

# Self-supervised learning of hologram reconstruction using physics consistency

Received: 17 September 2022

Luzhe Huang<sup>1,2,3,5</sup>, Hanlong Chen<sup>1,2,3,5</sup>, Tairan Liu<sup>1,2,3</sup> & Aydogan Ozcan<sup>1,2,3,4</sup>✉

Accepted: 6 July 2023

Published online: 7 August 2023

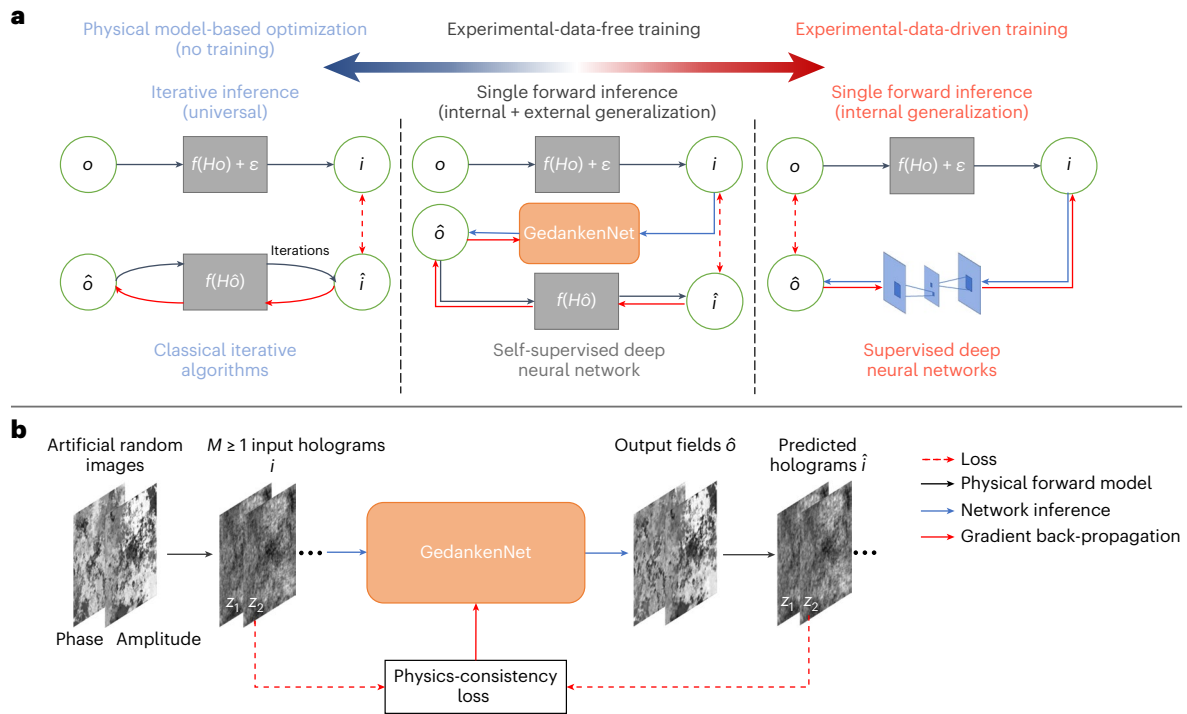
 Check for updates

Existing applications of deep learning in computational imaging and microscopy mostly depend on supervised learning, requiring large-scale, diverse and labelled training data. The acquisition and preparation of such training image datasets is often laborious and costly, leading to limited generalization to new sample types. Here we report a self-supervised learning model, termed GedankenNet, that eliminates the need for labelled or experimental training data, and demonstrate its effectiveness and superior generalization on hologram reconstruction tasks. Without prior knowledge about the sample types, the self-supervised learning model was trained using a physics-consistency loss and artificial random images synthetically generated without any experiments or resemblance to real-world samples. After its self-supervised training, GedankenNet successfully generalized to experimental holograms of unseen biological samples, reconstructing the phase and amplitude images of different types of object using experimentally acquired holograms. Without access to experimental data, knowledge of real samples or their spatial features, GedankenNet achieved complex-valued image reconstructions consistent with the wave equation in free space. The GedankenNet framework also shows resilience to random, unknown perturbations in the physical forward model, including changes in the hologram distances, pixel size and illumination wavelength. This self-supervised learning of image reconstruction creates new opportunities for solving inverse problems in holography, microscopy and computational imaging.

Recent advances in deep learning have revolutionized computational imaging, microscopy and holography-related fields, with applications in biomedical imaging<sup>1</sup>, sensing<sup>2</sup>, diagnostics<sup>3</sup> and three-dimensional (3D) displays<sup>4</sup>, also achieving benchmark results in various image translation and enhancement tasks, for example, super-resolution<sup>5–12</sup>, image denoising<sup>13–16</sup> and virtual staining<sup>17–23</sup>, among others. The flexibility of deep learning models has also facilitated their widespread use in different imaging modalities, including bright-field<sup>24,25</sup> and fluorescence<sup>8,11,12,15,26,27</sup> microscopy. As another important example, digital

holographic microscopy, a label-free imaging technique widely used in biomedical and physical sciences and engineering<sup>28–37</sup>, has also remarkably benefited from deep learning and neural networks<sup>4,38–49</sup>. Convolutional neural networks<sup>38–41,43,45,46,50,51</sup> and recurrent neural networks<sup>47,52</sup> have been used for holographic image reconstruction, presenting unique advantages over classical phase retrieval algorithms, such as using fewer measurements and achieving an extended depth of field. Researchers have also explored deep learning-enabled image analysis<sup>53–58</sup> and transformations<sup>18,19,25,59,60</sup> on holographic images to

<sup>1</sup>Electrical and Computer Engineering Department, University of California Los Angeles, Los Angeles, CA, USA. <sup>2</sup>Bioengineering Department, University of California Los Angeles, Los Angeles, CA, USA. <sup>3</sup>California NanoSystems Institute (CNSI), University of California Los Angeles, Los Angeles, CA, USA. <sup>4</sup>David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. <sup>5</sup>These authors contributed equally: Luzhe Huang, Hanlong Chen. ✉ e-mail: [ozcan@ucla.edu](mailto:ozcan@ucla.edu)



**Fig. 1 | Diagrams of GedankenNet and other existing methods for solving holographic imaging problems. a**, Diagrams of classical iterative hologram reconstruction algorithms, the self-supervised deep neural network (GedankenNet) and existing supervised deep neural networks. **b**, Self-supervised training pipeline of GedankenNet for hologram reconstruction.

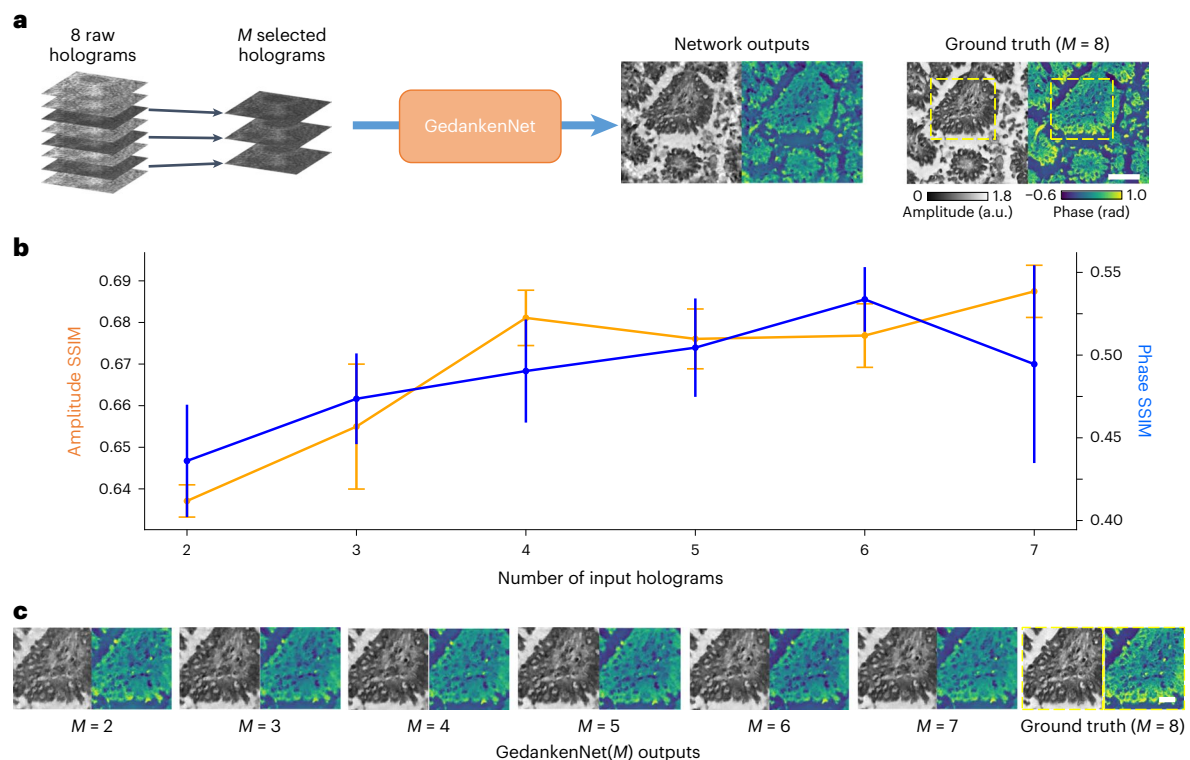
further leverage the quantitative phase information provided by digital holographic microscopy.

In these existing approaches, supervised learning models were utilized, demanding large-scale, high-quality and diverse training datasets (from various sources and types of object) with annotations and/or ground-truth experimental images. For microscopic imaging and holography, in general, such labelled training data can be acquired through classical algorithms that are treated as the ground-truth image reconstruction method<sup>38,39,43,47–49,52</sup>, or through registered image pairs (input versus ground truth) acquired by different imaging modalities<sup>8,17,18,25</sup>. These supervised learning methods require substantial labour, time and cost to acquire, align and pre-process the training images, and potentially introduce inference bias, resulting in limited generalization to new types of object never seen during the training. Generally speaking, existing supervised learning models demonstrated on microscopic imaging and holography tasks are highly dependent on the training image datasets acquired through experiments, which show variations due to the optical hardware, types of specimen and imaging (sample preparation) protocols. Although there have been efforts utilizing unsupervised learning<sup>61–67</sup> and self-supervised learning<sup>16,68–70</sup> to alleviate the reliance on large-scale experimental training data, the need for experimental measurements or sample labels with the same or similar features as the testing samples of interest is not entirely eliminated. Using labelled simulated data for network training is another possible solution; however, generating simulated data distributions to accurately represent the experimental sample distributions can be complicated and requires prior knowledge of the sample features and/or some initial measurements with the imaging set-up of interest<sup>6,10,71–74</sup>. For example, supervised learning-based deep neural networks for hologram reconstruction tasks demonstrated decent internal generalization to new samples of the same type as in the training dataset, while their external generalization to different sample types or imaging hardware was limited<sup>38,46,52</sup>.

A common practice to enhance the imaging performance of a supervised model is to apply transfer learning<sup>52,69,75–77</sup>, which trains the learned model on a subset of the new test data. However, the features learned through supervised transfer learning using a limited training data distribution, for example, specific types of sample, do not necessarily advance external generalization to other types of sample, considering that the sample features and imaging set-up may differ substantially in the blind testing phase. Furthermore, transfer learning requires additional labour and time to collect fresh data from the new testing data distribution and fine-tune the pre-trained model, which might bring practical challenges in different applications.

In addition, deep learning-based solutions for inverse problems in computational imaging generally lack the incorporation of explicit physical models in the training phase; this, in turn, limits the compatibility of the network’s inference with the physical laws that govern the light–matter interactions and wave propagation. Recent studies have demonstrated physics-informed neural networks<sup>70,78–83</sup>, where a physical loss was formulated to train the network in an unsupervised manner to solve partial differential equations. However, physics-informed neural network-based methods that can match (or come close to) the performance of supervised learning methods have not been reported yet for solving inverse problems in computational imaging with successful generalization to new types of sample.

Here we demonstrate a self-supervised learning (SSL)-based deep neural network for zero-shot hologram reconstruction, which is trained without any experimental data or prior knowledge of the types or spatial features of the samples. We term it GedankenNet as the self-supervised training of our network model is based on randomly generated artificial images with no connection or resemblance to real samples at the micro- or macroscale, and therefore the spatial frequencies and the features of these images do not represent any real-world samples and are not related to any experimental set-up. As illustrated in Fig. 1a, the self-supervised learning scheme of GedankenNet adapts



**Fig. 2 | Hologram reconstruction performance of GedankenNet using multiple ( $M$ ) input holograms.** **a**,  $M$  holograms were selected from eight raw holograms as the inputs for GedankenNet. The ground-truth complex field (used only for comparison) was retrieved by MHPR using all the eight raw holograms. Scale bar, 50  $\mu\text{m}$ . **b**, The amplitude and phase SSIM values between the

reconstructed fields of GedankenNet and the ground-truth object fields. SSIM values were averaged on a testing set with 94 unique human lung tissue FOVs, and the SSIM standard deviations were calculated on 4 individual models for each  $M$ . **c**, Zoomed-in regions of the GedankenNet outputs and the ground-truth object fields. Scale bar, 20  $\mu\text{m}$ .

a physics-consistency loss between the input synthetic holograms of random, artificial objects and the numerically predicted holograms calculated using the GedankenNet output complex fields, without any reference to or use of the ground-truth object fields during the learning process. After its training, the self-supervised GedankenNet directly generalizes to experimental holograms of various types of sample even though it never saw any experimental data or used any information regarding the real samples. When blindly tested on experimental holograms of human tissue sections (lung, prostate and salivary gland tissue) and Pap smears, GedankenNet achieved better image reconstruction accuracy compared with supervised learning models using the same training datasets. We further demonstrated that GedankenNet can be widely applied to other training datasets, including simulations and experimental datasets, and achieves superior zero-shot generalization to unseen data distributions over supervised learning-based models.

As GedankenNet's self-supervised learning is based on a physics-consistency loss, its inference and the resulting output complex fields are compatible with the Maxwell's equations and accurately reflect the physical wave propagation phenomenon in free space. By testing GedankenNet with experimental input holograms captured at shifted (unknown) axial positions, we showed that GedankenNet does not hallucinate and the object field at the sample plane can be accurately retrieved through wave propagation of the GedankenNet output field, without the need for retraining or fine-tuning its parameters. These results indicate that in addition to generalizing to experimental holograms of unseen sample types without seeing any experimental data or real object features, GedankenNet also implicitly acquired the physical information of wave propagation in free space and gained robustness towards defocused holograms or changes in the pixel size

through the same self-supervised learning process. Furthermore, for phase-only objects (such as thin label-free samples), the GedankenNet framework also shows resilience to random unknown perturbations in the imaging system, including arbitrary shifts of the sample-to-sensor distances and unknown changes in the illumination wavelength, all of which make its generalization even broader without the need for any experimental data or ground-truth labels.

