SCIENTIFIC REPORTS
natureresearch

Check for updates

OPEN

DeepGhost: real-time computational ghost imaging via deep learning

Saad Rizvi, Jie Cao✉, Kaiyu Zhang & Qun Hao✉

The potential of random pattern based computational ghost imaging (CGI) for real-time applications has been offset by its long image reconstruction time and inefficient reconstruction of complex diverse scenes. To overcome these problems, we propose a fast image reconstruction framework for CGI, called "DeepGhost", using deep convolutional autoencoder network to achieve real-time imaging at very low sampling rates (10–20%). By transferring prior-knowledge from STL-10 dataset to physical-data driven network, the proposed framework can reconstruct complex unseen targets with high accuracy. The experimental results show that the proposed method outperforms existing deep learning and state-of-the-art compressed sensing methods used for ghost imaging under similar conditions. The proposed method employs deep architecture with fast computation, and tackles the shortcomings of existing schemes i.e., inappropriate architecture, training on limited data under controlled settings, and employing shallow network for fast computation.

Computational ghost imaging[1] acquires spatial information about an unknown target by illuminating it with a series of random binary patterns generated by a spatial light modulator (SLM). For each projected pattern, the light intensity back-reflected from the target plane is recorded by an ordinary photodiode. By correlating intensity measurements with corresponding projected patterns, the target image is reconstructed. One downside of CGI is the requirement of a large number of measurements to produce a good-quality image, which increases its imaging time. Despite the emergence of basis scan schemes[2], CGI (using random patterns) is still employed in many applications due to its simplicity, inherent encryption of patterns[3], and ease of deployment[4]. Therefore, it is important to improve the efficiency of CGI by integrating it with some optimization technique to avoid complex (hardware based) methods[5] that fail to reap the benefits of reduced cost and simplicity in ghost imaging (GI). Owing to its advantages of low cost, robustness against noise and scattering, and ability to operate over long spectral range, CGI is widely used in many applications[6–8].

In order to make CGI practical, more specifically for real-time imaging, it is important to reduce its imaging time. The imaging time of CGI can be sub-categorized as data acquisition time and image reconstruction time. The data acquisition time of CGI depends on the required number of measurements and mainly on the projection rate of SLM. Recent advances in SLM technology make it easy to reduce data acquisition time by employing commercially available high-resolution digital micromirror devices (DMDs) operating at ~ 20 kHz. The acquisition time can also be reduced by employing some simple yet novel solutions[9,10]. Therefore, the image reconstruction time remains the main bottleneck towards achieving high speed imaging in CGI. This image reconstruction time can be reduced by employing an efficient image reconstruction framework.

Recently, compressive sensing (CS) techniques[11] have been applied to recover an image with fewer (compressive) measurements. Although a promising technique, CS suffers from two inherent problems. First, to reconstruct an image from a few samples, CS algorithms require prior knowledge about the scene. However, for practical applications, images may not be sparse in a fixed basis, thereby limiting application flexibility. Second, the computational cost associated with most high-performance CS algorithms is very high, which increases reconstruction time, hence restricting their use in real-time applications. Although CS has been applied successfully in GI[12], fast image reconstruction requires an alternative advanced method.

Recent years have seen the rise of Deep learning (DL) as a powerful technique for solving complex problems in computational imaging[13]. DL has the potential to significantly enhance the performance of GI for real-time applications. For some years, the GI community remained skeptic about using DL for fast image reconstruction,

School of Optics and Photonics, Beijing Institute of Technology, Key Laboratory of Biomimetic Robots and Systems, Ministry of Education, Beijing 100081, China. ✉email: caojie@bit.edu.cn; qhao@bit.edu.cn

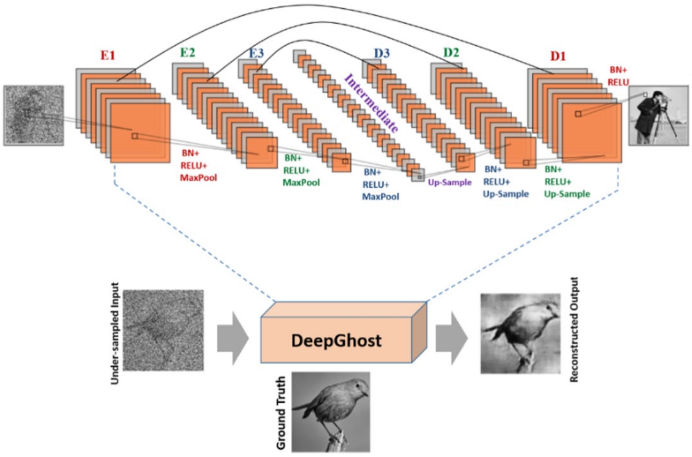

Figure 1. DeepGhost network architecture.

