nature machine intelligence

Article

https://doi.org/10.1038/s42256-023-00704-7

Self-supervised learning of hologram reconstruction using physics consistency

In the format provided by the authors and unedited

Contents

Supplementary Note 1: Comparison of GedankenNet models trained with different sample-to-sensor axial positions

We investigated the relationship between the reconstruction performance of GedankenNet and the axial positions (sample-to-sensor distances) of the input holograms. In addition to the GedankenNet model with $(z_1, z_2) = (300, 375)\mu\text{m}$ shown in Fig. 2 of the main text, we separately trained two new models with different sample-to-sensor distances, $(z_1, z_2) = (350, 425)\mu\text{m}$ and $(z_1, z_2) = (575, 650)\mu\text{m}$. The other training details remained identical to the model with $(z_1, z_2) = (300, 375)\mu\text{m}$. After the training, the three models were tested on the same field of views (FOVs) of experimental holograms captured at their corresponding axial positions. Extended Data Figure 2(a) reports the reconstruction performance of the three models on Pap smears. With increasing sample-to-sensor distances, one can notice that the reconstruction quality in terms of the structural similarity index (SSIM) and enhanced correlation coefficient (ECC) relatively deteriorated, which mainly results from the reduced signal-to-noise ratio (SNR) of the input holograms over larger sample-to-sensor distances and the enlarged diffraction patterns leaking out of the sensor active area. Extended Data Figure 2(b) further confirms the same conclusion on lung tissue sections.

Supplementary Note 2: Comparison of GedankenNet with the deep image prior (DIP)

In this experiment, we compared GedankenNet against another popular deep neural network-based algorithm for solving inverse problems, i.e., deep image prior (DIP)[1]. DIP-based approaches utilize a deep neural network as a function approximator to solve inverse problems without the supervision of a ground truth image[2,3]. Here we adapted the same U-Net architecture and Adam optimizer as in a previous study[3] of DIP for holographic phase retrieval. DIP took two experimental input holograms of a Pap smear sample captured approximately at 300μm and 375μm, shown as FOV 1 in Extended Data Fig. 3(a). After the iterative optimization of the network's parameters (~10K iterations), DIP was able to converge and retrieve the sample complex field, as shown in Extended Data Fig. 3(a); however, it resulted in a lower ECC value and worse reconstruction quality with noticeable artifacts compared to the outputs of MHPR[4] and GedankenNet, shown in Extended Data Fig. 3(a) and (b), respectively. Once optimized on a single FOV, DIP could not directly generalize to other FOVs (Pap smear FOV 2 and lung FOV 1) of the same sample type or a different sample type as illustrated in Extended Data Fig. 3(a).

Supplementary Note 3: Superior generalization of self-supervised learning over supervised learning

To further showcase the advantages of the self-supervised learning scheme, we trained GedankenNet and a supervised learning model (FIN) on the same artificial hologram datasets, and compared their external generalization performance in Extended Data Fig. 4. In this new analysis, we compared GedankenNet and the supervised learning model FIN that were trained with the same artificial hologram datasets generated from random synthetic images (Extended Data Fig. 4a) or natural images from COCO dataset (Extended Data Fig. 4b). The blind testing of these models used experimental holograms of Pap smear samples and lung tissue sections. The results of this comparison (summarized in Extended Data Fig. 4) reveal that the self-supervised learning scheme consistently achieved better reconstruction accuracy and enhanced ECC scores over the supervised learning scheme, further highlighting the superior external generalization of GedankenNet to experimental holograms of new types of samples.

In Extended Data Fig. 4(a), the two models ($M = 2$) were trained on the same artificial hologram dataset generated from random, synthetic images, and were blindly tested on experimental holograms of Pap smear and human lung tissue sections. Both models were able to generalize on experimental holograms and match the ground truth object fields shown in Extended Data Fig. 4(e), although they had never seen the distribution of the experimental data in the training phase. However, the outputs of FIN have more noise and background artifacts, which are suppressed in the outputs of GedankenNet. This conclusion can be further confirmed by (1) the visualization results shown in Extended Data Fig. 4(b), where the two models were trained on the same artificial hologram dataset generated from COCO natural images[5], and (2) the mean ECC values shown in Extended Data Fig. 4(c, d), corresponding to the generalization outputs in (a, b),

respectively. ECC values were averaged on Pap smear and lung testing datasets using 49 and 94 unique FOVs, respectively.

