

Dual-attention cGAN for high-fidelity ghost imaging at ultra-low sampling rates

Yuanyuan Ma(马园园), Yunfan He(何运繁), Le Wang(王乐), Shengmei Zhao(赵生妹)

Citation: Chin. Phys. B. 35, 094207 (2026). DOI: 10.1088/1674-1056/ae1b78 | CSTR: 32038.14.CPB.ae1b78
http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Ghost imaging based on the control of light source bandwidth

Zhao-Qi Liu(刘兆骐), Yan-Feng Bai(白艳锋), Xuan-Peng-Fan Zou(邹璇彭凡), Li-Yu Zhou(周立宇), Qin Fu(付芹), Xi-Quan Fu(傅喜泉)
Chin. Phys. B. 32, 034210 (2023). DOI: 10.1088/1674-1056/ac6edd

Orthogonal-triangular decomposition ghost imaging

Jin-Fen Liu(刘进芬), Le Wang(王乐), Sheng-Mei Zhao(赵生妹)
Chin. Phys. B. 31, 084202 (2022). DOI: 10.1088/1674-1056/ac5e99

Handwritten digit recognition based on ghost imaging with deep learning

Xing He(何行), Sheng-Mei Zhao(赵生妹), Le Wang(王乐)
Chin. Phys. B. 30, 054201 (2021). DOI: 10.1088/1674-1056/abd2a5

Experimental demonstration of influence of underwater turbulence on ghost imaging

Man-Qian Yin(殷曼倩), Le Wang(王乐), Sheng-Mei Zhao(赵生妹)
Chin. Phys. B. 28, 094201 (2019). DOI: 10.1088/1674-1056/ab33ee

Mask-based denoising scheme for ghost imaging

Yang Zhou(周阳), Shu-Xu Guo(郭树旭), Fei Zhong(钟菲), Tian Zhang(张天)
Chin. Phys. B. 28, 084204 (2019). DOI: 10.1088/1674-1056/28/8/084204

Dual-attention cGAN for high-fidelity ghost imaging at ultra-low sampling rates

Yuanyuan Ma(马园园)[1], Yunfan He(何运繁)[1], Le Wang(王乐)[1], and Shengmei Zhao(赵生妹)[1,2,†]

[1] *Institute of Signal Processing and Transmission, Nanjing University of Posts and Telecommunications, Nanjing* 210003, *China*
[2] *Key Laboratory of Broadband Wireless Communication and Sensor Network Technology (Ministry of Education), Nanjing* 210003, *China*

(Received 10 August 2025; revised manuscript received 3 November 2025; accepted manuscript online 5 November 2025)

Achieving high-quality ghost imaging (GI) reconstruction under extremely low sampling rates remains a significant challenge. This paper proposes a novel deep learning-based approach for high-fidelity GI reconstruction. Leveraging an end-to-end conditional generative adversarial network (cGAN), the method incorporates residual structures integrated with efficient channel attention (ECA) and convolutional block attention module (CBAM) mechanisms to significantly enhance feature extraction and representation. This framework enables direct reconstruction of target images from one-dimensional bucket detection signals. Pre-trained on simulated datasets, the proposed method demonstrates exceptional performance. The results under ideal simulation conditions demonstrate that the proposed method can achieve high quality (SSIM $\geq$ 0.8) even at the extremely low ratio of 0.29%. Experimental results on the MNIST and Fashion–MNIST benchmark datasets confirm that accurate image reconstruction can be obtained at remarkably low sampling rates (0.78%). Compared to traditional computational GI, baseline cGAN methods and GAN with integrated SENet attention module consistently demonstrate the superiority of our approach, establishing a new state-of-the-art in low-sampling-rate GI reconstruction.

Keywords: ghost imaging (GI), deep learning, conditional generative adversarial network (cGAN)

PACS: 42.30.Wb, 42.30.Va, 07.05.Pj, 02.70.–c

DOI: 10.1088/1674-1056/ae1b78 **CSTR:** 32038.14.CPB.ae1b78

1. Introduction

Ghost imaging (GI) represents a fundamentally distinct imaging paradigm compared to conventional CCD/CMOS sensors, reconstructing an object's image non-locally through second-order correlations within light intensity fluctuations.[1–3] Following its inaugural experimental realization with entangled photons in 1995,[4] the field has evolved substantially with techniques including: iterative denoising GI,[5] compressed sensing GI (CSGI),[6,7] differential GI,[8] and Hadamard basis GI.[9,10] Whereas traditional implementations require dual optical paths, Shapiro's computational GI (CGI) framework (2008) eliminated the reference arm requirement through a single-path architecture.[11] Subsequent research has extended GI methodologies to various domains such as terahertz,[12] x-ray,[13] and turbulent media imaging.[14]

Despite significant advancements, ghost imaging (GI) systems still face a fundamental challenge: their inherently high measurement requirements. Both imaging speed and reconstruction quality are directly constrained by the number of sampling measurements. While increasing the sampling rate typically enhances reconstruction quality, achieving high-fidelity reconstruction under low-sampling conditions necessitates meticulous optimization of the illumination patterns. Illumination strategies have evolved from initial random speckle patterns[15] to structured illumination schemes, encompassing advanced modalities such as Fourier basis patterns[16,17] and Walsh–Hadamard basis patterns.[18,19] Concurrently, reconstruction frameworks have undergone parallel development, particularly driven by the groundbreaking progress in compressed sensing (CS) algorithms,[20] which offer novel technical pathways for enhancing reconstruction efficiency.

