Optics Letters

Learning-based single-shot long-range synthetic aperture Fourier ptychographic imaging with a camera array

Bowen Wang,[1,2] Sheng Li,[1,2] Qian Chen,[1,2] and Chao Zuo[1,2,*]

[1]Smart Computational Imaging Laboratory (SCILab), School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[2]Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
*Corresponding author: zuochao@njust.edu.cn

Received 24 October 2022; revised 28 November 2022; accepted 28 November 2022; posted 29 November 2022; published 2 January 2023

In this Letter, we report a new long-range synthetic aperture Fourier ptychographic imaging technique, termed learning-based single-shot synthetic aperture imaging (LSS-SAI). LSS-SAI uses a camera array to record low-resolution intensity images corresponding to different non-overlapping spectral regions in parallel, which are synthesized to reconstruct a super-resolved high-quality image based on a physical model-based dual-regression deep neural network. Compared with conventional macroscopic Fourier ptychographic imaging, LSS-SAI overcomes the stringent requirement on a large amount of raw data with a high spectral overlapping ratio for high-resolution, high signal-to-noise imaging of reflective objects with diffuse surfaces, making single-shot long-range synthetic aperture imaging possible. Experimental results on rough reflective samples show that our approach can improve the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) by 10.56 dB and 0.26, respectively. We also demonstrate the single-shot ptychography capability of the proposed approach by the synthetic aperture imaging of a dynamic scene at a camera-limited speed (30 fps). To the best of our knowledge, this is the first demonstration of macroscopic Fourier ptychography to single-shot synthetic aperture imaging of dynamic events.

https://doi.org/10.1364/OL.479074

Acquiring high-resolution images has emerged as an indispensable requirement in application scenarios such as astronomy, remote sensing, and geological exploration. A major limitation of remote imaging detection is the spatial resolution, which is jointly capped by the finite aperture size and pixel size corresponding to the Nyquist sampling frequency. Astronomers attempted to extend the effective aperture of the system by either employing a well-designed large-aperture lens or splicing the primary mirror, posing significant challenges for lightweight designs. This approach also tends to introduce optical aberrations with bulky dimensions, which precludes its feasibility in practical imaging. Pioneering research has emerged to circumvent the inherent limitations of imaging systems, e.g., coherent optical detection [1] and flat plate interference [2].

As a promising and elegant computational imaging approach, Fourier ptychographic microscopy (FPM) [3], invented in 2013, breaks the trade-off between the large field of view and high-resolution (HR) with a combination of synthetic aperture radar (SAR) [4] and optical phase retrieval [5]. Combined with the concept of Fourier optics, the imaging process can be understood as sampling the different regions of the HR Fourier domain of an object. Its application potential has been demonstrated in both microscopic biomedical imaging [6–8], and remote sensing [9,10], and meanwhile, the technology has been incorporated in the latest Fourier optics publications. Undoubtedly, during each acquisition, a certain amount of redundant information (at least 35% aperture overlapping percentage [11] in the Fourier domain) needs to be leveraged to perform the lost phase information decoupling, as the sensor can only record intensity information. The converged intensity and phase images are yielded by iterative optimization, jointly imposing both space and frequency-domain constraints on the observed data. This, in turn, is laborious, which implies it is less suitable for dynamic scenes, hampering its application in dynamic scenarios (default: the observed scene remains stationary over a time-lapse). Adaptive compensation [12] and simulated annealing correction algorithms [13] have also been proposed successively to tackle the artifact phenomenon in the reconstruction results, providing fast convergence speed with few computational overheads. Motivated by the rise of convolutional neural network (CNN) techniques [14] and their flexibility to the prior latent features as network layers, many efforts [15–17] to refine the FPM framework have been catalyzed. It has been proven that Neural networks can simultaneously restore the envisaged images by aggregating multi-scale features and nonlinear mappings [18,19], notably in the field of phase recovery.

