Research Article Vol. 40, No. 8 / August 2023 / Journal of the Optical Society of America A 1491

Journal of the Optical Society of America A

OPTICS, IMAGE SCIENCE, AND VISION

Deep-learning blurring correction of images obtained from NIR single-pixel imaging

Carlos Osorio Quero,[1,*] Daniel Durini,[1] Jose Rangel-Magdaleno,[1] Jose Martinez-Carranza,[2] and Ruben Ramos-Garcia[3]

[1]Electronics Department in the Digital Systems Group, Instituto Nacional de Astrofísica Óptica y Electrónica, 72810 Puebla, Mexico
[2]Computer Department, Instituto Nacional de Astrofísica Óptica y Electrónica, 72810 Puebla, Mexico
[3]Optics Department, Instituto Nacional de Astrofísica Óptica y Electrónica, 72810 Puebla, Mexico
*caoq@inaoep.mx

Received 28 February 2023; revised 23 June 2023; accepted 24 June 2023; posted 26 June 2023; published 12 July 2023

In challenging scenarios characterized by low-photon conditions or the presence of scattering effects caused by rain, fog, or smoke, conventional silicon-based cameras face limitations in capturing visible images. This often leads to reduced visibility and image contrast. However, using near-infrared (NIR) light within the range of 850–1550 nm offers the advantage of reduced scattering by microparticles, making it an attractive option for imaging in such conditions. Despite NIR's advantages, NIR cameras can be prohibitively expensive. To address this issue, we propose a vision system that leverages NIR active illumination single-pixel imaging (SPI) operating at 1550 nm combined with time of flight operating at 850 nm for 2D image reconstruction, specifically targeting rainy conditions. We incorporate diffusion models into the proposed system to enhance the quality of NIR-SPI images. By simulating various conditions of background illumination and droplet size in an outdoor laboratory scenario, we assess the feasibility of utilizing NIR-SPI as a vision sensor in challenging outdoor environments.

https://doi.org/10.1364/JOSAA.488549

1. INTRODUCTION

Object detection under bad-weather conditions is a fundamental computer vision task widely used in autonomous robot systems, including self-driving vehicles [1] and autonomous drones [2]. In recent years, vision systems based on RGB sensors [3] that operate in the visible (VIS) part of the spectra (with wavelengths between 400 and 700 nm) have become an essential element for autonomous navigation systems, but they can be affected by scattering particles in the atmosphere, drastically reducing the quality and the depth of visibility [4]. Consequently, alternative vision systems are required to work under different weather conditions [5,6].

In rainy environments, the interaction of light with raindrops causes absorption, reflection, and scattering [7]. The light scattering is stronger in the VIS spectrum than in the near-infrared (NIR) part of the spectrum [8]. Due to this fact, image sensors' imaging and ranging capabilities under scattering conditions are degraded [3]. Therefore other technologies have been used as redundant sensors to mitigate the low performance of the vision sensors. Such is the case of silicon-based light detection and ranging (lidar) [9,10], typically using NIR radiation with 850 and 905 nm wavelengths, respectively, or radio detection and ranging (radar) [11].

In the spectral range between 0.8 and 15 μm [12], incident light can pass through the water droplets with less scattering (Mie effect [7,13]); these advantages mean that the IR bands can be exploited to be integrated into vision technologies as time of flight (TOF) and single-pixel imaging (SPI) for applications in an environment with scattering [14].

For applications where high spatial resolutions in both 2D and ranging are not crucial, a SPI vision system can be employed. This system utilizes active illumination in the NIR wavelength range of 850–1500 nm, using single InGaAs photodetectors [15]. NIR-SPI offers a viable solution that can be effectively used, even in scattering conditions. However, the captured NIR images may suffer from degradation and blurring. To address this issue and enhance image quality, a diffusion model neuronal network based on the UNet architecture has been proposed [16,17]. This network employs a training process where a reference image is input, and the image is gradually degraded to simulate the level of degradation caused by water droplets.

Considering the inherent symmetry of the diffusion network [16,18,19], it is possible to reverse the blurring effects by applying a deblurring process [20,21], thereby restoring the image to its optimal quality. Motivated by this observation, our work aims to propose the following approach:

- We present a novel NIR-SPI vision system designed to reconstruct images effectively in rainy conditions. Our aim is to

evaluate the performance of the NIR-SPI system in comparison to an RGB camera operating under identical conditions.

- We propose using an array of NIR-LEDs to substitute for the spatial light modulator (SLM) commonly used in SPI.
- We propose using diffusion networks to reconstruct the blurred image by applying a deblurring process.

2. SINGLE-PIXEL IMAGE RECONSTRUCTION

An SPI camera produces images by interrogating a scene with a sequence of spatially structured light patterns while measuring the correlated intensity on a detector without spatial resolution.

A key element of the SPI camera is the use of spatial light modulators (SLMs) such as digital micromirror devices (DMDs), as shown in Fig. 1. SPI is available in two architectures: structured detection Fig. 1(a), and structured illumination Fig. 1(b) [22].

In the case of structured detection, the object is illuminated by a light source, the reflected light is projected onto an SLM, and the light is captured by a bucket detector. In the case of structured illumination, the light source is spatially modulated by the SLM illuminating the object [see Fig. 1(b)]; the reflected light is again detected.

The correlation of the light spatial pattern Φ and the light reflected from the object O is equivalent to an inner product, i.e., an electrical signal y_i. Thus, projecting a sequence of $N \times M$ spatial patterns allows us to obtain a sequence of an electrical signal given by Eq. (1) [15],

$$y_i = \alpha \sum_{i=1}^{M} \sum_{j=1}^{N} O(i, j)\Phi(i, j), \tag{1}$$

where α is a constant factor that depends on the optoelectronic response of the photodetector. The image x_i is computationally reconstructed from the captured signal y_i, and the corresponding pattern Φ according to Eq. (2) [15],

$$x_i = \alpha \sum_{i=1}^{M} \sum_{j=1}^{N} y_i \Phi(i, j). \tag{2}$$

A. Generation of the Hadamard-like Patterns Using Active Illumination

In this study, we introduce a novel method utilizing structured illumination. The illumination is achieved by employing an array of 32 × 32 NIR-LEDs that emit at a wavelength of 1550 nm. The choice of this wavelength is motivated by its reduced scattering and absorption coefficients in water. The array of NIR-LEDs is strategically positioned perpendicular to the lens's optical axis, allowing for the projection of light patterns. However, due to the array's size, the patterns are projected within a range of 0.3–3 m. One drawback of active illumination is that it does not fully illuminate the object, potentially impacting the quality of the resulting image reconstruction. Nevertheless, we address this issue by utilizing the fast superresolution convolutional neural network (FSRCNN) technique [23], allowing us to reconstruct high-quality images. This active illumination approach offers numerous advantages, including its ability to operate effectively under various outdoor weather conditions, low-level illumination (scenarios involving dust, fog, rain, or smoke), and reduced sensitivity to background radiation noise [24]. Moreover, our proposed configuration requires fewer optical elements and is more cost-effective. Additionally, it offers the potential for a significantly higher modulation rate, as no moving parts are involved.

