. S4b, when  $\Delta_{xy,test} = 0.6\lambda$ , the output image PSNR value for the vaccinated design ( $\Delta_{xy,tr} = 0.4\lambda$ ) remains at 15.29 dB, while the PSNR value for the baseline design ( $\Delta_{xy,tr} = 0$ ) drops to 12.32 dB. A similar trend is also observed for axial misalignments as shown in Supplementary Fig. S5, where the models vaccinated with  $\Delta_{z,tr} = 0.4\lambda$  and  $0.8\lambda$  effectively mitigate the performance degradation associated with inter-layer spacing errors, reducing the output image PSNR degradation to  $\sim 0.1$  dB. Crucially, this enhanced robustness incurs only a negligible penalty ( $\sim 0.09$  dB) on the peak image quality under ideal alignment conditions. These findings confirm that incorporating stochastic 3D positional noise during training effectively desensitizes the diffractive decoder design to uncontrolled mechanical variations, thereby relaxing stringent fabrication tolerances and facilitating robust experimental deployment of the EDOF PSR system.

In addition to alignment imperfections, fabrication constraints introduce another practical challenge, which is particularly critical for visible-wavelength applications employing lithography-based manufacturing. Specifically, current fabrication technologies typically limit the achievable axial thickness modulation of each diffractive layer, resulting in finite phase quantization per diffractive feature, which can lead to performance degradation relative to the ideal numerical designs. To systematically assess the robustness of our hybrid image projection framework under a limited phase bit-depth, we evaluated the system performance using phase bit depth values selected from  $B \in \{16, 8, 6, 5, 4, 3, 2\}$ . As shown in Supplementary Fig. S6a, when a diffractive model is trained using a large phase bit-depth (e.g.,  $B_{train} = 16$ ), the image



projection performance gradually degrades as the phase quantization level is reduced during the testing ( $B_{test} \in \{16, 8, 6, 5, 4, 3, 2\}$ ). For instance, the average output image PSNR decreases from  $\sim 15.8$  dB at  $B_{test} = 16$  to  $\sim 14.4$  dB at  $B_{test} = 3$ , reflecting the increasing phase discretization error. Importantly, this degradation can be effectively mitigated by incorporating the phase quantization constraint directly into the training process. When the diffractive processor is trained and optimized with a reduced phase resolution (e.g.,  $B_{train} = 3$ ) and evaluated using the same bit depth ( $B_{test} = B_{train} = 3$ ), the average output PSNR improves to  $\sim 15.6$  dB, approaching the performance of higher phase bit-depth designs. The corresponding visual reconstructions shown in Supplementary Fig. S6b further demonstrate consistent image quality, preserved contrast, and stable feature resolution across the entire image-projection depth range, despite severe phase quantization. Together, these results highlight the robustness of our hybrid diffractive framework, confirming that high-fidelity and large-DOF image projection can be achieved even under stringent fabrication constraints by co-designing the training process with realistic hardware limitations in mind.

Further advances in this technology promise to unlock new paradigms across diverse fields ranging from display technologies to advanced scientific instrumentation and optical metrology. For example, deep learning-based metrology methods have recently attracted considerable attention<sup>36</sup>, owing to their advantages over traditional model-based approaches. Our diffractive image projection system, combining a CNN-based encoder with an all-optical decoder, might potentially offer new types of solutions to optical metrology tasks.

