0.5% Nyquist noise-free computational ghost imaging via non-experimental deep learning

HAOTIAN SONG[1,2,*], XIAOYU NIE[1,3,*], HAIRONG SU[4], HUI CHEN[5], YU ZHOU[1], XINGCHEN ZHAO[3], TAO PENG[3,†], AND MARLAN O. SCULLY[3,6,7]

[1] Physics Department, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[2] College of Physics & Astronomy, University of Manchester, Manchester M13 9PL, UK
[3] Department of Physics & Astronomy, Texas A&M University, College Station, Texas, 77843, USA
[4] National Engineering Laboratory for Big Data Analytics, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[5] Electronic Materials Research Laboratory, Key Laboratory of the Ministry of Education & International Center for Dielectric Research, Xi'an Jiaotong University, Xi'an, 710049, China
[6] Baylor Research and Innovatuon Collaborative, Baylor University, Waco, 76706, USA
[7] Princeton University, Princeton, New Jersey 08544, USA
[*] these two authors contribute equally to this work.
[†] taopeng@tamu.edu

We present a novel framework for computational ghost imaging based on deep learning and pink noise patterns, which substantially decreases the sampling ratio over 10 times smaller than previous sub-Nyquist computational ghost imaging works. Here, the deep neural network, which can learn the sensing model and enhance the quality of image reconstruction, is trained merely by simulation results. There is no necessity to conduct experiments to get training inputs (non-experimental) and add noise to customize with a real imaging system (noise-free). This one-time trained network can be applied to multiple environments and various situations. To demonstrate the sub-Nyquist level in our achievement, the conventional computational ghost imaging results, imaging results reconstructed using white noise, and pink noise via deep learning are compared in several sampling rates. To indicate its non-experimental and noise-free advantages, a group of results with strong environmental noise are presented. This method has great potentials in various applications that require a low sampling rate, quick reconstruction efficiency, and strong turbulence.

1. INTRODUCTION

Ghost imaging (GI) [1–4], a novel practice with measuring the spatial correlations between two light beams, where the signal light field interacts with the object then collected by a single-pixel detector, and the reference light field, which never interacts with the object, falls onto the imaging detector. Therefore, the image information is not present in either beam alone but only revealed in their correlations. To further ameliorate and simplify this framework, computational ghost imaging (CGI) [5, 6] was proposed. The reference light path that functions as recording speckles is replaced by loading pre-generated patterns directly onto the spatial light modulator or the digital micro-mirror device (DMD). The unconventional image is then revealed by correlating the sequentially recorded intensities at the single-pixel detector and their corresponding patterns. CGI finds a lot of applications such as wide spectrum imaging [7–9], remote sensing [10], and quantum-secured imaging [11].

However, CGI normally requires a large number of samplings to reconstruct a high-quality image, or the signal would have been submerged under correlation fluctuations and environmental noise. To suppress the environmental noise and correlation fluctuations, the required minimum number of sampling is proportionate to the total number of pixels within patterns that are going to be applied on DMD, *i.e.*, the Nyquist limit of sampling [12, 13].The image could have a meager signal-to-noise ratio (SNR) if the sampling number does not satisfy the requirement. This demanding requirement hindered CGI fully replacing the conventional photography. Previously, many schemes have been

proposed to improve CGI's speed and decrease the number of sampling (sub-Nyquist) [14–21]. Compressive sensing imaging is capable of reconstructing imaging with a relatively low sampling rate by exploiting the sparsity of the objects [22–24]. It nevertheless largely depends on the sparsity of objects and is sensitive to noise [25]. Orthonormalized noise patterns can be used to suppress the noise and improve the image's quality under a limited sampling number [16, 26], in which the orthonormalized colored noise patterns can break the Nyquist limit down to ∼ 5% [26]. Fourier and sequency-ordered Walsh-Hadamard patterns, which are orthogonal to each other in time or spatial domain, were also applied to the sub-Nyquist imaging [17–19]. The Russian doll [20] and cake-cutting [21] ordering of Walsh-Hadamard patterns were proposed to minimize the sampling ratio to 5%-10% Nyquist limit.