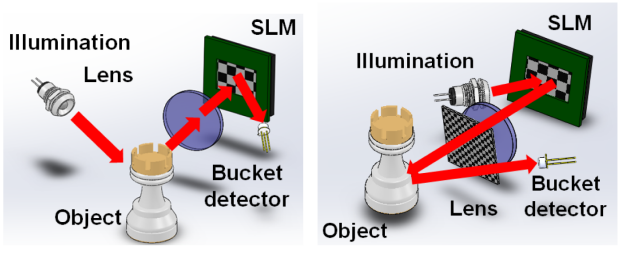

Fig. 1. Two different approaches applied to SPI. (a) Structured detection and (b) structured illumination [22].

B. SPI Camera

The NIR-SPI architecture presented in this study consists of two main components. First, we generate images based on the single-pixel principle. These SPI elements include an InGaAs photodetector (Thorlabs FGA015 diode operating at 1550 nm), an array of NIR-LEDs (1550 nm) for active illumination, a time-of-flight system, and an analog-to-digital converter (ADC) [see Fig. 2(a)]. Second, we introduce a subsystem responsible for processing the electrical output signal obtained from the bucket detector. The signal is digitized using the ADC, and the resulting data is processed using an embedded system-on-module (SOM) [25], specifically the GPU–Jetson Xavier NX shown in Fig. 2(b). The SOM performs multiple tasks, such as generating Hadamard-like patterns and processing the digitized data from the ADC. The orthogonal matching pursuit (OMP-GPU) algorithm [26] is implemented on the SOM, enabling the generation of 2D images (Table 1 presents the processing time for each stage involved in the 2D image reconstruction process). For further details on the SPI camera, we refer interested readers to [4].

C. 2D Reconstruction Algorithm

After acquiring the electrical signal y_i using an ADC, we utilized the OMP algorithm 1 to extract the image x_i. The OMP algorithm requires solving the equation $|y_i - \Phi(i, j)x_i| < \varepsilon$ [22]. To improve the efficiency of reconstructing the 2D SPI image using the OMP algorithm on a GPU unit, we employed the Cholesky method for matrix inversion as defined in [27,28]. This method required us to precalculate the symmetric and positive Gram matrix, defined as $G_i = \Phi^T\Phi$ [26]. Additionally, we carried out an initial projection $p^0 = \Phi^T y_i$ [28] (see Algorithm 1, line 3). This projection was performed to facilitate the implementation of the Cholesky method,

$$L_i = \begin{bmatrix} L_{i-1} & 0 \\ w_i^T & \sqrt{1 - w_i^T w_i} \end{bmatrix}. \tag{3}$$

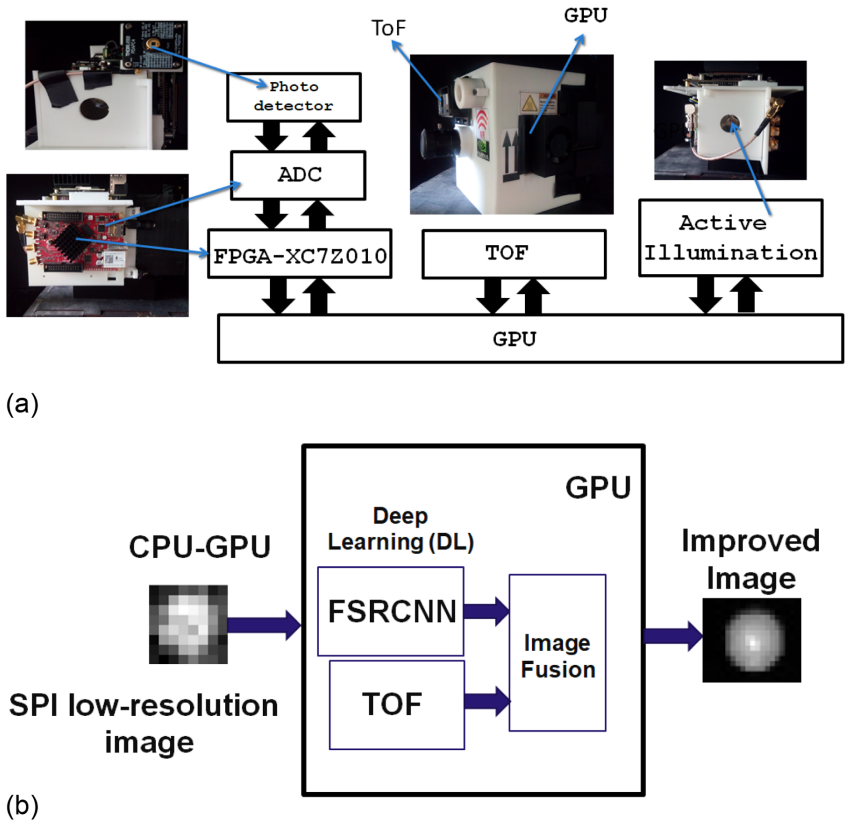

Fig. 2. Overall block diagram of the proposed vision system dimension 11 cm × 11 cm × 14 cm. It contains a lens to project the active illumination patterns (focal length of 20 cm), a weight of 1.2 kg, and a power consumption of 45 W. (a) The first stage module contains a photodiode, an active illumination source, and an InGaAs photodetector diode (FGA015) for TOF system [4]. (b) The second stage includes a GPU unit and ADC, a processing unit involving an FSRCNN network to improve the low-resolution SPI images and fusing them with TOF information captured.