Notably, the presented extended DOF PSR image projection approach can operate across different parts of the electromagnetic spectrum. The designed models can be scaled to any part of the spectrum, including the visible and infrared (IR) wavelengths, by physically scaling the dimensions of the spatial features with respect to the wavelength of light. Crucially, recent advances in nanofabrication techniques provide a practical pathway toward realizing large-scale diffractive decoders for visible display applications. State-of-the-art fabrication techniques, including advanced optical lithography and electron-beam lithography<sup>37–40</sup>, enable lateral feature sizes on the order of a few hundred nanometers, axial step sizes below 100 nanometers, and precise multi-level phase modulation, all readily applicable to visible holographic display applications. Furthermore, wafer-level processes have been demonstrated to support hundreds of millions to billions of nanoscale diffractive features per wafer<sup>40</sup>, with multi-layer alignment accuracy sufficient for complex optical processing tasks. These capabilities enable the fabrication of large-aperture, multi-layer diffractive

decoders with high uniformity, repeatability, and compatibility with mass production. Once integrated with these manufacturing capabilities, our hybrid electronic-optical approach can establish a compelling route toward high-resolution image projection and display platforms that surpass the SBP and power-efficiency limits of conventional holographic displays.

In summary, our EDOF PSR image projection system, enabled by the symbiosis of deep learning with diffractive optical processors, represents a significant advance toward versatile and high-performance light shaping, with far-reaching implications for fields that demand precise lateral and depth control of light, including microscopy and metrology.

## Materials and methods

### Digital encoder design

Our digital encoder, depicted in Fig. 1a, consists of a series of convolutional layers followed by a fully connected layer. It accepts a  $24 \times 24$  grayscale image as input. Each convolutional layer applies its filters to the input, followed by normalization and a PReLU<sup>25</sup> activation. Max-pooling operations between the layers progressively reduce the spatial dimensions while increasing the feature depth. As the input progresses through the network, the spatial dimensions are reduced while the number of channels increases. The final convolutional layer yields a  $3 \times 3 \times 128$  feature map, which is then flattened and processed by a fully connected layer with sigmoid activation, producing a compact encoded representation ( $\frac{24}{k} \times \frac{24}{k}$  pixels, where a PSR factor of  $k = 6$  is used in Fig. 1a). This encoding operation enables the digital encoder to efficiently compress the essential information of the input image into a phase-only format optimized for subsequent optical decoding.

### Optical decoder design

Our diffractive decoder employs passive diffractive layers,  $L \geq 1$ , and each layer  $l$  is then subdivided into a grid of diffractive features<sup>22,41,42,43</sup>. These diffractive features, characterized by their transmittance coefficients  $t_l[m, n]$ , serve as the fundamental building blocks of the all-optical decoder. The transmittance of an individual diffractive feature is defined as follows:

$$t_l[m, n] = \exp\left(j \frac{2\pi}{\lambda} (\tau(\lambda) - n_a) h_l[m, n]\right) \quad (1)$$

In this expression, the wavelength-dependent complex refractive index  $\tau(\lambda) = n(\lambda) + j\kappa(\lambda)$  and the refractive index of the surrounding medium  $n_a$  determine the phase modulation imparted by each diffractive feature. A key parameter in the modulation process is the physical thickness of each diffractive feature, i.e.,  $h_l[m, n]$ , which is



computed using an auxiliary variable  $o_l[m, n]$ :

$$h_l[m, n] = \frac{\tanh(o_l[m, n]) + 1}{2} (h_m - h_b) + h_b \quad (2)$$

This formulation allows for a continuous spectrum of thickness values, bounded by  $h_b$  and  $h_m$ , which can be tuned during the optimization process to achieve desired optical properties. These diffractive features give rise to a transmission modulation function  $T_l(x, y)$ , described by:

$$T_l(x, y) = \sum_m \sum_n t_l[m, n] p_l(x - mW_x, y - nW_y) \quad (3)$$

where the 2D rectangular sampling kernel  $p_l(x, y)$  is defined as:

$$p_l(x, y) = \begin{cases} 1, & |x| < \frac{W_x}{2}, |y| < \frac{W_y}{2} \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