Recently, deep learning (DL) is employed to improve the quality of images with the deep neural network (DNN) [27–35] or merely identify images [36, 37]. Specifically, computational ghost imaging via deep learning (CGIDL) has shown a minimum ratio of Nyquist limit down to ∼ 5% [29, 31]. However, such work's DNNs are trained by experimental CGI results. It is because only when the training environment is highly identical to the environment used for image reconstruction, can the DNN be effective. Otherwise, the training network without adding environmental noise cannot be accustomed to the real noisy situation. This limits its universal applications and restricts it to achieve quick reconstructions in that usually over thousands of inputs have to be generated, which would be very time-consuming if conducting training experiments. On the other hand, some work have attempted to try the effectiveness of non-experimental CGI training DNN [27, 30, 33], but the minimum ratios of the Nyquist limit were up to 12%, 6.25%, and 3.5%, respectively. Namely, the work reached a 3.5% Nyquist sampling rate inevitably added simulation noise [27], in other words, the customizing process is closed to the experimental DL above. Meanwhile, for an arbitrary object outside of training targets, the sampling ratio is much higher than those objects within the training group [29]. Therefore, those works applying objects within the training group in CGIDL experiments hardly stand the points [28, 33]. In a word, retrieving high-quality images outside of the training group with a meager Nyquist limit ratio by non-experimental training in a noisy environmental situation remains a challenge for the CGI system.

This letter aims to minimize the necessary sampling number further and improve the SNR with the combination of DL and pink noise CGI. The pink noise CGI owns positive a cross-correlations in the second-order correlation [38, 39]. It gives a good quality image under a boisterous environment or pattern distortion when the traditional CGI method fails. Combining DL with pink noise CGI shows that the imaging can be retrieved under a meager sampling rate (∼ 0.5%). We also show that we can get training patterns from the simulation without introducing the environmental noises, *i.e.*, there is no need to get DNN training with a large number of experimental training inputs. We can get training patterns from the simulation on the computer without worrying about environmental noise. Another novelty demonstrated in this work is that the object used in the experiment can be independent of the digit used in training, which can largely benefit CGIDL in the real application.

The experimental setup is shown in Fig. 2. A CW laser is used to illuminate the DMD, where the noise patterns are loaded on it. The pattern generated by the DMD is then projected onto the object. In our experiment, the size of the noise patterns is

216 by 392 pixels, in which the independent changeable mirrors count for 4 by 4 pixels. The DMD contains tiny pixel-mirrors each measuring $16\ \mu m \times 16 \mu m$.
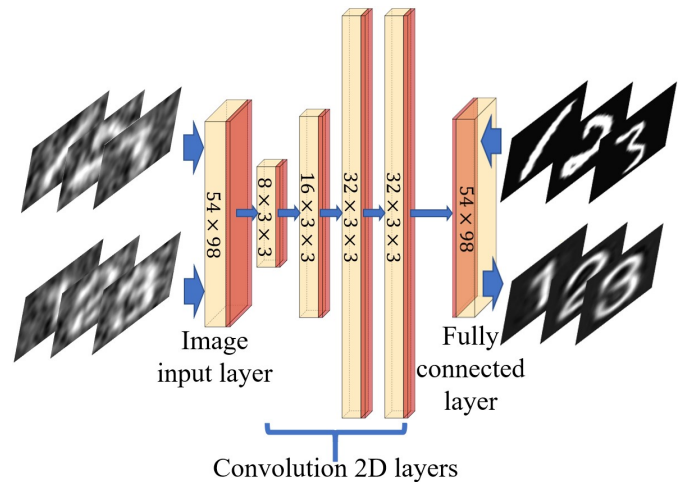

Fig. 1. Schematic of DNN. Our DNN consists of four convolution layers, one image input layer, one fully connected layer (in yellow), ReLU and BNL (in red). Patterns in the upper line are CGI results (training inputs) and handwriting ground truths (training outputs) respectively; Patterns in the bottom line are CGI results from the experiment (test inputs) and CGIDL results (test outputs) with printed body.