Figure 2. Sample images from 10 classes used for training.

relying on basic correlation and probabilistic methods for target detection[14,15]. Recently, there have been some interesting studies that explore the potential of DL for GI[16–20]. For GI, the most relevant deep neural network model is the denoising autoencoder[21]. An autoencoder can be used as an unsupervised feature learner to extract features from high-dimensional data in a systematic fashion. For GI, the autoencoder model can be used to recover a clean image from an undersampled ghost image reconstructed from fewer measurements, thus reducing reconstruction time.

The existing DL methods applied to CGI have limited applicability due to: (a) inappropriate architecture, (b) training on limited data or targets, and (c) employing shallow network for real-time operation. These schemes can work under controlled settings but fail when tested on a large dataset with complex scenes and measurement noise. For example, in Ref.[16] a stacked neural network model was used, confirming the potential of DL in CGI. The model employs a shallow fully connected network which is known to have computational complexity and is prone to data overfitting[22]. The model seems to work well with MNIST dataset, but its fully connected architecture is not suitable for complex image analysis. For image analysis, a more apt choice is the convolutional neural network (CNN)[23]. The work presented in Ref.[17] proposed a better (autoencoder) model based on CNN for CGI. However, the network was only trained for a particular object with limited training dataset, therefore not utilizing the true power of CNN.

In this paper, we demonstrate a CGI system that employs deep convolutional autoencoder network (DCAN) to reconstruct real-time images, using only a photodiode and random binary patterns for target scanning. The proposed DCAN (called "DeepGhost") strikes a balance between depth of layers and computation speed by employing a novel architecture for improved image recovery and fast network convergence. By employing innovations such as augmentation and transfer-learning, the proposed method can image complex unseen targets with high efficiency. Through simulations and experiments, we validate the superiority of our model by comparing it with existing DL[16,17] and state-of-the-art compressive sensing algorithms[24] used for GI under similar conditions.

Results

Simulations.

The network architecture for DeepGhost is shown in Fig. 1. The idea is to feed the network with undersampled (10%, 15%, 2 0%, and 25%) target images (acquired from CGI setup) for clear target reconstruction. The proposed network is optimized for physical imaging setup by exhaustively testing through numerical simulations. For training and testing, STL-10[25] dataset is used, which comprises of 10 classes: monkey, cat, dog, deer, car, truck, airplane, bird, horse, and ship. Sample image from each class is shown in Fig. 2.

Comparison with conventional and CS algorithms.

First, the performance of DeepGhost is evaluated through comparison with differential ghost imaging (DGI[26]) and compressive sensing methods[24]. The Deep-

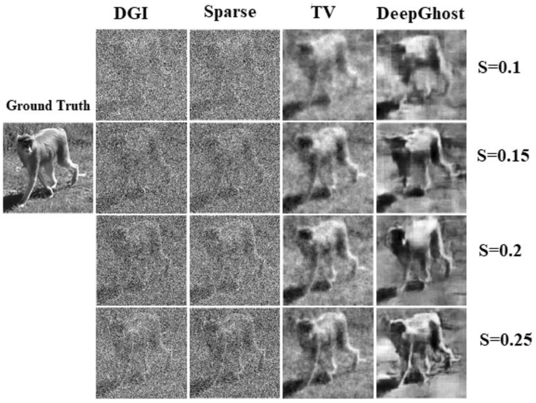

Figure 3. Qualitative comparison of reconstructions from different algorithms.

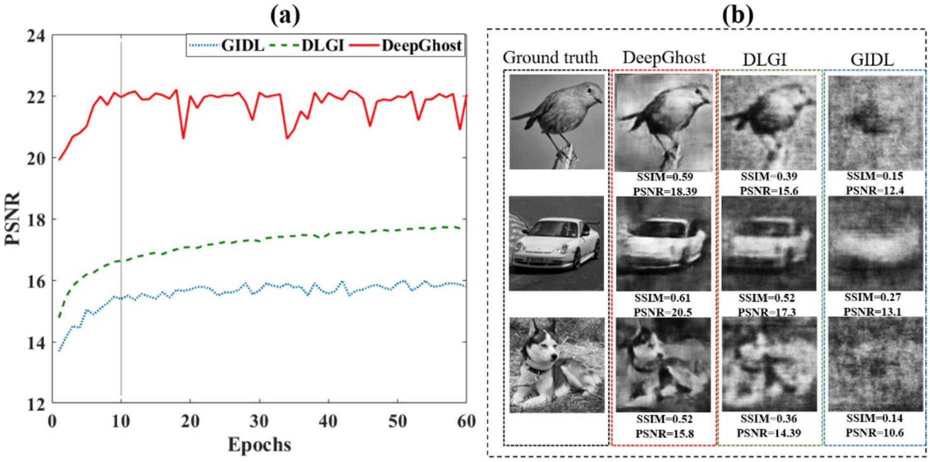

Figure 4. Performance comparison (for GIDL, DLGI, and DeepGhost) (**a**) on test set during training, (**b**) qualitative and quantitative comparison of reconstructions.