To model the propagation of spatially and temporally coherent light propagating through this system, the framework employs the angular spectrum method<sup>42</sup>, which is based on the fast Fourier transform-based implementation of the Rayleigh-Sommerfeld diffraction integral. This propagation is mathematically expressed as:

$$U_{l+1}(x, y) = U'_l(x, y) * w(x, y, z_{l+1} - z_l) \quad (5)$$

where  $U'_l(x, y) = U_l(x, y)T_l(x, y)$  is the output field of layer  $l$ , and the term  $z_{l+1} - z_l$  represents the axial propagation distance between the diffractive layer  $l$  and the next diffractive layer  $l + 1$ . The propagation kernel  $w(x, y, z)$  is given by:

$$w(x, y, z) = \frac{z}{r^2} \left( \frac{1}{2\pi r} + \frac{1}{j\lambda} \right) \exp\left(j \frac{2\pi r}{\lambda}\right) \quad (6)$$

with  $r = \sqrt{x^2 + y^2 + z^2}$ .

### Datasets and image preprocessing

The EMNIST handwritten letters dataset, consisting of 88,800 training and 14,800 testing samples, was used to train our numerical designs. We allocated 8800 samples from the training set for validation, yielding 80,000 training, 8800 validation, and 14,800 testing samples. Handwritten letter images ( $28 \times 28$  pixels) from this dataset were resized to  $24 \times 24$  pixels for terahertz experiments and  $32 \times 32$  pixels for visible-light experiments using bicubic downsampling. For the experimental phase, we utilized the same dataset to train the hybrid image projection system. To evaluate its performance, we

tested the model using handwritten letter images that are different from those in the training dataset.

For external generalization testing, we utilized the 'Quick, Draw!' dataset<sup>44</sup>, selecting 5000 images (50 randomly chosen samples from 100 different categories). These images were resized to  $32 \times 32$  pixels and subjected to a max-pooling operation with a  $3 \times 3$  kernel during preprocessing. This dataset was used solely for testing and was not included in any part of the model's training process.

### Optical decoder parameters

Our numerical designs in the terahertz spectrum incorporated diffractive layers spanning  $133.42\lambda \times 133.42\lambda$ , each discretized into a  $200 \times 200$  element grid. The size of diffractive features and the sampling period of the propagation model were set to  $0.667\lambda$ . The input and output fields of view of the decoder are  $96 \times 96$  pixels ( $64.04\lambda \times 64.04\lambda$ ), and aliasing was prevented through zero padding to  $400 \times 400$  pixels. The transmittance coefficient of each diffractive element was expressed as:

$$t_l[m, n] = \exp(j\theta_l[m, n]) = \exp\left(j \frac{2\pi}{\lambda} (n(\lambda) - n_a) h_l[m, n]\right) \quad (7)$$

Here, the phase coefficients  $\theta_l[m, n]$  were initialized to zero and subsequently refined through stochastic gradient-based error backpropagation<sup>45</sup>.

We explored different decoder configurations with  $L=1$ , 3 and 5 transmissive layers; for comparison, we also trained  $L=0$  designs without a diffractive decoder layer. By utilizing a grid search method, we determined the optimal distance from the input plane to the first diffractive layer ( $z_1$ ), the inter-layer spacing between the diffractive layers ( $z_2$ ), and the distance from the final layer to the first output plane ( $z_{3,1}$ ), as illustrated in Fig. 1b and Table 1. The final configuration was determined by maximizing the average peak signal-to-noise ratio (PSNR) across the validation dataset, ensuring optimized and robust performance under varied input conditions. By empirically optimizing both the axial placement and the number of diffractive layers, the presented hybrid diffractive system effectively exploits the available optical degrees of freedom, leading to improved image projection performance.

In our experimental results in the terahertz spectrum, a monochromatic terahertz source emitting at  $\lambda \approx 0.75$  mm was used. Our experiments focused on a single-layer ( $L=1$ ) diffractive decoder, with specific axial distances also detailed in Table 1.