2. DEEP LEARNING

The proposed scheme, as shown in Fig. 1, is a two-step process to reconstruct the CGIDL. Firstly, we construct a network framework which is going to be trained later. Specifically, we used a DNN model with four convolution layers, one image input layer, and one fully connected layer. Small 3×3 receptive fields were applied throughout the whole convolution layers, which have been proved to perform better by Google-Net [40]. Batch normalization layers (BNL) and rectified Linear Unit (ReLU) layers were added between each convolution layer. The BNL is functioned to avoid internal covariate shift during the training process and speeds up the training of DNN [41]. The ReLU layer applies a threshold operation to each element of the inputs [42]. To customize the size of training pictures, both the input and output layers were set as 54×98. The solver for training is employed by the Stochastic Gradient Descent with Momentum Optimizer (SGDMO), thus reducing the oscillation via using momentum. The parameter vector can be updated via equation Eq. (1), which demonstrates the updating process during the iteration.

$$\theta_{\ell+1} = \theta_\ell - \alpha \nabla E(\theta_\ell) + \gamma (\theta_\ell - \theta_{\ell-1}) \tag{1}$$

Here ℓ is the iteration number, α is the learning rate, θ is the parameter vector, and $E(\theta)$ is the loss function. The third part of the equation is the feature of SGDMO, analog to the momentum where γ determines the contribution of the previous gradient step to the current iteration [43]. Two strategies were applied to avoid over-fitting of training images. At the end of DNN, a dropout layer is aimed to reduce the connection between convolution layers and the fully connected layer [44]. Meanwhile, the learning rate dropped from 0.001 to 0.0001 after 75 epochs, which constrain the fitting parameters within reasonable region.

After the establishment of DNN, plenty of training images are reconstructed by the CGI algorithm, as is mentioned above. Then the training images and reconstruction training images feed the DNN model as inputs and outputs, respectively. Here we use a set of 10000 handwritten digits of 28×28 pixels in size from the MNIST handwritten digit database [45] as training images. Compared to the original training images, we resize images from 28×28 to 54×98 and normalize them to test a smaller sampling ratio. The maximum epochs were set as 600, and the training iteration is 46800. The program was implemented via MATLAB R2019a Update 5 (9.6.0.1174912, 64-bit), and the DNN was implemented through DL Toolbox. The GPU-chip NVIDIA GTX1050 was used to accelerate the speed of the computation.
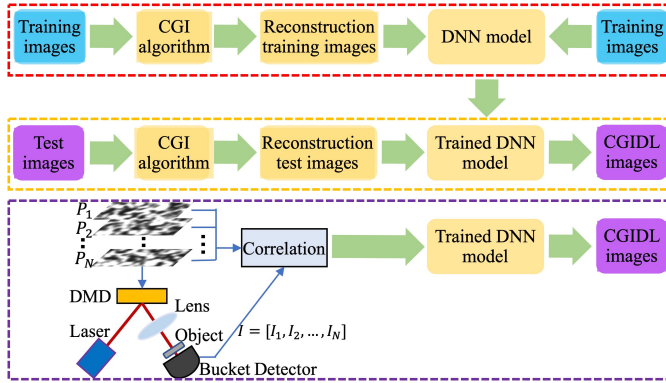

Fig. 2. The flow chart of CGIDL. The DNN model is trained in the training process (in red), and test images are for the testing stage (in orange). The experiment process is on the left bottom functioned as the input of trained DNN, and the experimental DL process is on the right bottom side (in purple).

After obtaining a trained DNN, the next step is testing the DNN by simulation and retrieving CGI results in the experiment. The schematic is shown in Fig. 2. In the testing part, the CGI algorithm generates reconstruction testing images from testing images different from images in the training group. The trained DNN, fed with reconstruction testing images, generates CGIDL results. Comparing the difference between CGIDL and testing images, we could measure the quality of the trained DNN. Well-performed DNN can be used for retrieving CGI in the experiment.