Supplementary Note 4: GedankenNet trained with artificial complex fields with correlated amplitude and phase channels

One contributing factor to the superior generalization of GedankenNet is the Gedanken training dataset, which consists of simulated holograms of artificial random complex fields. To further illustrate the relationship between the training dataset composed of artificial random images and the generalization performance of GedankenNet, we compared the standard GedankenNet model reported in the main text against a new GedankenNet model trained on artificial random complex fields with *correlated* amplitude and phase channels. The standard GedankenNet model was the same model used in Fig. 3 of the main text, and it was trained on simulated holograms generated using identically and independently distributed (i.i.d.) artificial random images that served as the amplitude and phase channels of the synthetic complex fields (see Extended Data Fig. 5(b)). In comparison, the newly trained GedankenNet model was based on the same artificial image dataset; however, the simulated holograms were generated from synthetic complex fields that used the same artificial image as its amplitude and phase channel (Extended Data Fig. 5(b)). After convergence, the two models were externally tested on experimental holograms of human lung tissue sections and Pap smears.

As summarized in Extended Data Fig. 5, this new GedankenNet model trained with correlated amplitude and phase images generalized relatively worse on the same external test datasets compared to the original GedankenNet model that did not use any correlation between the amplitude and phase channels. Extended Data Figure 5(a) visualizes the output fields of the two models, and Extended Data Fig. 5(c) quantitatively evaluates their reconstruction quality. The standard GedankenNet model demonstrated better generalization over the model trained with correlated field amplitudes and phases. These results further confirm that the large data

variations in the phase and amplitude channels of the artificially generated random training images greatly contribute to the superior generalization of GedankenNet models.

Supplementary Note 5: GedankenNet's resilience to random perturbations

During the training phase of GedankenNet, the physical forward model is given to the network as part of the *Gedankenexperiments*. However, perturbations, i.e., the mismatch between the *a priori* forward model and the *a posteriori* model in the experiments, could impact the performance of the learned GedankenNet. These sources of perturbations often include: (1) the measurement noise ϵ, (2) the modeling error of the sampling function f and (3) the error of the transfer function H. The first two sources can come from a combination of factors, e.g., thermal and shot noise, sensor nonlinearity, aberrations, etc., and can be properly handled by setting regularization terms in the loss function, e.g., the total variation (TV) loss[6]. The last source of perturbations may result from the assumptions when establishing the forward model and errors in the key parameters of the holographic imaging system, e.g., the sample-to-sensor distances, the illumination wavelength, the pixel pitch, etc. Through the self-supervised learning process, GedankenNet intrinsically acquired robustness to various types of random perturbations in the physical forward model (see e.g., Fig. 5) and implicitly learned the physics of wave propagation. Furthermore, in this Supplementary Note we will demonstrate that the GedankenNet framework could adapt to more complicated random, unknown perturbations, including the pixel pitch, SNR, axial distance and the illumination wavelength. These observations align with earlier reports, which showed self-supervised models to be more robust to adversarial attacks and input image/data corruptions and exceed the performance of fully supervised models on near-distribution outliers[7,8].

Firstly, in addition to the analyses reported in Fig. 5, where the input holograms were defocused by an axial distance of Δz, we report in Extended Data Fig. 6 and Supplementary Note 5.1 the resilience of GedankenNet models to different experimental input holograms with various pixel

pitches that are larger compared to the training pixel pitch. In these results, GedankenNet simultaneously implemented hologram reconstruction and pixel super-resolution without any fine-tuning or retraining of its model, which only utilized artificial random training images at a single pixel pitch (see the Extended Data Fig. 6). Secondly, to further explore the robustness of GedankenNet to variations in other physical features of the acquired holograms of interest, in Supplementary Note 5.2 and Extended Data Fig. 7, we studied the impact of the SNR of the input holograms on the image reconstruction quality, and compared GedankenNet's blind inference results against supervised models, further confirming its superiority and robustness to different sources of noise or perturbations.

