

Self-supervised high-pixel-resolution reconstruction for single-pixel imaging via measurement decoupling

Qihang Liang, Xin Xu, Jia Wu, Zichao Wei, Shovit A. Gupta, Jinyong Wang, Kun Li, Xiaoxi Chen*, Pinghe Wang*, Fake Lu* and Guohua Shi*

Citation: Liang QH, Xu X, Wu J et al. Self-supervised high-pixel-resolution reconstruction for single-pixel imaging via measurement decoupling. *Opto-Electron Adv* **9**, 260133 (2026).

<https://doi.org/10.29026/oea.2026.260133>

Received: 23 May 2026; Accepted: 27 July 2026; Published online: 8 September 2026

Related articles

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

Fujie Li, Haoyu Zhang, Zhilan Lu, et al

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

Pixelated BIC metasurfaces for terahertz integrated sensing and imaging

Zhanqiang Xue, Guizhen Xu, Junliang Chen, et al

Opto-Electronic Advances 2026, **9**(3): 250211 doi: [10.29026/oea.2026.250211](https://doi.org/10.29026/oea.2026.250211)

Multi-prior physics-enhanced neural network enables pixel super-resolution and twin-image-free phase retrieval from single-shot hologram

Xuan Tian, Runze Li, Tong Peng, et al

Opto-Electronic Advances 2024, **7**(9): 240060 doi: [10.29026/oea.2024.240060](https://doi.org/10.29026/oea.2024.240060)

More related articles in Opto-Electronic Journals Group website 



Self-supervised high-pixel-resolution reconstruction for single-pixel imaging via measurement decoupling

Qihang Liang¹, Xin Xu², Jia Wu¹, Zichao Wei¹, Shovit A. Gupta³, Jinyong Wang⁴, Kun Li¹, Xiaoxi Chen^{1*}, Pinghe Wang^{1*}, Fake Lu^{3*} and Guohua Shi^{2*}

Abstract: Single-pixel imaging (SPI) is constrained by the optical resolving power of the imaging system and the limitations of its sampling mechanism, making high-pixel-resolution (HPR) imaging difficult to achieve and leading to degradation in structural details in reconstructed images. In this work, we propose a self-supervised SPI reconstruction method. The method can directly generate a target HPR image using a neural network, and then downsample it to the measurement-pixel-resolution (MPR) grid for optimization. By imposing consistency constraints on the network in the measurement domain via the SPI physical model, it enables convergent reconstruction of HPR images without external supervision. Based on this framework, the dependency between physical MPR grid and the target reconstruction space is decoupled, allowing the algorithm to recover HPR images from low-resolution MPR data and alleviating the information representation limitation caused by the constrained MPR dimension. Simulation and experimental results demonstrate that the proposed method achieves stable reconstruction performance, preserves fine structural details effectively, and outperforms several existing methods in both visual quality and quantitative metrics, providing an effective imaging solution for applications such as biomedical imaging.

Keywords: single-pixel imaging; high-pixel-resolution reconstruction; self-supervised learning; physics-constrained; measurement decoupling

DOI: [10.29026/oea.2026.260133](https://doi.org/10.29026/oea.2026.260133) | CSTR: [32247.14.oea.2026.260133](https://cstr.cn/32247.14.oea.2026.260133)

Citation: Liang QH, Xu X, Wu J et al. Self-supervised high-pixel-resolution reconstruction for single-pixel imaging via measurement decoupling. *Opto-Electron Adv*, 260133 (2026).

1 Introduction

As an evolving computational imaging paradigm, single-pixel imaging (SPI) has attracted widespread attention¹⁻⁷. Unlike conventional imaging systems that rely on pixelated detector arrays, SPI reconstructs target scenes by correlating structured illumination patterns with intensity measurements acquired by a single-pixel detector. Benefiting from the superior sensitivity, broad spectral response, and low-cost characteristics of single-pixel detectors, SPI exhibits unique advantages in challenging imaging scenarios where conventional array sensors suffer from performance degradation, including three-dimensional imaging⁸, low-light imaging^{9,10}, non-visible band imaging¹¹⁻¹⁵, and imaging

through scattering media¹⁶. Despite its advantages, obtaining high-quality SPI reconstructions remains challenging in practice, as the imaging performance is highly dependent on the acquisition conditions. In practical systems, insufficient sampling rates, limited sampling resolutions, measurement noise, and adverse imaging environments often lead to degraded reconstruction quality. Moreover, reconstructing high-fidelity images from highly compressed and inherently ill-posed measurements remains a challenging task, and traditional reconstruction methods may struggle to faithfully recover fine image structures and textures. These limitations restrict the practical performance of SPI in demanding imaging applications.

To address these limitations, extensive efforts have been

Received: 23 May 2026

Accepted: 27 July 2026

Published online: 8 September 2026

¹School of Optoelectronic Science and Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China; ²Hangzhou Institute of Medicine, Chinese Academy of Sciences, Hangzhou 310002, China; ³Department of Biomedical Engineering, Thomas J. Watson College of Engineering and Applied Science, Binghamton University, State University of New York, Binghamton, New York 13902, USA; ⁴Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576, Singapore.

*Correspondence: XX Chen, E-mail: chenxiaoxi@uestc.edu.cn; PH Wang, E-mail: wphsci@uestc.edu.cn; F Lu, E-mail: fakelu@binghamton.edu; GH Shi, E-mail: ghshi_lab@126.com

devoted to improving SPI performance from both acquisition and reconstruction perspectives. Among these efforts, advances in reconstruction algorithms have played a crucial role in enhancing image quality and extending the applicability of SPI. Research has transitioned from traditional correlation-based algorithms to modern deep learning (DL) algorithms. Traditional correlation-based algorithms^{3,6} usually require oversampling to achieve acceptable image reconstruction. Compressed sensing (CS)^{4,5,17,18} enables image reconstruction from undersampled measurements and outperforms traditional methods. However, its ill-posed nature leads to information loss and degraded image quality. In recent years, DL has emerged as a powerful paradigm in computational imaging, enabling remarkable advances in addressing challenging imaging problems^{19–27}. In SPI, DL-based algorithms are typically divided into data-driven and physics-constrained algorithms. The data-driven algorithms^{7,28,29} learn a direct mapping from measurements to images using large paired datasets. They offer fast and high-quality reconstructions but their performance depends heavily on extensive training data. The physics-constrained algorithms^{30–35} embed the SPI forward model into the reconstruction process and use SPI measurements to iteratively guide the neural network. This self-supervised scheme makes large external datasets no longer essential and couples physical priors with the nonlinear representation ability of deep networks, yielding high-quality reconstructions while offering better generalization and interpretability. Furthermore, recent works are pushing the boundaries of DL-based methods by exploring new research avenues, such as fusing physical priors with data-driven methods^{36–38} and ensuring robustness in practical scenarios³⁹. Overall, DL-based SPI

methods show great potential for efficient and high-quality imaging in practical applications.

In SPI, a continuous scene $O(x_t, y_t)$ is typically discretized into an $N \times N$ spatial grid, and the measurement grid traditionally binds the baseline reconstruction resolution to the sampling dimensions, limiting the recovery of fine spatial details. To achieve high-pixel-resolution (HPR) reconstructions, the number of sampling pixels N^2 must be increased. However, with a fixed number of measurements M , this inevitably reduces the effective sampling rate ($SR = M/N^2$), resulting in degraded reconstruction quality. Due to the hardware limitations of practical imaging systems, directly increasing the acquisition resolution often introduces additional system complexity and cost. Therefore, improving the reconstruction resolution through computational algorithms has become an attractive strategy^{40–44}. In SPI, a two-stage strategy has been explored, where measurement-pixel-resolution (MPR) images are first reconstructed and subsequently enhanced using super-resolution algorithms to obtain HPR results^{45,46}. Such approaches have demonstrated the potential of combining SPI with image super-resolution techniques. However, the performance of SPI super-resolution is highly dependent on the quality of the initial reconstruction. When severe artifacts, noise, and information loss are present in the initial MPR reconstruction, the subsequent super-resolution process may struggle to recover reliable high-frequency details and can even propagate reconstruction errors, thereby limiting the achievable image quality.

To address the aforementioned issues, this paper proposes an adjustable-pixel-resolution image reconstruction method for SPI, termed ARSPI, as shown in Fig. 1. Unlike conven-

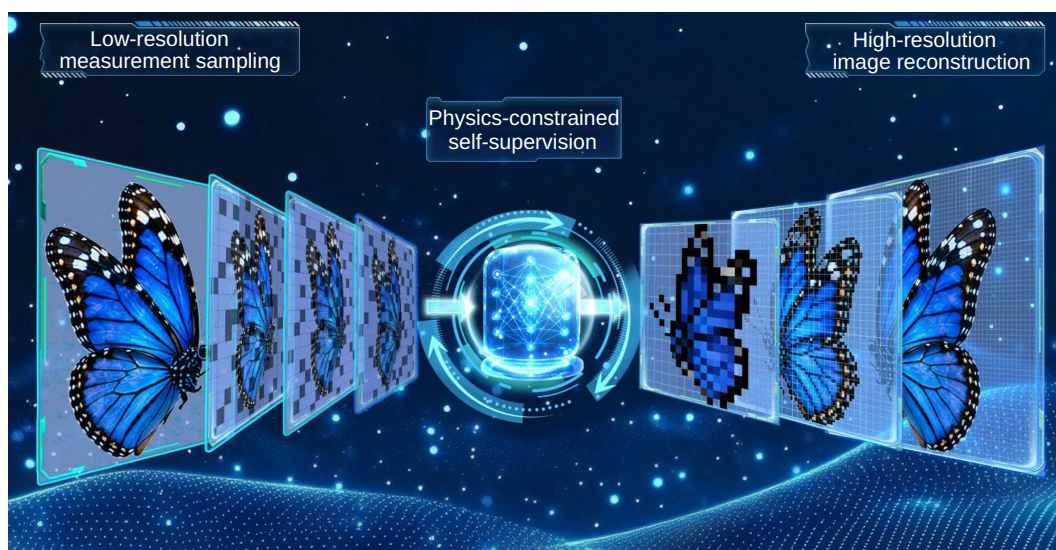


Fig. 1 | Conceptual illustration of the ARSPI method. A spatially continuous target scene is modulated and sampled by discrete modulation patterns through single-pixel imaging measurements, resulting in low-dimensional measurement-pixel-resolution data. A physics-constrained self-supervised neural network is employed for controllable high-pixel-resolution imaging. As the target reconstruction scale increases, image textures and structural details are progressively recovered with enhanced clarity.