Ghost model is first trained on STL-10 data set (10,000 images), and then evaluated over a validation dataset (1,000 images) which is not seen during training. The same validation dataset is used as target images for DGI and CS based methods. In this paper, the sampling ratio '***S***' is defined as the ratio between *Number of measurements* to *Image size in pixels*. For quantitative comparison, peak signal-to-noise ratio (PSNR) and Structural SIMilarity (SSIM)[27] metrics are used.

Results and analysis. For qualitative comparison, an image from the "*monkey*" class of validation dataset is chosen. We evaluate the reconstruction results of DGI, Sparse, total variation (TV), and DeepGhost algorithms (see details in "Methods" ****section) for sampling ratios ranging from 0.1 to 0.25. We use Sparse and TV algorithms which are well-known high performance algorithms for specifically comparing the reconstruction quality. By visual inspection, it can be seen from Fig. 3 that the reconstruction results for TV and DeepGhost are almost identical. For a low sampling ratio of 15%, we get a reasonable target reconstruction for complex scene using DeepGhost. However, to achieve better results on overall dataset and diverse scenes, we resort to ***S*** = 0.2–0.25 for practical imaging. At such low sampling rates, both *DGI* and Sparse (*DCT based*) algorithms fail to reconstruct a clear target.

Comparison with deep learning algorithms. Furthermore, we design an experiment to validate the superior performance of our deep learning network by comparing it with two existing deep learning networks used for CGI under similar settings. Specifically, we train the models of[16] (GIDL) and[17] (DLGI) along with DeepGhost on STL-10 dataset at a low sampling ratio of 0.2. For all three networks, we use similar network parameters (weights, strides, initializations, activations, learning rate etc.).

Results and analysis. The PSNR over the test set (1,000 images) is computed during training and plotted against training epochs, shown in Fig. 4a. The PSNR for the reconstructed image is calculated with respect to its

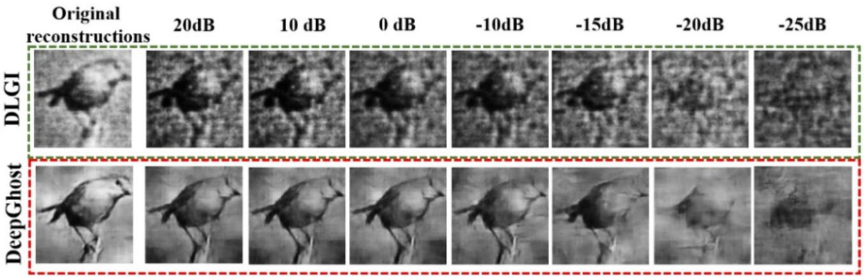

Figure 5. Qualitative comparison of DeepGhost with DLGI for noise robustness (at different noise levels, $\boldsymbol{S} = 0.2$).

ground truth counterpart. It can be seen from Fig. 4a that it is very challenging for the GIDL network to recover image details from an under sampled image, achieving low PSNR values throughout its training. This is easy to understand because fully-connected neural networks are not ideal for image analysis. Although they can perform well on simple (e.g., digits) dataset, it is difficult for them to achieve satisfactory performance on complex images. Moreover, the training time for the GIDL network is very long compared to DeepGhost due to its fully connected structure. Compared to GIDL, the DLGI employs a better network based on convolutional layers. However, from Fig. 4a, it can be seen that DeepGhost also outperforms DLGI in terms of image reconstruction quality with high PSNR values achieved within a few epochs.

It is important to highlight that the training convergence for DeepGhost is faster compared to both DLGI and GIDL networks. This points toward the fact that simply using deep networks for image reconstruction may not lead to a satisfactory performance. Since DeepGhost uses skip connections along with deep architecture, it can achieve better results with fast convergence. Keeping in view the long convergence times of other models compared to DeepGhost, we carry out comparison testing at a high learning rate ($lr = 0.001$). It can be seen from Fig. 4a that DeepGhost has a chirpy PSNR response after ~ 10 epochs. This is because our network converges faster at a high learning rate compared to DLGI and GIDL networks and then goes into overfitting mode. Therefore, we choose a lower learning rate ($lr = 0.0001$) for DeepGhost training. To further investigate performance differences between these networks, a qualitative comparison is presented in Fig. 4b.

From Fig. 4b, it can be seen that the GIDL network fails to reconstruct complex targets because of its fully connected architecture. Therefore, this kind of network is not suitable for dynamic CGI. Similarly, the DLGI network, by using shallow convolutional structure, roughly estimates the target, failing to provide a clear reconstruction. In contrast, DeepGhost provides much better reconstructions for complex diverse targets. This superior performance of DeepGhost can be attributed to its denoising autoencoder structure with skip connections, which achieves deep architecture with low computational time. The inclination towards using simple architecture, shallow network (to reduce computational time), and validating model on limited data results in poor performance of DLGI and GIDL.

