Supplementary Material for Reference-less decomposition of highly multimode fibers using a physics-driven neural network

Qian Zhang[1*†], Yuan Sui[1†], Stefan Rothe[2†],
Nektarios Koukourakis[1], Jürgen Czarske[1]

[1*]Laboratory of Measurement and Sensor System Technique, TU Dresden, Helmholtzstrasse 18, 01069, Dresden, Sachsen, Germany.
[2]Department of Applied Physics, Yale University, 06520, New Haven, CT, USA.

*Corresponding author(s). E-mail(s): qian.zhang@tu-dresden.de;
†These authors contributed equally to this work.

1 Decomposition on Synthetic Data without Noise

Multimode fibers (MMFs) have attracted significant research interest in various applications, such as telecommunications, ultra-thin endoscopy, and fiber lasers. For different scenarios, the number of modes required varies from a few modes for communication to thousands of modes for fiber imaging. Therefore, it is necessary to investigate the performance of the method on different numbers of modes. To train and test the neural network used in this work, we compute the eigenmode fields of the MMF by simulation. Once the prior knowledge of the MMF is known, the data can be generated by superposing the eigenmodes with pseudo-random mode weights. Benefiting from the easy accessibility and flexibility in simulation, the performance of the PhyMoNet can be quickly and comprehensively tested through synthetic data. In the first test scenario, only the data without any noise is considered. 100 images with the resolution of 64×64 pixel have been generated for all test cases. The test results are given in Table S1. It can be found that the predicted weights of the PhyMoNet can approximate the test light field well from 3 to 23 modes, where the average correlation coefficient exceeds 99%. The correlation coefficient gradually decreases as the number of modes increases which is related to the exponentially increasing complexity

Table S1: Performance of PhyMoNet on synthetic data without noise.

Modes	$\bar{\Gamma}$	σ_Γ	$\Delta\rho$	$\Delta\rho_r$	$\Delta\theta$
3	99.13%	3.45%	0.0358	4.71 %	0.0391
6	99.12%	2.17%	0.0748	18.33%	0.1112
8	99.01%	2.01%	0.0701	19.76%	0.1391
10	99.09%	1.37%	0.0596	21.29%	0.1582
15	99.06%	1.09%	0.0722	27.97%	0.2021
19	99.01%	0.94%	0.0729	32.07%	0.2115
23	99.04%	0.75%	0.0670	32.12%	0.2242
40	98.83%	0.65%	0.0614	40.33%	0.2606
55	98.88%	0.44%	0.0581	43.12%	0.2812
109	98.58%	0.39%	0.0482	49.89%	0.2980
142	98.52%	0.33%	0.0436	51.96%	0.2997
220	98.32%	0.27%	0.0367	54.40%	0.3033
409	97.74%	0.28%	0.0281	56.92%	0.3157
1033	98.42%	0.17%	0.0185	59.74%	0.3256

of the light field. It is interesting to note that in the 1033-mode, the PhyMoNet performs better than the 409-mode instead. The main reason is the limited resolution of the light fields. The loss of detail in the structure of the light field makes the correlation coefficient higher. This result indicates that for more modes, higher resolution is necessary. The increase in relative amplitude error and phase error also suggests that the neural network should give relatively poorer results on the 1033 modes.

In order to fully explore the effects of image size, additional tests of the relationship between the resolution of test images and performance were conducted with 1033 modes as given in Table S2. It can be found that as the image size increases, the results for high-resolution images get worse and worse for the same number of training epochs. The accuracy of the decomposition can be improved by increasing the number of training epochs. For example, with the resolution of 128 × 128, 1500 epochs are required to achieve the same correlation as with the resolution of 64 × 64, whereas the value is 2000 if the resolution is 300 × 300.

Table S2: Performance of PhyMoNet on synthetic data with different image sizes for 1033 modes.