tional SPI reconstruction methods that tightly couple the reconstruction grid with the measurement sampling, ARSPI decouples the reconstruction scale from the physical measurement dimensions, enabling flexible imaging at arbitrary target pixel-resolutions from low-dimensional measurements. Specifically, ARSPI directly reconstructs images on a defined HPR grid using a neural network, while enforcing measurement consistency through the SPI physical model on the MPR grid. This decoupled optimization strategy enables the simultaneous recovery of both MPR and HPR images from the same low-dimensional measurements, driving the reconstruction toward stable convergence. To achieve this, the framework adopts a self-supervised optimization strategy that exploits the implicit priors of deep neural network⁴⁷, without relying on any pre-training data. ARSPI is implemented using a multi-attention U-Net architecture, which supports effective feature extraction and detail modeling during the optimization process. Consequently, it produces finer pixel-level representations on the reconstruction grid and avoids the cumulative degradation commonly observed in two-stage super-resolution reconstruction strategies. Both numerical simulations and optical experiments indicate that ARSPI enables reconstruction at a finer pixel grid, thereby alleviating the inherent trade-off between measurement efficiency and imaging quality in SPI. This provides a reliable approach for imaging systems limited by optical resolving power, substantially enhancing structural details and image smoothness.

2 Principles and methods

2.1 SPI measurement model

In SPI, the measurement acquisition process is performed at a MPR level primarily determined by the spatial resolution of the modulation patterns rather than the physical pixel resolution of the spatial light modulator (SLM). Specifically, as shown in Fig. 2(a), the continuous scene $O(x_t, y_t)$ is illuminated by a sequence of spatially modulated patterns $P_m(x_t, y_t)$, $m = 1, 2, \dots, M$. Each P_m is generated from a predefined $w \times w$ discrete modulation matrix and displayed on the SLM by expanding each pattern element into a corresponding block of SLM pixels according to the optical system configuration. The displayed patterns are then optically projected onto the continuous scene coordinates (x_t, y_t) . For each modulation, the detector records only the total light intensity, yielding a one-dimensional bucket measurement^{3,48}:

$$I_m = \iint P_m(x_t, y_t) O(x_t, y_t) dx_t dy_t. \quad (1)$$

The measurements $\mathbf{I} = \{I_1, I_2, \dots, I_M\}$ together with the corresponding modulation patterns $\mathbf{P} = \{P_1, P_2, \dots, P_M\}$ form the complete dataset required for image reconstruction. Traditional SPI methods³ reconstruct images by correlating the P_m with the I_m :

$$X = \frac{1}{M} \sum_{m=1}^M P_m I_m - \frac{1}{M^2} \left(\sum_{m=1}^M P_m \right) \left(\sum_{m=1}^M I_m \right), \quad (2)$$

where $X \in \mathbb{R}^{w \times w}$. As a result, it can only reconstruct images on the MPR grid defined by the modulation patterns, which limits the achievable pixel grid density.

2.2 Proposed ARSPI

We propose the ARSPI framework that decouples the target reconstruction scale from the physical sampling grid in SPI, as shown in Fig. 2(c). Specifically, a fixed random noise tensor $Z \in \mathbb{R}^{W \times W}$ is used as the input to the generator to produce the HPR estimated image:

$$\tilde{X}^{\text{HPR}} = G_\theta(Z), \quad (3)$$

where $\tilde{X}^{\text{HPR}} \in \mathbb{R}^{W \times W}$, $W = w \cdot \alpha$, α is the adjustable scaling factor, and $G_\theta(\cdot)$ represents the generative network parameterized by the learnable weights θ . Since SPI measurements are acquired in a low-dimensional $w \times w$ measurement space, the reconstruction is represented in a $W \times W$ high-dimensional discretization space, resulting in an ill-posed inverse problem.

The reconstruction is therefore formulated as an optimization problem over the parameters of the network. During optimization, the network output is evaluated only through its consistency with the physical SPI measurement process. To bridge the resolution discrepancy between the reconstruction domain and the experimental acquisition process, a differentiable average pooling operator $D_\alpha(\cdot)$ is introduced to model the α -fold resolution reduction, as illustrated in Fig. 2(b). Specifically, the MPR image is obtained as

$$\tilde{X}^{\text{MPR}} = D_\alpha(\tilde{X}^{\text{HPR}}), \quad (4)$$

where $\tilde{X}^{\text{MPR}} \in \mathbb{R}^{w \times w}$. The average pooling operation divides the HPR image into non-overlapping $\alpha \times \alpha$ regions and calculates the mean intensity within each region, which is expressed as

$$\tilde{X}^{\text{MPR}}(i, j) = \frac{1}{\alpha^2} \sum_{m=0}^{\alpha-1} \sum_{n=0}^{\alpha-1} \tilde{X}^{\text{HPR}}(\alpha i + m, \alpha j + n), \quad (5)$$

$$i, j = 0, \dots, w-1.$$

This operation provides a differentiable approximation of the spatial integration process in the SPI system, where each projected pattern element corresponds to a finite illumination area on the object plane and the detected signal is determined by the integrated light intensity. Therefore, average pooling provides a more physically meaningful resolution mapping between the HPR and MPR domains while preserving differentiability for gradient-based optimization.

Then, the $\tilde{X}^{\text{MPR}}$ together with the discrete modulation patterns $P_m(i, j)$ are used to compute the simulated bucket signal $\tilde{I}_m$:

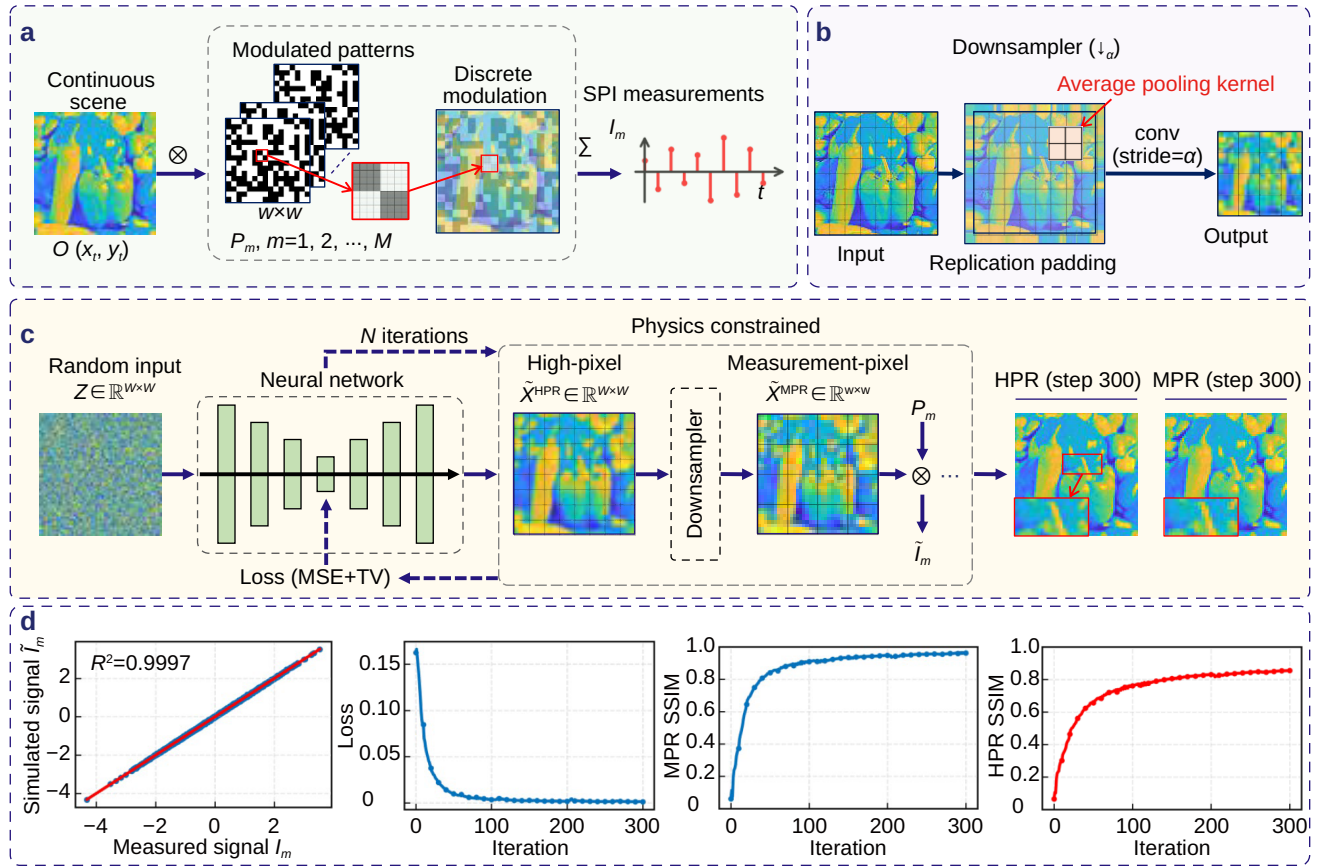


Fig. 2 | Overview of ARSPI. (a) Schematic diagram of the signal acquisition process in single-pixel imaging. (b) Downsampling operator: performs image downsampling according to the adjustable scaling factor α . (c) The proposed ARSPI algorithm. A fixed random noise tensor is used to generate the high-pixel-resolution image, which is then downsampled by the average pooling operator to obtain the measurement-pixel-resolution image. The simulated measurements are generated through the SPI physical model. The network parameters are updated via the loss function, and the image quality is iteratively optimized. (d) Reconstruction results together with the optimization dynamics.

