



High-quality and high-diversity conditionally generative ghost imaging based on denoising diffusion probabilistic model

SHUAI MAO,¹ YUCHEN HE,^{2,*} HUI CHEN,² HUAIBIN ZHENG,² JIANBIN LIU,² YUAN YUAN,² MINGNAN LE,³ BIN LI,^{4,5} JUAN CHEN,¹ AND ZHUO XU²

¹School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China

²Electronic Materials Research Laboratory, Key Laboratory of the Ministry of Education and International Center for Dielectric Research, School of Electronic Science and Engineering, Xi'an Jiaotong University, Xi'an 710049, China

³School of Information Science and Technology, Northwest University, Xi'an 710127, China

⁴Key Laboratory of Biomedical Information Engineering of Ministry of Education, School of Life Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China

⁵Bioinspired Engineering and Biomechanics Center (BEBC), Xi'an Jiaotong University, Xi'an 710049, China

*yuchenhe@xjtu.edu.cn

Abstract: Deep-learning (DL) methods have gained significant attention in ghost imaging (GI) as promising approaches to attain high-quality reconstructions with limited sampling rates. However, existing DL-based GI methods primarily emphasize pixel-level loss and one-to-one mapping from bucket signals or low-quality GI images to high-quality images, tending to overlook the diversity in image reconstruction. Interpreting image reconstruction from the perspective of conditional probability, we propose the utilization of the denoising diffusion probabilistic model (DDPM) framework to address this challenge. Our designed method, known as DDPMGI, can not only achieve better quality but also generate reconstruction results with high diversity. At a sampling rate of 10%, our method achieves an average PSNR of 21.19 dB and an SSIM of 0.64, surpassing the performance of other comparison methods. The results of physical experiments further validate the effectiveness of our approach in real-world scenarios. Furthermore, we explore the potential application of our method in color GI reconstruction, where the average PSNR and SSIM reach 20.055 dB and 0.723, respectively. These results highlight the significant advancements and potential of our method in achieving high-quality image reconstructions in GI, including color image reconstruction.

© 2023 Optica Publishing Group under the terms of the [Optica Open Access Publishing Agreement](#)

1. Introduction

Ghost imaging (GI) has garnered significant attention in recent years after its initial experimental completion by the esteemed team of T.B. Pittman and Y.H. Shih in 1995 [1]. In contrast to conventional imaging methods, which rely on first-order correlation, GI utilizes the second-order correlation among fluctuations in the intensity of the light field to achieve imaging. The unique imaging mechanism of GI offers several advantages, including turbulence-free, lens-less, and high resolution. These characteristics make this technology highly valuable in scattering media imaging [2–5], lidar [6–8], encryption [9–11], and correlated areas. However, high sampling rates are required to achieve high-quality image reconstruction. Compressed sensing (CS) [12–15] and deep learning (DL) [16–22] are two main research directions to improve the quality of reconstruction images at low sampling rates. CS-based GI algorithms face the problems of high complexity and intricate selection of hyperparameters. In previous DL-based research, image

reconstruction was seen as a regression task. Various network structures like CNN [16–18], RNN [19,20], and Transformer [21,22], have been dug deep to map inputs, usually bucket signals or low-quality images, to high-quality images.

Despite the positive outcomes of the studies mentioned above, they all suffer from a common limitation: they are inherently deterministic in nature. Once inputs are given, corresponding outputs are determined. However, it is important to note that under undersampling conditions, a group of bucket signals may correspond to multiple plausible target images. Deterministic models, which typically focus on learning one-to-one mappings, often overlook the inherent diversity involved in image reconstruction. To illustrate this deficiency, consider a scenario where the task is to detect whether the system is risk-free given bucket signals, regression-based models can obtain high-quality images for risk assessment. Regardless of whether the model's output accurately reflects the true target, there is only one result available for risk judgment. In contrast, if given bucket signals, multiple output results can be gotten, then we can assess the accuracy of the judgment from different perspectives. This example emphasizes the significance of enhancing the diversity of reconstructed images while simultaneously ensuring the quality of image reconstruction. In summary, while deterministic models have their merits, it is crucial to recognize the importance of incorporating diversity into the reconstructed images when dealing with bucket signals.

To achieve this goal, generative models are considered in this work to handle GI image reconstruction tasks. Instead of learning direct mapping, generative models learn the data distribution on the basis of probabilistic and statistical knowledge. By leveraging the learned data distribution, it becomes possible to sample a wide variety of images from it. The most common generative model is the generative adversarial network (GAN). In 2022, GAN was designed to enable recognition that is not based on target image information [23]. Ming Zhao et al. proposed the use of conditional GAN to achieve high-quality computational ghost imaging (CGI) in 2023 [24]. However, the input random noise in native GAN is replaced by bucket signals or low-quality GI images, which means these models are only similar to GAN but decay into regression-based models. Besides, due to the need to train both generator and discriminator, GANs are prone to imbalance and thus difficult to converge when dealing with complex targets, which increases the difficulty of using [25]. In 2020, another outstanding generative model, Denoising Diffusion Probabilistic Model (DDPM) [26], was proposed. Based on the process of noise diffusion and Bayesian theory, DDPM can help to generate target images that conform to learned data distribution from a Gaussian distribution.