For evaluating noise robustness, the performance of DeepGhost is compared with DLGI (which gives slightly better reconstruction than GIDL). In this experiment, the detection fluctuations are simulated by adding noise (using *awgn()* function in Matlab) to measurement data (intensity values), resulting in different SNRs. The reconstruction results for the 'bird' image at $\boldsymbol{S} = 0.2$ are shown in Fig. 5. From qualitative comparison in Fig. 5, it can be seen that the DLGI network fails to combat noise with poor reconstruction quality at different SNRs. This indicates that the convolutional layers (of DLGI) with no mechanism to suppress noise fail to recover a clean target. On the other hand, the DeepGhost network based on denoising autoencoder architecture, learns to suppress noise using compressing/decompressing stages, recovering clean targets at different SNRs. This noise suppression is further aided by skip connections, which provide high frequency information across different layers, to recover fine details which are lost during noise suppression. From overall comparison, it can be concluded that the DeepGhost model is more suitable for practical CGI compared to existing networks. The reconstruction results for DeepGhost at different sampling ratios are shown in Fig. 6.

Physical experiments. The experimental arrangement of CGI setup is shown in Fig. 7. A series of random binary patterns is projected using a custom-made projection system. Light from the source LED is modulated by a TI DLP6500 DMD. A projection lens with focusing dial is used to project sharp patterns on the target plane. Target scenes are printed on an A4-sized white paper (using a regular printer). The target is placed at a distance of 500 mm from the plane of projection and detection. Light back-reflected from the scene is collimated on the photodetector (Thorlabs; 21 mm^2 active area) by a 5 mm imaging lens. Intensity measurements captured by the photodetector are digitized by a 16-bit data acquisition (DAQ) card (Sampling at 2 MS/s). A customized software is used to project patterns and acquire intensity values (using a synchronous trigger) for computation. The rudimentary image reconstructed by the software is passed down to DeepGhost for clean undersampled reconstruction. The data collection and preparation (of experimental and synthetic data) for training takes a week.

Experiment-1 results. In the first experiment, we directly apply the DeepGhost model trained on simulation dataset to reconstruct target images acquired from random image datasets (airplane and dog image[28], standard mandrill test image, and our university logo). It is observed that the application of simulation-trained model

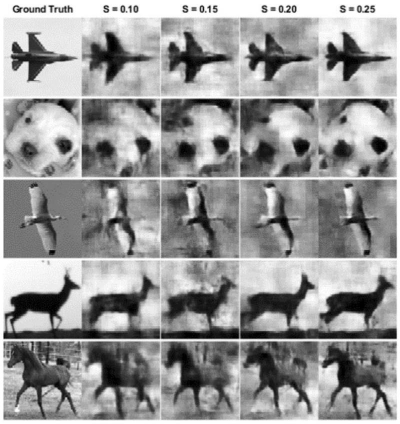

Figure 6. Simulation based image reconstruction using DeepGhost for different sampling ratios.

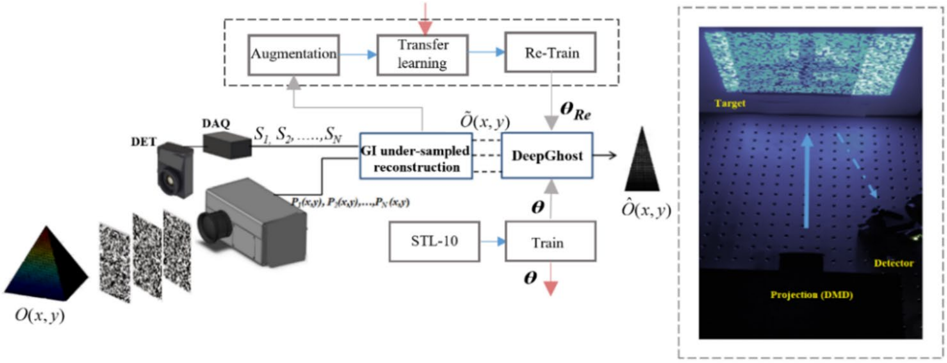

Figure 7. DeepGhost experimental setup.

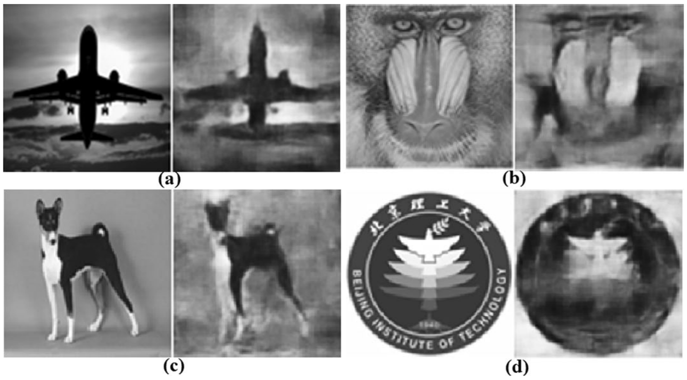

Figure 8. Reconstructions by simulation-trained model on diverse images at $\boldsymbol{S=0.4}$. (**a**) SSIM = 0.5521, PSNR = 17.20 dB, (**b**) SSIM = 0.4812, PSNR = 13.22 dB, (**c**) SSIM = 0.6014, PSNR = 19.91 dB, (**d**) SSIM = 0.4613, PSNR = 14.56 dB.