Table 1. Total Processing Time, Denoted as T_{2D}, Varies Depending on the Image Capture and Reconstruction Stages[a]

T_{pre} (ms)	T_{OMP} (ms)	T_{DL} (ms)	T_{dm} (ms)	T_{2D} (ms)
18–24	20–25	9	–	45–58
18–24	20–25	9	15	62–76

[a]This includes considering the acquisition time ADC (T_{aq}), the Hadamard time projection array NIR-LEDs (T_{Had}), the exposure time of the bucket detector (T_{ext}), which is defined as the sum of pre-acquisition stage times ($T_{pre} = T_{aq} + T_{Had} + T_{ext}$), reconstruction time ($T_{OMP}$), TOF fusion image-processing time (T_{DL}), and the application of a diffusion model (T_{dm}).

The matrix G can be decomposed into two triangular matrices using Cholesky decomposition, represented as $L_i L_i^T$ [Eq. (3)]. Here, L_i is a triangular Cholesky factor [29] (refer to Algorithm 1, line 8). To solve this matrix, we define a system $L_i L_i^T x_i = \Phi^T y_i$. This system can be solved by treating it as a triangular system, where we express the system in the form $L_i u = b$ with $b = \Phi y_i$ and $L_i^T x_i = u$ (see Algorithm 1, line 10). The matrix L_i can be calculated using the formulation in Eq. (3) [26], where $w_i = L_i^{-1} G_i$ (see Algorithm 1, line 7). To obtain the reconstructed signal x_i (which contains the vector image reconstruction and needs to undergo a reshape operation to convert it into an $N \times N$ matrix), we define a stopping criterion to compare the norm of the residual with a threshold ε (see Algorithm 1, line 14), eliminating the need to calculate the residual δ (see Algorithm 1, lines 11-13). To enhance the efficiency of the algorithm, we propose implementing it on compute unified device architecture (CUDA) to parallelize the reconstruction operation [25,30] (see Algorithm 1).

To generate the final 2D image, we employ a combination of the SPI image (obtained using Algorithm 1) and postprocessing the depth information from a TOF system. To enhance the depth of data, we apply a normalization technique. Initially, the input image undergoes fusion by incorporating data from the TOF system. This fusion process is achieved using the FSRCNN network method, as detailed in [4]. The result is an improved image with 4 times the original resolution. As a result, we obtain a final high-resolution image with dimensions of 64 pixels × 64 pixels.

D. Image Acquisition Protocol for Artificial Rain Scenarios

For the design of the SPI camera, two parameters are considered when capturing SPI images under rainy conditions: the exposure time (T_{ext}) in the detector and the frequency of the pattern projection [31]. In the first case, the exposure time needed to acquire the minimum number of photons to construct an image was estimated using theoretical modeling of the NIR-SPI under rainy conditions; this is summarized in the Algorithm 2 and was calculated assuming Mie scattering [32], rain speed effect

Algorithm 1. OMP-GPU algorithm [26], Input: OMP-GPU algorithm input data: Patterns Φ, input signal y_i, target sparsity K, Output: OMP-GPU algorithm output data: sparse representation x_i that fulfills the relation $y_i \approx \Phi x_i$

procedure OMP-GPU Φ, y_i, K:
  set: $L_1 = [1], i = 1, p^0 = \Phi^T y_i$
  set: $\varepsilon = y_i y_i^T, G_i = \Phi^T \Phi, p = p^0$
  while $\varepsilon_{i-1} > \epsilon$ **do**
    $k = \text{argmax}_K |p|$ ▷ Finding the new atom
    if $i > 1$ **then**
      $w_i = \{L_{i-1} w_i = G_{i-1,K}\}$ ▷ Solver w_i
      $L_i = \begin{bmatrix} L_{i-1} & 0 \\ w_i^T & \sqrt{1 - w_i^T w_i} \end{bmatrix}$ ▷ Update of Cholesky

    $x_i = \{L_i L_i^T x_i = p^o\}$ ▷ Solver x_i
  $\beta = G_i x_i$ ▷ Matrix-sparse-vector product for each path
  $p = p^o - \beta$
  $\delta^k = x_i^T \beta$ ▷ Calculate error
  $\varepsilon^k = \varepsilon^{k-1} - \delta^k + \delta^{k-1}$ ▷ Calculate norm ε

  $i = i + 1$ ▷ increasing iteration
return x_i

modeling [32], noise floor NIR-SPI to determine [33]. The exposition time T_{ext} was estimated to be between 8 and 25 μs, for distances between 0.3 and 1 m.

Based on the exposition time T_{ext}, the minimum frequency range of the ADC is $F_{\text{ADC}} =$ 125 MHz. The frequency patterns projection are defined through the Eq. (4) [34],

$$F_{\min} = \frac{F_{\text{ADC}}}{F_{\text{patterns}}}, \tag{4}$$

where the parameter $F_{\min}$ determines the efficiency in pixels (InGaAs with efficiency pixel 0.9 [34]). An optimal design to be considered is $F = F_{\min}$ (F corresponding to the number of real pixels of the sensor) for which the highest ADC measurement rate with the lowest resolution of the sensor can be obtained to increase the signal-to-noise ratio under outdoor measurement conditions. For the design condition $F < F_{\min}$, the measurement resolution is limited by the frequency patterns F_{patterns}, and for the condition $F > F_{\min}$, the measurement resolution is limited by the sensor change-collection or signal exposure time (T_{ext}). Based on this condition and considering the effect in the rainy speed [35], the pattern generation frequency will be defined between the range of $F_{\text{patterns}} = 25$ MHz (the modulation rate of the LED array is represented by the time parameter, which is indicated in column 1 variable T_{Had} of Table 1) and $F_{\text{ADC}} = 125$ MHz, can obtain an improved measurement rate of 5 times faster Eq. (4).

To determine the setup parameters for designing the NIR-SPI system, we evaluate the exposure time versus visibility depth in two conditions: light (diameter 0.5 mm) and heavy (2 mm) rain, which is necessary to determine the rain rate approximate and extinction factor through the methods defined in [36] (see Table 2).

