Optics and Lasers in Engineering 165 (2023) 107546

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

FocusNET: An autofocusing learning-based model for digital lensless holographic microscopy

Manuel Montoya [a,*], Maria J. Lopera [a], Alejandra Gómez-Ramírez [b], Carlos Buitrago-Duque [b], Adriana Pabón-Vidal [c], Jorge Herrera-Ramirez [d], Jorge Garcia-Sucerquia [b], Carlos Trujillo [a,*]

[a] Applied Optics Group, School of Applied Science and Engineering, Universidad EAFIT, Medellín, Colombia
[b] School of Physics, Universidad Nacional de Colombia sede Medellín, A.A: 3840-Medellín-050034, Colombia
[c] Grupo Malaria, Facultad de Medicina, Universidad de Antioquia (UdeA), Sede de Investigación Universitaria (SIU), Medellín, Colombia
[d] AEyCC Group, School of Engineering, Instituto Tecnológico Metropolitano, Calle 54a No. 30 - 01, Medellín, Colombia

ARTICLE INFO

Keywords:
Lensfree microscopy
Autofocusing
Deep learning
Convolutional neural network
Digital Gabor holography

ABSTRACT

This paper reports on a convolutional neural network (CNN) – based regression model, called FocusNET, to predict the accurate reconstruction distance of raw holograms in Digital Lensless Holographic Microscopy (DLHM). This proposal provides a physical-mathematical formulation to extend its use to different DLHM setups than the optical and geometrical conditions utilized for recording the training dataset; this unique feature is tested by applying the proposal to holograms of diverse samples recorded with different DLHM setups. Additionally, a comparison between FocusNET and conventional autofocusing methods in terms of processing times and accuracy is provided. Although the proposed method predicts reconstruction distances with approximately 54 μm standard deviation, accurate information about the samples in the validation dataset is still retrieved. When compared to a method that utilizes a stack of reconstructions to find the best focal plane, FocusNET performs 600 times faster, as no hologram reconstruction is needed. When implemented in batches, the network can achieve up to a 1200-fold reduction in processing time, depending on the number of holograms to be processed. The training and validation datasets, and the code implementations, are hosted on a public GitHub repository that can be freely accessed.

1. Introduction

Digital Lensless Holographic Microscopy (DLHM) is an imaging technique that uses computational methods to retrieve the complex wavefield information of light scattered by micrometer-sized samples [1]. DLHM is based on the in-line Gabor holographic setup with spherical illumination [2,3]. Its simplicity is highlighted by the required hardware: a digital sensor, the sample to be studied, and a divergent spherical wavefront source. In DLHM, the magnification of the recorded hologram is controlled by the location of the sample: the closer the sample to the point source, the higher the magnification [1]. In other lens-free microscopy techniques, such as computational microscopy, the magnification is fixed to the unit [4]. These imaging modalities rely on highly resource-demanding super-resolution techniques and other learning-based approaches [5,6] which are not needed in DLHM. DLHM allows the retrieval of amplitude and phase information of the object by numerically backpropagating the recorded hologram to an accurate reconstruction distance [1]. In general, lens-free microscopy implementations allow non-destructive and label-free imaging, which has propelled its study and research around biomedical applications, providing polarization sensitivity [7], *in vivo* analysis [3,8], and disease diagnosing [8,9], among others.

In Digital Holographic Microscopy (DHM), a significant challenge is determining the samples' exact location within the inspection volume without any extended procedure [10]. For samples containing weak scatterers axially non-connected specimens, a digital hologram provides plane-by-plane information about the said volume. Nevertheless, there is no straightforward way to establish the in-focus reconstruction plane of the studied object [11,12]. Autofocusing methods have been proposed mainly in Digital Holography (DH) and lens-based DHM to tackle this challenge [10,13,14]. The Dubois' metric is one of the most quoted methods, consisting of a pixel-wise summation of the amplitude module of each reconstructed hologram in the inspection volume [15]. As this one, many other proposals can be found in the literature to solve this problem in lens-based DHM [10,12]. The application of these proposals has allowed flow analysis [16], tracking of living cells [17], and living cell analysis [18], among other high-impact applications [19]. Although highly effective, all these methods require hundreds, and sometimes

* Corresponding authors.
E-mail addresses: mmonto95@eafit.edu.co (M. Montoya), catrujilla@eafit.edu.co (C. Trujillo).

https://doi.org/10.1016/j.optlaseng.2023.107546
Received 4 December 2022; Received in revised form 29 January 2023; Accepted 21 February 2023
Available online 3 March 2023

thousands, of sequential numerical reconstructions to provide accurate in-focus reconstruction distances. Thus, conventional methods are computationally demanding and time-consuming [20].

Beyond being computationally demanding, conventional autofocusing proposals fail for DLHM holograms because they do not consider this technique's illumination source wavefront sphericity [21]. In [12,21], the authors developed an autofocusing method based on finding the reconstruction plane in which the smallest area around the object to be focused encloses a given amount of irradiance. For each reconstruction plane, the enclosed irradiance is weighted by the inverse of the spherical illumination radius at that plane. The proposal successfully autofocuses intricate inner-structure samples, at the expenses of requiring high computational complexity preventing its implementation for *in vivo* applications [12].

Deep learning-based methods are powerful tools for solving computer vision problems and image analysis [22,23]. These methods provide promising strategies to obtain the in-focus reconstruction distance directly from the hologram without requiring several back-propagation operations. Some learning-based proposals can be encountered in holographic imaging to solve the autofocusing challenge. For instance, Shimobaba et al. [24] proposed a CNN-based regression to automatically focus DH holograms with millimeter precision directly from the power spectrum of the holograms. Rembo et al. reported an autofocusing proposal for DH using a CNN-based classification model by splitting the reconstruction distances into categories [25]. The same authors later reported a CNN-based regression to directly predict the reconstruction distance from the hologram [26]. Pitkäaho et al. proposed the first approach to deep learning autofocusing in lens based DHM, successfully avoiding stack reconstruction procedures. Although these deep learning-based models have been successfully implemented and tested for DH and lens-based DHM, to the authors' best knowledge, these methods have not been applied yet to DLHM, where the autofocusing problem is more challenging because of its required spherical illumination and the wide range of routinely attained reconstruction distances.

This paper proposes a convolutional neural network (CNN) model called FocusNET to autofocus raw DLHM holograms. Directly from the hologram, the model uses a regression strategy to predict the in-focus reconstruction distance, thus the object's location within the inspection volume. After describing the DLHM fundamentals and the proposed method in detail, a validation of its accuracy after proper training is presented. This validation evaluates the FocusNET-yielded intensity reconstructions for holograms in the test dataset and holograms recorded with different geometrical and optical parameters. For the latter case, a physical-mathematical formulation is provided using the scalar diffraction integral that describes the reconstruction procedure in DLHM to extend the use of FocusNET to DLHM holograms with arbitrary recording conditions. Finally, the method performance is contrasted against conventional methods considering the computation times to reach accurate in-focus reconstruction distances [8]. Experimental results show that FocusNET predicts the object's in-focus reconstruction distance directly from raw DLHM holograms without any numerical reconstructions at remarkably reduced processing times. The datasets and the implemented codes are publicly available via the GitHub repository "https://github.com/mmonto95/focusnet".

2. Digital lensless holographic microscopy

As previously mentioned, DLHM only requires a spherical wavefront source, a sample, and a digital sensor [3]. The sample is placed after the illumination source to produce a magnified diffraction pattern of the specimen at the sensor plane. Fig. 1 shows a basic DLHM setup.