$$\tilde{I}_m = \sum_{i=1}^w \sum_{j=1}^w P_m(i, j) \tilde{X}^{\text{MPR}}(i, j), \quad m = 1, 2, \dots, M. \quad (6)$$

The mean squared error (MSE) between $\tilde{I}_m$ and I_m is used as the loss function to update the network weights θ . To further suppress high-frequency artifacts and promote spatial smoothness in the reconstructed image, a total variation (TV) regularization term is incorporated into the loss function. Thus, the entire optimization process is formulated as:

$$\text{Loss} = \arg\min_{\theta} \sum_{m=1}^M \|\tilde{I}_m - I_m\|^2 + \beta \text{TV}[G_{\theta}(Z)], \quad (7)$$

where β is a weighting factor controlling the contribution of the TV regularization. Therefore, the optimized parameters θ^* are obtained, which are subsequently used to generate the final HPR reconstruction results.

Although the optimization in Eq. (7) involves only low-dimensional measurement constraints, the feasible solution space is progressively restricted during optimization by the joint effect of the network parameterization and the physi-

cal forward model. In this way, the MPR and HPR structural representations are tightly coupled within a unified framework. This coupling not only enforces physical consistency but also effectively preserves fine structural details. Figure 2(d) presents the final reconstruction results together with the optimization dynamics. The simulated signals agree closely with the measured signals, achieving an R^2 value of 0.9997. Meanwhile, the loss function decreases steadily during optimization, while both the MPR and HPR SSIM values increase progressively and eventually converge. These observations demonstrate the measurement consistency, reconstruction accuracy, and stable convergence behavior of the proposed ARSPI framework.

To enhance the representation capability of the generative model for SPI, we adopt a U-Net architecture⁴⁹ and design a variant that integrates multiple attention mechanisms, referred to as MAU-Net. Inspired by general attention concepts widely used in prior studies^{50–52}, MAU-Net incorporates attention modules at different stages of the encoder-decoder network to facilitate effective feature interaction and fusion. The code and neural network architec-

ture have been released to an open-source repository to support reproducibility. Computational tasks are performed on a system featuring 16 GB of RAM and an NVIDIA GeForce RTX 3060 GPU. All algorithms are developed in Python using the PyTorch framework, leveraging CUDA acceleration to ensure computational efficiency during the optimization and reconstruction processes.

3 Results and discussions

3.1 Comparison of measurement-pixel-resolution reconstructions

We conduct comparative analyses with four representative methods: computational single-pixel imaging (CSPI)³, compressive sensing single-pixel imaging (CSSPI)¹⁷, a physics-driven single-pixel imaging (PDSPI) approach inspired by ref.³⁰, and a dual-domain constrained single-pixel imaging (DDCSPI) approach inspired by ref.³¹. For a fair comparison, PDSPI and DDCSPI are implemented using the same MAU-Net architecture and optimization strategy as ARSPI, with only the corresponding physical constraints modified. All methods are tested using random modulation patterns with binary values $\{-1, 1\}$ as the sampling matrices, as their noise robustness and structure-agnostic nature provide an objective setting for evaluating both reconstruction quality and generalization performance. To quantitatively evaluate the reconstruction quality in the image domain, we employ three complementary metrics: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM)⁵³, and learned perceptual image patch similarity (LPIPS)⁵⁴. Higher values indicate better performance for PSNR and SSIM, while lower values indicate better performance for LPIPS.

We test three 64×64 grayscale images with diverse textures ("butterfly", "pepper" and "house") at $SR = 25\%$ and 100% for quantitative and visual comparison. Figure 3 presents the visual reconstruction results and quantitative metrics. CSPI suffers from severe noise contamination, yielding poor structural fidelity. CSSPI reconstructs coarse outlines but suffers from oversmoothing at low sampling rates, losing high-frequency details. PDSPI shows improved image quality but still suffers from residual noise. DDCSPI enhances reconstruction robustness through dual-domain constraints. However, when using random modulation patterns, the reconstructed images exhibit increased noise contamination, which leads to degraded image quality. In contrast, ARSPI (with the scaling factor of $\alpha = 4$) preserves fine structural details and sharp edges at both sampling rates, achieving superior image quality compared with all competing methods. Quantitatively, ARSPI achieves superior overall reconstruction performance across different sampling rates based on the average results of the three test images. In particular, ARSPI demonstrates clear advantages under low sampling conditions, achieving superior recon-

struction quality compared with other methods. At higher sampling rates, although PDSPI achieves a slightly better PSNR performance, ARSPI provides better structural preservation and perceptual fidelity, as reflected by the SSIM and LPIPS metrics. These results indicate that ARSPI can effectively recover image details and maintain high-quality reconstructions under different measurement conditions.

3.2 Measurement-decoupled high-pixel-resolution reconstructions

To further demonstrate the capability of ARSPI in reconstructing images beyond the MPR grid, HPR reconstructions are performed using two representative targets, including a binary "UESTC" pattern and a grayscale "bird" image. Both targets are tested under 64×64 measurements and $SR = 100\%$. The MPR results and enhanced resolutions with scaling factors of $\alpha = 2$ and $\alpha = 4$ are presented in Fig. 4. For comparison, CSSPI and PDSPI are selected as representative conventional and self-supervised deep learning-based SPI reconstruction methods, respectively. Since these methods inherently recover MPR images, two representative super-resolution approaches are further applied to their reconstructed results, including bicubic interpolation⁴⁰ and zero-shot super-resolution (ZSSR)⁴². Specifically, bicubic interpolation represents conventional image upscaling without additional prior information, whereas ZSSR utilizes an internal learning strategy to enhance image details from a single input image. In contrast, ARSPI directly reconstructs HPR images through the proposed measurement-decoupled framework without requiring any subsequent super-resolution processing.

As shown in Fig. 4(a), the differences among the methods are already observable at the MPR grid. CSSPI exhibits noticeable over-smoothing effects, resulting in blurred boundaries and loss of structural details. In contrast, PDSPI and ARSPI both provide relatively accurate MPR reconstructions. Furthermore, ARSPI successfully reconstructs the MPR "UESTC" images with clear boundaries and fine structural details at both $\alpha = 2$ and $\alpha = 4$. In comparison, bicubic interpolation increases only the spatial size of the CSSPI and PDSPI reconstructions while introducing significant smoothing effects due to the lack of high-frequency information recovery. Although ZSSR provides improved visual quality by exploiting internal image priors, the recovered edges remain less sharp than those obtained by ARSPI. These results indicate that post-processing super-resolution methods cannot fundamentally recover details that are absent from the initial MPR reconstruction. Similar observations can be found for the grayscale "bird" image in Fig. 4(b). Compared with the binary target, the grayscale image contains richer intensity variations and more complex texture information, making HPR reconstructions more challenging. Although PDSPI combined with ZSSR achieves relatively favorable visual quality, its performance remains