Algorithm 2. Pseudocode to estimate the maximum capture distance of SPI camera under rainy conditions, Input: L_b background radiance (noise), $QE(\lambda)$ photodetector's quantum efficiency, R material reflection index, λ wavelength, T_{ext} exposure time, T_{patterns} time it takes to project the active illumination patterns, Z_{max} field-far measurement, Z_h horizontal polarized reflectivity, Z_v vertical polarized reflectivity, D diameter of the water droplet, Output: z Maximum measurement distance

procedure EstimateDistance (L_b, QE (λ), R, λ, T_{ext}, $T_{patterns}$, Z_{max}, Z_h, Z_v, D):
  set: transmittance τ_{lens}, effective photosensitive area A_{pixel}
  set: focal aperture angle α_{FOV}, irradiation level of the active illumination source $\Phi_{e\lambda}$, and focal number of the system $f_\#$.
  ii=0 ▷ Initialization iteration ii
  z=0 ▷ Initialization distance z
  $\Delta z = (Z_{max}/10)$ ▷ Initialization step Δz
  Weather parameters
  $Z_{dr} = 10 \log(Z_h/Z_v)$ ▷ Reflectivity of the medium [37].
  Λ ▷ Raindrop concentration distribution slope [32]
  μ ▷ Raindrop shape parameter [32]
  $N(D)$ ▷ The raindrop distribution [32]
  $\mu_s(\lambda, D)$ ▷ Scattering cross section [32]
  Mie scattering
  $\mu_a(\lambda, D)$ ▷ Absorption coefficient [32,38]
  $\alpha(\lambda, D) = \mu_a(\lambda, D) + \mu_s(\lambda, D)$ ▷ Attenuation coefficient
  Rain speed effect.
  $v(D)$ ▷ Rain speed [32]
  $\tau_{rain} = 2D/v(D)$ ▷ Rain fall time constant [32].
  $0 < \beta < \frac{\sqrt{D}}{50 T_{patterns}}$ ▷ Factor of the raindrop [32].
  while $z < Z_{max}$) **do**
  $\Phi_{e\lambda} = \Phi_{e\lambda} e^{-\alpha(\lambda, D)z}$ ▷ Active illumination source.
  $A = \frac{\pi f_{foc}^2 D^2}{z}$ ▷ Factor A corresponds to focal parameters.
  $L_{r_i} = L_{r_{i-1}} e^{-\beta z} + L_b(1 - e^{-\beta z})$ ▷ Brightness of rain droplets.
  $E_{e\lambda\_sum(\lambda)}$ ▷ Background irradiance level [32]
  $E(N)_{ii}(A, E_{e\lambda\_sum(\lambda)}, z, \lambda, \Phi_{e\lambda}, T_{ext}, T_{patterns}, \alpha_{FOV}, \tau_{lens}, A_{pixel})$
  ▷ Number of photons impinging on the photodetector [32]
  $\sigma_{Noise\ floor}$ ▷ Calculation of the general electrical noise floor [33].
  if $|E(N)_{ii} - \sigma_{Noise\ floor}| < \delta_{th}$ **then**
    Break
  ii=ii+1
  $z = z + \Delta z$ ▷ Increase the step measurement
  return z

Table 2. Relation Exposure Time (T_{ext}), Rain Rate (R), and Capture Distance under Conditions of Light and Heavy Artificial Rain Based on Algorithm 2

Rainfall	T_{ext} (μs)	R (mm/h)	Distance (m)
Light	8–25	1.93	0.1–2
Heavy	8–25	5.6	0.2–1.5

3. DIFFUSION MODELS

In computer vision, a neural network is trained to denoise blurred images with Gaussian noise [39]. The range for defining the level of degradation in the image is determined by applying the Gaussian filter [40,41]. The neural network learns to reverse the diffusion process [17]. The diffusion model is

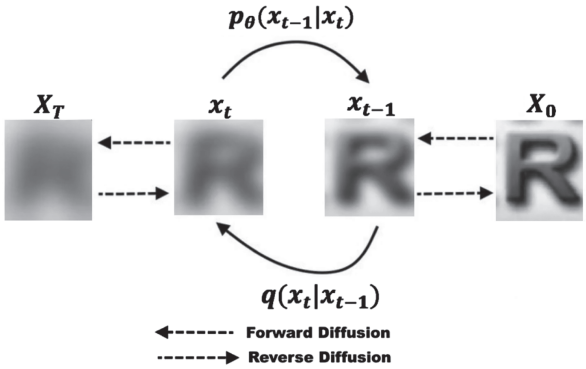

Fig. 3. Forward and reverse diffusion process of generating a sample by slowly adding and removing blur [16].

defined by a forward process that gradually degrades [42] (the degradation can be randomized or deterministic) the input data $x_0 \sim q(x)$, and we add a small amount of Gaussian noise and blur to the sample in steps, producing a sequence of affected samples (see Fig. 3). The step sizes are controlled by a variance schedule with noise over the course of T time steps and through a restoration operator [Eq. (5)], the original data can be restored (see Fig. 3) [17]. The idea of using the diffusion model is to train a neural network to recognize objects after the light is affected by light scattering or low-level light.

A. Forward Diffusion Process

For each training data set (see Fig. 3), we add a blur to the sample in T steps, producing a sequence of images with blur $x_1, \ldots, x_T$; the step sizes are controlled by a variance schedule $\{\beta_t \varepsilon(0, 1)\}_{t=1}^{T}$; the forward diffusion process defined as $(q(x_t|x_{t-1}))$ is defined from the following Markov chain Eq. (5), with $z_t \sim N(0, I)$, where N represent the noise caused by blur in the samples; $\{\beta_t\}_{t=1}^{T}$ is the predefined blur schedule; $\alpha_t = 1 - \beta_t$, and $\bar{\alpha}_t = \prod_{t=1}^{T} \alpha_t$ [17],

$$x_t = x_{t-1} - \sqrt{1-\beta_t} x_{t-1} + \sqrt{\beta_t} z_t$$

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{\bar{\alpha}_t} x_0, (1-\bar{\alpha}_t)^2 I). \quad (5)$$

B. Reverse Diffusion Process

The reverse process (see Fig. 3) requires the estimation of probability density $q(x_t|x_{t-1})$ when $t = T$ (the variable T represents the maximum number of iterations that the sample is affected by blur), which implicates generating a data sample from isotropic Gaussian noise. Therefore, we shall have to train a neural network model that estimates the $p_\theta(x_{0:T})$ [Eq. (6)] based on learned weights θ and the current state at time t [Eq. (6)] [17], where μ_Θ is the parameterization of the mean [Eq. (7)] [17], and $\sum_\theta(x_t, t)$ as variance function. For the estimation of $\mu_\theta(x_t, t)$, we apply different ways of training. In this work, we use UNet as a neural network, trained to predict the noise ϵ from the earlier formulation of $q(x_t|x_{t-1})$,

$$p_\theta(x_{0:T}) = p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1}|x_t)$$

$$p_\theta(x_{t-1}|x_t) = N(x_{t-1}; \mu_\theta(x_t, t), \sum_\Theta(x_t, t)), \quad (6)$$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_t \right). \quad (7)$$