Next, in Supplementary Note 5.3 and Extended Data Fig. 9, we further enhanced the robustness of GedankenNet to axial defocusing using phase-only object priors, and introduced GedankenNet-*Phase* that achieved both autofocusing and hologram reconstruction within its training axial distance range. Lastly, following the concept of GedankenNet-*Phase*, we extended its resilience to unknown changes in the illumination wavelength, and created an additional model, GedankenNet-*Phaseλ*. GedankenNet-*Phaseλ* is showcased in Supplementary Note 5.4 and Extended Data Fig. 10.

Supplementary Note 5.1: Generalization of GedankenNet to various hologram pixel pitches

Here, we report the intrinsic robustness of GedankenNet (trained on simulated random holograms at a single, fixed pixel pitch) to input holograms with different pixel pitches; we show that a trained GedankenNet model could achieve pixel super-resolution on low-resolution (LR) input holograms without any retraining or fine-tuning of its parameters. Extended Data Figure 6(a) illustrates the process of obtaining LR input holograms from high-resolution (HR)

holograms through k-fold pixel binning. The effective pixel size of the resulting LR holograms is kp, where $p = 0.3733$ μm is the pixel pitch of the HR holograms. We simply upsampled LR holograms through k-fold bilinear interpolation and then fed the interpolated LR holograms to the trained GedankenNet model (the same model as in Fig. 3 of the main text). Our analyses revealed that a trained GedankenNet can adapt to LR input holograms with larger pixel pitches (never used in the training phase) and simultaneously implement hologram reconstruction and pixel super-resolution with minimal performance loss. Extended Data Figure 6(b) quantitatively assessed the reconstruction quality of GedankenNet for various pixel sizes at the input LR holograms, evaluated on a test set of 94 unique FOVs. The robustness of GedankenNet to hologram pixel pitch variations can also be visually confirmed through Extended Data Fig. 6(c).

Supplementary Note 5.2: Robustness of GedankenNet to the signal-to-noise ratio of the input holograms

The self-supervised experiment-free learning scheme of GedankenNet also brings robustness to perturbations caused by the imaging hardware, such as e.g., limited SNR or optical aberrations. In Extended Data Fig. 7, we validated this robustness of GedankenNet to data perturbations, modeled as random measurement noise in the forward imaging model. For this analysis, we trained GedankenNet and a supervised deep neural network model (FIN) using the artificial hologram dataset generated from random synthetic images, the same as before. The trained models were then tested on artificial holograms generated from natural macro-scale images (COCO dataset) with different levels of white Gaussian noise added, as shown in Extended Data Fig. 7(b). Extended Data Figure 7(c) illustrates the hologram reconstruction results of both models (GedankenNet vs. supervised learning-based FIN); Extended Data Figure 7(e) further

compares the root mean square error (RMSE) values for the amplitude and phase channels, as well as the ECC values of these reconstructions. The results revealed that GedankenNet hologram reconstruction quality relatively degraded with input holograms having more noise (lower SNR), but its performance was still superior compared to the supervised learning-based FIN model trained on the same artificial hologram dataset. Additionally, we tested another supervised FIN model (trained on experimental holograms of lung tissue sections, i.e., the same model used in Fig. 4) on the same simulated test holograms and summarized its inference outputs in Extended Data Fig. 7(d). Compared to the other two models trained on the artificial hologram dataset, the experimental-data-driven supervised FIN model exhibited inferior reconstruction quality and lower ECC values for unseen object features, demonstrating significantly poorer external generalization. In general, supervised hologram reconstruction models overfit to perturbations or noise present in their training, which makes them more susceptible to unseen perturbations from other sources or lower SNR input holograms compared to their training data. This superior generalization of GedankenNet to poor SNR input holograms stems from its self-supervised learning using the physical-consistency loss and artificial holograms, as opposed to directly learning a mapping between the input and the target image domains through experimental data.

Supplementary Note 5.3: GedankenNet-*Phase*

This robustness of GedankenNet and its external generalization performance reported in Fig. 5 can be further improved using some object priors, such as the assumption of phase-only objects. Different from the earlier models, here the GedankenNet framework uses phase-only artificial random complex fields (with unit amplitude) during its training. Apart from this phase-only

object assumption, there is *no* prior knowledge about the sample types to be imaged, and therefore the self-supervised learning model was trained using the same physics-consistency loss and artificial phase-only random objects that were synthetically generated without any experiments or resemblance to real-world samples. This new model, which we termed GedankenNet-*Phase*, exhibits enhanced adaptability to random, unknown perturbations in the forward optical model. Following a similar experimental-data-free training protocol as used in earlier GedankenNet models, GedankenNet-*Phase* was trained on $M = 2$ simulated holograms of artificial phase-only objects; however, these holograms were virtually located at independent random axial positions between 275 μm and 400 μm during the training, aiming to achieve both hologram reconstruction and autofocusing using self-supervised learning based on the same physics-consistency loss (see the Methods and Extended Data Fig. 8 for details on the training process and the architecture of GedankenNet-*Phase*). After its training with artificial random data generated without any experiments or resemblance to real-world samples, Extended Data Fig. 9 demonstrates the experimental hologram reconstruction and autofocusing performance of GedankenNet-*Phase* on unstained (label-free) human kidney tissue sections. Extended Data Figure 9(a) visualizes GedankenNet-*Phase* outputs corresponding to $M = 2$ input holograms independently captured at arbitrary and unknown axial positions within $[300, 400]$μm, and Extended Data Fig. 9(b) quantitatively evaluates the reconstruction quality of GedankenNet-*Phase* in terms of phase RMSE with respect to the ground truth over a test set of 98 unique FOVs. These results reveal that GedankenNet-*Phase* successfully achieved both autofocusing and hologram reconstruction within its training axial distance range. As expected, the reconstruction quality drops for $z_1 = z_2$ since it corresponds to $M = 1$, deviating from the training of GedankenNet-*Phase,* which used $M = 2$ random input holograms.

Supplementary Note 5.4: GedankenNet-*Phaseλ*

The concept of GedankenNet-*Phase* can be further expanded to bring additional resilience to its reconstructions and achieve broader external generalization for other types of perturbations in the physical forward model, such as unknown changes or shifts/drifts in the illumination wavelength. To showcase this, we created an additional model, which we termed GedankenNet-*Phaseλ* and trained it on $M = 2$ simulated holograms of artificial phase-only random fields, as illustrated in Extended Data Fig. 10(a). GedankenNet-*Phaseλ* was trained using simulated input holograms of artificially generated phase-only complex objects, illuminated by a random wavelength (λ) that is uniformly sampled between λ_1 and λ_2, i.e., $\lambda \in [\lambda_1, \lambda_2]$ nm. We chose $\lambda_1 = 520$ nm and $\lambda_2 = 540$ nm. Following a similar experimental-data-free training protocol as GedankenNet-*Phase*, GedankenNet-*Phaseλ* utilized no prior knowledge/assumptions about the sample fields other than the phase-only object assumption, and was solely trained using the same physics-consistency loss and artificial phase-only random complex objects without any experiments or resemblance to any real-world samples. The sample-to-sensor distances were set as $(z_1, z_2) = (300,375)$μm. The other training details (hyperparameters, network architectures, etc.) were kept the same as in GedankenNet-*Phase*. Typical training holograms under various illumination wavelengths are also shown in Extended Data Fig. 10(a).

After its training with artificial random data, GedankenNet-*Phaseλ* generalized to experimental data of unstained human kidney tissue sections and achieved stable reconstruction performance on holograms acquired at various illumination wavelengths, ranging from 500 nm to 560 nm, without knowing what the illumination wavelength is. Extended Data Figure 10(b,c) visualize and evaluate the reconstructed phase images by GedankenNet-*Phaseλ* on experimental holograms of unstained (label-free) human kidney tissue sections, and compare them against the

results of the standard GedankenNet model (trained with a single illumination wavelength, $\lambda_0 = 530$ nm). GedankenNet-*Phase*λ successfully generalized to experimental holograms of phase-only tissue sections, and generalized to various illumination wavelengths within and beyond its training range. Across the spectral range from $\lambda = 500$ nm to $\lambda = 560$ nm, the output phase images of GedankenNet-*Phase*λ showed consistent success, whereas the results of the standard GedankenNet model (trained with a single wavelength, $\lambda_0 = 530$ nm) displayed blurring artifacts for $\lambda \neq \lambda_0$ and reconstruction accuracy degradation in terms of the ECC values.

These results and analyses reported in Supplementary Note 5 demonstrate the superior robustness of the GedankenNet framework to various sources of perturbations in the physical forward model while also maintaining its broad external generalization performance.

Supplementary Note 6: Study on residual errors during the network training

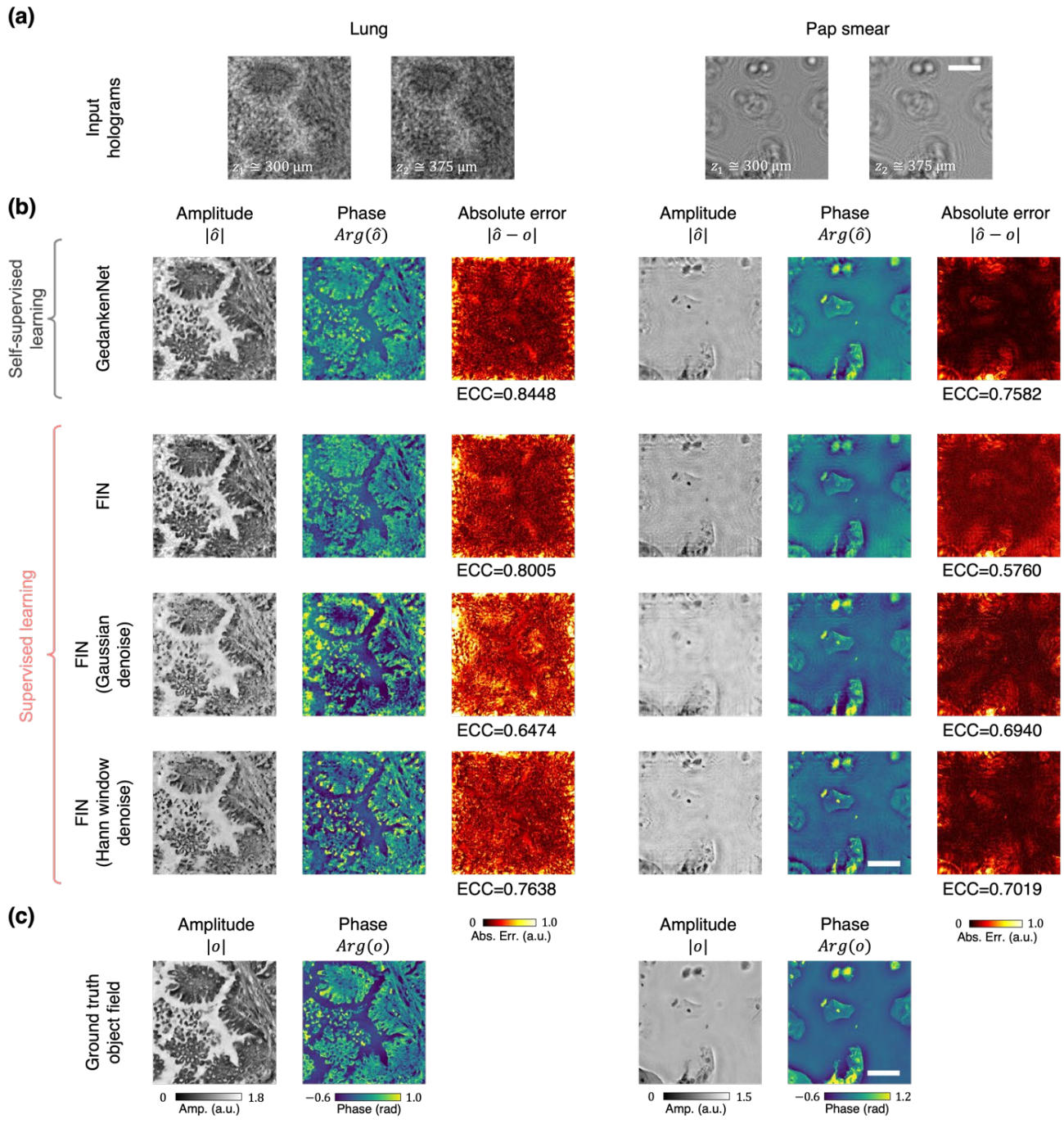
We further investigated the residual errors in the network training, and explored possible methods to improve the performance of supervised learning models. As elucidated in the Discussion, the traditional structural loss functions employed in supervised learning rely on the statistics of the training samples to penalize the residual errors of a supervised model, which could hinder its external generalization on unseen types of test samples. In contrast, GedankenNet applies a physics-consistency loss that universally penalizes the non-physical residual errors, helping us achieve superior external generalization. Considering that the residual errors during the network training are generally non-physical and may exhibit a similar distribution as random noise, we additionally trained two supervised learning models (following the FIN architecture) on the same artificial image dataset as used in GedankenNet's training, and applied two types of image denoisers when calculating the loss function to help filter residual errors. The first model employed a 3×3 Gaussian kernel with 0.5 pixel standard deviation for denoising, and the second model applied a Hann window in the Fourier domain. Denoising using the Hann window in Fourier domain can be expressed as:

$$D(x) = \mathcal{F}^{-1}\{\mathcal{F}\{x\} \cdot w\}$$

Here $x \in \mathbb{R}^{N\times N}$ is the 2D input image, $w \in \mathbb{R}^{N\times N}$ is the 2D Hann window, and $\mathcal{F}, \mathcal{F}^{-1}$ denote 2D Fourier transform and inverse Fourier transform operations, respectively. The other training setups, including the loss function, hyperparameters, network architectures, etc., were kept the same as the standard FIN model (see the Methods for details).

Supplementary Figure 1 summarizes the performance of the two additional FIN models and compares them with the standard FIN and GedankenNet models. All models were trained on the same artificial image dataset, and then externally tested on experimental holograms of human

lung tissue sections and Pap smears, as demonstrated in Supplementary Figure 1(a). Supplementary Figure 1(b) shows the reconstructed fields of the four network models corresponding to these experimental holograms and the resulting absolute errors, defined as the modulus of the difference between the reconstructed and the ground truth complex fields. These results indicate that all the supervised FIN models, with and without image denoising during training, showed inferior external generalization and scored lower ECC values compared to GedankenNet, which further supports the effectiveness of the physics-consistency-based self-supervised learning of GedankenNet framework.



Supplementary Figure 1. External generalization of FIN models trained with image denoisers, and their comparison against the standard FIN and GedankenNet models. (a) Experimental holograms of human lung tissue sections and Pap smears. (b) External generalization of GedankenNet, the standard FIN, FIN models trained with Gaussian denoising and Hann window denoising on the reconstruction of experimental holograms shown in (a). The absolute errors of the reconstructed fields are also shown, and ECC values were calculated for the shown FOV. (c) Ground truth sample fields obtained using M=8 raw holograms of each FOV. Scale bar: 50µm.

References

1. Lempitsky, V., Vedaldi, A. & Ulyanov, D. Deep Image Prior. in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9446–9454 (IEEE, 2018).
2. Bostan, E., Heckel, R., Chen, M., Kellman, M. & Waller, L. Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network. *Optica* **7**, 559 (2020).
3. Wang, F. *et al.* Phase imaging with an untrained neural network. *Light Sci. Appl.* **9**, 77 (2020).
4. Rivenson, Y. *et al.* Sparsity-based multi-height phase recovery in holographic microscopy. *Sci. Rep.* **6**, 37862 (2016).
5. Lin, T.-Y. *et al.* Microsoft COCO: Common Objects in Context. Preprint at http://arxiv.org/abs/1405.0312 (2015).
6. Rudin, L. I., Osher, S. & Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. Nonlinear Phenom.* **60**, 259–268 (1992).
7. Hendrycks, D., Mazeika, M., Kadavath, S. & Song, D. Using Self-Supervised Learning Can Improve Model Robustness and Uncertainty. in *Advances in Neural Information Processing Systems* (eds. Wallach, H. et al.) vol. 32 (Curran Associates, Inc., 2019).
8. Liu, H., HaoChen, J. Z., Gaidon, A. & Ma, T. Self-supervised Learning is More Robust to Dataset Imbalance. Preprint at http://arxiv.org/abs/2110.05025 (2022).