The recent integration of artificial intelligence (AI), particularly deep learning, offers further promise in improving GI efficiency under low sampling conditions.[21,22] Current deep learning approaches for GI reconstruction fall primarily into two categories: (i) Post-reconstruction enhancement: Employing deep neural networks to refine an initially blurred image obtained at low sampling rates;[23–25] (ii) End-to-end reconstruction: Directly mapping one-dimensional (1D) bucket detector signals to the target image without intermediate reconstruction.[26–28] Although end-to-end approaches demonstrate the potential for substantial sampling reduction—achieving reconstructions at 6.25% sampling for MNIST images or 5.45% with specialized architectures like DattNet,[21] high-fidelity reconstruction under ultra-low sampling rates (significantly below these thresholds) remains elusive. Generative adversarial networks (GANs) have emerged as powerful solutions, with capabilities that encompass full-sampling quality CSGI at 50% sampling[29] and motion-target GI using cross-correlation.[30] However, achieving high-quality imaging below the 5% sampling rates continues to pose significant

†Corresponding author. E-mail: zhaosm@njupt.edu.cn

© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

research challenges.

In this work, we address this critical challenge by proposing a novel deep learning-based GI reconstruction method specifically designed for high-quality image recovery under extremely sparse sampling. We propose a novel deep learning-based method built upon an end-to-end conditional generative adversarial network (cGAN) framework, pre-trained on simulated data for enhanced performance. The core innovation lies in the generator architecture, which integrates residual blocks with dual attention mechanisms: efficient channel attention (ECA) and convolutional block attention module (CBAM), strategically embedded at multiple deeper network layers. This design significantly strengthens feature extraction and representation by enabling dynamic focus on critical image regions. Extensive simulation and experimental evaluations on benchmark datasets (MNIST, Fashion–MNIST) demonstrate the method's exceptional effectiveness, achieving high-fidelity reconstruction even at remarkably low sampling rates down to 0.29% (in ideal simulation) and 0.78% (in experiment), significantly outperforming traditional and contemporary baseline methods.

2. Method

2.1. Deep learning-based GI reconstruction method

The principle of the proposed scheme is illustrated in Fig. 1. Our ghost imaging reconstruction framework utilizes a generative adversarial network (GAN)[31] architecture comprising two core components: a generator (G) and a discriminator (D).

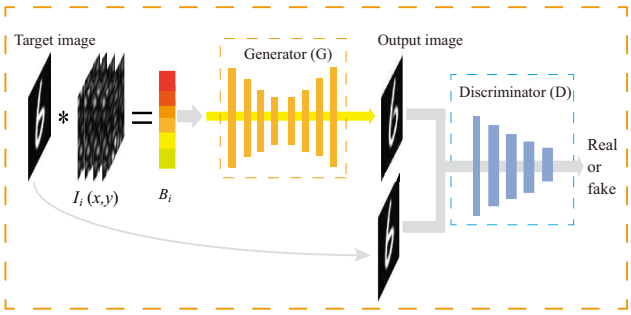


Fig. 1. The schematic diagram of the proposed GI method.

The generator processes the 1D bucket signals B_i acquired by the GI system. These signals first undergo dimensional expansion via a fully connected layer. The dimensionally expanded bucket signal is then reshaped into a feature tensor of the required size (say $1\times64\times64$). This tensor serves as the input to a U-net backbone network, which ultimately generates the final reconstructed image. The discriminator receives both the generated image from G and the corresponding ground-truth images as input. Its objective is to classify each input image as 'real' (ground-truth, output ~ 1) or 'fake' (generated, output ~ 0).

During adversarial training, G and D are optimized alternately within a minimax game framework. The generator G strives to produce realistic images capable of deceiving the discriminator D. Concurrently, the discriminator D learns to improve its ability to distinguish between real and generated images. This adversarial process drives both networks towards optimizing their respective objective functions. Theoretically, the training converges when a Nash equilibrium is reached, signifying that G produces images indistinguishable from the real data by D.

Within the ghost imaging framework, a sequence of spatially varying speckle patterns, $I_i(x,y)$ $(i=1,2,3,\ldots,M)$, is sequentially illuminated onto the target object $T(x,y)$. A bucket detector collects the transmitted or reflected intensity for each illumination event. This yields measurements $\{B_i\}$ $(i=1,2,3,\ldots,M)$, where M is the measurement number, relating to the sampling rate. The bucket measurement B_i for each speckle pattern is given by the spatial integral over the object plane as

$$B_i = \iint I_i(x,y)T(x,y)\,\mathrm{d}x\mathrm{d}y. \tag{1}$$

At the reconstruction stage, the object image is computationally retrieved from the bucket measurements $\{B_i\}$ and the known speckle patterns $\{I_i(x,y)\}$ using a correlation function. The reconstruction is expressed as

$$\hat{T}(x,y) = \frac{1}{M}\sum_{i=1}^{M}(B_i - \langle B\rangle)(I_i(x,y) - \langle I(x,y)\rangle), \tag{2}$$

where $\hat{T}(x,y)$ denotes the reconstructed target image, $\langle B\rangle = \frac{1}{M}\sum_{i=1}^{M}B_i$ denotes the mean bucket signal, and $\langle I(x,y)\rangle = \frac{1}{M}\sum_{i=1}^{M}I_i(x,y)$ corresponds to the ensemble-averaged speckle pattern.

The proposed scheme differs from traditional second-order correlation and compressive sensing methods by accomplishing direct image reconstruction from a 1D sequence of bucket detection values. This mapping is mathematically represented as,

$$\hat{T}(x,y) = \mathcal{G}\left(\{B;\Psi\}\right), \tag{3}$$

where $B=\{B_i\}_{i=1}^{M}$ is a bucket detection sequence of M measurements. The reconstruction model, parameterized by Ψ, implements an implicit mapping $\mathcal{G}(\cdot)$ from B to $\hat{T}(x,y)$. Since deriving an explicit expression for $\mathcal{G}(\cdot)$ and solving the inverse problem in Eq. (3) is challenging, a cGAN is introduced to approximate the inverse mapping. The training dataset is $(B^{(n)},T^{(n)})_{n=1}^{N}$, where N denotes the total number of training instances, $B^{(n)}$ is the n-th bucket sequence, and $T^{(n)}$ is the corresponding ground-truth image.