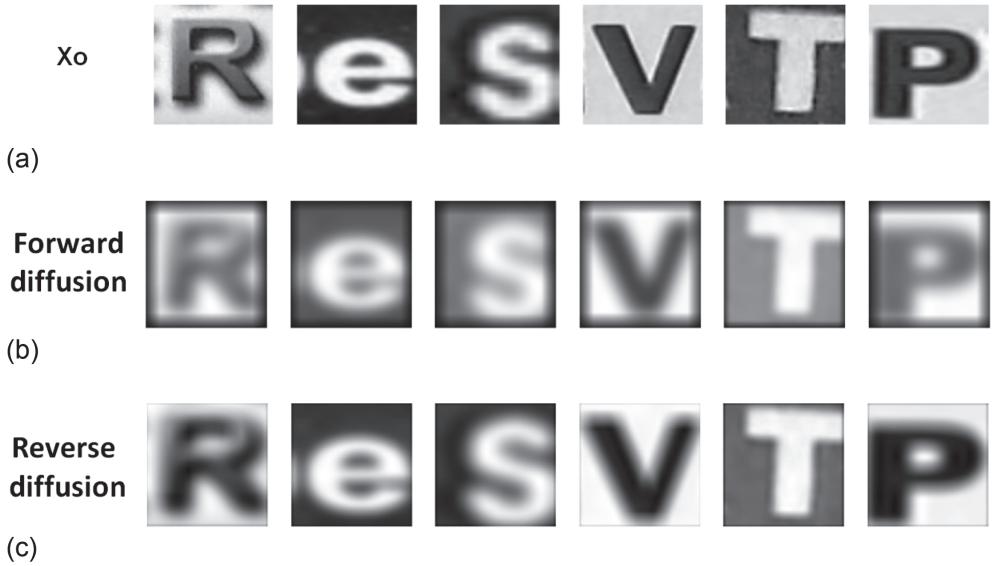

Fig. 4. Simulating blur removal using PyTorch CUDA with the data set Chars74K: a diffusion-based approach. (a) showcases the input image X_0, where the details are partially obscured due to blur. In panel (b), we apply a forward blurring process using the UNet architecture. This step emulates real-world scenarios where blur is introduced during image capture or transmission. The UNet helps us accurately simulate this blurring effect. In panel (c), we employ a backward blur diffusion module. This module utilizes the diffusion model, a powerful concept from image processing, to effectively reverse the blur and recover the lost details.

C. Training

During the training of the UNet procedure, we performed a total of 500 diffusion steps. The data set used was Chars74K [43], which contains 74,000 images comprising uppercase and lowercase letters, numbers, and symbols. The images had a resolution of 64 pixels × 64 pixels. To simulate rainy conditions, we applied a Gaussian filter to both the training and testing images, as shown in Fig. 4. We set the learning rate to 10^{-4} to expedite convergence. The implementation of the diffusion model involved the following steps [17,39]:

1. Sample image $x_0 \sim q(x_0)$.
2. Choose a certain step in the diffusion process $x_1, \ldots, x_T$.
3. Apply the bluring $\epsilon \sim N(0, I)$.
4. Try to estimate the blur.
5. $\epsilon(x_t, t) = \epsilon_\Theta(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon, t)$.
6. Learn the network UNet by gradient descent on loss $L\ \nabla_\theta||\epsilon - \epsilon_\theta(x_t, t)||^2$, where the loss can be nicely presented as Eq. (8),

$$L = \left(\left\|\epsilon - \epsilon_0\left(\sqrt{\bar{\alpha}_t}x_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t\right)\right\|^2\right). \quad (8)$$

4. SIMULATOR STRUCTURE DESIGN

For this project, we have developed a simulator designed to replicate rainy scenarios, including heavy rainfall with raindrop sizes of 2 mm (the operation of the rainfall simulation system and the experimental setup is described in [32]), as well as a half-cloudy scenario with an illumination intensity of 15 KLux. The dimensions of the simulator, as shown in Fig. 5, are 1.2 m in length, 60 cm in width, and 50 cm in height.

To enhance the simulation capabilities, we have incorporated a hydraulic lift mechanism that allows the simulator plot to be tilted up to 15°. The hydraulic lift is equipped with a flow rate range of 1 L/h, ensuring precise control over the simulation conditions. Furthermore, the simulator has six nozzles strategically placed in a random arrangement. Each nozzle has a spray radius ranging from 0.5 to 2 mm, as specified in the SPI rain simulation standard [32].

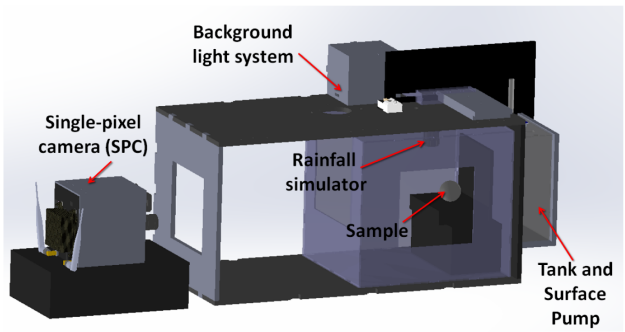

Fig. 5. Rainfall simulator for the proposed NIR-SPI system: The test bench incorporates a control system capable of emulating various rain conditions, with raindrop sizes ranging from 0.5 to 2 mm. Additionally, it allows for precise control over the level of background illumination. As illustrated, the test objects were strategically positioned within the enclosed glass box.

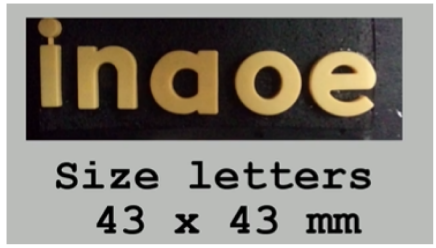

Fig. 6. Test structure size letters 43 mm × 43 mm.

A. Rainfall Simulator Calibration

To determine the distribution of rainwater across various points in the selected sample-placement area, the following steps were undertaken to improve accuracy testing:

1. To simulate rainfall at each point, containers were placed under individual nozzles to collect rainwater over a specific time period.
2. By collecting water samples from each container, we assessed the rainfall received by each sample location.
3. To ensure uniformity in the rainfall distribution, we employed the Christiansen uniformity coefficient (CUC) [44]. It is a widely used mathematical equation for evaluating water application uniformity from sprinkler irrigation systems.