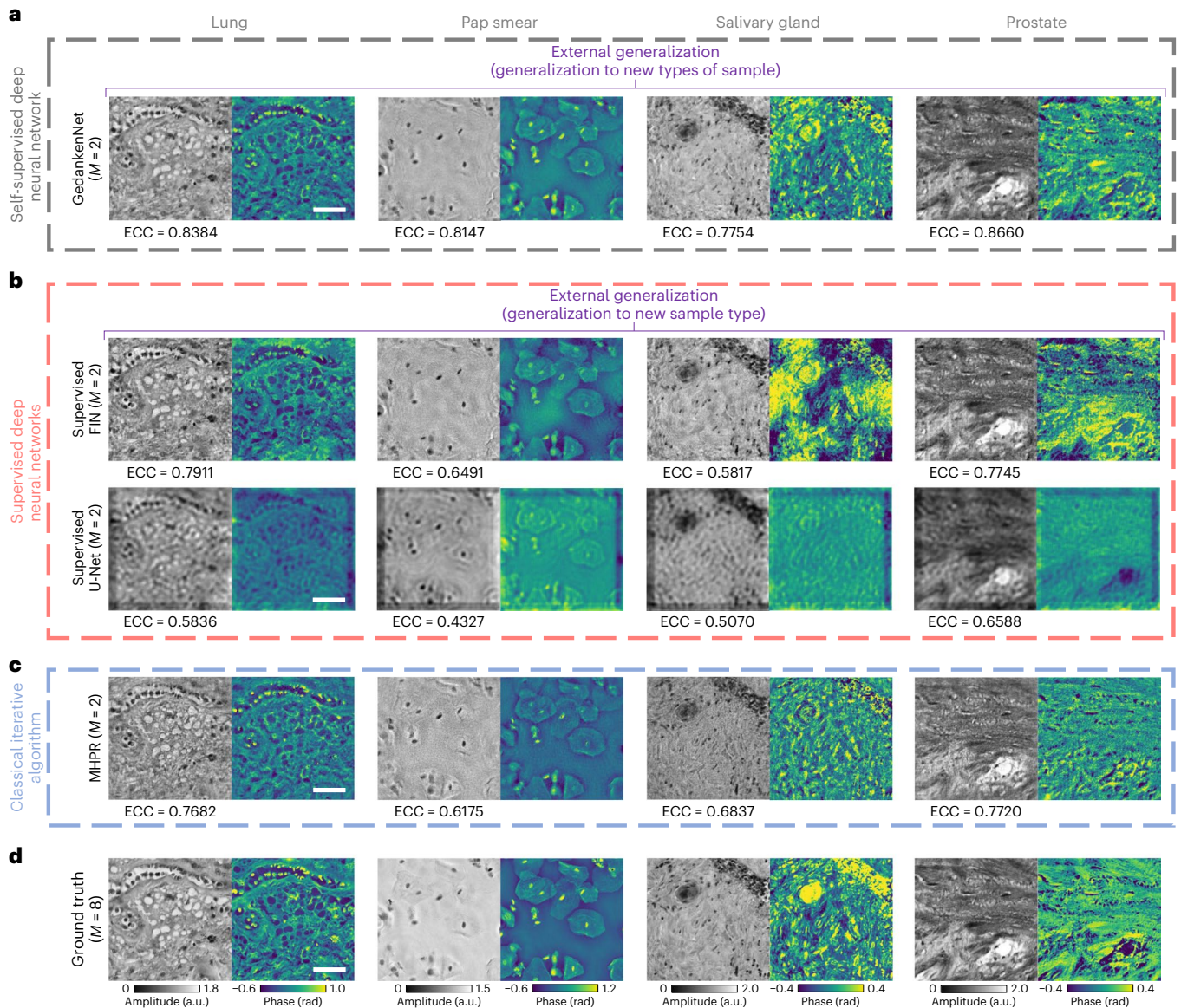
The success of GedankenNet eliminates three major challenges in existing deep learning-based holographic imaging approaches: (1) the need for large-scale, diverse and labelled training data, (2) the limited generalization to unseen sample types or shifted input data distributions, and (3) the lack of an interpretable connection and compatibility between the physical laws and models and the trained deep neural network. This work introduces a promising and powerful alternative to a wide variety of supervised learning-based methods that are currently applied in various microscopy, holography and computational imaging tasks.

## Self-supervised learning of hologram reconstruction

The hologram reconstruction task, in general, can be formulated as an inverse problem<sup>44</sup>:

$$\hat{o} = \arg \min_o L(H(o), i) + R(o)$$

where  $i \in \mathbb{R}^{MN^2}$  represents the vectorized  $M$  measured holograms, each of which is of dimension  $N \times N$  and  $o \in \mathbb{C}^{N^2}$  is the vectorized object complex field.  $H(\cdot)$  is the forward imaging model,  $L(\cdot)$  is the loss function and  $R(\cdot)$  is the regularization term. Under spatially and temporally coherent illumination of a thin sample,  $H(\cdot)$  can be simplified as:



**Fig. 3 | External generalization of GedankenNet on human tissue sections and Pap smears, and comparison with existing supervised learning models and MHPR.** **a**, External generalization results of GedankenNet on human lung, salivary gland, prostate and Pap smear holograms. **b**, External generalization results of supervised learning methods on the same test datasets. The supervised

models were trained on the same simulated hologram dataset as GedankenNet used. **c**, MHPR reconstruction results using the same  $M = 2$  input holograms. **d**, Ground-truth object fields retrieved using eight raw holograms of each FOV. Scale bar, 50  $\mu\text{m}$ .

$$H(o) = f(Ho) + \epsilon$$

where  $H \in \mathbb{C}^{MN^2 \times MN^2}$  is the free-space transformation matrix<sup>44,84</sup>,  $\epsilon \in \mathbb{R}^{MN^2}$  represents random detection noise and  $f(\cdot)$  refers to the (opto-electronic) sensor-array sampling function, which records the intensity of the optical field.

Different schemes for solving holographic imaging inverse problems are summarized in Fig. 1. Existing methods for generalizable hologram reconstruction can be mainly classified into two categories, as shown in Fig. 1a: (1) iterative phase retrieval algorithms based on the physical forward model and iterative error-reduction; and (2) supervised deep learning-based inference methods that learn from training image pairs of input holograms  $i$  and the ground-truth object fields  $o$ . Similar to the iterative phase recovery algorithms listed under category 1, deep neural networks were also used to provide iterative approximations to the object field from a batch of hologram(s); however, these network models were iteratively optimized for each hologram batch

separately, and cannot generalize to reconstruct holograms of other objects once they are optimized<sup>70,79,81</sup> (Supplementary Note 2 and Extended Data Fig. 3).

Different from existing learning-based approaches, instead of directly comparing the output complex fields ( $\hat{o}$ ) and the ground-truth object complex fields ( $o$ ), GedankenNet infers the predicted holograms  $\hat{i}$  from its output complex fields  $\hat{o}$  using a deterministic physical forward model, and directly compares  $\hat{i}$  with  $i$ . Without the need to know the ground-truth object fields  $o$ , this forward model–network cycle establishes a physics-consistency loss ( $L_{\text{physics-consistency}}$ ) for gradient back-propagation and network parameter updates, which is defined as:

$$L_{\text{physics-consistency}}(\hat{i}, i) = \alpha L_{\text{FDMAE}}(\hat{i}, i) + \beta L_{\text{MSE}}(\hat{i}, i),$$

where  $L_{\text{FDMAE}}$  and  $L_{\text{MSE}}$  are the Fourier domain mean absolute error (FDMAE) and the mean square error (MSE), respectively, calculated between the input holograms  $i$  and the predicted holograms  $\hat{i}$ .  $\alpha$  and

**Table 1 | Holographic image inference time (for 1 mm<sup>2</sup> sample area) for GedankenNet, supervised learning models and MHPR**

| Model  | GedankenNet ( $M=2$ ) | GedankenNet-Phase ( $M=2$ ) | FIN ( $M=2$ ) | U-Net ( $M=2$ ) | MHPR ( $M=2$ ) |
|--|-----------------------|-----------------------------|---------------|-----------------|----------------|
| Inference time without parallelization (s mm <sup>-2</sup> ) | 0.21                  | 1.36                        | 0.20          | 0.084           | 4.1            |
| Inference time with parallelization (s mm <sup>-2</sup> )    | 0.032                 | 0.53                        | 0.034         | 0.0062          | NA             |
| Number of trainable parameters (millions)                    | 39.4                  | 5.7                         | 11.5          | 31.0            | NA             |

NA, Not Available.

$\beta$  refer to the corresponding weights of each term (see Methods for the training and implementation details). The network architecture of GedankenNet is also detailed in Methods and Extended Data Fig. 1.

As emphasized in Fig. 1, GedankenNet eliminates the need for experimental, labelled training data and thus presents unique advantages over existing methods. The training dataset of GedankenNet only consists of artificial holograms generated from random images (with no connection or resemblance to real-world samples), which serve as the amplitude and phase channels of the object field (Methods and Fig. 1b). After its self-supervised training using artificial images without any experimental data or real-world specimens, GedankenNet can be directly used to reconstruct experimental holograms of various microscopic specimens, including, for example, densely connected tissue samples and Pap smears. GedankenNet also provides considerably faster reconstructions in a single forward inference without the need for numerical iterations, transfer learning or fine-tuning of its parameters on new testing samples.

### Superior generalization of GedankenNet

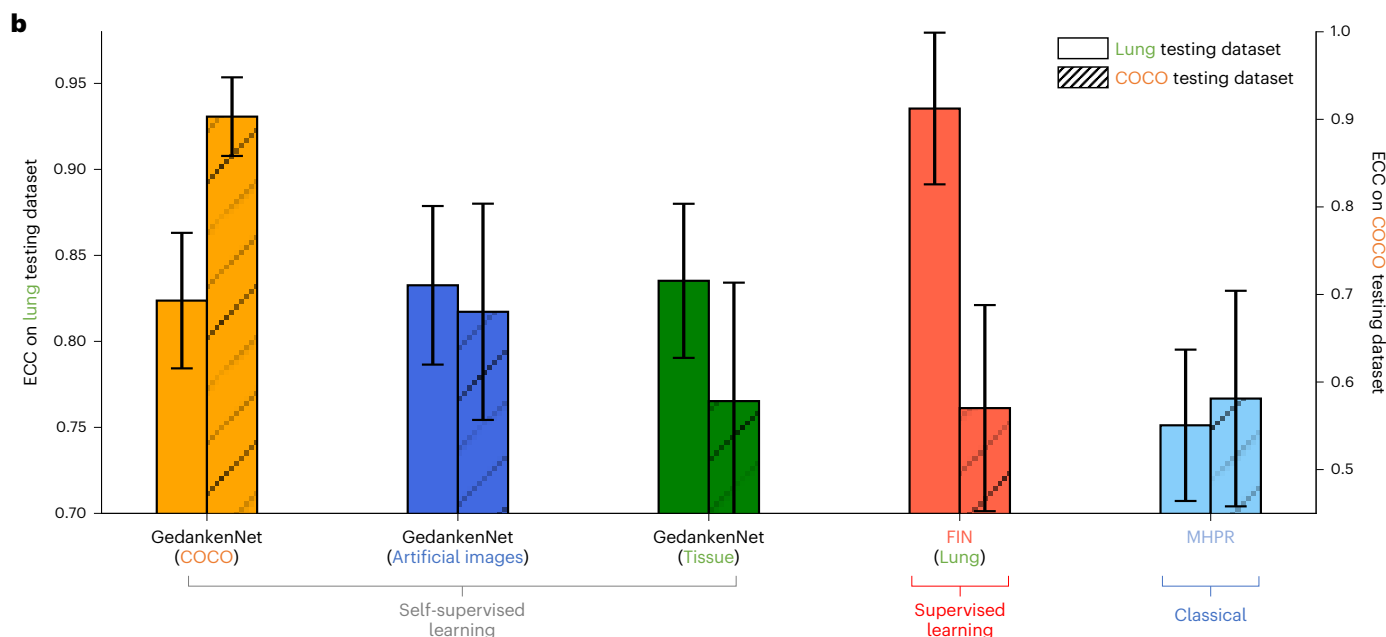
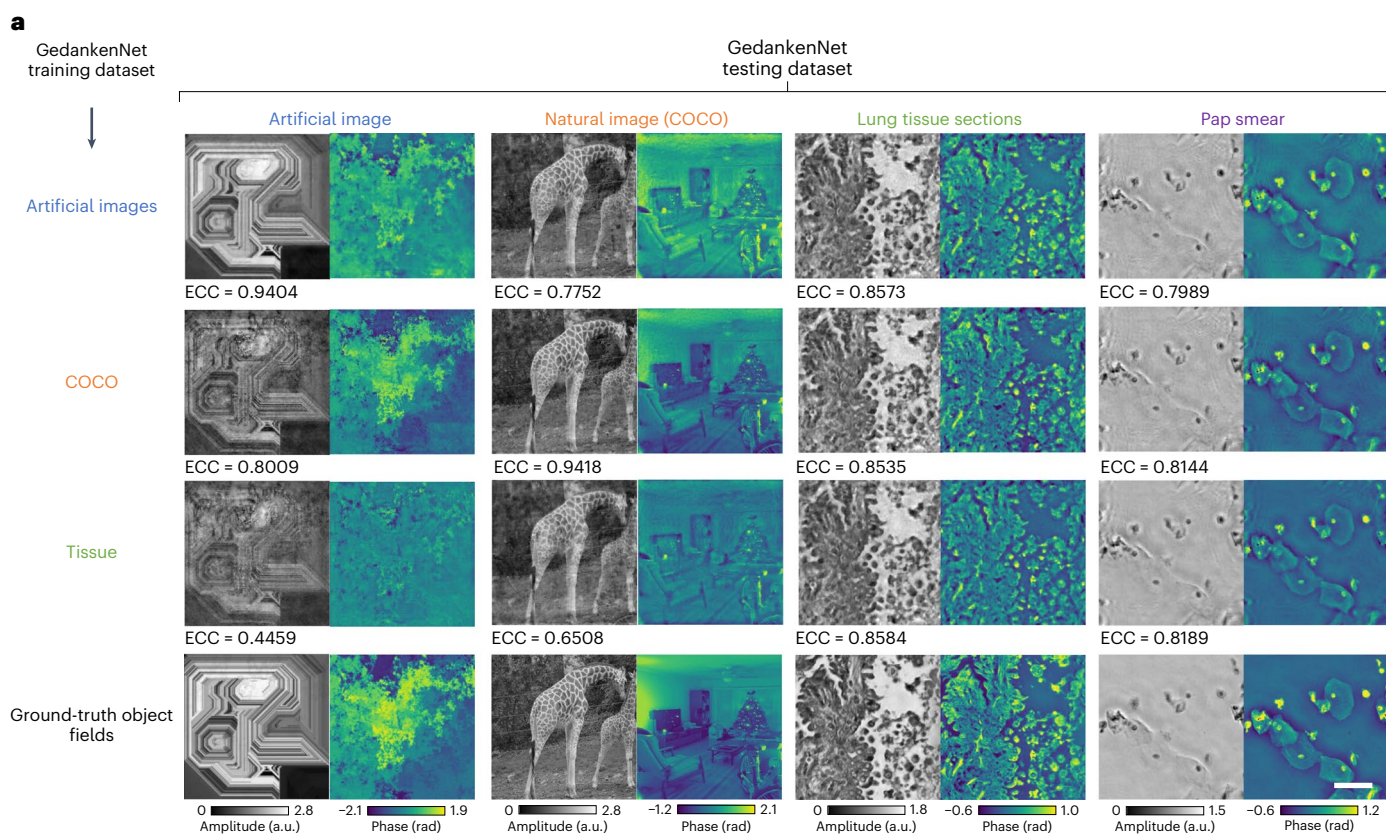
To demonstrate the unique features of GedankenNet, we trained a series of self-supervised network models that take multiple input holograms ( $M$  ranging from 2 to 7), following the training process introduced in Fig. 1. Each GedankenNet model for a different  $M$  value was trained using artificial holograms generated from random synthetic images based on  $M$  different planes with designated sample-to-sensor distances  $z_i$ ,  $i = 1, 2, \dots, M$ . In the blind testing phase illustrated in Fig. 2a,  $M$  experimental holograms of human lung tissue sections were captured by a lens-free in-line holographic microscope (see Extended Data Fig. 1b and Methods for experimental details). We tested all the self-supervised GedankenNet models on 94 non-overlapping fields-of-view (FOVs) of tissue sections and quantified the image reconstruction quality in terms of the amplitude and phase structural similarity index measure (SSIM) values with respect to the ground-truth object fields (Fig. 2b). The ground-truth fields were retrieved by the multi-height phase retrieval (MHPR)<sup>85–87</sup> algorithm using  $M = 8$  raw holograms of each FOV. Our results indicate that all the GedankenNet models were able to reconstruct the sample fields with high fidelity even though they were trained using random, artificial images without any experimental data (Fig. 2c). In addition, Fig. 2 shows that the reconstruction quality of GedankenNet models increased with increasing number of input holograms  $M$ , which inherently points to a general trade-off between the image reconstruction quality and system throughput; depending on the level of reconstruction quality desired and the imaging application needs,  $M$  can be accordingly selected and optimized. In addition to the number of input holograms, we investigated the relationship between the sample-to-sensor distances and the reconstruction quality of GedankenNet (Extended Data Fig. 2 and Supplementary Note 1). Due to the reduced signal-to-noise ratio of the experimental in-line holograms acquired at large sample-to-sensor (axial) distances, GedankenNet models trained with larger sample-to-sensor distances show a relatively reduced reconstruction quality compared with the GedankenNet models trained with smaller axial distances.