Image size	Epoch				
	500	1000	1500	2000	2500
64 ×64	98.02%	98.42%	98.54%	98.59%	98.62%
128 ×128	97.70%	97.96%	98.02 %	98.05%	98.08%
300 ×300	97.57%	97.88%	98.00%	98.03%	98.05%

To intuitively understand the performance and decomposition time of PhyMoNet on different modes, the training processes of 3-, 55-, and 109-mode are taken as examples and visualized in Fig. S1. It can be observed that the PhyMoNet converges fastest on 3 modes, reaching approximately 98% correlation after about 170 epochs. On 55 and 109 modes, it needs around 240 and 310 epochs, respectively, to achieve the same level of correlation. In this work, each epoch takes about 3.5 ms, which means that the characterization of a single channel of few-mode MMF can be done in less than one second during the characterization using digital phase conjugation. It is expected that this time can be shortened by FPGA acceleration [1, 2] or pre-training of the neural network.

In contrast to our method, data-driven neural networks have an advantage in the decomposing time [3–5]. However, only amplitude weights can be directly obtained from the prediction. To infer the phase information from the predicted vectors takes a longer time, which varies with the number of modes. Due to the ambiguity caused by phase, the cosine value of the phase is used as the label. Due to the ambiguity caused by phase, the cosine value of the phase is used as the label. This means that when getting the phase back via the arccos function, two possible phase values (with positive and negative signs) could be obtained. To determine the phase information from the predicted vectors, it is necessary to calculate the correlation coefficients between all possible intensity distributions and the target distribution. The correct combination of phase signs corresponds to the maximum correlation coefficient. For example, MTNet is able to decompose 23 modes at the speed of about 1.1 ms [5]. However, the comparison takes about 10 hours in our previous work, as 2^{21} combinations are compared with the target images [5].

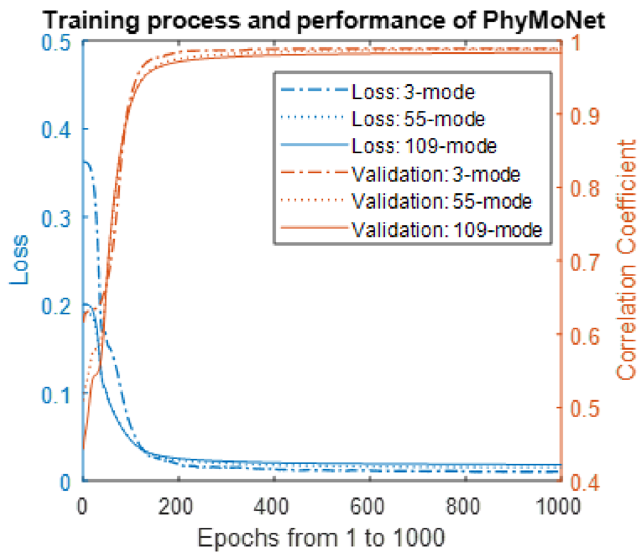

Fig. S1: Training process on selected mode cases. The shown training process and results are the average loss and average correlation coefficient for 100 test data.

2 Decomposition on Synthetic Data with Noise

In this work, a white Gaussian noise model is introduced for simulating the noise in experiments. The generation of noise data can be represented as:

$$I_{noisy} = min[max[0, I + G(0, \sigma_n^2) \cdot max(I)], 1], \tag{1}$$

where G is the white Gaussian noise model. To better mimic the actual noise, negative values in intensity images should be avoided. Therefore, the pixels with a value smaller than zero are set to 0, and larger than 1 are set to 1. Noise factor σ_n^2 determines the variance of the noise, which is obtained by reverse calculation from the signal-to-noise ratio (SNR):

$$\sigma_n^2 = \sqrt{\sigma_I^2 / 10^{\frac{SNR}{10}}}. \tag{2}$$

Here, σ_I^2 is the variance of the light field. The unit of SNR is decibels (dB). For a fair and comprehensive investigation, SNR varies from 1 dB to 30 dB for all mode cases. The correlation coefficient between the noise data and the reconstruction from the predicted mode weights is calculated. The standard deviation as an indicator of stability was also calculated. The figures-of-merit are calculated and averaged over these 100 decompositions. Fig. S2(a) shows the results. It can be seen that the larger the SNR of the noisy test image, the closer the reconstructed image is to the test image. It should be noted that when compared to the original noiseless data, the reconstructed image has a correlation coefficient over 95% even at an SNR of 1 dB as shown in Fig. S2(b). The difference between Γ_n and Γ_o is mainly attributed to the interference brought about by the noise information in the noisy image, which makes the reconstructed image different from the original image. The comprehensive investigation of different cases from 3-mode to 220-mode is shown in Table S3.