In this work, we introduce the DDPM model to the GI field and design a DDPM-based conditional computational ghost imaging method (DDPMGI). DDPMGI leverages a progressive sampling process to transform the Gaussian distribution into a conditional image distribution. DDPMGI offers two notable advantages. For one thing, high-quality GI reconstruction images can be obtained using bucket signals as guidance. And for another, it can be used to generate high-diversity reconstruction results. These different reconstruction results can reveal the characteristics of target images from different perspectives. To verify the effectiveness of our method, we conducted simulations at a sampling rate of 10%. The results of simulations indicate that DDPMGI consistently produced higher-quality reconstructions, surpassing the performance of other methods employed for comparison. Besides, the relationship between image quality and image diversity was explored. Our findings revealed an intriguing trend: as the different reconstruction results were averaged, the diversity of the resulting image decreased while the overall quality of the image reconstruction improved. This understanding emphasizes the importance of considering the specific requirements and objectives of the imaging task. Moreover, our method exhibited a remarkable capability to utilize bucket signals directly for guiding the generation of color GI reconstruction results. The average peak signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM) achieved 20.055 dB and 0.723,

respectively. These remarkable results serve as strong evidence of the effectiveness of our method. Furthermore, physical experiments were organized at sampling rates ranging from 10% to 5%. The experimental results obtained in our study successfully validate the practical applicability of our method, especially in the presence of noise and other environmental challenges.

2. Method

2.1. CGI framework

CGI was proposed by Shapiro in 2008 [27]. To achieve imaging, a series of speckles are projected from a digital mirror device (DMD) or spatial light modulator (SLM) onto the targets. Transmitted or reflected signals are then collected by a bucket detector without spatial resolution. One measurement can be modeled by:

$$y_i = \iint S_i(x, y) \times T(x, y) dx dy, \quad (1)$$

where S_i is the i^{th} illumination speckle, y_i is the i^{th} transmitted or reflected signal received by a bucket detector, $T(x, y)$ is the feature function that characterizes target image. For a target in the resolution of n^2 , m measurements will be organized. The corresponding sample rate (SR) is defined as $\frac{m}{n^2}$. Combining all these measurements together, the entire measurement process can be modeled by:

$$\mathbf{Y} = \mathbf{H}\mathbf{x}, \quad (2)$$

where $\mathbf{Y} \in \mathbb{R}^m$ is bucket detector signals, each row in $\mathbf{H} \in \mathbb{R}^{m,n \times n}$ is a measurement speckle, and $\mathbf{x} \in \mathbb{R}^{n \times n}$ is a target image reshaped into column vector form. Given \mathbf{y} and \mathbf{H} , CGI can reconstruct targets by the following equation:

$$\tilde{\mathbf{x}} = \frac{1}{m} \mathbf{H}^T \times (\mathbf{Y} - \langle \mathbf{Y} \rangle). \quad (3)$$

Some GI algorithms such as DGI [28], and NGI [29] can also be used to better reconstruct target images. To reduce sampling time, CGI is always performed under under-sampling conditions, which means m is smaller than n^2 . In these scenarios, given a \mathbf{y} , there is more than one possible \mathbf{x} satisfying Eq. (2). In other words, many target images may decay onto one \mathbf{y} at the same time. If regression-based models are used to map bucket signals or low-quality into high-quality reconstruction results, given a \mathbf{y} , only one \mathbf{x} can be gotten, and other possibilities are completely ignored. To solve this problem, instead of learning direct mapping, we turn to use generative models which can learn image distribution.

2.2. DDPM model

As shown in Fig. 1, the idea of the DDPM comes from the process of noise diffusion. For given images, if noise is constantly superimposed on them, eventually, they will be completely submerged in noise. This forward diffusion process can be modeled by a Markov process.

$$x_t = \sqrt{\alpha_t} x_{t-1} + \sqrt{1 - \alpha_t} z_t, \quad (4)$$

where $z_t \sim \mathcal{N}(0, I)$ is noise added in each step, α_t is a hyperparameter that determines the intensity of noise. According to Eq. (4), x_t can be calculated using x_{t-1} , x_{t-1} can be calculated using x_{t-2} Expanding this recursion relation, x_t can be calculated using x_0 . Eq. (4) can be converted to Eq. (5) after some simplification.

$$x_t = \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} z, \quad (5)$$

where $\bar{\alpha}_t = \prod_{i=1}^t \alpha_i$, $z \sim \mathcal{N}(0, I)$. From this equation, it can be found $x_t \sim \mathcal{N}(\sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)I)$. Thus, to make the final x_T a random noise, $\bar{\alpha}_t$ needs to be set small enough. Besides, the effect

of adding noise is not the same at different stages. Adding a small noise on a clean image at the beginning will have a noticeable effect, but a much larger noise is needed on a noisy image to achieve the same effect. Thus, α_t is always set to increase gradually. Through this forward diffusion process, the image distribution can be converted into a Gaussian distribution. Once the diffusion process can be reversed, the image distribution can be re-obtained with a Gaussian distribution. The reverse process is illustrated in Fig. 2. As every x_t can be calculated by adding Gaussian noise to the initial image x_0 , if the Gaussian noise mentioned earlier can be estimated by a neural network $G_\theta(x_t)$, then we can reverse x_t back to the initial image x_0 following black paths at the bottom of Fig. 2 using Eq. (5). But as we have seen, there are many steps in the diffusion process, and if we turn x_t into x_0 only through one step, it will inevitably cause a large error. So similar to the diffusion process, as shown in the blue paths in Fig. 2, the actual reverse process is achieved in the following way: given x_t at a moment, noise \tilde{z} can be estimated through network $G_\theta(x_t)$, and x_{t-1} can be calculated using them together. Based on Bayesian posterior probability, the reverse process can be described by the following equations:

$$\begin{aligned} q(x_{t-1}|x_t) &\sim \mathcal{N}(\mu, \sigma^2 I), \\ \sigma^2 &= \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t), \\ \mu &= \frac{1}{\sqrt{\alpha_t}} (x_t - \frac{1 - \alpha_t}{\sqrt{1 - \alpha_t}} \tilde{z}), \\ \tilde{z} &= G_\theta(x_t, t). \end{aligned} \quad (6)$$

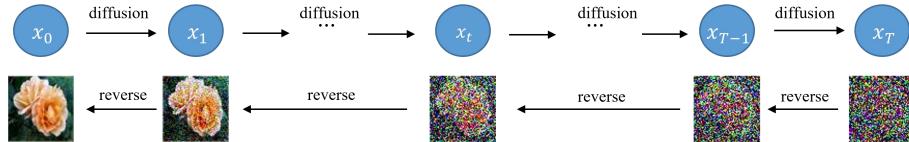


Fig. 1. DDPM, diffusion and reverse process.

