

Single-pixel imaging using physics enhanced deep learning

FEI WANG,^{1,2} CHENGLONG WANG,^{1,2}  CHENJIN DENG,^{1,2} SHENSHENG HAN,^{1,2,3} AND GUOHAI SITU^{1,2,3,*} 

¹Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China

²Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China

³Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China

*Corresponding author: ghsitu@siom.ac.cn

Received 16 August 2021; revised 3 November 2021; accepted 3 November 2021; posted 4 November 2021 (Doc. ID 440123); published 13 December 2021

Single-pixel imaging (SPI) is a typical computational imaging modality that allows two- and three-dimensional image reconstruction from a one-dimensional bucket signal acquired under structured illumination. It is in particular of interest for imaging under low light conditions and in spectral regions where good cameras are unavailable. However, the resolution of the reconstructed image in SPI is strongly dependent on the number of measurements in the temporal domain. Data-driven deep learning has been proposed for high-quality image reconstruction from a undersampled bucket signal. But the generalization issue prohibits its practical application. Here we propose a physics-enhanced deep learning approach for SPI. By blending a physics-informed layer and a model-driven fine-tuning process, we show that the proposed approach is generalizable for image reconstruction. We implement the proposed method in an in-house SPI system and an outdoor single-pixel LiDAR system, and demonstrate that it outperforms some other widespread SPI algorithms in terms of both robustness and fidelity. The proposed method establishes a bridge between data-driven and model-driven algorithms, allowing one to impose both data and physics priors for inverse problem solvers in computational imaging, ranging from remote sensing to microscopy. © 2021 Chinese Laser Press

<https://doi.org/10.1364/PRJ.440123>

1. INTRODUCTION

Single-pixel imaging (SPI) is an emerging computational imaging modality that utilizes the second-order correlation of quantum or classical light to reconstruct a two-dimensional (2D) image from a one-dimensional (1D) bucket signal [1–4]. As most of the photons that interact with the object are collected by the bucket detector, SPI has significant advantages in terms of the detection sensitivity, dark counts, and spectral range. Thus it has received increasing attentions over the past decade for the people working in the divergent fields of remote sensing [5,6], 3D imaging [7,8], spectral imaging [9,10], microscopy [11], and of the sort [3,12]. However, in SPI, each single-pixel measurement contains highly compressed information about the object, and one needs a large amount of such measurements to reconstruct an image with good resolution. This leads to a trade-off between the acquisition time and the image quality that hinders the practical application of SPI. Many studies have been carried out to address this issue. The solutions proposed so far can be categorized into two mainly aspects of strategies. The first one is to design the encoding patterns that ensures each single-pixel measurement contains as more information as possible [13–15]. The second one is to develop an optimization

algorithm to obtain better reconstruction using a smaller number of measurements [16,17].

Owing to its capability of solving various challenging problems in divergent fields [18,19], deep learning (DL) has also been adopted for SPI recently. Previous studies have shown that the DL-based SPI methods can dramatically reduce the sampling ratio, promising real-time performance [17,20,21]. Specifically, Lyu *et al.* [17] proposed a physics-informed deep learning method called ghost imaging (GI) using deep learning, in which the input of the deep neural network (DNN) is an approximant recovered using the conventional correlation algorithm. This method allows the reduction of the sampling ratio. However, as GIDL used speckle patterns to encode the object information, the modulation efficiency is not very high. Higham *et al.* [20] proposed a deep convolutional autoencoder network (DCAN) for this task, in which the trained binary weights in the encoding part of DCAN are used to scan the target. This allows an efficient encoding-decoding strategy for SPI. However, DCAN is a pure data-driven method, which suffers from common issues such as generalizability and interpretability [22]. Although our previous works [21,23] have shown that end-to-end DNN can be used to recover the object directly from the detected bucket signal without any physical

priors, recent studies have shown that blending the physics of the imaging system into DNN has the advantages in terms of data acquisition [21,24], generalization [25,26], and interpretability [27].

In this work, we report a physics enhanced deep learning technique for SPI. The physics prior we exploit mainly contains two aspects that rely on the forward propagation model of the SPI system, H , i.e., $I = Hx$. First, in contrast with end-to-end learning algorithms [21,23], the bucket signal I is used to estimate x_p with the knowledge of H ; the resulting x_p is used as the input of the DNN \mathcal{R}_θ . This allows us to optimize the encoding patterns and add an interpretable physics decoding layer before the DNN. Second, the difference between the acquired bucket signal I and estimated one $\hat{I} = H\mathcal{R}_\theta(x_p)$ is used to finely tune the weights θ of the DNN model. This allows us to correct the distortion of the DNN predictions due to insufficient generalization of the model. Numerical simulations and experiments are performed to demonstrate that the proposed strategy brings advantages in terms of both robustness and fidelity.

2. METHODS

As schematically presented in Fig. 1, the proposed method consists of two main steps: a physics-informed autoencoder DNN that generates a set of optimal encoding patterns H^* , and a model-driven fine-tuning process that enhances the reconstructed image.