A further study found that the impact of noise data on the convergence speed of the network model is almost negligible. Fig. S3 illustrates the training process with noisy data for 55-mode. For a more intuitive analysis, correlation coefficients were computed for the reconstructed light field and the noise light field. The results represent the average of 100 samples. It can be observed that all training processes converge rapidly within the first 200 epochs and gradually stabilize.

3 Decomposition on Experimental Data

MMFs with different numbers of modes can be used in different scenarios, such as in communication networks, image transmission, and high-power lasers. In this work, we have tested different fibers supporting few modes, tens of modes, and thousands of modes, respectively, In the current communication systems, few modes and tens of modes are usually employed. Here, 2 fibers on hand were used as test fibers to collect data. The first fiber has a core diameter of 10 μm and a numerical aperture of NA = 0.1. A generalized experimental setup design is used as shown in Fig. S4. With the laser source @532 nm (Spectra-Physics, Excelalor-532-15), this fiber with a length of 1 m supports 8 different spatial modes per polarization state. Another fiber, which

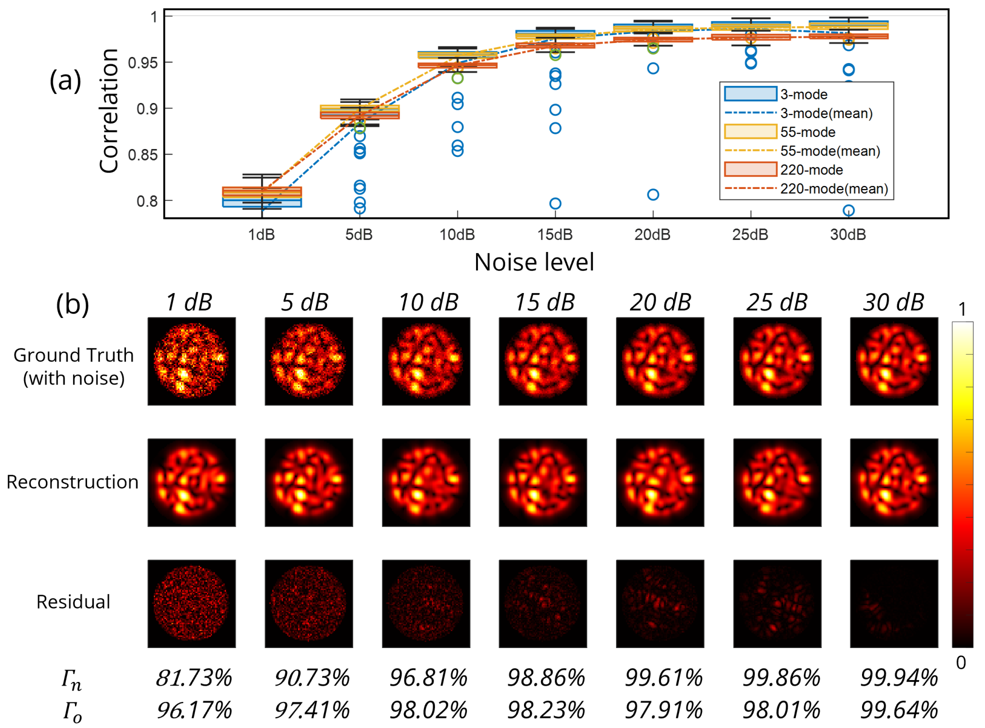

Fig. S2: Training results of synthetic data with noise. (a): Impact of noise levels on decomposition performance (the results shown are statistical results for 100 test samples); (b): examples of data with different noise levels on the 55-mode case. Γ_n: correlation between the prediction and noisy data; Γ_o: correlation between the prediction and original data without noise.

supports 65 modes and has a length of 1 km, fits closer to the requirement of the application scenario of short-distance communication for data centers.