To mathematically describe the DLHM hologram, consider that the sample transmittance $S(\vec{r_0})$, placed at a distance z from the point source, is illuminated by a diverging spherical wavefront $\exp[i\,\vec{k}\cdot\vec{r_0}]/|\,\vec{r_0}\,|$. The diffracted wavefield is magnified by free-space propagation until it reaches the digital sensor at a distance L from the point source. Vectors

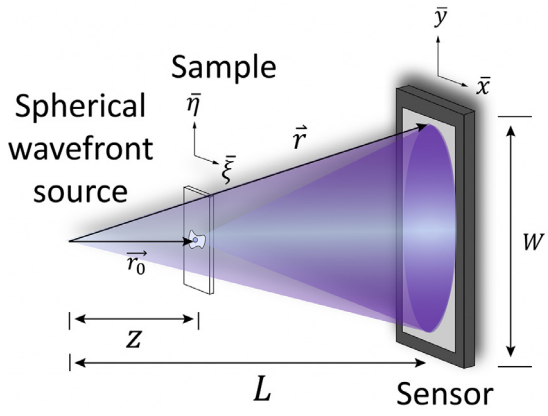

Fig. 1. Digital lensless holographic microscopy (DLHM) setup. The source-to-sample distance z, the source-to-sensor distance L and the sensor width W establish the geometry of the DLHM setup.

$\vec{r_0} = (\xi, \eta, z)$ and $\vec{r} = (x, y, L)$ denote locations at the sample and camera planes, respectively. The description of the complex-valued optical field that reaches the sensor is given by the Rayleigh-Sommerfeld diffraction formula given by Eq. (1) [27,28].

$$U\left(\vec{r}\right) = \int_{sample} S\left(\vec{r_0}\right) \frac{\exp\left[i\,\vec{k}\cdot\vec{r_0}\right]}{\left|\vec{r_0}\right|} \frac{\exp\left[i\,\vec{k}\cdot\left(\vec{r}-\vec{r_0}\right)\right]}{\left|\left(\vec{r}-\vec{r_0}\right)\right|} \mathrm{d}\,\vec{r_0}\,. \tag{1}$$

Then, the intensity distribution recorded at the sensor plane, known as the DLHM hologram, is described by Eq. (2),

$$I\left(\vec{r}\right) = U\left(\vec{r}\right)\cdot\ U^{*}\left(\vec{r}\right) \tag{2}$$

where $U^{*}(\vec{r})$ is the complex conjugate of the optical field, say, $U^{*}(\vec{r}) = \int_{sample} S^{*}(\vec{r_0}) \frac{\exp[-i\,\vec{k}\cdot\vec{r_0}]}{|\vec{r_0}|} \frac{\exp[-i\,\vec{k}\cdot(\vec{r}-\vec{r_0})]}{|(\vec{r}-\vec{r_0})|}\,\mathrm{d}\,\vec{r_0}$.

The information of the sample $S(\vec{r_0})$ is later retrieved by backpropagating the DLHM hologram as being illuminated with a converging spherical wavefront $\exp[-i\,\vec{k}\cdot\vec{r}]/|\,\vec{r}\,|$, which is the complex conjugate of the illumination wavefront used during recording. This diffraction process can be numerically described through a scalar diffraction formula; the sample's information is found via Eq. (3).

$$S\left(\vec{r_0}\right) = \int_{sensor} I\left(\vec{r}\right) \frac{\exp\left[-i\vec{k}\cdot\vec{r}\right]}{\left|\vec{r}\right|} \frac{\exp\left[-i\vec{k}\cdot\left(\vec{r}-\vec{r_0}\right)\right]}{\left|\left(\vec{r}-\vec{r_0}\right)\right|} \mathrm{d}\vec{r}. \tag{3}$$

A discrete version of Eq. (3) is used for the digital reconstruction of the DLHM holograms [28]. From Eq. (3), the complex-valued optical field of the sample $S(\vec{r_0})$ is recovered, and its amplitude $|S(\vec{r_0})|$, intensity $S(\vec{r_0})\cdot S^{*}(\vec{r_0})$ or phase $\phi(\vec{r_0}) = \arctan(\mathrm{Im}[S(\vec{r_0})]/Re\,[S(\vec{r_0})])$ can be retrieved [29,30]. As the information provided by $S(\vec{r_0})$ is affected by the DC term, it is routine to use a contrast reference hologram [1] prior to applying Eq. (3) to remove this noisy artifact. Additionally, this and other undesired artifacts can be removed from the resulting reconstructions by digitally pre-processing the acquired DLHM holograms [31] or by applying other numerical processing methods, such as numerical darkfield illuminations on the resulting images [32]. To address the aberration caused by the spherical illumination wavefront, a final point-wise multiplication of the resulting complex wavefield at the sample plane with a conjugated wavefront of the spherical illumination is used. This point-wise operation results in phase reconstructions of the sample without the undesired effects of the illumination wavefront.

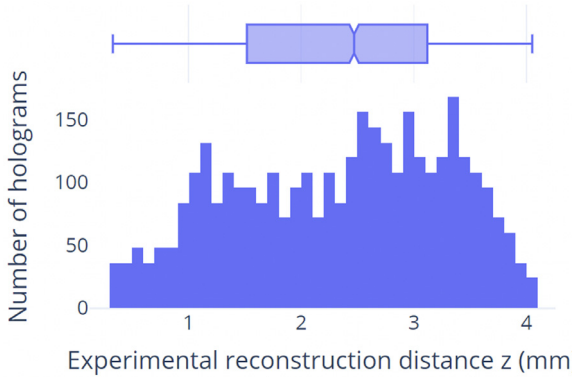

Fig. 2. Number of DLHM recorded holograms for each reconstruction distance z in the experimental dataset. A total of 3540 holograms constitute the dataset: 2832 for training and 708 for testing the model. The distribution of holograms per reconstruction distance is the same for the training and testing sets.

Regarding imaging performance, the system's resolution is limited by its numerical aperture (NA), which is defined as follows [1],

$$\mathrm{NA} = \frac{W}{2\sqrt{\left(\frac{W}{2}\right)^2 + L^2}}. \tag{4}$$

In Eq. (4), W is the sensor's width, and L is the source to camera distance, as illustrated in Fig. 1. Eq. (4) is valid as long as the sensor is fully illuminated. The lateral resolution δx can be approximated with the given expression for the numerical apertures as shown in Eq. (5) [1,30].

$$\delta x \geq \frac{\lambda}{2\mathrm{NA}}. \tag{5}$$

3. FocusNET

3.1. DLHM training dataset

A setup like the one presented in Fig. 1, has been implemented to record the DLHM dataset to train and test FocusNET. A laser source illumination (632.8 nm He-Ne) is expanded and later focused by an aspheric lens with a diameter of 6.33 mm and a numerical aperture of 0.67. A Panasonic MN34230 camera is used to record the holograms; the sensor width is 13.3 mm with a native pixel pitch of 3.8 μm and 3518 × 3518 square pixels. The holograms are rescaled to 1024 × 1024 pixels, changing the effective pixel pitch to 12.99 μm, which is used in this experiment. Although the latter preprocessing procedure produces low-pass filtering affecting the reconstructed images, this hologram size has been selected after considering a tradeoff between the model efficiency and accuracy; see Section 3.3. for further details on this topic. The source-to-camera distance L is fixed at 18.8 ± 0.1 mm. DLHM holograms with a numerical aperture of 0.33 have been recorded. Thus, a lateral resolution of 0.9 μm is attained. A total of 3540 holograms were acquired, with the only variable being the location of the sample (z), which was adjusted using a Vernier micrometer screw. Of those, 2832 are used for training, and 708 for testing the model, following a training-test ratio of 80:20. The distribution of the number of recorded holograms for each reconstruction distance z is shown in Fig. 2. These distances range from 0.3 mm to 4.1 mm.