under physical conditions (e.g., noise, target reflectivity) demands undersampled input to be reconstructed at $\boldsymbol{S} = 0.4$. Therefore, we capture input images at 40% sampling rate with respect to clear target reconstruction through our CGI (DGI) setup in this case. Figure 8(a,c: good case, b,d: worst case) shows the reconstructed images with corresponding PSNR and SSIM values. From Fig. 8, it can be seen that the network is able to reconstruct random images from different classes. However, the network is unable to correctly reconstruct all random targets with clarity because of limited data training and knowledge of physical imaging environment. In fact, it

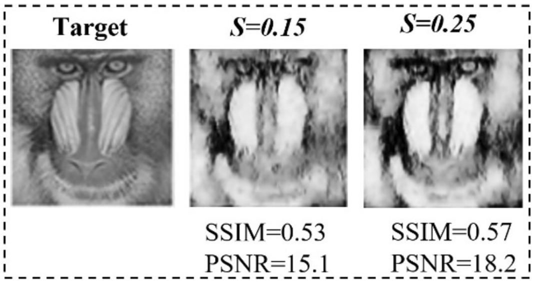

Figure 9. Results for experiment-2 on 'mandrill' test set image (as unseen target).

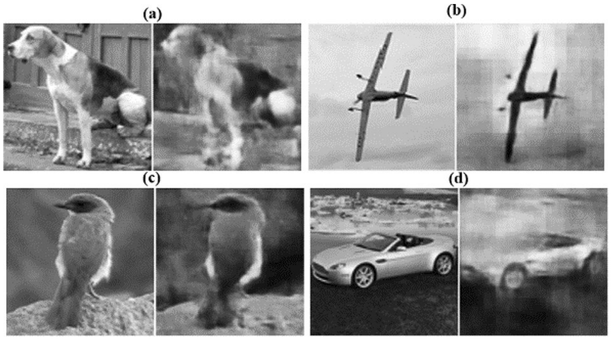

Figure 10. Validation set image reconstruction. (**a**) SSIM = 0.5214, PSNR = 17.62 dB, (**b**) SSIM = 0.6518, PSNR = 18.77 dB, (**c**) SSIM = 0.6913, PSNR = 18.79 dB, (**d**) SSIM = 0.4645, PSNR = 15.12 dB.

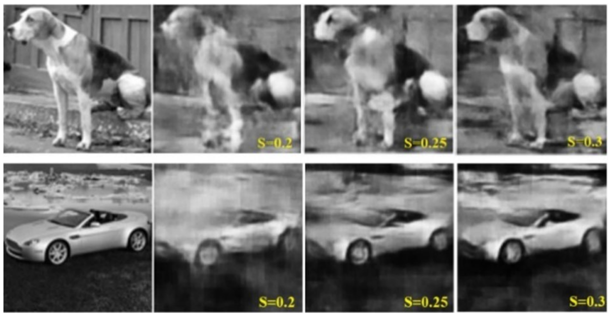

Figure 11. Improving image quality by increasing sampling ratio (***S*** = 0.2, 0.25, and 0.3).

is very challenging to optimize a DL model for CGI directly through simulation data for reconstructing diverse random scenes. To counter this problem, we apply augmentation and transfer-learning in our experiments.

Experiment-2 results. In the second experiment, the proposed network is trained on undersampled images acquired from the CGI setup (through DGI for different targets), with ground truth counterparts set as training output. To increase limited data acquired from physical setup, we apply data-augmentation technique (using Keras's DataGenerator module; by applying translation, rotation, and adding noise in the images). Even though, the data can be increased through augmentation, it is still prone to overfitting. Therefore, we further use transfer-learning to make the network highly-scalable. Transfer-learning is used to provide prior-knowledge from the large dataset (obtained during training) to the smaller augmented dataset to perfect imaging under physical conditions. The results for 'mandrill' test image are presented in Fig. 9. It can be seen that the results from experiment-2 (Fig. 9) are very clear compared to the result (Fig. 8b) from simulation based model. The results on validation dataset are understandably consistent, shown in Fig. 10. Overall, it is observed that simple targets with plain background are easily reconstructed at ***S*** = 0.2.

However, for some complex targets (e.g., Fig. 10a,d), better image quality is achieved at a slightly higher sampling ratio (Fig. 11). This is due to (1) practical system noise that can blur reconstructed images by corrupting feature extraction and/or (2) complex image features of random unseen images. The overall results indicate that the reconstruction quality with 20% sampling rate using binary random patterns based CGI is very promising.