The diffractive layers were constructed using diffractive features, each with dimensions of  $0.667\lambda$ , arranged in a  $66.7\lambda \times 66.7\lambda$  grid ( $50$  mm  $\times$   $50$  mm). Similarly, we

**Table 1** The axial distances used for the numerical and experimental evaluations of the diffractive decoders reported in this work

| Distance     | $z_1$                 | $z_2$                 | $z_{3,1}$               | $z_e$                   |
|--------------|-----------------------|-----------------------|-------------------------|-------------------------|
| Numerical    | 10 mm ( $13\lambda$ ) | 50 mm ( $66\lambda$ ) | 150 mm ( $200\lambda$ ) | 200 mm ( $267\lambda$ ) |
| Experimental | 10 mm ( $13\lambda$ ) | N/A                   | 155 mm ( $207\lambda$ ) | 50 mm ( $66\lambda$ )   |

established an effective pixel size of  $1.33\lambda$  at the measurement plane, while both the phase-only wavefront modulator and the output field of view were designed as  $40\lambda \times 40\lambda$  ( $30\text{ mm} \times 30\text{ mm}$ ). The LR wavefront modulator's pixel pitch was set to  $8\lambda \times 8\lambda$ , yielding a PSR factor of 6.

The diffractive components were fabricated using a 3D-printing material characterized by a complex refractive index of  $\tau(\lambda) \approx 1.6518 + j0.0612$  at the operational wavelength,  $\lambda \approx 0.75\text{ mm}$ . The thickness of each diffractive feature, denoted as  $h_l[m, n]$ , was optimized within the range of  $[0.5\text{ mm}, 1.644\text{ mm}]$ , corresponding to a phase modulation span of  $[-\pi, \pi]$ .

For the visible PSR EDOF projection design reported in Fig. 7a, we utilized phase-only SLMs and treated the diffractive component, including the encoded phase profile and the diffractive layer, as phase modulators, with a reflection coefficient of:

$$r(x_s, y_s, z_l; \lambda) = \exp(-j\phi(x_s, y_s; \lambda)) \quad (8)$$

where  $\phi(x_s, y_s; \lambda)$  is the phase modulation value in the range of  $[-\pi, \pi]$  for the corresponding diffractive feature. The diffractive features have dimensions of  $4.5\text{ }\mu\text{m}$ , arranged in an  $800 \times 800$  pixel grid. Both the phase-only wavefront modulator and the output field of view were designed as  $3628\lambda \times 3628\lambda$  ( $2.304 \times 2.304\text{ mm}$ ). The LR wavefront modulator's pixel pitch was set to  $441\lambda \times 441\lambda$  ( $0.288 \times 0.288\text{ mm}$ ), yielding a PSR factor of  $k = 4$ .

To enhance the robustness of our system, we implemented a vaccination process<sup>34</sup> during the training. This involved introducing random alignment errors to the positions of diffractive layers and the input patterns during the training phase. The random errors are represented by  $D^l = (D_x, D_y, D_z)$ , indicating deviations of the diffractive layer  $l$  from its ideal position. Here,  $D_x$ ,  $D_y$ , and  $D_z$  independently follow a uniform distribution, such that  $D_x \sim U(-\Delta_x, \Delta_x)$ ,  $D_y \sim U(-\Delta_y, \Delta_y)$ , and  $D_z \sim U(-\Delta_z, \Delta_z)$ , where the parameters  $\Delta_x$ ,  $\Delta_y$ , and  $\Delta_z$  denote the hyperparameters of the possible alignment errors along the x-, y-, and z-axes, respectively. Consequently, the position of the diffractive layer  $l$  at the  $i^{\text{th}}$  iteration,

denoted as  $L^{(l,i)}$ , is expressed as:

$$L^{(l,i)} = (L_x^l, L_y^l, L_z^l) + (D_x^{(l,i)}, D_y^{(l,i)}, D_z^{(l,i)}) \quad (9)$$

The alignment error ranges for our terahertz experiments were set to  $0.67\lambda$  for both  $\Delta x$  and  $\Delta y$ , and  $0.534\lambda$  for  $\Delta z$ . For experiments in the visible spectrum, we set  $\Delta z = 0$  and  $\Delta x = \Delta y = 0.75\lambda$ .