The ground truth reconstruction distance, which is the target variable used to train the network, is the location obtained by visually adjusting the reconstruction distance of each hologram during the discrete numerical reconstruction. This is achieved by performing a batch reconstruction procedure on every DLHM hologram in the dataset using the ImageJ plugin for DLHM [28]. The samples used to record the dataset holograms include an optical fiber, plastic circles, a fly's leg, a leaf, a skeletal muscle, a spirogyra, a honeybee wing, epithelial cheek cells, volvox algae, and worm eggs. The samples are uniformly distributed in the dataset, *i.e.*, the number of holograms per specimen is the same. In Fig. 3, three examples of holograms from the DLHM dataset are displayed. These include a hologram of epithelial cheek cells in (a), a plastic circle in (b), and a hologram of epithelial cheek cells with a lower concentration in (c).

3.2. FocusNET architecture

Convolutional Neural Networks (CNN) have played a major role in computer vision and image processing [33] because of their ability to infer features and learn abstract representations from a given set of images. These models have shown their capabilities in various recognition, classification, and segmentation tasks [22,34]. CNNs are an extension of the multilayer neural network [35], thus, their structure is typically a sequence of convolutional layers and filters, followed by one or more fully connected layers [34]. This architecture has the advantages of being shift, scale, and distortion invariant: these properties make CNNs suitable for image processing solutions [36].

Following [26,37], the proposed learning-based model has the architecture presented in Fig. 4. The input layer is composed of a single channel representing the raw DLHM hologram with 8-bit intensity values. Then, a resizing layer is added. The best model's accuracy is obtained when the images are resized to 256 × 256 because of the available training data: models with more parameters are expected to be more data-hungry [33]. This matter is analyzed in Section 3.3. After this step, a data augmentation layer is added; this layer randomly rotates the hologram by 0°, 90°, 180°, or 270°, helping to reduce overfitting. The following preprocessing layer adds a second channel, which represents the amplitude of the Fourier transform of the resized hologram. This layer plays a critical role in increasing the model's generalization, thus further reducing overfitting, as demonstrated by [24] and validated during the experiments in this proposal. Then, the two-channel layer is fed into the convolutional backbone, composed of five convolutional blocks, which is the feature extractor of the model. In this backbone, the model learns the underlying characteristics of the holograms, which will then be used to perform the autofocus estimation. A minimal change in this convolutional backbone structure will significantly diminish the model's performance. Each convolutional block includes a convolutional layer, followed by batch normalization and max pooling layers. As the network goes forward inside the convolutional backbone, the number of filters increases while the transformed image size decreases. Specifically, the first convolutional layer starts with eight neurons multiplied by a size factor that controls the final size of the model in terms of trainable parameters, and each subsequent layer doubles the neurons until reaching the regression head. The transformed data is fed to the regression head after being processed by a dropout layer and the flatten layer. The regression head includes two fully connected layers and an output layer with only one neuron. This output neuron provides Z_{FN}, that is, the reconstruction distance predicted by FocusNET. The *swish* activation function, an improvement of the very popular ReLU activation function [34], is used in all the network layers except for the output layer, which has a linear activation function. The Adam optimizer is used as the optimization algorithm with a variable learning rate. For the best performance models, the training requires 500 epochs, with the learning rate starting at a value of 0.001, decreasing by a factor of 2 every 100 epochs, and then decreasing exponentially after epoch 450. This variation of the learning rate proved to be very effective in improving both the model's flexibility and generalization.

To train FocusNET, the previously described DLHM dataset has been used. The loss function used for optimization is the *mean squared error (MSE)*, which is given by the average of squared errors in the estimation. These errors are the differences between the predicted reconstruction distance Z_{FN} that outputs the CNN model and the corresponding ground

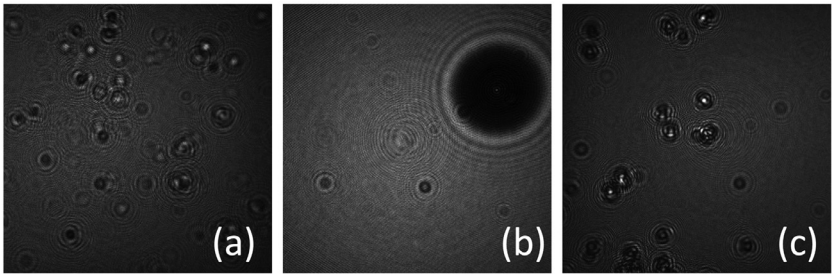

Fig. 3. Three examples of holograms in the DLHM dataset. (a) epithelial cheek cells, (b) plastic circle, and (c) epithelial cheek cells in a lower concentration.

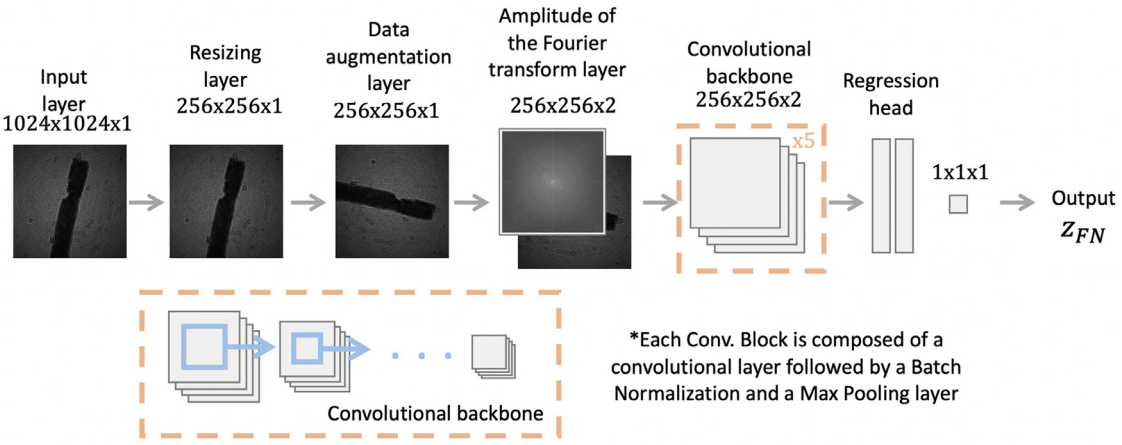

Fig. 4. FocusNET architecture. The network receives an input hologram of 1024 × 1024 pixels, then the preprocessing stage involves a resizing layer, a data augmentation layer and a Fourier transform layer. After the convolutional backbone a regression head outputting a single neuron provides the predicted reconstruction distance of the hologram.

truth value z; thus, the model's loss function is given by Eq. (6).

$$MSE(z,\ z_{FZ}) = \frac{1}{N}\sum_{i=1}^{N}\left(z - z_{\mathrm{FN}}\right)^2. \tag{6}$$

In Eq. (7), N is the total number of holograms processed in each batch. In this proposal, a batch size $N = 64$ has been set for training. Complex loss functions, including the computation of the information of the reconstructed image instead of only the reconstruction distance, could improve network training. However, these methods require larger datasets and more complex network architectures. In future work, we will study the tradeoff between these advanced training strategies, processing time, and prediction accuracy.

3.3. Evaluation metrics and training details

The MAE metric has been used to evaluate the model's performance. The MAE reports the average distance error between the model estimation and the ground truth reconstruction distances. Thus, MAE provides a physical measurement of the estimation error.

$$MAE(z,\ z_{FZ}) = \frac{1}{N}\sum_{i=1}^{N}\left|z - z_{FN}\right|. \tag{7}$$

The model has been evaluated using the typical train-validation approach. The dataset is randomly split into two subsets with a ratio of 80:20 for training and validation. The best-performing model is achieved using the hyperparameters presented in Table 1. Except for the resizing, all those hyperparameter values are adjusted empirically by evaluating the model performance with the MAE and the loss function (MSE) metrics on an epoch-by-epoch basis. The implemented method consists of a grid search with variations in the hyperparameter values, starting with the less-complex (fewer parameters) combinations and choosing the parameters to tune for the next iteration based on some characteristics of the training process, such as convergence rate, stability, overfitting, and best and final values found for the metrics in the training and validation datasets. See Supplementary 1 for further details. The best values for the model hyperparameters are presented in Table 1. "Changing Epoch" refers to the epoch when the exponential decay of the learning rate starts, and "Unchanging Period" refers to the number of epochs where the learning rate remains unaltered before halving its value.