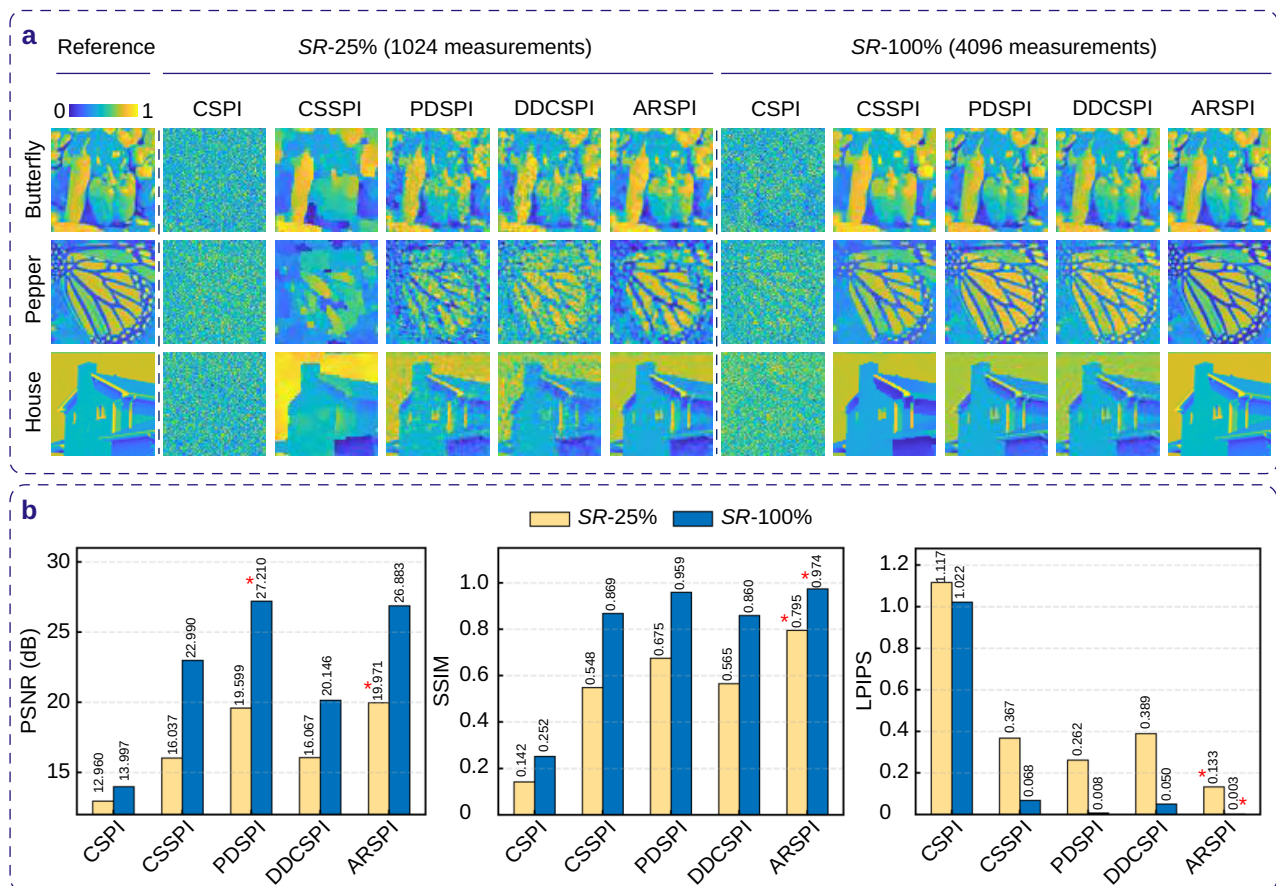


Fig. 3 | Simulation results of measurement-pixel-resolution grid. (a) Simulated results of three grayscale images under 64×64 measurements. Results obtained using CSPI, CSSPI, PDSPI, DDCSPI and ARSPI are shown at the sampling rates $SR=25\%$ and 100% . (b) Quantitative evaluation of the reconstruction performance using the average PSNR, SSIM, and LPIPS values over the three test images. Red asterisks "*" indicate the best performance among all compared methods.

strongly dependent on the fidelity of the initial MPR reconstruction. In particular, the super-resolution results derived from CSSPI suffer from blurred textures and distorted fine structures because the missing information in the initial reconstruction cannot be reliably recovered during the subsequent enhancement stage. By contrast, ARSPI reconstructs the HPR image directly under measurement consistency constraints, allowing fine textures and structural details to be progressively inferred from the acquired measurements rather than solely from image priors. As a result, ARSPI produces more faithful texture representation and improved structural consistency across different reconstruction scales.

Quantitative evaluations further support the visual observations. For the MPR results, ARSPI achieves the best reconstruction performance for both targets and consistently outperforms CSSPI and PDSPI in terms of PSNR, SSIM, and LPIPS. Moreover, the MPR reconstruction quality further improves as the scaling factor increases from $\alpha = 2$ to $\alpha = 4$. This improvement can be attributed to the enhanced representational capacity provided by the larger reconstruction

grid, which allows the network to satisfy the measurement consistency constraint more accurately after downsampling. For the HPR results, ARSPI achieves competitive or superior quantitative performance compared with the super-resolution-based approaches. Although PDSPI combined with ZSSR attains slightly better results for several individual metrics, its performance remains strongly dependent on the quality of the initial MPR reconstruction. In contrast, ARSPI directly reconstructs HPR images from the measurements and maintains stable reconstruction quality as the scaling factor increases. These results indicate that simply enlarging the MPR reconstruction cannot fundamentally overcome the information representation limitation imposed by the measurement process. These quantitative results further validate the effectiveness of ARSPI for HPR imaging beyond the physical sampling limit.

3.3 Analysis of the proposed reconstruction framework

The reconstruction performance of ARSPI is jointly deter-

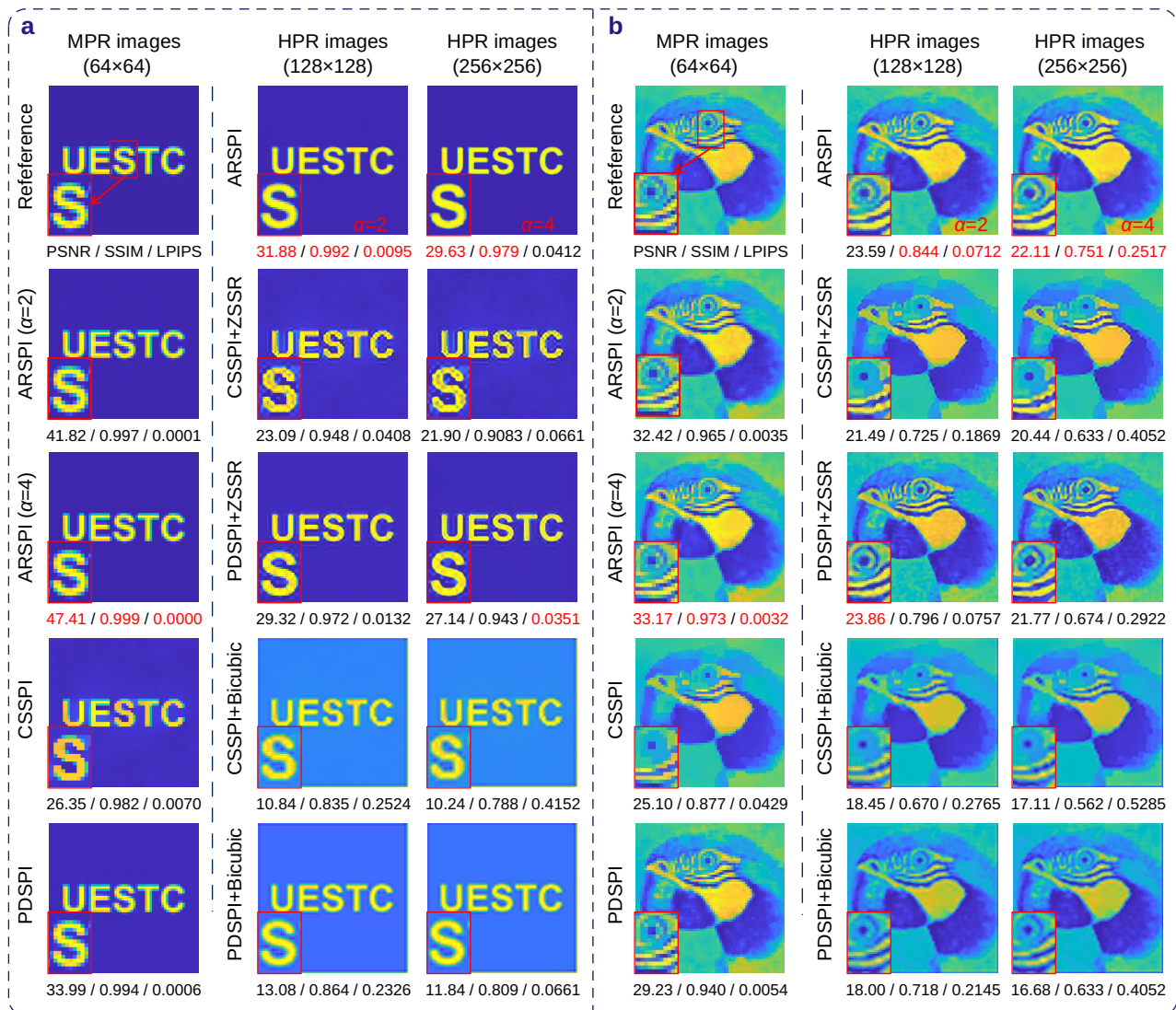


Fig. 4 | High-pixel-resolution reconstruction beyond the measurement-pixel-resolution grid. Reconstruction results of (a) binary "UESTC" and (b) grayscale "bird" images under 64×64 measurements with the sampling rate of $SR=100\%$. The measurement-pixel-resolution (MPR) and high-pixel-resolution (HPR) reconstructions with scaling factors of $\alpha=2$ and $\alpha=4$ are presented. ARSPI is compared with CSSPI and PDSPI, as well as their super-resolution-enhanced results using bicubic interpolation and zero-shot super-resolution (ZSSR) method. PSNR, SSIM, and LPIPS values are reported for quantitative evaluation. Zoomed-in views highlight selected fine structures.

mined by the measurement information and the reconstruction network. Therefore, the influence of the network architecture and sampling rate on the reconstruction quality is investigated through an ablation study. A conventional U-Net without attention modules is compared with the proposed MAU-Net under identical reconstruction conditions. Figure 5(a) presents the reconstructed images obtained by MAU-Net and U-Net at $SR = 12.5\%$ and 125% with scaling factors of $\alpha = 2$ and $\alpha = 4$, including both MPR and HPR results, while Fig. 5(b) summarizes the corresponding SSIM evaluations under different sampling rates. The two networks are evaluated using a 64×64 "pepper" image with SR ranging from 12.5% to 125% . As the SR increases, both MAU-Net and U-Net exhibit improved

SSIM values, indicating that additional measurements provide stronger constraints for image reconstruction. The improvement gradually saturates at high SR , suggesting that the reconstruction performance is increasingly limited by the network representation capability rather than measurement insufficiency. Meanwhile, MAU-Net consistently achieves higher SSIM values than U-Net for both MPR and HPR reconstructions across all SR . Although the visual differences between the two networks are relatively subtle, MAU-Net provides better intensity consistency and grayscale fidelity, which agrees with the quantitative evaluation. These results demonstrate that the introduced attention modules effectively enhance feature representation and improve reconstruction fidelity in ARSPI.