Critically, instead of processing the raw 1D bucket signal directly, the generator $\boldsymbol{G}$ first transforms it into a structured two-dimensional (2D) representation $Z = E(B^{(n)})$ through

fully connected layers. This design enables the optimization process to convert the temporal variations in the bucket signal into an organized spatial representation, thereby making it suitable for subsequent convolutional processing. The generator subsequently processes Z through a series of convolutional layers for deep feature extraction and spatial upsampling. The fundamental convolution operation at each layer is given by

$$Z^{\ell+1} = \varphi(W^{\ell} * Z^{\ell} + b^{\ell}), \tag{4}$$

where Z^{ℓ} denotes the feature map at layer ℓ, W^{ℓ} and b^{ℓ} are the convolutional kernel weights and biases, $*$ denotes convolution, and φ is a nonlinear activation function. Through this progressive layer-wise feature transformation, the network effectively maps the initial 2D representation Z to a high-fidelity reconstructed image.

2.2. Network structure

The designed network employs a cGAN architecture,[32] extending the standard GAN framework through its dual adversarial components: a generator dedicated to image synthesis and a discriminator responsible for authenticity assessment. Figure 2 shows the designed network architecture, with Fig. 2(a) detailing the generator, and Fig. 2(b) illustrating the discriminator.

As illustrated in Fig. 2(a), the generator utilizes an encoder–decoder structure initiated by a 1D bucket detector signal. Its encoder employs four progressive downsampling stages with $4\times$ 4 convolutional kernels (stride 2, padding 1) to extract hierarchical features. To enhance architectural performance, a residual block is utilized in the second stage to strengthen feature continuity, while the third stage incorporates ECA for efficient channel dependency modeling. The fourth stage further integrates the CBAM to jointly optimize spatial and channel feature representation. To balance computational efficiency and generalization, attention mechanisms are selectively applied only in later layers. The decoder reconstructs images through four transposed convolutional stages with skip connections preserving critical high-frequency details between corresponding encoder–decoder layers. All layers use ReLU activations except the final output layer, which employs a sigmoid function. This strategic implementation of attention mechanisms primarily in later stages effectively balances computational efficiency against potential overfitting risks. This strategic placement of attention mechanisms optimizes performance while mitigating overfitting risks.

In contrast, the discriminator (Fig. 2(b)) adopts a PatchGAN architecture to evaluate the authenticity of reconstructed images. Unlike conventional discriminators that perform global classification, PatchGAN divides the input image into multiple local patches and independently assesses the realism of each patch. Consequently, it outputs a probability feature map of size $1 \times 3 \times 3$, where each element represents the likelihood that the corresponding image patch originates from real data. The output values range from 0 to 1, with values closer to 1 indicating a highly realistic region, and values closer to 0 suggesting synthetic or unrealistic content. By enabling fine-grained discrimination of local structures, this design enhances the model's sensitivity to texture and structural details, thereby improving the overall perceptual quality of the reconstructed images.

Adversarial training progressed through three phases. Initially, the discriminator underwent dedicated pre-training to establish baseline discrimination capability between early generator outputs and real images. Subsequently, both networks engaged in alternating min–max optimization: the generator progressively refined its synthetic samples while the discriminator adaptively enhanced its discrimination capability. Training terminated when the accuracy of the discriminator approached 50%, indicating the generator produced samples indistinguishable from real data.

Under ultra-low sampling rates, the training of generative adversarial networks (GANs) becomes highly unstable and prone to mode collapse. To address this issue, we adopt a Wasserstein GAN with gradient penalty (WGAN–GP), which enforces Lipschitz continuity to stabilize gradient updates. Additionally, spectral normalization is applied to both the generator and the discriminator, constraining model capacity and further preventing training divergence.

2.3. Loss function

Neural network training optimizes the parameters governing the weights and biases to enhance predictive capability. Within our supervised learning framework, this optimization relies on paired samples of input bucket signals and corresponding ground-truth images that constrain the model during iterative refinement. The network thereby learns to reconstruct high-quality images from input signals.

Employing a cGAN framework, our model's objective function integrates complementary loss components that collectively guide the generator toward structurally faithful and visually realistic reconstructions. The adversarial loss L_{GAN} promotes distributional alignment by training the generator to produce outputs indistinguishable from real images, formulated as follows:

$$L_{\mathrm{cGAN}}(G,D) = E\left[\log D(x,y)\right] + E\left[\log\left(1 - D(x,G(x))\right)\right], \tag{5}$$

where x represents the input bucket signal reshaped into a 2D image, and y is the corresponding ground-truth image. $G(x)$ denotes the generated output, and $D(\cdot)$ denotes the discriminator's authenticity probability. This loss conditions the generator output on the input signal while matching the real image distribution.

Structural precision is enforced through pixel-level L_1 loss, which constrains the mean absolute error between generated and target images, expressed as

$$L_{\mathrm{L1}}(G) = E\left[\|y - G(x)\|_1\right]. \tag{6}$$

This component ensures geometric alignment by minimizing pixel-space deviations.

The complete objective function combines these complementary mechanisms as

$$G^* = \arg\min_G \max_D L_{\mathrm{cGAN}}(G,D) + \lambda L_{\mathrm{L1}}(G), \tag{7}$$

where $\lambda = 100$ is used to balance the loss contributions. The $\arg\min_G \max_D$ operator represents the adversarial min–max optimization, stabilizing training by simultaneously refining the generator's synthesis capability and the discriminator's assessment accuracy.

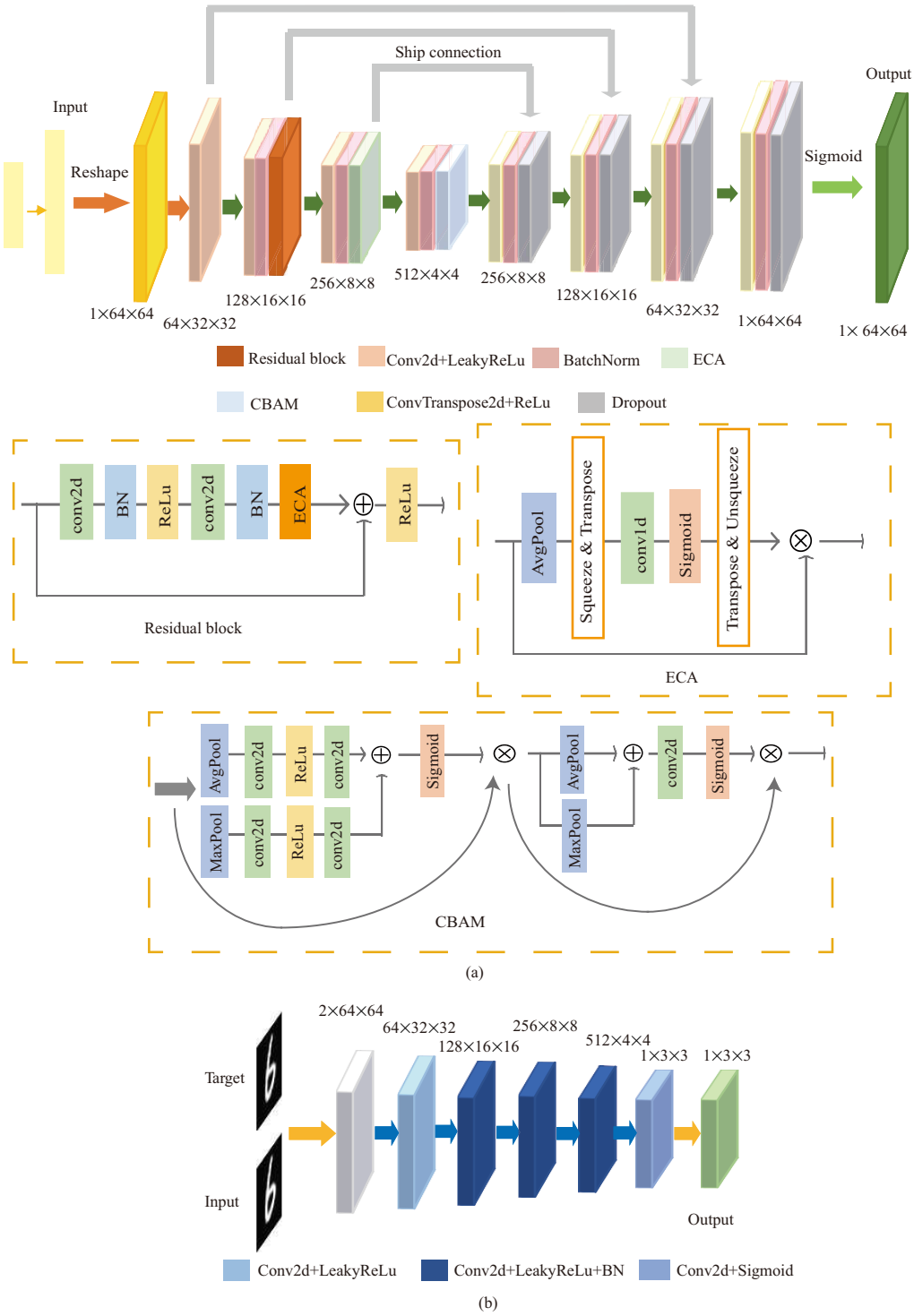


Fig. 2. Network organization of the proposed ghost imaging system implemented via cGAN. (a) Generator architecture: Comprises an encoder–decoder topology incorporating residual blocks and attention mechanisms. (b) Discriminator configuration: Structural design of the adversarial evaluator.

3. Numerical simulation and experimental results

This section details the numerical validation framework and experimental verification of the proposed GI methodology.

For the numerical simulations, a total of 2×10^4 images were randomly selected from the MNIST training set for model training, while an additional 2500 images were randomly sampled from the MNIST test set for evaluation. All images were resized to 64×64 spatial resolution. Network optimization employed the Adam algorithm with learning rate 10^{-4} and weight decay 10^{-4}, implemented on an Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.5 GHz and NVIDIA TITAN RTX 24 G GPU platform.

The experimental configuration, detailed in Fig. 3, employs a digital micromirror device (Texas Instruments DLP LightCrafter 4500) to project structured speckle patterns onto target specimens. Backscattered photons were collected by a single-element photodetector (Thorlabs PMM02-1) devoid of spatial discrimination capability. The photocurrent signals were digitized using a 16-bit analog-to-digital conversion unit (National Instruments USB-6341). An adversarial neural network architecture performed direct end-to-end reconstruction of object intensity distributions from the temporally sampled bucket detection sequences.

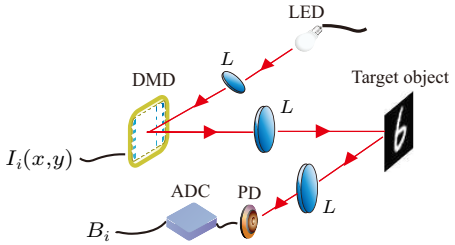


Fig. 3. Experimental setup for the proposed ghost imaging scheme. DMD: digital micromirror device, PD: photodetector, ADC: analog-to-digital converter.

Quantitative evaluation of reconstruction fidelity employs two complementary metrics: peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM), which provide distinct yet synergistic assessments of image quality. Both metrics evaluate fidelity between reconstructed image $M(i,j)$ and ground truth $N(i,j)$ with dimensions $m \times n$. The PSNR quantifies pixel-level accuracy through mean squared error

$$\mathrm{MSE}(M,N) = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} [M(i,j) - N(i,j)]^2, \quad (8)$$

$$\mathrm{PSNR}(M,N) = 10 \log_{10} \left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}} \right), \quad (9)$$

where MAX denotes the maximum possible pixel value in the image, and higher PSNR values indicate reduced distortion. The SSIM metric evaluates perceptual similarity by comparing luminance, contrast, and structural covariance as

$$\mathrm{SSIM}(M,N) = \frac{(2\mu_M\mu_N + C_1)(2\sigma_{MN} + C_2)}{(\mu_M^2 + \mu_N^2 + C_1)(\sigma_M^2 + \sigma_N^2 + C_2)}, \quad (10)$$

where μ_M and μ_N represent the mean values, σ_M and σ_N denote the standard deviations of the image intensity distributions $M(i,j)$ and $N(i,j)$, respectively, and σ_{MN} is their covariance. The terms $C_1 = (k_1 L)^2$ and $C_2 = (k_2 L)^2$ ensure numerical stability, where L represents the dynamic range of pixel values ($L = 255$ for 8-bit grayscale images). The scalar constants k_1 and k_2 (satisfying $k_1, k_2 \ll 1$) prevent division by near-zero denominators. An SSIM value approaching unity indicates perfect structural correspondence between the images.

3.1. The performance

Figure 4 presents simulated and experimental reconstruction results for handwritten digits across progressively reduced sampling ratios. The simulation results are shown for sampling ratios of 1.2%, 0.78%, 0.39%, and 0.29%, respectively. The results under ideal simulation conditions demonstrate robust performance, achieving SSIM ≥ 0.8 even at the extremely low ratio of 0.29%. Experimental validation at 1.2% and 0.78% sample ratios confirms this capability. Reconstructions remain fully successful at 1.2%, while most digits retain recognizability at 0.78%, but digits '6' and '9' exhibit quality degradation at 0.78%, attributable to measurement errors from environmental noise. Collectively, these results confirm the method's strong low-sampling reconstruction capability, supporting its feasibility for ghost imaging. Relevant demonstrations have been conducted subsequently to verify this phenomenon.

Figure 5 illustrates the comprehensive validation results on the structurally more complex Fashion–MNIST dataset. Quantitative evaluations under sub-Nyquist sampling rates confirm the method's capability for high-fidelity reconstruction of complex targets, demonstrating statistically significant performance improvements compared to conventional ghost imaging methodologies. Particularly noteworthy is the superior performance of the proposed method on the Fashion–MNIST dataset at ultra-low sampling rates (0.29% and 0.39%) compared to its performance on the MNIST dataset. This enhancement can be attributed to the effectiveness of the attention mechanism in extracting salient features from complex target objects, while the attention mechanism provides limited benefits for simple structures such as MNIST digits, it significantly enhances the reconstruction of rich visual details in Fashion–MNIST images.

To bridge the discrepancy between simulation and experiment, we introduced varying levels of Gaussian noise into the simulations to emulate experimental noise disturbances. As shown in Fig. 6, the reconstruction performance of the proposed method was evaluated under different noise levels, characterized by the standard deviation σ, at a sampling rate of

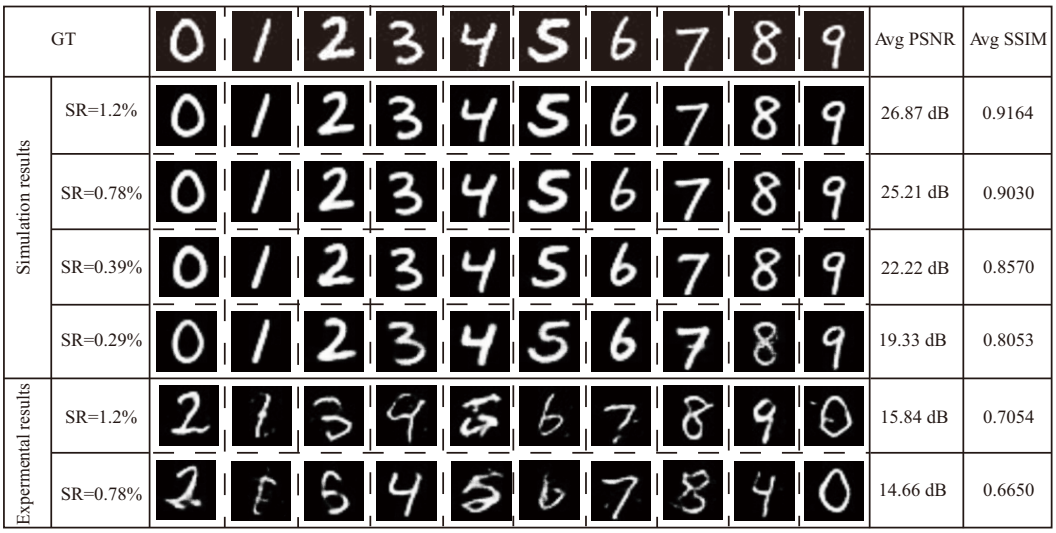


Fig. 4. Quantitative reconstruction performance of handwritten digits (0–9) under sub-Nyquist sampling using proposed ghost imaging method.



Fig. 5. Quantitative assessment of Fashion–MNIST image reconstructions under sub-Nyquist sampling using proposed ghost imaging method.

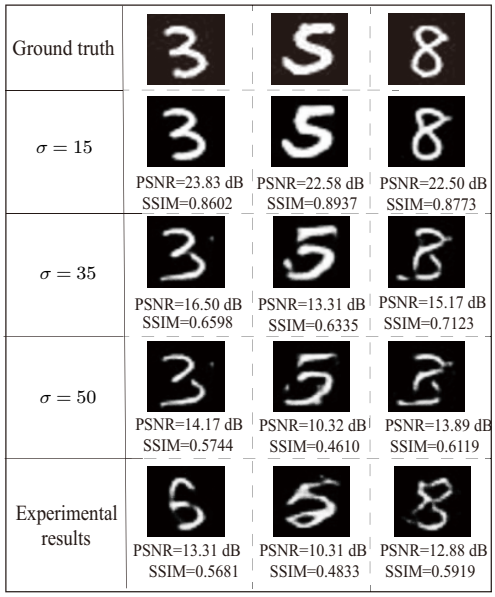


Fig. 6. Reconstruction performance under different noise intensities at a 1.2% sampling rate.

1.2%. The results demonstrate a decline in both PSNR and SSIM with increasing noise variance. When the standard deviation σ reaches 50, the simulation results closely approximate the experimental outcomes, exhibiting nearly identical PSNR and SSIM values. This indicates that environmental noise is a primary factor degrading the reconstructed image quality in practical experiments.

3.2. Comparisons

Figure 7 compares simulation-based reconstruction results on the MNIST dataset across multiple GI methods at progressively reduced sampling rates (1.2% to 0.29%). They are conventional GI (GI), a baseline cGAN without residual blocks and attention mechanisms,[32] a GAN integrated with the SENet attention module (SE-GAN),[33] and our proposed method. To quantitatively assess the reconstruction performance, the average SSIM and PSNR for reconstructed images are evaluated for digits (‘3’, ‘6’, and ‘8’) under each sampling

rate, as shown in Fig. 7.

The results demonstrate that under extremely low sampling rates (*e.g.*, 0.39% and 0.29%), conventional GI fails to produce recognizable reconstructions. While both the baseline cGAN and SE-GAN remain functional under these conditions, our method consistently achieves superior reconstruction quality. More importantly, this performance gain comes with a significant efficiency advantage. The integration of the SENet module in SE-GAN leads to substantial computational overhead, resulting in a training time approximately double that of our method. Consequently, our approach delivers not only enhanced performance but also markedly reduced training costs. This performance advantage enlarges significantly as sampling rates decrease further, demonstrating our method's enhanced capability under severely under-sampled conditions.

On the Fashion–MNIST dataset, the proposed imaging scheme demonstrates notably superior performance compared to other ghost imaging methods. As shown in Fig. 8(a), when the sampling rate is as low as 0.39% and 0.29%, the conventional cGAN fails to reconstruct images from the "sneaker" category effectively. By contrast, the proposed method consistently yields high-fidelity reconstruction results. Although the reconstructions of SE-GAN and the proposed method appear visually comparable, the PSNR and SSIM trends across different sampling rates in Figs. 8(b) and 8(c) quantitatively validate the performance advantage of our scheme.

To evaluate the respective contributions of the ECA module and the CBAM in our framework, we conduct an ablation study at 0.78% sampling rate, comparing the performance when both modules are employed versus when only one is used. All models were trained under identical settings in terms of dataset size and number of epochs. Figure 9 illustrates the reconstruction results for the full model (dual attention), CBAM-only, and ECA-only configurations. Quantitatively, the proposed method with dual attention mechanisms achieves higher PSNR and SSIM values than the single-attention variants. The ECA-only model performs the worst across most categories, with the exception of the "NUPT" image. We attribute the effectiveness of the dual-attention design to the complementary roles of each module: CBAM, with its spatial attention mechanism, is critical in preserving structural details during reconstruction. In our network, CBAM is applied at the fourth layer, while ECA operates at the third layer. This combination enables joint optimization across channel and spatial dimensions, leading to enhanced feature representation and superior reconstruction performance.

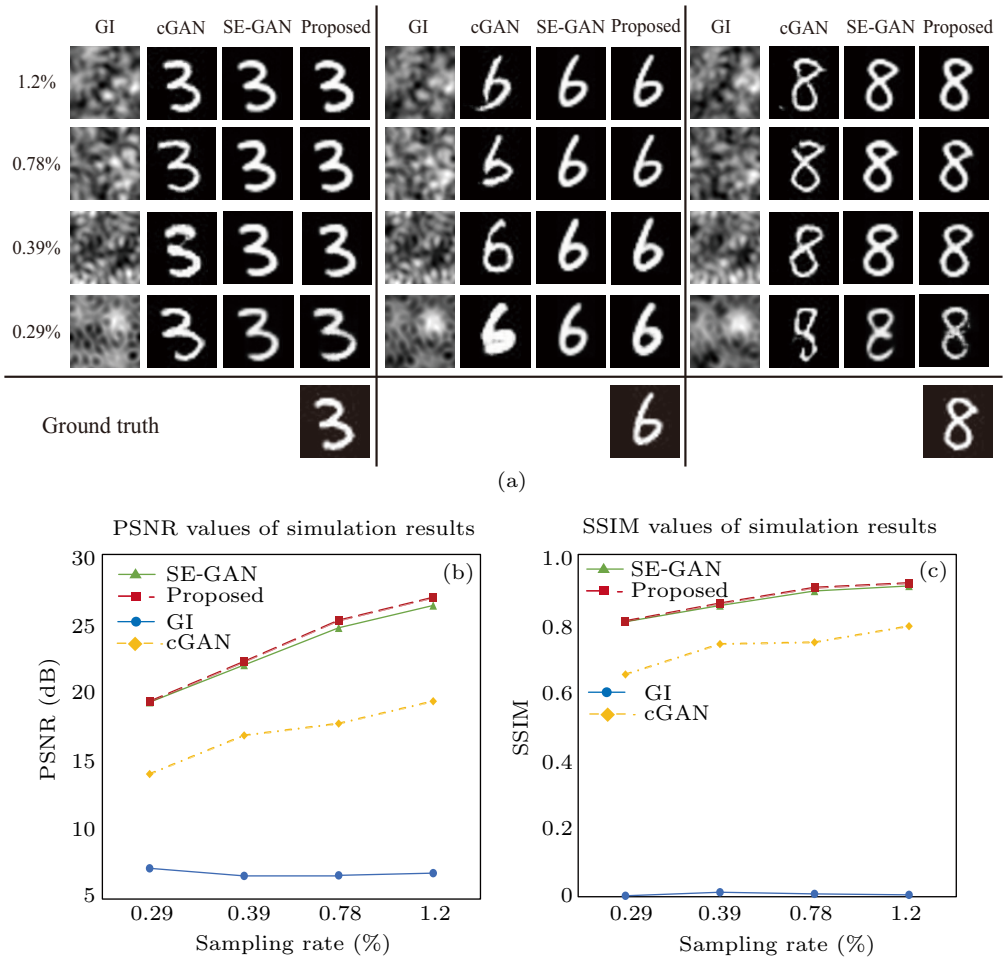


Fig. 7. The MNIST dataset analysis: (a) qualitative comparison of four reconstruction algorithms, (b) quantitative PSNR evaluation across test samples, and (c) quantitative SSIM evaluation across test samples.

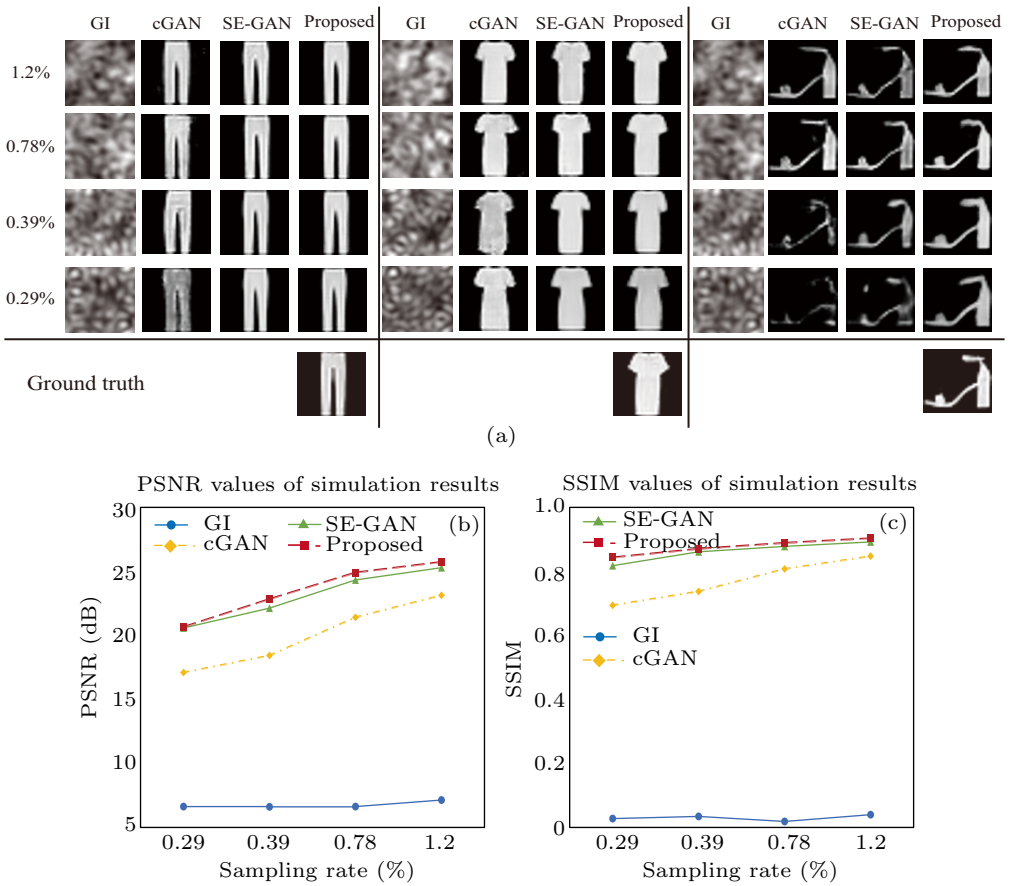


Fig. 8. The Fashion–MNIST dataset analysis: (a) qualitative visual comparison of reconstruction algorithms, (b) quantitative PSNR assessment across test samples, and (c) quantitative SSIM assessment across test samples.

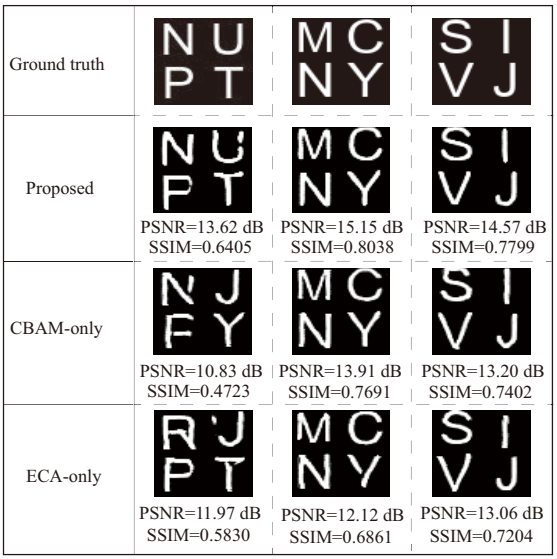


Fig. 9. The ablation experiments for the reconstruction results by the the full model (dual attention), CBAM-only, and ECA-only configurations at 0.78% sampling rate.

Furthermore, we conducted comparative evaluations of four distinct GI methodologies on both the MNIST and Fashion–MNIST datasets at a sampling rate of 1.2%. Quantitative analysis of the reconstructed images, with visual results for the MNIST dataset and Fashion–MNIST dataset presented in Fig. 10, demonstrates that, compared to the baseline model, the cGAN, and the SE-GAN, the proposed method achieves statistically significant improvements in both perceptual clarity and structural integrity. These qualitative enhancements are further corroborated by the quantitative metrics in Table 1, which show that the proposed method attains the highest mean PSNR and SSIM across both datasets. Collectively, these experimental results confirm that the proposed framework exhibits superior performance in three critical aspects of low-sampling-rate ghost imaging: reconstruction fidelity, computational efficiency, and noise robustness.

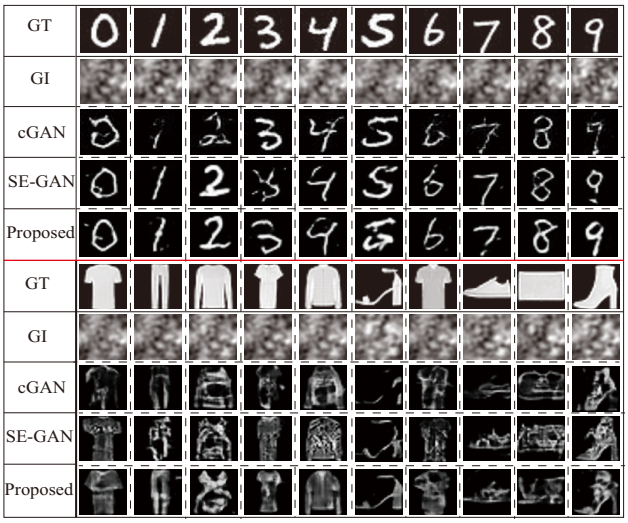


Fig. 10. Experimental results of different imaging methods at 1.2% sampling rate on MINST and Fashion–MNIST datasets.