We also compared the generalization performance of self-supervised GedankenNet models against other supervised learning models and iterative phase recovery algorithms using experimental

holograms of various types of human tissue section and Pap smears (Fig. 3). Although only seeing artificial holograms of random images in the training phase, GedankenNet ( $M = 2$ ) was able to directly generalize to experimental holograms of Pap smears and human lung, salivary gland and prostate tissue sections. For comparison, we trained two supervised learning models using the same artificial image dataset, including the Fourier Imager Network (FIN)<sup>48</sup> and a modified U-Net<sup>88</sup> architecture (Methods). These supervised models were tested on the same experimental holograms to analyse their external generalization performance. Compared with these supervised learning methods, GedankenNet exhibited superior external generalization on all four types of sample (lung, salivary gland and prostate tissue sections and Pap smears), scoring higher enhanced correlation coefficient (ECC) values (Methods). A second comparative analysis was performed against a classical iterative phase recovery method, that is, MHPR<sup>85–87</sup>. GedankenNet inferred the object fields with less noise and higher image fidelity compared with MHPR ( $M = 2$ ) that used the same input holograms (Fig. 3a,c). In addition, we compared GedankenNet image reconstruction results against deep image prior-based approaches<sup>70,79,81,89</sup>, also confirming its superior performance (Extended Data Fig. 3 and Supplementary Note 2).

The inference time of each of these hologram reconstruction algorithms is summarized in Table 1, which indicates that GedankenNet accelerated the image reconstruction process by ~128 times compared with MHPR ( $M = 2$ ). These holographic imaging experiments and resulting analyses successfully demonstrate GedankenNet's unparalleled zero-shot generalization to experimental holograms of unknown, new types of sample without any prior knowledge about the samples or the use of experimental training data or labels.

GedankenNet's strong external generalization is due to its self-supervised learning scheme that employs the physics-consistency loss, which is further validated by the additional comparisons we performed between self-supervised learning and supervised learning schemes (Extended Data Fig. 4 and Supplementary Note 3). In addition to GedankenNet's superior external generalization (from artificial random images to experimental holographic data), this framework can also be applied to other training datasets. To showcase this, we trained three GedankenNet models using (1) the artificial hologram dataset generated from random images, same as before; (2) a new artificial hologram dataset generated from a natural image dataset (common objects in context, COCO)<sup>90</sup>; and (3) an experimental hologram dataset of human tissue sections (see Methods for dataset preparation). Each one of these training datasets had ~100,000 training image pairs with  $M = 2$ ,  $z_1 = 300 \mu\text{m}$  and  $z_2 = 375 \mu\text{m}$ . As shown in Fig. 4, these three individually trained GedankenNet models were tested on four testing datasets, including artificial holograms of (1) random synthetic images and (2) natural images as well as experimental holograms of (3) lung tissue sections and (4) Pap smears. Our results reveal that all the self-supervised GedankenNet models showed very good reconstruction quality for both internal and external generalization (Fig. 4a,b). When trained using the experimental holograms of lung tissue sections, the supervised hologram reconstruction model FIN (solid red bar) scored higher ECC values ( $P$  value of  $7.5 \times 10^{-38}$ ) than the GedankenNet (solid blue bar) on the same testing set of the lung tissue sections. However, when it comes to external generalization,

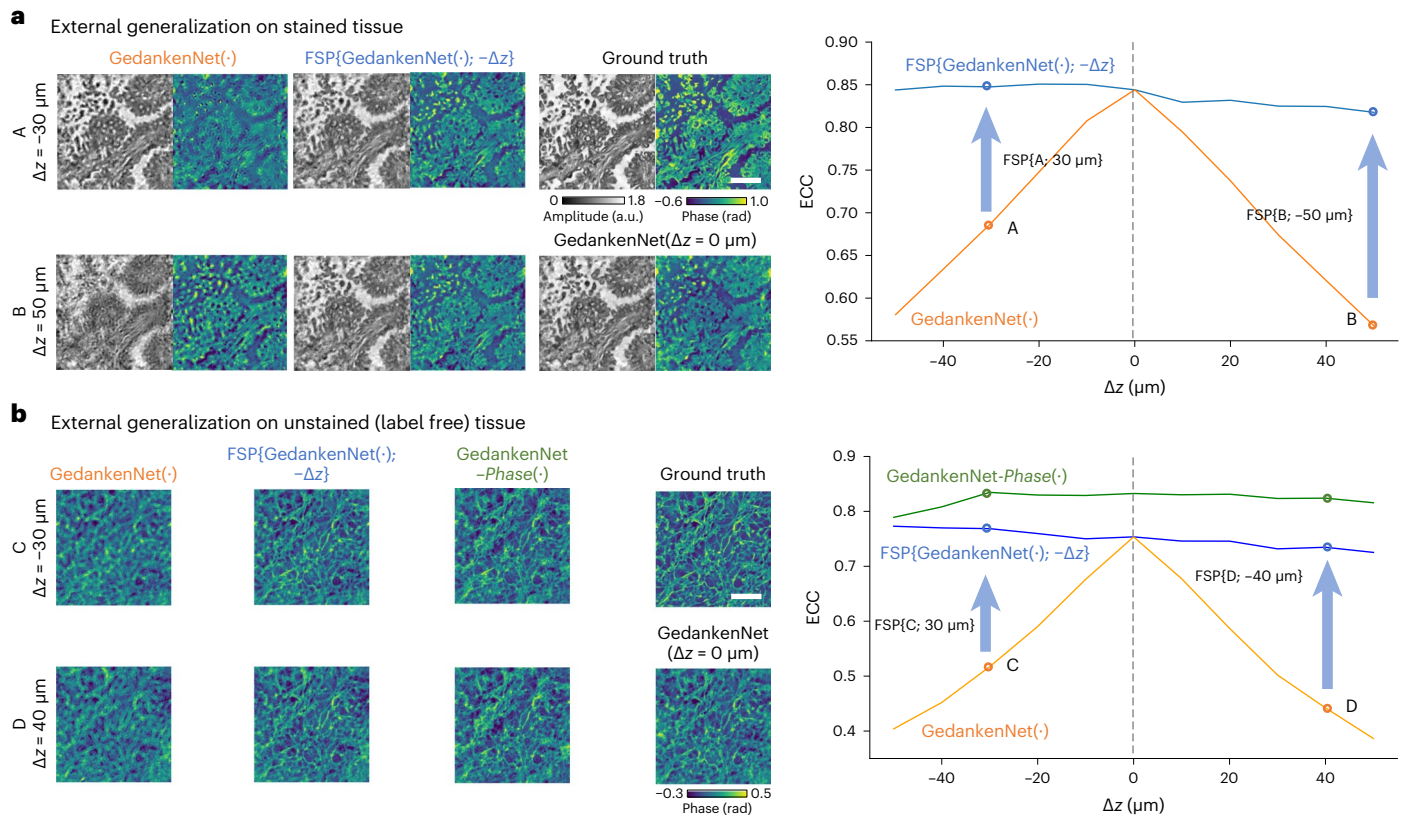


**Fig. 4 | Generalization of GedankenNet trained with different training datasets to various new testing datasets. a**, Outputs of GedankenNet models trained on three different training datasets (artificially generated random synthetic images, natural images (COCO) and tissue sections, respectively). Scale bar, 50  $\mu$ m. **b**, Quantitative performance analysis of GedankenNet models

trained on three different datasets. The performances of a supervised deep neural network (trained on lung tissue sections) and MHPR are also included for comparison purposes. ECC mean  $\pm$  s.d. values are presented and were calculated on lung and COCO test datasets with 94 and 100 unique FOVs, respectively.

as shown in Fig. 4b, GedankenNet (the blue shadow bar) achieved superior imaging performance ( $P$  value of  $8.5 \times 10^{-10}$ ) compared with FIN (the red shadow bar) on natural images (from the COCO dataset). One can also notice the overfitting of the supervised model (FIN) by the large performance gap observed between its internal and external

generalization performance shown with the red bars in Fig. 4b. On the contrary, the self-supervised GedankenNet trained with artificial random images (the blue bars) showed very good generalization performance for both test datasets covering natural macroscale images as well as microscale tissue images. Also see Extended Data Fig. 5 and



**Fig. 5 | Compatibility of GedankenNet output images with the wave equation in free space.** The GedankenNet model was trained to reconstruct  $M = 2$  input holograms at  $z_1 = 300 \mu\text{m}$  and  $z_2 = 375 \mu\text{m}$ , but blindly tested on input holograms captured at  $z'_1 = 300 + \Delta z \mu\text{m}$  and  $z'_2 = 375 + \Delta z \mu\text{m}$  (orange curve). The resulting GedankenNet output complex fields are propagated in free space by  $-\Delta z$  using the wave equation, revealing a very good image quality (blue curve) across a wide range of axial defocus distances. The GedankenNet-Phase (green curve) was trained to reconstruct sample fields with  $M = 2$  input holograms at arbitrary, unknown axial positions within  $[275, 400] \mu\text{m}$ . **a**, External generalization on

stained human lung tissue sections. **A**, **B** represent testing results with  $\Delta z = -30$  and  $50 \mu\text{m}$ , respectively. **b**, External generalization on unstained, label-free human kidney tissue sections. **C**, **D** represent testing results with  $\Delta z = -30$  and  $40 \mu\text{m}$ , respectively. These results demonstrate that the GedankenNet framework not only has a superior external generalization to experimental holograms (using experiment- and data-free training) but also very well generalized to work with defocused experimental holograms, and encoded the wave equation into its inference process using the physics-consistency loss. Scale bars,  $50 \mu\text{m}$ .

Supplementary Note 4 for additional results supporting the superior generalization performance of GedankenNet.

## Compatibility of GedankenNet with the wave equation

Besides its generalization to unseen testing data distributions and experimental holograms, the inference of GedankenNet is also compatible with the wave equation. To demonstrate this, we tested the GedankenNet model (trained with the artificial hologram dataset generated from random synthetic images) on experimental holograms captured at shifted unknown axial positions  $z'_1 \cong z_1 + \Delta z$  and  $z'_2 \cong z_2 + \Delta z$ , where  $z_1$  and  $z_2$  were the training axial positions and  $\Delta z$  is the unknown axial shift amount. The same model as in Fig. 3 was used for this analysis and blindly tested on lung tissue sections (that is, external generalization). Due to the unknown axial defocus distance ( $\Delta z$ ), the direct output fields of GedankenNet do not match well with the ground truth, indicated by the orange curve in Fig. 5a. However, as GedankenNet was trained with the physics-consistency loss, its output fields are compatible with the wave equation in free space. Thus, the object fields at the sample plane can be accurately retrieved from the GedankenNet output fields by performing wave propagation by the corresponding axial defocus distance. After propagating the output fields of GedankenNet by  $-\Delta z$  using the angular spectrum approach, the propagated fields (blue curve) matched very well with

the ground-truth fields across a large range of axial defocus values,  $\Delta z$ . These results are important because (1) they once again demonstrate the success of GedankenNet in generalizing to experimental holograms even though it was trained only by artificial holograms of random synthetic images; and (2) the physics-consistency based self-supervised training of GedankenNet encoded the wave equation into its inference process so that instead of hallucinating and creating non-physical random optical fields when tested with defocused holograms, GedankenNet outputs correct (physically consistent) defocused complex fields. In this sense, GedankenNet not only shows superior external generalization (from experiment- and data-free training to experimental holograms) but also very well generalized to work with defocused experimental holograms. To the best of our knowledge, these features have not been demonstrated before for any hologram reconstruction neural network in the literature.

Figure 5b shows another example of GedankenNet's superior external generalization and its compatibility with the wave equation. The same trained GedankenNet model of Fig. 5a was blindly tested on experimental holograms of unstained (label free) human kidney tissue sections, which can be considered phase-only samples. Besides the success of GedankenNet's generalization to experimental data of biological samples, the results shown in Fig. 5b demonstrate GedankenNet's zero-shot generalization to another physical class of objects (that is, phase-only samples) that exhibit different physical properties

than the synthetic, artificial random complex fields used in the training, which included random phase and amplitude patterns. Stated differently, although GedankenNet's artificially generated random training images did not include any phase-only objects, it successfully reconstructed the experimental holograms of phase-only objects—the first time that they were seen. Furthermore, similar to Fig. 5a, we observe in Fig. 5b that by digitally propagating the GedankenNet outputs for defocused input holograms of the label-free tissue samples (orange curve) by an axial distance of  $-\Delta z$ , the resulting phase reconstructions (blue curve) showed good fidelity to the ground-truth phase images of the same samples.

Also refer to the Supplementary Note 5 and Extended Data Figs. 6–10 on the analysis of GedankenNet's resilience to various sources of unknown perturbations, including the pixel pitch, signal-to-noise ratio, axial distances and the illumination wavelength. As detailed in Supplementary Note 5, this resilient performance of GedankenNet can be further enhanced using phase-only object priors; we termed this new model for phase-only object reconstruction as *GedankenNet-Phase*, which shows external generalization, reconstructing experimental holograms of unseen sample types, while simultaneously achieving autofocusing (Supplementary Note 5.3 and Extended Data Fig. 9).

## Discussion

Compared with the existing supervised learning methods, GedankenNet has several unique advantages. It eliminates the dependence on labelled experimental training data in computational microscopy, which often come from other imaging modalities or classical algorithms and, therefore, inevitably introduce biases for external generalization performance of the trained network. The self-supervised, zero-shot learning scheme of GedankenNet also considerably relieves the cost and labour of collecting and preparing large-scale microscopic image datasets. For the inverse problem of hologram reconstruction, the reported physics-consistency loss that we used in self-supervised learning outperforms traditional structural loss functions commonly employed in supervised learning as they often overfit to specific image features that appear in the training dataset, resulting in generalization errors, especially for new types or classes of sample never seen before (Extended Data Fig. 4). In general, the residual errors that stochastically occur during the network training would be non-physical errors that are incompatible with the wave equation, for example, noise-like errors that do not follow wave propagation. In contrast to traditional structural loss functions that penalize these types of residual error based on the statistics of the sample type of interest (which requires experimental data and/or knowledge about the samples and their features), the physics-consistency loss function that we used focuses on physical inconsistencies, which is at the heart of the superior external generalization of GedankenNet framework as such physical errors are universally applicable, regardless of the type of sample or its physical properties or features (also refer to Supplementary Note 6). Furthermore, this physics-consistency loss benefits from multiple hologram planes (that is,  $M \geq 2$ ) so that it can also filter out twin-image-related artefacts that would normally appear in conventional in-line hologram reconstruction methods due to lack of direct phase information; stated differently, an artificial twin image that would be superimposed onto the complex-valued true image of the sample would be attacked by our self-consistency loss as it will create physical inconsistencies on at least  $M - 1$  hologram planes as a result of the wave propagation step for  $M \geq 2$  planes. In addition to this, the large degrees of freedom provided by the artificially synthesized image datasets, with random phase and amplitude channels, also contribute to the effectiveness of the GedankenNet framework, as also highlighted in the 'Superior Generalization of GedankenNet' subsection, Extended Data Fig. 5 and Supplementary Note 4. Limited by the optical system, the experimental holographic imaging process applies a low-pass filter to the ground-truth object fields. Furthermore, the recurrent spatial features within the same type

of sample further reduce the diversity of the experimental datasets. Thus, adapting simulated holograms of random, artificial image datasets presents a more effective solution when access to large amounts of experimental data is impractical (Fig. 4 and Extended Data Fig. 5). In addition, GedankenNet exhibits superior generalization to unseen data distributions than supervised models, and achieves better holographic image reconstruction for unseen, new types of sample (see, for example, Figs. 3 and 4).

## Methods

### Sample preparation and imaging

Human tissue samples used in this work were prepared and provided by the University of California Los Angeles (UCLA) Translational Pathology Core Laboratory. A fraction of tissue slides were stained with haematoxylin and eosin to reveal structural features in the amplitude channel and the other slides remained unstained to serve as phase-only objects. Stained Pap smears were acquired from the UCLA Department of Pathology. All slides were deidentified and prepared from existing specimens without links or identifiers to the patients.