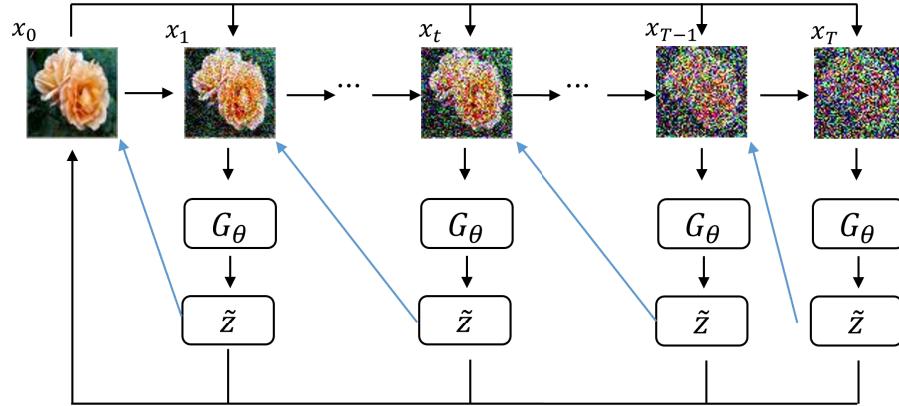


Fig. 2. Illustration of reverse process.

2.3. DDPMGI

As illustrated in the previous part, given images, classical DDPM can convert a Gaussian distribution to their image distribution $p(x)$. However, for such a distribution, the final results

only depend on the initial sampling noises but ignore the information on bucket signals. Taking this relevant information into account, GI image reconstruction pays more attention to what conditional image distribution $p(x|y)$ will be when bucket signals are given. As a result, instead of being converted into $p(x)$, the Gaussian distribution is converted into $p(x|y)$ by DDPMGI. To achieve this, the forward diffusion process is kept unchanged. And the reverse process is modified to the following:

$$\begin{aligned} q(x_{t-1}|x_t) &\sim \mathcal{N}(\tilde{\mu}, \sigma^2 I), \\ \sigma^2 &= \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} (1 - \alpha_t), \\ \tilde{\mu} &= \frac{1}{\sqrt{\alpha_t}} (x_t - \frac{1 - \alpha_t}{\sqrt{1 - \bar{\alpha}_t}} \tilde{z}), \\ \tilde{z} &= \tilde{G}_\theta(x_t, t, y). \end{aligned} \quad (7)$$

As shown in Eq. (7), DDPMGI treats bucket signals y as another input for the network G_θ . Inspired by one work in the super-resolution area [30], G_θ is modified to \tilde{G}_θ . The flow and change of data are shown in Fig. 3.

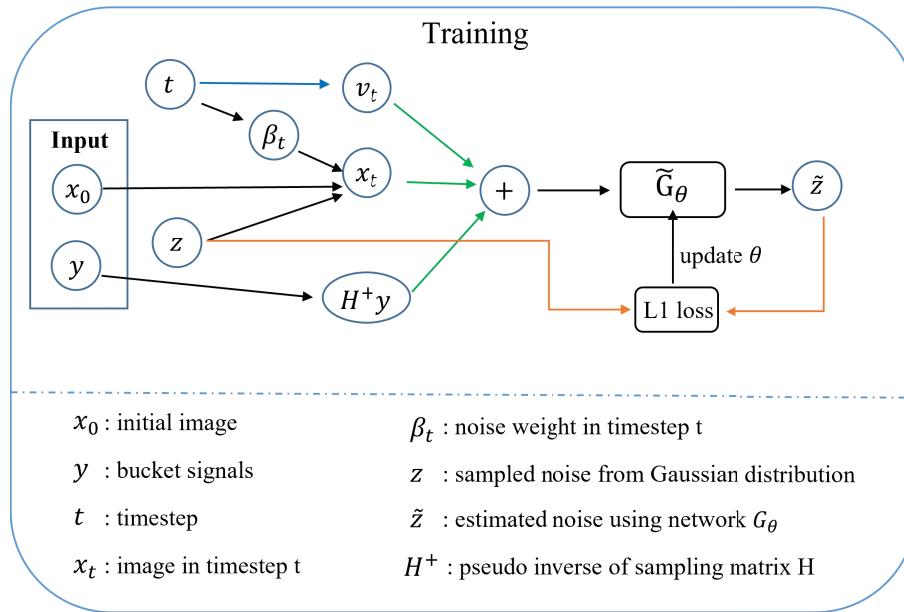


Fig. 3. Training process of network \tilde{G}_θ .

In the training phase, for a given pair (x_0, y) , a timestep t will first be selected. After a Gaussian noise z is sampled from Gaussian distribution, x_t can be calculated using Eq. (5) and predefined noise schedule α_t . As the network needs to be able to process input images at any given time, it lacks knowledge of the specific timestep to which an input image belongs. Therefore, similar to the positional encoding [31] in Transformer, timestep information is additionally added into the network in the way of time embedding. The corresponding time-embedding vector v_t will be superimposed on x_t on each pixel. For ease of operation, y will first be converted into image form by the pseudo-inverse of the sampling matrix H , then be overlayed with x_t in the channel dimension, and fed into the network to obtain estimated added noise \tilde{z} . Mean Absolute loss (L1 loss) between \tilde{z} and z is calculated and then used to update network parameters through gradient descent. The whole training process can be summarized as Algorithm 1.