As shown in Fig. 1(a), the physics-informed autoencoder DNN contains three parts. The first part is a set of M patterns $H_m(u, v)$ that are used to encode an object $x(u, v)$ to a 1D bucket signal $I_m = H_m(u, v)x(u, v)$ with the length of M . The second part is to reconstruct a rough estimation of the object x_p by using any conventional GI algorithm from I and H . In this study, we employ differential ghost imaging (DGI) [29,30] for this job:

$$x_p = \text{DGI}(H, I) = \langle H_m I_m \rangle - \frac{\langle H_m \rangle}{\langle S_m \rangle} \langle S_m I_m \rangle, \quad (1)$$

where $\langle \cdot \rangle$ denotes the ensemble average approximately defined as $\langle H_m \rangle = \frac{1}{M} \sum_{m=1}^M H_m$, $\langle I_m \rangle = \frac{1}{M} \sum_{m=1}^M I_m$, and $S_m = \sum_{u,v} H_m(u, v)$ is used to normalize the illumination patterns so as to improve the robustness. To proceed, we need to define the sampling ratio $\beta = M/N$, where N represents the number of pixels that represent x . The DGI algorithm described in Eq. (1) is a noniterative one and is thus fast and robust to execute [31]. In this way, one can physically map the features in the measurement space to the estimated image space, which provides an interpretable feature extraction layer without complicated calculations. The third part is a DNN model \mathcal{R}_θ that is used to perform further image enhancement. It takes x_p as its input and produces a high-quality estimation $\mathcal{R}_\theta(x_p)$ at the output layer.

Apparently, both the DNN model \mathcal{R}_θ and the encoding patterns H should be trained, for example, on a set of training data $\mathcal{S}_T = \{x^k | k = 1, 2, \dots, K\}$. With random initialization, the patterns H and the weights parameters θ in the DNN model \mathcal{R}_θ can be optimized by solving

$$\{\mathcal{R}_{\theta^*}, H^*\} = \arg \min_{\theta \in \Theta, H \in \mathcal{H}} \|\mathcal{R}_\theta(x_p^k) - x^k\|^2, \quad \forall x^k \in \mathcal{S}_T, \quad (2)$$

where $x_p^k = \text{DGI}(H, Hx^k)$. One can see that the most significant difference between the proposed framework and DCAN [20] is that a physics informed layer (i.e., DGI [29,30]) is blended into the DNN model. This generates a set of optimized patterns H^* that can be used to encode a real-world object to be imaged.

Encoding a real-world target by using a typical SPI system shown in Fig. 1(b), one can acquire a 1D raw bucket signal I . This is the input of the second component of the proposed method, a model-driven fine-tuning process, which essentially

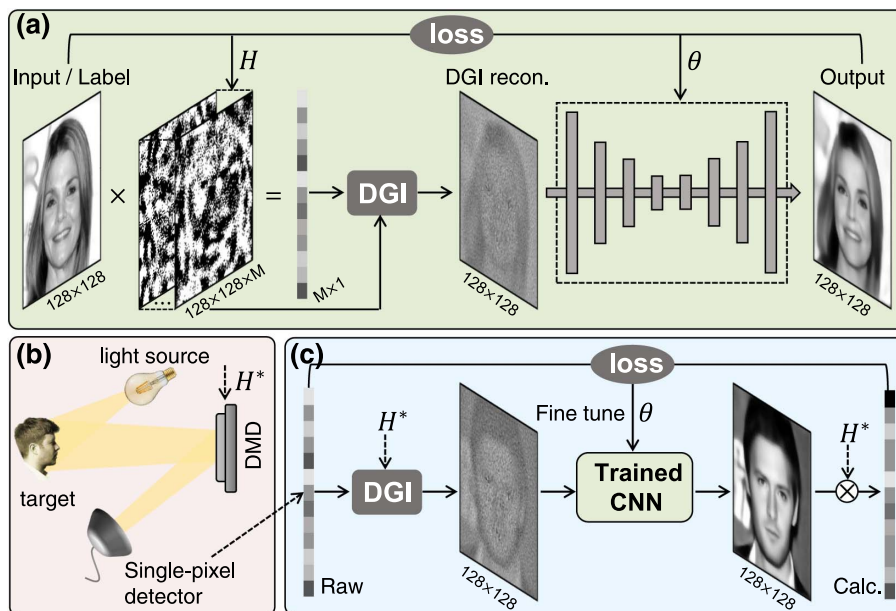


Fig. 1. Schematic diagram of the physics enhanced deep learning approach for SPI. (a) The physics-informed DNN. (b) The SPI system. (c) The model-driven fine-tuning process. The face images were taken from CelebAMask-HQ [28].