Following the decoder optimization, we converted the thickness maps of the diffractive layers and the corresponding phase-encoded representations into STL files using MATLAB. These designs were then physically fabricated using an Objet30 Pro 3D printer (Stratasys). To precisely adjust the distance between the decoder and the fixed terahertz wave detector at the output plane, a slidable structure, illustrated in Fig. 5b, was designed and 3D-printed. This configuration enables output image measurements over an extended DOF by controlling the distance from the diffractive layer to the output plane ( $z_{3,i}$ ).

### Experimental setup

The experimental apparatus is depicted in Fig. 5c. At its core, a Virginia Diodes Inc. WR2.2 modular amplifier/multiplier chain (AMC) serves as the radiation source. This device transforms an  $11.111\text{ GHz}$  ( $f_{\text{RF1}}$ ) input, amplified to  $10\text{ dBm}$ , into a  $0.4\text{ THz}$  continuous wave through a 36-fold frequency multiplication process. To enhance detection sensitivity, the output undergoes 1-kHz amplitude modulation. Terahertz waves emanate from a diagonal horn antenna, traversing  $60\text{ cm}$  before encountering the target object and, subsequently, a 3D-printed diffractive decoder. A Virginia Diodes Inc. single-pixel Mixer/AMC, operating with an  $11.083\text{ GHz}$  ( $f_{\text{RF2}}$ ) local oscillator at  $10\text{ dBm}$ , captures and down-converts the diffracted field to a  $1\text{ GHz}$  intermediate frequency. A detector ( $0.5\text{ mm} \times 0.25\text{ mm}$ ) is mounted on Thorlabs NRT100 motorized linear stages, enabling precise X-Y positioning. It scanned the field of view in  $0.75\text{ mm}$  steps, and subsequently,  $3 \times 3$  bilinear upsampling and  $4 \times 4$ -pixel binning were employed to increase the signal-to-noise ratio and approximate the system's  $\sim 1.33\lambda$  ( $1\text{ mm}$ ) output pixel size. Mini-Circuits ZRL-1150-LN+ amplifiers provide  $40\text{ dB}$  gain, followed by a KL Electronics 3C40-1000/T10-O/O bandpass filter ( $1\text{ GHz} \pm 10\text{ MHz}$ ).



An HP 8495B attenuator allows for system calibration before the signal reaches a Mini-Circuits ZX47-60 power detector. Final processing occurs in a Stanford Research SR830 lock-in amplifier synchronized to the 1 kHz modulation. The resulting data undergoes linear scaling and dynamic range adjustment, where the extreme 5% of values at both ends are saturated, with the remainder mapped to a 0-1 range.

For the PSR EDOF projection experiment shown in Fig. 7a, a laser source (Fianium Ltd, Southampton, UK) provided the illumination at 635 nm. Phase patterns were displayed onto two SLMs that served as the encoded phase pattern and the pre-trained diffractive layer: a HOLOEYE PLUTO-2.1 device (8  $\mu\text{m}$  pixel pitch,  $1920 \times 1080$ ) and a HOLOEYE LUNA device (4.5  $\mu\text{m}$  pixel pitch,  $1920 \times 1080$ ), respectively. A complementary metal-oxide-semiconductor (CMOS) camera (Basler ace acA1920-40gm) was used for capturing the output images. Light propagated 17.5 mm between the two SLMs and covered an output DOF of  $\sim 30$  mm.

