



Improving NIR single-pixel imaging: using deep image prior and GANs

CARLOS OSORIO QUERO,^{1,*} IRVING RONDON,² AND JOSE MARTINEZ-CARRANZA^{1,3}

¹Electronics Department in the Digital Systems Group, Instituto Nacional de Astrofísica Óptica y Electrónica, 72810 Puebla, Mexico

²School of Computational Sciences, Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

³carranza@inaoep.mx

*caoq@inaoep.mx

Received 9 September 2024; revised 27 November 2024; accepted 1 December 2024; posted 2 December 2024; published 27 January 2025

We introduce a hybrid approach that combines deep image prior (DIP) with generative adversarial networks (GANs) to improve the resolution of single-pixel imaging (SPI). SPI excels in challenging conditions such as low light or limited spectral camera availability, particularly in the near-infrared (NIR) range from 850 to 1550 nm. By employing an unsupervised image super-resolution technique based on DIP, we reduce the need for extensive direct SPI image datasets. This innovation simplifies enhancing image quality in specific NIR bands. We provide numerical and experimental evidence to support our method and detail the enhancements in UNet and GAN architectures across four neural network configurations. © 2025 Optica Publishing Group. All rights, including for text and data mining (TDM), Artificial Intelligence (AI) training, and similar technologies, are reserved.

<https://doi.org/10.1364/JOSAA.541763>

1. INTRODUCTION

Single-pixel imaging (SPI) is an advanced imaging technology that captures images using only a single photodetector [1]. This innovative technique utilizes a series of spatial modulation patterns to acquire images [2] meticulously. SPI has proven to be highly versatile, adapting seamlessly to various spectrum bands across various applications, as documented in the literature. These applications include 2D reconstruction in bands such as X-rays [3], ghost imaging [4,5], infrared [6], terahertz [7], underwater imaging [8], imaging in scattering environments [9], and three-dimensional imaging [10].

In SPI, achieving a high-quality image typically requires acquiring $N \times N$ measurements to achieve a full sampling rate of 100%. However, current limitations exist due to the constraints of commercial spatial light modulators (SLMs), such as the digital micromirror device (DMD), with a frequency range of 10–30 kHz [2], or array LEDs, which allow projection frequencies greater than 100 kHz [11]. These limitations pose a bottleneck for SPI applications, where the objective is to implement reconstruction methods with fewer samples to obtain low-resolution images. The primary challenge in SPI lies in developing strategies that can produce images of acceptable quality while operating at a reduced sampling rate [12]. Numerous studies have made significant efforts to address the challenge of image recovery. For instance, researchers have leveraged the inherent sparsity in natural scenes and applied techniques like compressed sensing (CS). Moreover, the use of orthogonal sub-sampling methods, such as the deterministic Hadamard [13], Fourier [14], wavelet bases [15], and deep

learning [2], has shown promise in reconstructing high-quality images, even below the Nyquist limit [16].

Due to the inherent low quality, artifacts, and noise present in SPI images, deep learning techniques offer promising solutions for their enhancement, including denoising [17], deblurring [18], and super-resolution [19]. Various models have been employed to tackle these challenges, such as convolutional neural networks (CNNs) [20], generative adversarial networks (GANs) [21], and deep convolutional autoencoder networks (DCANs) [22,23]. However, collecting dataset images from various visible spectrum bands presents significant challenges, making traditional training-based methods less feasible. Unsupervised image super-resolution techniques [24] offer a new perspective on improving image quality without the need for extensive training datasets. A particularly notable method in this category is deep image prior (DIP) [25], which uses untrained neural networks as a flexible form of image prior.

In DIP, an image is represented as the output of a neural network, and an optimization process refines the network's weights to minimize the discrepancy between the predicted image and the observed measurements. This approach exploits the inherent structure of CNNs as an image prior, eliminating the need for ground-truth images and large datasets typically required for neural network training. As a result, DIP produces refined, detailed, and more natural-looking images with enhanced visual quality. DIP has proven effective in various image-related inverse problems, including super-resolution [25], inpainting [26], deblurring [27], compressive sensing [28], phase retrieval [29], and computer-generated holography [30]. While DIP

effectively enhances resolution, sharpness, and overall image appearance, it faces challenges when dealing with noisy low-resolution images. Recent advancements have addressed this issue by integrating noise estimation techniques with DIP, utilizing GANs and self-supervised learning (SSL) [31].

To improve the quality of 2D SPI image reconstruction in the near-infrared (NIR) imaging application spectrum, we comprehensively evaluated various strategies. The wavelength range of focus was 850–1550 nm. The review was conducted, taking into account limitations imposed by the camera design. We incorporated the deep prior method and employed neural networks such as U2Net, GAN, StyleGAN, and StyleGAN2 [32–34]. Furthermore, we developed a novel architecture that combines different UNet models with GAN and StyleGAN components. To assess the improvement in 2D image quality, we study rigorous metrics such as a peak signal-to-noise ratio (PSNR) greater than 20 dB, a structural similarity index measure (SSIM) exceeding 0.5, and a Fréchet inception distance (FID) below 40, as detailed in Ref. [35]. Our results show valuable insights into the performance, computational efficiency, and quality of the reconstructed 2D SPI images [11]. The main contributions can be summarized as follows:

- A super-resolution deep prior model for SPI is proposed, based on the combination of UNet and GAN networks [21,36]. The performance of the model is validated through both simulations and real experiments.
- A novel end-to-end SPI reconstruction network has been developed to more effectively capture features in SPI measurements, leading to the reconstruction of high-quality images from under-sampled measurements.
- By utilizing a deep prior model, a high-resolution (HR) SPI image is generated, enhancing the quality of the low-resolution (LR) SPI image.

The paper is structured as follows: Section 2 provides a detailed discussion of the proposed SPI reconstruction method, including an overview of each system component. Section 3 outlines the methodology for super-resolution using deep prior. Section 4 presents the integration of DIP with GANs for enhanced SR. The experimental results and discussions can be found in Section 5. Finally, Section 6 concludes the paper.

2. FUNDAMENTAL CONCEPTS IN SINGLE-PIXEL IMAGE RECONSTRUCTION

A SPI camera captures images by lighting a scene with carefully structured light patterns. Simultaneously, it measures the corresponding intensity correlations using a bucket detector without spatial resolution. One crucial component of the SPI camera is the utilization of SLMs, such as the DMD, as illustrated in Fig. 1. SPI technology offers two distinct architectural approaches: structured detection, as depicted in Fig. 1(a), and structured illumination, as showcased in Fig. 1(b) [2].

Structured detection is a process that involves illuminating an object with light from a source. The reflected light is then directed onto a SLM. Detection is carried out using a bucket detector. Structured illumination, conversely, involves using modulated light patterns, such as Hadamard sequences, random patterns, or other variations, denoted as pattern modulation

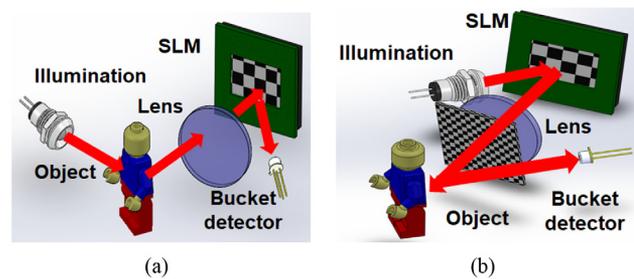


Fig. 1. Two different approaches applied to SPI: (a) structured detection and (b) structured illumination [2].