consists of the DGI model, H^* , and the trained DNN \mathcal{R}_{θ^*} . As both H^* and \mathcal{R}_{θ^*} have been trained in the first step, one expects to have a good reconstructed image of the target [19]. However, since the network model \mathcal{R}_{θ^*} is trained on a data set, it has a strong bias to reconstruct an image that is statistically similar to those in the training set [22]. We thus hypothesize that one can get further image enhancement by fitting the measurements as in conventional model-driven optimization methods [32]. This can be formulated as

$$\mathcal{R}_{\theta^{**}} = \arg \min_{\theta^* \in \Theta} \|H^* \mathcal{R}_{\theta^*}(x_p) - I\|^2, \quad (3)$$

where $x_p = \text{DGI}(H^*, I)$. By fine tuning the weights in the first three layers in the pre-trained network \mathcal{R}_{θ^*} , the optimization converges quickly along with the drop of error between the acquired I and the estimated one \hat{I} . This yields a fine-tuned reconstructed image $\mathcal{R}_{\theta^{**}}(x_p)$. Compared with traditional transfer learning [33], the proposed fine-tuning strategy does not need to use any labeled training data. All it needs to input is the raw bucket signal I , from which one expects to reconstruct an image of the target of interest. We will show that this target does not have to be similar to those in the training set \mathcal{S}_T that is used to train \mathcal{R}_{θ^*} . In the case that H^* cannot be obtained precisely; one can also include parameters that represent model uncertainty into the objective function Eq. (3) as trainable weights as in Ref. [34]. Here we simply use the ideal SPI model.

The network architecture we used to implement \mathcal{R}_{θ} is illustrated in Fig. 2. It actually has a U-net-like structure that contains five downsampling layers and five upsampling layers. In order to adapt to data/images of different lengths/sizes, one does not need to change the network hyperparameters but the size of the feature maps. We would also like to emphasize that there is no limitation to choose the neural network architecture for the proposed physics-enhanced deep learning framework, although properly adjusting of the network architecture may get better results. In this work, we simply employ the one shown in Fig. 2. In the implementation of the neural networks, we used the following parameter setting: the learning rate was 0.0002, and the momentum and epsilon parameters in the batch normalization were 0.99 and 0.001. The leaky ReLU

with the leak parameter of 0.2 was used as the activation function. The training set for \mathcal{R}_{θ^*} was formed by using 29,000 128×128 -pixel images from CelebAMask-HQ [28]. The training was conducted in a computer with an Intel Xeon E5-2696 V3 CPU, 64 GB RAM, and an NVIDIA Quadro P6000 GPU. It converged within 64 epoches.

3. RESULTS AND DISCUSSION

Here we perform a comparative study on the effectiveness of the proposed method. For the sake of quantitative evaluation, we first examine its performance by using simulation data. Then we demonstrate its practical applications in laboratory and outdoor experiments.

A. Simulations

First let us examine the effectiveness of the physics-informed layer that we add to the DNN. The results are plotted at the fifth column in Fig. 3(a). It is clearly seen that the DGI reconstructed image with the learned patterns is far better than the one with random illumination. This conclusion retains even if Gaussian white noise (with the variance of δ) is added to the bucket signal. We use the signal-to-noise ratio ($\text{SNR} = 10 \log_{10}[(I - \hat{I})^2 / \delta]$) of the bucket signal to measure the noise level. From the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) of the reconstructed images are plotted in Figs. 3(b) and 3(c) (the solid gray curves in contrast to the dashed ones), respectively, and one can confirm that the learned patterns are more effective.

One can also see that the two gray curves are fairly flat with respect to noise level. This suggests that the DGI reconstruction algorithm is immune to the additive noise [29,30], no matter whether the physics-informed layer is used or not. This robustness is important for the downstream decoding DNN, as it takes the DGI reconstructed image as its input.

Now we proceed to compare the performance of the proposed methods with some of the widespread SPI methods, namely, DCAN [20], the reordering Hadamard SPI (HSI) [14], the compressed sensing based total variation (TV) regularization [31], and the Fourier domain regularized inversion (FDRI) methods [35,36]. The results are plotted in Fig. 3.