The ARSPI framework employs MAU-Net as the reconstruction network, which contains 60,798 trainable parameters. The relatively compact network architecture enables efficient optimization. To evaluate the computational efficiency of ARSPI, the reconstruction time is measured using the 64×64 "pepper" image under different sampling rates and scaling factors, with a fixed optimization iteration number of 300. Specifically, the effects of sampling rate and scaling factor are evaluated separately by varying SR with fixed $\alpha = 2$, and varying α with fixed SR = 100%. As shown in Fig. 5(c), the reconstruction time remains

nearly constant with varying sampling rates, ranging from 12.5% to 125%, with an average value of approximately 22.5 s. This is because the sampling rate only changes the dimensionality of the measurement vector, while the reconstruction network architecture and optimization process remain unchanged. In contrast, increasing the scaling factor α introduces additional computational costs due to the enlarged reconstruction space. Specifically, the reconstruction time increases from 22.47 s at $\alpha = 2$ to 31.47 s at $\alpha = 8$. Nevertheless, the additional computational cost remains moderate, enabling ARSPI to achieve flexible HPR reconstruction

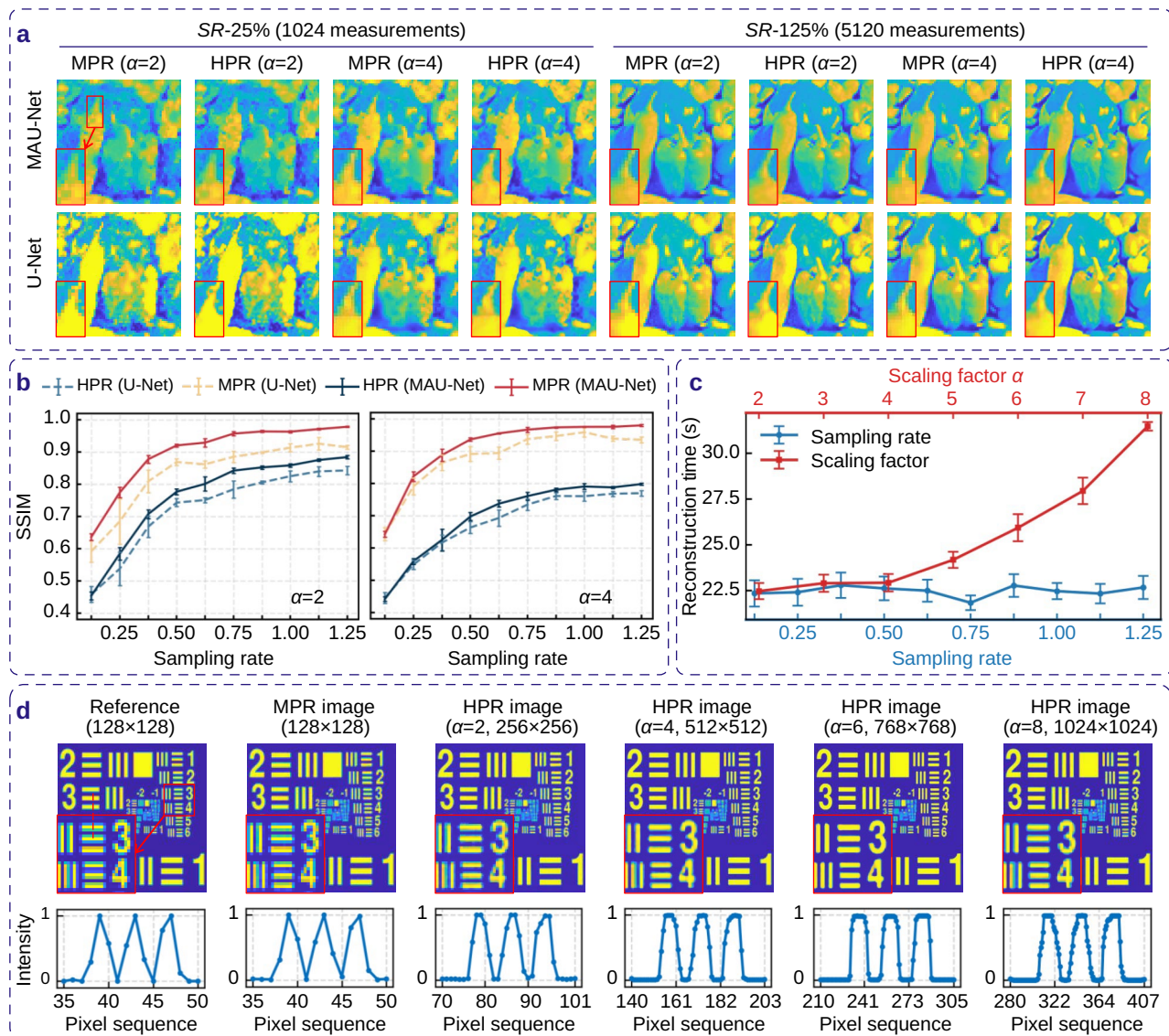


Fig. 5 | Performance analysis of the proposed ARSPI framework. (a) MPR and HPR reconstruction results of the "pepper" image obtained by MAU-Net and U-Net under 64×64 measurements, with scaling factors of $\alpha=2$ and $\alpha=4$, under sampling rates of SR=12.5% and 125%. (b) SSIM comparison of MAU-Net and U-Net for MPR and HPR reconstructions under sampling rates ranging from 12.5% to 125%. (c) Computational efficiency analysis of ARSPI in terms of reconstruction time under different sampling rates and scaling factors. (d) Reconstructions of a resolution target under 128×128 measurements, with pixel distribution analysis in detailed regions. Zoomed-in views highlight selected fine structures.

with acceptable computational efficiency.

Figure 5(d) presents simulations conducted using a resolution target under 128×128 measurements and $SR = 100\%$. The reconstruction performance is evaluated at $\alpha = 2, 4, 6, 8$, and the pixel distribution in detailed regions is further visualized and analyzed. The results show that the reconstructions on the baseline MPR grid are limited by the pixel size, leading to noticeable blurring in detailed regions and a relatively disordered pixel distribution. As α increases, the target structures of images become clearer, and the pixel distribution in detailed regions tends to be more uniform with smoother transitions, effectively recovering the fine stripe structures of the resolution target. These results indicate that, although ARSPI cannot surpass the optical resolution limit, it can effectively enhance image quality and detail representation at the pixel level, demonstrating practical value in scenarios with constrained imaging conditions. It is worth noting that when the scaling factor becomes sufficiently large (e.g., $\alpha = 8$), slight overfitting artifacts may appear, indicating a practical upper bound under fixed measurement conditions. This can be explained by the fact that, under a fixed number of measurements, expanding the reconstruction grid dimensions leads to a more underdetermined inverse problem. In this case, the reconstruction relies more strongly on the implicit prior of the network, which may generate plausible but insufficiently constrained high-frequency details, leading to reduced reconstruction

stability at larger scaling factors.

The downsampling operator plays a critical role in ARSPI because it establishes the relationship between the reconstructed HPR and MPR images constrained by the acquired single-pixel measurements. To further investigate the influence of the downsampling operator, average pooling, lanczos interpolation, and bilinear interpolation are compared under 64×64 measurements and $SR = 100\%$. Figure 6(a) presents the reconstruction results for $\alpha = 2$ and $\alpha = 4$. As shown by the SSIM values, all three kernels achieve comparable performance for reconstructions on the MPR grid, indicating that the measured information is well preserved regardless of the downsampling operator. In contrast, more significant differences are observed in HPR reconstructions, where the average kernel consistently provides the highest SSIM values. This behavior can be attributed to the imaging mechanism of the SPI system, in which each projected pixel corresponds to spatial integration over a finite area. Therefore, average pooling provides a more physically meaningful approximation of the pixel aperture response than interpolation-based kernels. The average pooling operator is adopted as the downsampling operator in ARSPI, offering both improved reconstruction performance and better consistency with the imaging mechanism.

To determine the appropriate TV regularization weight β in Eq. (7), a sensitivity analysis is conducted using a grayscale "house" image. The MPR grid size is fixed at