Φ [2]. This modulated light is used to illuminate the object, O (which can be defined as a 2D matrix $M \times N$), and a bucket detector detects the reflected light. The bucket detector converts the reflected light into an electrical signal, S_i [as shown in Eq. (1)] [2]:

$$S_i = \alpha \sum_{x=1}^M \sum_{y=1}^N O(x, y) \Phi_i(x, y), \quad (1)$$

where the constant factor α depends on the optoelectronic response of the photodetector. The image $I(x, y)$ is computationally reconstructed from the captured signal S_i and the corresponding pattern Φ_i , as represented in Eq. (2) [2]:

$$I(x, y) = \alpha \sum_{x=1}^M \sum_{y=1}^N S_i \Phi_i(x, y). \quad (2)$$

In our application, we employed Hadamard-like pattern [37] sequences generated through active illumination using an array of 32×32 near-infrared light emitting diodes (NIR-LEDs) that emit radiation with a peak wavelength of 1550 nm for this study. The NIR-LED array is perpendicular to the lens's focal length to project the light pattern to infinity. However, due to the array's size, the patterns are projected at distances ranging from 0.3 to 3 m.

A. Experimental Setup: SPI Camera

Our research proposes employing structured illumination to enhance image quality in challenging lighting conditions, such as solid backlight and stray light interference. To accomplish this, we leverage a time-of-flight (TOF) system operating at a wavelength of 850 nm in conjunction with an InGaAs photodiode as the bucket detector, which operates at 1550 nm. The architecture introduced in this study, named NIR-SPI, comprises two primary components. First, we utilize critical elements based on the single-pixel principle for image generation. These components consist of the InGaAs photodiode, specifically the *Thorlabs FGA015* diode operating at 1550 nm, an array of NIR-LEDs for emission, a time-of-flight system, and an analog-to-digital converter (ADC) (see Fig. 2). Second, we integrate a subsystem responsible for processing the electrical output signal acquired from the bucket detector. The signal is digitized using the ADC, and the resulting data are processed by an embedded system-on-module (SOM) [38], specifically the GPU-Jetson Xavier NX as depicted in Fig. 2. The SOM carries

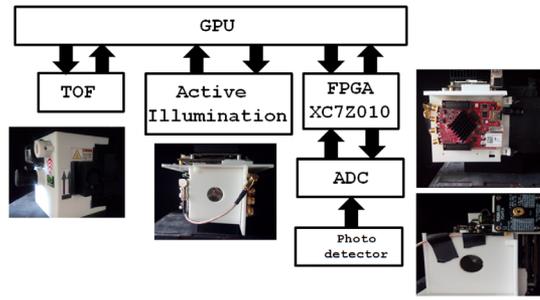


Fig. 2. Comprehensive block diagram of the proposed vision system measures 11 cm × 11 cm × 14 cm in size. It encompasses multiple components, including a lens with a 20 cm focal length. The system has a weight of 1.2 kg and a power consumption of 45 W. In the initial stage module, three key elements are present: a photodiode, an active illumination source, and an InGaAs photodetector diode (FGA015), which is employed for the time-of-flight (TOF) system, as detailed in Ref. [2].

out several tasks, including generating Hadamard-like patterns and processing the digitized data from the ADC. We implement the OMP-GPU algorithm [39] on the SOM to facilitate 2D image generation. Additionally, we provide the processing times for each stage involved in the 2D image reconstruction process. For a more in-depth exploration of the SPI camera, we recommend interested readers to consult Ref. [11].

B. 2D SPI Algorithm Reconstruction Method

The reconstruction of a 2D image begins with an SPI image reconstructed at low sampling rates to maintain processing times within the range of 50–80 ms. This low-resolution SPI image is then improved by incorporating data from ITOF using a technique outlined in a previous study [39]. Initially, the orthogonal matching pursuit (OMP) algorithm is employed to reconstruct the SPI image. The OMP algorithm utilizes Cholesky methods to efficiently compute $(\Phi^T \Phi)^{-1}$ for matrix inversion. During the Cholesky factorization process [see Eq. (3)], it is imperative to precompute the Gram matrix $G = \Phi^T \Phi$ and ensure its symmetry and positivity. Subsequently, the initial projection $p^0 = \Phi^T y$ is calculated following the steps outlined in the OMP-GPU algorithm, specifically at line 3:

$$L_{\text{new}} = \begin{bmatrix} L & 0 \\ w^T & \sqrt{1 - w^T w} \end{bmatrix}. \quad (3)$$

Equation (3) represents the Cholesky decomposition as a factorization of a Hermitian, positive-definite matrix into the product of a lower triangular matrix and its conjugate transpose. This matrix can be broken down into two triangular matrices through the Cholesky decomposition, represented as LL^T , where L is the lower triangular Cholesky factor (refer to Algorithm 1, line 8). We describe the system as $LL^T x = \Phi^T \bar{y}$, where $b = \Phi \bar{y}$. This setup is then approached as a triangular system using $Lu = b$ and $L^T x = u$ (refer to Algorithm 1, line 10). To derive L , we apply the equation in Eq. (3) [39], and define $w = L^{-1} \Phi^T$ (see Algorithm 1, line 7). The reconstructed signal x_i implements a stopping criterion that involves comparing the residual norm to a threshold ε , thus bypassing the direct calculation of the residual δ (refer to Algorithm 1, lines

Algorithm 1. OMP-GPU algorithm [39], Input: OMP-GPU algorithm input data: Hadamard patterns Φ , input signal y , Output: OMP-GPU algorithm output data: sparse representation x that fulfils the relation $y \approx \Phi x$

```

1: procedure OMP-GPU( $\Phi, y$ ):
2:   set:  $L = [1], i = 1$ 
3:   set:  $p^0 = \Phi^T y, \varepsilon = yy^T, G = \Phi^T \Phi, p = p^0$ 
4:   while  $\varepsilon_{i-1} > \varepsilon$  do
5:      $M = \arg \max_K |p|$  ▷ Finding the new atom
6:     if  $M > 1$  then
7:        $w = \{L_{i-1} w = G_{i-1, K}\}$  ▷ Solver  $w$ 
8:        $L_i = \begin{bmatrix} L_{i-1} & 0 \\ w^T & \sqrt{1 - w^T w} \end{bmatrix}$  ▷ Update of Cholesky
9:
10:       $x_i = \{LL^T x_i = p^o\}$  ▷ Solver  $x_i$ 
11:       $\beta = G x_i$  ▷ Matrix-sparse-vector product for each path
12:       $p = p^o - \beta$ 
13:       $\delta^i = x^T \beta$  ▷ Calculate error
14:       $\varepsilon^i = \varepsilon^{i-1} - \delta^i + \delta^{i-1}$  ▷ Calculate norm  $\varepsilon$ 
15:       $i = i + 1$  ▷ increasing iteration
16:   return  $x$ 

```

11–13). To improve the algorithm efficiency, the use of Compute Unified Device Architecture (CUDA) is recommended for parallelizing the reconstruction steps (see Algorithm 1). This enhanced image is then combined with the ITOF sensor data to produce an image of 32×32 .