5. EXPERIMENTAL RESULTS

To evaluate the performance of the prototype NIR-SPI system, we utilized a rainfall simulator to generate rainy scenes. The system was calibrated following the steps outlined in Section 4.A, and samples were taken. An image letter of dimensions 43 mm × 43 mm was reconstructed (see Fig. 6) at distances of 30 cm, 60 cm, and 1 m. For testing purposes, we evaluated the enhanced SPI image using a generative adversarial network (GAN) [45] and our diffusion model (see Fig. 7). To validate the performance of the NIR-SPI system, we conducted two tests:

(i) We evaluated various parameters to assess the quality of the reconstructed images. Specifically, we considered the peak signal-to-noise ratio (PSNR) to be desirable at >20 dB, indicating a low level of noise in the images [33]. The structural similarity index measure (SSIM) was also examined, with a value >0.5, suggesting a reasonable level of similarity between the processed and original images [33]. Additionally, we utilized the Fréchet inception distance (FID) as a measure of image postprocessing quality using neural networks, aiming for a value <20 for desirable results [46]. These objective metrics were chosen to quantitatively evaluate the quality of the reconstructed images, as summarized in Table 3).

(ii) We measured the improvement in spatial resolution [47] of the NIR image by applying both the GAN and diffusion model (see Table 4), using the parameters PSNR, SSIM, and FID.

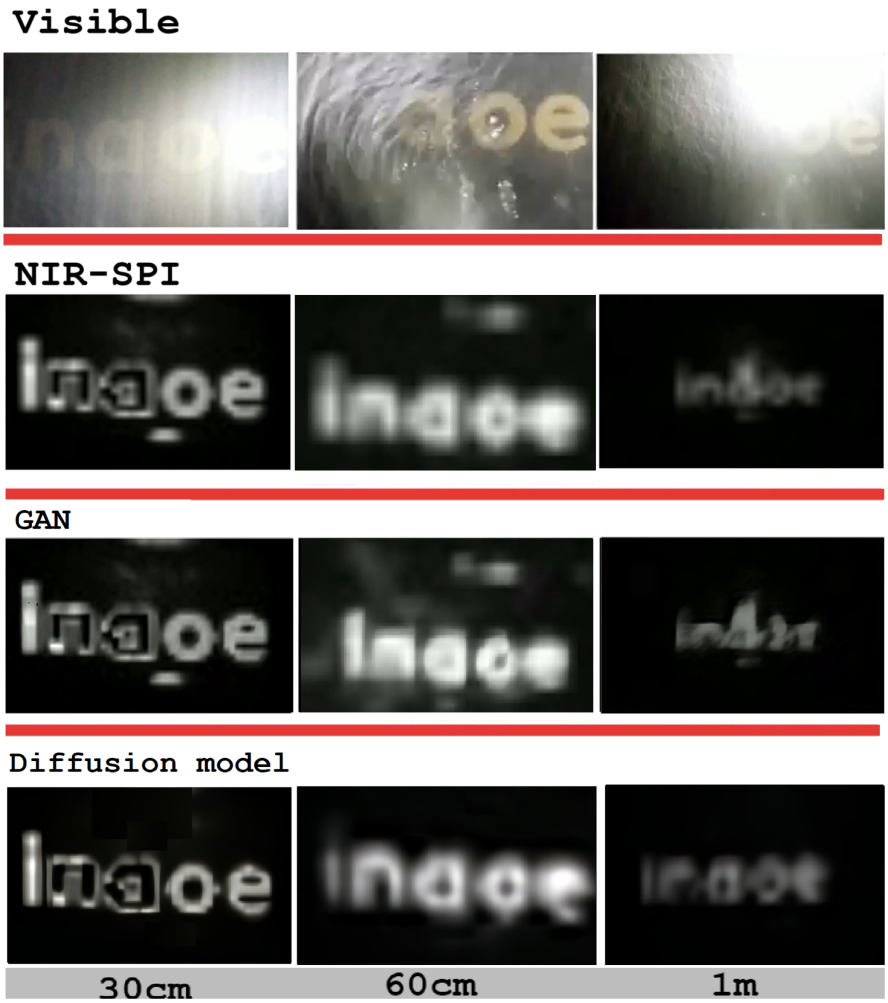

Fig. 7. VIS and NIR-SPI images obtained measuring distance 30, 60, and 1 m, under a scenario of half-cloudy (15 KLux) and heavy rainy conditions. In the case of NIR-SPI, the image is improved by applying the GAN and diffusion model (see Visualization 1).

Table 3. Evaluation of Image Reconstruction Quality Using NIR-SPI Compared to VIS under Rainy Conditions, with 2 mm Droplet Diameters and a Background Illumination of 15 KLux (Half-Cloudy)[a]

Image	PSNR (dB) ↑	SSIM ↑	FID ↓
VIS	10	0.3	-
NIR-SPI	24	0.7	14
NIR − SPI + GAN	23	0.75	38
NIR − SPI + diffusion model	26	0.84	11

[a]We measured PSNR, SSIM, and FID to assess the performance. The network model GAN [45] was applied, along with our diffusion model, to test the defined objects as shown in Fig. 6.

A. Discussion: Testing the Proposed NIR-SPI System with and without Rain

We conducted tests on various SPI image reconstructions with and without utilizing GAN and diffusion models for rainy scenarios for evaluation purposes. These scenarios involved droplet diameters of approximately 2 mm, representing heavy rain and half-cloudy background illumination conditions. The tests were conducted at distances of 30 cm, 60 cm, and 1 m. Initially, we compared the images captured in the VIS and NIR-SPI spectra under rainy conditions (refer to Fig. 7). The reconstructed image of the VIS spectrum was adversely affected by the Rayleigh effect, leading to degraded image quality values of PSNR = 10 dB and SSIM = 0.3. Additionally, the spatial resolution varied between 45 and 60 mm.