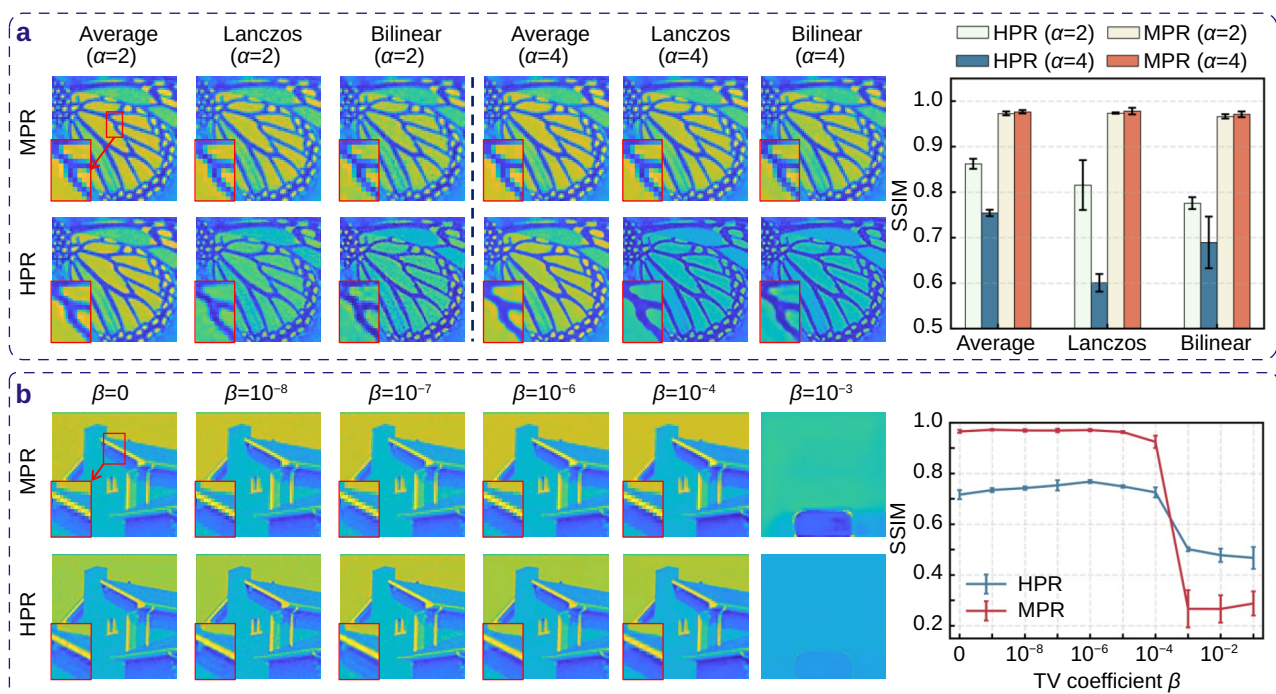


Fig. 6 | Analysis of key components in the proposed ARSPI framework. (a) Comparison of ARSPI reconstructed images using different downsampling kernels, including average pooling, lanczos interpolation, and bilinear interpolation. (b) Sensitivity analysis of the TV regularization weight β . The corresponding reconstruction results and SSIM values under different settings are presented. Zoomed-in views highlight selected fine structures.

64×64 , with $SR = 100\%$ and $\alpha = 4$, β is varied from 0 to 10^{-1} . As shown in Fig. 6(b), excessively small β provide insufficient regularization and lead to increased reconstruction noise, although the SSIM remains at a consistently high level when β is smaller than 10^{-5} . When β increases to 10^{-4} , the SSIM begins to decrease, accompanied by noticeable loss of fine image details. For $\beta \geq 10^{-3}$, the reconstructed images become severely over-smoothed and fail to preserve the target structures, resulting in a significant degradation of reconstruction quality. These results indicate that excessively large weights impose overly strong smoothness constraints, causing detail loss and structural degradation. Considering the trade-off between noise suppression and detail preservation, $\beta = 10^{-6}$ is adopted as the default setting for the experiments in this work.

3.4 Experimental demonstrations

The experimental setup is illustrated in Fig. 7. A He-Ne laser ($\lambda = 632.8$ nm) is first expanded and collimated by a spatial filter system (SFS, including a $20\times$ microscope objective and a $15\ \mu\text{m}$ pinhole) and a lens (L_1 , $f = 200$ mm). A diaphragm (D_1) is used to control the beam diameter. To achieve intensity modulation with a phase-only liquid-crystal-on-silicon spatial light modulator (SLM, HDSLM64R), the SLM is positioned between a pair of crossed polarizers (P_1 and P_2). The modulated patterns are loaded onto the SLM and imaged onto the target object plane (O) by a lens (L_2 , $f = 100$ mm). A diaphragm (D_2) at the back focal plane of lens (L_2) acts as a Fourier spatial filter, blocking high-order diffraction and passing only the zero-order component. Finally, the light is collected by a lens (L_3 , $f = 65$ mm) and a complementary metal oxide semiconductor (CMOS) detector. A computer controls the data acquisition process, synchronizing the pattern display on the SLM with signal integration at the detector. In this experiment, the CMOS detector is used as an array-based bucket detector. The intensity measurements are obtained by summing the pixel values over the

entire detector area, which is equivalent to the spatial integration of the incident light intensity performed by a conventional single-pixel detector. The pixel size of the CMOS sensor determines the spatial sampling interval on the detector plane and the light collection capability, while the measured signals are determined by the integrated optical intensity over the whole sensor.

Optical experiments are conducted using the cropped bar elements from a USAF resolution target under 32×32 measurements. To avoid the influence of undersampling-induced quality degradation in the MPR reconstructions, a saturated sampling rate of $SR = 200\%$ is employed. This setting minimizes the impact of insufficient measurements and enables a more faithful evaluation of the HPR reconstruction performance of different methods. The corresponding pixel-intensity distributions are further analyzed to quantitatively compare the reconstructed bar structures at different image resolutions. Figure 8(a) and 8(b) show results obtained from two regions. The first column presents the MPR reconstructions of different methods. The second and third columns show the corresponding HPR reconstructions, where ARSPI results are directly reconstructed, whereas the competing results are generated by applying super-resolution methods to the MPR images. For the MPR results, ARSPI, CSSPI, and PDSPI can all recover the overall bar structures. Nevertheless, ARSPI produces images with higher intensity responses, reduced noise and reconstruction imperfections compared with CSSPI and PDSPI. The differences become more evident as the target reconstruction scale increases. For CSSPI and PDSPI, bicubic interpolation only enlarges the image grid without introducing additional structural information. Consequently, the reconstructed bar edges become blurred, leading to smoother intensity variations and less steep transitions in the corresponding pixel-intensity profiles. Although ZSSR improves edge sharpness compared with bicubic interpolation, noticeable distortions and intensity fluctuations emerge at $\alpha = 4$. This is because ZSSR enhances high-frequency details based on the internal image statistics of the input reconstruction, without considering the measurement degradation inherent in SPI reconstruction, which may amplify existing noise and non-uniform intensity variations. In contrast, ARSPI maintains clear structural boundaries and well-separated bar patterns across different reconstruction scales, while preserving stable intensity distributions without amplifying reconstruction imperfections. These results demonstrate that ARSPI can effectively preserve structural fidelity during resolution enhancement and provide more reliable HPR representations than post-processing super-resolution approaches. Furthermore, in practical optical experiments, the acquired measurements inevitably contain noise and system imperfections. Therefore, the optimization iterations are determined based on the convergence behavior to obtain stable reconstructions and avoid overfitting of measurement noise.

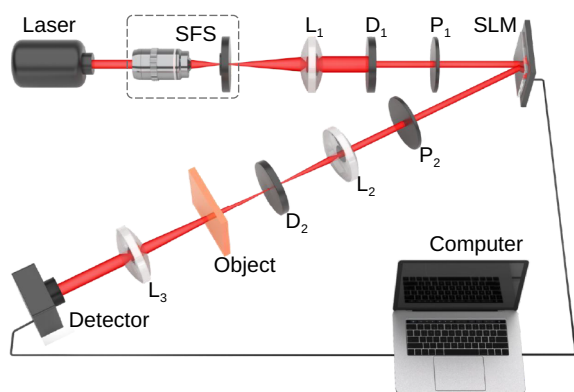


Fig. 7 | Experimental SPI system for validating ARSPI. SFS, spatial filter system; L, convex lens; D, diaphragm; P, polarizer; SLM, phase-only liquid-crystal-on-silicon spatial light modulator.

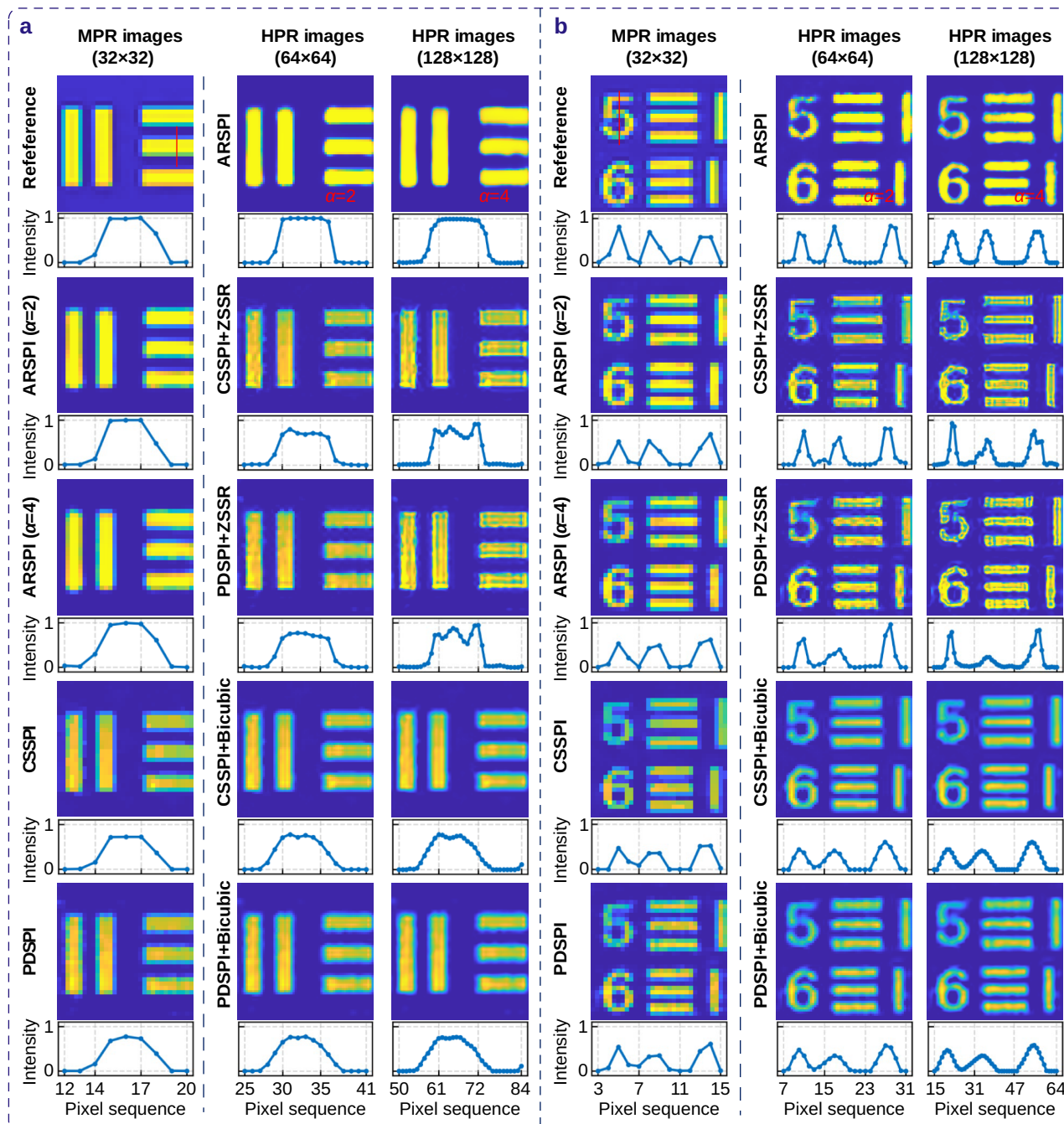


Fig. 8 | Experimental reconstruction results of the USAF resolution target obtained by CSSPI, PDSPI, and ARSPI under 32×32 measurements with the sampling rate $SR=200\%$, with corresponding pixel-intensity profiles along the marked column.