In this Letter, we report a new super-resolution technique, termed learning-based single-shot synthetic aperture imaging, which is capable of "regenerating" the lost spatial resolution with deep learning. The proposed method leverages the advantages of deep learning data fitting to address the problem of

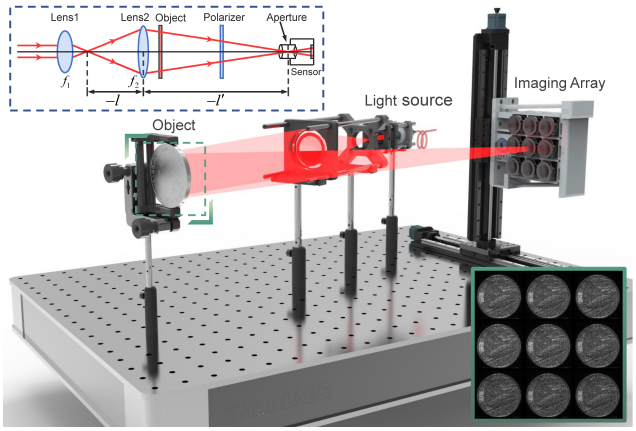

Fig. 1. Overview of the proposed LSS-SAI framework. The object is illuminated by a fiber laser, and the Fourier spectrum is formed at the aperture plane. The specific optical path tracing diagram is shown in the upper left-hand corner.

speckles and artifacts in reconstruction, rekindling sparse aperture, single shot, and ambiguity-free super-resolved imaging. Inspired by a priori knowledge in image processing, we could impute the extension of the Fourier spectrum to the physical prior of four elaborate-designed parallel network architectures. CNN can excavate more texture features from the original multiple encoded images, refining the details while balancing the speckle interference. The iterative non-negativity constraint was further employed to compute the filling of the missing information, yielding optimal outcomes. As a similar concept, the learning-based network [20,21] has achieved tremendous success in searching the map functions (statistical model of desirable target and the observational data) for various underdetermined imaging problems, verifying the feasibility of constraining the inverse problem through the network.

In the construction system, as shown in Fig. 1, a coherent laser is introduced to illuminate the 3×3 camera array on the receiver side. The sub-aperture will receive the wave vector of the incident beam from different angles, and the specific imaging process can be expressed as follows:

$$o_{output}(x,y) = h(x,y) \otimes \left[o_{input}(x,y)\, e^{iu_{m,n}x + iv_{m,n}y} \right], \quad (1)$$

where $o_{input}(x,y)$ and $o_{output}(x,y)$ represent the complex amplitudes of the input and output optical fields, respectively; $\otimes$ represents the two-dimensional convolution process; $e^{iu_{m,n}x+iv_{m,n}y}$ denotes the wave vector of the mth row and nth column of the angled incident plane light wave. The light field at the imaging aperture plane can be described as $\Psi(u,v)$ for the sensor, which can only record intensity information, and the measurements are phaseless, whereupon the intensity information recorded is formulated as follows:

$$I(x,y) = \left| \mathcal{F}^{-1} \left[\Psi(u,v) \cdot P(u-u_c, v-v_c) \right] \right|^2, \quad (2)$$

where (u_c, v_c) is the center of the aperture; $P(u,v)$ is the Fourier transform of $h(x,y)$, i.e., the coherent transfer function of the optical imaging system (NA/λ). It is also the intrinsic concept of FP to increase the resolution by obtaining the sub-spectrum at different locations, thus extending the range of the equivalent spectrum and widening the size of the equivalent aperture. We mainly focus on the reconstruction quality enhancement of the long-range rough reflective samples. Therefore, it is necessary to consider the influence of phase fluctuations on the rough surface, which implies that a random phase distribution will be integrated. The ingenious exploitation of angular illumination to mitigate the influence of speckle noise is a thoughtful approach to coherent synthetic aperture imaging. Object information is encrypted in disordered-seeming speckles, especially those related to the phase fluctuations of the surface, which can be decoded by inverse transmission matrices and CNN. From the perspective of information optics, the whole reflection process can be analogized to the dot product between the scattering layer and the smooth non-diffuse target [9]. We set the scattering intensity A and the corresponding phase φ randomly distributed in the interval $[0, 1]$ and $[0, 2\pi]$, respectively. Thereby, the complex amplitude distribution S can be expressed as $S = A\exp(i\varphi)$.

The proposed verification platform contains nine imaging sensors (pixel size of 1.85 μm) equipped with the FUJINON lens (75 mm focal length, F# from 2.8 to 16). Coherent illumination conditions are required to be satisfied in FP imaging. The illumination source employed in the demonstration is a semiconductor laser with a wavelength of 632 nm and a maximum power of 5 mW. The distance between the measured object and our system is 2.3 m. The specific imaging optical path is shown in Fig. 1, and we can observe that each sub-camera is tightly arranged (non-overlapping) with the corresponding acquired images presented in the right-hand corner. For captured image pairs, the sequence captured by the camera with a lens (F-number 4) is recorded as the ground truth HR sequence, and the sequence captured by the camera with a lens (F-number 12) is adapted to generate the corresponding LR sequence, generating a dataset for 3× super-resolved imaging. Based on this setup, we built a dataset containing 1000 raw data tailored by off-the-shelf detectors for the proposed network. Figure 2 reveals the intrinsic model of the proposed network, which is a nine-path CNN. It is noteworthy that the method takes the form of a "forward generation–reverse regression" procedure. The network attempts to recover an estimate of the envisioned object from the degraded image by prior mapping knowledge (e.g., the system transfer function). Physics-informed learning seamlessly incorporates both data and mathematical models to address the under-determined problem, even in noisy and high-dimensional contexts. Nine non-overlapping intensity data were fed into the network simultaneously to derive a high-resolution super-resolved image with high signal-to-noise. We hypothesize that the training process can be regarded as a prior learning

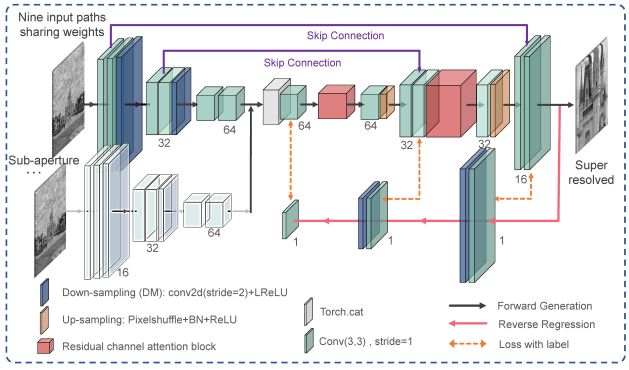

Fig. 2. Proposed network follows the form of a "forward generation–reverse regression" procedure. The proposed method is a nine-input, single-output supervised network.

process, including illumination angle, speckle noise, and acquisition position, which will further bolster the interpretability of the model.

The proposed algorithm is flexible since once the training task is completed, no further manual adjustment of parameters is required to optimize the reconstruction performance. In order to yield reasonable predictions, the forward generation process and the reverse regression process are simultaneously constrained in the network model, and the dual loss functions compensate each other to produce the entire loss function balance. In the constructed network environment, the Adam optimizer is installed to implement structure feedback, with initial learning rate, batch size, and epoch setting to 10^{-4}, 4, and 200, respectively. The model was operated on the configured computer with an Nvidia RTX2080Ti graphics card and an Intel Core ™ i7-9700K CPU @ 3.60GHz×8 processor. The training and testing times for the entire network are 7.7 hours and 0.2 seconds, respectively. The network feeds back/refines the fitting errors between the data through the optimizer to avoid producing contrived results, thus, effectively retaining latent texture information. The mathematical expression of the loss function is presented as

$$Loss = \sum_{i=1}^{N} \mathrm{Loss}_1 \left[F\left(x_i\right), y_i\right] + \lambda\, \mathrm{Loss}_2 \left[D\left(y_i\right), x_i\right], \quad \textbf{(3)}$$

where x_i and y_i represent the input LR and output HR images, respectively; $\mathrm{Loss}_1 \left[F\left(x_i\right), y_i\right]$ and $\mathrm{Loss}_2 \left[D\left(y_i\right), x_i\right]$ describe the loss functions of forward regression and inverse regression tasks, respectively. The super-resolved image $F(x_i)$ is constantly approaching the similarity with its corresponding HR image in the training process. Simply put, the similarity between the predicted map $D(y_i)$ and the forward-fed map is continuously approached during the regression process. Hinting that the forward loss value is preferred, hereby, we set the weight distribution λ of the hybrid loss function to 0.1.

We evaluated the proposed method on both synthetic and real-world datasets. To test the effectiveness of the proposed method, we first reconstructed the low-resolution scene without coherent speckles (dataset was created by DIV2K [22]). To establish the unique advantages of the proposed method over traditional ptychography imaging, we conduct quantitative analyses in terms of input images having different overlaps, as illustrated in Figs. 3(a1)–3(a4). The bottom row of Figs. 3(a1)–3(a5) presents the line profile along the red dotted line. As one would expect, the robustness of the network is boosted, and more texture components of the image are reproduced with the increasing amount of data fed into the network. We also perform the reconstruction of diffuse reflective objects (rough paper), as shown in Fig. 3(b), and the corresponding zoomed-in areas are shown in Figs. 3(b1)–3(b5) and Figs. 3(c1)–3(c5). It is noted that the proposed network results still defeat the other network method [23], such as Generative Adversarial Network, in terms of the maximum improvement of 3.38 dB in peak signal-to-noise ratio (PSNR). The learning-based single-shot synthetic aperture imaging (LSS-SAI) approach supports significant improvement (extreme reduction from 50 minutes to 0.2 seconds) in imaging speed with a negligible decline in reconstruction quality against the related methods.

Furthermore, we selected a coin made of metal alloy with a diameter of 27 mm as the object, as shown in Fig. 4. Figures 4(c1)–4(c5) demonstrate the different reconstruction results

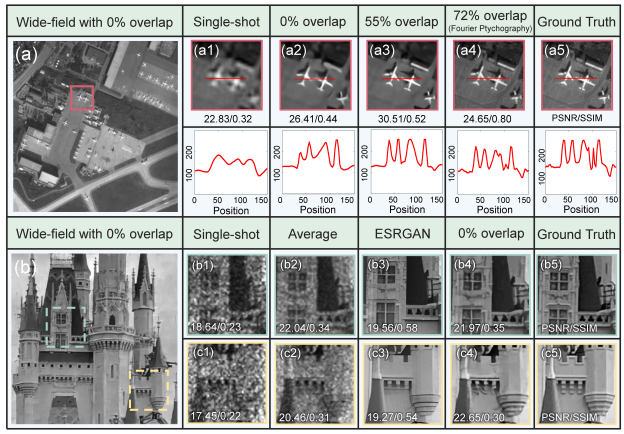

Fig. 3. Comparison of network reconstruction results. (a) Simulation reconstruction results of specular objects. (b) Performance of the LSS-SAI platform for the smooth object. (a1)–(a5), (b1)–(b5), (c1)–(c5) Corresponding region reconstruction results.

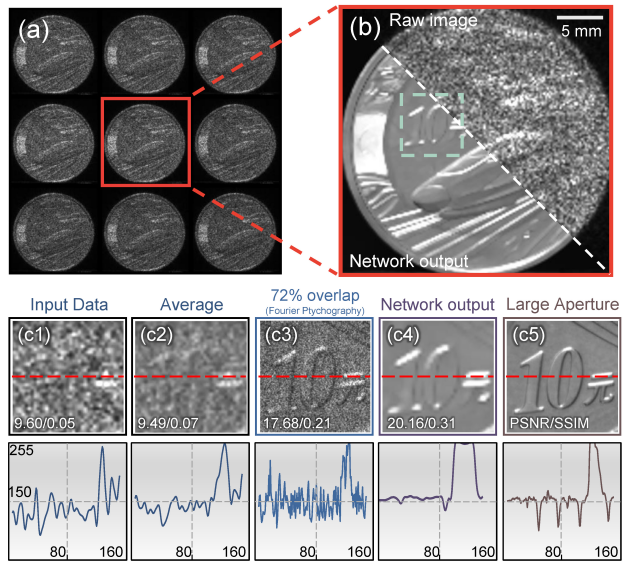

Fig. 4. Comparison of reconstruction results in the case of diffuse reflection. (a) The original images fed into the network. (b) Predicted reconstruction results of the network. (c1)–(c5) Results of image quality evaluations for the region of interest in different cases/methods (the reference image with an aperture F-number of 4 is selected as the label).

in the magnified region of interest for the commemorative coin. The gray scale profile of the magnified region is plotted below the corresponding one, for which the smooth profile indicates lower scattering noise. Figure 4(c1) illustrates that the signal-to-noise ratio and the resolution of the coin are too inferior to distinguish the features. As shown in Fig. 4(c2), although the noise is partially suppressed as a result of taking the cumulative average of the nine sub-aperture maps, the detailed components of the images are not yet reproducible. Theoretically, with enough low-resolution images, this method is able to increase the resolution of the image by a factor of two. Although the high-frequency components of the image are a super-resolved reconstruction, there is still significant speckle

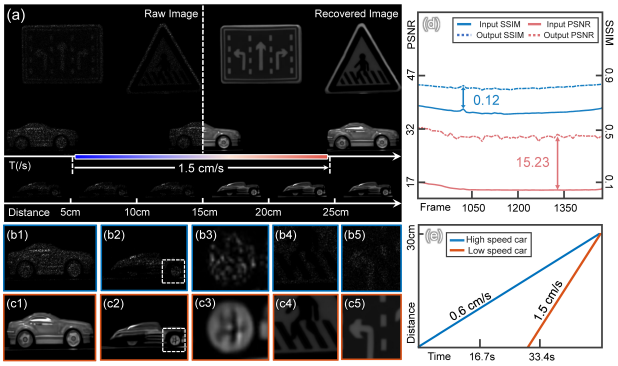

Fig. 5. Constructed vehicle dynamic pursuit imaging results. (a) Comparison of the recorded low-resolution image (under F-number 12) with the predicted super-resolved image. (b1)–(b5) Magnification of the regions in the low-resolution images. (c1)–(c5) Super-resolution reconstruction results of the corresponding regions. (d) Comparison of the PSNR and SSIM curves of the original image and the reconstructed image in the dynamic experiment. (e) The corresponding curves of displacement distance versus time for the two vehicles.

noise in Fig. 4(c3). Moreover, the same trade-off between massive data volume (temporal resolution) and spatial resolution has to be compromised. On the contrary, the reconstruction result of the proposed method is presented in Fig. 4(c4), improving 10.56 and 0.26 in both PSNR and structural similarity (SSIM) indexes, respectively. The proposed method demonstrates efficient noise suppression capabilities while improving image resolution, elegantly solving the problems of conventional methods.

Moreover, we demonstrate the high temporal resolution of LSS-SAI by performing super-resolved videography of a dynamic scenario containing two isolated samples (See Visualization 1 for the whole video recording). The established system is depicted in Fig. 5, which employs a linear displacement stage to push the movement of both toy vehicles separately and places a stationary sign at the rear. Figure 5(b) shows that the reconstructed result from raw low-resolution data is too coarse to identify the model tire and steering sign details, which is capped by the combined limitation of aperture diffraction and laser speckle noise. In contrast, the proposed LSS-SAI yielded the best super-resolved reconstruction with well-reproduced surface details, as shown in Fig. 5(c), which is almost reproduced to the ground-truth data. The reconstruction method improves the PSNR and SSIM metrics by up to 0.12 and 15.23, respectively, compared with the original image for 1500 consecutive frames. Furthermore, we also performed the corresponding linear fits for the two moving targets, which can be inversely calculated as 0.6 cm/s and 1.5 cm/s for the two vehicles, respectively. Experiments show that the proposed algorithm is a powerful approach for improving the performance of Fourier ptychography even if containing complex speckle noise.

In this Letter, we have presented a learning-based single-shot synthetic aperture imaging, endowing the capability to overcome the reconstruction quality deterioration and stringent overlapping ratio constraints in conventional FP. Moreover, thanks to its single-shot nature, LSS-SAI is fundamentally immune to artifact induced by object motion. The proposed method has great potential for performing super-resolution imaging of macroscopic diffuse reflectance observations. More modifications and innovations remain to be implemented in further, e.g., whether it is promising to reconstruct the phase information of far-field diffuse scattering objects.

Funding. National Natural Science Foundation of China (U21B2033, 61905115, 62105151, 62175109); Leading Technology of Jiangsu Basic Research Plan (BK20192003); National Major Scientific Instrument Development Project (62227818); Space Optoelectronic Measurement and Perception Lab (No.LabSOMP-2022-05); Youth Foundation of Jiangsu Province (BK20190445, BK20210338).

Disclosures. The authors declare no conflicts of interest.

Data availability. Data 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. S. Ralston, D. L. Marks, P. S. Carney, and S. A. Boppart, Nat. Phys. **3**, 129 (2007).
2. Q. Chu, Y. Shen, M. Yuan, and M. Gong, Opt. Commun. **405**, 288 (2017).
3. G. Zheng, R. Horstmeyer, and C. Yang, Nat. Photonics **7**, 739 (2013).
4. A. Moreira, P. Prats-Iraola, M. Younis, G. Krieger, I. Hajnsek, and K. P. Papathanassiou, IEEE Geosci. Remote Sens. Mag. **1**, 6 (2013).
5. L.-H. Yeh, J. Dong, J. Zhong, L. Tian, M. Chen, G. Tang, M. Soltanolkotabi, and L. Waller, Opt. Express **23**, 33214 (2015).
6. P. Song, S. Jiang, T. Wang, C. Guo, R. Wang, T. Zhang, and G. Zheng, Photonics Res. **10**, 1624 (2022).
7. G. Zheng, C. Shen, S. Jiang, P. Song, and C. Yang, Nat. Rev. Phys. **3**, 207 (2021).
8. L. Tian, Z. Liu, L.-H. Yeh, M. Chen, J. Zhong, and L. Waller, Optica **2**, 904 (2015).
9. J. Holloway, Y. Wu, M. K. Sharma, O. Cossairt, and A. Veeraraghavan, Sci. Adv. **3**, e1602564 (2017).
10. S. Dong, R. Horstmeyer, R. Shiradkar, K. Guo, X. Ou, Z. Bian, H. Xin, and G. Zheng, Opt. Express **22**, 13586 (2014).
11. J. Sun, Q. Chen, Y. Zhang, and C. Zuo, Opt. Express **24**, 15765 (2016).
12. Y. Shu, J. Sun, J. Lyu, Y. Fan, N. Zhou, R. Ye, G. Zheng, Q. Chen, and C. Zuo, PhotoniX **3**, 24 (2022).
13. J. Sun, Q. Chen, Y. Zhang, and C. Zuo, Biomed. Opt. Express **7**, 1336 (2016).
14. C. Zuo, J. Qian, S. Feng, W. Yin, Y. Li, P. Fan, J. Han, K. Qian, and Q. Chen, Light: Sci. Appl. **11**, 39 (2022).
15. S. Jiang, K. Guo, J. Liao, and G. Zheng, Biomed. Opt. Express **9**, 3306 (2018).
16. L. Boominathan, M. Maniparambil, H. Gupta, R. Baburajan, and K. Mitra, "Phase retrieval for fourier ptychography under varying amount of measurements," arXiv, arXiv:1805.03593 (2018).
17. C. Wang, M. Hu, Y. Takashima, T. J. Schulz, and D. J. Brady, Opt. Express **30**, 2585 (2022).
18. J. Tang, K. Wang, Z. Ren, W. Zhang, X. Wu, J. Di, G. Liu, and J. Zhao, Opt. Lasers Eng. **139**, 106463 (2021).
19. K. Wang, J. Dou, Q. Kemao, J. Di, and J. Zhao, Opt. Lett. **44**, 4765 (2019).
20. B. Wang, Y. Zou, L. Zhang, Y. Li, Q. Chen, and C. Zuo, Opt. Lasers Eng. **156**, 107078 (2022).
21. K. Wang, M. Zhang, J. Tang, L. Wang, L. Hu, X. Wu, W. Li, J. Di, G. Liu, and J. Zhao, PhotoniX **2**, 8 (2021).
22. E. Agustsson and R. Timofte, in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (2017).
23. X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, in *Proceedings of the European Conference on Computer Vision (ECCV) Workshops* (2018).