EasyChair Preprint

№ 1476

Deep Learning based Fourier Spectrum Sampling Strategy for Improving Performance of Imaging

Kaiyu Zhang, Jie Cao, Lei Yan, Fanghua Zhang and Qun Hao

September 5, 2019

Deep Learning based Fourier Spectrum Sampling Strategy for Improving Performance of Imaging

Kaiyu Zhang, Jie Cao, Lei Yan, Fanghua Zhang, Qun Hao

Abstract— **Single-pixel imaging is a framework that reconstructs the target information only with a single bucket detector. The principle of the single-pixel imaging is correlating the measurements of a single bucket detector and the corresponding 2D light field distributions modulated by an optical field device in the scene. Single-pixel imaging has a good prospect in various imaging applications. To improve the imaging quality and speed, the compressed sensing and the basis scan strategies are demonstrated at the current stage. Based on the Fourier single-pixel imaging, the representative one of the basis scan strategies, and the deep learning, we propose a deep learning based Fourier spectrum sampling strategy for Fourier single-pixel imaging. Our goal is to predict and acquire the significant Fourier coefficients instead of the traditional sampling strategies to recover higher quality image under the same measurements. The simulation results demonstrate that the reconstruction image of the proposed strategy outperforms others. Applications to high quality and speed imaging could benefit from our strategy.**

I. INTRODUCTION

Single-pixel imaging (SPI) reconstructs the target information by correlating the measurements of a single bucket detector and the corresponding 2D light field distributions modulated by an optical field device in the scene. The SPI architecture enables to build low-cost, high image quality and small imaging system. [1] Therefore, SPI has been studied widely in static imaging [2], dynamic imaging [3, 4], multidimensional imaging [5], microscopic imaging [6], remote sensing [7-9], scatter imaging [10, 11] and others [12, 13].

However, the contradiction between the imaging quality and speed limits the development of the SPI [13]. To solve this problem, the compressed sensing paradigm and basis scan based methods are main solutions. [14] On the one hand, the theory of compressed sensing can be used for the reconstruction of an image from a small set of measurements at the sub-Nyquist sampling frequency, but it suffers from the huge computing time. [15] On the other hand, the basis scan strategies avoid expense of the computation and recover an image by inverse transforming (e.g. Hadamard, Fourier or wavelet) the measured data.[16-18] This kind of methods well balances the trade-off between the acquisition time and computation time. While it has the disadvantage that long acquisition time since the number of measurements is given by the number of pixels of the desired image.

In the recent years, adaptive and optimization schemes for basis scan based single-pixel imaging have emerged. [1, 11-13, 19] It depends on the predictions of the most significant basis functions for the scene or the optimization of the projected sequence of the basis functions. These strategies have been deeply studied in the frameworks of wavelet single-pixel imaging and Hadamard single-pixel imaging.

In comparison, less similar schemes are studied in Fourier single-pixel imaging (FSI). The traditional common sampling strategies in FSI are circular, spiral and diamond on basis of the uniform spectrum sampling. [18] Due to the natural image signal energy is mainly concentrated in the low-frequency, thus this implies FSI sampling the Fourier coefficients from low-frequency band to high-frequency band for reconstruction is considered as an efficient strategy. Meanwhile, owing that the maximum spatial resolution is decided by the cut-off spatial frequency given by the diffraction limit in the common optical system, the Fourier spectrum coefficients in the low-frequency band stopping at the cutoff frequency or designed upper limit frequency are often acquired for construction from the practical use of FSI. Nevertheless, these traditional common strategies have the disadvantage that only the coarse image can be reconstructed because of sampling the low-frequency coefficients prior. Moreover, the most significant Fourier coefficients are not totally concentrated in the middle and low frequency. It means that acquiring the most significant coefficients can improve the reconstruction quality than the traditional strategy under the same sampling ratio (the ratio between the number of measurements and the number of total pixels in the reconstructed image).

Deep learning is a kind of machine learning to learn the features for data modeling, sorting and decision with a deep neural network trained by a huge amount of data. [20] In the last decade, the deep learning has been developing rapidly and achieving excellent performance for scatter imaging [21], image processing [22], target classification and recognition based on image [23-26].