C. SPI Imaging Acquisition Protocol

In developing the SPC camera, we focused on two critical parameters essential for capturing SPI images: the exposure time of the detector T_{ext} and the frequency of pattern projection F_{patterns} . We utilized a theoretical model of the NIR-SPI system outlined in Ref. [11] to set the appropriate exposure time. This model considers various factors, such as the maximum measurement distance, scattering effects, and correlation between photon incidence on the sensor and the noise threshold. We established the exposure time, denoted as T_{ext} , to vary between 80 and 120 μs , which is optimal for measurement distances ranging from 0.3 to 1 m. From this exposure time, we derived that the minimum frequency for the ADC must be at least 60 kHz. The frequency patterns are based on Eq. (4) [40], employing the parameter F_{min} to assess efficiency at the individual pixel level. The ideal configuration occurs at $F = F_{\text{min}}$ (where F denotes the actual pixel count of the sensor), facilitating the highest ADC measurement rate at the lowest sensor resolution F_{ADC} . This setup significantly improves the signal-to-noise ratio under outdoor conditions. When the design condition falls below $F < F_{\text{min}}$, the frequency patterns limit the measurement resolution. Conversely, exceeding this value ($F > F_{\text{min}}$), the pattern generation frequency will be defined between the range of $F_{\text{patterns}} = 40$ kHz and $F_{\text{ADC}} = 125$ MHz, allowing us to obtain an improved measurement rate of three times faster:

$$F_{\text{min}} = \frac{F_{\text{ADC}}}{F_{\text{patterns}}}. \quad (4)$$

3. METHODS APPLYING DEEP IMAGE PRIOR TO SUPER-RESOLUTION IN SPI

The proposed super-resolution (SR) technique primarily focuses on the task of reconstructing an image of high resolution (HR) from a low-resolution (LR) SPI image. When employing deep learning for an SR image, we achieve image generation by training generator/decoder networks, denoted as $\hat{x} = f_{\theta}(z)$, where z represents a random code vector and $\hat{x} \in \mathbb{R}^{H \times W}$ is the resulting image. Here, θ denotes the network parameters, and H and W represent the image's dimensions. This network consists of convolution, upsampling, and nonlinear activation functions to generate the HR image, as illustrated in Eq. (5) [29], where various factors are optimized to improve the super-resolution output [29]:

$$\hat{x}^* = \min_{\hat{x}} E(\hat{x}; \hat{x}_0) + U(\hat{x}), \quad (5)$$

where the function $E(x; x_0)$ is defined as

$$E(\hat{x}; \hat{x}_0) = \|d(\hat{x}) - \hat{x}_0\|^2, \quad (6)$$

where $d(\cdot): \mathbb{R}^{tH \times tW} \rightarrow \mathbb{R}^{H \times W}$ represents the downsampling operator that resizes an image by factor t , while x_0 denotes a low-resolution image. The term $U(x, y) = \sum(\nabla_x I(x, y) + \nabla_y I(x, y))$ represents the total variation (TV) of the image $I(x, y)$ [41,42], which is utilized in single-image super-resolution for the convergence of TV, encouraging solutions to contain uniform regions [43]. The objective is to find the high-resolution image \hat{x} that, when downsampled, matches the low-resolution image \hat{x}_0 . The minimization θ^* is obtained using an optimizer such as gradient descent, with parameters initialized randomly. After minimization, the restored high-resolution image \hat{x}^* is received and updated iteratively during restoration.

4. INTEGRATING DIP WITH GANS FOR ENHANCED SUPER-RESOLUTION

In the context of GANs [44], the SR technique relies on the integration of a generator and a discriminator. The generator network, denoted as G , creates the high-resolution image from a low-resolution input. The discriminator network, denoted as D , aims to distinguish between real high-resolution images and those generated by the generator. The generator can be expressed as $\hat{x} = G_{\theta_G}(z)$, where θ_G represents the parameters of the generator network, and z is a random input vector. The discriminator can be expressed as $D_{\theta_D}(\hat{x})$, where θ_D represents the parameters of the discriminator network. The discriminator is trained to maximize the following objective function [45]:

$$L_D = \mathbb{E}[\log D_{\theta_D}(x_{\text{real}})] + \mathbb{E}[\log(1 - D_{\theta_D}(\hat{x}))], \quad (7)$$

where x_{real} represents real SR images, while \hat{x} denotes the images generated by the generator. In the process of discrimination between real or fake, the discriminator assigns a high probability (close to 1) to the term $\log D_{\theta_D}(x_{\text{real}})$, indicating that they are indeed real, and assigns a low probability (close to 0) to the term $\log(1 - D_{\theta_D}(\hat{x}))$, indicating that they are fake. During training, it iteratively updates its parameters θ_D to maximize L_D , enhancing its ability to differentiate real images from generated

ones. The GAN framework merges the objectives of both the generator and the discriminator into a single min-max optimization problem, enhanced by the integration of DIP, as shown in Eq. (8):

$$\min_{\theta_G} \max_{\theta_D} L_D + E(\hat{x} - \hat{x}_0) + U(\hat{x}). \quad (8)$$

The optimization in Eq. (8) allows for the adjustment of the generator network's weights based on the combined objective, incorporating the prior knowledge encoded in the DIP framework.

A. Image Enhancement through Deep Learning Architectures

To enhance the processing of SR images through neural networks, various model types were considered, including UNet, U2Net, Res-U2Net, and combinations of GAN with StyleGAN or StyleGAN2. These models were trained using the PyTorch framework in Python, with the Adam optimizer and a stopping criterion set below 10^{-4} . Initially, we focus on the implementation and optimization of UNet, U2Net, and Res-U2Net models in the generator stage:

- **Case 1:** The UNet model was utilized for SR-SPI [see Fig. 3(a)]. The model's architecture was customized to handle varying resolutions, featuring interconnected downsampling and upsampling blocks. The model consisted of two main components:

1. **Downsampling:** The encoder section utilized convolutional layers and max-pooling operations to decrease the spatial dimensions of the input image, with a sequential resolution reduction from 32, 64, 128, to 256.
2. **Upsampling:** The decoder portion gradually enhanced the spatial resolution of the feature maps.

- **Case 2:** The U2Net model was utilized for super-resolution tasks, providing flexibility and configurability [refer to Fig. 3(b)]:

1. **Initialization:** Various parameters were defined for the U2Net class, including the number of input and output channels and feature scaling.
2. **Layer Definitions:** The network was structured with encoding (downsampling) and decoding (upsampling) stages.
3. **Encoding (Downsampling):** This component utilized max-pooling operations, convolutional layers, and filters to reduce the spatial dimensions of the input image.
4. **Upsampling:** Transposed convolutional layers and nearest-neighbor upsampling were employed to increase feature maps' spatial dimensions from 1024 to 128.
5. **Final Layer:** A convolutional layer with a specified number of output channels applied the sigmoid activation function to the output.
6. **Forward Pass:** The input was propagated through the network in a cascading manner, with feature maps from downsampling stages being concatenated.

