Supplemental Document

Optics Letters

Fourier phase retrieval using physics-enhanced deep learning: supplement

ZIKE ZHANG,[1,2,3] FEI WANG,[1,2,*] QIXUAN MIN,[1,2] YING JIN,[1,2] AND GUOHAI SITU[1,2,3,4,5]

[1] *Wangzhijiang Innovation Center for Laser, Aerospace Laser Technology and System Department, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China*
[2] *Key Laboratory of Space Laser Communication and Detection Technology, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China*
[3] *Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China*
[4] *Hangzhou Institute for Advanced Study, University of Chinese Academy of Sciences, Hangzhou 310024, China*
[5] *ghsitu@siom.ac.cn*
[*] *wangfei@siom.ac.cn*

This supplement published with Optica Publishing Group on 23 October 2024 by The Authors under the terms of the Creative Commons Attribution 4.0 License in the format provided by the authors and unedited. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

Supplement DOI: https://doi.org/10.6084/m9.figshare.27135180

Parent Article DOI: https://doi.org/10.1364/OL.537792

Fourier phase retrieval using physics-enhanced deep learning: supplemental document

This document complements Fourier Phase Retrieval Using Physics-Enhanced Deep Learning by providing detailed information on the neural network structure, quantitative metrics, robustness analysis, results on more complex datasets, and preprocessing of experimental data.

1. The neural network structure

Fig. S1 shows the structure of our network. The network comprises an encoder and a decoder. The encoder performs downsampling through convolutions with a stride of two, while the decoder performs upsampling using transposed convolutions. Skip connections are utilized to preserve detailed information and make more effective use of the data. More details about the neural network structure and hyperparameters can be found in the released source code.

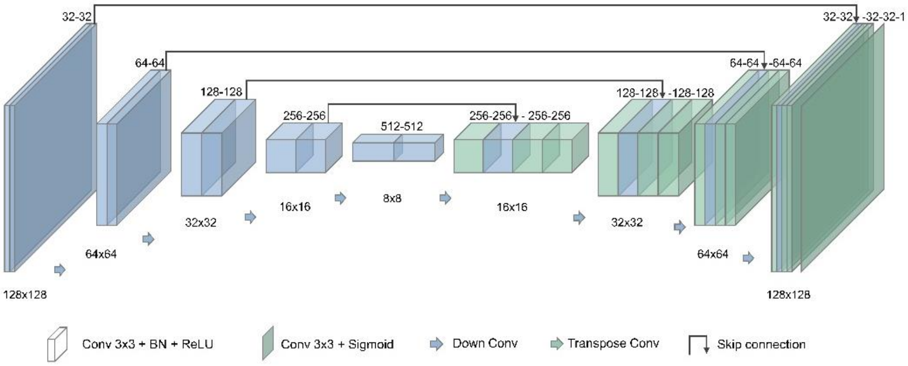

Fig. S1. Neural network structure and hyperparameters.

2. Quantitative metrics on EMNIST

To eliminate potential variability in reconstruction metrics due to specific test data, we randomly selected 30 unseen test samples from the EMNIST dataset for evaluation. As shown in Fig. S2, our method achieves the best image reconstruction metrics.

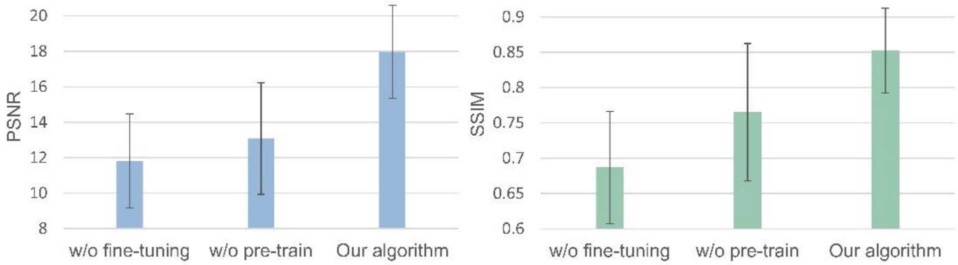

Fig. S2. Averaged metrics (PSNR/SSIM) of 30 randomly selected EMNIST images using different reconstruction algorithm.

3. Time efficiency

As shown in Table S1, end-to-end methods provide the fastest image reconstruction. However, our method, while slower due to the physics-driven fine-tuning process, achieves better image reconstruction results (see Visualization 1). It is worth noting that for test objects similar to the training data (MNIST), our method requires fewer iterations, leading to shorter computational times for the data from EMNIST compared to the "trile-slit" data.

Table S1. Time consumption statistics for ten repetitions. w/o FT: without fine-tuning. w/o PT: without pre-training.