Table 1
The best values found for the model hyperparameters.

Hyperparameter	Value
Training Epochs	500
Batch Size	64
Initial Learning Rate	0.001
Changing Epoch	450
Unchanging Period	100
Decay	0.1
Size Factor	4
Dropout Rate	0.2
Distance Unit	mm
Resizing	256 × 256
Fourier Transform Feature	Amplitude

Because the purpose of FocusNET is to predict the accurate in-focus reconstruction distance without high inference times, the resizing layer is crucial. As the size of the trained model affects this inference time, it becomes an important factor to consider when selecting the final best-performance model. Fig. 5, panel (a) presents the model's prediction time as its size increases on a batch of 50 holograms. The variation of the MAE metric as a function of different model sizes is presented in Fig. 5, panel (b). For smaller models, overfitting is more likely to occur. Conversely, no significant improvement of the metric is achieved

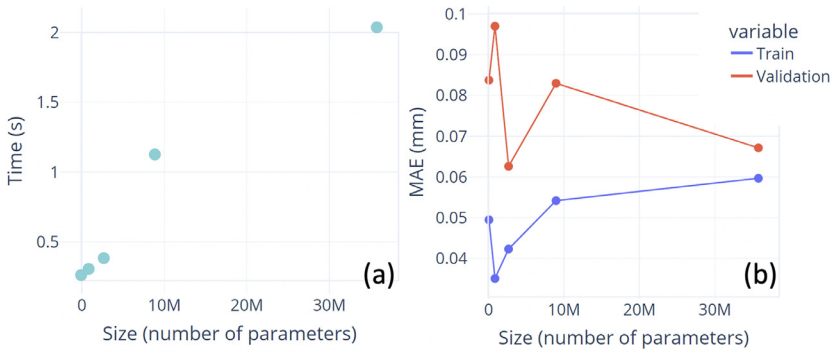

Fig. 5. Performance of FocusNET in terms of the number of parameters. (a) Inference times (seconds) of the proposed network at different model sizes. (b) Errors attained (MAE) at different model sizes.

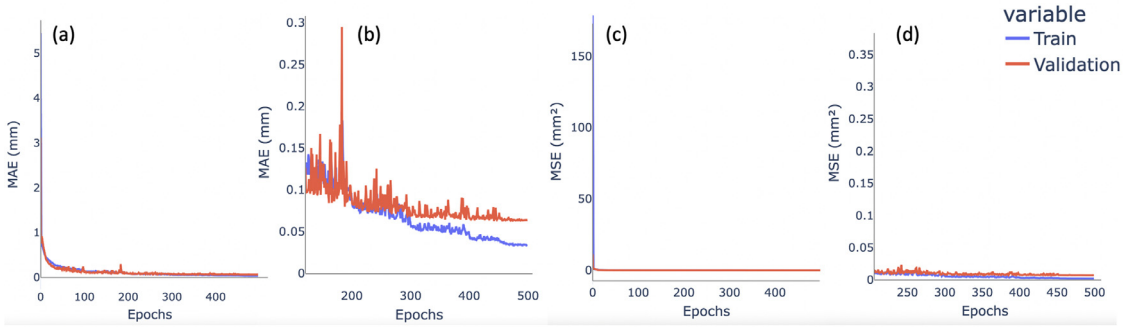

Fig. 6. Training history for the best FocusNET model. (a) Complete MAE history. (b) Zoom in MAE history. (c) Complete loss (MSE) history. (d) Zoom in MSE history.

Table 2
Quality metrics of FocusNET.

Metric	Training	Validation
MSE (mm^2)	0.0019	0.0070
MAE (mm)	0.034	0.064

for larger models. From these results, choosing the 256 × 256 model (third from left to right) is adequate, as it has the lowest combination of training and validation MAE values. The model using this resizing value presents a better generalization while reducing inference time.

All the models have been trained using the Apolo high-performance computer provided by Universidad EAFIT [38]. The acceleration node is powered by three NVIDIA V100 GPUs of 32GB each [39]. The training time varied from 2 h for the less-complex models with fewer epochs to 16 h for the largest models. FocusNET has been trained with the hyperparameters presented in Table 1; the training history is displayed in Fig. 6. Panel (a) presents the complete MAE training history and panel (b) from 150 to 500 epochs. Panel (c) shows the complete loss history (MSE) and panel (d) a zoom in from 150 to 500 epochs. This history shows a healthy training process that presents a higher variance at the beginning and then progressively stabilizes with each iteration. After 400 epochs, the metrics for training continue decreasing, but the ones for validation stop decreasing, causing the series to diverge. At that point, the training process is stopped to avoid overfitting. The results for MSE and MAE metrics are presented in Table 2.

4. Results

4.1. Validation: test dataset

Three random images are initially selected from the validation dataset and then reconstructed based on the FocusNET predicted reconstruction distances. The intensity information of the samples is presented in Fig. 7: human saliva specimens in panel (a), a plastic circle in panel (b), and epithelial cheek cells in panel (c). The standard deviation of the measurements provided by FocusNET over these holograms is 54 μm. The predicted reconstruction distance allows the proper observation of details of the sample, as can be observed in the zoomed-in areas of Fig. 7.

Two random holograms have been picked from the validation dataset to study the quality of the retrieved images. Table 3, Figs. 8 and 9 present the results.

Fig. 8 shows the intensity reconstruction for the optical fiber sample with the reconstruction distance predicted by FocusNET [panel (a)] and using the ground truth reconstruction distance [panel (b)]. In the zoomed-in area of the FocusNET prediction, one can observe diffraction effects at the edge of the sample due to the focusing distance error (observe the green arrow). These effects are not present in the reconstruction presented in panel (b). For a better comparison, panels (c) and (d) present horizontal (c) and vertical (d) intensity profiles of the optical fiber denoted by orange for the FocusNET reconstruction and blue for the ground truth reconstruction lines. The FocusNET-yielded reconstruction presents more diffraction effects resulting in a narrower fiber profile. Nevertheless, the measured width of the fiber only differs by six pixels, corresponding to approximately 6 μm (panel (d)). Also, the background (BG) noise is approximately the same for both reconstructions, and the normalized gradient in the fiber's edge is higher in the ground truth as the edge is sharper than the FocusNET reconstruction.

Fig. 9 presents the intensity reconstruction of a fly's leg hologram with the FocusNET reconstruction distance [panel (a)] and with the ground truth distance [panel (b)] where the *tarsal claws* are focused ($z = 2.38 \pm 0.05$ mm). One interesting effect can be observed in this reconstruction: several in-focus planes can be found. The ground truth distance is relative depending on which part of the sample is observed. For instance, this kind of sample presents a range of well-focused values. If the purple area is analyzed, more details of the sample's hairs can be detailed at the FocusNET distance than at the considered ground truth distance (observe the orange arrows inside the purple square). The op-

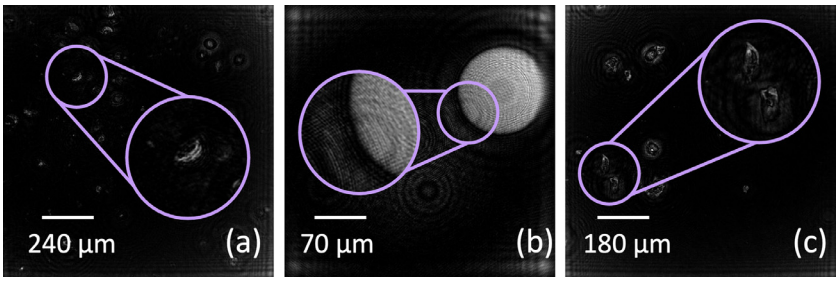

Fig. 7. Intensity reconstructions of three random holograms of the validation dataset with the FocusNET predicted reconstruction distances. (a) human saliva specimens, (b) plastic circle and (c) epithelial cheek cells.

