Holographic reconstruction enhancement via unpaired image-to-image translation

Romane Scherrer, Thomas Quiniou, Thierry Jauffrais, Hugues Lemonnier, Sophie Bonnet, Nazha Selmaoui-Folcher

Applied Optics
2022, Volume 61 Pages 9807-9816
https://doi.org/10.1364/AO.471131
https://archimer.ifremer.fr/doc/00804/91576/

Holographic reconstruction enhancement via unpaired image-to-image translation

Scherrer Romane [1, *], Quiniou Thomas [1], Jauffrais Thierry [2], Lemonnier Hugues [2], Bonnet Sophie [3], Selmaoui-Folcher Nazha [1, *]

[1] ISEA, Université de la Nouvelle-Calédonie, Nouméa, New Caledonia
[2] Ifremer, UMR9220 Entropie, Nouméa, New Caledonia
[3] Aix Marseille University, Université de Toulon, CNRS, IRD,MIO,Marseille, France

* Corresponding authors : Romane Scherrer, email address : romane.scherrer@hotmail.fr ; Nazha Selmaoui-Folcher, email address : nazha.selmaoui@unc.nc

Abstract :

Digital holographic microscopy is an imaging process that encodes the 3D information of a sample into a single 2D hologram. The holographic reconstruction that decodes the hologram is conventionally based on the diffraction formula and involves various iterative steps in order to recover the lost phase information of the hologram. In the past few years, the deep-learning-based model has shown great potential to perform holographic reconstruction directly on a single hologram. However, preparing a large and high-quality dataset to train the models remains a challenge, especially when the holographic reconstruction images that serve as ground truth are difficult to obtain and can have a deteriorated quality due to various interferences of the imaging device. A cycle generative adversarial network is first trained with unpaired brightfield microscope images to restore the visual quality of the holographic reconstructions. The enhanced holographic reconstructions then serve as ground truth for the supervised learning of a U-Net that performs the holographic reconstruction on a single hologram. The proposed method was evaluated on plankton images and could also be applied to achieve super-resolution or colorization of the holographic reconstructions.

doi.org/10.1364/

1 Introduction

The hologram, which encodes the 3D information of a sample into a single 2D image, is recorded on a camera and then digitally processed to find the phase and amplitude of the objects. Since all the image processing steps (e.g., focusing, segmentation) are performed after the holograms acquisition and that the imaging device can be compact [1], digital holography has gained growing interest in many fields of application such as 3D particle tracking [2, 3], tomography [4, 5] or label-free cell analysis [6]. However, the hologram, which is created by the superposition of an unaffected reference wave and the wave diffracted by the object, must be numerically reconstructed before detecting, counting or analysing the objects of interest.

The holographic reconstruction is composed of several steps which depend on the holographic device architecture (in-line or off-axis) and the type of imaging technique (static or dynamic) but usually includes the focusing and the phase recovery of the hologram. These two steps, which are iterative, require the back-propagation of the hologram at several planes in order to calculate certain metrics to search for the focus or to recover the phase of the object.

To accelerate the processing of holograms, several studies have been successfully carried out to perform the holographic reconstruction directly from a single hologram using deep learning models [7, 8, 6]. In practice, most of the proposed methods have one point in common: the models are trained in a supervised setting to transform a raw or de-focused hologram (input) into an amplitude and/or phase image of the object (output) [9]. For this type of learning, the preparation of a large, paired and artifact-free dataset to train the models is challenging and requires considerable human and technical efforts. Indeed, as the model learns a mapping function between inputs examples and the target images, its outputs on unseen data once trained will have a visual quality and resolution at best similar to that of the target images in the training dataset. As a result, acquiring high-quality target images is often a prerequisite to obtain a robust and efficient model.

To obtain clean and high-quality target images, the first strategy consists in improving the acquisition process of the holograms (multi-height phase recovery [10], super-resolution [11], ...) and/or in optimizing the holographic device (smaller pixel size, increased coherence of the light source,...). However, theses approaches, which further complicate the acquisition and processing of holograms, do not always guarantee the good quality of the reconstructed images, which can be out-of-focus, noisy or can be deteriorated by interference fringes that are difficult to compensate (twin image or out-of-focus objets for e.g.).

To overcome the difficulty of obtaining a clean dataset to train a model, some recent methods rely on the use of a generative adversarial network (GAN). In [12], brightfield microscopy images are used as ground-truth and a GAN learns to transform a hologram into a brightfield-like image. However, a paired and aligned dataset is needed, which requires a great effort of co-registration between the holograms and the corresponding brightfield images that are acquired with two different microscopy imaging modalities. In [13], a CycleGAN [14] is trained with unpaired images to transform a hologram into a phase image. This scenario is particular because the holograms are generated by a spatial light modulator (2D sample) and the distribution of the target images are known beforehand. More recently, [15] demonstrated that it is possible to transform a hologram into brightfield-like images by training a CycleGAN with unpaired

data but the generator tends to output brightfield-like objects with random focal representation.

These recent works have demonstrated that the use of a GAN can overcome the problem of low-quality reference images or even the absence of the latter. By using images acquired by another imaging modalities or during a previous experiment, it is possible to considerably improve the visual quality of the images reconstructed by deep learning models. However, an important limitation of GANs, and particularly of CycleGAN, is that the model learns one-to-one mappings, i.e. the model associates each input image with a single output image. Applied to images that encode 3D information, such as a hologram or a CT scan, CycleGAN can generate slice-to-slice inconsistencies [16] or random focal representations [15]. In the case of holography, it is desirable to have a model that learns many-to-one mappings as multiple defocused holograms of the same region of interest (ROI) would still correspond to the same reconstructed image. Another limitation of GANs is their number of parameters, which can be substantial and make inference computationally expensive and memory-intensive. As a results, it can be complicated to deploy it on low-power edge devices.

In this paper, we propose a novel approach that aim to solve the problem of random focal representation and that can be used to obtain a small model with an optimized structure for edge computing. Unlike the previously mentioned methods, we propose to train 2 models in a Teacher-Student setting, which together are able to learn many-to-one mappings between two microscopy imaging modalities. The first model, a CycleGAN, is the Teacher model and learns one-to-one mappings between unpaired holographic reconstructed images and brightfield microscopy images. Once trained, the Teacher model translates the high-noise and low-resolution holographic reconstructed images into low-noise and high-contrast brightfield-like images that are used as target images to guide the second model training. The Student model, a small Unet, is trained with paired data (holograms/brightfield-like images) that includes several defocused holograms paired up with the same target image (many-to-one mapping) provided by the Teacher model. Our framework is demonstrated experimentally using a species of diatom that was imaged with a lensless holographic device and a bright-field microscope. The results show that our method, which takes advantage of unpaired images from another microscopy modality, allows to obtain holographic reconstructed images of better quality than those obtained with a model that is trained in a directly supervised way with an aligned dataset. Moreover, the final Student model trained with our framework has fewer parameters than the CycleGAN and is robust in its holographic reconstruction. The output images are not compromised by distortion or artifacts unlike those that could be obtained with the CycleGAN and the model is able to perform many-to-one mappings which solves the problem of random focal representation.

2 METHODS

2.1 Holography principle

The principle of holography is based on the recording of the fringe pattern resulting from the superposition and the interference of two light waves. Mathematically, these two waves, often referred to as reference R and object O, are expressed respectively by:

$$R(X,Y) = A_R(X,Y)e^{-i\phi_R(X,Y)} \tag{1}$$

$$O(X,Y) = A_O(X,Y)e^{-i\phi_O(X,Y)} \tag{2}$$

where $A(X,Y)$ and $\phi(X,Y)$ are the amplitude and phase distributions, respectively. At the imaging plane, where a CCD camera can be positioned and which is placed at a distance d from the object to be observed, the hologram H is expressed by:

$$\begin{aligned} H_d(X,Y) &= |R(X,Y) + O(X,Y)|^2 \\ &= |R|^2 + |O|^2 + RO^* + OR^* \end{aligned} \tag{3}$$

where * represents the complex conjugate. The holographic reconstruction, which aims to recover the original object from its hologram, is often composed of two steps. The first step is the autofocusing during which the object-sensor distance d is searched. This step often requires several back-propagation of the hologram at different planes in which different sharpness criteria can be calculated on the resulting complex-valued image (Gini Index [17], Tamura Coefficient [18], AMP [19], SoG [20],...). The second step is the phase recovery which consists in the reconstruction of the phase distribution of the wavefront at the imaging plane. This iterative step is often based on the Gerchberg-Saxton algorithm [21] and requires the acquisition of several holograms of the same object with, for example, different wavelengths [22], different object-sensor distances [10] or different angles [23].

Although digital holography offers several advantages over other microscopy imaging modalities (such as a wide field-of-view [24] or a the possibility to obtain a compact imaging device [25, 26]), the relatively complex imaging set-ups and algorithms that are required to perform the holographic reconstruction have limited its suitability for many applications. Even-though many deep-learning based-methods have emerged to simplify the holographic image reconstruction steps, the constitution of an aligned and artifact-free dataset to train those models remains a challenge, especially when the reconstructed images are hard to acquire and can have a deteriorated quality.

Several factors will impact the overall quality of the holographic reconstructed images. The first ones are related to physical constraints and mainly depends on the configuration and parameters of the holographic device. For example, the spatial and temporal coherence of the light source, the absence of lens and the pixel size of the camera limit the final contrast and resolution of the images [27]. The others factors are related to the algorithms used to process the holograms (autofocusing, phase recovery routine or co-registration) whose performances and robustness may depend on the nature of the object to be observed [20] (amplitude-contrast object, phase-contrast object, sparse or dense sample,...) and the execution mode (automatic or supervised by expert). Motivated by the fact that it is cumbersome and not always feasible to obtain a large and high quality holographic dataset, we propose to use images acquired by another microscopic modality which already have the desired properties in term of contrast or resolution and that might be easier to acquire and process.

2.2 Proposal of a new method

The objective of our study is to obtain a small model that is able perform the image reconstruction from a single hologram. The

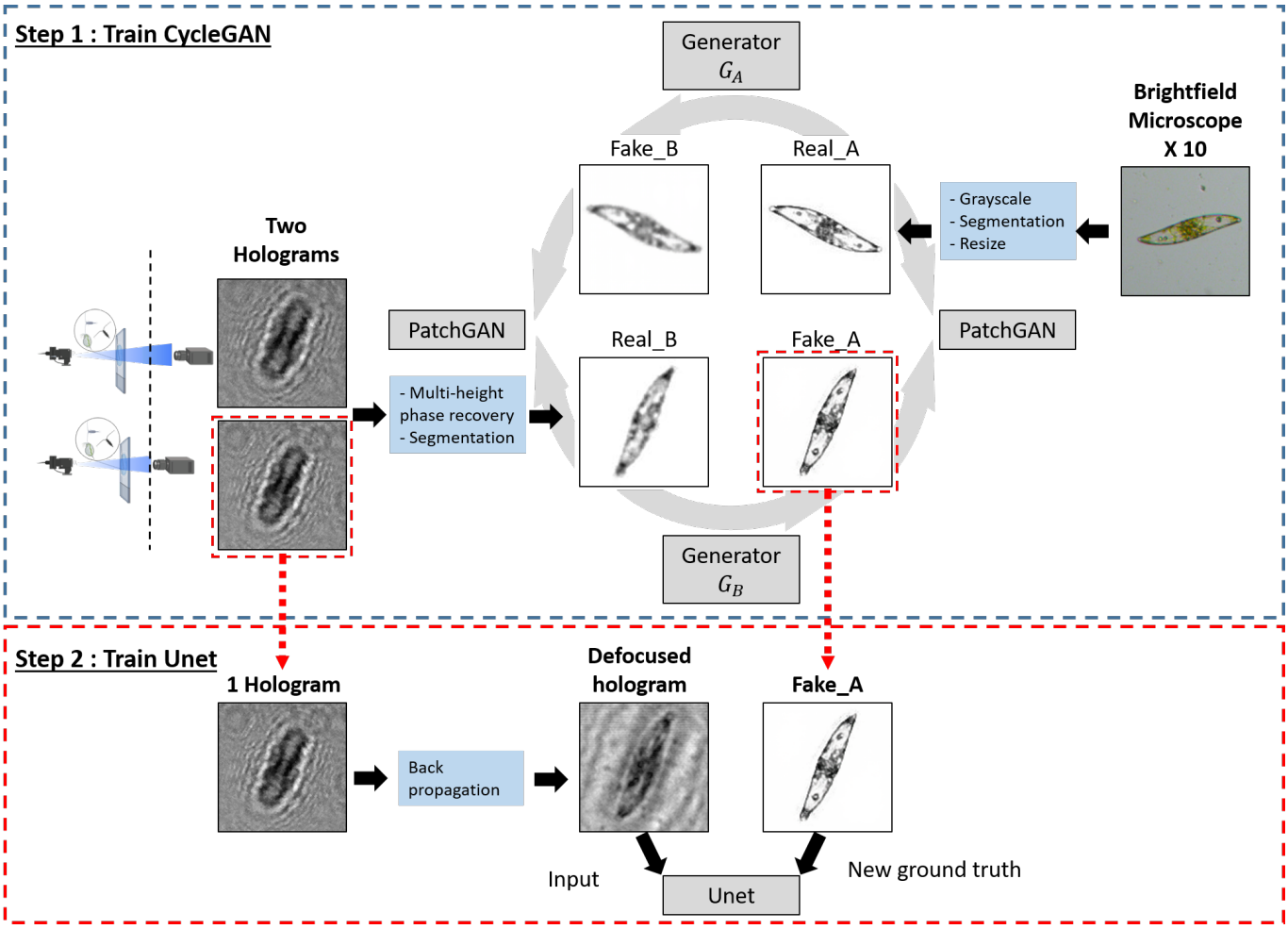

Figure 1: General diagram of our 2-step method. A CycleGAN is trained to generate bright-field like images that are then used as ground-truth to train a second model.

output of the trained model should be similar to the reconstructed in-focus amplitude images obtained after a conventional phase retrieval/auto-focusing algorithms but must have a better visual quality, with notably less noise and more contrast. However, this task is not feasible if the training of the model is carried out in a supervised setting with a paired holograms/reconstructed images dataset that has low quality images as ground-truth images. To improve the quality of the holographic reconstructed images without complicating the experimental setup to acquire holograms, we propose to use microscopic images acquired by another imaging modality which already possess certain desired properties in terms of resolution, contrast and noise reduction.
Our method, depicted in Fig. 1, is composed of two steps. The first one aims at enhancing the visual quality of the holographic reconstructed images which will then serve as ground-truth to train a second and smaller model in a supervised way. For this step, a cycle generative adversarial network [14] (CycleGAN) is trained with unpaired brightfield microscopy images (domain A) and holographic reconstructed images (domain B) obtained by a conventional phase retrieval algorithm to remove the twin image artifact.
The CycleGAN is composed of two generators that translate the images from one domain to another and two discriminators that determine if the generated images are structurally identical to the true images of domains A and B. The 4 models are trained simultaneously in a minimax game and learn 2 mapping functions between domains A and B. One generator $G_A : A \rightarrow B$ learns to translate an image $a \in A$ into a fake image $G_A(a)$ that fool a discriminator D_A. The discriminator goal is to determine if the image $G_A(a)$ is a true image from domain B or a generated one. This opposite objective is expressed by the adversarial loss that is used to match the distribution of generated images to the data distribution in the target domain :

$$\mathcal{L}_{GAN}(G_A, D_A, A, B) = \mathbb{E}_b[\log D_A(b)] + \mathbb{E}_a[\log(1 - D_A(G_A(a)))] \quad (4)$$

Similarly, the second generator $G_B : B \rightarrow A$ learns to translate an image $b \in B$ into a synthetic image $G_B(b)$ that is similar to the images from target domain A and fools the discriminator D_B. During training, the cycle consistency loss is minimize to ensure that any input image and output image (generated) are paired up in a meaningful way by reducing the number of possible

mappings between domains A and B:

$$\begin{aligned}\mathcal{L}_{cyc}(G_A, G_B) = \mathbb{E}_a[\|G_B(G_A(a)) - a\|_1] \\ + \mathbb{E}_b[\|G_A(G_B(b)) - b\|_1]\end{aligned} \tag{5}$$

Finally, the full loss can be written as :

$$\begin{aligned}\mathcal{L}(G_A, G_B, D_A, D_B) = \mathcal{L}_{GAN}(G_A, D_A, A, B) \\ + \mathcal{L}_{GAN}(G_B, D_B, B, A) \\ + \lambda\mathcal{L}_{cyc}(G_A, G_B)\end{aligned} \tag{6}$$

where λ controls the relative importance of the two objectives. The CycleGAN is trained to solve :

$$G_A^*, G_B^* = \arg\min_{G_A,G_B}\max_{D_A,D_B}\mathcal{L}(G_A, G_B, D_A, D_B) \tag{7}$$

At the end of the first step of our method, the degraded holographic reconstructed images are transformed into better quality images $G_B(b)$ that should have similar characteristics to bright-field microscopy images (domain A). Note that the second CycleGAN generator G_A, which translates the bright-field images into a fake holographic reconstructed image, is not used for the remainder of our framework.

The second step of our method consists in the supervised learning of a small Student model that performs the image reconstruction from a single hologram. To train the model, we build a new dataset in which the fake brightfield images generated by the CycleGAN generator G_B serve as ground truth. Note that even though the CycleGAN is trained with unpaired images (i.e. the images from domains A and B are independently acquired by two different microscopy modalities on distinct samples), the generated bright-field like images $G_B(b)$ are aligned with the reconstructed image b. Since b are already aligned with their corresponding holograms, the resulting paired hologram/fake bright-field images dataset is perfectly aligned without the need to perform a co-registration step. During the second model training, the raw holograms are first back-propagated close to focus plane and the amplitude images are used as inputs. This step is performed for two reasons. First, it allows to increase the number of images in the dataset by coupling it with a classical data augmentation routine (random cropping, rotation, flip). More precisely, a single raw hologram can be back-propagated to an infinite number of planes to generate defocused holograms. We chose to randomly back-propagated every raw hologram at 30 different planes within $\pm 100\ \mu m$ from the correct global focus plane. Secondly, we used defocused holograms as inputs to ensure that the trained model is robust in its holographic reconstruction regardless of the axial defocus distance (many-to-one mappings). At the end of step 2, the trained Student model should be able to transform a defocused hologram into a enhanced and segmented amplitude image.

2.3 Models Architecture

2.3.1 CycleGAN for step 1

The CycleGAN is composed of two generators and two discriminators. Both generators have the same U-Net [28] structure. The first part of the generator, the encoder, is composed of 7 convolutional layers with a kernel size of 4 and a stride of 2. Each convolution layer is followed by a leaky-ReLu activation function and a batch normalization. The number of filters for each convolution layer is {64,128,256,512,512,512}. The second part of the U-Net, the decoder, is composed of 7 transposed convolutional layers with a kernels size of 4, a stride of 2 and a number of filters set to {512,512,512,256,128,64,3}. Skip connections are added after each transposed convolutional layer in order to concatenate the feature maps of the decoder part with those of the encoder. The ReLu activation function is used in the decoder part. The two discriminators have a patchGAN [29] structure and are composed of 5 convolutional layers with a number of filters set to {64,128,256,512,1}, a stride of {2,2,2,2,1,1} and a kernel of size 4. The leaky-ReLu activation function is used for all convolutions except the last one which is linear.

2.3.2 U-net for step 2

For the second step of our method, we trained a Student Unet-like model that has fewer parameters than the CycleGAN generator (4M VS 41.8 M) and has a slightly different architecture. The model is composed of 4 convolutional blocks and 4 transposed convolutional blocks in its encoder and decoder parts, respectively. Each block consists of 3 convolution layers (resp. transpose convolution) followed by a normalization batch layer and the ReLu activation function. The first two convolutional (resp. deconvolutional) layers in each block have the same number of filters and the last one doubles (resp. halves) the number of feature maps. The downsampling and upsampling in the endoder and decoder are provided by a Maxpooling and Upsampling layer, respectively. The skip connections between the encoder and decoder parts are Add layers that sum the feature maps instead of concatenating them. The last convolution layer is activated with a sigmoid so that the output images have pixels values between 0 and 1. The Unet-like model is trained to minimized the binary cross-entropy loss which is expressed as :

$$\mathcal{L}_{BCE} = -\frac{1}{N}\sum_{i=1}^{N} y_i \log(p_i) + (1 - y_i)\log(1 - p_i) \tag{8}$$

where N represents the total number of images, y is the reference image and p is the predicted image.

2.4 Data acquisition and preparation

To evaluate our method, we imaged a plankton species from New Caledonia (*Pleurosigma sp.*) with a brightfield microscope and a holographic device. All the images were acquired independently on different samples. The plankton was imaged with a brightfield microscope at a $\times 10$ magnification. The images were automatically segmented with a Sobel-based edge detection algorithm and converted into grayscale images. The images have been resized so that the effective size of the pixels is identical to that of the holograms. The holograms were acquired with an experimental digital in-line holographic microscopy setup. Our lensfree system is composed of a green LED ($520nm$) that is coupled to a $100\mu m$ pinhole to increase its spatial coherence. The plankton sample is placed at approximately 10 cm away from the partially coherent light source and $\leq 1mm$ away from the camera (IMX219). Since the sample is close to the sensor surface of the camera, the imaging field-of-view is equivalent to the sensor area ($\approx 10mm^2$) and the holograms are acquired with unit optical magnification. Note that the student network *only* uses the fake brightfield images generated by the CycleGAN as ground-truth during training. As a result, the input images (B-domain) used

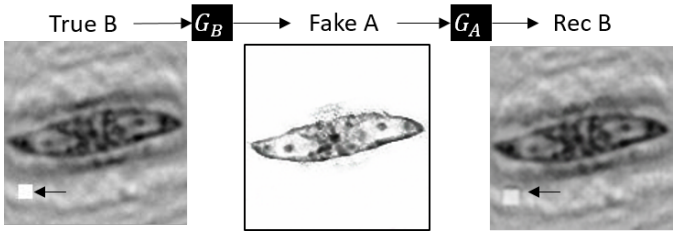

Figure 2: Generator G_B encoded the background information of $TrueB$ inside the object support of $FakeA$ so that G_A can recover the background. The arrows indicates where the white square was added prior to inference.

to train the CycleGAN and the Unet can be a different representation of the holographic images (raw or in-focus/defocused backpropagated hologram for example). We conducted preliminary studies (that can be found in the Supplementary Document) and trained several CycleGANs with different B-domain images to find the one that outputs the most realistic fake brightfield images. We found that CycleGANs trained with backpropagated (infocus or defocused) holograms have a unstable training (Fig.S1) and tend to produce images with artifacts due to the cycle consistency that is enforced at the pixel level (Fig.S2). For the hologram → brightfield image translation, the background information around the objects and in particular the twin image and noise must be encoded in the fake bright field image to ensure the cycle consistency. However, since the fake brightfield images are segmented, the background information is somehow encoded in the pixels inside the object support which cause unrealistic artifacts (see Fig.2 where a trained G_B encodes a white square that we added in the background). To correct this flaw, we chose to carry out a phase recovery routine on the holograms before backpropagated them at the correct focal plane and segmenting them to retain only the amplitude image of the plankton. This process ensure that there is no background information in the B-images and that the images are in-focused. To carry out the recovery phase, we acquired 2 holograms of the same sample by changing the object/camera distance (by adding a coverslip between the sample and the camera). The holograms are then processed as follows: First, the two holograms of the same sample were backpropagated to their respective focus plane that was automatically determined with the calculation of the Gini Index. Then, an affine transformation was build to co-registered the two holograms. Indeed, during the holograms acquisition of different heights, lateral shifts and rotations are unavoidable because the placement of the sample after changing the height is not exactly the same. This step, which is essential before the phase recovery, was carried out by an automatic algorithm (Surf [30]) that determines common control points in the holograms before calculating the affine transformation to compensate for the deformations. Finally, the dual-planes phase recovery is performed using an iterative Gerchberg–Saxton based algorithm [31]. The method starts from the first intensity measurement (ie hologram #1 denoted H_1) and assumes a complex wavefront $E_1 = \sqrt{H_1}\exp i\phi_1$ where the phase distribution ϕ_1 is initialized to zero. The wavefront is then numerically propagated to the other hologram plane to obtain the complex wavefront there as $E_2 = A_2\exp i\phi_2$. Then, the amplitude is replaced by the second intensity measurement (i.e. hologram #2 denoted H_2) $A_2 = \sqrt{H_2}$ while the phase distribution remains unchanged. The updated wavefront E_2 is then numerically backpropagated to the first hologram plane where the amplitude is again replaced by the intensity measurement and the phase is retained. The algorithm continues to numerically propagate the complex wavefront back and forth among the two heights until the mean squared error between $|E_1|^2$ and H_1 is smaller than a set threshold. Once the process terminates, the phase of each lensfree hologram is recovered and the first hologram is backpropagated to the object plane. The amplitude images of the objects are then segmented to remove the background and the detected regions of interest are saved to constitute B-domain image for the CycleGAN.

2.5 *Evaluation metrics*

Several metrics are used to evaluate the performance of the models.

To evaluate the CycleGAN outputs on unpaired data, the Frechet Inception Distance (FID) was used. This metric, proposed by [32] was specially developed to evaluate the performances of generative adversarial networks. Since GANs are composed of several networks that are playing against each other in a zero-sum game, the generators must be evaluated during training to identify the best epoch at which the synthesized images coming from the generators are the most similar to the real images from the data distributions. The FID score uses statistics on feature vectors generated by a Inception V3 model trained on ImageNet :

$$FID(r,g) = \|m_r - m_g\|_2^2 + Tr[C_r + C_g - 2(C_rC_g)^{\frac{1}{2}}] \quad (9)$$

where (m_r, C_r) and (m_g, C_g) are the mean and covariance of the real data and generated data distributions, respectively. A smaller FID score indicates that the distribution of the generated images is similar to the distribution of real images that were used to train the model.

To evaluate the Unet-like model that performs the holographic image reconstruction form a single hologram, we used the measure of enhancement (EME) [33] and a perceptual blurness metric [34].

The EME is a no-reference image quality assessment metric and is expressed as :

$$EME(x) = \frac{1}{k_1k_2}\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} 20\ln\frac{I_{max}^{i,j}}{I_{min}^{i,j}} \quad (10)$$

where the image is divided in k_1k_2 blocks and where $I_{min}^{i,j}$ and $I_{max}^{i,j}$ are the minimum and maximum pixels values in each block of the input image x. A higher EME score indicates better perceptual quality.

The no-reference blur metric proposed by [34] compares the intensity variations of the neighboring pixels between the image to be evaluated and the same image blurred with a low-pass filter. The blur estimation score ranges from 0 to 1 and a smaller score indicates better quality in term of blur perception.

We also used the well-known SSIM score [35], which is a full reference image quality metric:

$$SSIM(r,e) = \frac{(2\mu_r\mu_e + C_1)(2\sigma_{re} + C_2)}{(\mu_r^2 + 2\mu_e^2 + C_1)(\sigma_r^2 + \sigma_e^2 + C_2)} \quad (11)$$

where μ_r,μ_e, σ_r,σ_e and σ_{re} are the local means, standard deviations, and cross-covariance for images r and e.

3 RESULTS

3.1 CycleGan

The CycleGAN was trained for 200 epochs with a training dataset consisting of 800 segmented brightfield microscopy images (domain A) and 600 holographic reconstructed images (domain B). The image size in both domains is 128x128 pixels and a batch size of 1 is used. During training, the images are randomly cropped and data augmented (rotations and flips) before being used as inputs. Every 5 epochs, the generators and discriminators are saved. To evaluate the performance of the CycleGAN during training and to determine at which epoch the generated brightfield-like images are most similar to the true brightfield images, the Frechet Inception Distance (FID) score was calculated. The best FID score (53) was obtained at epoch 95 of the training (see Fig.S1) and the model at this epoch was used to transform the real holographic reconstructed images into fake brightfield images. Fig. 3 shows examples of generated brightfield images. The generator outputs are characterized by sharper edges, less noise and an increased contrast.

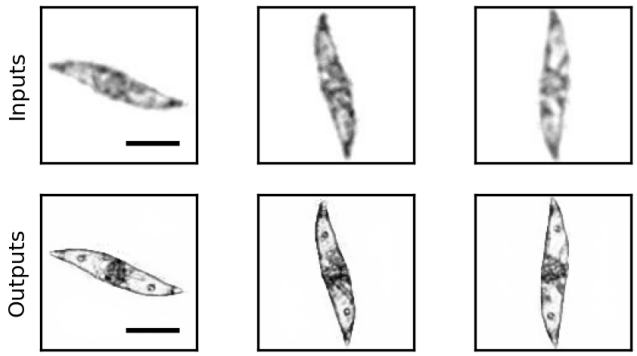

Figure 3: CycleGAN outputs. First row: inputs (real holographic reconstructed images). Second row: corresponding outputs (fake brightfield images). Scale bar is 50μm.

At this stage, which corresponds to the end of step 1 of our method, we show that it is possible to transform a real holographic reconstructed image into a fake brightfield image using a generative model trained with unpaired data. Although the generator is able to enhance the image and even restores/hallucinates some details of the plankton (chloroplasts for e.g.), it should be noted that not all the images are perfectly translated by the generator. Indeed, some generated images can have some important deformations and distortions that can occur if the input images are not correctly segmented or possess multiple reflection interference. Although these distorted images will be found in the dataset (as ground-truth) of step 2, supervised learning coupled with a high batch size allows to limit the possible negative impact of these images on the training.

3.2 New enhanced dataset

Once the CycleGAN was trained, the real holographic reconstructed images were transformed into fake bright-field microscopy images using the G_B generator. Fig. 4 presents a sample of the new dataset.

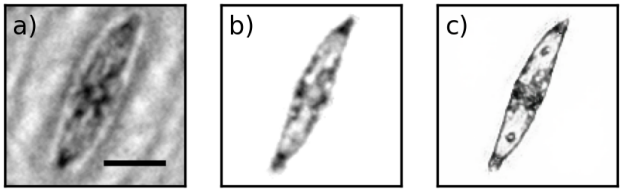

Figure 4: New dataset. a) defocused hologram, b) dual-plane phase recovered amplitude image and c) enhanced fake brightfield image generated from image b) by the CycleGAN. Scale bar is 50μm.

With this new aligned dataset (*hologram / real reconstruction/ fake brightfield image*), two Unet-like models are trained in a supervised setting to perform the segmented holographic image reconstruction from a single hologram. The first model (*holo* $\rightarrow$ *rec*), whose results will serve as a baseline, is trained with the dataset *hologram / real reconstruction*. Its performances correspond to the case where the CycleGAN would not have been used beforehand to enhance the visual quality of the holographic reconstructed images. The second model (*holo* $\rightarrow$ *fake*) is trained with the dataset *hologram / fake brightfield image*. Both models were trained under the same conditions during 100 epochs with the same hyperparameters (batch size of 16, learning rate of 0.00015, Adam optimizer) and with the same defocused holograms as inputs (80% of the dataset).

The Fig. 5 shows examples of predictions of the two models on test holograms. The outputs of the model trained with the data *hologram / fake brightfield image* have a better visual quality than the outputs of the baseline model with notably less blur and noise. The background subtraction is also more efficient and the segmentation of the object is sharper. Fig. 6 displays an example of the 2 models outputs for the same input and their corresponding gradient magnitude images from Sobel operators. It is clear that the *holo* $\rightarrow$ *fake* model is able to produce an image with sharper edges and restored details (chloroplasts).

To further evaluate and compare the models performances, we computed the EME and Blurness metrics in a pairwise manner: for every hologram in the testing dataset, we computed the scores on the outputs of the baseline model (*holo* $\rightarrow$ *rec*) and the final model (*holo* $\rightarrow$ *fake*), respectively. The mean scores evaluated from 100 test holograms are summarized in Table 1 and indicate that the holographic reconstructed images generated by the *holo* $\rightarrow$ *fake* model are characterized by less blur and an increased contrast, which is consistent with the qualitative comparison reported in Fig. 5. We also report the image quality assessment metrics of the holographic reconstructed images that were obtained with a dual-plane phase recovery routine on the test holograms. The results confirmed that the outputs of the *holo* $\rightarrow$ *rec* model have a much lower quality than the holo-

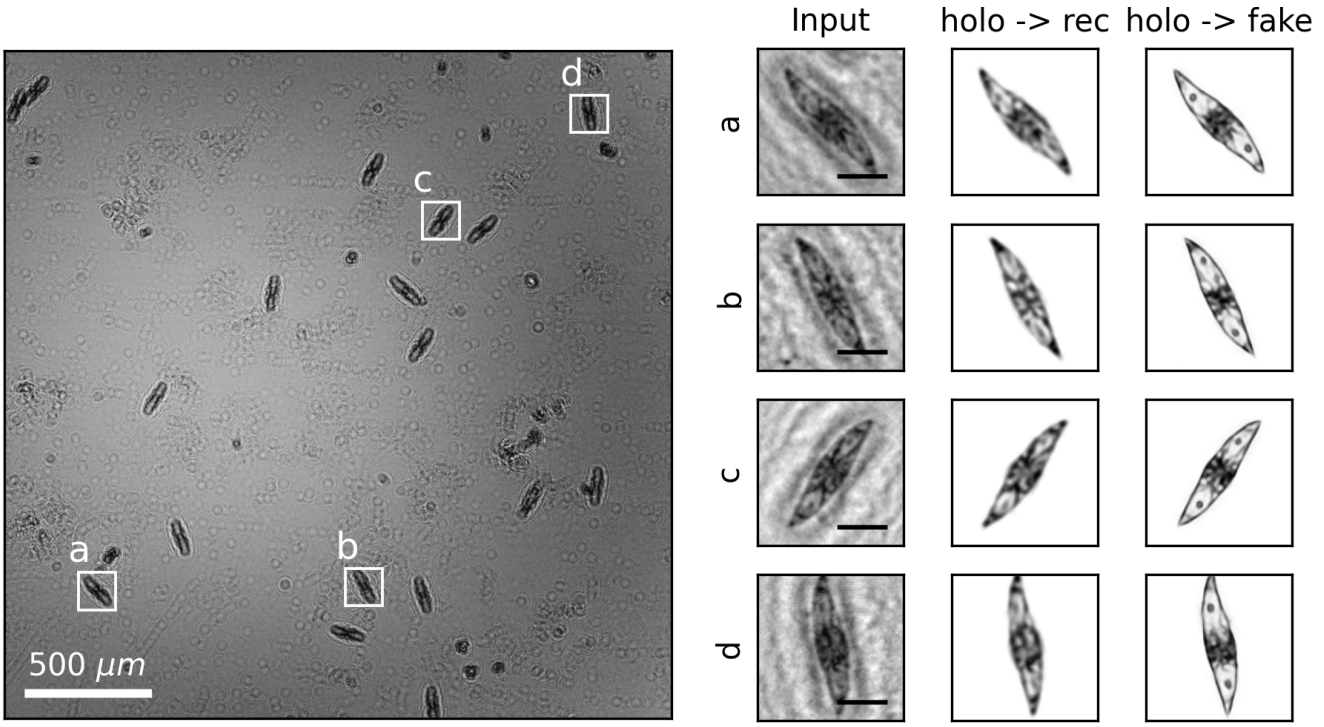

Figure 5: Models comparison on a test hologram (left image). The hologram is first backpropagated near the global focus plane and the models perform the holographic reconstruction on small regions of interest (a,b,c,d). First column (right images) : models input, second and third columns : corresponding outputs of *holo* → *rec* and *holo* → *fake* models, respectively. Black scale bar is 50μm.

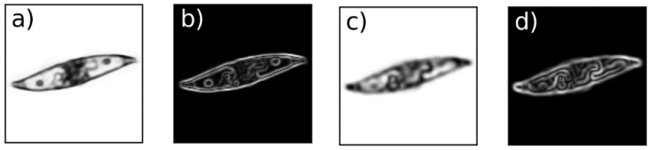

Figure 6: Sobel edges images. Images a) and c) are the outputs of the *holo* → $fake$ and *holo* → *rec* models, respectively. Images b) and d) are the corresponding absolute magnitude of the gradient (Sobel operators).

graphic reconstructions obtained by dual-plane phase recovery. This is explained by the fact that for supervised learning, the quality of the outputs of the trained model is limited by the quality of the ground-truth images during training and that it cannot be miraculously better. While *holo* → $fake$ model is also trained in a supervised setting, the ground-truth images are brightfield-like images that are, in a way, a better and enhanced version of the holographic reconstructed images. As a result, the model is able to generate images of higher quality than the images reconstructed by dual-plane phase recovery.

Table 1: Outputs scores evaluated on 100 test holograms

Images	Blurness ↓	EME ↑
holo → *rec* (baseline) outputs	0.5429	2.6829
phase recovered amplitude image	0.4179	3.2050
holo → $fake$ outputs	**0.3894**	**3.3702**

3.3 *Model prediction on de-focused holograms*

During the training of the Student (*holo* → $fake$) model, the raw holograms were randomly backpropagated close to the correct focus plane in order to (1) increase the number of images in the dataset and (2) obtain a model which is robust to various defocus distances. This aspect is an important element to study because in order to speed up the holographic reconstruction it is possible to backpropagate once the full hologram (FOV of $10mm^2$) before performing the holographic reconstruction with the model on small ROIs. Since the different plankton within the FOV do not have exactly the same object-camera distance, the ROIs will have different defocus variations. Moreover, the model should be able to produce the same output for a given ROI regardless of the axial defocus distance. Fig. 7 shows examples of model predictions for different defocused holograms of the same ROI (backpropagated to a plane that is placed 0,+15μm and +30μm away from the correct global focus plane). We can see that the model outputs are very similar to each other, even when the defocus distance is large. To evaluated the model con-

sistency for various defocused holograms, we have calculated the SSIM index across an axial defocus range of -30 μm and 30 μm, which was averaged over 100 test ROIs. The results, which are displayed in Fig. 8, show that the SSIM index between the model output of a in-focus hologram and the model output of the same but de-focused hologram is greater than 0.98, suggesting that the difference is visually negligible. Hence, the model is able to produce the same reconstructed image for multiple defocused holograms of the same ROI and has learned many-to-one mappings.

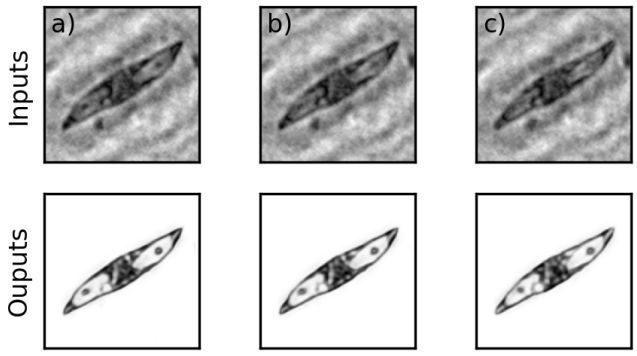

Figure 7: *holo* $\rightarrow$ *fake* outputs for different degree of defocus. a) Inplane hologram, b) 15μm defocused hologram , c) 30μm defocused hologram.

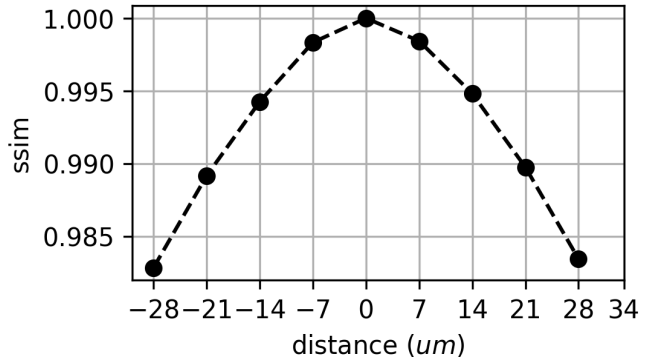

Figure 8: SSIM index value as a function of the axial defocus distance. The reference image for the calculation is considered to be the model output of the in-plane hologram for each ROI. The SSIM curve is averaged over 100 test ROIs (128x128 pixels).

3.4 Model restoration performances

To further evaluate the CycleGAN and the Unet, we measured a set of the same samples under a brightfield microscope and a holographic device. The image registration between the two imaging modalities was manually performed using the imageJ plugin TrackEM2. A visual comparison can be found Fig.9 where the predicted images (CycleGAN and Student model) are compared to the references images in the brightfield and holographic domains, respectively. The mean SSIM, MSE and FID scores on 20 paired images are reported in Table 2. One can notice that the *holo* $\rightarrow$ *fake* outputs look much more like true holographic reconstructed (domain B) images than real brightfield microscopy images (domain A), and this despite the fact that the Student model was trained on synthetic brightfield-like images provided by the CycleGAN. Indeed, the images predicted by the Unet obtain a SSIM/FID scores of 0.911/26.4 with the reference images of domain B against a SSIM/FID scores of 0.844/60.5 with the reference images of domain A. The Unet output seems de-noised compared to the true reconstructed image and that some details are restored but globally the object is not misrepresented and the Unet does not hallucinate too much features as a GAN could do. This result was expected because the Unet is trained with a standard supervised learning setting and therefore, has to minimize its loss per batch on the entire training dataset which reduces its hallucination capacities. However, training a Unet in a supervised setting with the images generated by the CycleGAN as ground-truth has many advantages. First, even though the CycleGAN is able to generate images that look like brightfield images (FID of 49 on the test paired data), it is not capable of many-to-one mappings and some of its outputs may have significant distortions when the input images are not correctly segmented or if they already possess significant interference (twin-image artifacts, cross-interference between objects). The Unet model, although trained on the images generated by the CycleGAN, does not output images that display these deformations or artifacts as the supervised setting during training tends to "smooth" the possible solutions. Moreover, the Unet is trained with a multitude of input images that display different noise level, interference and contrast. As a result, the domain shift between the training dataset and new/unseen holograms is limited and the model is more robust.

Table 2: CycleGAN and Unet evaluation on 20 paired.

Reference Image	Scores	CycleGan (Fake A)	Unet
True Bright-field (True A)	MSE	0.0130	0.0105
	SSIM	0.833	0.844
	FID	49.98	60.46
True amplitude (True B)	MSE	0.0047	0.0035
	SSIM	0.895	0.911
	FID	55.05	26.43

Secondly, CycleGANs are known to be complicated to train due to the minimax game and they generally possess a large number of layers and parameters that may be oversized for the task at hand. However, once trained, the CycleGAN can be seen as a preprocessing step that are apply to the holographic reconstructed images to produce new enhanced ground-truth images to guide the training of a smaller model. Once the synthetic images are generated and the new dataset is constituted (which is a one-time effort), it is possible to train and optimize the structure of the Unet over several runs. Indeed, since the Unet is trained with paired data, it is also more prone to converge than the CycleGAN and therefore, it is also easier to optimize its structure (number of layers, number of filters, etc.) to the requested task.

Finally, since the CycleGAN in our method can be considered

as a pre-processing step, it is possible to incorporate various additional image processing operations. As the objective is to transform the holographic reconstructed image (domain B) into a synthetic brightfield image (domain A), it is possible to process the images of domain A as desired before the CycleGAN training. For example, in this paper, we have segmented and converted in grayscale the images of domain A. As a result, the Student model will output a grayscale and segmented reconstructed image. Our method could be applied for other computer vision tasks, for example colorization (grayscale hologram → colored image) or super-resolution (LR hologram → HR image), which will be simplified because all the image processing steps are applied on the brightfield images before the CycleGAN training and not on holograms, which are more difficult to acquire and process. However, it is important to note that the domains A et B must share a similar pixel-level structure to ensure that the CycleGAN outputs consistent images. For example, a CycleGAN can transform an image of a zebra into an image of a horse but tends to fail to translate an image A into an image B when the domain shift is significant (for e.g. cat and dog) [36]. Since the brightfield images (domain A) are used because they already have the desired perceptual features (contrast, resolution,...), it is necessary to adapt the images of domain B so that they are structurally similar to the images of domain A. In particular, raw holograms that exhibit interference patterns are far from having pixel-level similarity with brightfield images, and must, therefore be reconstructed in their respective focus plane before constituting B-domain images.

Even if the Unet hallucinates fewer features than the CycleGAN, it is important to remember that the task of the two models is not the same. The CycleGAN learns one-to-one mappings between in-focus phase recovered images and brightfield microscopy images, whereas the Unet learns many-to-one mappings between raw defocused holograms and enhanced reconstructed images. The prior use of CycleGan can, in a way, correct the holographic reconstructed images before they are used as target images. In particular, The generator G_B allows to de-noise them and to recover the characteristic features of the object (e.g. chloroplasts) which appear poorly on the reconstructed images but are clearly visible on the brightfield images.

4 CONCLUSIONS

In this paper, we propose a novel Teacher-Student approach to improve the visual quality of holographic reconstructions performed by deep learning from a single hologram. The method relies on the use of images acquired by other imaging methods, which reduces the negative impact that degraded holographic reconstructed images could have if they were used as target images in a classical supervised learning setting. A Teacher model, a CycleGAN, is first trained with unpaired images to transform holographic reconstruction images into synthetic brightfield microscopy images. These generated images, which have similar features to brightfield images in term of contrast, noise or resolution, are then used as target images in supervised learning of a second (Student) model that has fewer parameters than the Teacher model and can be optimized to be deployed into edge devices. Our experiments conducted on sparse samples (plankton images) have shown that the images reconstructed by our method have a better visual quality than those obtained by a classical holographic image reconstruction technique (dual-plan phase recovery) or those obtained by a trained model without the prior use of CycleGAN. Our method allows to restore and improve the quality of the images without hallucinating features. We have also demonstrated that the model has learned many-to-one mappings, which ensures that different defocused holograms of the same object give the same reconstructed image and that the model does not output random focal representation. Our method based on the use of unpaired images limits the complexity of the holographic device or holographic image reconstruction algorithms and has a great potential for other computer vision tasks in other coherent imaging fields, such as super-resolution or colorization. In future work, we attend to instigate the generality of the proposed approach on others objects and in particular dense samples.

REFERENCES

[1] Alborz Feizi, Yibo Zhang, Alon Greenbaum, Alex Guziak, Michelle Luong, Raymond Yan Lok Chan, Brandon Berg, Haydar Ozkan, Wei Luo, Michael Wu, Yichen Wu, and Aydogan Ozcan. Rapid, portable and cost-effective yeast cell viability and concentration analysis using lensfree on-chip microscopy and machine learning. *Lab on a Chip*, 16(22):4350–4358, 2016. ISSN 14730189. doi: 10.1039/c6lc00976j.

[2] M Malek, D Allano, S Coëtmellec, C Özkul, and D Lebrun. Digital in-line holography for three-dimensional–two-components particle tracking velocimetry. *Measurement Science and Technology*, 15(4):699–705, mar 2004.

[3] Xiao Yu, Jisoo Hong, Changgeng Liu, and Myung K. Kim. Review of digital holographic microscopy for three-dimensional profiling and tracking. *Optical Engineering*, 53(11):1 – 22, 2014.

[4] Francesco Merola, Pasquale Memmolo, Lisa Miccio, Roberto Savoia, Martina Mugnano, Angelo Fontana, Giuliana D'Ippolito, Angela Sardo, Achille Iolascon, Antonella Gambale, and Pietro Ferraro. Tomographic flow cytometry by digital holography. *Light: Science and Applications*, 6(4):1–7, 2017.

[5] Gregor Scholz, Qifeng Xu, Torben Schulze, Heidi Boht, Kai Mattern, Jana Hartmann, Andreas Dietzel, Stephan Scherneck, Ingo Rustenbeck, Joan Prades, Sönke Fündling, Hutomo Wasisto, and Andreas Waag. LED-Based Tomographic Imaging for Live-Cell Monitoring of Pancreatic Islets in Microfluidic Channels. *Proceedings*, 1(10):552, 2017.

[6] Zoltán Göröcs, Miu Tamamitsu, Vittorio Bianco, Patrick Wolf, Shounak Roy, Koyoshi Shindo, Kyrollos Yanny, Yichen Wu, Hatice Ceylan Koydemir, Yair Rivenson, and Aydogan Ozcan. A deep learning-enabled portable imaging flow cytometer for cost-effective, high-throughput, and label-free analysis of natural water samples. *Official journal of the CIOMP*, 7:2047–7538, 2018.

[7] Hao Wang, Meng Lyu, and Guohai Situ. eHoloNet: a learning-based end-to-end approach for in-line digital holographic reconstruction. *Optics Express*, 26(18):22603, 2018.

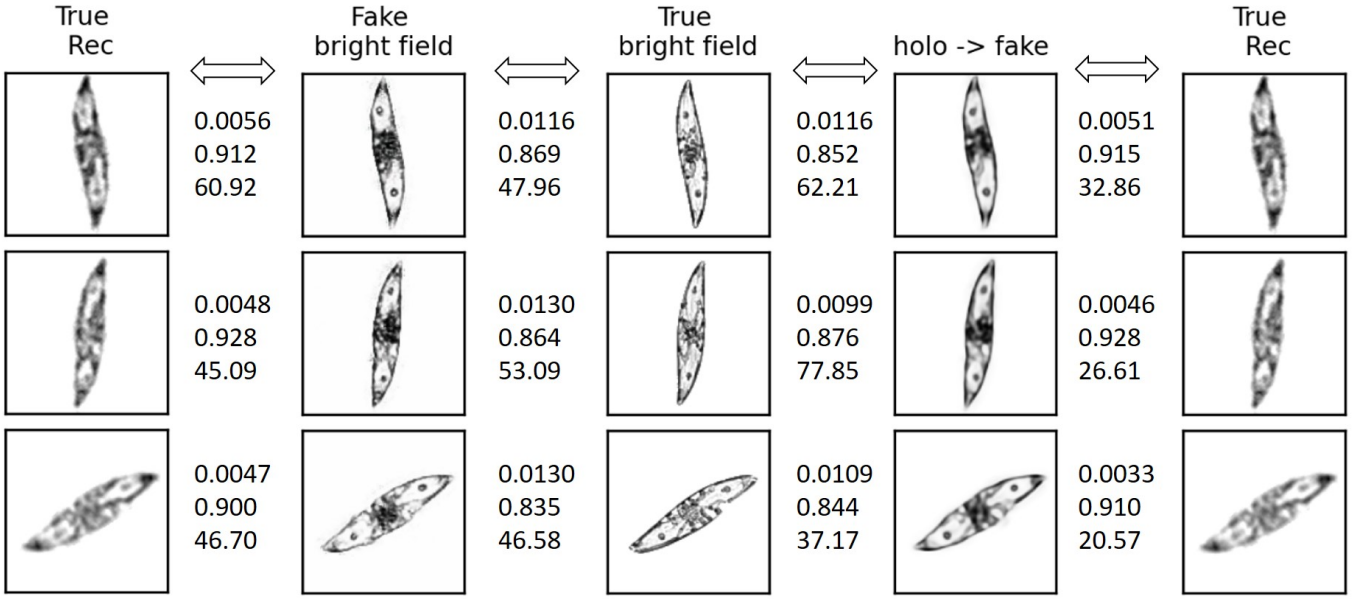

Figure 9: Comparison between the CycleGAN and the Unet *holo* $\rightarrow$ *fake* outputs. The CycleGAN generated the Fake brightfield images from the True Rec images. The MSE, SSIM and FID scores are displayed in this ordre between the predicted images (CycleGAN or Unet) and the reference images in the A and B domains, respectively.

[8] Yair Rivenson, Yibo Zhang, Harun Günaydın, Da Teng, and Aydogan Ozcan. Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light: Science and Applications*, 7(2):17141, 2018.

[9] Tianjiao Zeng, Yanmin Zhu, and Edmund Y. Lam. Deep learning for digital holography : a review. *Opt. Express*, 29(24):40572–40593, 2021.

[10] Yair Rivenson, Yichen Wu, Hongda Wang, Yibo Zhang, Alborz Feizi, and Aydogan Ozcan. Sparsity-based multi-height phase recovery in holographic microscopy. *Scientific Reports*, 6(November):1–9, 2016.

[11] Tairan Liu, Zhensong Wei, Yair Rivenson, Kevin de Haan, Yibo Zhang, Yichen Wu, and Aydogan Ozcan. Deep learning-based color holographic microscopy. *Journal of Biophotonics*, 12(11):1–12, 2019. ISSN 18640648. doi: 10.1002/jbio.201900107.

[12] Yichen Wu, Yilin Luo, Gunvant Chaudhari, Yair Rivenson, Ayfer Calis, Kevin de Haan, and Aydogan Ozcan. Bright-field holography: cross-modality deep learning enables snapshot 3D imaging with bright-field contrast using a single hologram. *Light: Science and Applications*, 8(1), 2019.

[13] Da Yin, Zhongzheng Gu, Yanran Zhang, Fengyan Gu, Shouping Nie, Jun Ma, and Caojin Yuan. Digital Holographic Reconstruction Based on Deep Learning Framework With Unpaired Data. *IEEE Photonics Journal*, 12(2): 1–12, 2020.

[14] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision*, 2017-October:2242–2251, 2017.

[15] Dániel Terbe, László Orzó, and Ákos Zarándy. Deep-learning-based bright-field image generation from a single hologram using an unpaired dataset. *Optics Letters*, 46 (22):5567, 2021.

[16] Veit Sandfort, Ke Yan, Perry Pickhardt, and Ronald Summers. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*, 9, 11 2019. doi: 10.1038/s41598-019-52737-x.

[17] Pasquale Memmolo, Melania Paturzo, Bahram Javidi, Paolo A. Netti, and Pietro Ferraro. Refocusing criterion via sparsity measurements in digital holography. *Optics Letters*, 39(16):4719, 2014. ISSN 0146-9592. doi: 10.1364/ol.39.004719.

[18] P. Memmolo, C. Distante, M. Paturzo, A. Finizio, P. Ferraro, and B. Javidi. Automatic focusing in digital holography and its application to stretched holograms. *Opt. Lett.*, 36(10):1945–1947, May 2011. doi: 10.1364/OL.36.001945.

[19] Frank Dubois, Cédric Schockaert, Natcaha Callens, and Catherine Yourassowsky. Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Optics Express*, 14(13):5895, 2006.

[20] Yibo Zhang, Hongda Wang, Yichen Wu, Miu Tamamitsu, and Aydogan Ozcan. Edge sparsity criterion for robust holographic autofocusing. *Optics Letters*, 42(19):3824, 2017.

[21] R. W. Gerchberg. A practical algorithm for the determi-

nation of phase from image and diffraction plane pictures. *Optik*, 35:237–246, 1972.

[22] Peng Bao, Fucai Zhang, Giancarlo Pedrini, and Wolfgang Osten. Phase retrieval using multiple illumination wavelengths. *Optics Letters*, 33(4):309, 2008.

[23] Wei Luo, Alon Greenbaum, Yibo Zhang, and Aydogan Ozcan. Synthetic aperture-based on-chip microscopy. *Light: Science and Applications*, 4(3), 2015.

[24] Yibo Zhang, Alon Greenbaum, Alborz Feizi, Ping Luen Chung, Wei Luo, Shivani R. Kandukuri, and Aydogan Ozcan. Wide-field imaging of pathology slides using lensfree onchip microscopy. *CLEO: Applications and Technology, CLEO-AT 2015*, 6(267), 2015.

[25] Onur Mudanyali, Derek Tseng, and Chulwoo Oh. Compact, Light-weight and Cost-effective Microscope based on Lensless Incoherent Holography for Telemedicine Applications. *Lab on a chip*, 10(11):1417–1428, 2010.

[26] Ting-Weig Su, Anthony Erlinger, Derek Tseng, and Aydogan Ozcan. Compact and Light-Weight Automated Semen Analysis Platform Using Lensfree on-Chip Microscopy. *Anal Chem*, 82(19):1–7, 2010.

[27] Jialin Zhang, Jiasong Sun, Qian Chen, and Chao Zuo. Resolution Analysis in a Lens-Free On-Chip Digital Holographic Microscope. *IEEE Transactions on Computational Imaging*, 6:697–710, 2020.

[28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9351:234–241, 2015.

[29] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei Efros. Image-to-image translation with conditional adversarial networks. pages 5967–5976, 07 2017. doi: 10.1109/CVPR.2017.632.

[30] Herbert Bay, Tinne Tuytelaars, and Luc" Van Gool. SURF: Speeded Up Robust Features. In *Computer Vision – ECCV 2006*, pages 404–417. Springer Berlin Heidelberg, 2006.

[31] Alon Greenbaum and Aydogan Ozcan. Maskless imaging of dense samples using pixel super-resolution based multi-height lensfree on-chip microscopy. *Optics Express*, 20(3):3129, 2012.

[32] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 6629–6640, Red Hook, NY, USA, 2017. Curran Associates Inc.

[33] Sos S. Agaian, Karen P. Lentz, and Artyom M. Grigoryan. A new measure of image enhancement. In *IASTED International Conference on Signal Processing & Communication*, 2000.

[34] Frederique Crete, Thierry Dolmiere, Patricia Ladret, and Marina Nicolas. The blur effect: perception and estimation with a new no-reference perceptual blur metric. In Bernice E. Rogowitz, Thrasyvoulos N. Pappas, and Scott J. Daly, editors, *Human Vision and Electronic Imaging XII*, volume 6492, pages 196 – 206. International Society for Optics and Photonics, SPIE, 2007.

[35] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

[36] Amélie Royer, Konstantinos Bousmalis, Stephan Gouws, Fred Bertsch, Inbar Mosseri, Forrester Cole, and Kevin Murphy. XGAN: Unsupervised Image-to-Image Translation for Many-to-Many Mappings, 2018.