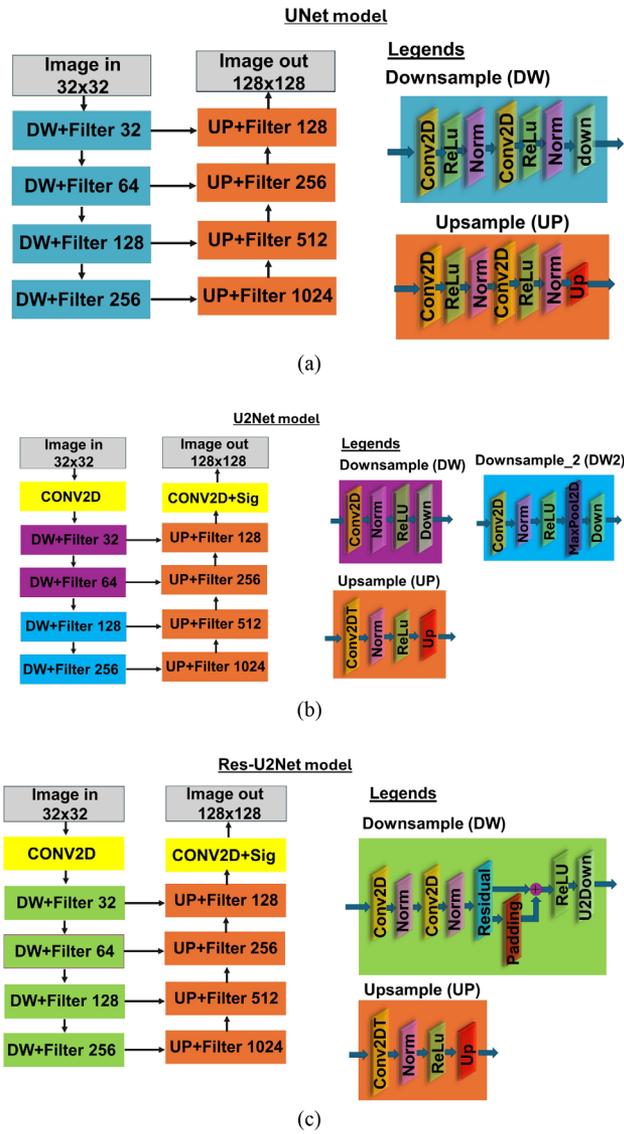


Fig. 3. SR-SPI neural network model architecture for cases 1 to 3: (a) UNet model, (b) U2Net model, and (c) Res-U2Net model.

7. **Output:** The network returned an enhanced-resolution image.

- **Case 3:** The ResU2Net-based architecture was used, offering flexibility to add more layers, concatenate the input image with feature maps, and customize the network for super-resolution image tasks [see Fig. 3(c)]:

1. **Initialization:** Parameters were set to configure the network's architecture, including input and output channels and feature scaling.
2. **Additional Layers:** Extra downsampling and upsampling layers were added to capture finer image details.
3. **Upsampling:** Four upsampling blocks increased spatial dimensions while decreasing channel numbers using sequence filters from 32 to 1024.
4. **Forward Pass:** The input image was processed through the network, with the downsampling results being concatenated.

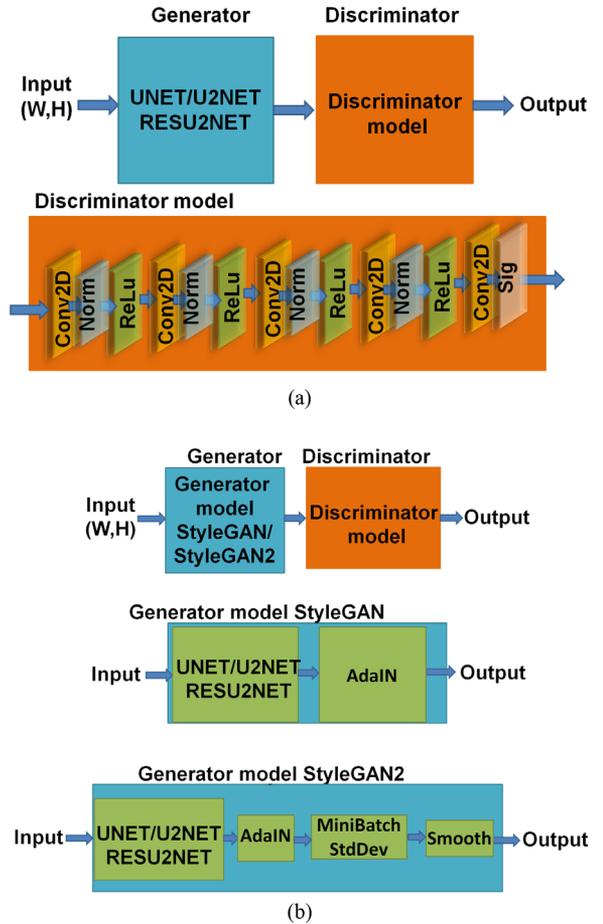
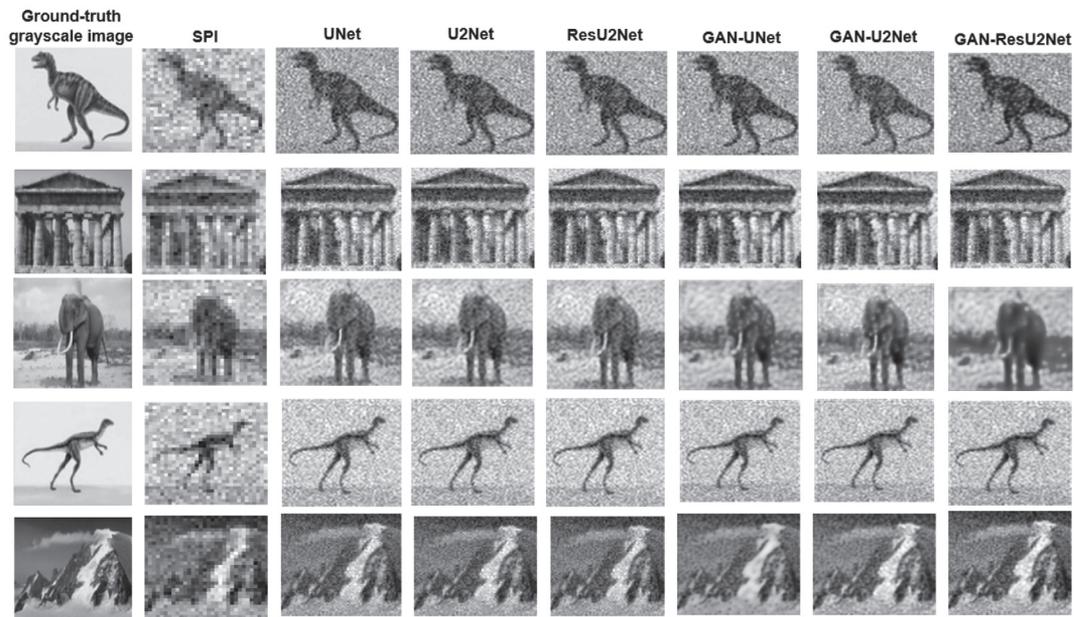


Fig. 4. SR-SPI neural network model architecture for cases 4 and 5: (a) GAN model and (b) StyleGAN/StyleGAN2.