Algorithm 1. Training process for a DDPMGI model.

```

1: repeat
2:    $t \sim U(1, T)$                                 ▷ Select a timestep from 1 to T
3:    $(x_0, y) \sim p(x, y)$                          ▷ Select  $x_0$  and corresponding bucket signal y from dataset
4:    $z \sim \mathcal{N}(0, I)$                            ▷ Sample a random noise from standard Gaussian distribution
5:    $Loss \leftarrow \|\tilde{G}_\theta(x_t, t, y) - z\|_1$ 
6:   Update network parameters  $\theta$  using gradient descent
7: until network converges

```

The whole test or sampling process can be summarized as Algorithm 2. When network \tilde{G}_θ is well-trained, it can be used to generate high-quality GI reconstruction images given bucket signals. The detailed process is as follows. Starting from timestep T, a noise will first be sampled from standard Gaussian distribution as x_T . In each timestep t, x_t , timestep t, and bucket signals y will be fed into network \tilde{G}_θ to obtain estimated added noise \tilde{z} . The image of the previous moment, x_{t-1} , can be calculated by Eq. (6). Repeating this process, high-quality GI reconstruction images x_0 can finally be achieved. From this process, it can be found that x_T is sampled from a standard Gaussian distribution, and each x_T corresponds to a reconstructed image. By sampling different x_T from the standard Gaussian distribution, reconstruction results with high diversity can be finally obtained.

Algorithm 2. Sampling/Test process for a DDPMGI model.

Require: : bucket signal y

```

 $t \leftarrow T$                                      ▷ Begin from timestep T
2:  $x_t \sim \mathcal{N}(0, I)$ 
for  $t > 0$  do
4:    $\tilde{z} \leftarrow \tilde{G}_\theta(x_t, t, y)$            ▷ Estimate added noise
       $z \sim \mathcal{N}(0, I)$ 
6:    $x_{t-1} = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\tilde{z}) - \sqrt{\frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1-\alpha_t)}z$     ▷ Calculate x based on the posterior distribution Equ.7
       $t \leftarrow t - 1$ 
8: end for

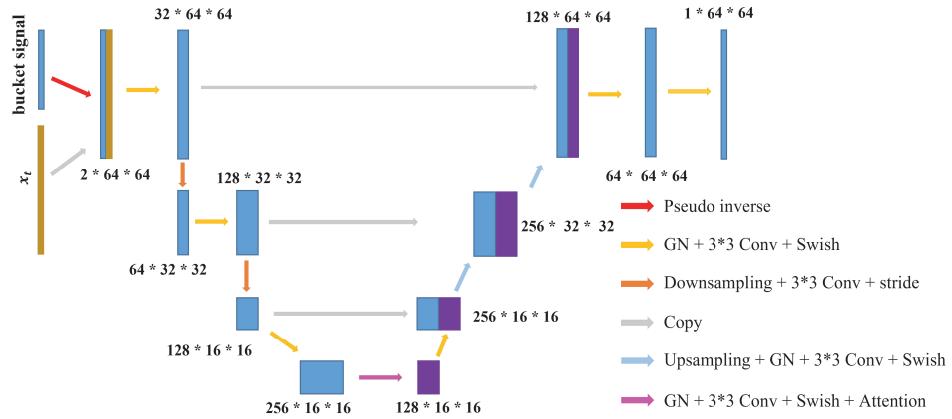
```

3. Demonstration based on simulation results

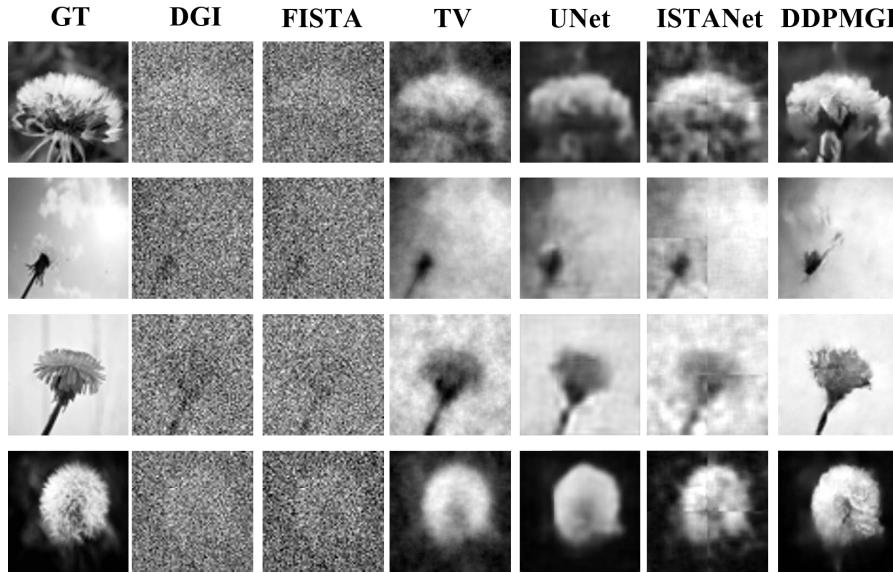
A network \tilde{G}_θ is needed to be trained to estimate added noise on each pixel of images. While in principle all networks can be used to achieve this, UNet is finally adopted in DDPMGI because the output of network \tilde{G}_θ is the same size as target images. The U-net structure used by us is shown in Fig. 4. below. The input image x_t , the bucket signals y, and the timestep t will be superimposed together in the same way as mentioned before. In order to introduce nonlinearity, the Swish activation function [32] is employed. Group normalization (GN) [33] and Attention [31] are utilized to enhance the network's capability to process data. 3539 flower images in the size of 64×64 are used to train the network. Adam optimizer and L1 loss are selected to optimize network parameters. For DDPMGI, the total diffusion step T is set to 1000, and noise schedule α_t is set to grow linearly from 1e-5 to 5e-2.