In the CGI process, the SNR of images is proportional to the measurement ratio. Therefore, the ratio between the number of illumination patterns $N_{pattern}$ and the (average) number of speckle in each of these patterns N_{pixel} [46, 47], namely,

$$\beta = N_{pattern} / N_{pixel} \qquad (2)$$

is used in our work to reflect SNR. Here we present several results concerning a couple of β by applying pink noise and conventional white noise patterns. Meanwhile, we also calculate the Mean Square Error (MSE) to justify the quality of the reconstructed image by CGI and CGIDL with pink and white noise, respectively. The MSE is defined as

$$MSE = \frac{1}{N_{pixel}} \sum_{i=1}^{N_{pixel}} [\frac{G_i - X_i}{\langle G_{(o)} \rangle}]^2 \qquad (3)$$

Here, X is the reference matrix calculated by

$$X_i = \begin{cases} \langle G_{(o)} \rangle & , \text{Transmission} = 1 \\ \langle G_{(b)} \rangle & , \text{Transmission} = 0 \end{cases} \qquad (4)$$

$G_{(o)}$ represents pixels in the correlation results that the light ought to be transmitted, *i.e.*, the object area, while $G_{(b)}$ represents pixels in the correlation results that the light ought to be blocked, *i.e.*, the background area.
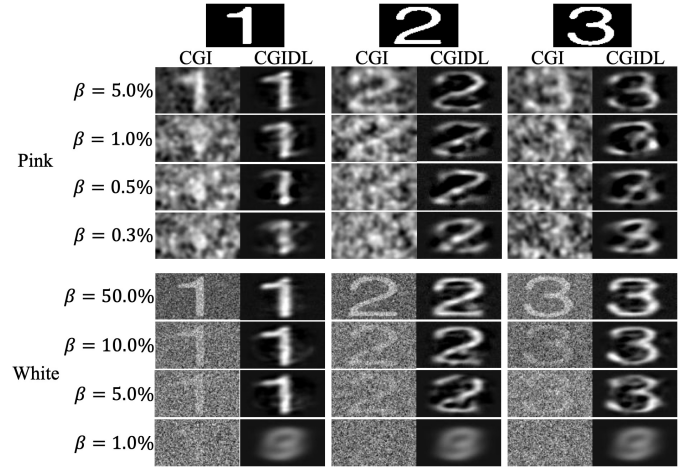
3. SIMULATION

Fig. 3. Main simulation results. For pink noise, we select $\beta = 0.05, 0.01, 0.005, 0.003$. For white noise when β equals to these small values, the results are totally smeared both under CGI and CGIDL. Thus, here we present $\beta = 0.5, 0.1, 0.05, 0.01$ for white noise to make comparison.
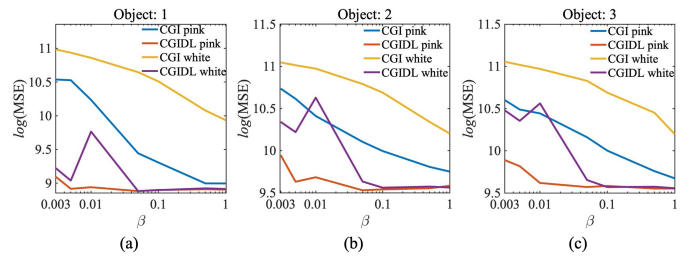

Fig. 4. The MSE of object (a): 1, (b): 2, and (c): 3 in the simulation.

To obtain the trained DNN, we carried out simulation using 10000 training images with different β. Patterns used in our simulation were white noise and pink noise. Because white noise pattern is globally used in CGI while pink noise pattern can significantly improve the image quality, we select β as 0.5, 0.1, 0.05, and 0.01 for white noise, 0.05, 0.01, 0.005, and 0.003 for pink noise. Then several testing images (digits '1', '2', and '3'), which are completely independent of training images, were chosen as examples in our simulation. These images have 28×28 pixels, and are resized into 54×98 by widening and amplification. As shown in Fig. 3, both CGI and CGIDL images are de-noised and become clearer and brighter with increasing β. For smaller β, CGIDL can be visualized while CGI can not, reflecting the advantage of DL. Meanwhile, testing images using white noise patterns became unrecognizable when $\beta = 0.05$, and we can