The experimental holograms were captured on stained human lung, prostate, salivary gland, kidney, liver and oesophagus tissue sections, Pap smears, and unstained label-free human kidney tissue sections. Holographic microscopy imaging was implemented using a lens-free, in-line holographic microscope as illustrated in Extended Data Fig. 1b. The custom-designed microscope was equipped with a tunable light source (WhiteLase Micro, NKT Photonics) and an acousto-optic tunable filter. In the reported experiments, the acousto-optic tunable filter was set to filter the illumination light at  $\lambda_0 = 530$  nm wavelength unless otherwise specified. Raw holograms ( $-4,600 \times 3,500$  pixels) were recorded by a complementary metal-oxide semiconductor with a pixel size of  $1.12 \mu\text{m}$  (IMX 081 RGB, Sony). A 6-axis 3D positioning stage (MAX606, Thorlabs) controlled the complementary metal-oxide semiconductor sensor to capture raw holograms consecutively for each FOV at various sample-to-sensor distances, which ranged from  $-300 \mu\text{m}$  to  $-600 \mu\text{m}$  in this work with an axial spacing of  $-10$ – $15 \mu\text{m}$ . A computer connected all the devices and automatically controlled the image acquisition process through a LabView script (LabView 2012, version 12.0) and AYA software tool (Sony).

### Artificial hologram preparation and preprocessing

The artificial holograms used in this work for the training were simulated from either random images or natural images (from the COCO dataset). Random images (with no connection or resemblance to real-world samples) were generated using a Python package *rand-image*, which coloured the pixels along a path found from a random grey-valued image to generate an artificial RGB image. Then we mapped the generated random RGB images to greyscale. Two independent images randomly selected from the dataset served as the amplitude and phase of the complex object field, and a small constant was added to the amplitude channel to avoid zero transmission and undefined phase issue. For the artificial random phase-only object fields, only the phase image was selected, and the amplitude was set as 1 everywhere. The object field was then propagated by the given sample-to-sensor distances using the angular spectrum approach<sup>84</sup>, and the intensity of the resulting complex field was calculated. The resulting holograms were cropped into  $512 \times 512$  patches. Each of the two datasets (from either random images or COCO natural images) used  $-100,000$  images for training and a set of 100 images for validation and testing, which were excluded from training. All models in this work used the amplitude of the measured fields as the inputs.

Given a randomly selected amplitude ( $A$ ) image and phase ( $\phi$ ) image, the simulated hologram  $i(x, y; z)$  at axial position  $z$  is generated by free-space propagation (FSP):

$$i(x, y; z) = |\text{FSP}((A + \delta) \odot e^{i\phi}, z) + \epsilon|$$



where  $\delta \in \mathbb{R}$  stands for the added small constant,  $\odot$  represents element-wise multiplication and  $\epsilon \in \mathbb{C}^{N \times N}$  is the additional white Gaussian noise. For the phase-only objects, the simulated holograms can be expressed as:

$$i(x, y; z) = |\text{FSP}(e^{i\phi}; z) + \epsilon|$$

The FSP is implemented based on the angular spectrum propagation method<sup>84</sup> by taking into account all the travelling waves in free space. The angular spectrum of a light field  $U(x, y; z_0)$  at the axial position  $z_0$  can be expressed as

$$\tilde{A}(\xi, \eta; z_0) = \mathcal{F}\{U(x, y; z_0)\}$$

The angular spectrum of the propagated field at  $z$  is related to  $A(\xi, \eta; z_0)$  by

$$\tilde{A}(\xi, \eta; z) = \begin{cases} e^{i(z-z_0)\sqrt{k^2-\xi^2-\eta^2}} \tilde{A}(\xi, \eta; z_0), & \text{if } \xi^2 + \eta^2 < k^2 \\ 0, & \text{otherwise} \end{cases}$$

Here  $\mathcal{F}$  and  $\mathcal{F}^{-1}$  are the fast Fourier transform (FFT) pairs (forward versus inverse).  $x, y$  and  $\xi, \eta$  are spatial and frequency domain coordinates, respectively.  $k$  is the wave number of the illumination light in the medium. The FSP then infers the propagated field at  $z$  by:

$$\text{FSP}(U(x, y; z_0); z - z_0) = \mathcal{F}^{-1}\{\tilde{A}(\xi, \eta; z)\}.$$

### Experimental hologram dataset preparation and processing

Raw experimental holograms were pre-processed through pixel super-resolution and autofocusing algorithms to retrieve subpixel features of the samples. For this, a pixel super-resolution algorithm<sup>85,91</sup> was applied to raw experimental holograms to obtain high-resolution holograms, resulting in a final effective pixel size of 0.37  $\mu\text{m}$ . Then, an edge sparsity-based autofocusing algorithm<sup>92</sup> was employed to determine the sample-to-sensor distances for each super-resolved hologram. The ground-truth sample field was retrieved from  $M = 8$  super-resolved holograms of the same FOV using the MHPR algorithm<sup>85-87</sup>. The MHPR algorithm retrieves the sample complex field through iterations between eight input holograms. The initial guess of the sample complex field is propagated to each measurement plane using FSP and the corresponding sample-to-sensor distance. Then, the propagated field is updated by replacing the amplitude with the measured one and retaining the phase. One iteration is completed after all eight holograms have been used. The algorithm generally converges after 100 iterations.

Input-target pairs of  $512 \times 512$  pixels were cropped from the super-resolved holograms and their corresponding retrieved ground-truth fields, forming the experimental hologram datasets. Standard data augmentation techniques were applied, including random rotations by  $0^\circ, \pm 45^\circ$  and  $\pm 90^\circ$ , and random vertical and horizontal flipping. The multi-height experimental hologram dataset of tissue sections contains ~100,000 input-target pairs of stained human lung, prostate, salivary gland, kidney, liver and oesophagus tissue sections. A subset of the lung, prostate, salivary gland slides from new patients and Pap smears were excluded from the training dataset and used as testing datasets, containing 94, 49, 49 and 47 unique FOVs, respectively. The holograms of the unstained (label free) kidney tissue thin sections (~3–4  $\mu\text{m}$  thick) were used as our phase-only object test dataset containing 98 unique FOVs.

### Network architecture

A sequence of  $M$  holograms is concatenated as the input image with  $M$  channels and the real and imaginary parts of the object complex field are generated at the output of GedankenNet. GedankenNet contains a series of spatial Fourier transformation (SPAF) blocks and a large-scale residual connection, in addition to two  $1 \times 1$  convolution layers at the head and the tail of the network (Extended Data Fig. 1a). In each SPAF

block, input tensors pass through two recursive SPAF modules with residual connections, which share the same parameters before entering the parametric rectified linear unit (PReLU) activation layer<sup>93</sup>. The PReLU activation function with respect to an input value  $x \in \mathbb{R}$  is defined as:

$$\text{PReLU}(x) = \max(0, x) + a \times \min(0, x)$$

where  $a \in \mathbb{R}$  is a learnable parameter. Another residual connection passes the input tensor after the PReLU layer. The SPAF module consists of a  $3 \times 3$  convolution layer and a branch performing linear transformation in the Fourier domain (Extended Data Fig. 1a). The input tensor with  $c$  channels to the SPAF module is first transformed into the frequency domain by a two-dimensional FFT and truncated by a window with a half size  $k$  to filter out higher-frequency components. The linear transformation in the frequency domain is realized through pixel-wise multiplication with a trainable weight tensor  $W \in \mathbb{R}^{c \times (2k+1) \times (2k+1)}$ , that is

$$F'_{j,u,v} = W_{j,u,v} \cdot \sum_{i=1}^c F_{i,u,v}, \quad u, v = 0, \pm 1, \dots, \pm k, \quad j = 1, \dots, c$$

where  $F \in \mathbb{C}^{c \times (2k+1) \times (2k+1)}$  are the truncated frequency components. The resulting tensor  $F'$  is then transformed into the spatial domain through an inverse two-dimensional FFT. The same pyramid-like setting of half window size  $k$  as in ref. 82 was applied here such that  $k$  decreases for deeper SPAF blocks. This pyramid-like setting provides a mapping of the high-frequency information of the holographic diffraction patterns to low-frequency regions in the first few layers and passes this low-frequency information to the subsequent layers with a smaller window size, which better utilizes the spatial features at multiple scales and at the same time considerably reduces the model size, avoiding potential overfitting and generalization issues.

The architecture of GedankenNet was extended for two additional models, namely GedankenNet-Phase and GedankenNet-Phase $\lambda$ , as detailed in the Supplementary Note 5 and Extended Data Fig. 8a. Similar to GedankenNet, these models use a sequence of  $M$  holograms concatenated as the input image with  $M$  channels, but, instead of outputting real and imaginary parts, the GedankenNet-Phase and GedankenNet-Phase $\lambda$  only generate phase-only output images. The dynamic SPAF (dSPAF) modules<sup>49</sup> inside GedankenNet-Phase and GedankenNet-Phase $\lambda$  exploit a shallow U-Net to dynamically generate weights  $\mathbf{W}$  for each input tensor, and enable the capabilities of autofocusing and adapting to unknown shifts or changes in the illumination wavelengths. The dense links provide an efficient flow of information from the input layer to the output layer, so that every output tensor of the dSPAF group is appended and fed to the subsequent dSPAF groups, resulting in an economic and powerful network architecture.

### Algorithm implementation

GedankenNet, GedankenNet-Phase and GedankenNet-Phase $\lambda$  were implemented using PyTorch<sup>94</sup>. We calculated the loss values based on the hologram amplitudes, that is:

$$\hat{i} = |\text{FSP}(\hat{o}; z_1, z_2, \dots, z_M)|$$

The training loss consists of three individual terms: (1) FDMAE loss between the predicted holograms  $\hat{i}$  and the input holograms  $i$ ; (2) MSE loss between  $\hat{i}$  and  $i$ ; and (3) total variation (TV) loss on the output complex field  $\hat{o}$ . The first two terms constitute the physics-consistency loss, and the total loss is a linear combination of the three terms, expressed as:

$$\begin{aligned} L_{\text{total}} &= L_{\text{physics-consistency}}(\hat{i}, i) + \gamma L_{\text{TV}}(\hat{o}) \\ &= \alpha L_{\text{FDMAE}}(\hat{i}, i) + \beta L_{\text{MSE}}(\hat{i}, i) + \gamma L_{\text{TV}}(\hat{o}) \end{aligned}$$

where  $\alpha, \beta$  and  $\gamma$  are loss weights empirically set as 0.1, 1 and 20.

The FDMAE loss is calculated as:

$$L_{\text{FDMAE}}(\hat{i}, i) = \frac{1}{N^2} \sum_{\xi=1}^N \sum_{\eta=1}^N |\mathcal{F}\{\hat{i}\}(\xi, \eta) \cdot w(\xi, \eta) - \mathcal{F}\{i\}(\xi, \eta) \cdot w(\xi, \eta)|$$

Here  $w \in \mathbb{R}^{N \times N}$  is a two-dimensional Hann window<sup>95</sup>, and  $\xi, \eta$  are indices of frequency components. MSE and TV losses are computed using:

$$L_{\text{MSE}}(\hat{i}, i) = \frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N |\hat{i}(x, y) - i(x, y)|^2$$

$$L_{\text{TV}}(\hat{o}) = \frac{1}{2N^2} \sum_{x=1}^N \sum_{y=1}^N (|\nabla_x \text{Re}\{\hat{o}\}(x, y)| + |\nabla_y \text{Re}\{\hat{o}\}(x, y)| + |\nabla_x \text{Im}\{\hat{o}\}(x, y)| + |\nabla_y \text{Im}\{\hat{o}\}(x, y)|)$$

Here  $x, y$  are spatial indices,  $\nabla_x, \nabla_y$  refer to the differentiation operation along the horizontal and vertical axes,  $\text{Re}\{\cdot\}, \text{Im}\{\cdot\}$  return the real and imaginary parts of the complex fields, respectively.

For the GedankenNet-Phase and GedankenNet-Phase $\lambda$ , which generate phase-only output fields, the predicted hologram was calculated using:

$$\hat{i} = |\text{FSP}(e^{i\hat{p}}; z_1, z_2, \dots, z_M)|$$

where  $\hat{p}$  is the output phase field. The TV loss was calculated by using:

$$L_{\text{TV}}(\hat{p}) = \frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N (|\nabla_x \hat{p}(x, y)| + |\nabla_y \hat{p}(x, y)|)$$

To avoid trivial ambiguities in phase retrieval<sup>96–98</sup>, the GedankenNet’s output was normalized using its complex mean; the outputs of GedankenNet-Phase and GedankenNet-Phase $\lambda$  were subtracted from their corresponding mean.

All the trainable parameters in GedankenNet were optimized using the Adam optimizer<sup>99</sup>. The learning rate follows a cosine annealing scheduler with an initial rate of 0.002. All the models went through ~0.75 million batches (equivalent to ~7.5 epochs) and the best model was preserved with the minimal validation loss. The training takes ~48 h for an  $M=2$  model on a computer equipped with an i9–12900F central processing unit, 64 GB random-access memory and an RTX 3090 graphics card. The inference time measurement (Table 1) was done on the same machine with GPU acceleration and a test batch size of 20 for GedankenNet, 12 for both GedankenNet-Phase and GedankenNet-Phase $\lambda$ .

The supervised FIN adopted the same architecture and parameters as in ref. 48. The U-Net architecture employed four convolutional blocks in the down-sampling and up-sampling paths separately, and each block contained two convolutional layers with batch normalization and ReLU activation. The input feature maps of the first convolutional block had 64 channels and each block in the down-sampling path doubled the number of channels. Supervised FIN and U-Net<sup>88</sup> models adopted the same loss function as in ref. 48. The same Adam optimizer and learning rate were applied to the supervised learning models. Deep image prior adopted a U-Net architecture, an Adam optimizer and the loss function used in ref. 81.

### Image reconstruction evaluation metrics

SSIM, root mean square error (RMSE) and ECC were used in our work to evaluate the reconstruction quality of the output fields with respect to the ground-truth fields. SSIM and RMSE are based on single-channel images. Denoting  $\hat{o} \in \mathbb{R}^{N \times N}$  as the reconstructed amplitude or phase image, and  $o \in \mathbb{R}^{N \times N}$  as the ground-truth amplitude or phase

image, SSIM and RMSE values were calculated using the following equations:

$$\text{SSIM}(\hat{o}, o) = \frac{(2\mu_{\hat{o}}\mu_o + c_1)(2\sigma_{\hat{o}o} + c_2)}{(\mu_{\hat{o}}^2 + \mu_o^2 + c_1)(\sigma_{\hat{o}}^2 + \sigma_o^2 + c_2)}$$

$$\text{RMSE}(\hat{o}, o) = \sqrt{\frac{1}{N^2} \sum_{x=1}^N \sum_{y=1}^N (\hat{o}(x, y) - o(x, y))^2}$$

Here  $\mu_{\hat{o}}$  and  $\mu_o$  stand for the mean of  $\hat{o}$  and  $o$ , respectively.  $\sigma_{\hat{o}}^2$  and  $\sigma_o^2$  stand for the variance of  $\hat{o}$  and  $o$ , respectively, and  $\sigma_{\hat{o}o}$  is the covariance between  $\hat{o}$  and  $o$ .  $c_1 = 2.55^2$  and  $c_2 = 7.65^2$  are constants used for 8-bit images.  $x$  and  $y$  are two-dimensional coordinates of the image pixels.

The ECC is calculated based on the reconstructed complex field and the ground-truth field.  $\hat{o}' \in \mathbb{C}^{N \times N}$  is the reconstructed field obtained by subtracting  $\hat{o}$  with its mean value.  $o' \in \mathbb{C}^{N \times N}$  is the corresponding ground-truth field. The ECC can be calculated as:

$$\text{ECC}(\hat{o}', o') = \text{Re} \left\{ \frac{\text{vec}(\hat{o}')^H \cdot \text{vec}(o')}{\|\text{vec}(\hat{o}')\| \cdot \|\text{vec}(o')\|} \right\}$$

Here  $\text{vec}(\hat{o}')^H$  is the conjugate transpose of the vectorized  $\hat{o}'$ , and  $\|\cdot\|$  is the Euclidean norm.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

### Data availability

This study involved simulation data generated from the COCO dataset (available at <https://cocodataset.org/#home>) and an artificial image dataset, as well as an experimental image dataset of human samples (from existing, anonymized specimens collected before this work). Simulation data used in the paper can be generated using the open-source code of the paper, public datasets and open-source Python libraries. The training artificial image dataset is available at <https://github.com/PORPHURA/GedankenNet> (ref. 100). A portion of the experimental dataset corresponding to human tissue samples is also shared and referenced in the code repository without any links or identifiers to the patients, which is made available at <https://github.com/PORPHURA/GedankenNet> (ref. 100). Additional data are available from the corresponding author upon reasonable request.