3.1. Results at low sampling rate

Although increasing the diversity of reconstructed GI images is a potential goal, the first thing that must be guaranteed is the image reconstruction results at low sampling rates. Therefore, our method is first tested at the sampling rate of 10%. In order to verify the effectiveness of our

**Fig. 4.** UNet structure.

method, we also tested the results of DGI [28], FISTA [34], TV [35], UNet [30], and ISTANet [36] at 10% sampling rate for comparison. Among the five methods compared, DGI is an imaging method based on traditional GI, FISTA and TV are GI methods based on compressed sensing, UNet and ISTANet are GI methods based on deep learning. The results of these methods are shown in Fig. 5.

**Fig. 5.** Reconstruction results when SR = 10%.

To evaluate the quality of the results, the PSNR and SSIM are used. Their definitions are given below:

$$PSNR(x, y) = 10 \log_{10} \frac{(255)^2}{MSE(x, y)}, \quad (8)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (9)$$

where MSE is the mean square error, C_1 and C_2 are small non-zero valued regularization constants to avoid instability in image regions where the mean or standard deviation is close to zero, and μ is image mean, σ is image standard deviation and σ_{xy} is cross-covariance between two images. An overview of the PSNR and SSIM is provided in Table 1. From this table, it can be found our DDPMGI has excellent reconstruction performance. The average PSNR of DDPMGI and ISTANet results is close (0.11dB apart), but the SSIM of DDPM results is much higher than ISTANet (0.09). The SSIM results of TV reconstruction results have two results higher than DDPMGI (0.05 and 0.02 respectively), but the average value is smaller than DDPMGI (0.03 smaller), and the average value of PSNR is also much smaller than DDPMGI (2.13dB smaller).

Table 1. PSNR and SSIM results of different methods

Method	PSNR(dB)					SSIM				
	flower1	flower2	flower3	flower4	average	flower1	flower2	flower3	flower4	average
DGI [28]	10.76	9.28	8.7	8.64	9.35	0.13	0.08	0.1	0.06	0.09
FISTA [34]	11.45	9.33	9.28	9.28	9.83	0.13	0.07	0.09	0.06	0.09
TV [35]	19.32	18.19	18.53	20.22	19.06	0.56	0.70	0.50	0.53	0.57
UNet [30]	15.98	16.93	21.52	22.67	19.28	0.46	0.69	0.66	0.64	0.61
ISTANet [36]	18.69	24.00	20.78	21.73	21.30	0.52	0.65	0.50	0.54	0.55
DDPMGI	18.02	22.74	21.52	22.46	21.19	0.51	0.68	0.72	0.66	0.64

In order to gain further insights, we conducted ablation experiments from various perspectives using the same four test data as shown in Fig. 5. The average PSNR and SSIM of these experiments are outlined in Table 2. In the first experiment, the network was trained using L2 (Mean Square Error) Loss. The second experiment involved the incorporation of an additional down-sampling and corresponding up-sampling layer to increase the complexity of the network. The third and fourth experiments focused on modifying the total timestep: in the third experiment, the total timestep was reduced to 500, while in the fourth experiment, it was increased to 2000. From the results, the following findings were observed: Firstly, L1 loss yielded better results compared to L2 loss, as indicated by higher PSNR and SSIM values. Furthermore, the DDPMGI model with 1000 total steps demonstrated the best performance in terms of PSNR and SSIM. However, increasing the complexity of the network, as seen in Experiment 2 and Experiment 4, resulted in inadequate training and consequently poorer results. These findings provide valuable insights into the impact of loss functions, network complexity, and the total number of timesteps on the performance of the method. The results will guide further improvements in the research and refinement of the proposed approach.

Table 2. Results of ablation experiments

Methods	PSNR (dB)	SSIM
DDPMGI	21.19	0.64
DDPMGI + L2 loss	16.93	0.54
DDPMGI + an extra layer	20.62	0.62
DDPMGI + 500 steps	20.81	0.62
DDPMGI + 2000 steps	14.80	0.44

3.2. Impact of image diversity

As we mentioned earlier, one advantage of our method is that it can achieve high-diversity image reconstructions through learning conditional image distribution $p(x|y)$ instead of a direct mapping from y to x . Through this conditional image distribution $p(x|y)$, many possible results can be

obtained. To verify the diversity, for two target images, we generated 4 images respectively using our method. The corresponding results are shown in Fig. 6. Since the bucket signals cannot be displayed intuitively, H^+y are shown here to represent the information contained in the bucket signals. Although the same bucket signals were used, the resulting images differ in detail, which can help us understand target images from more perspectives. Additionally, we calculated the mean and variance of PSNR and SSIM of them. From Fig. 6, it can be found the averaged PSNR and SSIM are still high but the variance is kept at a low level. Thus, it can be inferred that though they fluctuate within a certain range, they are all concentrated at a high level. This effectively proved that bucket signals can be used to guide image generation and that the conditional image distribution is as we expected – the more it matches that target image, the greater the probability of being sampled.