#### Loss function and DOF-aware optimization strategy

The optimization of our EDOF PSR image projection system involves the joint training of an electronic CNN-based encoder and an optical diffractive decoder. We formulated a comprehensive loss function that balances the image projection accuracy with the output diffraction efficiency:

$$\mathcal{L}_{\text{sample}} = \frac{1}{N} \sum_{j=1}^N \left( y_j - \sigma \hat{y}_{j,z_{3,i}} \right)^2 + \alpha e^{-\eta} \quad (10)$$

This loss function comprises two key components: the mean squared error (MSE) between the target high-resolution image ( $y_j$ ) and the normalized decoder output at a randomly chosen  $z_{3,i}$  distance ( $\sigma \hat{y}_{j,z_{3,i}}$ ), and a diffraction efficiency-related penalty term ( $e^{-\eta}$ ). Here,  $N$  denotes the number of pixels in each image,  $\sigma = \frac{\sum_{j=1}^N y_j}{\sum_{j=1}^N \hat{y}_{j,z_{3,i}}}$  represents a normalization factor, and  $\alpha$  is a weighting coefficient for the diffraction efficiency-related term. During the numerical validations, we set  $\alpha$  to 0, and in our experimental designs, we set  $\alpha$  to 0.01 to achieve  $\sim 2\%$  output diffraction efficiency. The diffraction efficiency term,  $\eta$ , is defined as  $100 \times \frac{P_{\text{out}}}{P_{\text{in}}}$ , where  $P_{\text{in}}$  and  $P_{\text{out}}$  represent the input and output optical powers, respectively, and it enables fine-tuning of the diffractive decoder's power efficiency through Eq. (10). For each training iteration, the total loss  $L_{\text{batch}}$  is calculated as the mean of all  $L_{\text{sample}}$  values within that batch.

To assess the quality of our projected output images, we employed two widely recognized metrics: SSIM<sup>46</sup> and

PSNR. SSIM evaluates the similarity between the reconstructed and the ground truth images using:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (11)$$

where  $\mu_x$  and  $\mu_y$  are the mean values of image  $x$  and  $y$ ,  $\sigma_x^2$  and  $\sigma_y^2$  are the variances of image  $x$  and  $y$ ,  $\sigma_{xy}$  is the covariance of  $x$  and  $y$ , respectively, and  $c_1$  and  $c_2$  are constants for numerical stability. PSNR quantifies the ratio of the maximum signal power to noise power:

$$\text{PSNR} = 10 \log_{10} \left( \frac{\text{MAX}_I^2}{\text{MSE}} \right) \quad (12)$$

with  $\text{MAX}_I$  denoting the maximum possible pixel intensity.

Additionally, to evaluate the modulation depth of the projected patterns, we utilized the Michelson Contrast metric. Prior to the metric calculation, the projected image intensity  $I$  is normalized using min-max normalization. The contrast is then computed based on the regional statistics defined by a binary target (ground truth) mask  $T$ :

$$\text{MC} = \frac{\mu_w - \mu_b}{\mu_w + \mu_b + \epsilon} \quad (13)$$

where  $\mu_w$  and  $\mu_b$  represent the mean pixel intensities of the bright regions (where the target mask  $T = 1$ ) and the dark regions (where  $T = 0$ ), respectively, and  $\epsilon$  is a small constant introduced to ensure numerical stability.

We implemented our training framework using Python 3.9.13 and TensorFlow 2.10, leveraging the Adam optimizer with a Cosine Annealing learning schedule incorporating a warm-up phase. The optimization process utilized a maximum learning rate of  $1 \times 10^{-3}$ , with 25 warm-up epochs and a total of 400 training epochs. All networks were trained on an NVIDIA GeForce RTX 3090 GPU, with each design requiring approximately 28 hours of training time using a batch size of 200.

#### Acknowledgements

The authors acknowledge the support of the US NSF (Grant # 2401393).

#### Author details

<sup>1</sup>Electrical and Computer Engineering Department, University of California, Los Angeles, CA, USA. <sup>2</sup>Bioengineering Department, University of California, Los Angeles, CA, USA. <sup>3</sup>California NanoSystems Institute (CNSI), University of California, Los Angeles, CA, USA

#### Author contributions

H.C., C.I., and C.S. conducted numerical simulations. H.C. and C.I. designed and built the experimental setup, and H.C., C.S., S.C., and T.G. carried out the experiments. H.C., C.S., C.I., M.J., and A.O. contributed to the preparation of the manuscript. A.O. supervised the research.