In this paper, combining with the FSI and deep learning, we propose a deep learning based Fourier spectrum sampling strategy for FSI. The proposed method uses deep learning to get high-quality image reconstruction under the low sampling ratio. The comparisons of the image quality between our strategy and traditional strategies are discussed. The simulation results show that the proposed strategy promises huge potentials in applications such as high-speed imaging and sensing.

Kaiyu Zhang, Jie Cao, Lei Yan, Fanghua Zhang and Qun Hao are now with the School of Optics and Photonics, Beijing Institute of Technology, Key Laboratory of Biomimetic Robots and Systems, Ministry of Education, Beijing, 100081, CHINA (corresponding author: Jie Cao, phone: +86 18810328027; corresponding e-mail: ajieanyyn@163.com, others: dxzky@126.com, 805428679@qq.com, zfh135@163.com, qhao@bit.edu.cn).

II. METHOD AND SCHEME

In this study, we combine the FSI and the convolutional neural network (CNN) to improve the imaging quality of FSI by predicting and acquiring the most important Fourier coefficients.

On the one hand, due to the theorem of 2D Fourier transform, the reconstruction result can be expressed as a series of weighted sum of $M \times N$ 2-D sinusoid patterns by the corresponding Fourier coefficients, which can be written as [16]

$$f(x,y) = \frac{1}{MN} \sum_{u=0,v=0}^{M-1,N-1} F(u,v) \cdot e^{j2\pi(\frac{xu}{M}+\frac{yv}{N})}, \quad (1)$$

where $M \times N$ is the resolution of image f, (u,v) represents the spatial frequency, and $F(u,v)$ is the corresponding Fourier coefficient. Therefore, this can be seen as acquiring the $F(u,v)$ based on the Fourier basis patterns to reconstruct the target.

According to the FSI, a Fourier basis pattern $I_F(x,y)$ can be generated by applying an 2D-inverse Fourier transform to a delta function $\delta_F(u,v,\varphi)$ [18]

$$I_F(x,y) = \frac{1}{2} + \frac{1}{2} real(F^{-1}(\delta_F(u,v,\varphi))), \quad (2)$$

where $real(\ .\ )$ represents the real part, $F^{-1}(\ .\ )$ denotes an inverse Fourier transform, the parameter φ represents the initial phase and

$$\delta_F(u,v,\varphi) = \begin{cases} \exp(j\varphi), & u = u_0, v = v_0 \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

Combining (1) - (3), FSI reconstructs the target information under view.

On the other hand, we propose a CNN to predict the significant coefficient. Fig. 1 shows that our proposed network structure.

Our proposed CNN network consists of a constructing and expansive path. The path includes convolution, rectified linear units (ReLU), max-pooling and fully connect layers. The convolution layers use convolution operations to extract the feature maps from the input images, which are used commonly in image processing. The convolution layers are denoted as 512 × 512 × 32, 256 × 256 ×32, 128 × 128 ×64, 64 × 64 × 64 and 32 × 32 × 64, respectively. Taking the first convolution layer as instance, 512 ×512 ×32 means that the resolution and channels of the outputs of this layer are 512 × 512 and 32. The convolution layers have convolution filters with a kernel size of 5 × 5, and the kernel weights are learned from the training dataset. The convolution layers reduce the number of pixels and generate the features of the input. Additionally, the zero padding are used to keep the number of pixels unchanged.

The max-pooling layers down-sample the input data from the ReLU layers to reduce the effects of changes in position and size. The down-sample rate of 2 × 2 are applied. This means the output data size is K/2 × K/2 assuming the input size is K × K.

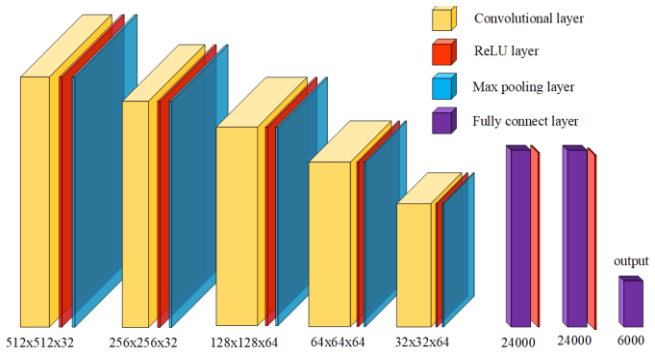

Figure 1. The proposed CNN structure .