Table 3
FocusNET prediction for two random holograms.

Sample	Prediction time (s)	FocusNET distance (mm)	Reconstruction distance [visual inspection] (mm)	Label distance [ground truth] (mm)
Optical fiber	0.10 ± 0.02	2.055 ± 0.054	2.10 ± 0.05	2.10 ± 0.01
Fly's leg	0.10 ± 0.02	2.308 ± 0.054	[2.26 - 2.38] ± 0.05 (several focal planes)	2.38 ± 0.01

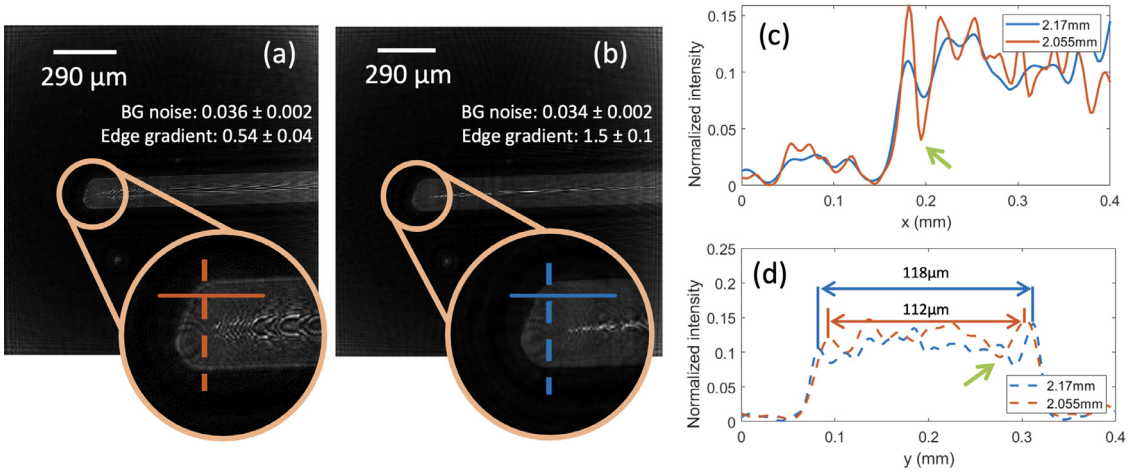

Fig. 8. Intensity reconstruction of the fiber hologram at the FocusNET predicted distance (a) and ground truth reconstruction distance (b). Crossed sections of the fiber reconstruction with the ground truth distance (blue) and the FocusNET distance (red) at x axis (c) and y axis (d).

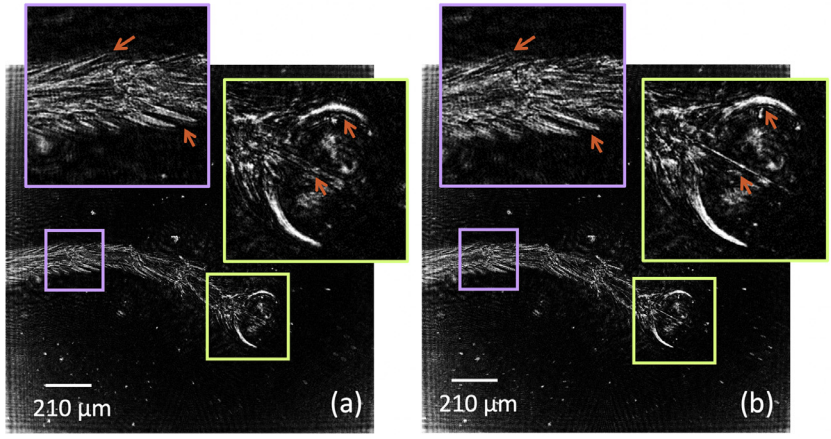

Fig. 9. Intensity reconstruction of the fly's leg hologram at the FocusNET predicted distance (a) and ground truth reconstruction distance (b).

posite occurs for the area enclosed by the green square. More details are observed with the ground truth distance than with the FocusNET-based reconstruction. This result concurs with what is stated in [40]: the accuracy of the autofocusing models depends on the studied sample. When considering the proper reconstruction distance of this hologram after visual inspection, FocusNET predicts its in-focus reconstruction with no error (see Table 3).

4.2. Validation with holograms acquired with different DLHM setups

FocusNET has been trained with a dataset containing holograms acquired with a single DLHM geometry ($L_{FN} = 18.8 \pm 0.1$ mm, $W_{FN} = 13.3mm$, $k_{FN} = 2\pi/\lambda_{FN} = 9.9$ μm^{-1}). Therefore, the predicted distance Z_{FN} corresponds to holograms registered with those geometrical parameters. For instance, considering a hologram I registered with a different

Table 4
FocusNET prediction values for an experimental hologram with different DLHM optical parameters.

Sample	L(mm)	W(mm)	λ(nm)	Ground truth z (mm)	NA	NA difference	FocusNET raw Z_{FN} (mm)	FocusNET corrected z (mm)
Red blood cells	8.0	4.71	532	0.730 ± 0.005	0.31 ± 0.01	0.02 ± 0.01	1.373 ± 0.054	0.7135 ± 0.055
Amplitude positive USAF test	16.5	3.30	405	10.50	0.10 ± 0.01	0.23 ± 0.01	2.727 ± 0.054	10.550 ± 0.055
FocusNET setup	18.8	13.3	632.8	–	0.33±0.01	–	–	–

DLHM geometry ($\boldsymbol{L}$, $\boldsymbol{W}$, $\boldsymbol{k}$), the recovered information of the sample is given by Eq. (3). The reconstruction distance of the hologram I, with its original reconstruction parameters is z_i. If the FocusNET parameters are used to reconstruct this hologram, the information of the sample can be found through Eq. (8).

$$S\left(\vec{r_{0FN}}\right) = \int_{sensor} I \frac{\exp\left[-i\,\vec{\mathbf{k}_{\mathbf{FN}}}\cdot\vec{r_{FN}}\right]}{\left|\vec{r_{FN}}\right|} \frac{\exp\left[-i\,\vec{\mathbf{k}_{\mathbf{FN}}}\cdot\left(\vec{r_{FN}}-\vec{r_{0FN}}\right)\right]}{\left|\left(\vec{r_{FN}}-\vec{r_{0FN}}\right)\right|}\,\mathbf{d}\,\vec{r_{FN}}\,. \tag{8}$$

Let us impose that the information of the sample retrieved with the original geometrical recording parameters ($\vec{r}$ **and** $\vec{r_0}$, $[\boldsymbol{L_i}$ **and** $\boldsymbol{z_i}]$) is the same as the one recovered with the FocusNET parameters ($\vec{r_{FN}}$ **and** $\vec{r_{0FN}}$, $[\boldsymbol{L_{FN}}$ **and** $\boldsymbol{z_{FN}}]$). This means that the magnification of the DLHM hologram in both geometries is the same.

$$\mathbf{M_{FN}} = \frac{\boldsymbol{L_{FN}}}{\boldsymbol{Z_{FN}}} = \boldsymbol{M} = \frac{\boldsymbol{L_i}}{\boldsymbol{zi}} \tag{9}$$

The latter assumption is valid if there is a slight numerical aperture difference between the geometries. In this proposal, the validity of Eq. (10) has been tested up to a 0.23 numerical aperture difference. From Eq. (9), the reconstruction distance with the actual geometrical DLHM parameters (without considering the illumination wavelength λ) can be found from the FocusNET reconstruction distance.

$$z_i = \frac{L_i}{L_{FN}} Z_{FN} \tag{10}$$

Moreover, to retrieve the full same sample information, including λ, Eq. (12) must be fulfilled.

$$\frac{\exp\left[-i\,\vec{\mathbf{k}_{\mathbf{FN}}}\cdot\vec{r_{FN}}\right]}{\left|\vec{r_{FN}}\right|} \frac{\exp\left[-i\,\vec{k_{FN}}\cdot\left(\vec{r_{FN}}-\vec{r_{0FN}}\right)\right]}{\left|\left(\vec{r_{FN}}-\vec{r_{0FN}}\right)\right|} = \frac{\exp\left[-i\,\vec{k}\cdot\vec{r}\right]}{\left|\vec{r}\right|} \frac{\exp\left[-i\,\vec{k}\cdot\left(\vec{r}-\vec{r_0}\right)\right]}{\left|\left(\vec{r}-\vec{r_0}\right)\right|}. \tag{11}$$

Because the phase factors in (11) must be equal on both sides of the equation, then z of the actual DLHM setup is found, Eq. (13).

$$z_i = 2L - \frac{\lambda}{\lambda_{FN}}\left(2L_{FN} - Z_{FN}\right) \tag{12}$$

From this analysis, the following procedure must be followed to obtain the reconstruction distance of a DLHM hologram, regardless of its physical recording parameters.