optimize them using DL, which is similar to work with [28]. But when β is 0.01, both CGI and CGIDL failed because of the lack of averaging during correlation. Nevertheless, the β of CGIDL using pink noise can reach 0.003, through which we can still distinguish these numbers. To directly justify the quality of reconstructed images, we compare the MSE of all figures above in Fig. 4. We can clearly see that the CGI with pink noise patterns already owns a great advantage in sub-Nyquist imaging than conventional white noise. With the combination of DL, it can reach to very small β with low MSE. The MSE of CGIDL with white noise substantially increases when $\beta < 0.05$ and oscillate due to the network of DL could not distinguish the shape of the original image. Therefore it becomes a random process in amelioration, and would be some probability that results with a lower sampling ratio have better quality after the DL process.

In a word, we realize 0.3% Nyquist reconstruction via CGIDL and pink noise patterns in the simulation.
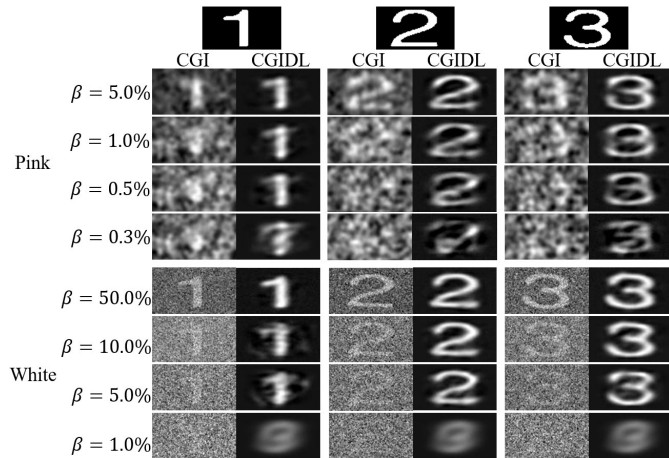
4. EXPERIMENT

Fig. 5. Main experimental results with the SNR of 14.68dB. For pink noise, we select $\beta = 0.05, 0.01, 0.005, 0.003$. For white noise when β equals to these small values, the results are totally smeared both under CGI and CGIDL. Thus, we present $\beta = 0.5, 0.1, 0.05, 0.01$ for white noise to make comparison.

To further demonstrate the advantage and applicability of CGIDL with pink noise, we performed experiments with two core directions. First of all, we explored the bottom limits of the sampling ratio where we control the noise to the bottom limit in the experiment. Then, we increase the environmental noise to test the non-experimental and one-time training eligibility and its noise-free feature. The main results are shown in Fig. 5. The signal to noise ratio (SNR) can be calculated by

$$SNRdB = 10log\frac{P_{signal}}{P_{background}} \qquad (5)$$

Here the average signal $P_{signal} = 247$, and the average background $P_{background} = 8$, so the SNR is $14.90dB$. Intuitively, we could claim that we can still get great results with pink noise CGIDL when $\beta = 0.005$. Meanwhile, we can see that the CGI without DL results can be clearly distinguished while $\beta = 0.05$ and $\beta = 0.1$ respectively with pink and white noise patterns. We present the MSE as well in Fig. 6.
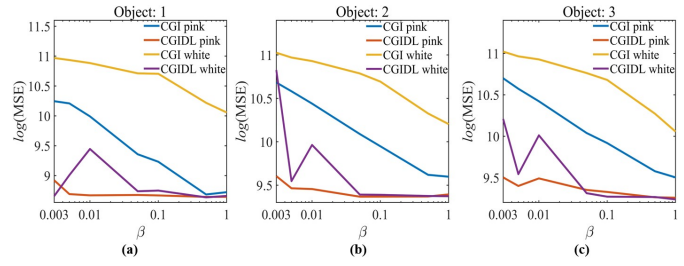