Table 4. Evaluation of Spatial Resolution (mm) under Rainy Conditions with 2 mm Droplet Diameters, Background Illumination of 15 KLux (Half-Cloudy) for the Distance Measurement $d_1 = 30$ cm, $d_2 = 60$ cm, and $d_3 = 1$ m Applying the Network Model GAN [45] and Our Diffusion Model

Image	d_1 (mm)	d_2 (mm)	d_3 (mm)
VIS	45	48	60
NIR-SPI	5	25	45
NIR − SPI + GAN	5	18	40
NIR − SPI + diffusion model	5	15	25

On the other hand, employing the NIR-SPI system allowed us to obtain an image with lower scattering than the VIS image (see Fig. 7). The NIR-SPI image exhibited improved image quality with PSNR = 24 dB and SSIM = 0.8 (refer to Table 3), while the spatial resolution ranged between 5 and 45 mm.

By utilizing diffusion models on the NIR-SPI image, we observe a noticeable enhancement in the performance of reconstructed images under rainy conditions. Using the diffusion model, we achieve PSNR = 26 dB, SSIM = 0.85, and FID = 11 (refer to Table 3). Additionally, there are significant improvements in spatial resolution (refer to Table 4), with an increase from 60 to 25 mm compared to VIS scenarios. However, when employing GAN, convergence limitations become apparent due to the scattering effect of light. This leads to image deformations and lower values of PSNR and SSIM.

6. CONCLUSION

In this work, we present a NIR-SPI vision system that utilizes active illumination provided by an array of NIR-LEDs. This system allows for the reconstruction of single-pixel images even in rainy scenarios with a background illumination of 15 KLux. To enhance the image quality of the NIR-SPI, we propose using a diffusion model to eliminate the scattering effect caused by the interaction between light and water droplets. During testing, we observed that the diffusion model performs better than using GAN to remove the blur effect in the image (refer to Fig. 7). The diffusion model significantly increases the level of detail in the NIR-SPI image (refer to Table 3), especially when compared to the VIS image, which is more adversely affected by the scattering effect. At distances greater than 50 cm, both the NIR-SPI and VIS images exhibit limitations regarding image quality and spatial resolution (refer to Table 4). However, with the application of the diffusion model, we have successfully improved the NIR-SPI image, resulting in a clearer representation of the test structure (refer to Fig. 7). This improvement makes the system suitable for various applications, such as autonomous navigation for unmanned aerial vehicles (UAVs) [48], particularly in GPS-denied or hazardous scenarios where RGB sensors face limitations in capturing images [49].

Funding. National Council for Science and Technology (251992); Consejo Nacional de Ciencia y Tecnología (661331).

Acknowledgment. The first author is thankful to Consejo Nacional de Ciencia y Tecnología (CONACYT) for his scholarship. The authors wish to thank Ing. Humberto García Flores, Head of the Illumination and Energy Efficiency (LIEE), Dr. Rafael Izazaga Perez, Head of Optical Workshop, and M.Sc. Francisco Barbosa, Head of the Engineer Center at INAOE, for their most appreciated help; they facilitated the development of the test bench and performance of the experimental testing of the NIR-SPI system. The Ph.D. studies of the first author, Carlos Alexander Osorio Quero, were funded by the Mexican Government through the National Council for Science and Technology-CONACyT.

Disclosures. The authors declare no conflicts of interest.

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

REFERENCES

1. S. Grigorescu, B. Trasnea, T. Cocias, and G. Macesanu, "A survey of deep learning techniques for autonomous driving," J. Field Rob. **37**, 362–386 (2020).
2. H. Moon, J. Martinez-Carranza, T. Cieslewski, M. Faessler, D. Falanga, A. Simovic, D. Scaramuzza, S. Li, M. Ozo, C. De Wagter, G. de Croon, S. Hwang, S. Jung, H. Shim, H. Kim, M. Park, T.-C. Au, and S. J. Kim, "Challenges and implemented technologies used in autonomous drone racing," Intell. Serv. Rob. **12**, 137–148 (2019).
3. Y.-T. Lin, M.-D. Yang, J.-Y. Han, Y.-F. Su, and J.-H. Jang, "Quantifying flood water levels using image-based volunteered geographic information," Remote Sens. **12**, 706 (2020).
4. C. Osorio Quero, D. Durini, J. Rangel-Magdaleno, J. Martinez-Carranza, and R. Ramos-Garcia, "Single-pixel near-infrared 3D image reconstruction in outdoor conditions," Micromachines **13**, 795 (2022).
5. S. Nayar and S. Narasimhan, "Vision in bad weather," in *Proceedings of the 7th IEEE International Conference on Computer Vision* (1999), Vol. **2**, pp. 820–827.
6. M. R. Ibrahim, J. Haworth, and T. Cheng, "Weathernet: recognising weather and visual conditions from street-level images using deep residual learning," ISPRS Int. J. Geo-Inf. **8**, 549 (2019).
7. D. J. Lockwood, "Rayleigh and Mie scattering," in *Encyclopedia of Color Science and Technology* (Springer, 2016), pp. 1097–1107.
8. J. Wojtanowski, M. Zygmunt, M. Kaszczuk, Z. Mierczyk, and M. Muzal, "Comparison of 905 nm and 1550 nm semiconductor laser rangefinders' performance deterioration due to adverse environmental conditions," Opto-Electron. Rev. **22**, 183–190 (2014).
9. J. Vargas Rivero, O. Schubert, H.-M. Kroll, B. Buschardt, M. Berk, M. Dura, and D. Straub, "A stochastic physical simulation framework to quantify the effect of rainfall on automotive lidar," SAE Int. J. Adv. Curr. Pract. Mobility **1**, 531–538 (2019).
10. A. Ronen, E. Agassi, and O. Yaron, "Sensing with polarized lidar in degraded visibility conditions due to fog and low clouds," Sensors **21**, 2510 (2021).
11. R. Rasshofer, M. Spies, and H. Spies, "Influences of weather phenomena on automotive laser radar systems," Adv. Radio Sci. **9**, 49–60 (2011).
12. M. Vollmer, K.-P. Möllmann, and J. A. Shaw, "The optics and physics of near infrared imaging," Proc. SPIE **9793**, 97930Z (2015).
13. A. Kokhanovsky, *Aerosol Optics: Light Absorption and Scattering by Particles in the Atmosphere*, Springer Praxis Books (Springer Berlin Heidelberg, 2010).
14. M. Byeon and S. W. Yoon, "Analysis of automotive lidar sensor model considering scattering effects in regional rain environments," IEEE Access **8**, 102669–102679 (2020).
15. G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," Opt. Express **28**, 28190–28208 (2020).
16. F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: a survey," IEEE Trans. Pattern Anal. Mach. Intell. (to be published).
17. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, eds. (Curran Associates, 2020), Vol. **33**, pp. 6840–6851.
18. S. Rissanen, M. Heinonen, and A. Solin, "Generative modelling with inverse heat dissipation," in *International Conference on Learning Representations (ICLR)* (2023).
19. B. Jing, G. Corso, R. Berlinghieri, and T. Jaakkola, "Subspace diffusion generative models," in *Computer Vision–ECCV*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds. (Springer Nature Switzerland, 2022), pp. 274–289.
20. J. Anger, M. Delbracio, and G. Facciolo, "Efficient blind deblurring under high noise levels," in *11th International Symposium on Image and Signal Processing and Analysis (ISPA)* (2019), pp. 123–128.
21. S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *IEEE/CVF International Conference on Computer Vision (ICCV)* (2021), pp. 4621–4630.
22. C. A. Osorio Quero, D. Durini, J. Rangel-Magdaleno, and J. Martinez-Carranza, "Single-pixel imaging: an overview of different methods to be used for 3d space reconstruction in harsh environments," Rev. Sci. Instrum. **92**, 111501 (2021).
23. C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Computer Vision–ECCV*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. (Springer International Publishing, 2016), pp. 391–407.