(i). FocusNET predicts the reconstruction distance Z_{FN}.

(ii). To fulfill the magnification assumption, Eq. (10) is applied over the distance found in i), and z_i is computed.

(iii). Then, Eq. (12) must be applied. If the wavelength, $\lambda = \lambda_{FN}$, then this step can be avoided.

Two holograms acquired with different DLHM setups and diverse samples have been studied to validate the latter analysis. The sample contains red blood cells (RBC) and an amplitude USAF target. The DLHM setup and the reconstruction distance results are displayed in Table 4 together with the NA difference between the experimental setup and the FocusNET setup. In Table 4, the FocusNET dataset parameters are also presented for comparison purposes.

The amplitude and phase reconstructions of the red blood cells sample are presented in Fig. 10. Despite the prediction difference against the visually attained ground truth, the amplitude reconstruction, panel (b) in Fig. 10, allows the identification of the cells. This sample was recorded with a numerical aperture of 0.31, yielding a 0.85 μm lateral resolution. For this sample, the phase map obtained with the FocusNET predicted reconstruction distance also allows the visualization of the cells with some deformations. The phase changes inside the cells allow observing the typical donut-like shape of these specimens, as can be observed in the 3D phase maps of panels (c) and (d).

A hologram of a USAF target is also tested to observe the spatial frequency loss due to defocus. The DLHM geometry used to record this hologram is presented in Table 4. The NA of this setup is approximately 0.10, implying a system with a 2.03 μm of lateral resolution. This resolution can be corroborated by correctly visualizing the USAF test target's element 7–6 (2.2 μm, according to the manufacturer). Fig. 11 shows the reconstruction of the hologram. Panel (a) depicts the amplitude information of the sample reconstructed using the FocusNET corrected distance, and panel (b) shows the reconstruction at the ground truth distance. As can be observed in the orange zoomed-in areas, element 7–6 of the USAF target can be retrieved in the FocusNET prediction, as it is obtained with the ground truth reconstruction. The latter validates FocusNET regarding the proper recovery of spatial frequency information.

Table 5
Conventional autofocusing methods and FocusNET computing times for the optical fiber hologram.

Method	Time (s)
Dubois	57.1 ± 0.1
Gradient	60.7 ± 0.1
Power spectra	69.1 ± 0.1
Variance	64.4 ± 0.1
FocusNET	0.1 ± 0.02

4.3. Comparison with conventional autofocusing algorithms

The main advantage of FocusNET is the computing time required to retrieve the proper reconstruction distance. Once the model is trained, the reconstruction distance prediction is highly efficient. As no numerical reconstructions are needed, changing plane-by-plane reconstruction-based methods (stack methods) for a one-time kernel-image multiplication significantly reduces the computational complexity of the proposal when compared to the conventional approaches. Dubois, Gradient-based, Power spectra, and variance are the conventional methods to validate the latter assumption [12]. The hologram of the optical fiber presented in Fig. 8 is used for this experiment. The value of each metric applied to this hologram as a function of z is presented in Fig. 12. In this figure, the y-axis shows the normalized metrics values, and the x-axis is in millimeters. The distance at which the minimum value is attained is the predicted reconstruction distance by the metrics.

The prediction times obtained for the four conventional techniques and FocusNET are presented in Table 5. The axial step for the stack reconstruction methods is 2 μm. All algorithms run on an intel core i7,

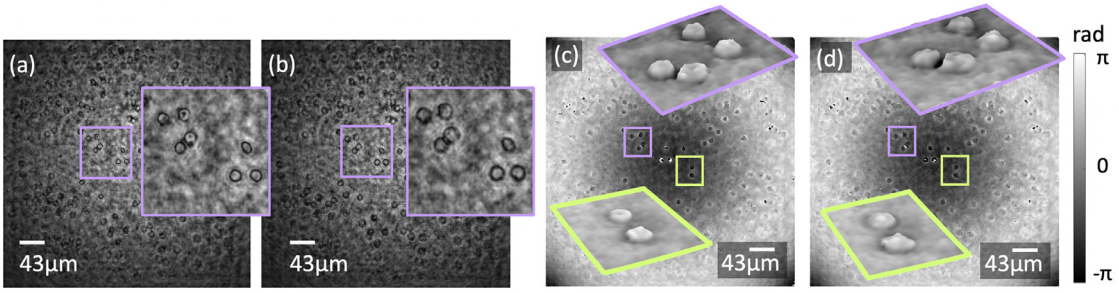

Fig. 10. Amplitude reconstruction of a DLHM hologram. (a) With the ground truth reconstruction distance. (b) With the corrected FocusNET reconstruction distance. Phase map reconstruction of the RBC sample with (c) ground truth distance and (d) corrected FocusNET distance.

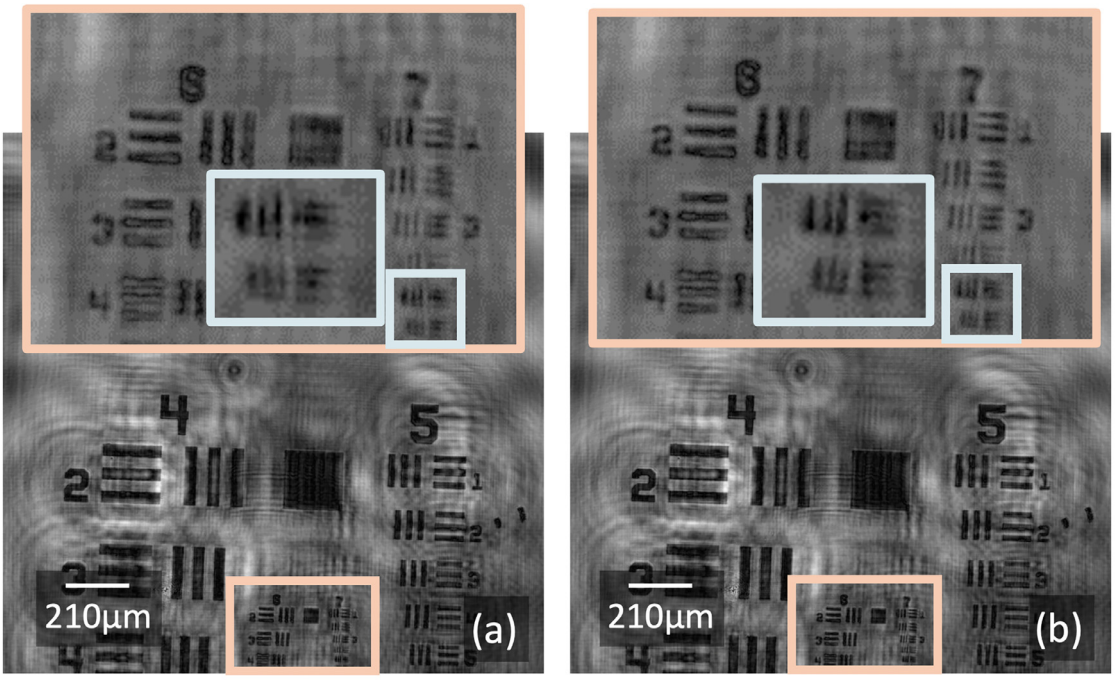

Fig. 11. Amplitude reconstruction of a DLHM hologram of a USAF test target. (a) With the corrected FocusNET's predicted distance. (b) With the ground truth reconstruction distance.

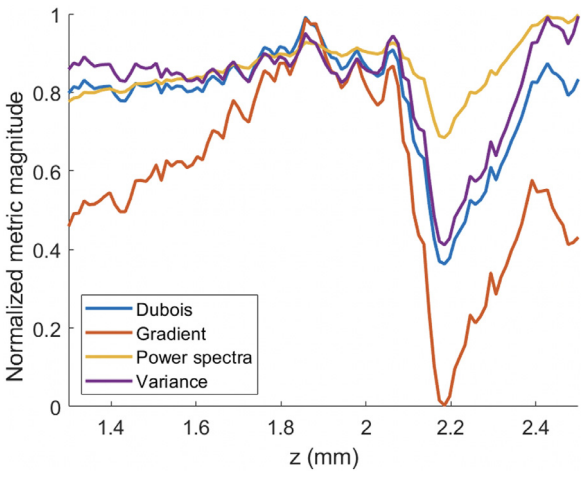

Fig. 12. Performance of the conventional autofocusing metrics for the optical fiber DLHM hologram.

eighth gen with 4 CPU cores. As can be observed, the prediction time of FocusNET is about six hundred times faster than the average time of the classical metrics.

If we dive into the video-rate domain, where a series of holograms need to be processed sequentially in a minimal amount of time, the difference becomes even more prominent. A deep learning-based autofocusing approach can handle batch inference and is highly parallelizable, so the computation time does not increase linearly with the number of inputs, as opposed to the conventional techniques, in which it would be much harder to implement the same level of parallelism. A batch of 10 validation holograms has been tested to prove the latter with FocusNET and the Dubois metric. To test the Dubois metric with the batch of holograms, a parallelized loop is implemented in MATLAB; each hologram is independently reconstructed for each reconstruction distance. As these processes are not sequential, the computing tasks can be easily split into each CPU core. Both algorithms run in a 4 CPU intel core i7 computer. Table 6 presents the prediction time required by the methods. The prediction time with FocusNET for a batch of 10 holograms is about 1.200 times faster than the Dubois-based method in a machine without GPU. These values enable accurately reconstructing in-focus DLHM holograms at video rates. The latter is confirmed when FocusNET

Table 6
FocusNET and Dubois autofocusing methods for a batch of ten holograms.

Method	Total prediction time (s)
Batch Dubois	541 ± 1
Batch FocusNET	0.44 ± 0.02

is tested using a Mac's M1 chip powered by a GPU unit. In this case, the prediction time for a single hologram decreases to 58 milliseconds, and for a batch of 50 holograms, the whole prediction time goes down to 420 milliseconds, this means 8.4 milliseconds per single hologram of 1024×1024 pixels.

The remarkable performance of FocusNET compared to conventional approaches is valid under the network's trained and modeled conditions, say, reconstruction distances ranging from 0.3 mm to 4.8 mm and numerical apertures ranging from 0.1 to 0.33, no matter the imaged specimens. If one desires to implement FocusNET in different conditions, the network training time and data collection steps will increase the overall execution time of the proposal.

5. Conclusions

This paper reports on a regression learning-based autofocusing model, FocusNET, to attain the in-focus reconstruction distances of DLHM holograms. The model is built doing a data augmentation stage followed by a two-channel input including the amplitude of the Fourier transform of the raw DLHM holograms, and a convolutional backbone composed of convolutional layers ending in a conventional regression head. The model is trained with DLHM holograms recorded with a specific geometrical setup in which the sample location ranges from 0.3 mm to 4.8 mm, with a numerical aperture of 0.33. The dataset, training algorithms and the trained FocusNET model are available on the public GitHub repository https://github.com/mmonto95/focusnet. FocusNET predicts reconstruction distances leading to focused phase and amplitude images of the samples with a standard deviation of 54 μm for DLHM holograms of the validation dataset. The difference between the predictions provided by FocusNET and the visually attained ground truth distances is relative, as some samples may present different in-focus planes, as this study corroborates. A physical-mathematical model is also provided to correct the predicted FocusNET reconstruction distance of DLHM holograms acquired at different recording conditions. This strategy is validated with holograms of a biological sample and a USAF test target, yielding successful results at the autofocusing task, demonstrating the proposed learning-based method's generalization to different samples and setups. Compared to conventional autofocusing methods - Dubois, spectrum-based, variance-based, gradient-based –considering only the prediction time, FocusNET is 600-time faster in predicting the correct reconstruction distance of single holograms and 1200-time faster for hologram batches without any stack numerical reconstructions. Therefore, FocusNET paves the way to autofocusing video-rate applications of DLHM, facilitating the usability of the technique in motility analysis for microorganism tracking and cell counting, among others.

Funding

Universidad EAFIT under grants: "*Beca de reconocimiento a la investigación", "Beca de reconocimiento a la excelencia académica"*. Universidad Nacional de Colombia Sede Medellin, grant Hermes 50,069. Universidad de Antioquia. Fundación UdeA. Instituto Tecnológico Metropolitano, grant PE21102. SAPIENCIA (Agencia de Educación Superior de Medellín), grant 358. G8 Medellin Metropolitan Area universities funding "*Proyectos de I + D + i en el marco de la agenda regional I + D->i*".

Disclosures

The authors declare no conflicts of interest.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Manuel Montoya: Methodology, Software, Validation, Writing – original draft. **Maria J. Lopera:** Methodology, Software, Validation, Visualization, Writing – original draft. **Alejandra Gómez-Ramírez:** Data curation. **Carlos Buitrago-Duque:** Data curation. **Adriana Pabón-Vidal:** Resources, Funding acquisition. **Jorge Herrera-Ramirez:** Resources, Funding acquisition. **Jorge Garcia-Sucerquia:** Funding acquisition, Conceptualization, Writing – review & editing. **Carlos Trujillo:** Funding acquisition, Conceptualization, Supervision, Writing – review & editing.

Data availability

The code implementation, documentation, and holograms used are publicly available on GitHub [41]. The GitHub repository also includes troubleshooting guidelines for the correct use of FocusNET.

Acknowledgments

The authors acknowledge supercomputing resources made available by the *Centro de Computación Científica Apolo* at Universidad EAFIT (http://www.eafit.edu.co/apolo) to conduct the research reported in this scientific product. M. Montoya, M. J. Lopera, and C. Trujillo acknowledge the support provided by Vicerrectoría de Ciencia, Tecnología e Innovación from Universidad EAFIT. The authors acknowledge the valuable help of Heberley Tobón-Maya, Samuel Zapata-Valencia, Abraham Camelo, Juan Aguirre, Carlos Mena, Luis Gómez, José Pamplona and José Sánchez, on constructing the training data set.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.optlaseng.2023.107546.