Fig. 6. The MSE of object (a): 1, (b): 2, and (c): 3 in the experiment with SNR at 14.90dB.
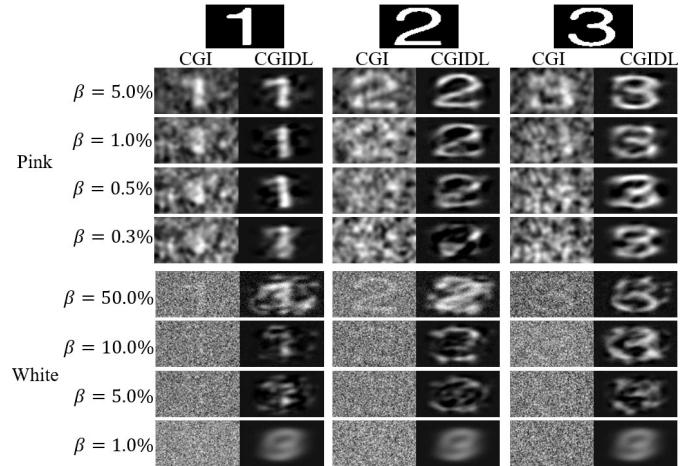

Fig. 7. Main experimental results with the SNR of 4.14dB. For pink noise, we select $\beta = 0.05, 0.01, 0.005, 0.003$. However, for white noise when β equals to these small values, the results are totally smeared both under CGI and CGIDL. Thus, here we present $\beta = 0.5, 0.1, 0.05, 0.01$ for white noise to make comparison.

Secondly, we increase the ratio of environmental noise in which the average signal P_{signal} equals to 18 and the average background $P_{background}$ equals to 6. Hence, the SNR is 4.77dB. The main results are shown in Fig. 7. We could see a distinctive change both in white noise CGI and CGIDL results. The CGIDL results with pink noise still maintain their MSE value around 9, while the CGIDL results with white noise raise around 0.5 correspondingly. We cannot get clear numbers even when β reaches to 1. In contrast, the CGI and CGIDL results with pink noise patterns always maintain their shapes no matter whether there exists noise or not. This demonstrates the noise-free feature of pink noise patterns [38], and we utilize this feature to make up the shortcomings of DL substantially. We present the MSE as well in Fig. 8. The MSE of white noise CGIDL changes randomly in that the DNN cannot readout any feature from the input of CGI with white noise, which is smeared by environmental noise. The DNN stochastically changes the structure of input so that the MSE randomly oscillates with probability. We cannot even say at what β it will give the optimum results.

In a word, experimental CGIDL with pink noise patterns can reconstruct the images at 0.5% Nyquist limit both the low and high noise level. However, the conventional CGIDL with white noise patterns only can reach to 5% Nyquist limit in the condition of low noise level. Otherwise, the conventional CGIDL method fails.
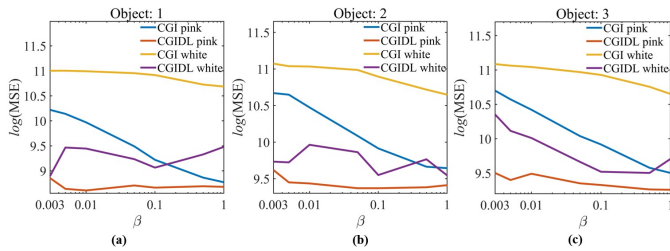
Fig. 8. The MSE of object (a): 1, (b): 2, and (c): 3 in the experiment with SNR at 4.77dB.

5. CONCLUSION

In conclusion, our new method of CGIDL with pink noise patterns achieves one-tenth of the traditional CGIDL sampling rate both in simulation and experiments. Rather than other previous works on CGIDL, our method does not require experimental training. Besides, our work is robust to noise so one group of trained networks is eligible to be implemented in various situations. This one-time, noise-free, and non-experimental training CGIDL has a wide range of application prospects. The pink noise patterns, trained DNN with various sampling rates, and their raw encoding programs are encapsulated and uploaded online [1].People who need a quick sampling function on CGIDL can utilize this universal system to get ameliorated results in other CGIDL systems.