To further evaluate the practical applicability of ARSPI under realistic optical conditions, experiments are conducted on two biological specimens, including a "housefly leg" (results shown in Fig. 9(a)) and a "honeybee mouthpart" (results shown in Fig. 9(b)). The MPR grid size is fixed at 64×64 with $SR = 200\%$, and the HPR images are reconstructed with scaling factors of $\alpha = 2$ and $\alpha = 4$. For the housefly leg, all methods recover the overall morphology of the sample. However, differences become apparent in the

preservation of local structural details. The narrow separation between adjacent regions is not resolved in the CSSPI reconstruction, and subsequent super-resolution processing cannot recover this missing information, demonstrating the limitation of the reconstruction-then-super-resolution strategy. In comparison, PDSPI and ARSPI successfully preserve this subtle structural feature. However, the ZSSR-enhanced PDSPI result exhibits local distortions due to the amplification of reconstruction imperfections during post-processing.

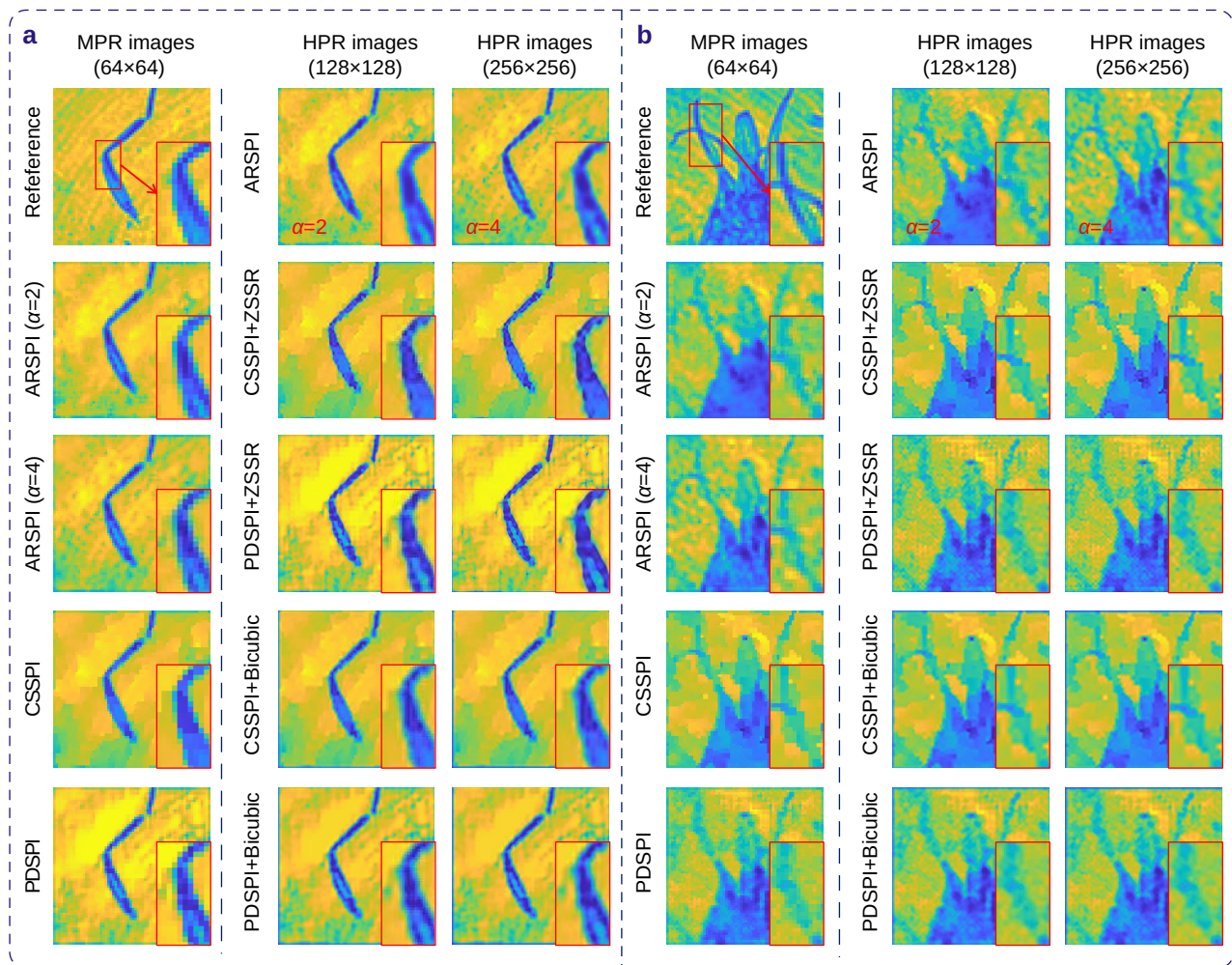


Fig. 9 | Experimental reconstruction results of (a) "housefly leg" and (b) "honeybee mouthpart" obtained by CSSPI, PDSPI, and ARSPI under 64×64 measurements with the sampling rate $SR=200\%$. Zoomed-in views highlight selected fine structures.

By contrast, ARSPI maintains consistent structural representations across different reconstruction scales, providing more reliable HPR reconstructions of the specimen. For the honeybee mouthpart, the comparison is more challenging due to the weak contrast, semi-transparent characteristics, and fine structures of the specimen, as well as noise and system imperfections in practical optical measurements. Therefore, the performance differences among different methods are less evident than those observed in the simulation and resolution-target experiments. Nevertheless, ARSPI provides more consistent reconstructions across different reconstruction scales. In particular, the crossing filament-like structure at the tip of the mouthpart is more clearly resolved in the $\alpha = 4$ ARSPI reconstruction, while it remains less distinguishable in the corresponding results obtained by the other methods. These results demonstrate that ARSPI can maintain stable HPR imaging performance for complex biological specimens under realistic optical conditions.

4 Conclusion

In this work, we have proposed a self-supervised framework for adjustable-pixel-resolution single-pixel imaging, denoted as ARSPI, which enables decoupling the target reconstruction scale from the physical sampling grid and alleviates the trade-off between acquisition efficiency and reconstruction quality by allowing the reconstruction scale to be specified independently of the sampling grid. Specifically, this framework achieves the joint reconstruction of HPR and MPR images under the constraints of the SPI physical forward model, without requiring external training data. As the scaling factor α increases, the enlarged reconstruction grid enhances the network's representational capability, enabling the progressive inference of fine structures while strictly satisfying measurement consistency. The HPR reconstructions generated by this framework can clearly distinguish the microstructures that are blurred on the MPR grid. It should be emphasized that this improvement does not arise from an enhancement of the system's optical resolving capability, but

from algorithmic reconstruction that yields finer image representations under fixed optical constraints, providing a highly practical pixel-resolution enhancement under fixed sampling resources. The optical experiments further validate its practical feasibility. The HPR capability and noise robustness of ARSPI originate from an architectural prior captured by a multi-attention U-Net under strict physical constraints, consistent with the deep image prior concept⁴⁷. In summary, ARSPI provides a versatile and efficient solution for high-quality imaging, and advances in single-pixel imaging are likely to enable broader applications in challenging fields, including biomedical microscopy, industrial non-destructive testing, and remote sensing under low-light conditions.