METHOD	# of projections	I_A (s)	I_R(s)	I_T (s)	*fps*
Sparse	–	–	32.61	–	–
TV	–	–	18.63	–	–
DEEPGHOST 15%	1,382	0.069	0.13	0.199	5
DEEPGHOST 20%	1843	0.092	0.13	0.222	4
DEEPGHOST 25%	2,304	0.115	0.13	0.245	4

Table 1. Time breakdown for practical imaging.

Although the network can produce better quality reconstructions at higher sampling ratios, it can further be trained on more data to achieve high-quality and reliability at lower sampling rates.

Imaging time.

To quantify imaging time, different values of time for the DeepGhost model are presented in Table 1. The imaging time is based on reconstructing 96 × 96 images at ~ 20 kHz modulation rate. The total imaging time (I_T) is equal to data acquisition time (I_{AQ}) + reconstruction time (I_R). The reconstruction time (I_R) is the combined time of DGI (undersampled reconstruction) + DCAN processing. The reconstruction time remains the same for different sampling ratios, which is an attractive feature of DL based model. It can be seen from Table 1 that DeepGhost can achieve real-time frame rates (fps) compared to conventional methods with high reconstruction overhead only.

Methods

Principles and methods of CGI.

In computational ghost imaging, a target scene $O(x, y)$ is reconstructed by correlating a series of modulation patterns $P_i(x, y)$ with intensity measurements S_i at the bucket detector. The target scene can be reconstructed by[29]:

$$O(x,y) = \left\langle (S_i - \langle S_i\rangle)\left(P_i(x,y) - \left\langle P_i(x,y)\right\rangle\right)\right\rangle \tag{1}$$

where S_i is the ith measurement, $P_i$is the ith modulation pattern, and the ensemble average for N iterations is given by: $\langle t_i\rangle = \frac{1}{N}\sum_{i=1}^{N} t_i$. To reconstruct high quality image, a large number of measurements are required.

To improve the performance of correlation based GI, DGI has been proposed[26]. Figure 3 shows images reconstructed using DGI defined by Eq. (2), where, R_i is the reference signal. It is evident that even with these methods, GI still requires a large number of measurements (long imaging time) to produce quality image.

$$O(x,y) = \left\langle P_i(x,y)S_i\right\rangle - \frac{\langle S_i\rangle}{\langle R_i\rangle}\left\langle R_iP_i(x,y)\right\rangle \tag{2}$$

To reduce reconstruction time for CGI, compressive sensing methods have been applied to ghost imaging[11,30,31]. The CS theory allows an object (target scene) $O(x, y)$ to be reconstructed from a set of undersampled measurements S, assuming that object is sparse within a fixed basis. For evaluation, we process our GI data with two commonly used priors for natural images: the sparse prior and the total variation (TV) regularization prior. The sparse representation prior[32] considers natural image to be represented by an orthogonal basis (discrete cosine transform) transform matrix **D** and coefficient vector **c**. The reconstruction for CGI is achieved by minimizing the following function:

$$\min_{O}\left\{\mathrm{f} = \|c\|_{l_1} + \frac{\mu_1}{2}\left\|DO - c + \frac{y_1}{\mu_1}\right\|_{l_2}^2 + \frac{\mu_2}{2}\left\|PO - S + \frac{y_2}{\mu_2}\right\|_{l_2}^2\right\} \tag{3}$$

where y is the Lagrange multiplier and μ is the balancing parameter. The above $l1$-minimization problem can be solved by using augmented lagrange multiplier (ALM) method[33]. The TV regularization prior is related to the gradient of an image. If **G** is the gradient matrix of an image, the TV regularization prior based reconstruction is given by solving the following minimization:

$$\min_{O}\left\{\mathrm{f} = \|c\|_{l_1} + \frac{\mu_1}{2}\left\|GO - c + \frac{y_1}{\mu_1}\right\|_{l_2}^2 + \frac{\mu_2}{2}\left\|PO - S + \frac{y_2}{\mu_2}\right\|_{l_2}^2\right\} \tag{4}$$

DeepGhost.

The proposed deep convolutional autoencoder architecture is shown in Fig. 1. The network employs convolutional layers with trainable filters for extracting features and filtering corruptions from the image. The encoding stages use 32, 64, and 128 (Conv2D) filters for scaling down the data. The compressed data is grouped at an “intermediate” layer with 256 conv-filters. The decoding stages use 128, 64, and 32 filters for reconstructing the encoded image. The output is reconstructed using a single conv-filter at the end. To visualize data processing at each layer, the feature maps for an unseen target (pepper test image) through the network pipeline are shown in Fig. 12. To prevent network operation in saturated or dead regions of activation, the network is initialized with Xavier initialization[34]. After every convolutional layer, batch normalization layer[35] is used to achieve training efficiency. The data along the pipeline is scaled into different dimensions using max-pooling