References

[1] Garcia-Sucerquia J, Xu W, Jericho SK, Klages P, Jericho MH, Kreuzer HJ. Digital in-line holographic microscopy. Appl Opt 2006;45:836–50. doi:10.1364/AO.45.000836.
[2] Leith EN, Upatnieks J. Reconstructed wavefronts and communication theory*. JOSA 1962;52(10):1123–30 52:1123–30. doi:10.1364/JOSA.52.001123.
[3] Kreuzer HJ, Meinertzhagen IA, Jericho MH, Xu W. Digital in-line holography of microspheres. Appl Opt 2002;41(25):5367–75 VolIssue41:5367–75. doi:10.1364/AO.41.005367.
[4] Ozcan A, McLeod E. Lensless imaging and sensing. Annu Rev Biomed Eng 2016;18:77–102. doi:10.1146/annurev-bioeng-092515-010849.
[5] Bishara W, Su TW, Coskun AF, Ozcan A. Lensfree on-chip microscopy over a wide field-of-view using pixel super-resolution. Opt Express 2010;18:11181. doi:10.1364/oe.18.011181.
[6] Bishara W, Sikora U, Mudanyali O, Su TW, Yaglidere O, Luckhart S, et al. Holographic pixel super-resolution in portable lensless on-chip microscopy using a fiber-optic array. Lab Chip 2011;11:1276–9. doi:10.1039/c0lc00684j.
[7] Lopera MJ, Trujillo C. Linear diattenuation imaging of biological samples with digital lensless holographic microscopy. Appl Opt 2022;61:B77. doi:10.1364/AO.440376.
[8] Seo S, Su TW, Tseng DK, Erlinger A, Ozcan A. Lensfree holographic imaging for on-chip cytometry and diagnostics. Lab Chip 2009;9:777–87. doi:10.1039/b813943a.
[9] Seo S, Isikman SO, Sencan I, Mudanyali O, Su TW, Bishara W, et al. High-throughput lens-free blood analysis on a chip. Anal Chem 2010;82:4621–7. doi:10.1021/AC1007915/SUPPL_FILE/AC1007915_SI_001.PDF.

[10] Langehanenberg P, von Bally G, Kemper B. Autofocusing in digital holographic microscopy. 3D Research 2011;2(1):1–11 2011;2. doi:10.1007/3DRES.01(2011)4.
[11] Trujillo Anaya C.A. Improved-performance digital holographic microscopy 2018.
[12] Trujillo C, Garcia-Sucerquia J. Cooperative execution of auto-focusing metrics in digital lensless holographic microscopy for internal-structured samples. Appl Opt 2017;56(21):5877–82 VolIssuePp56:5877–82. doi:10.1364/AO.56.005877.
[13] Lyu M, Yuan C, Li D, Situ G. Fast autofocusing in digital holography using the magnitude differential. Appl Opt 2017;56:F152. doi:10.1364/ao.56.00f152.
[14] Huang L., Liu T., Yang X., Luo Y.I., Rivenson Y., Ozcan A. Holographic image reconstruction with phase recovery and autofocusing using recurrent neural networks 2021.
[15] Yourassowsky C, Schockaert C, Dubois F, Callens N. Focus plane detection criteria in digital holography microscopy by amplitude analysis. Opt Express 2006;14(13):5895–908 VolIssuePp14:5895–908. doi:10.1364/OE.14.005895.
[16] Yourassowsky C, Dubois F, Campbell L, Olson RJ, Sosik HM, Abraham A, et al. High throughput holographic imaging-in-flow for the analysis of a wide plankton size range. Opt Express 2014;22(6):6661–73 VolIssuePp22:6661–73. doi:10.1364/OE.22.006661.
[17] Langehanenberg P, Ivanova L, Bernhardt I, Ketelhut S, Vollmer A, Dirksen D, et al. Automated three-dimensional tracking of living cells by digital holographic microscopy. J Biomed Opt 2009;14:014018. doi:10.1117/1.3080133.
[18] Dirksen D, Droste H, Kemper B, Deleré H, Deiwick M, Scheld HH, et al. Lensless Fourier holography for digital holographic interferometry on biological samples. Opt Lasers Eng 2001;36:241–9. doi:10.1016/S0143-8166(01)00053-7.
[19] Finizio A, Pierattini G, Coppola G, Ferraro P, De Nicola S. Digital holographic microscope with automatic focus tracking by detecting sample displacement in real time. Opt Lett 2003;28(14):1257–9 VolIssuePp28:1257–9. doi:10.1364/OL.28.001257.
[20] Lam EY, Ren Z, Xu Z. Learning-based nonparametric autofocusing for digital holography. Optica 2018;5(4):337–44 VolIssue5:337–44. doi:10.1364/OPTICA.5.000337.
[21] Trujillo C, Garcia-Sucerquia J. Comparative analysis of the modified enclosed energy metric for self-focusing holograms from digital lensless holographic microscopy. Appl Opt 2015;54:5102. doi:10.1364/ao.54.005102.
[22] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44 2015 521:7553. doi:10.1038/nature14539.
[23] Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks 2022.
[24] Shimobaba T, Kakue T, Ito T. Convolutional neural network-based regression for depth prediction in digital holography. IEEE Int Symp Ind Electr 2018:1323–6 2018-June. doi:10.1109/ISIE.2018.8433651.
[25] Xu Z., Ren Z., Lam E.Y.M. Autofocusing in digital holography using deep learning 2018:56. doi:10.1117/12.2289282.
[26] Lam EY, Ren Z, Xu Z. Learning-based nonparametric autofocusing for digital holography. Optica 2019;5(4):337–44 VolIssuePp5:337–44. doi:10.1364/OPTICA.5.000337.
[27] Goodman J.W. Introduction to Fourier optics 1968.
[28] Trujillo C, Piedrahita-Quintero P, Garcia-Sucerquia J. Digital lensless holographic microscopy: numerical simulation and reconstruction with ImageJ. Appl Opt 2020;59:5788. doi:10.1364/ao.395672.
[29] Piedrahita-Quintero P., Trujillo C., García-Sucerquia J. Digital Lensless Holographic Microscopy plugin for ImageJ 2022 https://unal-optodigital.github.io/DLHM/.
[30] Kreuzer HJ, Pomerleau N, Blagrave K, Jericho MH. Digital in-line holography with numerical reconstruction. SPIE Interferometry '99: Tech Technol 1999;3744:65–74. doi:10.1117/12.357756.
[31] Tobon H, Garcia-Sucerquia J, Trujillo C. Preprocessing in digital lensless holographic microscopy for intensity reconstructions with enhanced contrast. Appl Opt 2021;60(Issue 4):A215–21 VolPp60:A215–21. doi:10.1364/AO.404297.
[32] Trujillo C, Garcia-Sucerquia J. Numerical dark field illumination applied to experimental digital lensless holographic microscopy for reconstructions with enhanced contrast. Opt Lett 2018;43. doi:10.1364/OL.43.004096.
[33] Aloysius N, Geetha M. A review on deep convolutional neural networks. In: Proceedings of the IEEE international conference on communication and signal processing; 2018. p. 588–92. ICCSP 20172018-January. doi:10.1109/ICCSP.2017.8286426.
[34] Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks 2022.
[35] Cun Y Le, Denker JS, Henderson D., Howard R.E., Hubbard W., Jackel L.D. Handwritten Digit Recognition with a Back-Propagation Network 2022.
[36] Goodfellow Ian. Yoshua Bengio and AC. Deep lLearning. Genet Program Evolvable Mach 2017;19:305–7.
[37] Xu Z., Ren Z., Lam E.Y.M. Autofocusing in digital holography using deep learning 2018:56. doi:10.1117/12.2289282.
[38] EAFIT U. Home - Apolo - Universidad EAFIT n.d. https://www.eafit.edu.co/apolo (accessed September 19, 2022).
[39] EAFIT U. Apolo Scientific Computing Center's Documentation — apolo-docs 0.1 documentation n.d. https://apolo-docs.readthedocs.io/en/latest/index.html (accessed September 19, 2022).
[40] Buitrago-Duque C, Garcia-Sucerquia J. Sizing calibration in digital lensless holographic microscopy via iterative Talbot self-imaging. Opt Lasers Eng 2020;134:106176. doi:10.1016/J.OPTLASENG.2020.106176.
[41] Montoya Zuluaga M., Lopera M.J., Trujillo C. FocusNET repository. 2022 n.d. https://github.com/mmonto95/focusnet (accessed September 26, 2022).