**Data availability**

All the data and methods needed to evaluate the conclusions of this work are presented in the main text. Additional data can be requested from the corresponding author (A.O.). The codes used in this work use standard libraries and scripts that are publicly available in TensorFlow.

**Conflict of interest**

A.O. serves as an Editor for the Journal; no other author has reported any competing interests.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41377-026-02320-7>.

Received: 5 October 2025 Revised: 11 April 2026 Accepted: 15 April 2026  
Published online: 18 May 2026

**References**

- Xiong, J. H. et al. Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light Sci. Appl.* **10**, 216 (2021).
- Venkatesan, M. et al. Virtual and augmented reality for biomedical applications. *Cell Rep. Med.* **2**, 100348 (2021).
- Kooi, F. L. & Toet, A. Visual comfort of binocular and 3D displays. *Displays* **25**, 99–108 (2004).
- Blanche, P. A. Holography, and the future of 3D display. *Light Adv. Manuf.* **2**, 28 (2021).
- Maimone, A., Georgiou, A. & Kollin, J. S. Holographic near-eye displays for virtual and augmented reality. *ACM Trans. Graph.* **36**, 85 (2017).
- Yaraş, F., Kang, H. & Onural, L. State of the art in holographic displays: a survey. *J. Disp. Technol.* **6**, 443–454 (2010).
- Chang, C. L. et al. Toward the next-generation VR/AR optics: a review of holographic near-eye displays from a human-centric perspective. *Optica* **7**, 1563–1578 (2020).
- Blinder, D. et al. The state-of-the-art in computer generated holography for 3D display. *Light Adv. Manuf.* **3**, 35 (2022).
- Park, J. H. & Lee, B. Holographic techniques for augmented reality and virtual reality near-eye displays. *Light Adv. Manuf.* **3**, 9 (2022).
- Gilles, A. & Gioia, P. Real-time layer-based computer-generated hologram calculation for the Fourier transform optical system. *Appl. Opt.* **57**, 8508–8517 (2018).
- Blinder, D., Nishitsuji, T. & Schelkens, P. Real-time computation of 3D wireframes in computer-generated holography. *IEEE Trans. Image Process.* **30**, 9418–9428 (2021).
- Horisaki, R., Takagi, R. & Tanida, J. Deep-learning-generated holography. *Appl. Opt.* **57**, 3859–3863 (2018).
- Chakravarthula, P. et al. Learned hardware-in-the-loop phase retrieval for holographic near-eye displays. *ACM Trans. Graph.* **39**, 186 (2020).
- Peng, Y. F. et al. Neural holography with camera-in-the-loop training. *ACM Trans. Graph.* **39**, 185 (2020).
- Shi, L. et al. Towards real-time photorealistic 3D holography with deep neural networks. *Nature* **591**, 234–239 (2021).
- Peng, Y. F. et al. Speckle-free holography with partially coherent light sources and camera-in-the-loop calibration. *Sci. Adv.* **7**, eabg5040 (2021).
- Wu, J. C. et al. High-speed computer-generated holography using an autoencoder-based deep neural network. *Opt. Lett.* **46**, 2908–2911 (2021).
- Situ, G. H. Deep holography. *Light Adv. Manuf.* **3**, 13 (2022).
- Shi, L., Li, B. C. & Matusik, W. End-to-end learning of 3D phase-only holograms for holographic display. *Light Sci. Appl.* **11**, 247 (2022).
- Choi, S. et al. Time-multiplexed neural holography: a flexible framework for holographic near-eye displays with fast heavily-quantized spatial light modulators. *ACM SIGGRAPH 2022 Conference Proceedings*. Vancouver, BC, Canada: Association for Computing Machinery, 2022, 32.