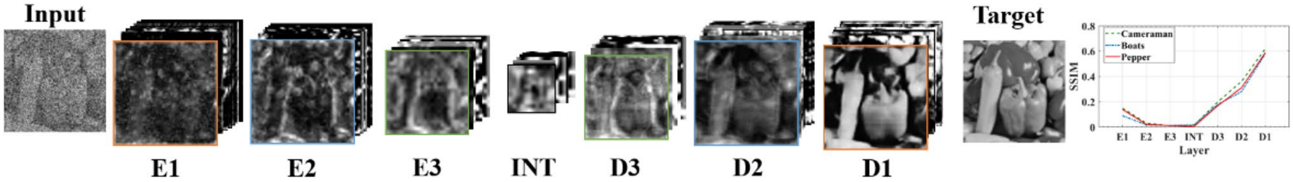

Figure 12. End-to-end Visualization of activation feature maps at different layers in the network. The SSIM plot for different standard test set targets is given to quantify SSIM at different layers. The SSIM increases when the decoding layers start reconstructing.

and Up-sampling operations. To counter data over-fitting, Gaussian noise layers are used to apply regularization through additive Gaussian noise in the hidden layers. The image reconstruction quality is improved by training the network with noisy data traversed via skip connections between similar scale stages. The nonlinearity between layers is created using a nonlinear activation (ReLU).

In general, the autoencoder serves the purpose of image denoising. If O(*x, y*) is assumed to be the target, then the target imaged by CGI using undersampled measurements is a corrupted version of the target $g\left(O\left(x, y\right)\right) + n$ added with noise, represented by $\tilde{O}\left(x, y\right)$. The inverse problem of recovering the original image from an undersampled image is solved by applying DL. Through training, the network learns an end-to-end mapping from $\tilde{O}\left(x, y\right)$ to $O\left(x, y\right)$. For the reconstructed target $\hat{O}\left(x, y\right)$, the network is trained on a set S = {*DGI* $_{\text{undersampled}}$, *Ground truth* }, to minimize the loss function expressed as:

$$\ell(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left[\hat{O}(x, y) - O(x, y)\right]^2 \tag{5}$$

The network is fed with an undersampled ghost image reconstructed from CGI data using iterative DGI algorithm (Eq. (2). For further time reduction and fast reconstruction, a compressive sensing algorithm can also be used to preprocess CGI data[17]. The network parameters are updated using Adaptive moment estimation optimization[36] with standard back propagation on mini-batch(es)$\underline{S}$. The learning rate for each layer = 10^{-4}. The proposed network is trained on gray-scaled STL-10[25] 96 × 96 images. All images are preprocessed using standard normalization procedure. The training set has 10,000 images, whereas both test and validation image sets have 1,000 images each. The network is implemented with Keras (TensorFlow support) on an Intel i7 CPU with 32 GB memory.

Conclusion

In this paper, we demonstrate a DL based imaging framework to improve the performance of random-pattern based CGI. DL can learn features from a large dataset and is more flexible compared to CS optimization techniques based on fixed priors and rigid calculations. The proposed method is capable of reconstructing good-quality 96 × 96 target with 80% compression at 4-5 Hz frame rates. Optimizing random-pattern based CGI for real-time application is very challenging because of its long reconstruction time. Even if the reconstruction time is reduced by means of undersampling, the reconstruction quality of undersampled CGI (through CS or DL) for diverse unseen targets is poor. The main objective in this paper is to reconstruct diverse unseen targets with accuracy. By importing prior knowledge from a large dataset, and training a network on physical data, this objective is achieved. The core component of our imaging framework is the DCAN. The network uses an encoding–decoding architecture combined with skip connections to reconstruct good quality image from an undersampled input. Deep learning combined with GI is a good choice in order to avoid complex methods that fail to reap the benefits of GI i.e., reduced cost and simplicity. By further training our algorithm on a larger dataset (more classes), we can enhance its feature learning ability, which would increase reconstruction reliability and quality. Experimental results show that the proposed method achieves better performance than compressive sensing and existing deep learning methods used for computational ghost imaging.

Received: 3 March 2020; Accepted: 21 May 2020
Published online: 09 July 2020

References

1. Shapiro, J. Computational ghost imaging. *Phys. Rev. A* **78**, 061802 (2008).
2. Zhang, Z., Wang, X., Zheng, G. & Zhong, J. Hadamard single-pixel imaging versus Fourier single-pixel imaging. *Opt. Express* **25**, 19619–19639 (2017).
3. Zhang, Z., Jiao, S., Yao, M., Li, X. & Zhong, J. Secured single-pixel broadcast imaging. *Opt. Express* **26**, 14578–14591 (2018).
4. Gong, W. *et al.* Three-dimensional ghost imaging lidar via sparsity constraint. *Sci Rep* **6**, 26133 (2016).
5. Satat, G., Tancik, M. & Raskar, R. Lensless imaging with compressive ultrafast sensing. *IEEE Trans. Comput. Imaging* **3**(3), 398–407 (2017).
6. Sun, M.-J. & Zhang, J.-M. Single-pixel imaging and its applications in three-dimensional reconstruction: A brief review. *Sensors* **19**(3), 732 (2019).
7. Wang, Y., Suo, J., Fan, J. & Dai, Q. Hyperspectral computational ghost imaging via temporal multiplexing. *IEEE Photon. Tech. Lett.* **28**(3), 288–291 (2016).
8. Gibson, G. *et al.* Real-time imaging of methane gas leaks using a single-pixel camera. *Opt. Express* **25**, 2998–3005 (2017).
9. Xu, Z. H., Chen, W., Penulas, J., Padgett, M. J. & Sun, M. J. 1000 fps computational ghost imaging using LED-based structured illumination. *Opt. Express* **26**, 2427–2434 (2018).