The three fully connect layers are used to generate the indexes of the significant Fourier coefficients. The sizes of first two layers and output layer are 1 × 24000, 1 × 24000 and 1 × 6000.

The output of each convolution layer and the first two fully connect layers utilizes a ReLU layer as the activation function which allows for faster and effective training of CNN on large and complex datasets compared with the sigmoid function. [27]

To optimize the kernel weights and the parameters of the proposed CNN, the Peak Signal to Noise Ratio (PSNR) (between the recovery image that are reconstructed by the sampled and predicted coefficients and the truth image) is used as the loss function to train the network. The reconstructed images are calculated using (1). The Adam optimizer is utilized to minimize the PSNR using stochastic gradient descent (SGD). A set of datasets is selected randomly in SGD. In our simulations, the size of partial dataset is 50 and equals to the batch size. Simultaneously, the number of epochs is set as 5 to optimize the network parameters.

Combining the two frameworks above, the proposed strategy can be expressed as two-step cycle. First, the image is reconstructed from the acquired Fourier coefficients directly by (1). Then the reconstruction image inputs the CNN to predict the indexes of significant Fourier coefficients. The two steps above are looped until finishing the sampling times. Lastly, the final image can be reconstructed by the sampled data.

III. SIMULATIONS AND RESULTS

A. Training and simulation setup

Firstly, the parameters of the CNN are needed to train. In this study, the Caltech-256 [28] is chose as the dataset consisting of nearly 30000 general images. In order to simulate the acquisition, the 1% coefficients at the low frequency in the Fourier spectrum are sampled to construct the image as the inputs of the CNN and the resolution of the recovery image is 512 × 512 pixels. Therefore, we selected 15,000 images randomly from Caltech-256 and resized them to 512 × 512 pixels. Then these images are reconstructed by FSI, using (1) with sampling 1% coefficients at the low Fourier frequency. Moreover, for a more realistic simulation, the reconstruction process is generated using our optical virtual simulation system based on a real FSI system, as illustrated in Tab. 1.

TABLE 1. THE SETUP OF THE SIMULATIONS

Parameter	Value
Light source	3W LED
Angle of divergence	3mrad
Wavelength	532nm
DMD	DLP Discovery 4100 development kit (0.7-inch DMD, containing 1024×768 micro mirrors, maximum 22.7kHz binary modulation rate, 13.6×13.6 μm2 each mirror)
Detector	Thorlabs PDA36A(-EC)
Acquisition Card	Gage CSEG8 (8-Bit 2.0GS/s)

Additionally, the simulations are executed on an Intel Core i5-7600@3.40 GHz computer with 8 GB of RAM memory and Windows 10 operating system.

B. Qualitative and quantitative comparison

For quantitative comparison, we compare the proposed method with other traditional strategies under the different sampling ratio. Due that the cameraman image is the natural image used widely in imaging system tests and has multiple gray levels, we choose cameraman image as the example image. Fig. 2 shows the comparisons of the example image that are produced using the proposed strategy and other traditional strategies under the different sampling ratio. The first column shows the images recovered by the circular strategy, the second shows the images reconstructed by the spiral strategy, the third shows the results obtained by the diamond strategy, and the fourth shows results produced by the proposed strategy. The quality and contrast of the images obtained by the proposed strategy outperform others subjectively.

Then, for further quantitative evaluation, the PSNR and structural similarity index (SSIM) are selected to evaluate the performance. The PSNR and SSIM are calculated by:

$$PSNR = 10\log_{10}\left(\frac{(2^n-1)^2}{MSE}\right), \tag{4}$$

$$MSE = \frac{1}{H \times W}\sum_{I=1}^{H}\sum_{j=1}^{W}(r(i,j)-t(i,j)) \tag{5}$$

and

$$SSIM(r,t) = \frac{(2\mu_r\mu_t + c_1)(2\sigma_{rt} + c_2)}{(\mu_r^{\,2} + \mu_t^{\,2} + c_1)(\sigma_r^2 + \sigma_t^2 + c_2)}. \tag{6}$$

Here, r and t are the reconstruction result and truth, μ_r and μ_t are their means, σ_r^2 and σ_t^2 are the corresponding variances, σ_{rt} is the covariance of r and t, $c_1=(k_1L)^2$ and $c_2=(k_2L)^2$ are two constants, additionally, $L=255$, $k_1=0.01$ and $k_2=0.03$.