For MMF-based endoscopy, thousands of modes are required for transmitting imaging information. A fiber with a core diameter of 200 μm and a numerical aperture of NA = 0.39 supports 5796 spatial modes at 1550 nm (ID Photonics CoBrite DX1) is then used in the work for testing. These fibers are named as MMF_8 MMF_{65}, and MMF_{5796}, respectively. The same optical system design was used, but the optics were changed depending on the wavelength. The data from MMF_8 was collected by a CMOS camera. Before being sent into the PhyMoNet, each intensity-only image from MMF_8 was cropped and resized to the resolution of 64 $\times$ 64 pixels which is mainly for a fair comparison with previous results. Images from MMF_{65} and MMF_{5796} were cropped and resized to 128 x 128 to maintain the detailed information. The size of the images depends on the actual need for precision and the resolution of the camera in the optical system. Also, larger images tend to require more computational resources. As a proof of concept, we chose a balanced resolution. The results are given in Table S4. It can be found that among these 3 different networks, PhyMoNet achieved the best

Table S3: Performance of PhyMoNet on synthetic data with noise.

Modes	1 dB		5 dB		10 dB		15 dB		20 dB		25 dB		30 dB	
	Γ	σ_Γ	Γ	σ_Γ	Γ	σ_Γ	Γ	σ_Γ	Γ	σ_Γ	Γ	σ_Γ	Γ	σ_Γ
3	78.84%	4.01%	88.38%	4.28%	94.90%	3.99%	97.57%	2.41%	98.38%	3.05.%	98.62%	2.67%	98.15%	4.69%
5	79.02%	2.58%	88.87%	1.47%	94.48%	2.31%	97.19%	1.54%	97.94%	1.59%	97.79%	2.76%	98.15%	1.94%
8	80.43%	1.06%	89.62%	1.28x%	95.90%	0.72%	97.87%	1.73%	98.76%	1.52%	98.99%	1.26%	98.75%	1.98%
10	79.89%	1.34%	89.34%	1.15%	95.30%	1.09%	97.26%	2.06%	98.25%	1.36%	98.69%	1.14%	98.75%	1.18%
15	80.10%	0.93%	89.54%	0.76%	95.37%	1.17%	97.83%	0.88%	98.69%	0.72%	98.93%	0.72%	98.89%	1.09%
19	80.44%	0.74%	89.80%	0.68%	95.66%	0.63%	98.04%	0.59%	98.68%	0.71%	98.85%	0.80%	98.92%	0.99%
23	80.50%	0.80%	89.86%	0.68%	95.61%	0.61%	97.99%	0.53%	98.71%	0.76%	98.83%	0.75%	99.00%	0.76%
40	80.50%	0.74%	89.79%	0.53%	95.54%	0.58%	97.82%	0.50%	98.44%	0.56%	98.80%	0.60%	98.77%	0.64%
55	80.82%	0.73 %	89.96%	0.50 %	95.67%	0.44 %	97.77%	0.40%	98.57%	0.40%	98.76%	0.44%	98.83%	0.52%
109	80.53%	0.65%	89.50%	0.50%	95.19%	0.41%	97.34%	0.44%	98.11%	0.42%	98.37%	0.46%	98.45%	0.42%
142	80.17%	0.58%	88.97%	0.40%	94.69%	0.39%	96.83%	0.45%	97.69%	0.41%	97.85%	0.38%	97.99%	0.38%
220	80.96%	0.55%	89.18%	0.45%	94.63%	0.36%	96.78%	0.35%	97.45%	0.33%	97.67%	0.39%	97.76%	0.33%

reconstruction results. In the test case of 8-mode, PhyMoNet was optimized for 1000 epochs. Compared with the results of synthetic data, the performance decreased by less than 3%. It should not be overlooked that MTNet's results are not as good as Phymonet's, but it is more stable with a standard deviation of 0.12%. This is mainly due to the data-driven mechanism. Physical drive-based PhyMoNet may vary slightly with each training.