24. R. Lange, S. Böhmer, and B. Buxbaum, "CMOS-based optical time-of-flight 3D imaging and ranging," in *High Performance Silicon Imaging*, D. Durini, ed., 2nd ed., Woodhead Publishing Series in Electronic and Optical Materials (Woodhead Publishing, 2020), Chap. 11, pp. 319–375.
25. P. Kang and S. Lim, "A taste of scientific computing on the GPU-accelerated edge device," IEEE Access **8**, 208337 (2020).
26. C. O. Quero, D. Durini, R. Ramos-Garcia, J. Rangel-Magdaleno, and J. Martinez-Carranza, "Hardware parallel architecture proposed to accelerate the orthogonal matching pursuit compressive sensing reconstruction," Proc. SPIE **11396**, 56–63 (2020).
27. B. L. Sturm and M. G. Christensen, "Comparison of orthogonal matching pursuit implementations," in *Proceedings of the 20th European Signal Processing Conference (EUSIPCO)* (2012), pp. 220–224.
28. J. Chen and Z. Chen, "Cholesky factorization on heterogeneous CPU and GPU systems," in *9th International Conference on Frontier of Computer Science and Technology* (2015), pp. 19–26.
29. R. Zheng, W. Wang, H. Jin, S. Wu, Y. Chen, and H. Jiang, "GPU-based multifrontal optimizing method in sparse Cholesky factorization," in *IEEE 26th International Conference on Application-Specific Systems, Architectures and Processors (ASAP)* (2015), pp. 90–97.
30. H. Choi and J. Lee, "Efficient use of GPU memory for large-scale deep learning model training," Appl. Sci. **11**, 10377 (2021).
31. P. Duthon, F. Bernardin, F. Chausse, and M. Colomb, "Methodology used to evaluate computer vision algorithms in adverse weather conditions," Transp. Res. Procedia **14**, 2178–2187 (2016).
32. C. O. Quero, D. D. Romero, J. Rangel-Magdaleno, J. Martinez-Carranza, and R. Ramos-Garcia, "2D/3D single-pixel NIR image reconstruction method for outdoor applications in presence of rain," Proc. SPIE **11914**, 224–243 (2021).
33. C. A. Osorio Quero, D. D. Romero, R. Ramos-Garcia, J. de Jesus Rangel-Magdaleno, and J. Martinez-Carranza, "Towards a 3D vision system based on single-pixel imaging and indirect time-of-flight for drone applications," in *17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)* (2020), pp. 1–6.
34. 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.
35. K. Garg and S. Nayar, "Detection and removal of rain from videos," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (2004), Vol. **1**, p. I.
36. C. H. Bahnsen and T. B. Moeslund, "Rain removal in traffic surveillance: does it matter?" IEEE Trans. Intell. Transp. Syst. **20**, 2802–2819 (2019).
37. S.-G. Park, V. N. Bringi, V. Chandrasekar, M. Maki, and K. Iwanami, "Correction of radar reflectivity and differential reflectivity for rain attenuation at x band. Part I: Theoretical and empirical basis," J. Atmos. Ocean. Technol. **22**, 1621–1632 (2005).
38. C. Matzler, "MATLAB functions for Mie scattering and absorption," IAP Res Rep 2002-08 (Institut für Angewandte Physik, 2002).
39. E. Hoogeboom and T. Salimans, "Blurring diffusion models," in *11th International Conference on Learning Representations* (2023).
40. S. Rizvi, J. Cao, K. Zhang, and Q. Hao, "Deringing and denoising in extremely under-sampled Fourier single pixel imaging," Opt. Express **28**, 7360–7374 (2020).
41. S. Rizvi, J. Cao, and Q. Hao, "Deep learning based projector defocus compensation in single-pixel imaging," Opt. Express **28**, 25134–25148 (2020).
42. J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei, eds. (2015), Vol. **37**, pp. 2256–2265.
43. T. E. de Campos, B. R. Babu, and M. Varma, "Character recognition in natural images," in *Proceedings of the International Conference on Computer Vision Theory and Applications*, Lisbon, Portugal, 2009.
44. J. Van Boxel, "Numerical model for the fall speed of raindrop," I.C.E. Special Report (1998).
45. O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: deblurring (orders-of-magnitude) faster and better," in *IEEE/CVF International Conference on Computer Vision (ICCV)* (IEEE Computer Society, 2019), pp. 8877–8886.
46. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "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.
47. C. A. O. Quero, D. Durini, J. D. J. Rangel-Magdaleno, J. Martinez-Carranza, and R. Ramos-Garcia, "2d NIR-SPI spatial resolution evaluation under scattering condition," in *19th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)* (2022), pp. 1–6.
48. J. Martinez-Carranza and L. O. Rojas-Perez, "Warehouse inspection with an autonomous micro air vehicle," Unmanned Syst. **10**, 329–342 (2022).
49. C. A. Quero, D. Durini, J. de Jesús Magdaleno, J. Martinez-Carranza, and R. Ramos-Garcia, "Fast NIR-single-pixel-imaging enhancement under scattering environment," in *13th International Micro Air Vehicle Conference*, G. de Croon and C. D. Wagter, eds. (2022), pp. 53–58, paper IMAV2022-6.