- Gopakumar, M. et al. Full-colour 3D holographic augmented-reality displays with metasurface waveguides. *Nature* **629**, 791–797 (2024).
- Işıl, Ç. et al. Super-resolution image display using diffractive decoders. *Sci. Adv.* **8**, eadd3433 (2022).
- Tseng, E. et al. Neural étendue expander for ultra-wide-angle high-fidelity holographic display. *Nat. Commun.* **15**, 2907 (2024).
- Ulyanov, D., Vedaldi, A. & Lempitsky, V. Instance normalization: the missing ingredient for fast stylization. Print at <https://doi.org/10.48550/arXiv.1607.08022> (2016).
- He, K. M. et al. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 IEEE International Conference on Computer Vision (ICCV). 1026–1034 (IEEE, 2015).
- Cohen, G. et al. EMNIST: an extension of MNIST to handwritten letters. Print at <https://doi.org/10.48550/arXiv.1702.05373> (2017).
- Greenbaum, A. et al. Increased space-bandwidth product in pixel super-resolved lensfree on-chip microscopy. *Sci. Rep.* **3**, 1717 (2013).
- Goodman, J. W. Introduction to Fourier Optics. (McGraw-Hill Book Co., 1968).
- Dowski, E. R. & Cathey, W. T. Extended depth of field through wave-front coding. *Appl. Opt.* **34**, 1859–1866 (1995).
- Rai, M. R. & Rosen, J. Depth-of-field engineering in coded aperture imaging. *Opt. Express* **29**, 1634–1648 (2021).
- Akpınar, U. et al. Learning wavefront coding for extended depth of field imaging. *IEEE Trans. Image Process.* **30**, 3307–3320 (2021).
- Xu, Z. Q. et al. DSCCNet for high-quality 4K computer-generated holograms. *Opt. Express* **33**, 13733–13747 (2025).
- Liu, N. H. et al. Propagation-adaptive 4K computer-generated holography using physics-constrained spatial and Fourier neural operator. *Nat. Commun.* **16**, 7761 (2025).
- Mengu, D. et al. Misalignment resilient diffractive optical networks. *Nano-photonics* **9**, 4207–4219 (2020).
- Mengu, D., Rivenson, Y. & Ozcan, A. Scale-, shift-, and rotation-invariant diffractive optical networks. *ACS Photonics* **8**, 324–334 (2021).
- Zuo, C. et al. Deep learning in optical metrology: a review. *Light Sci. Appl.* **11**, 39 (2022).
- Goi, E. et al. Nanoprinted high-neuron-density optical linear perceptrons performing near-infrared inference on a CMOS chip. *Light Sci. Appl.* **10**, 40 (2021).
- Chen, H. et al. Diffractive deep neural networks at visible wavelengths. *Engineering* **7**, 1483–1491 (2021).
- Bai, B. J. et al. Data-class-specific all-optical transformations and encryption. *Adv. Mater.* **35**, 2212091 (2023).
- Shen, C. Y. et al. Broadband unidirectional visible imaging using wafer-scale nano-fabrication of multi-layer diffractive optical processors. *Light Sci. Appl.* **14**, 267 (2025).
- Mengu, D. et al. Analysis of diffractive optical neural networks and their integration with electronic neural networks. *IEEE J. Sel. Top. Quantum Electron.* **26**, 3700114 (2020).
- Lin, X. et al. All-optical machine learning using diffractive deep neural networks. *Science* **361**, 1004–1008 (2018).
- Chen, S. Q. et al. Optical generative models. *Nature* **644**, 903–911 (2025).
- Ha, D. & Eck, D. A neural representation of sketch drawings. Proceedings of the 6th International Conference on Learning Representations. OpenReview.net: Vancouver, BC, Canada, 2018.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
- Wang, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**, 600–612 (2004).