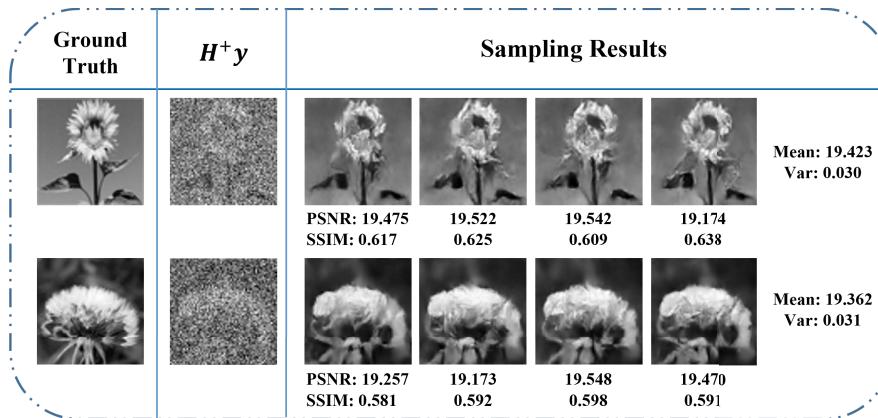


Fig. 6. Possible reconstruction results using the same bucket signals.

From the point of view of probability theory, when we sample, add, and average a random variable multiple times, we can reduce the variance without changing the mean. Following this idea, we also averaged the resulting images in groups of k ($k = 2, 4, 8$). Corresponding results are shown in Fig. 7. It turns out that averaging multiple resulting images can improve the quality of the final images. This is because target image information is correlated but noise is not. In fact, this process can be modeled in the following equation:

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i, \quad (10)$$

When we increase the number of samples so large that we can traverse all possibilities, the upper equation changes to the following equation:

$$\bar{x} = \int p(x_i|y)x_i, \quad (11)$$

Then our model can be seen as a probabilistic regression model. Although it can obtain images with better quality, it lost diversity due to pattern averaging. As a result, a trade-off between diversity and image quality exists, which needs to be weighted according to the actual application scenario.

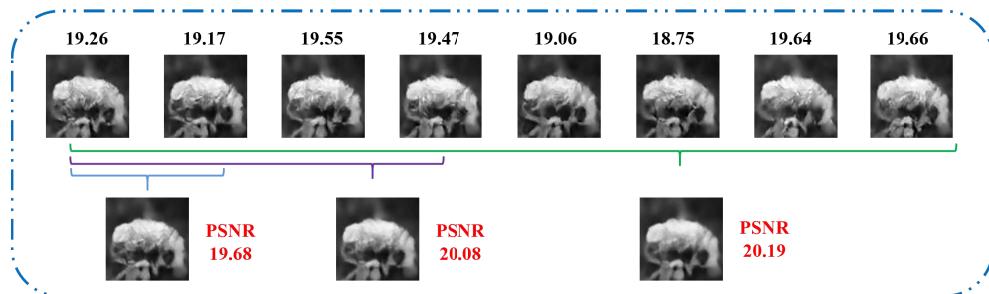


Fig. 7. Results of averaging possible reconstruction images.

4. Color GI reconstruction

Some research has been carried out to achieve color ghost imaging [37–39]. Their methods are to use three beams of light to measure the information of target objects on the RGB three channels respectively, and then perform imaging based on the information of the three channels. This is because in the measurement process of GI, if there is no special design, the information on different colors is difficult to distinguish. Thus the results of GI are usually presented in grayscale. The previous part has demonstrated our method works well when dealing with images in grayscale. Since color image space is much larger than grayscale image space, traditional regression-based methods are hard to find a direct mapping from grayscale images to corresponding color images. However, by learning conditional image distribution $p(x|y)$, in which x is color images, our method can achieve color GI image reconstruction. At a sampling rate of 20%, we retrained our model for color images. Except that the input x_t of the network is changed from grayscale to RGB three-channel color map, the other settings do not change. Using bucket signals as guidance, the final reconstruction color images are shown in Fig. 8. As illustrated in Fig. 8, The average PSNR achieved in our simulation was 20.055 dB, indicating a favorable level of reconstruction quality. Additionally, the average SSIM reached 0.723, further affirming the satisfactory similarity between the reconstructed and ground truth images. These metrics serve as quantitative measures

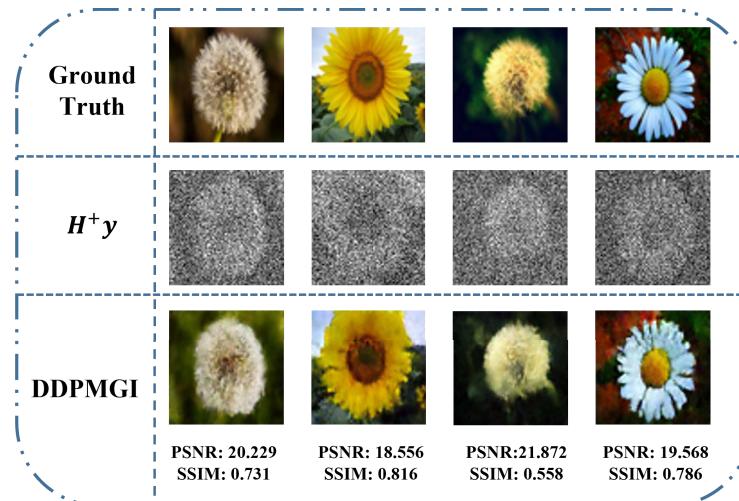


Fig. 8. DDPMGI results in color image reconstruction.

to evaluate the fidelity and similarity of the reconstructed images, highlighting the effectiveness of our approach. More research will be done in this direction to explore the potential of our models.

5. Demonstration based on physical experiment results

To further validate the effectiveness of the DDPMGI method, a physical experiment based on pseudo-thermal GI architecture is organized. In this experiment, two objects are employed as objects. The first object was a 0.6 mm-sized letter "GI," while the second object was a 1 mm-sized Chinese character "JIAO". The experiment setup is illustrated in Fig. 9. The experiment is organized on Zolix SC300-3B electronically controlled displacement platform. we maintained a distance of 1 meter between the light source and the target. The laser from the 532nm semiconductor laser is modulated by one 220 mesh rotated ground glass and divided into two beams through the polarization beam splitter. One beam is directed toward the target and captured by the Thorlabs DET025A/M bucket detector. The other beam is recorded directly by the CCD to obtain the corresponding speckle information. We demonstrated DDPMGI at sampling rates from 10% to 5% and compared results with other traditional methods. The corresponding results are depicted in Fig. 10. In the actual experiment, despite the interference factors such as environmental noise, DDPMGI consistently achieved remarkable results in the actual experiment compared with other methods, which further verifies the effectiveness of our method.