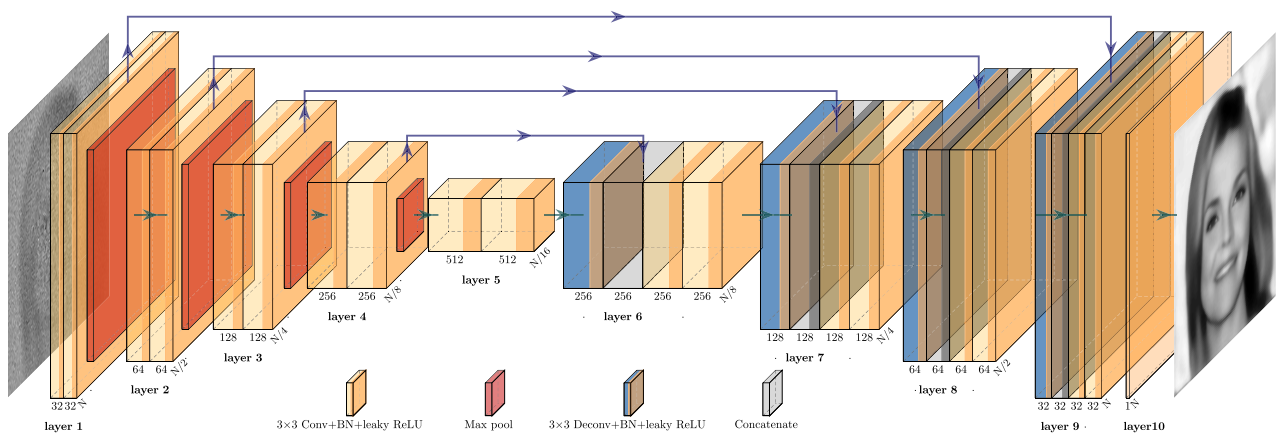


Fig. 2. Diagram of the DNN structure we designed. It consists of an encoder path that takes the low-quality image reconstructed by DGI as its input and a decoder path that outputs an enhanced one.

As a learning-based end-to-end SPI method, DCAN [20] outperforms the other existing methods (i.e., HSI, TV, and FDRI) except for some high-noise-level cases. The proposed physics-informed method has a similar performance to DCAN when the SNR of the bucket signal is 20 dB and higher, but is much better when the noise level increases. As the DNN parts of the proposed physics-informed method and DCAN do not have much difference, it must be the physics-informed layer described by Eq. (1) that contributes to the high performance [see, for example, the reconstructed images at row 3, columns 4 and 6, in Fig. 3(a)]. The reconstruction is quite time efficient. It takes only 0.32 s to reconstruct a 128×128 -pixel image using the trained model. This includes the time for DGI and the physics-informed DNN inferring. Note that the proposed algorithm was implemented in Python. It can be sped up by, for example, implementing it with a more efficient programming language like C/C++. This suggests that the proposed method has potential for SPI in real time, in addition to its robustness against noise. However, we found that the image reconstructed by the physics-informed DNN still has noticeable artefacts and thus proceeded to further enhance it by the second step, the model-driven fine-tuning process.

The results shown in Fig. 3 suggest that fine-tuning the trained DNN model \mathcal{R}_{θ^*} helps enhance the image quality when the SNR is high but is trivial otherwise. This is reasonable as the fine-tuning may also fit the noise. To see how it happens, one can examine the behavior of the objective function defined in Eq. (3). Clearly, the error between the noisy bucket signal I_{noise} and the estimated one $\hat{I} = H^*x_i$, where the subscript i denotes the iteration step, does decrease as the iteration proceeds, no matter what the noise level is [Fig. 4(a)]. However, we observe an interesting turnover phenomenon: the error function between the estimated image x_i and the ground truth x jumps steeply at the beginning and then turns over to increase as the iteration proceeds. This enforces x_i to gradually step away from x . The better H^*x_i is fit to I_{noise} , the larger the error $\|x_i - x\|$. We observe that the turnover occurs sooner when the SNR of the acquired bucket signal is low [indicated by the arrow in Fig. 4(b)], and vice versa.

It takes a lot more iterations to occur when the SNR is high. Such a turnover phenomenon is also observed in the error function between H^*x_i and the clean bucket signal I_{clean} as shown in Fig. 4(c). The main reason why the turnover phenomenon happens is that a properly designed DNN inherently regularizes the objective function because of the deep image prior [37]. That is, there is a competition for a DNN to fit the data between natural image-related content and noise (if it exists). Apparently, natural image-related contents have the priority at the beginning, but eventually the noise wins. So one can set up a trick like early stopping to obtain a better reconstructed image, in particular when the bucket signal SNR is low. More details on this matter can be found in Visualization 1.

B. Experiments

Now we proceed to demonstrate the proposed method with in-house experiments. We built a typical passive modulated SPI system as the one schematically shown in Fig. 1(b). Three real-world objects were used in our proof-of-principle experiments. They were illuminated by a thermal light source and imaged by an imaging optic with the focal length of 85 mm to a digital micromirror device (DMD, DLP7000, TI). On the DMD, the learned binary patterns H^* were sequentially displayed so as to encode the scenarios projected onto it. The encoded light was then focused on a single pixel detector (H10721, Hamamatsu) by using a $4f$ system composed of two lenses whose focal lengths are $f_1 = 75$ mm and $f_2 = 50$ mm, respectively. In all the three experiments, we acquired $M = 1024$ measurements for each object. Each pattern in H^* has a pixel number of $N = 128 \times 128$, meaning that the sampling ratio $\beta = M/N = 6.25\%$.

We reconstruct the images following the aforementioned pipeline. First we correlated H^* and the three bucket signals using the DGI algorithm according to Eq. (1). The three images reconstructed from DGI are shown in the top row of Fig. 5. Given the fact that β is as low as 6.25%, the images reconstructed by DGI alone are not too bad. But we can further improve them by feeding them into the trained physics-informed decoding network \mathcal{R}_{θ^*} . The corresponding outputs

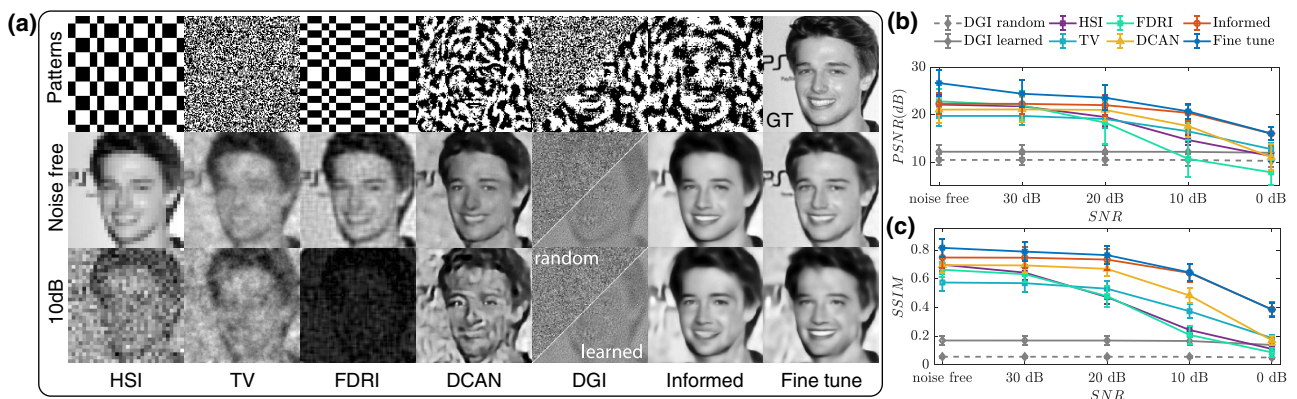


Fig. 3. Comparative study of the proposed method with some other fast SPI algorithms with a low sampling ratio ($\beta = 1024/16384 = 6.25\%$). (a) The images reconstructed by HSI [14], TV [31], FDRI [35], DCAN [20], DGI with and without learned patterns illumination, physics-informed DNN, and the fine-tuning process. (b) PSNR and (c) SSIM of the reconstructed images are used to quantitatively evaluate the performance of different methods under different SNR levels. The PSNR and SSIM metrics were averaged over 30 randomly selected images from the test set. The face image was selected from CelebAMask-HQ [28].

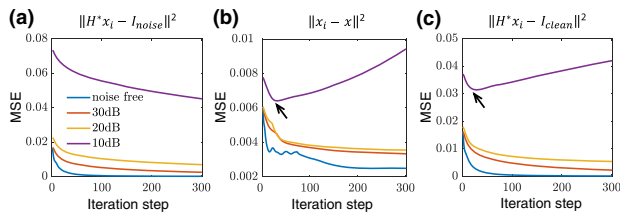


Fig. 4. Convergence behavior of different error functions that measure (a) the objective function, (b) the prediction error, and (c) the error between the estimated bucket signal and the ideal one.

of the neural network are shown in the second row of Fig. 5. One can clearly see that the noise has been significantly reduced and the contrast increased. However, as a data-driven method, the physics-informed decoding network was trained on the CelebAMask-HQ dataset [28] and thus could not recover the object images with high fidelity in our experiments. Indeed one can see that obnoxious artefacts appear in the reconstructed images. These artefacts were eliminated via the model-driven fine-tuning process according to Eq. (3) as shown in the last row of Fig. 5. This suggests that the fine-tuning process has great potential to address the generalization problem of conventional data-driven DL methods [17,20,21].

Next we endeavor to demonstrate that the proposed fine-tuning method outperforms the other widespread SPI algorithms such as DGI [29,30], HSI [14], DCAN [20]; the total variation minimization by augmented Lagrangian and alternating direction algorithms (TVAL3) [38]; and randomly initialized fine-tuning on the same set of experimental data. The data were acquired with the same SPI system we built. This time we replaced the previous objects with the badge of our institute printed on a white paper for the sake of quantitative analysis. For this purpose, we took the image reconstructed by HSI with

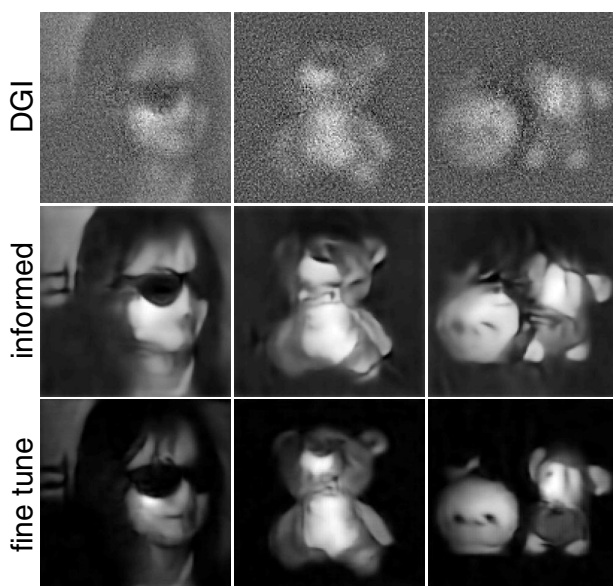


Fig. 5. Experimental results. The images reconstructed by DGI alone, DGI with physics-informed DNN, and the fine-tuning method. The sampling ratio $\beta = 6.25\%$.

full sampling ($\beta = 100\%$) as the ground truth [Fig. 6(a)] because it in principle guarantees closed form solutions [14]. However, in the comparative study, only $\beta = 6.25\%$ out of the 128×128 samples were used for image reconstruction. The images reconstructed with all these algorithms are plotted in Figs. 6(b)–6(h), respectively. Apparently, the proposed fine-tuning approach has the best performance in terms of both visual effect and quantitative metrics (PSNR and SSIM). One can find more information about the iteration process in Visualization 2.

To demonstrate the practical application of the proposed method, we incorporated the proposed method into a single-pixel LiDAR system upgraded from the one we built previously [5]. As schematically shown in Fig. 7(a), the upgrade was mainly done by replacing the active modulation module based on a rotating ground glass in Ref. [5] by a DMD-based passive one. The light source is a solid-state pulsed laser with the center wavelength of 532 nm and the pulse width of 8 ns at the repetition rate of 10 kHz. The laser light was first collimated and expanded, and then it was sent out to illuminate a remote target. The echo light scattered back from the target was collected by an imaging optic ($f = 313$ mm) with the angular field of view (FOV) of 1.5° and projected onto the DMD, on which it was encoded by the learned patterns H^* . Finally, the encoded light was focused to a photomultiplier tube (PMT, H10721-01, Hamamatsu). The PMT provides a time-resolved signal that can be used to calculate each depth slice of a 3D object. The single-pixel LiDAR experiment was performed in an outdoor environment. As shown in Fig. 7(b), the object to be imaged was a TV tower located at about 570 m away from the LiDAR system. It is practically reasonable to assume that different depth slices of the object do not have spatial overlap, and the reflectivity is real and non-negative.

To obtain a more general model for the remote sensing task, we retrained the same decoding DNN on a training set composed of 90,000 images (64×64 pixels in size) taken out of the STL10 dataset [39]. In this DNN, the size of the feature maps should be changed in accordance to the image size. Thus, the pattern H^* generated by the DNN to encode the echoed light has the dimension of $64 \times 64 \times 1024$.

For each measurement, the PMT was triggered with a time delay of 3700 ns with respect to that of the laser emission so

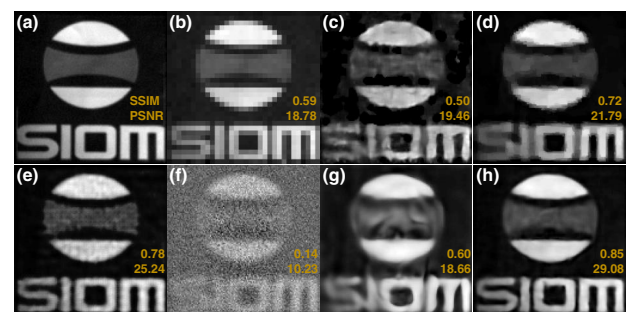


Fig. 6. Experimental results: images of the badge of our institute reconstructed by (a) HSI with $\beta = 100\%$ (it serves as the ground truth), (b) HSI with $\beta = 6.25\%$, (c) DCAN, (d) TVAL3, (e) fine-tuning with random initialization, (f) DGI with learned patterns, (g) DGI with physics-informed DNN, and (h) the fine-tuning process.

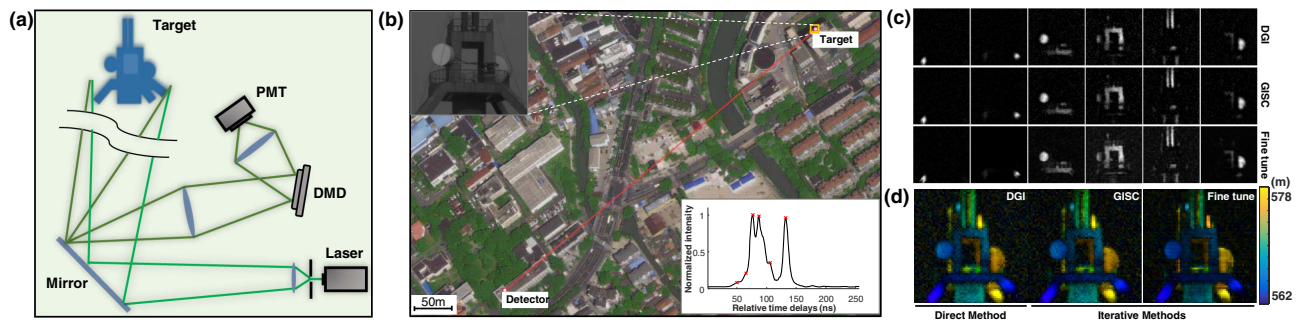


Fig. 7. Experimental results for single-pixel LiDAR. (a) Schematic diagram of the single-pixel LiDAR system. (b) Satellite image of our experiment scenario. The inset in the top left is the target imaged by a telescope, whereas the one in the bottom right is one of the echoed light signals. (c) Six typical 2D depth slices of the 3D object reconstructed by DGI with the learned patterns illumination, GISC [5], and the proposed fine-tuning method. (d) 3D images of the object reconstructed by the three aforementioned methods.