6. FUNDING.

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**, R3429–R3432 (1995).
2. R. S. Bennink, S. J. Bentley, and R. W. Boyd, ""two-photon" coincidence imaging with a classical source," Phys. Rev. Lett. **89**, 113601 (2002).
3. A. Valencia, G. Scarcelli, M. D'Angelo, and Y. Shih, "Two-photon imaging with thermal light," Phys. Rev. Lett. **94**, 063601 (2005).
4. X.-H. Chen, Q. Liu, K.-H. Luo, and L.-A. Wu, "Lensless ghost imaging with true thermal light," Opt. Lett. **34**, 695–697 (2009).
5. Y. Bromberg, O. Katz, and Y. Silberberg, "Ghost imaging with a single detector," Phys. Rev. A **79**, 053840 (2009).
6. J. H. Shapiro, "Computational ghost imaging," Phys. Rev. A **78**, 061802 (2008).
7. D. Shrekenhamer, C. M. Watts, and W. J. Padilla, "Terahertz single pixel imaging with an optically controlled dynamic spatial light modulator," Opt. Express **21**, 12507–12518 (2013).
8. R. S. Aspden, N. R. Gemmell, P. A. Morris, D. S. Tasca, L. Mertens, M. G. Tanner, R. A. Kirkwood, A. Ruggeri, A. Tosi, R. W. Boyd *et al.*, "Photon-sparse microscopy: visible light imaging using infrared illumination," Optica **2**, 1049–1052 (2015).
9. Y. Klein, A. Schori, I. Dolbnya, K. Sawhney, and S. Shwartz, "X-ray computational ghost imaging with single-pixel detector," Opt. Express **27**, 3284–3293 (2019).
10. N. D. Hardy and J. H. Shapiro, "Computational ghost imaging versus imaging laser radar for three-dimensional imaging," Phys. Rev. A **87**, 023820 (2013).
11. P. Clemente, V. Durán, E. Tajahuerce, J. Lancis *et al.*, "Optical encryption based on computational ghost imaging," Opt. Lett. **35**, 2391–2393 (2010).
12. R. L. Cook, "Stochastic sampling in computer graphics," ACM Transactions on Graph. (TOG) **5**, 51–72 (1986).
13. J. A. Tropp, J. N. Laska, M. F. Duarte, J. K. Romberg, and R. G. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," IEEE transactions on information theory **56**, 520–544 (2009).
14. S. Jiang, X. Li, Z. Zhang, W. Jiang, Y. Wang, G. He, Y. Wang, and B. Sun, "Scan efficiency of structured illumination in iterative single pixel imaging," Opt. Express **27**, 22499–22507 (2019).
15. Z.-H. Xu, W. Chen, J. Penuelas, M. Padgett, and M.-J. Sun, "1000 fps computational ghost imaging using led-based structured illumination," Opt. Express **26**, 2427–2434 (2018).
16. Bin, Luo, Pengqi, Yin, Longfei, Guohua, Wu, Hong, and Guo, "Orthonormalization method in ghost imaging." Opt. Express (2018).
17. Z. Zhang, X. Ma, and J. Zhong, "Single-pixel imaging by means of fourier spectrum acquisition," Nat. communications **6**, 1–6 (2015).
18. L. Wang and S. Zhao, "Fast reconstructed and high-quality ghost imaging with fast walsh–hadamard transform," Photonics Res. **4**, 240–244 (2016).
19. Z. Zhang, X. Wang, G. Zheng, and J. Zhong, "Hadamard single-pixel imaging versus fourier single-pixel imaging," Opt. Express **25**, 19619–19639 (2017).
20. M.-J. Sun, L.-T. Meng, M. P. Edgar, M. J. Padgett, and N. Radwell, "A russian dolls ordering of the hadamard basis for compressive single-pixel imaging," Sci. Reports **7**, 1–7 (2017).
21. W.-K. Yu, "Super sub-nyquist single-pixel imaging by means of cake-cutting hadamard basis sort," Sensors **19**, 4122 (2019).
22. O. S. Magana-Loaiza, G. A. Howland, M. Malik, and J. C. Howell, "Compressive object tracking using entangled photons," Appl. Phys. Lett. **102** (2013).
23. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," Appl. Phys. Lett. **95**, 739 (2009).
24. C. Yi, C. Zhengdong, F. Xiang, C. Yubao, and L. Zhenyu, "Compressive sensing ghost imaging based on image gradient," Optik - Int. J. for Light. Electron Opt. (2019).
25. J. Du, W. Gong, and S. Han, "The influence of sparsity property of images on ghost imaging with thermal light," Opt. Lett. **37**, 1067–1069 (2012).
26. X. Nie, X. Zhao, T. Peng, and M. O. Scully, "Sub-nyquist computational ghost imaging with orthonormalized colored noise pattern," arXiv preprint arXiv:2012.07250 (2020).
27. F. Wang, H. Wang, H. Wang, G. Li, and G. Situ, "Learning from simulation: An end-to-end deep-learning approach for computational ghost imaging," Opt. Express **27**, 25560–25572 (2019).
28. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," Sci. Reports **7**, 17865 (2017).
29. H. Wu, R. Wang, G. Zhao, H. Xiao, D. Wang, J. Liang, X. Tian, L. Cheng, and X. Zhang, "Sub-nyquist computational ghost imaging with deep learning," Opt. Express **28**, 3846–3853 (2020).
30. T. Shimobaba, Y. Endo, T. Nishitsuji, T. Takahashi, Y. Nagahama, S. Hasegawa, M. Sano, R. Hirayama, T. Kakue, A. Shiraki *et al.*, "Computational ghost imaging using deep learning," Opt. Commun. **413**, 147–151 (2018).
31. Y. He, G. Wang, G. Dong, S. Zhu, H. Chen, A. Zhang, and Z. Xu, "Ghost imaging based on deep learning," Sci. Reports **8**, 6469 (2018).
32. S. Rizvi, J. Cao, K. Zhang, and Q. Hao, "Deepghost: real-time computational ghost imaging via deep learning," Sci. Reports **10**, 1–9 (2020).
33. H. Wu, R. Wang, G. Zhao, H. Xiao, J. Liang, D. Wang, X. Tian, L. Cheng, and X. Zhang, "Deep-learning denoising computational ghost imaging," Opt. Lasers Eng. **134**, 106183 (2020).
34. T. Bian, Y. Yi, J. Hu, Y. Zhang, Y. Wang, and L. Gao, "A residual-based deep learning approach for ghost imaging," Sci. Reports **10**, 1–8 (2020).
35. X. Zhai, Z.-d. Cheng, Y. Chen, Z.-y. Liang, Y. Wei *et al.*, "Foveated ghost imaging based on deep learning," Opt. Commun. **448**, 69–75 (2019).