Table S4: Performance of PhyMoNet on experimental data on different mode cases.

Modes	Model	$\bar{\Gamma}$	σ_Γ
8	VGG[6]	94.38%	2.80%
	MTNet[5]	96.09%	0.12%
	PhyMoNet	97.28%	0.36%
65	PhyMoNet	98.01%	0.30%
5796	PhyMoNet	98.77%	0.33%

Another setup for manipulating the input light of an active MMF amplifier is illuminated in Fig. 5. A 1064 nm CW single-frequency laser is preamplified to 10 Watts. An SLM is used to shape the wavefront of the expanded laser source to generate tailored input of an MMF amplifier. Each pump diode provides up to 10 W power. The output end of the gain fiber is coupled and split into 2 parts, one for the measurement

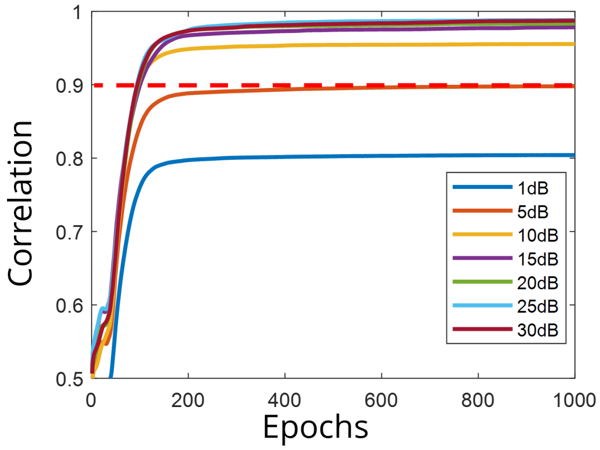

Fig. S3: Training process of synthetic data with noise. X-axis: training epochs; Y-axis: average correlation coefficient for 100 test samples. The figure illustrates the training results under various levels of noise from 1 dB to 30 dB for 55-mode.

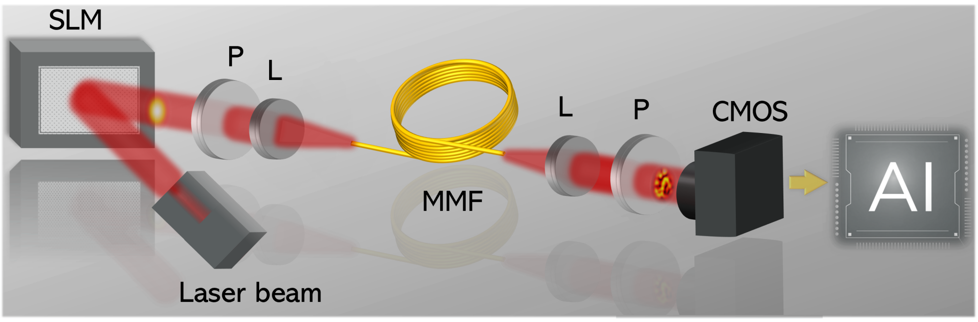

Fig. S4: Scheme of the experimental setup. L: lens; P: polarizer; SLM: spatial light modulator.

of power using a power meter, and another one is imaged on a camera to capture the near-field beam profile. 2 beam splitters and mirrors are used to split and orientate the reference beam for performing MD using digital holography (DH).

MD of the amplifier has been investigated both with and without gain. When the gain unit is off, 200 random phases are displayed on the SLM in sequence. On the CMOS camera, the distal light field distribution interferes with the reference beam. In the investigations shown, the amplitude distribution reconstructed from the hologram is also used as the input for the PhyMoNet ensuring a fair comparison. Captured images need to be properly cropped. Here, the resolution of the natively cropped image is 128x128 which is used further in the test of PhyMoNet. Similar measurements were done when the amplifier pumps to 10 W amplified signal power. Fig. S6 shows