that the echoed light contains the reflectivity information of the object within the FOV. The echoed signal measured by the PMT has the dimension of 1×256 . This corresponds the imaging range from 555 to 593.4 m, which is enough to contain the whole 3D volume of the object within the FOV. The PMT measurements produced 256 1×1024 bucket signals, from each of which one can recover a depth slice of the 3D object. We plot 6 out of them in Fig. 7(c), corresponding to the time bins marked in the echoed light in the inset of Fig. 7(b). Stacking all the depth slices together, one can reconstruct the 3D image of the object [Fig. 7(d)].

For comparison, we also plot the images reconstructed by DGI with learned pattern illumination, ghost imaging via sparsity constraint (GISC) [5] side by side in Figs. 7(c) and 7(d). These two images were post-processed by use of median filtering and non-negative constraint. It is apparent that the proposed method has the best performance as evidenced by the clean background, high contrast, the fine details of the reconstructed image.

Finally, let us analyze the time efficiency. First we note that the time period to display all the 1024 learned patterns on DMD, and the DGI reconstruction for each depth-slice image is at the scale of tens of milliseconds. It is therefore in principle possible to perform 3D LiDAR imaging in real time. Comparing with the scanning imaging LiDAR [40], the proposed method has the potential to operate in a more time-efficient way.

4. CONCLUSION

We have proposed a physics enhanced deep learning framework for SPI. The incorporation of physics mainly brings two aspects of advantages. First, the physics informed decoding layer allows us to optimize the illumination patterns and improve the performance of the decoding DNN. Second, the model-driven fine-tuning process imposes an interpretable constraint to the DNN output, so that it is not restricted by the issue of generalization.

We have demonstrated the proposed methods with simulation, in-house, and outdoor experimental data. In particular, we have shown that it allows high quality SPI with β as low as 6.25%. The 3D SPI LiDAR experiment demonstrated that

the proposed framework has great potential for 3D remote sensing in real time.

In comparison to conventional data-driven deep learning [20,21] and physics-driven [26,27] optimization approaches, the proposed fine-tuning process takes advantage of both of them, making it possible to use data prior information, i. e., characteristics of the objects, for solving ill-posed inverse problems. Besides, the issue of generalization in conventional learning-based methods can be eliminated at the cost of iterative calculations. As a result, the proposed framework should be applicable for diverse computational imaging systems, not just limited to the SPI we discussed here.

However, it is worth pointing out that the proposed method relies on the accurate model of the forward propagation, making it difficult to use in the cases that the physical model cannot be accurately modeled, e.g., imaging through optically thick scattering media. Further efforts should be made to solve this problem.

Funding. National Natural Science Foundation of China (61991452, 62061136005); Key Research Program of Frontier Sciences of the Chinese Academy of Sciences (QYZDB-SSW-JSC002); Chinesisch-Deutsche Zentrum für Wissenschaftsförderung (GZ1391).

Disclosures. The authors declare that there are no conflicts of interest related to this paper.

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

REFERENCES

1. T. B. Pittman, Y. H. Shih, D. V. Strekalov, and A. V. Sergienko, "Optical imaging by means of two-photon quantum entanglement," *Phys. Rev. A* **52**, R3429–R3432 (1995).
2. B. I. Erkmen and J. H. Shapiro, "Ghost imaging: from quantum to computational," *Adv. Opt. Photon.* **2**, 405–450 (2010).
3. M. P. Edgar, G. M. Gibson, and M. J. Padgett, "Principles and prospects for single-pixel imaging," *Nat. Photonics* **13**, 13–20 (2019).