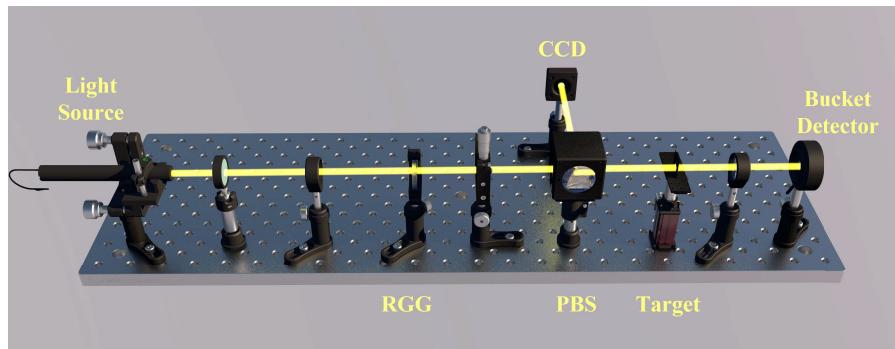


Fig. 9. Schematic diagram of the physical experiment.

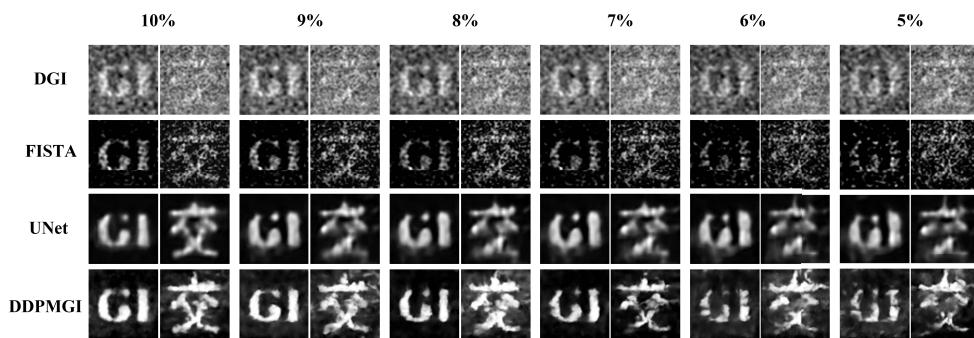


Fig. 10. DDPMGI physical experiments results at different sampling rates.

6. Conclusion

In this paper, we address the limitation of single deterministic output in regression-based deep learning models. To overcome this limitation, we propose the application of DDPM to the field of GI for the first time and design a corresponding DDPMGI method. Simulations and physical experiments have proven that with this method, high-quality GI image reconstruction can be achieved under low sampling rates. At the same time, it effectively resolves the issue of multiple high-quality images degrading into the same bucket signals. At 10% sampling rate, DDPMGI outperforms other comparison methods with an average PSNR of 21.19 dB and SSIM of 0.64. Through our research, we have discovered a trade-off between diversity and image quality, which should be weighted according to real scenarios. Furthermore, with a small change, our method can be used directly to handle color GI image reconstruction task. And the average PSNR and the SSIM achieved 20.055 dB and 0.723, respectively. At the current stage, the training process still requires a long period of time to be completed. Additionally, the way to fuse bucket signals is slightly simple, just concatenation in the channel dimension. And there might be deviations in hue during the reconstruction of color ghost imaging. However, we are aware of these limitations and plan to address them in our future research endeavors. Further efforts will be dedicated to resolving these issues and improving the accuracy and robustness of the method.

Funding. National Natural Science Foundation of China (61901353); Key Research and Development Projects of Shaanxi Province (2021GXZH-Z-012); Fundamental Research Funds for the Central Universities (xhj032021005); 111 Project (B14040); JD AI Research (202105127); Key Innovation Team of Shaanxi Province (2018TD-024).