### Code availability

The codes for the deep learning models used in this work (written in Python 3.9.6 and PyTorch 1.9.0) are publicly available at <https://github.com/PORPHURA/GedankenNet> (ref. 100). The trained model and demo data are uploaded and available in the same code repository. The code for analysing the results was written in Python using standard, open-source Python libraries.

### References

1. Suzuki, K. Overview of deep learning in medical imaging. *Radiol. Phys. Technol.* **10**, 257–273 (2017).
2. Ma, L. et al. Deep learning in remote sensing applications: a meta-analysis and review. *ISPRS J. Photogramm. Remote Sens.* **152**, 166–177 (2019).
3. Bakator, M. & Radosav, D. Deep learning and medical diagnosis: a review of literature. *Multimodal Technol. Interact.* **2**, 47 (2018).
4. Shimobaba, T. et al. Deep-learning computational holography: a review (invited). *Front. Photon.* **3**, 854391 (2022).
5. Rivenson, Y. et al. Deep learning microscopy. *Optica* **4**, 1437 (2017).

6. Nehme, E., Weiss, L. E., Michaeli, T. & Shechtman, Y. Deep-STORM: super-resolution single-molecule microscopy by deep learning. *Optica* **5**, 458 (2018).
7. Ouyang, W., Aristov, A., Lelek, M., Hao, X. & Zimmer, C. Deep learning massively accelerates super-resolution localization microscopy. *Nat. Biotechnol.* **36**, 460–468 (2018).
8. Wang, H. et al. Deep learning enables cross-modality super-resolution in fluorescence microscopy. *Nat. Methods* **16**, 103–110 (2019).
9. Liu, T. et al. Deep learning-based super-resolution in coherent imaging systems. *Sci. Rep.* **9**, 3926 (2019).
10. Nehme, E. et al. DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning. *Nat. Methods* **17**, 734–740 (2020).
11. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
12. Xiong, H. et al. Super-resolution vibrational microscopy by stimulated Raman excited fluorescence. *Light Sci. Appl.* **10**, 87 (2021).
13. Weigert, M. et al. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).
14. de Haan, K., Rivenson, Y., Wu, Y. & Ozcan, A. Deep-learning-based image reconstruction and enhancement in optical microscopy. *Proc. IEEE* **108**, 30–50 (2020).
15. Chen, J. et al. Three-dimensional residual channel attention networks denoise and sharpen fluorescence microscopy image volumes. *Nat. Methods* **18**, 678–687 (2021).
16. Lequyer, J., Philip, R., Sharma, A., Hsu, W.-H. & Pelletier, L. A fast blind zero-shot denoiser. *Nat. Mach. Intell.* **4**, 953–963 (2022).
17. Rivenson, Y. et al. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**, 466–477 (2019).
18. Rivenson, Y. et al. PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl.* **8**, 23 (2019).
19. Nygate, Y. N. et al. Holographic virtual staining of individual biological cells. *Proc. Natl Acad. Sci. USA* **117**, 9223–9231 (2020).
20. Zhang, Y. et al. Digital synthesis of histological stains using micro-structured and multiplexed virtual staining of label-free tissue. *Light Sci. Appl.* **9**, 78 (2020).
21. Liu, Y., Yuan, H., Wang, Z. & Ji, S. Global pixel transformers for virtual staining of microscopy images. *IEEE Trans. Med. Imaging* **39**, 2256–2266 (2020).
22. de Haan, K. et al. Deep learning-based transformation of H&E stained tissues into special stains. *Nat. Commun.* **12**, 4884 (2021).
23. Zhang, Y. et al. Virtual staining of defocused autofluorescence images of unlabeled tissue using deep neural networks. *Intell. Comput.* **2022**, 9818965 (2022).
24. Rivenson, Y. et al. Deep learning enhanced mobile-phone microscopy. *ACS Photon.* **5**, 2354–2364 (2018).
25. Wu, Y. et al. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light Sci. Appl.* **8**, 25 (2019).
26. Luo, Y., Huang, L., Rivenson, Y. & Ozcan, A. Single-shot autofocusing of microscopy images using deep learning. *ACS Photon.* **8**, 625–638 (2021).
27. Huang, L., Chen, H., Luo, Y., Rivenson, Y. & Ozcan, A. Recurrent neural network-based volumetric fluorescence microscopy. *Light Sci. Appl.* **10**, 62 (2021).
28. Mudanyali, O. et al. Compact, light-weight and cost-effective microscope based on lensless incoherent holography for telemedicine applications. *Lab Chip* **10**, 1417–1428 (2010).
29. Popescu, G. *Quantitative Phase Imaging of Cells and Tissues* (McGraw-Hill, 2011).
30. Osten, W. et al. Recent advances in digital holography [invited]. *Appl. Opt.* **53**, G44 (2014).
31. Memmolo, P. et al. Recent advances in holographic 3D particle tracking. *Adv. Opt. Photon.* **7**, 713 (2015).
32. Merola, F. et al. Tomographic flow cytometry by digital holography. *Light Sci. Appl.* **6**, e16241 (2017).
33. Wu, Y. & Ozcan, A. Lensless digital holographic microscopy and its applications in biomedicine and environmental monitoring. *Methods* **136**, 4–16 (2018).
34. Paturzo, M. et al. Digital holography, a metrological tool for quantitative analysis: trends and future applications. *Opt. Lasers Eng.* **104**, 32–47 (2018).
35. Park, Y., Depeursinge, C. & Popescu, G. Quantitative phase imaging in biomedicine. *Nat. Photon.* **12**, 578–589 (2018).
36. Javidi, B. et al. Roadmap on digital holography [invited]. *Opt. Express* **29**, 35078 (2021).
37. Balasubramani, V. et al. Roadmap on digital holography-based quantitative phase imaging. *J. Imaging* **7**, 252 (2021).
38. Rivenson, Y., Zhang, Y., Günaydin, H., Teng, D. & Ozcan, A. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
39. Wu, Y. et al. Extended depth-of-field in holographic imaging using deep-learning-based autofocusing and phase recovery. *Optica* **5**, 704–710 (2018).
40. Zhang, G. et al. Fast phase retrieval in off-axis digital holographic microscopy through deep learning. *Opt. Express* **26**, 19388–19405 (2018).
41. Wang, K., Dou, J., Kemao, Q., Di, J. & Zhao, J. Y-Net: a one-to-two deep learning framework for digital holographic reconstruction. *Opt. Lett.* **44**, 4765–4768 (2019).
42. Jo, Y. et al. Quantitative phase imaging and artificial intelligence: a review. *IEEE J. Sel. Top. Quantum Electron.* **25**, 1–14 (2019).
43. Rivenson, Y., Wu, Y. & Ozcan, A. Deep learning in holography and coherent imaging. *Light Sci. Appl.* **8**, 85 (2019).
44. Barbastathis, G., Ozcan, A. & Situ, G. On the use of deep learning for computational imaging. *Optica* **6**, 921–943 (2019).
45. Ren, Z., Xu, Z. & Lam, E. Y. End-to-end deep learning framework for digital holographic reconstruction. *Adv. Photon.* **1**, 016004 (2019).
46. Deng, M., Li, S., Goy, A., Kang, I. & Barbastathis, G. Learning to synthesize: robust phase retrieval at low photon counts. *Light Sci. Appl.* **9**, 36 (2020).
47. Huang, L. et al. Holographic image reconstruction with phase recovery and autofocusing using recurrent neural networks. *ACS Photon.* **8**, 1763–1774 (2021).
48. Chen, H., Huang, L., Liu, T. & Ozcan, A. Fourier Imager Network (FIN): a deep neural network for hologram reconstruction with superior external generalization. *Light Sci. Appl.* **11**, 254 (2022).
49. Chen, H., Huang, L., Liu, T. & Ozcan, A. eFIN: Enhanced Fourier Imager Network for generalizable autofocusing and pixel super-resolution in holographic imaging. *IEEE J. Sel. Top. Quantum Electron.* **29**, 1–12 (2023).
50. Goy, A., Arthur, K., Li, S. & Barbastathis, G. Low photon count phase retrieval using deep learning. *Phys. Rev. Lett.* **121**, 243902 (2018).
51. Byeon, H., Go, T. & Lee, S. J. Deep learning-based digital in-line holographic microscopy for high resolution with extended field of view. *Opt. Laser Technol.* **113**, 77–86 (2019).
52. Huang, L., Yang, X., Liu, T. & Ozcan, A. Few-shot transfer learning for holographic image reconstruction using a recurrent neural network. *APL Photon.* **7**, 070801 (2022).
53. Jo, Y. et al. Holographic deep learning for rapid optical screening of anthrax spores. *Sci. Adv.* **3**, e1700606 (2017).
54. Pavillon, N., Hobro, A. J., Akira, S. & Smith, N. I. Noninvasive detection of macrophage activation with single-cell resolution through machine learning. *Proc. Natl Acad. Sci. USA* **115**, E2676–E2685 (2018).

55. Go, T., Kim, J. H., Byeon, H. & Lee, S. J. Machine learning-based in-line holographic sensing of unstained malaria-infected red blood cells. *J. Biophoton.* **11**, e201800101 (2018).
56. Choi, G. et al. Cycle-consistent deep learning approach to coherent noise reduction in optical diffraction tomography. *Opt. Express* **27**, 4927 (2019).
57. Ozaki, Y. et al. Label-free classification of cells based on supervised machine learning of subcellular structures. *PLoS ONE* **14**, e0211347 (2019).
58. Rubin, M. et al. TOP-GAN: stain-free cancer cell classification using deep learning with a small training set. *Med. Image Anal.* **57**, 176–185 (2019).
59. Chen, L., Chen, X., Cui, H., Long, Y. & Wu, J. Image enhancement in lensless inline holographic microscope by inter-modality learning with denoising convolutional neural network. *Opt. Commun.* **484**, 126682 (2021).
60. Kandel, M. E. et al. Phase imaging with computational specificity (PICS) for measuring dry mass changes in sub-cellular compartments. *Nat. Commun.* **11**, 6256 (2020).
61. McRae, T. D., Oleksyn, D., Miller, J. & Gao, Y.-R. Robust blind spectral unmixing for fluorescence microscopy using unsupervised learning. *PLoS ONE* **14**, e0225410 (2019).
62. Abdolghader, P. et al. Unsupervised hyperspectral stimulated Raman microscopy image enhancement: denoising and segmentation via one-shot deep learning. *Opt. Express* **29**, 34205 (2021).
63. Zhang, Y. et al. PhaseGAN: a deep-learning phase-retrieval approach for unpaired datasets. *Opt. Express* **29**, 19593 (2021).
64. Li, X. et al. Unsupervised content-preserving transformation for optical microscopy. *Light Sci. Appl.* **10**, 44 (2021).
65. Yao, Y. et al. AutoPhaseNN: unsupervised physics-aware deep learning of 3D nanoscale Bragg coherent diffraction imaging. *npj Comput. Mater.* **8**, 124 (2022).
66. Zhang, Y. et al. High-throughput, label-free and slide-free histological imaging by computational microscopy and unsupervised learning. *Adv. Sci.* **9**, 2102358 (2022).
67. Lee, C., Song, G., Kim, H., Ye, J. C. & Jang, M. Deep learning based on parameterized physical forward model for adaptive holographic imaging with unpaired data. *Nat. Mach. Intell.* **5**, 35–45 (2023).
68. Krull, A., Vičar, T., Prakash, M., Lalit, M. & Jug, F. Probabilistic Noise2Void: unsupervised content-aware denoising. *Front. Comput. Sci.* **2**, 5 (2020).
69. Wang, Y. et al. Image denoising for fluorescence microscopy by supervised to self-supervised transfer learning. *Opt. Express* **29**, 41303 (2021).
70. Wu, X., Wu, Z., Shanmugavel, S. C., Yu, H. Z. & Zhu, Y. Physics-informed neural network for phase imaging based on transport of intensity equation. *Opt. Express* **30**, 43398 (2022).
71. Boyd, N., Jonas, E., Babcock, H. & Recht, B. DeepLoco: fast 3D localization microscopy using neural networks. Preprint at <https://doi.org/10.1101/267096> (2018).
72. Zhang, H. et al. High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. *Biomed. Opt. Express* **10**, 1044 (2019).
73. Sekh, A. A. et al. Physics-based machine learning for subcellular segmentation in living cells. *Nat. Mach. Intell.* **3**, 1071–1080 (2021).
74. Wang, Z. et al. Real-time volumetric reconstruction of biological dynamics with light-field microscopy and deep learning. *Nat. Methods* **18**, 551–556 (2021).
75. Chen, J. et al. A transfer learning based super-resolution microscopy for biopsy slice images: the joint methods perspective. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **18**, 103–113 (2020).
76. Christensen, C. N., Ward, E. N., Lu, M., Lio, P. & Kaminski, C. F. ML-SIM: universal reconstruction of structured illumination microscopy images using transfer learning. *Biomed. Opt. Express* **12**, 2720 (2021).
77. Zhang, Z. et al. Deep and domain transfer learning aided photoacoustic microscopy: acoustic resolution to optical resolution. *IEEE Trans. Med. Imaging* **41**, 3636–3648 (2022).
78. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
79. 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).
80. Jagtap, A. D. & Em Karniadakis, G. Extended physics-informed neural networks (XPINNs): a generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations. *Commun. Comput. Phys.* **28**, 2002–2041 (2020).
81. Wang, F. et al. Phase imaging with an untrained neural network. *Light Sci. Appl.* **9**, 77 (2020).
82. Li, Z. et al. Fourier neural operator for parametric partial differential equations. In *2021 Internal Conference on Learning Representations (ICLR)* (2021).
83. Karniadakis, G. E. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
84. Goodman, J. W. *Introduction to Fourier Optics* (Roberts & Co, 2005).
85. Greenbaum, A. & Ozcan, A. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Opt. Express* **20**, 3129–3143 (2012).
86. Greenbaum, A. et al. Wide-field computational imaging of pathology slides using lens-free on-chip microscopy. *Sci. Transl. Med.* **6**, 267ra175 (2014).
87. Rivenson, Y. et al. Sparsity-based multi-height phase recovery in holographic microscopy. *Sci Rep.* **6**, 37862 (2016).
88. Ronneberger, O., Fischer, P. & Brox, T. U-Net: convolutional networks for biomedical image segmentation. Preprint at <https://arxiv.org/abs/1505.04597> (2015).
89. Lempitsky, V., Vedaldi, A. & Ulyanov, D. Deep Image Prior. In *Recognition 9446–9454* (IEEE, 2018).
90. Lin, T.-Y. et al. Microsoft COCO: common objects in context. In *Computer Vision-European Conference on Computer Vision (ECCV) 740–755* (Springer, 2014).
91. Greenbaum, A., Sikora, U. & Ozcan, A. Field-portable wide-field microscopy of dense samples using multi-height pixel super-resolution based lens-free imaging. *Lab Chip* **12**, 1242 (2012).
92. Zhang, Y., Wang, H., Wu, Y., Tamamitsu, M. & Ozcan, A. Edge sparsity criterion for robust holographic autofocusing. *Opt. Lett.* **42**, 3824–3827 (2017).
93. He, K., Zhang, X., Ren, S. & Sun, J. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In *2015 IEEE International Conference on Computer Vision (ICCV) 1026–1034* (IEEE, 2015).
94. Paszke, A. et al. PyTorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* 8024–8035 (Curran Associates, 2019).
95. Gottinger, H.-W. *Elements of Statistical Analysis* (W. de Gruyter, 1980).
96. Fienup, J. R. Phase retrieval algorithms: a comparison. *Appl. Opt.* **21**, 2758 (1982).
97. Miao, J., Sayre, D. & Chapman, H. N. Phase retrieval from the magnitude of the Fourier transforms of nonperiodic objects. *J. Opt. Soc. Am. A* **15**, 1662 (1998).

98. Bendory, T., Beinert, R. & Eldar, Y. C. in *Compressed Sensing and its Applications* (eds Boche, H. et al.) 55–91 (Springer, 2017).
99. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *International Conference on Learning Representations (ICLR)* (2015).
100. Huang L., Chen H., Liu T. & Ozcan A. Self-supervised learning of hologram reconstruction using physics consistency. *Zenodo* <https://doi.org/10.5281/zenodo.8035499> (2023).