10. Salvador-Balaguer, E. *et al.* Low-cost single-pixel 3D imaging by using an LED array. *Opt. Express* **26**, 15623–15631 (2018).
11. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**(4), 1289–1306 (2006).
12. Katkovnik, V. & Astola, J. Compressive sensing computational ghost imaging. *J. Opt. Soc. Am. A* **29**, 1556–1567 (2012).
13. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**(8), 921–943 (2019).
14. Chen, Z., Shi, J. & Zeng, G. Object authentication based on compressive ghost imaging. *Appl. Opt.* **55**, 8644–8650 (2016).
15. Chen, W. & Chen, X. Object authentication in computational ghost imaging with the realizations less than 5% of nyquist limit. *Opt. Lett.* **38**, 546–548 (2013).
16. Lyu, M. *et al.* Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
17. He, Y. *et al.* Ghost imaging based on deep learning. *Sci. Rep.* **8**, 6469 (2018).
18. Higham, C. F., Murray-Smith, R., Padgett, M. J. & Edgar, M. P. Deep learning for real-time single-pixel video. *Sci. Rep.* **8**, 2369 (2018).
19. Rizvi, S., Cao, J., Zhang, K. & Hao, Q. Deringing and denoising in extremely under-sampled Fourier single pixel imaging. *Opt. Express* **28**, 7360–7374 (2020).
20. Rizvi, S., Cao, J., Zhang, K. & Hao, Q. Improving imaging quality of real-time Fourier single-pixel imaging via deep learning. *Sensors* **19**, 4190 (2019).
21. Vincent, P., Larochelle, H., Bengio, Y. & Manzagol, P. A. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th International Conference on Machine Learning* (ACM 2008), pp. 1096–1103.
22. Mousavi, A. & Baraniuk, R. G. Learning to invert: Signal recovery via deep convolutional networks. In *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE 2017*)*, pp. 2272–2276.
23. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**(11), 2278–2324 (1998).
24. Bian, L., Suo, J., Dai, Q. & Chen, F. Experimental comparison of single-pixel imaging algorithms. *J. Opt. Soc. Am. A* **35**, 78–87 (2018).
25. Coates, A., Lee, H. & Ng, A. Y. An analysis of single layer networks in unsupervised feature learning. *AISTATS* **20**, 20 (2011).
26. Ferri, F., Magatti, D., Lugiato, L. & Gatti, A. Differential ghost imaging. *Phys. Rev. Lett.* **104**, 253603 (2010).
27. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **3**, 600–612 (2004).
28. Khosla, A. Jayadevaprakash, N., Yao, B. & Fei-Fei, L. Novel dataset for fine-grained image categorization. In *First Workshop on Fine-Grained Visual Categorization (FGVC), IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011).
29. Bromberg, Y., Katz, O. & Silberberg, Y. Ghost imaging with a single detector. *Phys. Rev. A* **79**, 053840 (2009).
30. Katz, O., Bromberg, Y. & Silberberg, Y. Compressive ghost imaging. *Appl. Phys. Lett.* **95**, 131110 (2009).
31. Candès, E. J. & Wakin, M. B. An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008).
32. Duarte, M. F. *et al.* Single-pixel imaging via compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 83–91 (2008).
33. Lin, Z., Chen, M., Wu, L. & Ma, Y. The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices. UIUC Technical Report UILU-ENG-09-2215 (2009).
34. Glorot, X. & Bengio, Y. *Understanding the Difficulty of Training Deep Feedforward Neural Networks*. In *AISTATS* (2010).
35. Ioffe, S. & Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *Proceedings of International Conference on Machine Learning* (2015), pp. 448–456.
36. Kingma, D. & Ba, J. *A Method for Stochastic Optimization* (ICLR, Adam, 2015).

Acknowledgements

This research is supported by National Natural Science Foundation of China (NSFC) (61875012, 61871031), and Natural Science Foundation of Beijing Municipality (4182058). The authors appreciate valuable suggestions by F. Zia.

Author contributions

S.R. and Z.K. conceived the system. S.R. and C.J. proposed the use of deep networks to solve the inverse problem. S.R. developed the deep learning algorithm. C.J. and Q.H. supervised the work. S.R. wrote the manuscript and C.J. reviewed it.

Competing interests

The author declares no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.C. or Q.H.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.