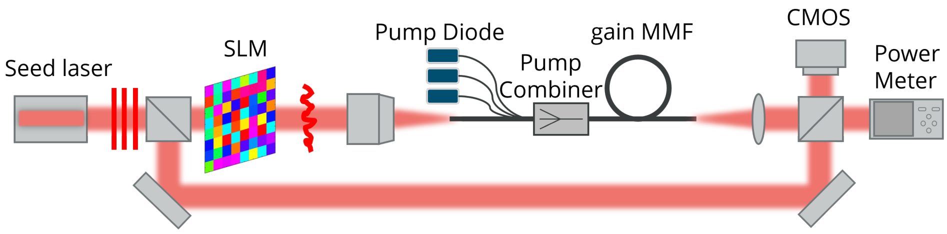

Fig. S5: (a) Scheme of the experimental setup.

examples of reconstructed mode fields from MD and corresponding ground truths. The comparison between the reconstructions shows that the decomposition accuracy of the neural network is slightly better than that of the DH. This is mainly due to the imperfect optical experimental bringing noise into the data.

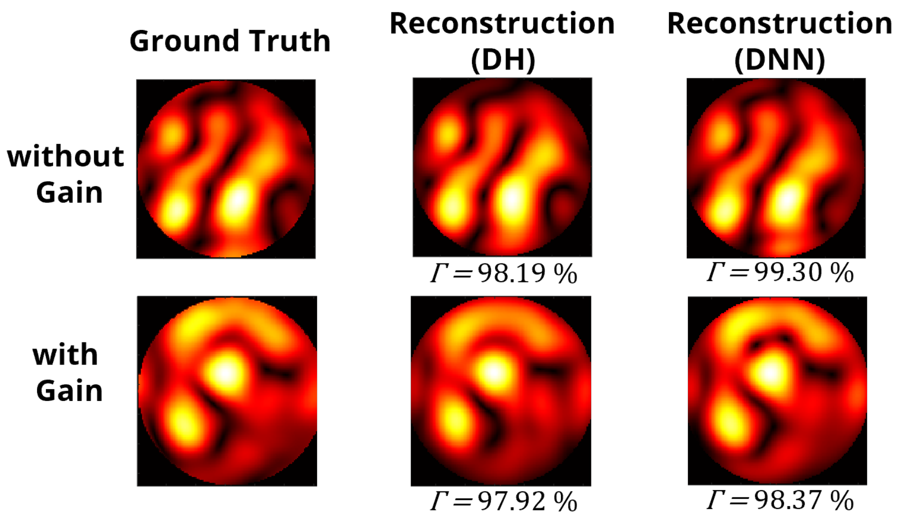

Fig. S6: Comparison between the DH-based MD and PhyMoNet-based MD on experimental data from the MMF amplifier.

References

[1] Caravaca-Aguirre, A.M., Niv, E., Conkey, D.B., Piestun, R.: Real-time resilient focusing through a bending multimode fiber. Optics express **21**(10), 12881–12887 (2013)

[2] Guo, K., Zeng, S., Yu, J., Wang, Y., Yang, H.: [dl] a survey of fpga-based neural network inference accelerators. ACM Transactions on Reconfigurable Technology and Systems (TRETS) **12**(1), 1–26 (2019)

[3] An, Y., Huang, L., Li, J., Leng, J., Yang, L., Zhou, P.: Learning to decompose the modes in few-mode fibers with deep convolutional neural network. Optics express **27**(7), 10127–10137 (2019)

[4] Rothe, S., Zhang, Q., Koukourakis, N., Czarske, J.: Intensity-only mode decomposition on multimode fibers using a densely connected convolutional network. Journal of Lightwave Technology **39**(6), 1672–1679 (2021)

[5] Zhang, Q., Rothe, S., Koukourakis, N., Czarske, J.: Learning the matrix of few-mode fibers for high-fidelity spatial mode transmission. APL Photonics **7**(6), 066104 (2022)

[6] Rothe, S., Zhang, Q., Koukourakis, N., Czarske, J.W.: Deep learning for computational mode decomposition in optical fibers. Applied Sciences **10**(4), 1367 (2020)