References

- Pittman TB, Shih YH, Strekalov DV et al. Optical imaging by means of two-photon quantum entanglement. *Phys Rev A* **52**, R3429(R)–R3432(R) (1995).
- Bennink RS, Bentley SJ, Boyd RW. “Two-photon” coincidence imaging with a classical source. *Phys Rev Lett* **89**, 113601 (2002).
- Shapiro JH. Computational ghost imaging. *Phys Rev A* **78**, 061802(R) (2008).
- Duarte MF, Davenport MA, Takhar D et al. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Mag* **25**, 83–91 (2008).
- Katz O, Bromberg Y, Silberberg Y. Compressive ghost imaging. *Appl Phys Lett* **95**, 131110 (2009).
- Ferri F, Magatti D, Lugiato LA et al. Differential ghost imaging. *Phys Rev Lett* **104**, 253603 (2010).
- Lyu M, Wang W, Wang H et al. Deep-learning-based ghost imaging. *Sci Rep* **7**, 17865 (2017).
- Sun MJ, Edgar MP, Gibson GM et al. Single-pixel three-dimensional imaging with time-based depth resolution. *Nat Commun* **7**, 12010 (2016).
- Liu XL, Shi JH, Wu XY et al. Fast first-photon ghost imaging. *Sci Rep* **8**, 5012 (2018).
- Zhao YN, Hou HY, Han JC et al. Full-color photon-counting single-pixel imaging. *Opt Lett* **46**, 4900–4903 (2021).
- Zanotto L, Piccoli R, Dong JL et al. Single-pixel terahertz imaging: a review. *Opto-Electron Adv* **3**, 200012 (2020).
- Wang YQ, Huang K, Fang JA et al. Mid-infrared single-pixel imaging at the single-photon level. *Nat Commun* **14**, 1073 (2023).
- Chen XY, Li J, Du, LH et al. Near-field terahertz single-pixel imaging with ultralow sampling ratio. *Opt Express* **32**, 35126–35138 (2024).
- Xue ZQ, Xu GZ, Chen JL et al. Pixelated BIC metasurfaces for terahertz integrated sensing and imaging. *Opto-Electron Adv* **9**, 250211 (2026).
- Sun B, Huang K, Zhao ZB et al. Infrared single-pixel hyperspectral imaging via spatial-temporal multiplexing. *Laser Photonics Rev* **20**, e01321 (2026).
- Li FJ, Zhang HY, Lu ZL et al. Unsupervised learning enabled label-free single-pixel imaging for resilient information transmission through unknown dynamic scattering media. *Opto-Electron Adv* **8**, 250013 (2025).
- Li CB, Yin WT, Jiang H et al. An efficient augmented Lagrangian method with applications to total variation minimization. *Comput Optim Appl* **56**, 507–530 (2013).
- Assmann M, Bayer M. Compressive adaptive computational ghost imaging. *Sci Rep* **3**, 1545 (2013).
- Barbastathis G, Ozcan A, Situ GH. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
- Tian X, Li RZ, Peng T et al. Multi-prior physics-enhanced neural network enables pixel super-resolution and twin-image-free phase retrieval from single-shot hologram. *Opto-Electron Adv* **7**, 240060 (2024).
- Ma HG, Ren SL, Wei X et al. Enhanced photoacoustic microscopy with physics-embedded degeneration learning. *Opto-Electron Adv* **8**, 240189 (2025).
- Gigan S. Data-driven polarimetric approaches fuel computational imaging expansion. *Opto-Electron Adv* **7**, 240158 (2024).
- Wang F, Czarke JW, Situ GH. Deep learning for computational imaging: from data-driven to physics-enhanced approaches. *Adv Photonics* **7**, 054002 (2025).
- Song K, Bian YX, Wang D et al. Advances and challenges of single-pixel imaging based on deep learning. *Laser Photonics Rev* **19**, 2401397 (2025).
- Xu DL, Cao LC, Sun HB. Adaptive aberration-corrected quantitative phase microscopy via physics-informed deep learning. *Adv Photonics* **7**, 066013 (2025).
- Chen YT, Cao LC. Polarization-guided diffusion prior for eyeglass reflection removal. *Opto-Electron Adv* **9**, 250249 (2026).
- Gao YH, Cao LC. Model-based deep learning enables time-resolved computational microscopy. *Photonix* **7**, 3 (2026).
- Higham CF, Murray-Smith R, Padgett MJ et al. Deep learning for real-time single-pixel video. *Sci Rep* **8**, 2369 (2018).
- Wang F, Wang H, Wang HC et al. Learning from simulation: an end-to-end deep-learning approach for computational ghost imaging. *Opt Express* **27**, 25560–25572 (2019).
- Wang F, Wang CL, Chen ML et al. Far-field super-resolution ghost imaging with a deep neural network constraint. *Light Sci Appl* **11**, 1 (2022).
- Chang XY, Wu Z, Li DY et al. Self-supervised learning for single-pixel imaging via dual-domain constraints. *Opt Lett* **48**, 1566–1569 (2023).
- Liu SP, Meng XF, Yin YK et al. Computational ghost imaging based on an untrained neural network. *Opt Lasers Eng* **147**, 106744 (2021).
- Wang Q, Mi JS, Shi HR et al. Ghost imaging by single-layer neural network based on forward physical model. *IEEE Trans Instrum Meas* **72**, 4505310 (2023).
- Liang QH, Zhang ZL, Wang XK et al. Single-pixel complex-amplitude imaging based on untrained complex-valued convolutional neural network. *Opt Express* **32**, 29656–29665 (2024).
- Song K, Liu HR, Bian YX et al. Physics-guided deep learning for enhanced single-photon single-pixel imaging. *Photonics Res* **14**, 1588–1603 (2026).
- Wang F, Wang CL, Deng CJ et al. Single-pixel imaging using physics enhanced deep learning. *Photonics Res* **10**, 104–110 (2021).
- Zhang XY, Deng CJ, Wang CL et al. VGenNet: variable generative prior enhanced single pixel imaging. *ACS Photonics* **10**, 2363–2373 (2023).
- Zhang YF, Li YX, Liu ZH et al. Physics and data-driven alternative optimization enabled ultra-low-sampling single-pixel imaging. *Adv Photonics Nexus* **4**, 036005 (2025).
- Liu ZH, Yang BH, Zhang YF et al. Comprehensive compensation of real-world degradations for robust single-pixel imaging. *Light Sci Appl* **14**, 365 (2025).
- Keys R. Cubic convolution interpolation for digital image processing. *IEEE Trans Acoust Speech Signal Process* **29**, 1153–1160 (1981).
- Ledig C, Theis L, Huszar F et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recogni-*

- tion (CVPR) 105–114 (IEEE, 2017). <http://doi.org/10.1109/CVPR.2017.19>.
42. Shocher A, Cohen N, Irani M. Zero-shot super-resolution using deep internal learning. In *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 3118–3126 (IEEE, 2018). <http://doi.org/10.1109/CVPR.2018.00329>.
 43. Wang XT, Xie LB, Dong C et al. Real-esrgan: training real-world blind super-resolution with pure synthetic data. In *Proceedings of 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* 1905–1914 (IEEE, 2021). <http://doi.org/10.1109/ICCVW54120.2021.00217>.
 44. Chen JY, Pan JS, Dong JX. FaithDiff: unleashing diffusion priors for faithful image super-resolution. In *Proceedings of 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 28188–28197 (IEEE, 2025). <http://doi.org/10.1109/CVPR52734.2025.02625>.
 45. Jiang PF, Liu JL, Wu L et al. Fourier single pixel imaging reconstruction method based on the U-net and attention mechanism at a low sampling rate. *Opt Express* **30**, 18638–18654 (2022).
 46. Li MD, Mathai A, Lau SLH et al. Underwater object detection and reconstruction based on active single-pixel imaging and super-resolution convolutional neural network. *Sensors* **21**, 313 (2021).
 47. Lempitsky V, Vedaldi A, Ulyanov D. Deep image prior. In *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9446–9454 (IEEE, 2018). <http://doi.org/10.1109/CVPR.2018.00984>.
 48. Gibson GM, Johnson SD, Padgett MJ. Single-pixel imaging 12 years on: a review. *Opt Express* **28**, 28190–28208 (2020).
 49. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (Springer, 2015). http://doi.org/10.1007/978-3-319-24574-4_28.
 50. Woo S, Park J, Lee JY et al. CBAM: convolutional block attention module. In *Proceedings of the 15th European Conference on Computer Vision* 3–19 (Springer, 2018). http://doi.org/10.1007/978-3-030-01234-2_1.
 51. Oktay O, Schlemper J, Le Folgoc L et al. Attention U-net: learning where to look for the pancreas. arXiv (2018-04-11). <https://arxiv.org/abs/1804.03999>
 52. Zhang H, Goodfellow I, Metaxas D et al. Self-attention generative adversarial networks. In *Proceedings of the 36th International Conference on Machine Learning* 7354–7363 (PMLR, 2019).
 53. Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. In *Proceedings of the 2010 20th International Conference on Pattern Recognition* 2366–2369 (IEEE, 2010). <http://doi.org/10.1109/ICPR.2010.579>.
 54. Zhang R, Isola P, Efros AA et al. The unreasonable effectiveness of

deep features as a perceptual metric. In *Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 586–595 (IEEE, 2018). <http://doi.org/10.1109/CVPR.2018.00068>.

Acknowledgements

We are grateful for financial supports from the National Natural Science Foundation of China (NSFC) (62675042).

Author contributions

Qihang Liang conceived the original idea, developed the methodology and reconstruction framework, implemented the algorithms, conducted the experiments and validation, analyzed the results, prepared the figures, and wrote the original manuscript. Xin Xu contributed to figure optimization, experimental analysis, and manuscript revision. Jia Wu and Zichao Wei assisted with experimental implementation, figure preparation, and technical discussions. Jinyong Wang contributed to manuscript improvement, technical discussions, and constructive suggestions on the early version of the work. Kun Li participated in manuscript discussion. Shovit A. Gupta and Fake Lu provided theoretical discussions, valuable insights into the presentation and positioning of the methodology, and contributed to manuscript refinement, academic communication, and the overall improvement of the manuscript quality from an international collaborative perspective. Xiaoxi Chen and Pinghe Wang supervised the research process, provided experimental guidance, and contributed extensively to manuscript revision and project coordination. Guohua Shi supervised the overall project, provided strategic guidance for the study and publication process, and finalized the manuscript for submission. All authors reviewed and approved the final manuscript.

Competing interests

The authors declare no competing financial interests.

Data availability

The neural network architecture, the proposed algorithms, and part of the data used in this study have been made publicly available at <https://github.com/QHANG-LIANG/AR-SPI>.



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

©The Author(s) 2026.

Published by Editorial Office of *Opto-Electronic Advance*, Institute of Optics and Electronics, Chinese Academy of Sciences.