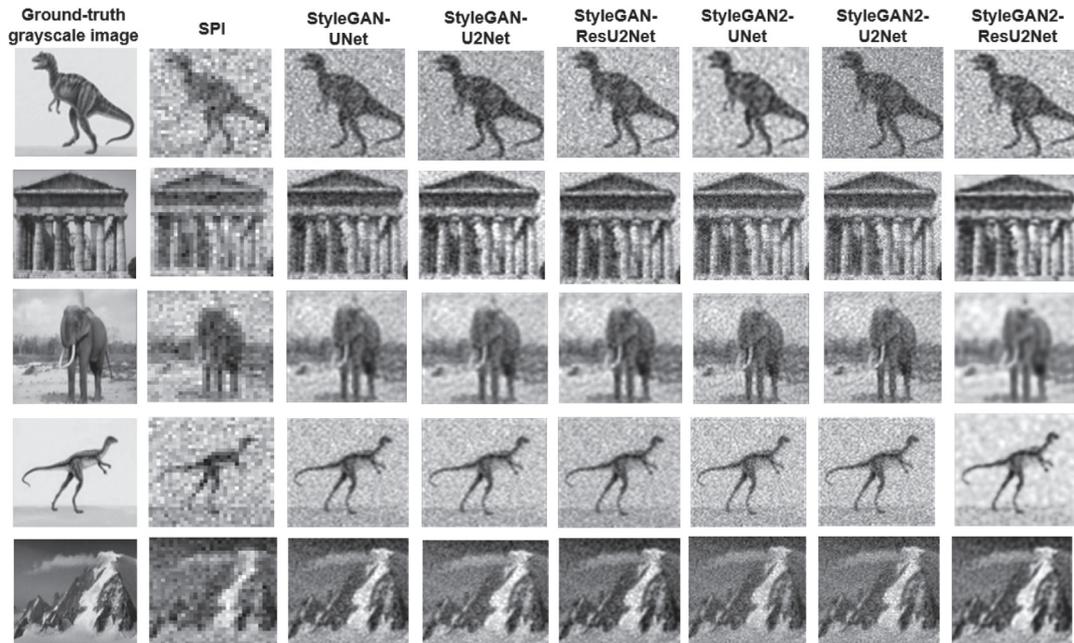
5. **Interpolation:** An interpolation step ensured a minimum input size of 128×128 pixels.
6. **Final Layer:** The final convolutional layer produced the output with the desired number of channels, applying a sigmoid activation function.

For the GAN, StyleGAN, and StyleGAN2 models, we have defined the following structure:

- **Case 4:** This super-resolution model employs a GAN, as illustrated in Fig. 4(a). The model features two primary components: the generator and the discriminator. The generator's role is to enhance low-resolution images by creating high-resolution counterparts, while the discriminator aims to distinguish between genuine and generated high-resolution images. Through adversarial training, these components collaborate to forge high-resolution images that appear exceedingly realistic to the extent that they can deceive the discriminator [see Eq. (7)]. The architecture of the generator incorporates multiple UNets, based on a UNet design, which effectively captures and reconstructs image details. The discriminator, conversely, consists of convolutional layers that assess the authenticity of images, outputting a probability value ranging from 0 to 1.
- **Case 5:** StyleGAN and StyleGAN2 are advanced architectures [refer to Fig. 4(b)] that seamlessly integrate the generator and discriminator components. For this model, we define the style mean as 0.3 and the standard deviation as 0.01, based on



(a)



(b)

Fig. 5. Simulation enhancing SPI reconstruction through deep prior methods using 30% patterns from Corel-1K dataset: (a) Cases 1 to 3 involve UNet, U2Net, and ResU2Net with 1500 iterations, and case 4 utilizes GAN-UNet, GAN-U2Net, and GAN-ResU2Net with 500 iterations. (b) In case 5, StyleGAN is applied with 250 iterations, while StyleGAN2 utilizes 150 iterations.

laboratory testing. During training, these models take as input an image along with style mean and standard deviation values. The generator in StyleGAN produces super-resolution images with unique styles by applying Adaptive Instance Normalization (AdaIN) based on the provided style parameters. StyleGAN2 includes enhancements such as MiniBatchStdDev, improving image quality and discriminator performance. The generator in StyleGAN can have multiple heads, each utilizing a UNet-like architecture, with their outputs combined to produce stylized super-resolution images. The discriminator in StyleGAN and StyleGAN2 employs convolutional layers with leaky ReLU

activation and instance normalization to determine the probability of an image being real or fake. Additionally, StyleGAN2 introduces a smoothing operation to generate smoother images. This process involves using AdaIN to stylize the content image and MiniBatchStdDev to further enhance StyleGAN2.

B. Assessing the Efficacy of SR-SPI Methods and Simulations

The DIP method is a systematic approach to improving SPI image reconstruction. This study examined six cases to

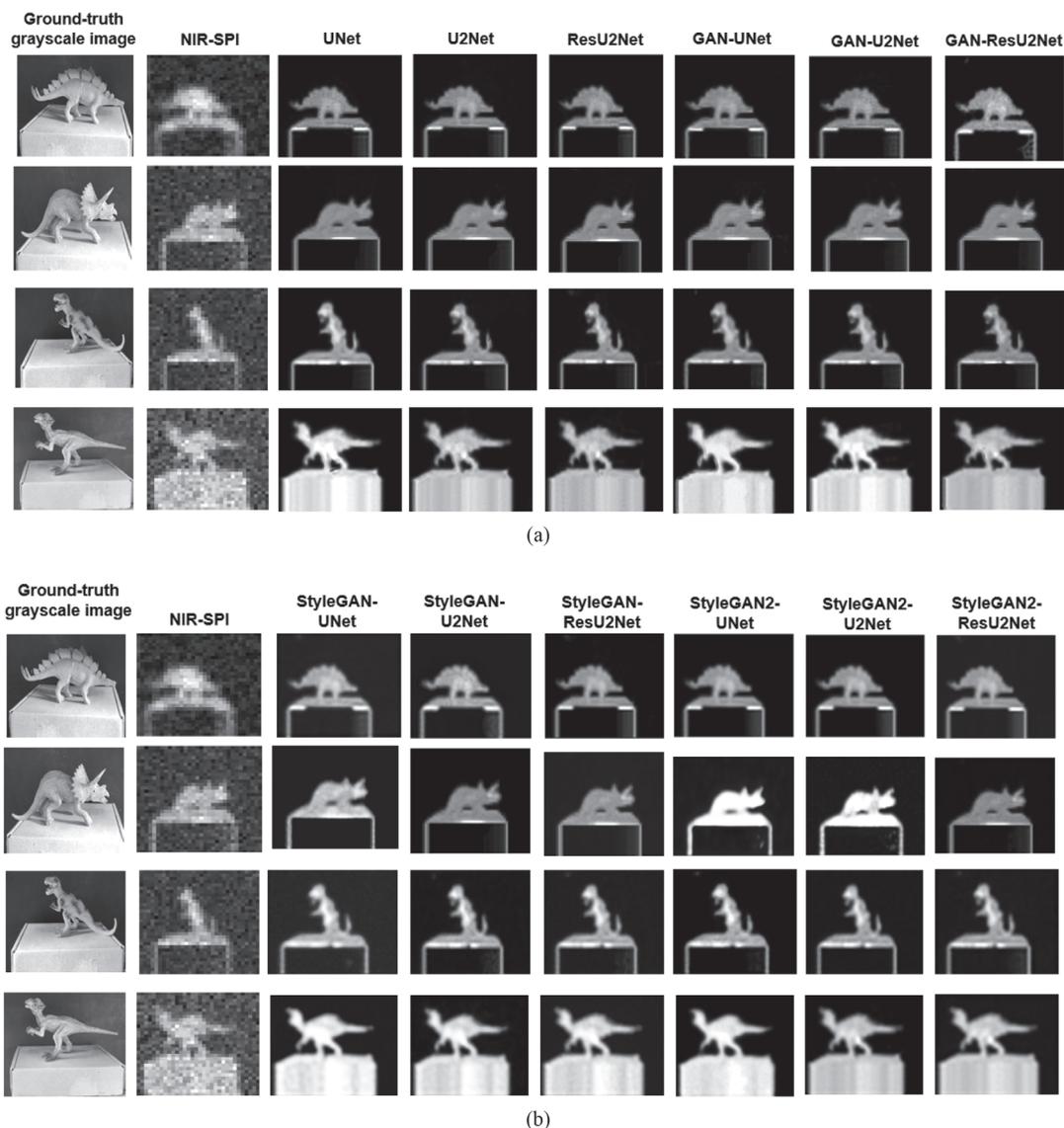


Fig. 6. In our experimental study, we significantly improved the resolution of NIR-SPI images from 32×32 to 128×128 using advanced deep image prior (DIP) techniques. More details can be found in Section 4.A. These models successfully achieved a fourfold increase in resolution at a distance of 50 cm. We test experimentally several network designs, adjusting the number of training iterations accordingly. The UNet and U2Net models were trained for 1500 iterations, while ResU2Net only required 500 iterations. GAN-based models like GAN-UNet, GAN-U2Net, and GAN-ResU2Net needed 500 iterations. The StyleGAN and StyleGAN2 models were trained much faster, completing their training in just 250 and 150 iterations, respectively.

determine the effect of different neural network architectures and iteration counts on image quality improvement. The first four cases involved UNet, U2Net, ResU2Net, GAN-UNet, GAN-U2Net, and GAN-ResU2Net [see Fig. 5(a)]. In the simulation results, we observe artifacts that were removed during testing by applying a combination of low-pass filtering and 2D edge-adaptive filtering [46]. In the test, we determined that this type of modeling can achieve enhanced image quality in terms of PSNR and SSIM using a low number of iterations, with iteration counts ranging from 500 to 1500. The complexity of the neural network was closely observed throughout the simulation, resulting in significant improvements in image quality and

a reduction in the required number of iterations while maintaining high-quality images. In case 5, StyleGAN/StyleGAN2 were used with iteration counts set at 250 and 150, respectively [see Fig. 5(b)]. Our investigation into neural network complexity significantly enhanced image quality and validated the efficacy of DIP methods for SPI image reconstruction. We employed objects from the Corel-1K image dataset [47], resized to 32×32 pixels for our simulations. We conducted performance simulations using 30% Hadamard patterns [1], employing approximately 320 such patterns. These simulations were rigorously assessed using metrics such as PSNR, SSIM, and FID to verify their effectiveness across the various configurations outlined in our study (refer to Fig. 6).

5. RESULTS FROM EXPERIMENTAL VALIDATIONS

We evaluated the performance of our NIR-SPI vision system using a specialized test setup. A comprehensive description of the experimental laboratory configuration can be found in [Supplement 1](#), Section 1. Our main goal was to assess the image reconstruction capabilities of the prototype NIR-SPI system. For this purpose, we tested objects measuring 140×70 mm (refer to Fig. 6) at a distance testing of 50 cm to the NIR-SPI system. Subsequently, we evaluated the enhanced SPI image using the neural network model introduced in Section 4. To validate the NIR-SPI performance using the DIP method in conjunction with various neural networks (see Table 1), we implement two tests, and parameters such as PSNR > 20 dB, SSIM > 0.5, and FID < 20 were evaluated [35].

A. Testing the Enhanced NIR-SPI Imaging Using the SR Deep-Prior Method

We comprehensively evaluated different deep-prior network models in the context of NIR-SPI testing and present our findings in this study (see Fig. 6). Our primary focus was on critical performance metrics and the number of iterations required for convergence. The PSNR metric improved from NN-UNet to StyleGAN2-UNet. StyleGAN2-UNet achieved the highest PSNR score of 22.1 dB, indicating that it can

produce high-quality images compared to other methods. Similarly, SSIM, a measure of structural similarity, consistently improved as we used more advanced models, demonstrating that more profound and sophisticated models create images that closely resemble ground truth images. StyleGAN2-UNet also performed the best, with an SSIM score of 0.64. In contrast, FID, which measures the dissimilarity between generated and authentic images, showed that StyleGAN-ResU2Net outperformed its counterparts with a lower FID value of 29.25. These values suggest that StyleGAN-ResU2Net is better at generating images that closely mimic authentic images (see Table 1). Our analysis also showed that GANs have faster convergence rates than non-GAN models, with lower values indicating speedier convergence. Therefore, StyleGAN2-UNet is best for image quality, while StyleGAN-ResU2Net is best for image fidelity. The choice between these models is based on the task's specific requirements, carefully balancing image quality and fidelity considerations with computational resources and convergence speed. Compared to other GAN-based SR models using deep learning (DL) (see Table 2), this model achieves acceptable SR quality from low-resolution SPI images (see Fig. 7). Its performance is evaluated in scale, PSNR, SSIM, and the number of iterations. Unlike other methods requiring numerous iterations to enhance image quality, this model delivers comparable results without requiring an extensive dataset.

Table 1. PSNR, SSIM, and FID Score Improvements Were Observed in the Images Processed by the NIR-SPI Testing Simulation (Sim) with Dataset Corel-1K and Laboratory (Lab) Using DIP Models with 30% Hadamard Patterns^a

Method	PSNR Sim (dB) ↑	SSIM Sim ↑	FID Sim ↓	PSNR Lab (dB) ↑	SSIM Lab ↑	FID Lab ↓	Iter
NN-UNet	17.90	0.60	45.18	20.36	0.60	46.05	1500
GAN-UNet	18.10	0.61	43.71	20.94	0.62	44.12	500
StyleGAN-UNet	18.23	0.62	42.83	21.96	0.63	40.35	250
StyleGAN2-UNet	18.30	0.63	40.00	22.10	0.64	41.81	150
NN-U2Net	19.46	0.64	36.79	19.92	0.62	27.86	1500
GAN-U2Net	20.50	0.65	35.81	19.50	0.65	31.88	500
StyleGAN-U2Net	21.90	0.67	33.47	20.61	0.67	34.14	250
StyleGAN2-U2Net	22.08	0.69	32.41	21.23	0.68	35.90	150
NN-ResU2Net	22.97	0.72	31.65	21.29	0.69	34.12	1500
GAN-ResU2Net	23.13	0.73	30.78	21.86	0.71	38.61	500
StyleGAN-ResU2Net	25.38	0.75	29.90	22.00	0.72	31.00	250
StyleGAN2-ResU2Net	28.12	0.77	23.83	22.45	0.74	29.25	150

^aThe bold values correspond to the optimized model proposed. This model effectively upscaled images from a resolution of 32×32 to a high resolution of 128×128 , achieving a fourfold increase in image resolution.