4. M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Process Mag.* **25**, 83–91 (2008).
5. W. Gong, C. Zhao, H. Yu, M. Chen, W. Xu, and S. Han, "Three-dimensional ghost imaging lidar via sparsity constraint," *Sci. Rep.* **6**, 26133 (2016).
6. C. Wang, X. Mei, L. Pan, P. Wang, W. Li, X. Gao, Z. Bo, M. Chen, W. Gong, and S. Han, "Airborne near infrared three-dimensional ghost imaging lidar via sparsity constraint," *Remote Sens.* **10**, 732 (2018).
7. B. Sun, M. P. Edgar, R. Bowman, L. E. Vittert, S. Welsh, A. Bowman, and M. J. Padgett, "3D computational imaging with single-pixel detectors," *Science* **340**, 844–847 (2013).
8. M. Sun, M. P. Edgar, G. M. Gibson, B. Sun, N. Radwell, R. Lamb, and M. J. Padgett, "Single-pixel three-dimensional imaging with time-based depth resolution," *Nat. Commun.* **7**, 12010 (2016).
9. L. Bian, J. Suo, G. Situ, Z. Li, J. Fan, F. Chen, and Q. Dai, "Multispectral imaging using a single bucket detector," *Sci. Rep.* **6**, 24752 (2016).
10. F. Magalhães, F. M. Araújo, M. Correia, M. Abolbashi, and F. Farahi, "High-resolution hyperspectral single-pixel imaging system based on compressive sensing," *Opt. Eng.* **51**, 071406 (2012).
11. N. Radwell, K. J. Mitchell, G. M. Gibson, M. P. Edgar, R. Bowman, and M. J. Padgett, "Single-pixel infrared and visible microscope," *Optica* **1**, 285–289 (2014).
12. G. M. Gibson, S. D. Johnson, and M. J. Padgett, "Single-pixel imaging 12 years on: a review," *Opt. Express* **28**, 28190–28208 (2020).
13. Z. Zhang, X. Wang, G. Zheng, and J. Zhong, "Fast Fourier single-pixel imaging via binary illumination," *Sci. Rep.* **7**, 12029 (2017).
14. M. Sun, L. Meng, M. P. Edgar, M. J. Padgett, and N. Radwell, "A Russian dolls ordering of the Hadamard basis for compressive single-pixel imaging," *Sci. Rep.* **7**, 3464 (2017).
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. O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Appl. Phys. Lett.* **95**, 131110 (2009).
17. M. Lyu, W. Wang, H. Wang, H. Wang, G. Li, N. Chen, and G. Situ, "Deep-learning-based ghost imaging," *Sci. Rep.* **7**, 17865 (2017).
18. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**, 436–444 (2015).
19. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," *Optica* **6**, 921–943 (2019).
20. F. H. Catherine, M.-S. Roderick, J. P. Miles, and P. E. Matthew, "Deep learning for real-time single-pixel video," *Sci. Rep.* **8**, 2369 (2010).
21. 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).
22. B. Neyshabur, S. Bhojanapalli, D. McAllester, and N. Srebro, "Exploring generalization in deep learning," in *Advances in Neural Information Processing Systems (NIPS)* (2017), pp. 1–10.
23. R. Shang, K. Hoffer-Hawlik, F. Wang, G. Situ, and G. P. Luke, "Two-step training deep learning framework for computational imaging without physics priors," *Opt. Express* **29**, 15239–15254 (2021).
24. A. Goy, G. Rughoobur, S. Li, K. Arthur, A. I. Akinwande, and G. Barbastathis, "High-resolution limited-angle phase tomography of dense layered objects using deep neural networks," *Proc. Natl. Acad. Sci. USA* **116**, 19848–19856 (2019).
25. A. Goy, K. Arthur, S. Li, and G. Barbastathis, "Low photon count phase retrieval using deep learning," *Phys. Rev. Lett.* **121**, 243902 (2018).
26. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," *Light Sci. Appl.* **9**, 77 (2020).
27. R. Iten, T. Metzger, H. Wilming, L. del Rio, and R. Renner, "Discovering physical concepts with neural networks," *Phys. Rev. Lett.* **124**, 010508 (2020).
28. C.-H. Lee, Z. Liu, L. Wu, and P. Luo, "Maskgan: towards diverse and interactive facial image manipulation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2020), pp. 5548–5557.
29. F. Ferri, D. Magatti, L. A. Lugiato, and A. Gatti, "Differential ghost imaging," *Phys. Rev. Lett.* **104**, 253603 (2010).
30. W. Gong and S. Han, "A method to improve the visibility of ghost images obtained by thermal light," *Phys. Lett. A* **374**, 1005–1008 (2010).
31. L. Bian, J. Suo, Q. Dai, and F. Chen, "Experimental comparison of single-pixel imaging algorithms," *J. Opt. Soc. Am. A* **35**, 78–87 (2018).
32. S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University, 2004).
33. F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE* **109**, 43–76 (2020).
34. X. Zhang, F. Wang, and G. Situ, "BlindNet: an untrained learning approach toward computational imaging with model uncertainty," *J. Phys. D* **55**, 034001 (2022).
35. K. M. Czajkowski, A. Pastuszczak, and R. Kotyński, "Real-time single-pixel video imaging with Fourier domain regularization," *Opt. Express* **26**, 20009–20022 (2018).
36. A. Pastuszczak, R. Stojek, P. Wróbel, and R. Kotyński, "Differential real-time single-pixel imaging with fourier domain regularization: applications to VIS-IR imaging and polarization imaging," *Opt. Express* **29**, 26685–26700 (2021).
37. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 9446–9454.
38. C. Li, *An Efficient Algorithm for Total Variation Regularization with Applications to the Single Pixel Camera and Compressive Sensing* (Rice University, 2010).
39. A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *14th International Conference on Artificial Intelligence and Statistics* (2011), pp. 215–223.
40. Z.-P. Li, X. Huang, Y. Cao, B. Wang, Y.-H. Li, W. Jin, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon computational 3D imaging at 45km," *Photon. Res.* **8**, 1532–1540 (2020).