Table 1. Average PSNR and SSIM values of different ghost imaging methods on the MNIST and Fashion–MNIST datasets at a sampling rate of 1.2%.

Dataset	Metric	GI	cGAN	SE-GAN	Proposed
MNIST	Avg PSNR	6.90	13.73	15.05	**15.84**
	Avg SSIM	0.0050	0.6288	0.6830	**0.7054**
Fashion–MNIST	Avg PSNR	7.13	7.39	8.32	**8.70**
	Avg SSIM	0.0364	0.3604	0.3704	**0.4013**

4. Conclusion

In summary, this paper presents a deep learning-based GI reconstruction method designed to achieve high-quality image recovery under low sampling rate conditions. The proposed approach builds upon a cGAN framework, in which residual blocks and attention mechanisms are strategically incorporated. Specifically, a U-net-style generator architecture is employed, with efficient attention modules embedded in the deeper layers of the generator to enhance the representation of critical image features. This design improves the network's expressiveness while maintaining computational efficiency and mitigating the risk of overfitting associated with excessive attention mechanisms.

The numerical simulations and experimental validations on the MNIST and Fashion–MNIST datasets demonstrate that the proposed method can effectively reconstruct clear and accurate images even at extremely low sampling rates (0.29% in ideal simulation and 0.78% in experiment). Compared to conventional GI techniques, the baseline cGAN model, and the SENet-integrated GAN (SE-GAN), the proposed method consistently achieves superior reconstruction quality. These results highlight the robustness and expressive capability of the proposed approach, indicating its strong potential for application in low-sampling-rate GI scenarios.

Acknowledgements

Project supported by the National Natural Science Foundation of China (Grant Nos. 62375140 and 62001249) and the 2024 Jiangsu Provincial Graduate Research Innovation Program (Grant No. KYCX24_1191).

References

[1] Cheng J and Han S 2004 *Phys. Rev. Lett.* **92** 093903
[2] Yu S, Kong L J, Li Y N, Tu C H and Wang H T 2016 *Chin. Phys. Lett.* **33** 034203
[3] Wang L and Zhao S M 2020 *Chin. Phys. B* **29** 024204
[4] Pittman T B, Shih Y H, Strekalov D V and Sergienko A V 1995 *Phys. Rev. A* **52** R3429
[5] Yao X, Yu W, Liu X, Li L, Li M, Wu L and Zhai G 2014 *Opt. Express* **22** 24268
[6] Wu H, Zhang X, Gan J, Luo C and Ge P 2016 *Laser Phys. Lett.* **13** 115205
[7] Zhang H and Duan D Y 2021 *Chin. Opt. Lett.* **19** 101101
[8] O-oka Y and Fukatsu S 2017 *Appl. Phys. Lett.* **111** 061106
[9] Guo H, Wang L and Zhao S M 2022 *Chin. Phys. B* **31** 084201
[10] Zhao Y N, Chen L S, Chen L Y, Kong L, Wang C, Ren C, Zhang S H and Cao D Z 2024 *Opt. Lasers Eng.* **181** 108408
[11] Shapiro J H 2008 *Phys. Rev. A* **78** 061802
[12] Stantchev R I, Sun B, Hornett S M, Hobson P A, Gibson G M, Padgett M J and Hendry E 2016 *Sci. Adv.* **2** e1600190
[13] Yu H, Lu R, Han S, Xie H, Du G, Xiao T and Zhu D 2016 *Phys. Rev. Lett.* **117** 113901
[14] Meyers R E, Deacon K S and Shih Y 2012 *Appl. Phys. Lett.* **100** 131114
[15] Chen L Y, Zhao Y N, Chen L S, Wang C, Ren C and Cao D Z 2024 *Opt. Laser Technol.* **169** 110007
[16] Chen J L, Zhu W, Wang L and Zhao S M 2025 *Chin. Phys. B* **34** 094210
[17] Wang L and Zhao S M 2021 *Opt. Lasers Eng.* **139** 106473
[18] Ren H, Zhao S M and Gruska J 2018 *Opt. Express* **26** 5501
[19] Wang L and Zhao S M 2018 *Photonics Res.* **4** 240
[20] Liu Y, Yu Z, Bai X, *et al.* 2023 *Phys. Scr.* **98** 115119
[21] Wu H, Wang R, Zhao G, *et al.* 2020 *Opt. Express* **28** 3846
[22] Li Y, Zhang L, Zhang D, *et al.* 2025 *Opt. Lasers Eng.* **190** 108953
[23] Lyu M, Wang W, Wang H, Wang H, Li G, Chen N and Situ G 2017 *Sci. Rep.* **7** 17865
[24] He X, Zhao S M and Wang L 2021 *Chin. Phys. B* **30** 054201
[25] Zhao S M, Cui Y, He X and Wang L 2022 *Front. Optoelectron.* **15** 31
[26] Ni Y, Zhou D, Yuan S, Bai X, Xu Z, Chen J, Li C and Zhou X 2021 *Opt. Lett.* **46** 1840
[27] Cao Y, Xiao Y, Pan Z, *et al.* 2022 *Opt. Lasers Eng.* **158** 107141
[28] Liu Y, Yu Z, Li X J, Sun M Z, Zhao D, Zhang J L, Li Y, Yuan S and Xin Z 2024 *Phys. Scr.* **99** 075114
[29] Ye H L and Guo D D 2023 *Opt. Lasers Eng.* **171** 107790
[30] Ge H Y, Yin L F, Cui X K, Zhu L Y, Chen L and Wu G H 2025 *Opt. Laser Technol.* **184** 112542
[31] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2020 *Commun. ACM* **63** 139
[32] Isola P, Zhu J Y, Zhou T and Efros A A 2017 2017 *IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 5967–5976
[33] Wang M, Jang L and Wang H 2024 *The 5th International Symposium on Computer Engineering and Intelligent Communications* (ISCEIC), pp. 567–572