---

[1]https://github.com/XJTU-TAMU-CGI/CGIDL

36. X. He, S.-M. Zhao, and L. Wang, "Handwritten digit recognition based on ghost imaging with deep learning," Chin. Phys. B (2020).

37. J. Li, M. Le, J. Wang, W. Zhang, B. Li, and J. Peng, "Object identification in computational ghost imaging based on deep learning," Appl. Phys. B **126**, 1–10 (2020).

38. X. Nie, F. Yang, X. Liu, X. Zhao, R. Nessler, Z. Li, T. Peng, M. S. Zubairy, and M. O. Scully, "Noise-free computational ghost imaging with pink noise speckle patterns," (2020).

39. X. Nie, X. Zhao, T. Peng, and M. O. Scully, "Moving object captured with pink noise pattern in computational ghost imaging," arXiv preprint arXiv:2012.07284 (2020).

40. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition,* (2015), pp. 1–9.

41. S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167 (2015).

42. V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th International Conference on International Conference on Machine Learning,* (Omnipress, Madison, WI, USA, 2010), ICML'10, pp. 807–814.

43. K. P. Murphy, *Machine learning: a probabilistic perspective* (MIT press, 2012).

44. N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," J. Mach. Learn. Res. **15**, 1929–1958 (2014).

45. L. Deng, "The mnist database of handwritten digit images for machine learning research [best of the web]," IEEE Signal Process. Mag. **29**, 141–142 (2012).

46. F. Ferri, D. Magatti, L. Lugiato, and A. Gatti, "Differential ghost imaging," Phys. review letters **104**, 253603 (2010).

47. W. Wang, X. Hu, J. Liu, S. Zhang, J. Suo, and G. Situ, "Gerchberg-saxton-like ghost imaging," Opt. Express **23**, 28416–28422 (2015).