Table 2. Numerical Evaluation in Various GANs for Image Super-Resolution Based on Different Training Parameters Including Scale, PSNR, SSIM, Iteration, and Training Methods

Method	Scale	PSNR (dB) ↑	SSIM ↑	Iter	Training
DGAN [48]	$\times 6$	28.62	0.90	25 K	Supervised
ESRGAN [49]	$\times 2$	31.99	0.66	300 K	Supervised
SRGAN [50]	$\times 2$	25.80	0.70	200 K	Supervised
PSSR [51]	$\times 4$	21.32	0.55	300 K	Semi-supervised
CTGAN [52]	$\times 4$	27.90	0.74	60 K	Semi-supervised
DNSR [53]	$\times 2$	26.15	0.70	10 K	Unsupervised
CycleSR [54]	$\times 4$	23.80	0.59	500 k	Unsupervised
Our	$\times 4$	22.45	0.74	150	Unsupervised

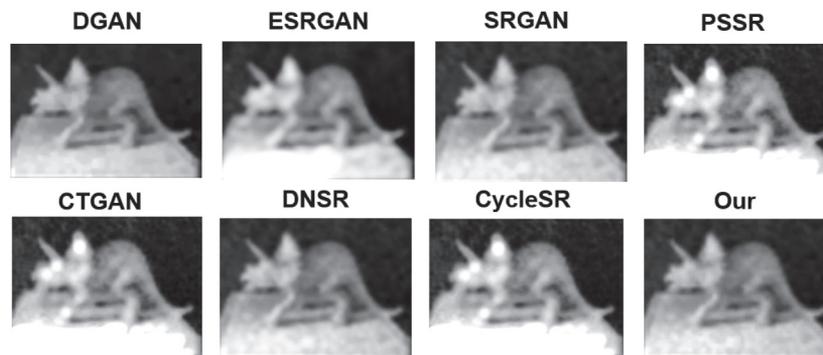


Fig. 7. Comparison of image super-resolution techniques applied to low-resolution SPI images. The figure presents results from various models: DGAN, ESRGAN, SRGAN, PSSR, CTGAN, DNSR, CycleSR, and the proposed model (“Our”). The proposed model demonstrates superior image clarity and detail preservation compared to the other approaches, showcasing its effectiveness in improving resolution in challenging low-light and noisy conditions.

6. CONCLUSION

In this paper, we evaluated the effectiveness of a DIP-based SR method in a NIR-SPI system with active illumination. We studied numerically and experimentally using UNet, U2Net, ResU2Net, GAN, and StyleGAN neural network configurations to identify the optimal setup for enhancing the resolution of SPI imaging. Our results showed that the resolution of NIR-SPI images was successfully increased from 32×32 to a super-resolved 128×128 , achieving a four-fold increase in resolution. Our results significantly enhance image quality for SPI low-resolution applications with a low number of iterations (as demonstrated in Table 2).

We evaluate our results based on key performance metrics such as PSNR, SSIM, FID, and the number of iterations required by the DIP method for each case. Our study suggests that the StyleGAN2 configuration, which uses a generator based on the ResU2Net neural network, outperforms other configurations by striking a remarkable balance between image quality and the number of iterations needed. This research advances our understanding of DIP in SPI and holds practical implications. It provides a significant advantage in contexts where datasets may be limited and when training a network for image quality enhancement is challenging.

Capturing images in low-contrast environments or high-noise conditions presents a challenge for visible light imaging. In contrast, NIR imaging delivers substantially better results in these challenging scenarios, enhancing image detail significantly. This advancement is made possible through the application of the proposed DIP-GAN method introduced in this study. We hope that the presented approach can be extended to study other network architectures, and it opens up promising avenues for future applications and research within SPI imaging applications, as well as new experimental advances in this field.

Funding. Consejo Nacional de Ciencia y Tecnología (Conacyt) (251992).

Acknowledgment. C. O. Q. acknowledges financial support from CONACYT.

Disclosures. The authors declare no conflicts of interest.

Data availability. The codes used in this project can be made available upon reasonable request.

Supplemental document. See Supplement 1 for supporting content.

REFERENCES

- C. O. Quero, D. Durini, J. de Jesús Rangel-Magdaleno, *et al.*, “Emerging vision technology: SPI camera an overview,” *IEEE Instrum. Meas. Mag.* **27**(2), 38–47 (2024).
- C. A. Osorio Quero, D. Durini, J. Rangel-Magdaleno, *et al.*, “Single-pixel imaging: an overview of different methods to be used for 3D space reconstruction in harsh environments,” *Rev. Sci. Instrum.* **92**, 111501 (2021).
- H. Yu, R. Lu, S. Han, *et al.*, “Fourier-transform ghost imaging with hard x rays,” *Phys. Rev. Lett.* **117**, 113901 (2016).
- F. Wang, C. Wang, M. Chen, *et al.*, “Far-field super-resolution ghost imaging with a deep neural network constraint,” *Light Sci. Appl.* **11**, 1 (2022).
- M. Lyu, W. Wang, H. Wang, *et al.*, “Deep-learning-based ghost imaging,” *Sci. Rep.* **7**, 17865 (2017).
- N. Radwell, K. J. Mitchell, G. M. Gibson, *et al.*, “Single-pixel infrared and visible microscope,” *Optica* **1**, 285–289 (2014).
- T. Vasile, V. Damian, D. Coltuc, *et al.*, “Single pixel sensing for THz laser beam profiler based on Hadamard transform,” *Opt. Laser Technol.* **79**, 173–178 (2016).
- M. Le, G. Wang, H. Zheng, *et al.*, “Underwater computational ghost imaging,” *Opt. Express* **25**, 22859–22868 (2017).
- Y. Cao, Y. Xiao, Z. Pan, *et al.*, “Direct generation of 2D arrays of random numbers for high-fidelity optical ghost diffraction and information transmission through scattering media,” *Opt. Lasers Eng.* **158**, 107141 (2022).
- C. Osorio Quero, D. Durini, J. Rangel-Magdaleno, *et al.*, “Single-pixel near-infrared 3d image reconstruction in outdoor conditions,” *Micromachines* **13**, 795 (2022).
- C. O. Quero, D. Durini, J. Rangel-Magdaleno, *et al.*, “Deep-learning blurring correction of images obtained from NIR single-pixel imaging,” *J. Opt. Soc. Am. A* **40**, 1491–1499 (2023).
- F. Sha, S. K. Sahoo, H. Q. Lam, *et al.*, “Improving single pixel imaging performance in high noise condition by under-sampling,” *Sci. Rep.* **10**, 19451 (2020).
- Z. Zhang, X. Wang, G. Zheng, *et al.*, “Hadamard single-pixel imaging versus Fourier single-pixel imaging,” *Opt. Express* **25**, 19619–19639 (2017).
- Z. Zhang, X. Ma, and J. Zhong, “Single-pixel imaging by means of Fourier spectrum acquisition,” *Nat. Commun.* **6**, 6225 (2015).
- H. Dai, G. Gu, W. He, *et al.*, “Adaptive compressed photon counting 3D imaging based on wavelet trees and depth map sparse representation,” *Opt. Express* **24**, 26080–26096 (2016).
- X. Yu, R. I. Stantchev, F. Yang, *et al.*, “Super sub-Nyquist single-pixel imaging by total variation ascending ordering of the Hadamard basis,” *Sci. Rep.* **10**, 9338 (2020).