Then we select 10 images randomly from the common gray-scale standard images as the test images. The average of 10 comparison results are shown in Table 2. Table 2 shows that sampling strategy influences the quality of the reconstructed image and different traditional strategies may get different image quality under the different sampling ratio. Moreover, this again shows that our proposed deep learning based strategy can reconstruct better results than the traditional strategies.

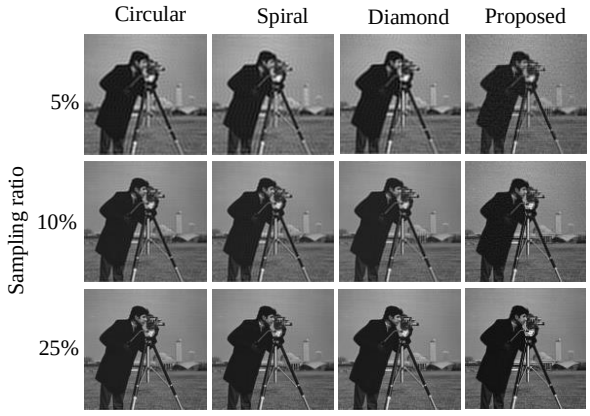

Figure 2. The comparison results by our proposed strategy and others

TABLE 2. THE AVERAGE OF THE QUANTITATIVE EVALUATION OF THE 10 RECONSTRUCTION RESULTS BY FOUR STRATEGIES

	Strategy	Sampling ratio		
		5%	10%	25%
PSNR (dB)	circular	21.34	23.10	25.23
	spiral	20.52	21.92	26.01
	diamond	21.64	21.20	25.56
	proposed	22.90	24.52	26.54
SSIM (%)	circular	62.94	71.14	80.50
	spiral	62.44	69.87	81.80
	diamond	64.13	71.94	82.93
	proposed	66.02	72.80	84.01

IV. CONCLUSION AND FUTURE WORK

In this paper, we propose a deep learning based Fourier spectrum sampling strategy for improving performance of images reconstructed with FSI and present the simulations where a CNN was trained to predict the indexes of the significant Fourier coefficients. We compare the images reconstructed by the proposed strategy with those obtained by traditional strategies. The simulation results show that the proposed deep learning based Fourier spectrum sampling strategy can get higher PSNR and SSIM. In our next work, we will try to do the physical experiments to verify and evaluate the effectiveness and performance of our strategy. It has a good application prospect to improve the image quality under the same sampling ratio and replace the traditional strategies used in FSI.

REFERENCES

[1] F. Rousset, N. Ducros, A. Farina, G. Valentini, C. D'Andrea, and F. Peyrin, "Adaptive Basis Scan by Wavelet Prediction for Single-pixel Imaging," *IEEE Transactions on Computational Imaging*, vol. PP, no. 99, pp. 1-1, 2016.

[2] S. S. Welsh, M. P. Edgar, B. Richard, J. Phillip, S. Baoqing, and M. J. Padgett, "Fast full-color computational imaging with single-pixel detectors," *Optics Express*, vol. 21, no. 20, pp. 23068-23074, 2013.

[3] M. P. Edgar *et al.*, "Simultaneous real-time visible and infrared video with single-pixel detectors," *Scientific reports*, vol. 5, p. 10669, 2015.

[4] M. P. Edgar, M.-J. Sun, G. M. Gibson, G. C. Spalding, D. B. Phillips, and M. J. Padgett, "Real-time 3D video utilizing a compressed sensing time-of-flight single-pixel camera," in *Optical Trapping and Optical Micromanipulation XIII*, 2016, vol. 9922, p. 99221B: International Society for Optics and Photonics.

[5] B. Sun, . *et al.*, "3D computational imaging with single-pixel detectors," *Science*, vol. 340, no. 6134, pp. 844-847, 2013.