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. T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Phys. Rev. A* **52**(5), R3429–R3432 (1995).
2. F. Li, M. Zhao, Z. Tian, F. Willomitzer, and O. Cossairt, "Compressive ghost imaging through scattering media with deep learning," *Opt. Express* **28**(12), 17395–17408 (2020).
3. Z. Gao, X. Cheng, J. Yue, and Q. Hao, "Extendible ghost imaging with high reconstruction quality in strong scattering medium," *Opt. Express* **30**(25), 45759–45775 (2022).
4. Y. Xiao, L. Zhou, and W. Chen, "High-resolution ghost imaging through complex scattering media via a temporal correction," *Opt. Lett.* **47**(15), 3692–3695 (2022).
5. L.-X. Lin, J. Cao, D. Zhou, H. Cui, and Q. Hao, "Ghost imaging through scattering medium by utilizing scattered light," *Opt. Express* **30**(7), 11243–11253 (2022).
6. S. Ma, Z. Liu, C. Wang, C. Hu, E. Li, W. Gong, Z. Tong, J. Wu, X. Shen, and S. Han, "Ghost imaging lidar via sparsity constraints using push-broom scanning," *Opt. Express* **27**(9), 13219–13228 (2019).
7. H.-Z. Lin, W.-T. Liu, S. Sun, and L.-K. Du, "Influence of pulse characteristics on ghost imaging lidar system," *Appl. Opt.* **60**(6), 1623–1628 (2021).
8. T. Jiang, Y. Bai, W. Tan, X. Zhu, X. Huang, S. Nan, and X. Fu, "Ghost imaging lidar system for remote imaging," *Opt. Express* **31**(9), 15107–15117 (2023).
9. S. Yuan, L. Wang, X. Liu, and X. Zhou, "Forgery attack on optical encryption based on computational ghost imaging," *Opt. Lett.* **45**(14), 3917–3920 (2020).
10. P. Zheng, Q. Tan, and H.-c. Liu, "Inverse computational ghost imaging for image encryption," *Opt. Express* **29**(14), 21290–21299 (2021).
11. P. Zheng, Z. Ye, J. Xiong, and H.-c. Liu, "Computational ghost imaging encryption with a pattern compression from 3d to 0d," *Opt. Express* **30**(12), 21866–21875 (2022).
12. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.* **95**(13), 131110 (2009).
13. L. Wang and S. Zhao, "Compressed ghost imaging based on differential speckle patterns," *Chin. Phys. B* **29**(2), 024204 (2020).
14. H. Zhang, Y. Xia, and D. Duan, "Computational ghost imaging with deep compressed sensing," *Chin. Phys. B* **30**(12), 124209 (2021).
15. Z. Yu, Y. Liu, X. Bai, X. Chen, Y. Wang, X. Li, M. Sun, and X. Zhou, "Bipolar compressive ghost imaging method to improve imaging quality," *Opt. Express* **31**(2), 3390–3400 (2023).
16. Y. He, G. Wang, G. Dong, S. Zhu, H. Chen, A. Zhang, and Z. Xu, "Ghost imaging based on deep learning," *Sci. Rep.* **8**(1), 6469 (2018).

17. S. Rizvi, J. Cao, K. Zhang, and Q. Hao, "Deepghost: real-time computational ghost imaging via deep learning," *Sci. Rep.* **10**(1), 11400 (2020).
18. F. Wang, C. Wang, M. Chen, W. Gong, Y. Zhang, S. Han, and G. Situ, "Far-field super-resolution ghost imaging with a deep neural network constraint," *Light: Sci. Appl.* **11**(1), 1 (2022).
19. Y.-Y. Huang, C. Ou-Yang, K. Fang, Y.-F. Dong, J. Zhang, L.-M. Chen, and L.-A. Wu, "High speed ghost imaging based on a heuristic algorithm and deep learning," *Chin. Phys. B* **30**(6), 064202 (2021).
20. Y. He, S. Duan, Y. Yuan, H. Chen, J. Li, and Z. Xu, "Semantic ghost imaging based on recurrent-neural-network," *Opt. Express* **30**(13), 23475–23484 (2022).
21. Y. He, Y. Zhou, Y. Yuan, H. Chen, H. Zheng, J. Liu, Y. Zhou, and Z. Xu, "Transunet-based inversion method for ghost imaging," *J. Opt. Soc. Am. B* **39**(11), 3100–3107 (2022).
22. W. Ren, X. Nie, T. Peng, and M. O. Scully, "Ghost translation: an end-to-end ghost imaging approach based on the transformer network," *Opt. Express* **30**(26), 47921–47932 (2022).
23. Y. He, Y. Chen, Y. Yuan, H. Chen, H. Zheng, J. Li, and Z. Xu, "Generative-adversarial-network-based ghost recognition," *Phys. Rev. A* **106**(2), 023710 (2022).
24. M. Zhao, X. Zhang, and R. Zhang, "High-quality computational ghost imaging with a conditional gan," in *Photonics*, vol. 10 (MDPI, 2023), p. 353.
25. D. Saxena and J. Cao, "Generative adversarial networks (gans) challenges, solutions, and future directions," *ACM Comput. Surv.* **54**(3), 1–42 (2022).
26. J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems* 33, 6840–6851 (2020).
27. J. H. Shapiro, "Computational ghost imaging," *Phys. Rev. A* **78**(6), 061802 (2008).
28. F. Ferri, D. Magatti, L. A. Lugiato, and A. Gatti, "Differential ghost imaging," *Phys. Rev. Lett.* **104**(25), 253603 (2010).
29. B. Sun, S. S. Welsh, M. P. Edgar, J. H. Shapiro, and M. J. Padgett, "Normalized ghost imaging," *Opt. Express* **20**(15), 16892–16901 (2012).
30. C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
31. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in Neural Information Processing Systems* 30, 1 (2017).
32. P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv* arXiv:1710.05941 (2017).
33. Y. Wu and K. He, "Group normalization," in *Proceedings of the European conference on computer vision (ECCV)*, (2018), pp. 3–19.
34. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.* **2**(1), 183–202 (2009).
35. X. Hu, J. Suo, T. Yue, L. Bian, and Q. Dai, "Patch-primitive driven compressive ghost imaging," *Opt. Express* **23**(9), 11092–11104 (2015).
36. C. Zhang, J. Zhou, J. Tang, F. Wu, H. Cheng, and S. Wei, "Deep unfolding for singular value decomposition compressed ghost imaging," *Appl. Phys. B* **128**(10), 185 (2022).
37. Y. Ni, D. Zhou, S. Yuan, X. Bai, Z. Xu, J. Chen, C. Li, and X. Zhou, "Color computational ghost imaging based on a generative adversarial network," *Opt. Lett.* **46**(8), 1840–1843 (2021).
38. D. Duan, R. Zhu, and Y. Xia, "Color night vision ghost imaging based on a wavelet transform," *Opt. Lett.* **46**(17), 4172–4175 (2021).
39. Z. Yu, Y. Liu, J. Li, X. Bai, Z. Yang, Y. Ni, and X. Zhou, "Color computational ghost imaging by deep learning based on simulation data training," *Appl. Opt.* **61**(4), 1022–1029 (2022).