## Acknowledgements

We acknowledge the funding of the US National Science Foundation (NSF) and Koç Group, and thank J. Hu (UCLA) for valuable discussions.

## Author contributions

A.O. and L.H. conceived the research. L.H. and H.C. conducted the numerical experiments and analysed the results. T.L. captured the experimental data. T.L. and L.H. processed and analysed the data. All the authors contributed to the preparation of the paper. A.O. supervised the research.

## Competing interests

The authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at <https://doi.org/10.1038/s42256-023-00704-7>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42256-023-00704-7>.

**Correspondence and requests for materials** should be addressed to Aydogan Ozcan.

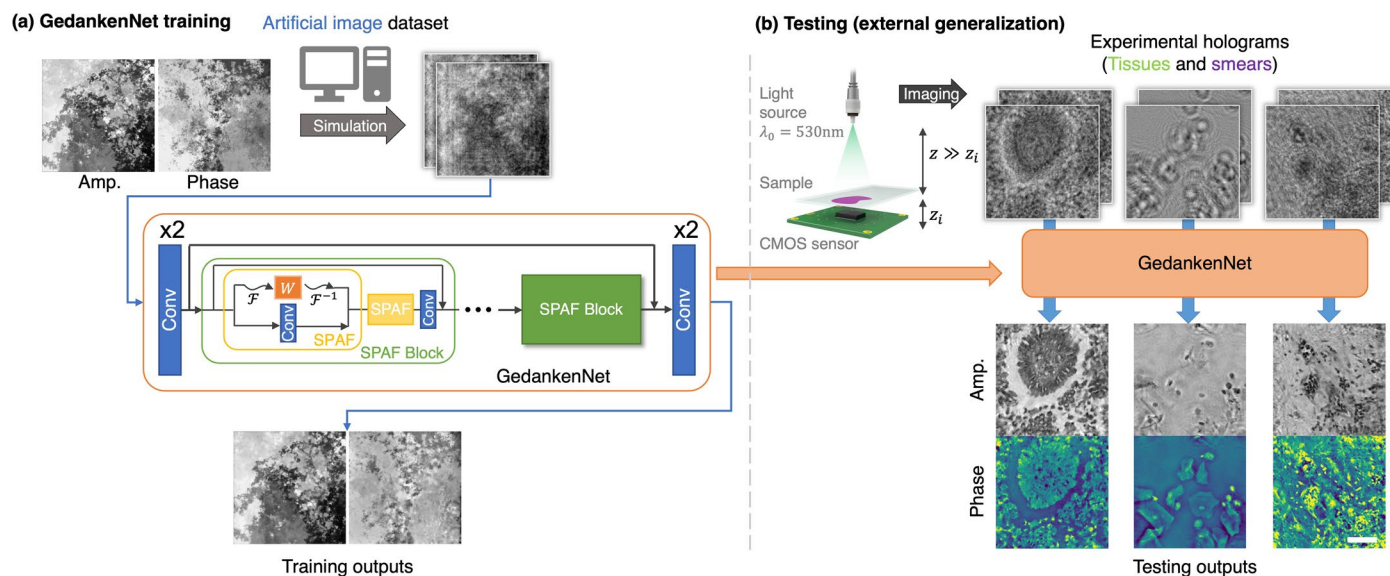
**Peer review information** *Nature Machine Intelligence* thanks Sang Joon Lee and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

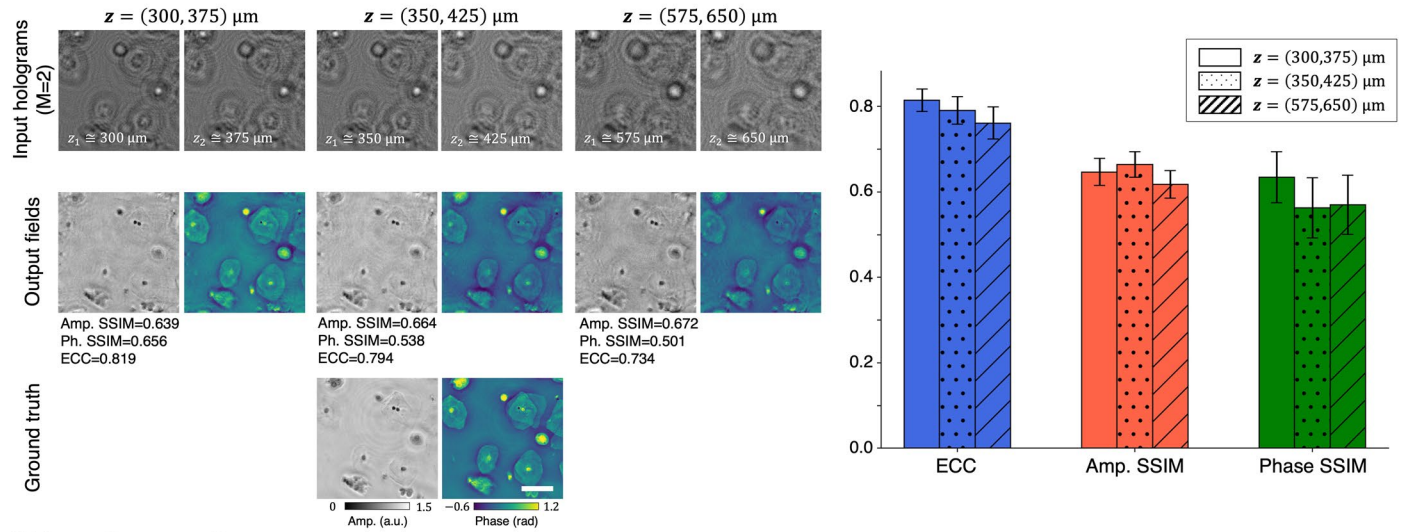
© The Author(s) 2023



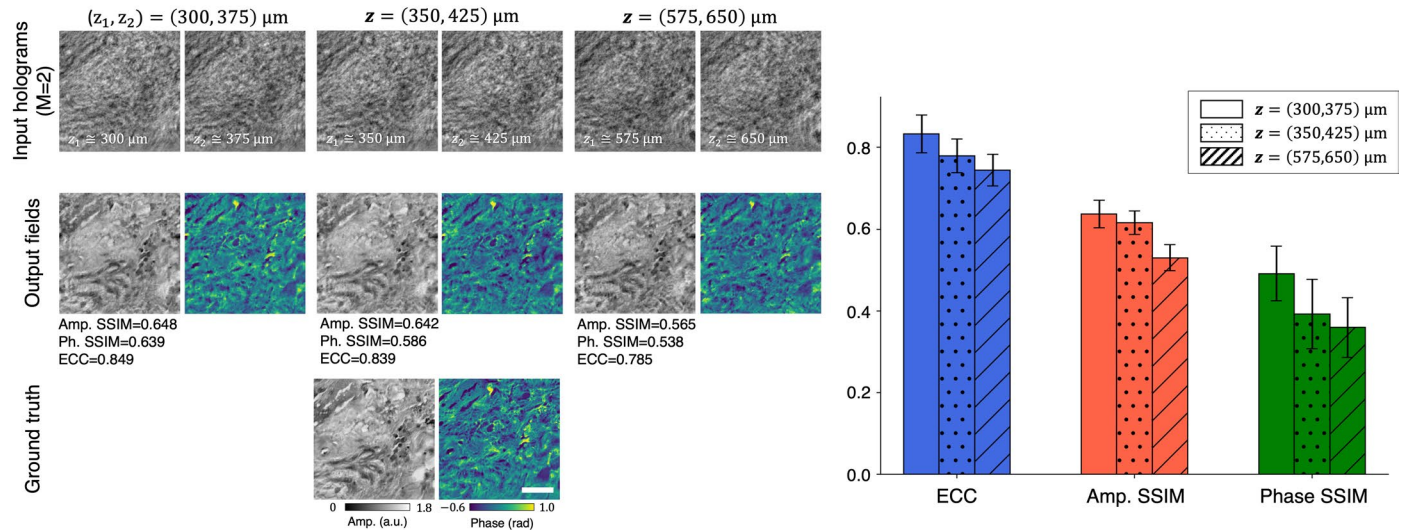
**Extended Data Fig. 1 | Training and testing pipelines, and the network architecture of GedankenNet.** (a) GedankenNet training using the artificial hologram dataset generated from random images. GedankenNet consists of a

series of spatial Fourier (SPAF) blocks and a long residual connection. (b) GedankenNet testing on experimental holograms of various human tissue sections and Pap smears. Scale bar: 50  $\mu\text{m}$ .

**(a) Pap smear**

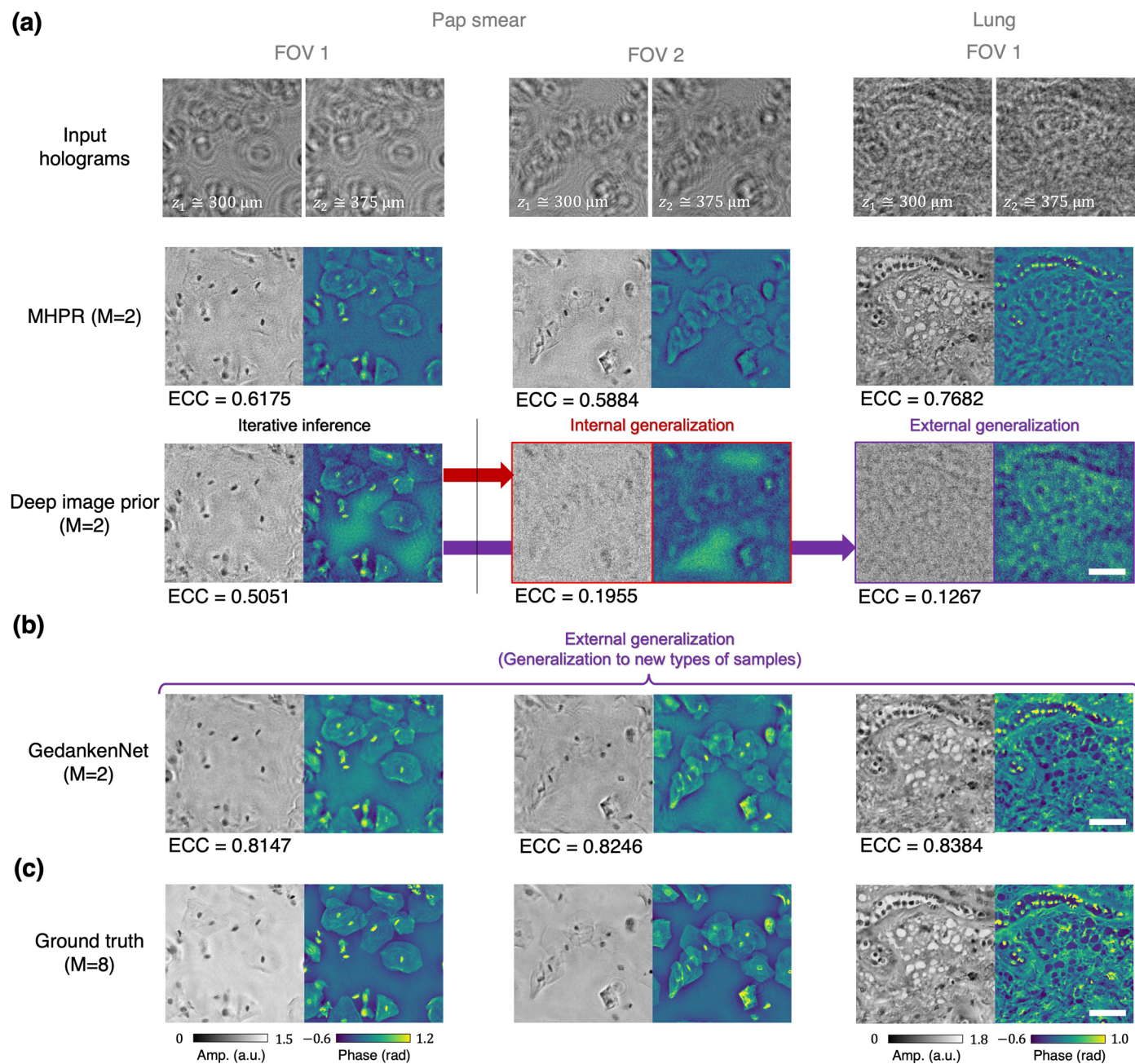


**(b) Lung tissue section**



**Extended Data Fig. 2 | Comparison of GedankenNet models trained with different sample-to-sensor axial distances.** Three GedankenNet models were trained on simulated holograms of the same artificial random complex fields propagated by  $(z_1, z_2) = (300, 375) \mu\text{m}$ ,  $(z_1, z_2) = (350, 425) \mu\text{m}$  and  $(z_1, z_2) = (575, 650) \mu\text{m}$ , respectively. **(a)** External generalization of these three GedankenNet models on experimental Pap smear holograms captured at different axial positions.

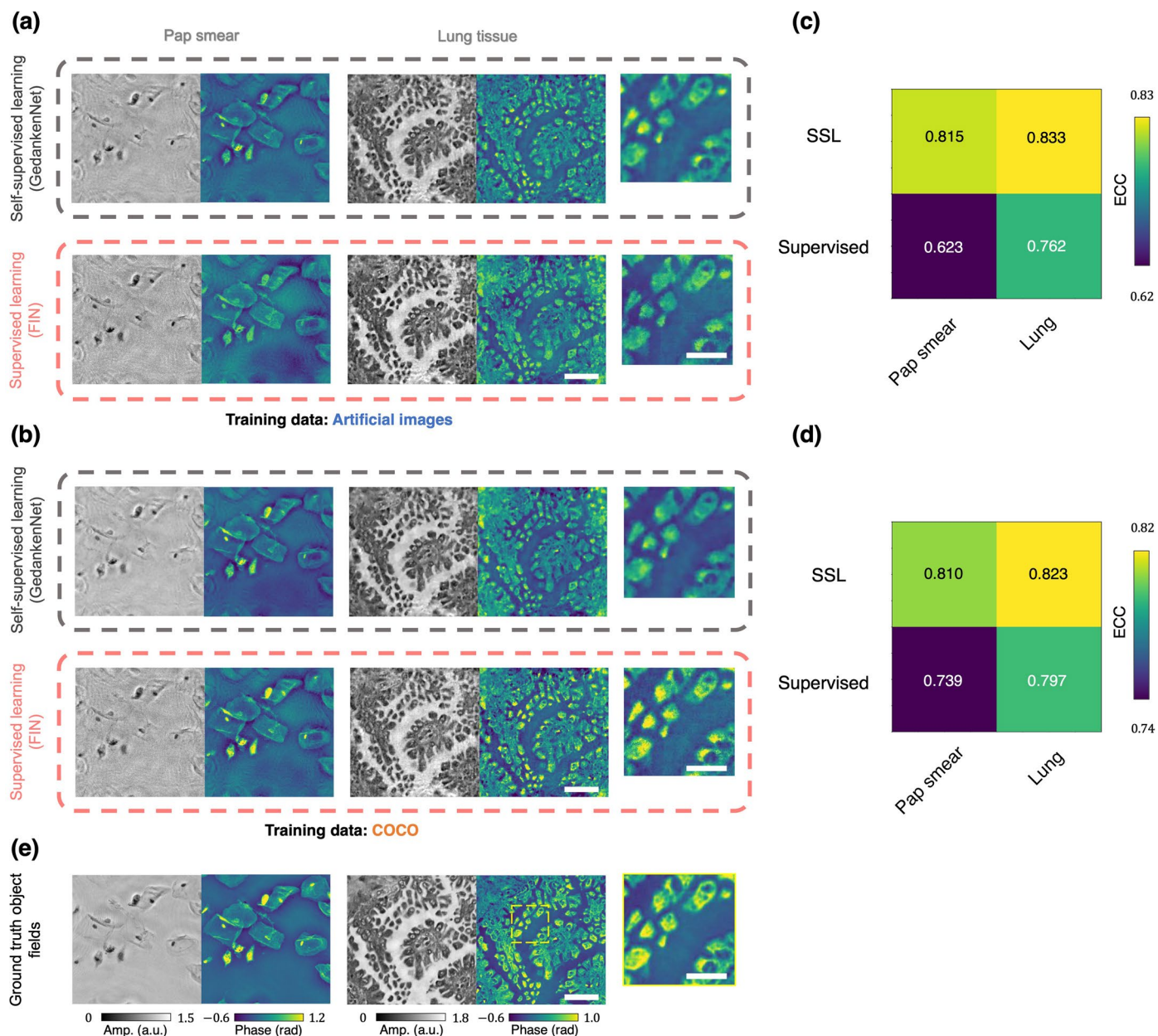
**(b)** External generalization of the three models on experimental holograms of human lung tissue sections. The ground truth object fields of each FOV were reconstructed by using 8 raw holograms. ECC mean  $\pm$  standard deviation values are presented and were calculated on Pap smear and lung test datasets with 47 and 94 unique FOVs, respectively. Scale bar: 50  $\mu\text{m}$ .