	HSE	**DNN-HIO**	**w/o FT**	**w/o PT**	**Ours**
EMINST	90.71±8.72 s	59.18±1.53 s	0.092±0.030 s	7.74±4.66 s	5.19±3.28 s
triple-slit	65.48±8.64 s	58.53±1.29 s	0.072±0.021 s	41.22±18.30 s	20.03±8.82 s

4. NOISE ROBUSTNESS

To analyze the noise robustness of the proposed method, we added varying levels of Gaussian white noise to a simulated Fourier intensity measurement, resulting in four different signal-to-noise ratios: noise-free, 20 dB, 10 dB, and 0 dB. We compared the results with those from typical reconstruction algorithms. The main visualization results are shown in Fig. S3, and the quantitative evaluation metrics are presented in Fig. S4. Clearly, although there is a slight decrease in image reconstruction quality when noise is present, our method still achieves the best reconstruction results, demonstrating its strong noise robustness.

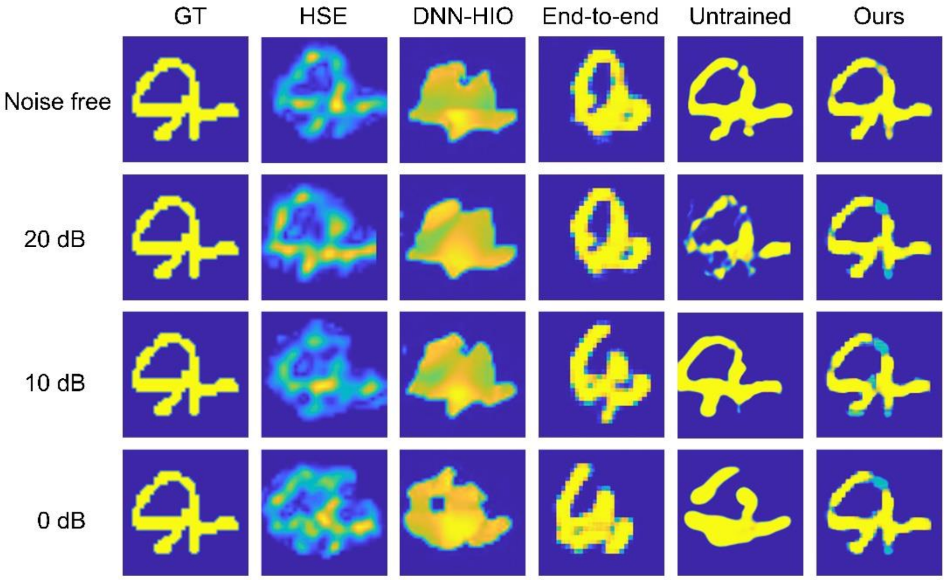

Fig. S3. Visualization results of different reconstruction algorithms at different SNRs.

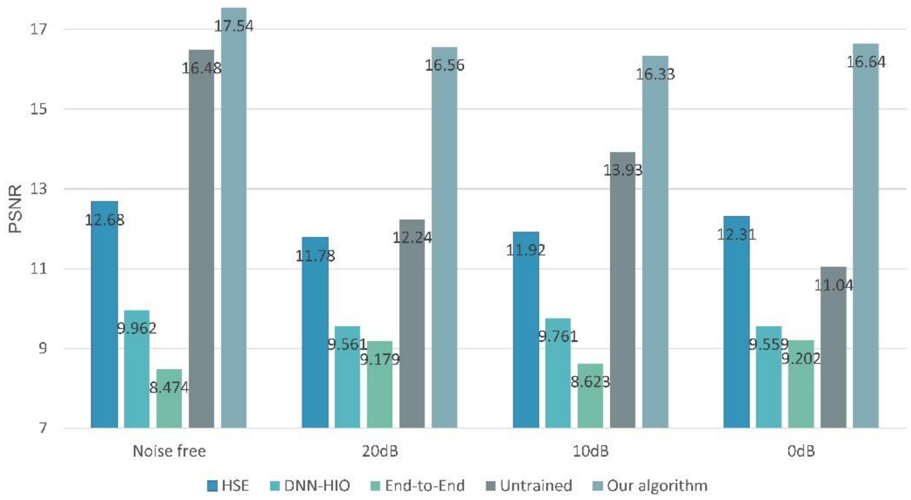

Fig. S4. Quantitative metrics of different reconstruction algorithms at different SNRs.

5. RESULTS ON COMPLEX OBJECTS

To demonstrate that our method can effectively reconstruct objects with complex structures, we selected 27,972 images from the CelebA dataset for pre-training. A circular mask was applied to extract the central region of each CelebA training image, which was then treated as a phase object with a phase range of 0 to 1 rad and resized to 128×128 pixels. The network was trained using the same hyperparameters as those applied for the MNIST dataset in the main text.

To evaluate the performance of our method, we randomly selected an unseen face image from the CelebA dataset. The main results are presented in Fig.~\ref{fig: complex face}, where our method successfully reconstructed the face image, while traditional methods failed to handle such intricate structures.

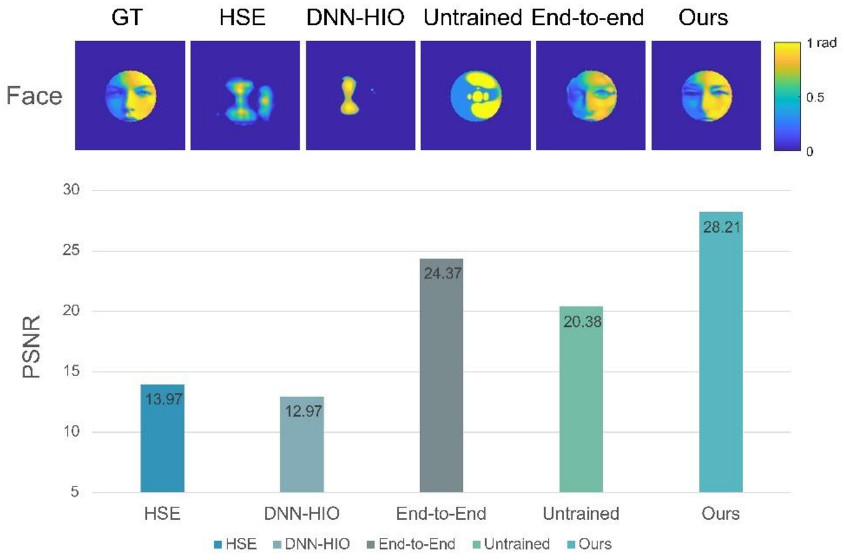

Fig. S5. Results on a complex phase object. The model was pre-trained on CelebA.

6. Data preprocessing

In the experiment, the raw Fourier intensity images were 2048 × 2048 pixels. To ensure consistency with the input size of the pre-trained model, as shown in Fig. S6, we first cropped the original images to 300 × 300 pixels and then resized them to 128 × 128 pixels.

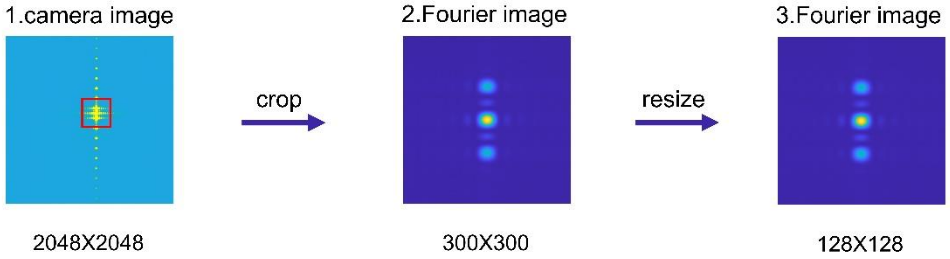

Fig. S6. Illustration of experiment data preprocessing

7. QUANTITATIVE METRICS OF RESULTS FROM CDI EXPERIMENT

Our approach leads in various metrics for different objects. Despite occasional superior metrics by the DNN-HIO method, our approach consistently delivers visually superior outcomes (Fig. 4). Table S2 is the detailed metrics of the CDI experiment.

Table S2. Metrics of results from CDI experiment

Object	Metrics	DNN-HIO	HSE	Untrained	End-to-end	Ours-w/o win.	Ours-w/ win
"4"	VMSE	**0.0126**	0.0153	0.0267	0.0525	0.0241	0.0184
	SSIM	**0.869**	0.8328	0.8371	0.7303	0.8328	0.8603
	PSNR	**18.99**	18.16	15.73	12.8	16.18	17.34
	MSSSIM	**0.9036**	0.8779	0.8319	0.7653	0.8603	0.8835
"5"	VMSE	0.0203	0.0188	0.0376	0.0643	0.0329	**0.0166**
	SSIM	0.8456	0.7585	0.7957	0.6945	0.8035	**0.8455**
	PSNR	16.93	17.27	14.25	11.92	14.83	**17.8**
	MSSSIM	0.8668	0.8434	0.8056	0.6736	0.8238	**0.9191**
"6"	VMSE	0.0233	0.0245	0.0300	0.0752	0.0302	**0.0194**
	SSIM	0.8403	0.7407	0.8152	0.6929	0.8229	**0.8437**
	PSNR	16.32	16.11	15.23	11.24	15.21	**17.13**
	MSSSIM	0.8468	0.7966	0.8537	0.5893	0.8505	**0.9035**
"三"	VMSE	**0.0451**	0.0655	0.0565	0.0890	0.0568	0.0523
	SSIM	0.6309	0.5673	0.5917	0.5823	0.5909	**0.65**
	PSNR	**13.46**	11.84	12.48	10.51	12.46	12.82
	MSSSIM	**0.8204**	0.7102	0.7788	0.6137	0.7958	0.7768