17. D. B. T. Le, M. Sadinski, A. Nacev, *et al.*, "Deep learning-based method for denoising and image enhancement in low-field MRI," in *IEEE International Conference on Imaging Systems and Techniques (IST)* (2021), pp. 1–6.
18. K. Y. Shin, K. R. Park, B. J. Kang, *et al.*, "Super-resolution method based on multiple multi-layer perceptrons for iris recognition," in *Proceedings of the 4th International Conference on Ubiquitous Information Technologies & Applications* (2009), pp. 1–5.
19. W. Yang, X. Zhang, Y. Tian, *et al.*, "Deep learning for single image super-resolution: a brief review," *Trans. Multim.* **21**, 3106–3121 (2019).
20. R. T. Schirmeister, J. T. Springenberg, L. D. J. Fiederer, *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapp.* **38**, 5391–5420 (2017).
21. S. Sun, Q. Yan, Y. Zheng, *et al.*, "Single pixel imaging based on generative adversarial network optimized with multiple prior information," *IEEE Photonics J.* **14**, 8538110 (2022).
22. S. Rizvi, J. Cao, K. Zhang, *et al.*, "DeepGhost: real-time computational ghost imaging via deep learning," *Sci. Rep.* **10**, 11400 (2020).
23. C. F. Higham, R. Murray-Smith, M. J. Padgett, *et al.*, "Deep learning for real-time single-pixel video," *Sci. Rep.* **8**, 2369 (2018).
24. Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: a survey," *IEEE Trans. Pattern Anal. Mach. Intell.* **43**, 3365–3387 (2021).
25. V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2018), pp. 9446–9454.
26. O. Sidorov and J. Y. Hardeberg, "Deep hyperspectral prior: single-image denoising, inpainting, super-resolution," in *IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)* (2019), pp. 3844–3851.
27. S. H. Jung, T. Bok Lee, and Y. S. Heo, "Deep feature prior guided face deblurring," in *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (2022), pp. 884–893.
28. Y. Peng, H. Tan, Y. Liu, *et al.*, "Structure prior guided deep network for compressive sensing image reconstruction from big data," in *6th International Conference on Big Data and Information Analytics (BigDIA)* (2020), pp. 270–277.
29. Y. Yang, Q. Lian, X. Zhang, *et al.*, "HioNet: deep priors based deep unfolded network for phase retrieval," *Digit. Signal Process.* **132**, 103797 (2023).
30. M. H. Eybposh, N. W. Caira, M. Atisa, *et al.*, "DeepCGH: 3D computer-generated holography using deep learning," *Opt. Express* **28**, 26636–26650 (2020).
31. S. Han, T. B. Lee, and Y. S. Heo, "Deep image prior for super resolution of noisy image," *Electronics* **10**, 2014 (2021).
32. X. Qin, Z. Zhang, C. Huang, *et al.*, "U2-Net: going deeper with nested u-structure for salient object detection," *Pattern Recognit.* **106**, 107404 (2020).
33. Y. Wei and W. Hou, "Blurring kernel extraction and super-resolution image reconstruction based on style generative adversarial networks," *Opt. Express* **29**, 44024–44044 (2021).
34. I. Skorokhodov, S. Tulyakov, and M. Elhoseiny, "StyleGan-V: a continuous video generator with the price, image quality and perks of StyleGan2," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2022), pp. 3616–3626.
35. M. Heusel, H. Ramsauer, T. Unterthiner, *et al.*, "GANS trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS)* (Curran Associates, 2017), pp. 6629–6640.
36. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (Springer International Publishing, 2015), pp. 234–241.
37. K. S. Charan, G. Rochan Ravi, T. N. Shashank, *et al.*, "Image super-resolution using convolutional neural network," in *IEEE 2nd Mysore Sub Section International Conference (MysuruCon)* (2022), pp. 1–7.
38. P. Kang and S. Lim, "A taste of scientific computing on the GPU-accelerated edge device," *IEEE Access* **8**, 208337–208347 (2020).
39. C. O. Quero, D. Durini, R. Ramos-Garcia, *et al.*, "Hardware parallel architecture proposed to accelerate the orthogonal matching pursuit compressive sensing reconstruction," *Proc. SPIE* **11396**, 113960N (2020).
40. J. Wang, M. Gupta, and A. C. Sankaranarayanan, "LiSens- a scalable architecture for video compressive sensing," in *IEEE International Conference on Computational Photography (ICCP)* (2015), pp. 1–9.
41. M. Sakurai, Y. Sakuta, M. Watanabe, *et al.*, "Super-resolution utilizing total variation regularization and a shock filter," in *19th IEEE International Conference on Image Processing* (2012), pp. 2221–2224.
42. L. Deng, Z. Zhou, G. Xu, *et al.*, "Tv2++: a novel spatial-temporal total variation for super resolution with exponential-type norm," *EURASIP J. Wireless Commun. Netw.* **2020**, 223 (2020).
43. K. Sun and S. Simon, "Bilateral spectrum weighted total variation for noisy-image super-resolution and image denoising," *IEEE Trans. Signal Process.* **69**, 6329–6341 (2021).
44. X. Xue, X. Zhang, H. Li, *et al.*, "Research on GAN-based image super-resolution method," in *IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)* (2020), pp. 602–605.
45. S. Liu, Y. Yang, Q. Li, *et al.*, "Infrared image super resolution using GAN with infrared image prior," in *IEEE 4th International Conference on Signal and Image Processing (ICSIP)* (2019), pp. 1004–1009.
46. G. Triantafyllidis, D. Tzovaras, D. Sampson, *et al.*, "A hybrid algorithm for the removal of blocking artifacts," in *Proceedings. IEEE International Conference on Multimedia and Expo* (2002), Vol. **1**, pp. 161–164.
47. J. Li and J. Z. Wang, "Real-time computerized annotation of pictures," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**, 985–1002 (2008).
48. M. Zareapoor, M. E. Celebi, and J. Yang, "Diverse adversarial network for image super-resolution," *Signal Process. Image Commun.* **74**, 191–200 (2019).
49. X. Wang, K. Yu, S. Wu, *et al.*, "ESRGAN: enhanced super-resolution generative adversarial networks," in *Computer Vision—ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, eds. (Springer International Publishing, 2019), pp. 63–79.
50. C. Ledig, L. Theis, F. Huszar, *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017), pp. 105–114.
51. S. Maeda, "Unpaired image super-resolution using pseudo-supervision," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 288–297.
52. X. Jiang, M. Liu, F. Zhao, *et al.*, "A novel super-resolution CT image reconstruction via semi-supervised generative adversarial network," *Neural Comput. Appl.* **32**, 14563–14578 (2020).
53. K. Prajapati, V. Chudasama, H. Patel, *et al.*, "Unsupervised single image super-resolution network (USISResNet) for real-world data using generative adversarial network," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2020), pp. 1904–1913.
54. S. Chen, Z. Han, E. Dai, *et al.*, "Unsupervised image super-resolution with an indirect supervised path," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2020), pp. 1924–1933.