**Extended Data Fig. 3 | Comparison between GedankenNet and a deep neural network-based algorithm (DIP, that is, deep image prior).** (a) The reconstruction output of DIP on Pap smear FOV 1 after ~10 K iterations. The internal and external generalization of the converged DIP on the other two FOVs of Pap smear sample and lung tissue section. The input holograms (M=2) and the

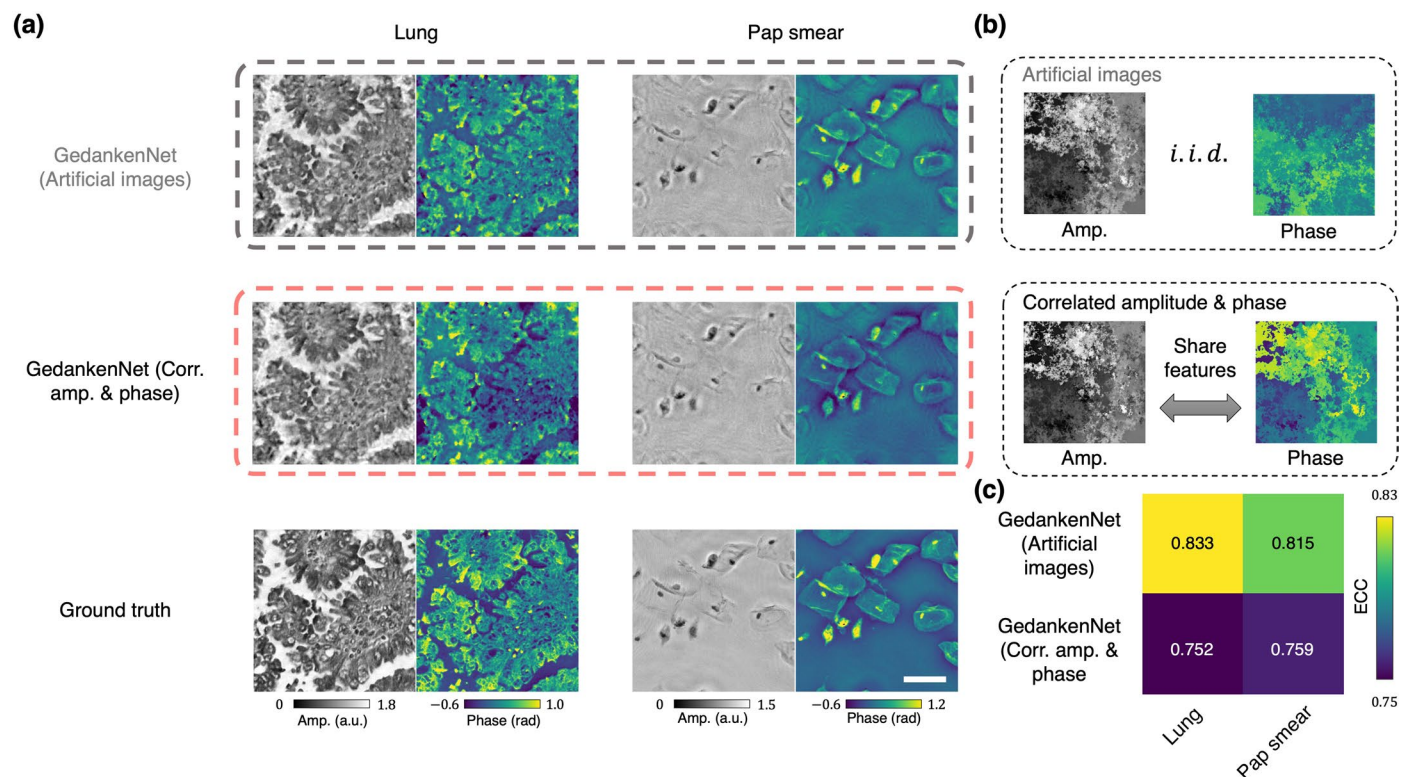
outputs of MHPR for each FOV are also shown for comparison. (b) External generalization results of GedankenNet (trained on artificial holograms of random synthetic objects) blindly tested on the same input holograms of each FOV. (c) The ground truth object fields were retrieved from 8 holograms of each FOV. Scale bar: 50  $\mu$ m.





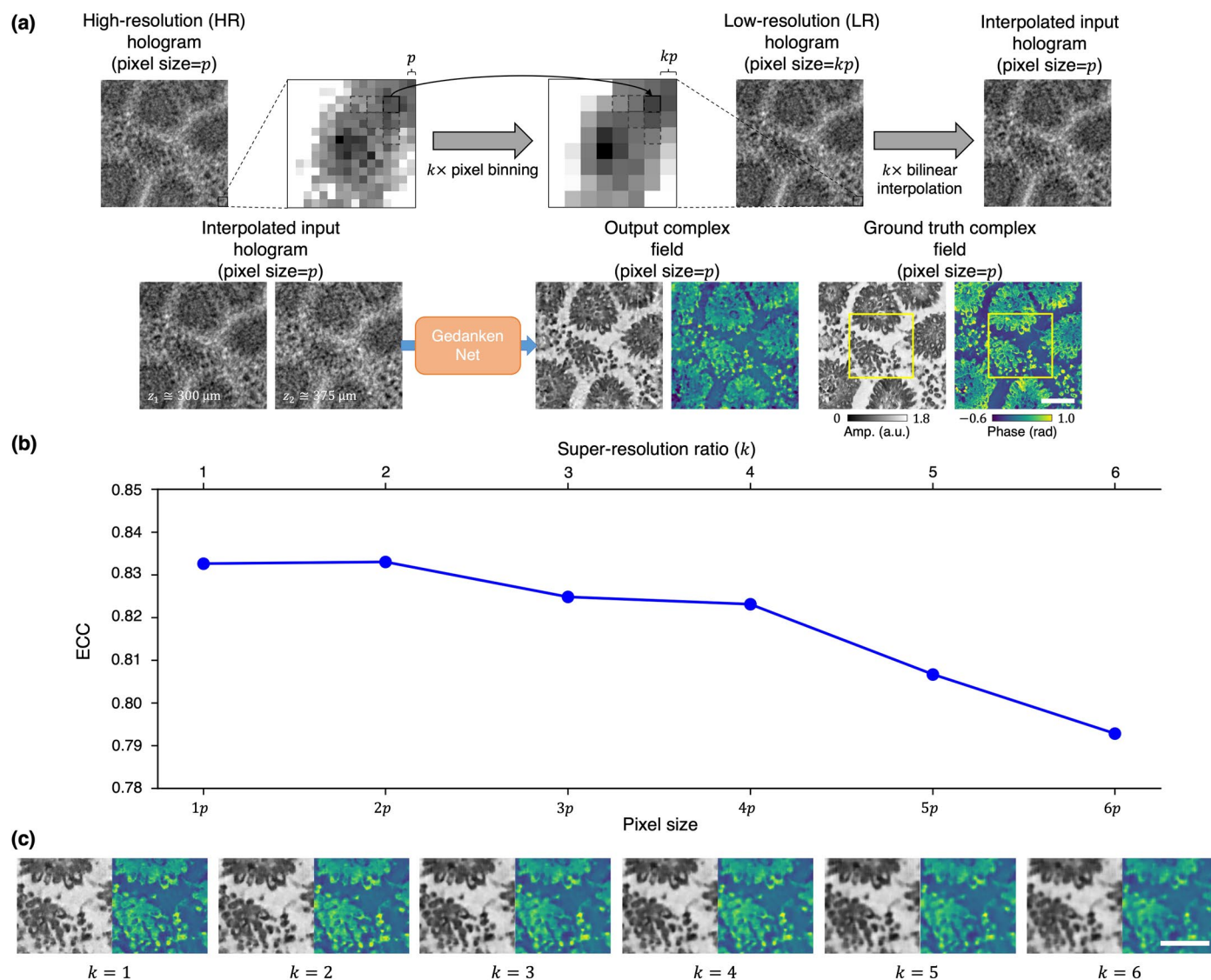
**Extended Data Fig. 4 | Generalization comparison between GedankenNet and a supervised learning model (FIN).** (a) GedankenNet and FIN were trained on the same simulated hologram dataset generated from random artificial images, and tested on experimental holograms of Pap smear and human lung tissue sections. (b) External generalization results of GedankenNet and FIN trained on the same simulated hologram dataset generated from natural images (COCO

dataset). (c, d) Mean ECC values of the outputs of the two models in (a, b) on the two testing datasets. The Pap smear and lung test datasets contain 47 and 94 unique FOVs, respectively. (e) The ground truth object fields of the testing FOVs were retrieved from 8 raw holograms of each FOV. Zoomed-in regions are highlighted by the yellow boxes. Scale bar: 50  $\mu\text{m}$  (20  $\mu\text{m}$  in the zoomed-in images).



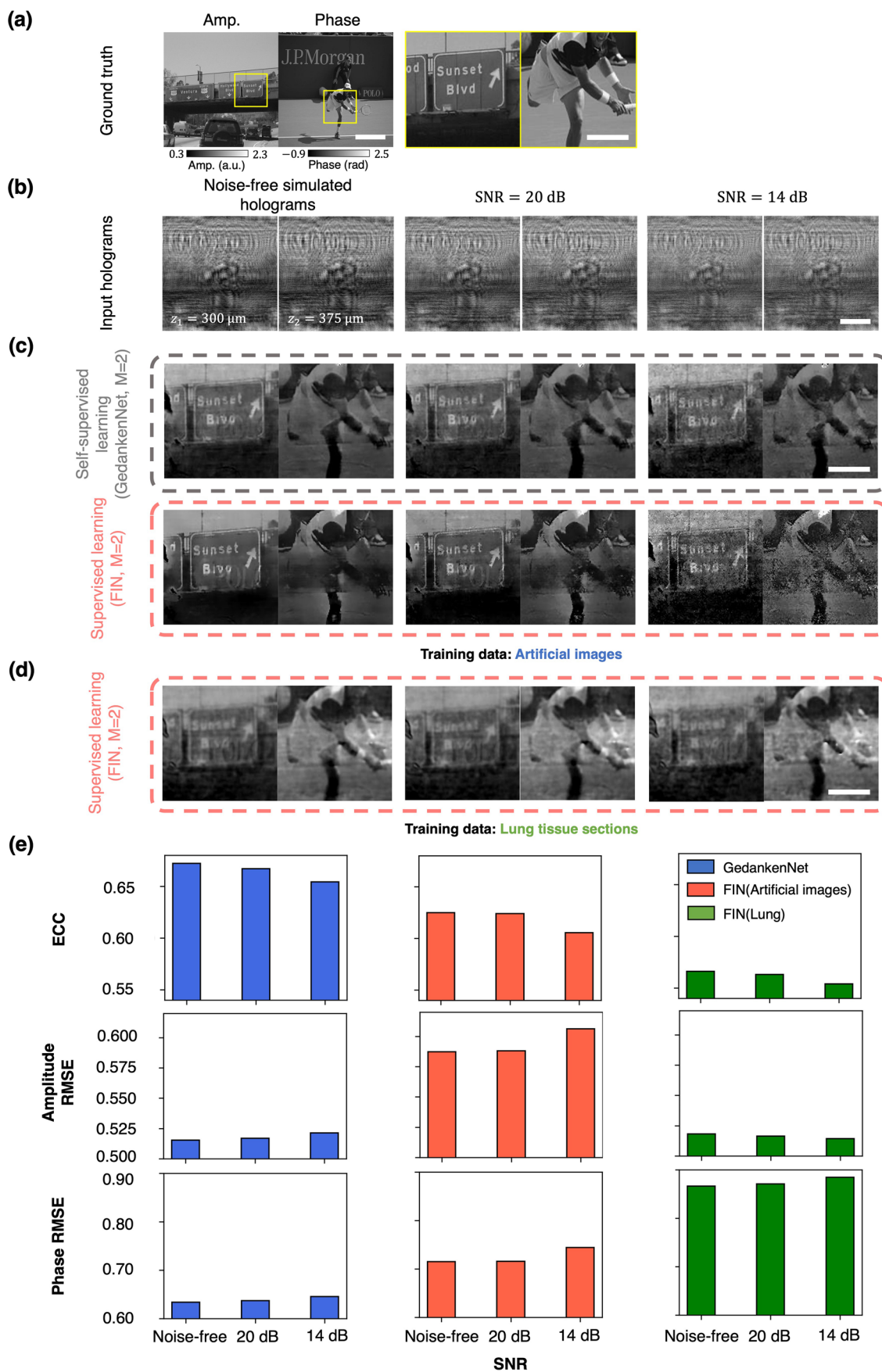
**Extended Data Fig. 5 | Comparison of the standard GedankenNet model trained using identically and independently distributed (i.i.d) random amplitude and phase profiles (synthetically generated without any experiments or resemblance to real-world samples) and a GedankenNet model trained with correlated amplitude and phase profiles. (a)** External generalization of the two models on experimental holograms of human lung tissue sections and Pap smear samples. Ground truth sample fields were

retrieved from  $M = 8$  raw holograms. **(b)** Schematic diagram illustrating artificial complex fields with i.i.d. amplitude and phase profiles, and artificial complex fields with correlated amplitude and phase profiles. **(c)** Quantitative evaluation of the reconstruction accuracy of the two GedankenNet models. ECC values were averaged on the lung test set with 94 unique FOVs and the Pap smear test set with 47 unique FOVs. Scale bar: 50  $\mu\text{m}$ .



**Extended Data Fig. 6 | Robustness of GedankenNet models to different pixel pitches.** (a) Pixel binning-based generation of lower-resolution (LR) holograms with a pixel size of  $kp$  from high-resolution (HR) holograms that have a pixel size of  $p$ . The LR holograms (after the resolution loss due to the pixel binning) were then interpolated and fed into a GedankenNet trained with pixel size  $p$ .

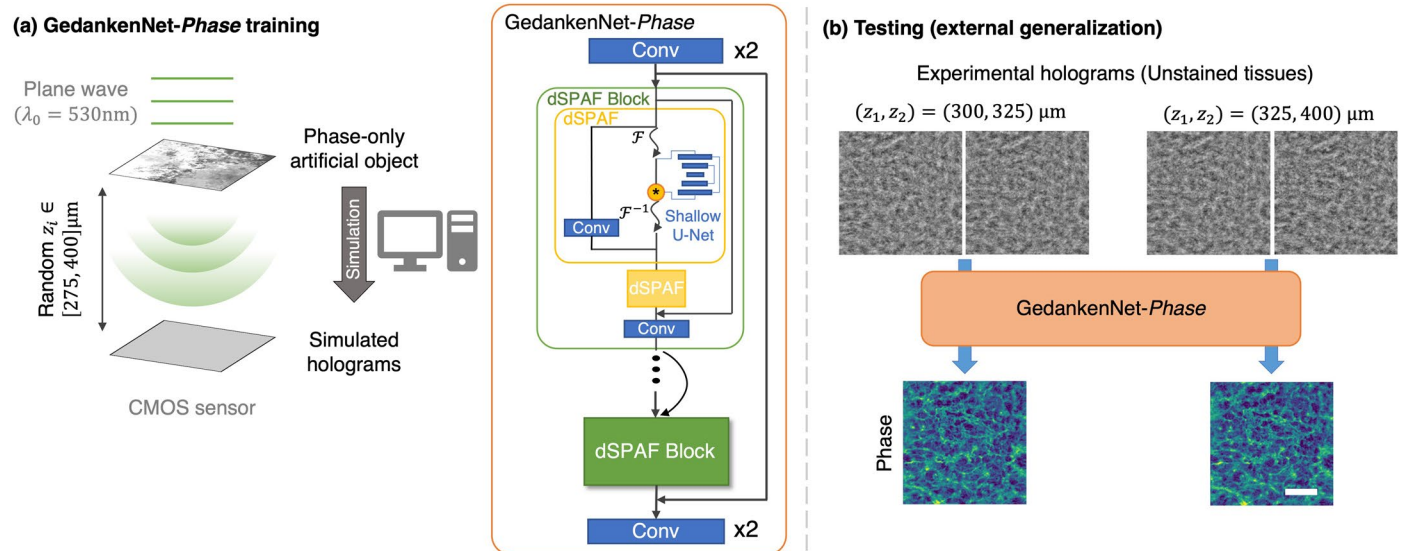
(b) Quantitative evaluation of the reconstruction accuracy of the GedankenNet with respect to the hologram pixel pitch. ECC values were calculated on the lung test dataset with 94 unique FOVs. (c) Zoom-in regions of the reconstructed sample fields by GedankenNet using LR holograms with various pixel pitches. Yellow squares highlight the zoom-in regions. Scale bar:  $50 \mu\text{m}$ .