[6] I.-H. Lee, M. T. Mahmood, and T.-S. Choi, "Robust focus measure operator using adaptive log-polar mapping for three-dimensional shape recovery," *Microscopy and Microanalysis*, vol. 21, no. 2, pp. 442-458, 2015.

[7] C. Zhao *et al.*, "Ghost imaging lidar via sparsity constraints," *Applied Physics Letters*, vol. 101, no. 14, pp. 139-R, 2012.

[8] H. Yu, E. Li, W. Gong, and S. Han, "Structured image reconstruction for three-dimensional ghost imaging lidar," *Optics express*, vol. 23, no. 11, pp. 14541-14551, 2015.

[9] W. Gong, C. Zhao, H. Yu, M. Chen, W. Xu, and S. Han, "Three-dimensional ghost imaging lidar via sparsity constraint," *Scientific reports*, vol. 6, p. 26133, 2016.

[10] M. Bina, D. Magatti, M. Molteni, A. Gatti, L. Lugiato, and F. Ferri, "Backscattering differential ghost imaging in turbid media," *Physical review letters*, vol. 110, no. 8, p. 083901, 2013.

[11] S. Sun *et al.*, "Multi-scale Adaptive Computational Ghost Imaging," *Scientific reports*, vol. 6, p. 37013, 2016.

[12] D. B. Phillips *et al.*, "Adaptive foveated single-pixel imaging with dynamic supersampling," *Science advances*, vol. 3, no. 4, p. e1601782, 2017.

[13] H. Jiang, S. Zhu, H. Zhao, B. Xu, and X. Li, "Adaptive regional single-pixel imaging based on the Fourier slice theorem," *Optics Express*, vol. 25, no. 13, pp. 15118-15130, 2017.

[14] O. Katz, Y. Bromberg, and Y. Silberberg, "Compressive ghost imaging," *Applied Physics Letters*, vol. 95, no. 13, p. 131110, 2009.

[15] V. Katkovnik and J. Astola, "Compressive sensing computational ghost imaging," *JOSA A*, vol. 29, no. 8, pp. 1556-1567, 2012.

[16] Z. Zhang, X. Ma, and J. Zhong, "Single-pixel imaging by means of Fourier spectrum acquisition," *Nature Communications*, vol. 6, p. 6225, 2014.

[17] K. M. Czajkowski, A. Pastuszczak, R. Kotynski, K. M. Czajkowski, A. Pastuszczak, and R. Kotynski, "Single-pixel imaging with Morlet wavelet correlated random patterns," *Scientific Reports*, vol. 8, no. 1, p. 466, 2018.

[18] Z. Zhang, X. Wang, G. Zheng, and J. Zhong, "Hadamard single-pixel imaging versus Fourier single-pixel imaging," *Optics Express*, vol. 25, no. 16, pp. 19619-19639, 2017.

[19] Z. Li, J. Suo, X. Hu, and Q. Dai, "Content-adaptive ghost imaging of dynamic scenes," *Optics express*, vol. 24, no. 7, pp. 7328-7336, 2016.

[20] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

[21] R. Horisaki, R. Takagi, and J. Tanida, "Learning-based imaging through scattering media," *Optics Express*, vol. 24, no. 13, p. 13738, 2016.

[22] H. Petersson, D. Gustafsson, and D. Bergstrom, "Hyperspectral image analysis using deep learning — A review," in *International Conference on Image Processing Theory Tools & Applications*, 2017.

[23] P. H. Liu, S. F. Su, M. C. Chen, and C. C. Hsiao, "Deep learning and its application to general image classification," in *International Conference on Informative & Cybernetics for Computational Social Systems*, 2015.

[24] L. W. Sommer, T. Schuchert, and J. Beyerer, "Deep learning based multi-category object detection in aerial images," in *Automatic Target Recognition XXVII*, 2017.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in *International Conference on Neural Information Processing Systems*, 2012.

[26] S. Chen and H. Wang, "SAR target recognition based on deep learning," in *International Conference on Data Science & Advanced Analytics*, 2015.

[27] V. Nair and G. E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," in *International Conference on International Conference on Machine Learning*, 2010.

[28] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," 2007.