Extended Data Fig. 7 | See next page for caption.

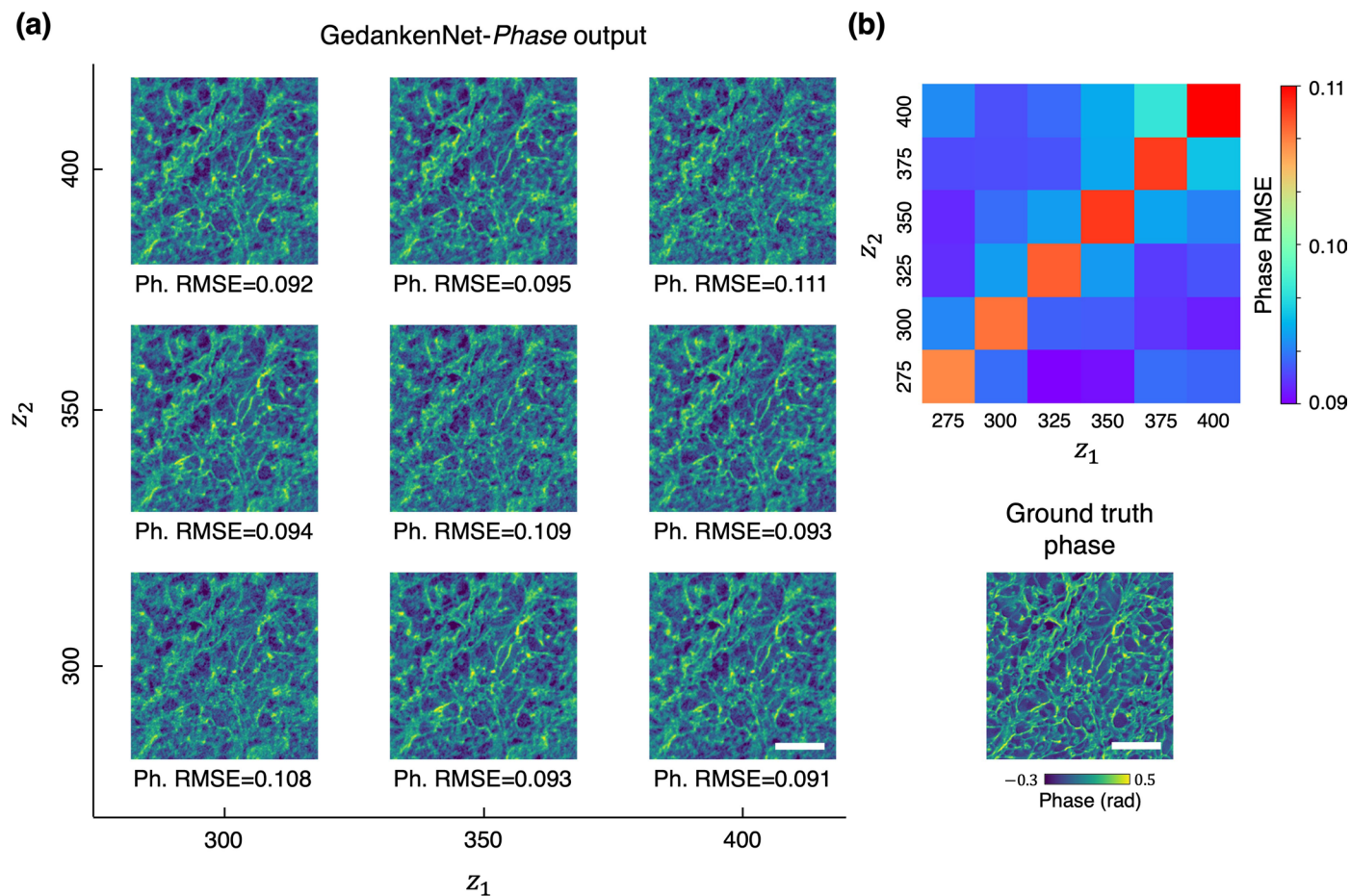
**Extended Data Fig. 7 | SNR comparison between GedankenNet and a supervised deep neural network FIN.** (a) The ground truth object fields and the zoomed-in regions marked by the yellow boxes. (b) Simulated holograms using natural images (from COCO dataset) with and without the presence of random noise. The noise-added holograms have 20 dB and 14 dB SNR, respectively. (c) The zoomed-in outputs of GedankenNet and FIN inference on the simulated test holograms shown in (b). Both models shared the same artificial hologram

training dataset (generated from random images) and the same network architecture. (d) The zoomed-in outputs of another supervised FIN model (trained using the experimental lung tissue holograms) blindly tested on the same simulated test holograms shown in (b). (e) Quantitative performance comparisons of GedankenNet and supervised FIN models. Metrics were averaged using 100 unique FOVs. Scale bar: 50  $\mu\text{m}$ , 20  $\mu\text{m}$  for zoomed-in images.



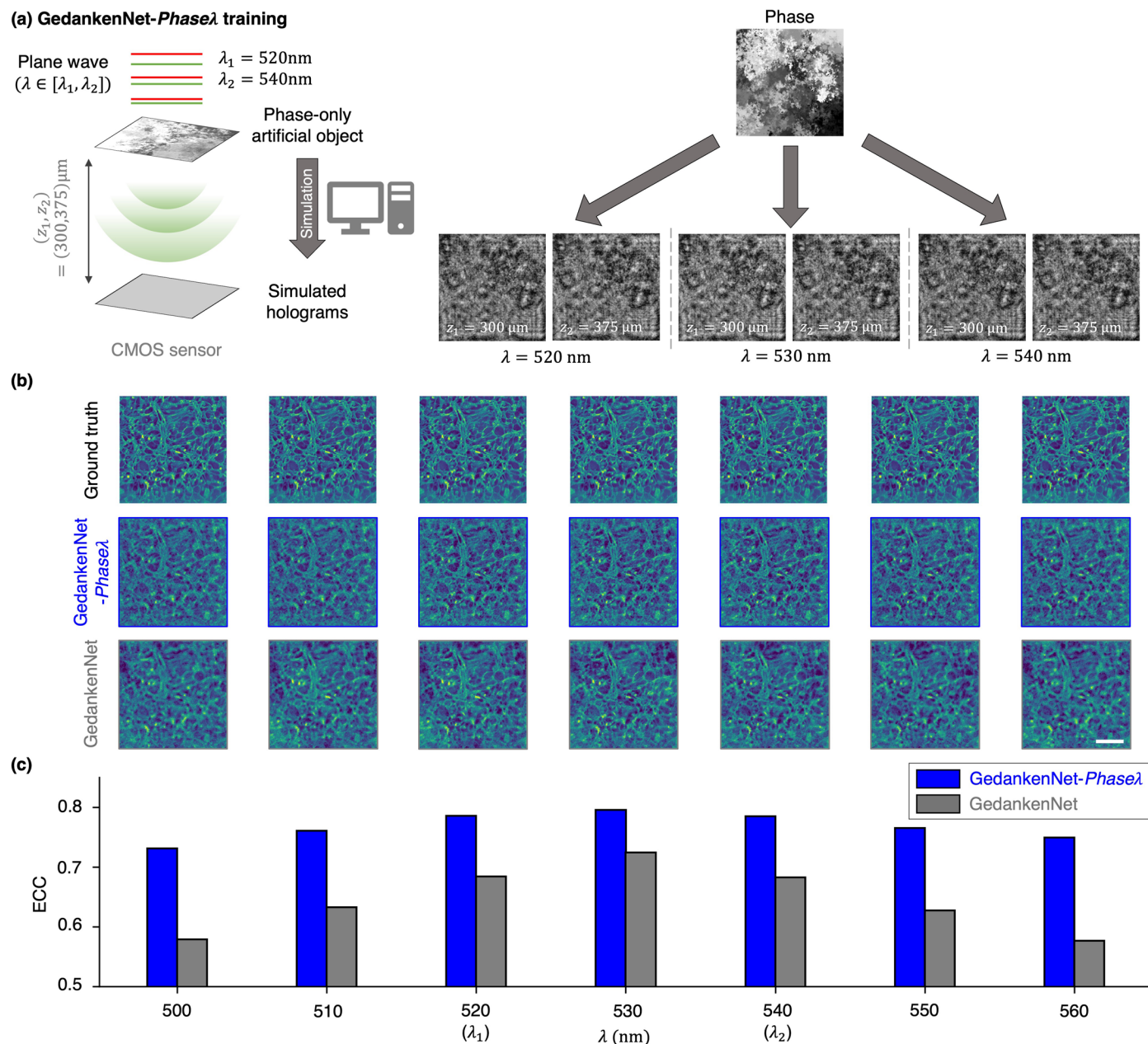
**Extended Data Fig. 8 | Training and testing workflow and the network architecture of GedankenNet-Phase.** (a) GedankenNet-Phase training used simulated holograms generated from random artificial phase-only objects at random sample-to-sensor distances within an axial range of  $[275, 400]\mu\text{m}$ .

GedankenNet-Phase is based on a DenseNet-like architecture consisting of a series of dSPAF blocks and dense links. (b) The trained GedankenNet-Phase can generalize to experimental holograms captured at random, unknown axial positions within the training range. Scale bar:  $50\mu\text{m}$ .



**Extended Data Fig. 9 | Autofocusing performance of GedankenNet-Phase for experimental holograms of unstained human kidney tissue sections captured at various sample-to-sensor distances. (a)** Reconstructed sample phase by GedankenNet-Phase. **(b)** Phase RMSE of the reconstructed phase images

by GedankenNet-Phase with respect to the ground truth. Phase RMSE values were averaged on the unstained kidney test dataset with 98 unique FOVs. The ground truth object fields were retrieved from 8 raw holograms. Scale bar: 50  $\mu\text{m}$ .



**Extended Data Fig. 10 | GedankenNet-Phase $\lambda$  training and generalization performance to various illumination wavelengths.** (a) The GedankenNet-Phase $\lambda$  model was trained on simulated holograms generated from random artificial phase-only objects using random illumination wavelengths within the spectral range  $[\lambda_1; \lambda_2] = [520\text{ nm}; 540\text{ nm}]$ . (b) External generalization of GedankenNet-Phase $\lambda$  (blue) to experimental holograms of unstained (label-free)

human kidney tissue sections captured with various, unknown illumination wavelengths. The results of the standard GedankenNet (gray, trained with  $\lambda_0 = 530\text{ nm}$ ) are also included for comparison. (c) Quantitative evaluation of the results of GedankenNet-Phase $\lambda$  (blue) and the standard GedankenNet (gray). ECC values were calculated on the unstained human kidney test dataset with 98 unique FOVs. Scale bar: 50  $\mu\text{m}$ .



## Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a                                 | Confirmed  |
|-------------------------------------|--|
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly  |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons   |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes  |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated  |

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

**Data collection** The software and algorithms for collecting and processing experimental data are detailed in the manuscript, with references to published literature. Experimental data were collected using a custom-designed holographic microscope. Raw holograms (~4.6K × 3.5K pixels) were recorded by a complementary metal-oxide semiconductor (CMOS) with a pixel size of 1.12 μm (IMX 081 RGB, Sony Corp.) using AYA software tool (Sony Corp.) and LabView 2012 (version 12.0).

**Data analysis** The codes for the deep learning models used in this work (written in Python 3.9.6 and PyTorch 1.9.0) can be accessed through: <https://github.com/PORPHURA/GedankenNet>  
The trained model and demo data are uploaded and available in the same code repository. The code for analyzing the results was written in Python using standard, open-source Python libraries.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

This study involved simulation data generated from the COCO dataset (available at <https://cocodataset.org/#home>) and the artificial image dataset, and experimental data of human samples. Simulation data used in the manuscript can be generated using the open-source code of the manuscript, public datasets and open-source Python libraries. The training artificial image dataset is available at: <https://github.com/PORPHURA/GedankenNet>. A portion of the experimental dataset corresponding to human tissue samples is also shared and referenced in the code repository without any links or identifiers to the patients, which is made available at: <https://github.com/PORPHURA/GedankenNet>. Additional data are available from the corresponding author upon reasonable request.

## Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

|                             |   |
|-----------------------------|---|
| Reporting on sex and gender | <a href="#">Sex and gender information of the human tissue slides was not collected/unknown, not used in this manuscript, and was irrelevant to this research.</a>  |
| Population characteristics  | All slides used in this manuscript were de-identified and prepared from existing specimens without links or identifiers to the patients. Therefore, this work did not interfere with standard practices of care or sample-collection procedures. The original tissue specimens had been archived before this work started, and had not been collected specifically for this research. |
| Recruitment                 | De-identified tissue slides were created before this work started and not collected specifically for this research.   |
| Ethics oversight            | Exempt: we used already existing, de-identified tissue samples collected before this work, without links or identifiers to the patients available.  |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

|                 |  |
|-----------------|--|
| Sample size     | To prepare the multi-height experimental hologram training dataset presented in the manuscript, we collected ~100K input-target pairs of stained human lung, prostate, salivary gland, kidney, liver and esophagus tissue sections, Pap smears, and unstained label-free human kidney tissue sections, each of which is of 512*512 pixels. The size of the experimental hologram training dataset was chosen as ~100K to match with the size of the artificial image training dataset. Empirically 100K training images are sufficient to reach a balance between the reported models' training speed and performance.<br>For the testing of the presented models, a subset of the lung, prostate, salivary gland slides from new patients and Pap smears were excluded from the training dataset and used as testing datasets, containing 94, 49, 49 and 47 unique 512*512 pixel field-of-views (FOVs), respectively. The holograms of the unstained (label-free) kidney tissue thin sections (~3-4 $\mu$ m thick) were used as our phase-only object test dataset containing 98 unique FOVs. The size of testing datasets was empirically determined so that it is sufficient to demonstrate the distribution of evaluation metrics of interest. |
| Data exclusions | The training FOVs and testing FOVs were strictly exclusive, and the testing FOVs were captured strictly from different sample slides from the training slides. The ground truth complex fields recovered from raw CMOS holograms (~4.6K*3.5K) were cropped and only the central regions (~4K*3K) were used for training and testing because of the missing side diffraction patterns out of the CMOS region.   |
| Replication     | The presented framework was validated by training on three training datasets, including artificial holograms generated from random synthetic images, natural images (COCO) and experimental holograms of stained tissue sections. After training, all the models were blindly tested on multiple testing datasets, including experimental holograms of lung, prostate, salivary gland slides from new patients and Pap smears, containing 94, 49, 49, and 47 unique FOVs. The trained models were also blindly tested on artificial holograms generated from 100 testing random images and 100 testing COCO natural images. The models for phase-only objects were tested on the testing dataset of unstained, label-free human kidney tissue sections, containing 98 unique FOVs.   |

The above model training and testing process was also repeated for four times with each selection of the number of input holograms (M). The ensemble models were evaluated and reported in Fig. 2.

## Randomization

The training, validation and testing image datasets were randomly partitioned. The artificial images used for training were randomly generated using a Python package (randimage) as described in the manuscript. The training process randomly selected mini-batches for optimization.

## Blinding

All the testing results of artificial and experimental holograms generated by the trained neural networks were blindly performed on new fields-of-view excluded from the training dataset. The blind testing experimental holograms were also captured on new slides from new patients that did not appear in the training dataset.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

| n/a                                 | Involvement in the study                               |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies                    |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms   |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data                 |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern  |

### Methods

| n/a                                 | Involvement in the study                        |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |