Research Article | Vol. 14, No. 7 / 1 Jul 2023 / Biomedical Optics Express 3172

Biomedical Optics EXPRESS

Fourier ptychographic microscopy image enhancement with bi-modal deep learning

LYES BOUCHAMA,[1,2,*] BERNADETTE DORIZZI,[1] MARC THELLIER,[3] JACQUES KLOSSA,[2] AND YANECK GOTTESMAN[1]

[1] *Samovar, Télécom SudParis, Institut Polytechnique de Paris, 91120 Palaiseau, France*
[2] *TRIBVN/T-life, 92800 Puteaux, France*
[3] *Sorbonne Université, INSERM, Institut Pierre Louis d'Epidémiologie et de Santé Publique, AP-HP, Hôpital Pitié-Salpêtrière, 75013 Paris, France*
[*] *lbouchama@tribvn.com*

Abstract: Digital pathology based on a whole slide imaging system is about to permit a major breakthrough in automated diagnosis for rapid and highly sensitive disease detection. High-resolution FPM (Fourier ptychographic microscopy) slide scanners delivering rich information on biological samples are becoming available. They allow new effective data exploitation for efficient automated diagnosis. However, when the sample thickness becomes comparable to or greater than the microscope depth of field, we report an observation of undesirable contrast change of sub-cellular compartments in phase images around the optimal focal plane, reducing their usability. In this article, a bi-modal U-Net artificial neural network (i.e., a two channels U-Net fed with intensity and phase images) is trained to reinforce specifically targeted sub-cellular compartments contrast for both intensity and phase images. The procedure used to construct a reference database is detailed. It is obtained by exploiting the FPM reconstruction algorithm to explore images around the optimal focal plane with virtual Z-stacking calculations and selecting those with adequate contrast and focus. By construction and once trained, the U-Net is able to simultaneously reinforce targeted cell compartment visibility and compensate for any focus imprecision. It is efficient over a large field of view at high resolution. The interest of the approach is illustrated considering the use-case of *Plasmodium falciparum* detection in blood smear where improvement in the detection sensitivity is demonstrated without degradation of the specificity. Post-reconstruction FPM image processing with such U-Net and its training procedure is general and applicable to demanding biological screening applications.

© 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

1. Introduction

The digitization of biological specimens with slide scanners at microscopic resolution over large dimensions has been experiencing an undeniable growing popularity for several years now. Such craze is explained by Whole-Slides-Imaging (WSI) practice that enables the storage, archive and exchange of samples between medical specialists, numerically. It is also due to the rapid emergence over the last ten years of new automated diagnosis approaches using, among others, Deep-Learning (DL) algorithms. As a result, digital pathology is promised to become more and more efficient. Nevertheless, scanners technologies still need to progress and some of their limitations are to be addressed. A first limitation concerns the nature of images (or modalities) that are traditionally recorded and exploited. In particular current slide scanners are dedicated to the acquisition of the light intensity that is transmitted through the sample. Other useful morphological information carried by phase such as sample optical thickness or dry-mass is left aside. Intensity and phase images represent two distinct modalities since attached to two distinct light-matter interaction processes. The first one is related to absorption whereas the second to optical path. These distinctive image modalities, when available, could be beneficially exploited jointly to further increase the sensitivity and efficiency of automated diagnosis (see

#489776 https://doi.org/10.1364/BOE.489776
Journal © 2023 Received 16 Mar 2023; revised 16 Apr 2023; accepted 4 May 2023; published 7 Jun 2023

Biomedical Optics EXPRESS

for example Optical Twin For Diagnosis project [1]). A second limitation concerns the ultimate image quality that can be achieved by current scanners. It is imposed by trade-off considerations between optical resolution, exploitable surface per field of view and image production throughput. In particular, high resolution imaging requires the employ of objective lens of high NA and important magnification G (typically G=40x and NA=0.9 or above). In such situation, the depth of field (DOF) is very limited. Depending on the nature of the specimen that is observed, all the objects in the field of view (FOV) are not guaranteed to be in focus. This problem is important and different post-acquisition image processing techniques together with some adaptation in the acquisition workflow have been developed. They permit to extend the microscope depth of field (see for example Bian et al. review article [2]). Images with extended DOF are obtained after correcting digitally focus imprecision with post-acquisition calculations or image processing. A popular approach consists in z-stacking microscopy where images are captured at various planes of focus (typically 10 planes or more). Extended DOF is the obtained by combining these different images adequately. An extension of the approach with higher throughput has been reported in [3] where only two out of focus images are recorded and fused with a deep neural network (DNN).

These limitations are undoubtedly to be reconsidered with recent microscopy advances [4]. Among others, one can think of Fourier Ptychographic Microscopy (FPM) [5]. Additionally to the traditional intensity, this microscopy allows to assess the phase information of biological samples in a simple way [6] (i.e. without any use of interference setup). Such information is theoretically sufficient to calculate numerically images at various planes of focus [7]. Also, images can be produced with a super-resolution factor that can be high [5] thanks to a numerical aperture synthesis that is at the core of FPM operating principles. When compared to conventional microscopy, a larger FOV is accessible for a fixed resolution because of the improved space-bandwidth product. At the same time, FPM depth of field (DOF) is natively important. It is directly related to the NA of the objective lens employed irrespective of the super-resolution factor. That means that the DOF of FPM can be much more important than in traditional microscopy (at comparable resolution).

Since the first demonstration of microscope operating principle in 2013 by Zheng et al. [5], a lot of progress has been made, as can be judged from the number of published results and the variety of topics that have been addressed. For instance, one can think of digital refocusing of images during their reconstruction (see digital wavefront correction section in [5]), reduction of acquisition time allowing high scan rates [8], embedded correction of the aberrations [9], reconstructed tomography of biological samples [10–12]. Nevertheless, the full exploitation of FPM images remains still challenging because of focus variations across large FOV. In [13], an approach permitting obtaining images with extended depth-of-field have been introduced. In [14] a lateral shift effect in raw intensity images is exploited to estimate Δz, the defocus distance. This value is then used as a reconstruction parameter to generate sharp images. Such process can be problematic when the sample thickness is comparable to the microscope DOF. In this case and as will be detailed, some important variability in the contrast and sharpness of sub-cellular components can be encountered in phase images. As a result, the Δz parameter that is estimated from intensity images only is not necessarily adapted to the reconstruction of phase images.

In this article we develop a deep neural network that takes advantage of intensity and phase images (in the same vein of [15]) combined to virtual z-stacking. The bi-modal DNN is trained to reinforce the contrast and sharpness of targeted sub-cellular compartments. To this end, z-stacking is realized virtually by calculating images at finely-spaced planes of focus with a reconstruction algorithm. Images of the highest quality are then isolated to construct a reference database for the DNN training. Here, the optimal focal plane is selected depending on the modality. We demonstrate quantitatively the impact of the DNN from an automated diagnosis point of view. For this, the use-case of *Plasmodium falciparum* detection in blood smears is undertaken as an

Biomedical Optics EXPRESS

example. It is chosen, noting that the problem of parasites detection with the required sensitivity and specificity has hitherto not been solved satisfactorily using blood smear images only [16]. It is also a typical problem where sample analysis over a large surface is needed and hence where local correction is particularly appreciable for a complete exploitation of FPM extended FOV. The paper is organized as follows: In section 2, we briefly recall the operating principle of an FPM imaging system and detail some specificities of the reconstructed intensity and phase images on stained parasitized cells under varied focus conditions. In particular, the contrast variability of phase images characteristics near optimal focus-plane is highlighted. In section 3, we introduce the proposed image post-processing treatment to enhance FPM microscopy image quality. The details of the artificial neural network architecture (U-Net) that is implemented is detailed together with its resulting autofocusing model. Also, the virtual z-stacking procedure used to constitute the (simplified) reference database from large FOV FPM images for training and evaluation is indicated. Finally, section 4 is devoted to the presentation of experimental results. In contrast to SSIM or image quality metrics traditionally used to evaluate the consequence of image correction, here the DL compensation is discussed quantitatively by studying its impact on a DNN-based parasitized red blood cell classifier. Sensitivity and specificity of parasites detection are evaluated with or without U-Net image enhancement. At last, perspectives of this work are indicated in section 5.

2. FPM Reconstructed image characteristics

2.1. Fourier Ptychographic Microscopy Principle

The Fourier Ptychographic Microscopy [5] technique is a computational microscopy that combines angular diversity illumination of samples and data calculations (reconstruction process) to access the phase of the sample in addition to its intensity. Images are reconstructed with super-resolution through synthetic aperture process. A matrix composed of n individual punctual led sources replaces the traditional Köhler illumination [17,18]. A set of n images $I^{(i)}$ (i varying from 1 to n) is acquired for different angular incidence of light on the specimen. In the limit of thin thickness approximation, the sample is modeled by a complex mask function T and the electromagnetic field E_o exiting out of the sample is:

$$E_o(x, y) = E_{in}(x, y) \cdot T(x, y) \tag{1}$$

where E_{in} represents the electromagnetic field immediately below the sample. For the ith led used and using further local plane wave approximation,

$$E_{in}^{(i)} = Ae^{j(k_x^{(i)} \cdot (x) + k_y^{(i)} \cdot (y))} \tag{2}$$

with , $k_x^{(i)} = \frac{2\pi\nu_{x_0}^{(i)}}{\lambda}, k_y^{(i)} = \frac{2\pi\nu_{y_0}^{(i)}}{\lambda}$ the projection of the k-vector along the x and y axes associated to the ith led illumination. The ith image $I^{(i)}$ recorded by the camera is:

$$I^{(i)} = |E_{in}^{(i)} \cdot T * PSF|^2 \tag{3}$$

Since the spectrum of $E_{in}^{(i)}$ is $A \cdot \delta(\nu_x - \nu_{x_0}^{(i)}, \nu_y - \nu_{y_0}^{(i)})$, Eq. (3) can be rewritten as

$$I^{(i)} = |A \cdot \mathcal{F}^{-1}[\widehat{T}(\nu_x - \nu_{x_0}^{(i)}, \nu_y - \nu_{y_0}^{(i)}) \cdot CTF(\nu_x, \nu_y)]|^2 \tag{4}$$

It is at this level that the principle of ptychography in Fourier space comes into play. We notice from Eq. (4) that each captured image is associated to a portion of the spectrum of T, centered in Fourier domain on $(\nu_{x_0}^{(i)}, \nu_{y_0}^{(i)})$. Hence the different angular illuminations that are employed permit to pave the Fourier space. It gives access to T over an extended spectral domain as

Biomedical Optics EXPRESS

compared to the one accessible by the microscope objective lens alone. The synthetic aperture mechanism is obtained after combining the different raw images and recovering the phase in the Fourier Domain. Phase recovery is indeed requested to stitch and combine the different information that are partial since sole the modulus of the filtered field is measured. The phase is reconstructed exploiting the fact that the CTF support associated with the microscope objective is bounded. Different ptychography reconstruction algorithms are available including the e-PIE [19] or Gerchberg–Saxton [20,21] algorithms. Both consist in applying constraints on T in the direct space (knowledge of the modulus of the field $I^{(i)}$) and in the Fourier space (knowledge of the support of the filter performed by the lens). This allows the complex spectrum reconstruction of T over the spectral domain paved by the different angles of illumination.

2.2. Microscope configuration

As mentioned in introduction, the samples considered hereafter consist in parasitized blood smears with *P. falciparum*. They are stained with May-Grünwald Giemsa coloration. The usual resolution required to detect parasites of $\sim 250nm$ [22,23] can be achieved with a numerical aperture NA=1. Considering commercially available microscope objective lenses, a good compromise between NA and measurement time, led us to use a 10x objective lens with a numerical aperture of NA=0.45. The light source consists of 12 LEDs arranged on a ring [6,24] and an additional central LED (see Fig. 1(a)). It is realized with a NeoPixel RGB 5050 ring of radius $r = 15mm$ with 12 LEDs (from Adafruit) and an additional NeoPixel central LED. This light source replaces the traditional Köhler illumination in a Nikon upright microscope (see Fig. 1(a)). The diodes-to-sample distance d is adjusted to achieve a synthetic numerical aperture $NA_{synth} = 0.9$, with d= 30 mm. The overlap in Fourier space is ~ 67 % as required [25]. The camera used (model IDS UI-3200SE) integrates a large surface sensor of 4104x3006 pixels (with a pixel pitch of $3.45\mu m$). This corresponds to a FOV of 1.41 X 1.03 mm^2. The choice of this microscope configuration is mainly motivated by the fact that it permits a good compromise between achievable optical resolution and FOV surface as well as a short acquisition time of 0.5s per field of view [26].

2.3. FPM depth of field

The theoretical depth-of-field of a conventional microscope is dictated by

$$DOF = \frac{\lambda \cdot n}{NA^2} + \frac{n \cdot \Delta x}{M \cdot NA} \tag{5}$$

where n is the refractive index of the medium between the sample and the objective lens, Δx the camera pixel pitch, M the objective lens magnification factor. It is evaluated to 3.35 µm at λ = 525 nm for our microscope configuration.

The previous formula should also provide a good approximation of the DOF for FPM microscopy although no equivalent formula has yet been derived theoretically, for the following reason: consider a planar sample placed in the vicinity of z=0. The raw images recorded for any value of z between -DOF/2 and DOF/2 cannot show any appreciable change. As a consequence, reconstructed images cannot exhibit appreciable changes either. To illustrate this point, we investigate the DOF experimentally using a USAF 1951 resolution chart for two different microscope configurations. The first one is related to a FPM setting. The second one consists in a conventional microscope under coherent illumination condition (along z direction) with an objective lens 40xNA0.95. These two configurations are used to compare the DOF of conventional and FPM microscopes of similar resolution. The procedure used is as follows: In a first step, the resolution chart is placed at z=0 and its smallest observable detail is identified. In a second step, the resolution chart is progressively moved along z until the previously identified

Biomedical Optics EXPRESS

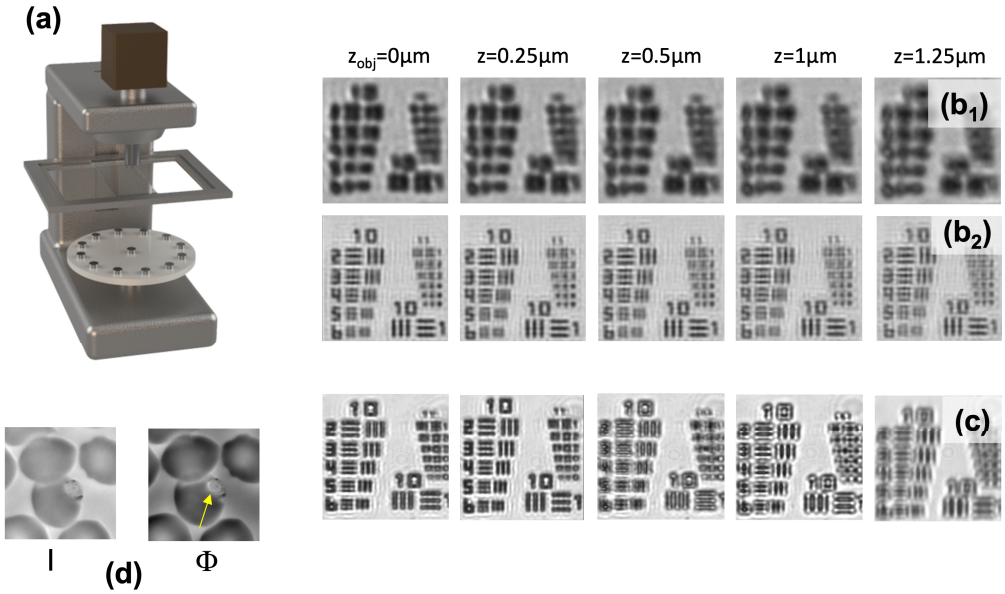

Fig. 1. a) sketch of a microscope equipped with its LEDs matrix. b) USAF1951 resolution chart for various positions of the z-objective lens 10xNA0.45, b_1) raw central led, b_2) intensity reconstructed FPM images. c) Resolution chart captured under central LED illumination with an objective lens 40xNA0.95. Comparison of b) and c) exhibits a different DOF at comparable resolution. d) FPM intensity and phase images obtained on a parasitized red blood cell. The arrow points to a compartment that is not visible in the intensity image.

smallest detail is no longer observable. The value z_{max} thus measured, provides an approximate value of DOF/2.

Some illustrative FPM experimental results are indicated in Fig. 1(b)) for different z. The Fig. 1$(b)_1$) corresponds to central LED illumination and Fig. 1$(b)_2$) to reconstructed images. The finest observable detail at $z = 0$ is attached to element 5 of group 10 for which the width of each bar is 308 nm. From the observation of the different reconstructed images, the value of $|z_{max}|$ is evaluated to $\sim 1.75\mu m$ (corresponding image not presented). The DOF is hence close to 3.5 μm, in accordance with Eq. (5). The same procedure is repeated for the conventional microscope configuration (under coherent illumination). Experimental images are presented Fig. 1(c)). This time, the DOF is here evaluated to $\sim 0.7\mu m$ (also in accordance with eq (5). It is interesting to observe that the DOF of FPM is essentially imposed by the native numerical aperture of the microscope objective rather than by its synthetic numerical aperture. Thus, for a fixed resolution, FPM can exhibit an increased DOF compared to conventional microscopy, provided a large super-resolution factor is used.

Since the thickness of a blood smear sample is $\sim 4\mu m$, an image refocusing is necessary to obtain in-focus images over its complete FOV. Raw images can be used to calculate images at a focus plan Δz away from the one used during acquisition, provided that an electromagnetic propagation model of an air layer of thickness Δz_{num} is introduced in the forward model of microscope image formation used for reconstruction. The original e-PIE [19] engine already includes such possibility using digital wavefront correction. The calculations are similar to those introduced in digital holography with back-propagation calculations. This possibility is useful to compensate for approximate focus conditions. It also permits a fine tuning of focus conditions a posteriori when necessary (for example for phase images exploration) around optimal focal plane.

Biomedical Optics EXPRESS

2.4. Characteristics of reconstructed images

Our primary motivation is oriented toward improvement in the performance of the classifiers that are used to establish automated diagnosis. Indeed and as can be beneficial, phase and intensity modalities are complementary since respectively attached to the dry-mass of the sample that is measured and its absorption. An illustrative example is presented Fig. 1(d)) belonging to a parasitized red-blood cell. As can be observed, a compartment of the parasite cell almost invisible in intensity image is clearly observable in the phase image. It is hence reasonable to imagine that these two types of images, if exploited conjointly, should help classifiers performances provided that their visibility are adequately controlled when necessary.

Figure 2 presents typical FPM images obtained on a parasite for different focus settings, i.e. position of the objective lens along vertical z-axis (ranging from $z = -3\mu m$ to $z = +3\mu m$, with an incremental step of $1\mu m$). Here $z = 0$ is defined by the optimal position deduced from direct observation of the sample using central led illumination. It is to be observed that these images reveal a striking contrast variability around the optimal z position of the microscope objective. In particular, for phase images, one cellular compartment of the parasite (hemozoin) changes from black to white contrast with z (see arrow in Fig. 2(b)) without important red-blood cells sharpness modifications. Such contrast variation is striking since a slow varying evolution of phase function is expected for small defocus distances along z-axis. Although such observation has not been reported to date by other authors (to the best of our knowledge), this effect is not so unusual. We also encountered it using other microscope objective lenses (20xNA0.75 and 40xNA0.90). We believe that this is due to the fact that the thin sample assumption at the core of the FPM principle is not fully adapted to blood smears description. As a consequence, the important phase variation is most likely to be interpreted as a reconstruction artefact. In order to further define the best image that is exploitable by a parasitologist, images obtained for $z = 0$ are explored using numerical focus compensation in the reconstruction (virtual z-stacking) with Δz_{num} ranging from $-3\mu m$ to $+3\mu m$ with a step of 0.1μm. The usability of the calculated images has been labeled by an expert parasitologist for ground truth definition considering focus, contrast and visibility of the different parasites cell compartments (cytoplasm, nucleus, hemozoin and vacuoles). The chosen I and Φ images that best reveal the parasites, are indicated at right hand side of Fig. 2 (reference). They have been obtained for $\Delta z_{num} = 0.7\mu m$ and $\Delta z_{num} = -1.7\mu m$ for the intensity and phase images respectively. It is not surprising that the optimal setting of z for central led illumination and FPM-reconstructed intensity images are close. The value of optimal Δz_{num} reconstruction parameter when considering FPM-phase image is however more singular and can be important (here $\Delta z_{num} = -1.7\mu m$). We presumably attribute this to a consequence of the sample preparation that has been employed (fixation, glue, slide cover-glass) which modify phase properties of the samples. Also, it may be influenced by the parasite cell compartment position within red-blood cell volume.

In other words, and regarding FPM acquisition, z-positioning of microscope objective lens can be delicate for several reasons: 1) The optimal z position of the microscope objective lens depends on the modality (intensity or phase). Since, the phase image cannot be observed directly, it requires many image reconstructions that are time consuming. 2) The random position of the parasites within the volume of the cells does not necessarily allow a systematic adjustment of focus settings. 3) The variability of the phase contrast, when not adequately controlled, is likely to reduce the performance of the classifiers. 4) When one is interested to produce an RGB image, the microscope objective lens exhibits some chromatic aberrations that should be compensated for each color channel.

From our point of view, exploring the samples with numerical focus compensation, although useful, is time-consuming and hence not really adapted for automatic sample diagnosis since the reconstruction of intensity and phase images implies calculations with a complexity C proportional to N_f, the number of planes of focus to be explored. C is $\sim 4nlog(n) * N * K * N_f$

Biomedical Optics EXPRESS

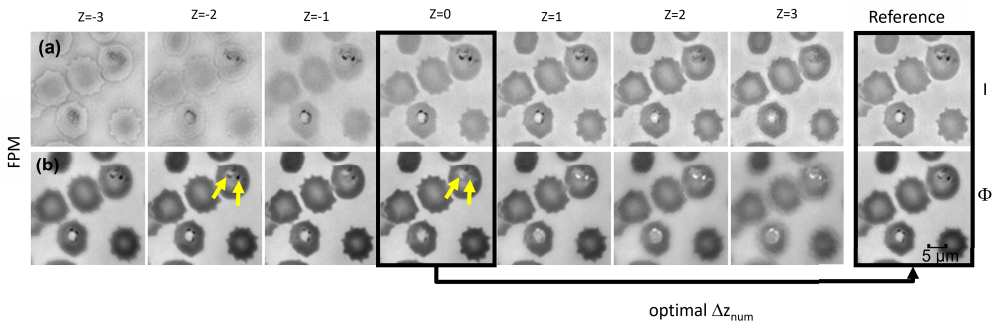

Fig. 2. FPM intensity (I) and phase (Φ) images of stained red blood cells acquired for different focus settings (ranging from $z = -3\mu m$ to $z = +3\mu m$) for $\lambda = 525nm$. Note the relative contrast evolution of one component of the parasite cell (hemozoin) indicated by the arrow. The images enclosed in the right box side correspond to the optimal images obtained with e-PIE numerical focus compensation ($\Delta z_{num} = 0.7\mu m$ for I, $\Delta z_{num} = -1.7\mu m$ for Φ).

where n represents the number of pixels composing a small image patch, N the number of LEDs used and K the number of patches.

3. Deep model

We introduce a strategy based on a deep-neural network model that is combined to virtual z-stacking calculations for the reference database constitution. Once trained, the DNN is in charge of correcting images locally (i.e. at sub-cellular compartments scale) rather than globally, in order to produce contrasted and sharp details of parasites. The locality of the correction is an important contribution of the model since, most of the time, focus needs to be adjusted depending on the cell position within the FOV of FPM. The definition of what is a sharp and contrasted is here learned intrinsically by examples rather than defined by formal properties of the image [27]. This is particularly advantageous for phase images. Furthermore, and in contrast to the possible image corrections discussed in section 2.3, the neural network approach proceeds differently. It does not seek to identify the optimal focal plane per modality but rather produces a clear image for each of the two modalities (intensity and phase) in a single step. The neural network model statistically learns the transformation that needs to be applied on the reconstructed images thanks to a database of reference images. This learning database is composed of pairs of reconstructed images associated with their corresponding refocused references. Once the parameters of the network, namely its weights have been learned, the network can be used on FPM images of any size.

3.1. Deep learning architecture

Among the different deep network architectures of the state of art, we chose a U-Net architecture [28], which has been shown to be very efficient on images for tasks such as segmentation, denoising, super-resolution, etc. [29]. One advantage of such model is that it is possible to consider and process jointly the 2 intensity-phase images, which will therefore be able to interact during the transformation process. The U-Net is an encoder-decoder model shown in Fig. 3, which takes as input non-focused (I, Φ) images and produces at the output the corresponding contrasted and focused images.

The first part of the network, the "compressive" part, (left side of the network) consists in several filtering with convolutions and down-sampling steps aiming at learning an abstract representation of the 2 input images. This resulting feature map is then used by the "expansive"

Biomedical Optics EXPRESS

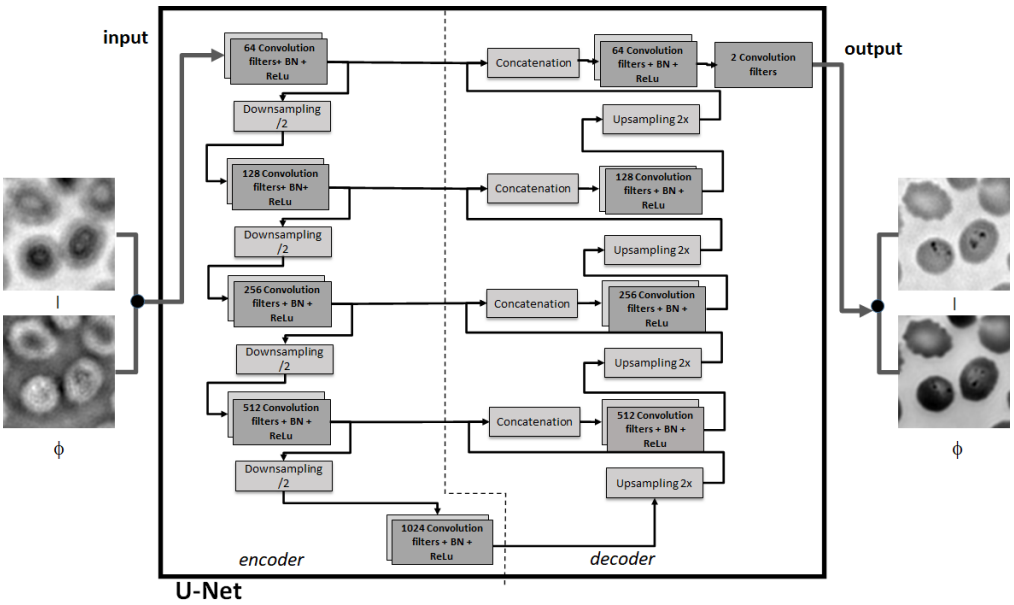

Fig. 3. Standard architecture of the U-Net. The compressive part of the U-Net (left-side) consists in three operations as in classical CNN that are repeated 4 times: two successive filter blocks of 3x3 convolution kernels, batch normalization (BN) and a non-linear ReLu activation function, followed by a max-pooling (2x2). At each step the dimension of the images is reduced by a factor 2. The expansive part of the U-Net (right-side) consists in (2x2) upsampling using transposed convolution, connections with left-part (through a skip-connection and concatenation) and two successive filter blocks of 3x3 convolution kernels, batch normalization and ReLu activation function. The numbers indicated in each convolution block represents the number of filters.

part of the network (right side of the network) composed of symmetric up-sampling steps in order to produce the 2 output images. Symmetric skip connections between the down sampling convolutional layers and the corresponding up-sampling convolutional layers, allow preserving the relevant details in the input image. For more details on the U-Net, see [28].

3.2. Training flowchart

The U-Net model needs to be trained in order to optimize its weights. All the steps involved for the construction of the training dataset are indicated in Fig. 4. In a first step the optimal position of the microscope objective lens is found with FPM central led. This defines the reference z-position, $z = 0$. The blood smear is then acquired for various focus conditions (z ranging from $z = -8\mu m$ to $8\mu m$ with a $2\mu m$ step) and using each time 13 successive angles of illuminations. During the 3rd step, the FOV restricted to its central 256x256 pixels (after cropping raw images) is exploited. These dimensions have been determined by considering the largest region over which no focus variation is visually appreciable and hence uniform. Since the defocus aberration law is spatially invariant, the training realized from this particular region is also applicable on the whole FOV. It is also the region where the other aberrations are generally the smallest, as beneficial for building reference images of the highest quality. As is appreciable, each cropped image contains a sufficiently large amount of red blood cells (typically ~ 100) to learn the properties of optimal red blood cell images (with or without parasites). After step 3, the data flowchart is divided in two branches: the left one is dedicated to the reconstruction of the images at varied z (step 4a-5a) with $\Delta z_{num} = 0$; the right one is dedicated to the production of the reference images (step

Biomedical Optics EXPRESS

4b-5b). Reconstructed images of the reference branch are here explored by varying finely Δz_{num} reconstruction parameter. The selection of reference images for intensity and phase modalities is finally determined by a parasitologist (step 5b).

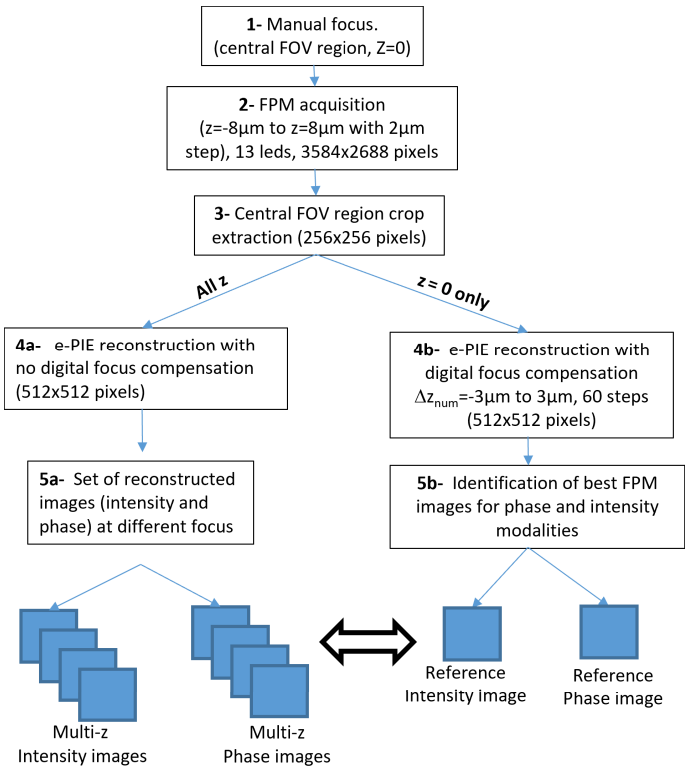

Fig. 4. U-Net training database construction. The samples are digitized (1-3) with FPM under different focus conditions (from $z = -8\mu m$ to $z = 8\mu m$). For all z, FPM images are reconstructed with no numerical focus compensation (step 4a and 5a) to produce the U-Net input examples. For each of these inputs, the reference I and Φ phase images are obtained after step 4b and 5b (the desired output image) and used to learn the Input-Output correspondence and transformation.

The flowchart of Fig. 4 has been conceived to limit the burden of the production of the training dataset using only coarse positioning of the microscope objective lens over a large range of z values. It is completed with virtual z-stacking calculations for various Δz_{num} for z=0. In test phase, the network is applied on complete FOV (I, Φ) images to produce corrected ones. We insist on the fact that even if only a restricted region of the FOV is used for the construction of the training dataset, the resulting trained neural network is general. In particular and as will be shown in the following section, the neural network is able to correct images over the complete microscope FOV and not only over its central zone.

3.3. Training database

In this study, anonymized blood-smear slides from two infected patients were used for the U-Net training. Although limited in terms of patient diversity, the database that has been realized is representative of parasites encountered in real situations in terms of size and age. For each patient 14 FOV central regions (of 512x512 pixels after reconstruction) were acquired. 20 FOV central regions were used for training, 4 for validation and 4 for test. Each reconstructed region

Biomedical Optics EXPRESS

has then been split into 25 small patches of 128x128 pixels. This choice has been made in order to constraint the processing time and to ensure a sufficient quantity of training data. Data were further augmented through random rotations and flipping to increase the size of the training dataset. The final size of the training database (including out-of-focus images and geometrical transformations and data augmentation) is 18000 patches. The validation and test datasets are each composed of 900 patches.

4. Results and evaluation

4.1. Visual evaluation

Some illustrative results obtained at the output of the trained U-Net are indicated in Fig. 5. The upper part corresponds to images reconstructed with $\Delta z_{num} = 0$ for z ranging from $-8\mu m$ to $8\mu m$ in a step of $2\mu m$. The reference images (upper-left hand-side) have been isolated from z-stacking reconstructions at $z = 0\mu m$. They have been obtained with $\Delta z_{num}= -0.6\ \mu m$ and $\Delta z_{num} = 0.8\mu m$ for intensity and phase images respectively. The same z-stacking approach is used to isolate the reference that would be obtained from the images recorded for $z = -8\mu m$. The resulting images are shown in the upper box on left hand-side of the figure. They have been obtained for $\Delta z_{num} = 4.8\mu m$ and $\Delta z_{num} = 6.5\mu m$. For these optimal values, focus compensation becomes imperfect. Refocusing limits with reconstruction algorithm is hence experimentally determined within z=$-6\ \mu m$ to z=$6\ \mu m$. For comparison, the outputs of the U-Net that has been applied to reconstructed images (with $\Delta z_{num} = 0\mu m$) for the various z are presented in the lower part of Fig. 5.

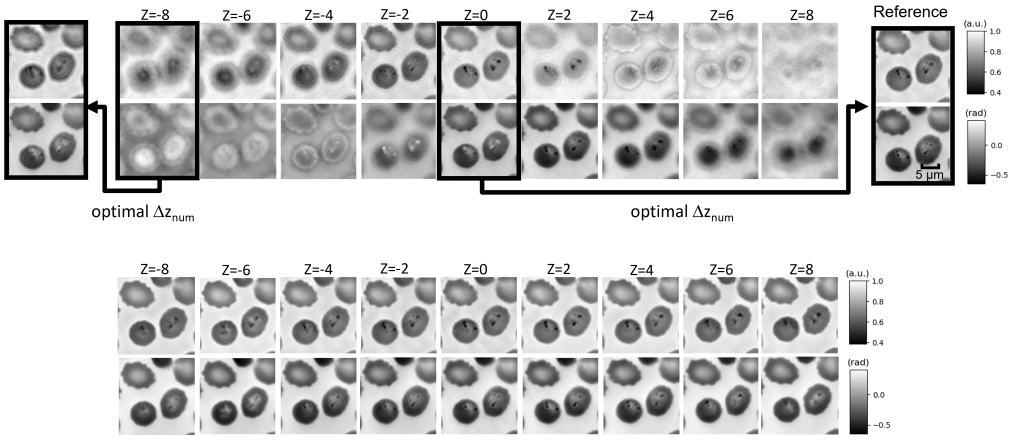

Fig. 5. (top) Illustrative examples of FPM images of blood smear recorded for different distances between the sample-microscope objective lens (from $z = -8\mu m$ to $z = 8\mu m$). References images (up-right box) are obtained following right branch of Fig. 3 using $\Delta z_{num} = -0.6\mu m$ and $\Delta z_{num} = 0.8\mu m$ for intensity and phase respectively; (bottom) corresponding U-Net-corrected images.

As can be observed, the produced images reveal a quality enhancement in terms of focus and contrast. They are adequately and automatically corrected with the U-Net regardless of z, a parameter that is unknown for the U-Net. It is interesting to note that all these images are very close to the reference even for $z = \pm 6\mu m$. Above $z = 6\mu m$ or below $z = -6\mu m$, some details on the parasites start to vanish (not presented). These first results permit to estimate the range over which the U-Net focus compensation is efficient. Indeed, this range is important ($\sim 12\mu m$) and comparable to the one accessible to e-PIE reconstruction with focus compensation.

Biomedical Optics EXPRESS

We now consider the complete microscope FOV. A representative result obtained with and without U-Net compensation is visible on Fig. 6. The reconstructed image exhibits some focus variations across the particularly large FOV (because of the sensor size). This comes from horizontal imperfections in the microscope slide-holder that varies from $\Delta z \sim -3\mu m$ (bottom-right) to $\sim 3\mu m$(top-left). The boxed region ***A*** corresponds to a region that is correctly focused, the boxed region ***B*** to a region that is out of focus ($\Delta z = 2.5\mu m$). The application of the trained U-Net on these two regions (see the 2 insets of Fig. 6) clearly confirms the expected behavior of the neural network: its ability to correct locally the reconstructed images. More precisely, the visibility of parasites is clearly enhanced across the entire FOV. This is particularly appreciable in region ***B*** ($\Delta z \sim 2.5\mu m$). Also, in region ***A*** ($\Delta z \sim 0\mu m$), the initial sharpness of the red blood cells, the contrast of targeted parasites sub-cellular compartments is enhanced in both intensity and phase images. One hemozoin crystal that was hardly visible in phase image becomes now clearly observable (pointed with red arrow).

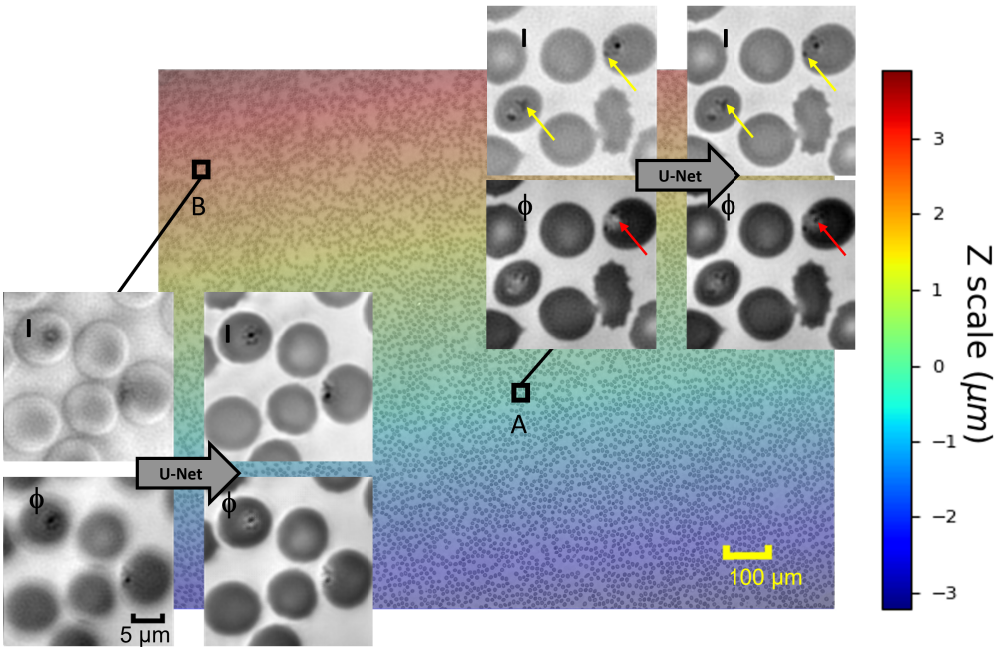

Fig. 6. Complete FOV of a blood smear sample captured by FPM. The superimposed colormap indicates the variation of sample-microscope objective distance separation caused by a tilt of the slide-holder. The two insets relative to areas A and B presents FPM images before and after U-Net correction. Region **A** is natively correctly focused ($z \sim 0$). Region **B** is initially out of focus ($z \sim 2.5\mu m$). The arrows indicate sub-cellular parasite compartments whose visibility is reinforced with the U-Net

We note that the U-Net extra processing time is of $\sim 5sec$ using an Nvidia RTX 3080 GPU. This value is to be compared to the reconstruction time of $\sim 7sec$ also implemented on the same GPU through TensorFlow framework [30].

4.2. Quantitative impact on parasite detection

The performance of an automated system dedicated to the detection of parasites is here considered to test quantitatively the impact the U-Net. Results obtained with and without the U-Net module

Biomedical Optics EXPRESS

are compared. More precisely, sensitivity (equivalently, the True Positive Rate TPR) and specificity (equivalently, the True Negative Rate TNR) are evaluated. The automaton is basic and integrates only the essential elements to its functioning, namely an FPM image reconstruction module, the U-Net image correction module and an image-based detection and classification module. Two experimental conditions are introduced to explore the limits of the U-Net module efficiency: for the first one, the z-axis position of the microscope objective is set optimally for the central region of the FOV. Due to inaccuracies in the horizontality of the microscope slide-holder, z varies between $-3\mu m$ and $3\mu m$ across the FOV ($z = 0$ in the center). For the second condition, the objective lens is moved by $2\mu m$ along z-axis. In this case, z varies between $-1\mu m$ and $5\mu m$ across the field of view.

By definition, $TPR = \frac{TP}{TP+FN}$ and $TNR = \frac{TN}{TN+FP}$ where TP represents the True Positive (parasites correctly detected), FN False Negative (parasites undetected), FP False Positive (healthy red blood cell detected as parasite), TN True Negative (healthy red blood cell correctly detected). TPR and TNR are two figures of merit of primary interest since used for the certification of diagnostic tests. The former qualifies the ability of the system to raise true alarms whereas the latter qualifies the ability of the system to not raise a false alarm. Ideally, the system has a sensitivity of 1 and a specificity of 1. In addition, we introduce 4 other metrics in order to produce more robust comparisons: the AUC (area under the ROC curve), Accuracy (percentage of overall (Infected and healthy RBCs) good classification, Precision (performance of the model at predicting the Infected class) and F1-score (harmonic mean of the model's precision and sensitivity) with F1-score= $\frac{TP}{TP+\frac{1}{2}(FP+FN)}$

The detection algorithm that is implemented in the automaton relies on a conventional YOLO [31] neural network evaluated with two training datasets (Refer to appendix 1 and 2 for further details). In order to benefit from the full information that is available on the sample, the intensity and phase images pair is used as input of the YOLO. The usual metric scores used to evaluate the detection performances are indicated in Table 1.

Table 1. *Plasmodium falciparum* detection performances on the test dataset, obtained on raw FPM reconstructed images with and without U-Net compensation module (for $z = 0\mu m$ and $z = 2\mu m$).

	with U-Net		without U-Net	
	$z = 0\mu m$	$z = 2\mu m$	$z = 0\mu m$	$z = 2\mu m$
Sensitivity	97.314	95.41	96.338	91.553
Specificity	99.08	99.07	99.281	98.055
AUC	0.989	0.982	0.989	0.967
Accuracy	98.998	98.9	99.144	97.753
Precision	83.739	83.326	86.725	69.625
F1-score	90.018	88.96	91.279	79.097

The sensitivity evaluated with and without U-Net for $z = 0$ and $z = 2\mu m$ are plotted Fig. 7(a)) as a function of a confidence threshold. The plain curves have been obtained with the application of the U-Net, the dashed curves without refocalisation. As could be expected, the best sensitivity curves are obtained when U-Net compensation module is implied in the workflow. Interestingly, the results obtained for $z = 2\mu m$ are very close to those obtained with $z = 0\mu m$. Comparatively and when no U-Net image correction is used, the sensitivity decreases even for $z = 0\mu m$. It is very degraded for $z = 2\mu m$. The complementary ROC (receiver operating characteristics) curve that represents the trade-off between sensitivity and specificity as a function of operating point is plotted in Fig. 7(b)). The optimal operating points are there indicated with circles.

When analyzing these results, we must take into account the fact that the 2 classes (healthy, infected) are largely disequilibrated (20 times healthier RBC, than infected ones; see Table 2).

Biomedical Optics EXPRESS

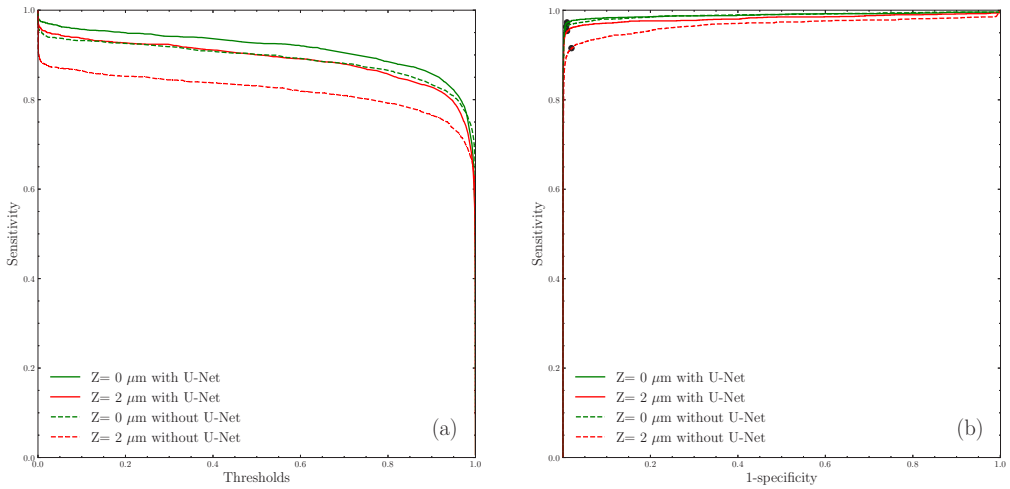

Fig. 7. a) Evolution of *Plasmodium falciparum* detection sensitivity as a function of the threshold parameter on the test database. The dashed curves correspond to results obtained on raw FPM reconstructed images without U-Net compensation module (for $z = 0\mu m$ and $z = 2\mu m$). Plain curves correspond to results after U-Net compensation of images. b) ROC curve revealing the best operating points (circles) defined by the best trade-off between sensitivity and specificity.

The metrics which focus on the recognition performance of the True Class (Infected), namely Sensibility, Precision and F1-score show an important decrease between Z=0 μm and Z=2 μm showing that parasites may be difficult to detect when defocus is present. This is particularly significant on precision and F1-score. Using a U-Net refocusing allows to increase the results at Z=2 in order to bring them to the same values as those at Z=0 μm. More specifically, the U-Net compensation leads to an enhancement of the sensitivity without degradation of the specificity. Considering the metrics which consider both classes without distinguishing them such as accuracy and AUC, they do not reveal significant differences between Z=0 μm and Z=2 μm, whether refocused or not). Considering the big number of healthy cells, we can deduce that healthy RBC can be detected even in the presence of some defocusing, contrary to what occur with parasites. These results demonstrate the positive impact of the U-Net. As a consequence, they attest that the tolerances regarding the precision of z positioning of the objective lens can be much relaxed. Recalling that $z = 2\mu m$ in the center of the FOV and the fact that slide holder is slightly tilted (see colormap of Fig. 6) the tolerance is roughly evaluated to $\sim \pm 5\mu m$.

Table 2. The number of healthy and infected red blood cells in training, validation and test sets

	Training	Validation	Test
Healthy RBC	212558	44333	41919
Infected RBC	9597	2023	2048

To analyze further these results, errors in the detection of red blood cells are indicated in Fig. 8. The complete test-dataset related to $z = 2\mu m$ is here considered. The crosses identify the positions of the undetected red blood cells in the FOVs. When no U-Net correction is implied (Fig. 8(a)), most of the errors are located near the left-up corner of the FOV, region where z is more important. The situation changes radically when images are corrected with the U-Net (Fig. 8(b)). Here errors are very few and almost homogeneously spatially distributed. Moreover,

Biomedical Optics EXPRESS

the black box in Fig. 8 indicates the region where the focus is natively correct. In this region and without any U-Net compensation, the density of the errors is small as could be expected. It is slightly further decreased after U-Net compensation. This suggests that the quality of images is indeed enhanced even though the focus was already correct.

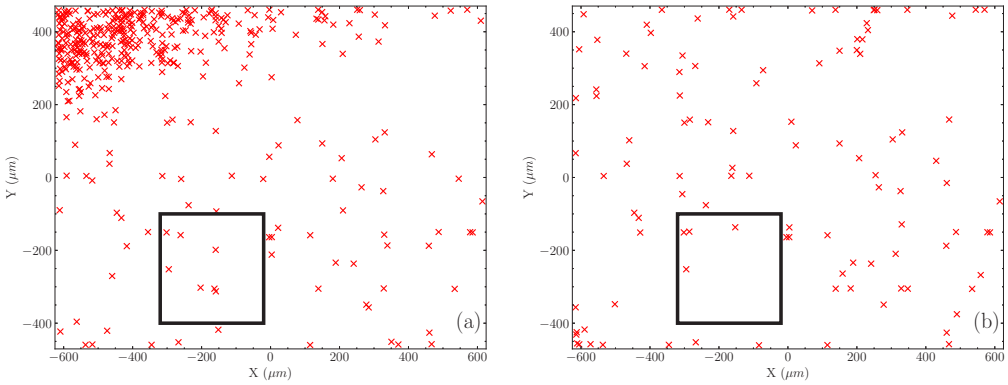

Fig. 8. Position of detection errors for $z = 2\mu m$ as evaluated by Yolo architecture. a) Without U-Net image correction (right). b) With U-Net-correction. The region delimited with the box corresponds to a region that is natively correctly focused.

Of course, the neural network classifier could benefit further optimizations (with a larger training dataset and an optimization of some of its internal parameters). Nevertheless, the presented results already demonstrate quantitatively the positive impact of the image correction brought by the U-Net treatment: it extends importantly the tolerances of the positioning of the microscope objective lens and permits also to extend the exploitable surface of the FOV which can be beneficial to reduce the sample acquisition with FPM apparatus over large surfaces.

5. Conclusion

At fixed resolution, the DOF of FPM is natively more important than the one obtained with a conventional microscope, rendering such microscopy attractive. However and when the sample thickness becomes comparable or higher than the DOF, undesirable variability in sub-cellular compartments visibility have been observed, particularly for phase images. To overcome such limitation, we have introduced a deep-learning approach aiming at enhancing targeted cellular compartment visibility, in a perspective of an automated diagnosis. The U-Net model is constructed to simultaneously reinforce targeted cell compartments visibility but also to compensate any focus imprecision. The image correction is done at sub-cellular scale. To this end, the neural network exploits jointly intensity and phase images (bi-modal U-Net) namely, the complete sample information accessible to FPM. During its training stage, the selection of optimal reference images is obtained with fine focus exploration. FPM reconstruction around optimal z setting, namely virtual z-stacking, is used. During the inference stage, images are corrected and the contrast of the targeted cell compartment is reinforced. To judge the impact of such approach and in contrast to what is usually done in the literature, a real-life problem is tested quantitatively (ROC curves) in complement of visual image quality metrics evaluation. Experimental results reveal that the sensitivity of detection is significantly improved without degradation of its specificity. Also the FPM depth of field is extended.

Although the training of the network is indeed based on focus compensation, the statistical framework exploited during the training stage renders the interpretation of the results difficult. Future experiments will be conducted to separate the contributions of focus correction, contrast enhancement and bi-modal complementary information.

Biomedical Optics EXPRESS

A. Appendix 1 : Brief description of the YOLO algorithm

Yolo (You Only Look Once) has been introduced in [31] in its original version. Since then, several improved versions have been proposed such as Yolo V2, V3, V4 and a slightly smaller version YOLO-tiny, which is specifically designed to achieve an incredibly high speed of 220fps. This paragraph is largely inspired by [31].

The aim of YOLO is to perform object detection and classification in an image in a single step of the algorithm. Compared to other algorithms of the same type, such as Fast-R-CNN [32] or [33], its rapidity allows real-time object detection. Moreover, it can be used when there are multiple objects in the image, at different locations and sizes. Yolo has become nowadays a standard in real-time object detection.

YOLO algorithm is based on regression, it predicts classes and bounding boxes for the whole image in **one run of the Algorithm**. The processing part of the algorithm relies on a Convolutional Neural Network. Yolo aims to predict a class of an object and the bounding box specifying the location of this object. Each bounding box can be described using four descriptors: center of the box (**bx**, **by**), width (**bw**), height (**bh**), and a value c corresponding to the class of an object. Apart from that, Yolo predicts a real number **bc**, which is the probability that there is an object in the bounding box (confidence).

The image is split into cells. Each cell is then responsible for predicting K bounding boxes and a confidence score for each box. An object is considered to lie in a specific cell only if the center coordinates of the anchor box lie in that cell. During the one pass of propagation, YOLO determines the probability that the cell contains an object of a certain class. The confidence score reflects how confident the model is that the box contains an object and also how the box accurately represents the object. It is calculated by the following formula (1) on all the boxes of the cell.

$$pc = Pr(Object)xIOU(pred, truth) \tag{6}$$

where IOU measures the Intersection over Union between the predicted box and the ground truth (varies from 0 to 1). Each cell also predicts conditional class probabilities for each object:

$$Pr(Class = i/object) \tag{7}$$

The retained probability Pr is the one that corresponds to the highest value among all the boxes in the cell. Following Bayes rule, the probability that the object be in class= i is given by:

$$Ci = Pr(class = i/object)xPr(object))xIOU(pred, truth) \tag{8}$$

The class of each box is attributed depending on a decision threshold applied on Ci. The present process relying on the class probabilities may lead to many anchor boxes as can be seen on Fig. 9. It is therefore necessary to suppress the un-necessary ones. To solve this problem Non-max suppression eliminates the bounding boxes that are very close, by performing the IoU (Intersection over Union) with the one having the highest class probability among them; the algorithm then rejects the bounding boxes whose value of IoU is greater than a threshold. This threshold is fixed empirically to 0.5 in the literature.

Finally, the algorithm outputs are the required vector indicating the parameters of the bounding box associated with the object of the respective class. The architecture of the network is composed of several layers of convolutions and its output consists of the parameters of the different boxes. This way the different boxes will be able, during the training phase, to fit at best the objects present in the images. The activation function in the first layers of convolutions are chosen as ReLu but in order to estimate probabilities (real numbers), they are chosen as sigmoids or

Biomedical Optics EXPRESS

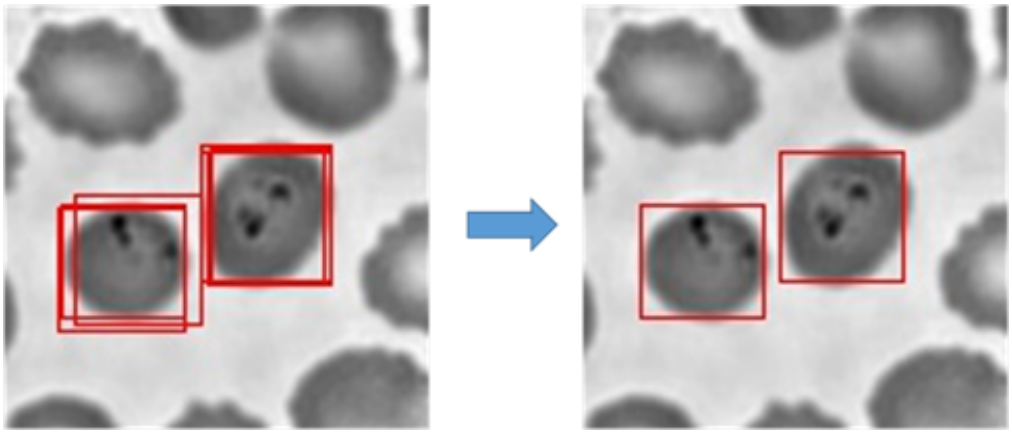

Fig. 9. (left) zoom on part of the image containing detected boxes which are too close to be all kept. (right) same image after boxes suppression

loglikelihoods in the last layers. The Loss function has the following value:

$$\begin{aligned} Loss = \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(x_i - \widehat{x}_{(i)})^2 + (y_i - \widehat{y}_{(i)})^2] \\ + \lambda_{coord} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} [(\sqrt{w_i} - \sqrt{\widehat{w}_{(i)}})^2 + (\sqrt{h_i} - \sqrt{\widehat{h}_{(i)}})^2] \\ + \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{obj} (C_i - \widehat{C}_i)^2 + \lambda_{NoObj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} 1_{ij}^{NoObj} (C_i - \widehat{C}_i)^2 \\ \sum_{i=0}^{S^2} 1_{ij}^{obj} \sum_{c \in classes} (P_i(c) - \widehat{P}_i(c))^2 \end{aligned} \tag{9}$$

where "obj" denotes if object appears in cell i and "1_{ij}^{obj}" denotes that the jth bounding box predictor in cell i is "responsible" for that prediction. One can note that the function loss only penalizes the classification error if an object is present in that grid cell. It also only penalizes the bounding box coordinates error if that predictor is "responsible" for the ground truth box (i.e. has the highest IOU of any predictor in that grid cell).

In order to jointly process the intensity and phase FPM images, the number of YOLO input channels is configured to two. The first channel is fed with an intensity image, the second with a phase image.

B. Appendix 2 : YOLO training datasets and parameters

To optimize the YOLO model's hyperparameters, we experimented different learning strategies, such as varying the optimizer, adjusting the learning rate, threshold IOU, and training the model with different strategies. For YOLO detection and classification, the best results were obtained for a batch size of 48 images and a threshold of 0.5 (as suggested in [34]), with or without U-Net enhancement. We used the Adam optimizer with $beta_1 = 0.9$ and $beta_2 = 0.999$. The training process was carried out in two stages, similar to [31]. Our training process involved training the model for 50 epochs with a learning rate of 10^{-3} in the first stage, as the weights were initialized randomly. We then continued the learning process using a lower learning rate of 10^{-5},in the second stage. We observed that the initial stage helped to achieve an accurate detection of the red blood cells quickly, while the second stage improved the model's classification performance.

Biomedical Optics EXPRESS

Two YOLO models are realized using the above described hyper-parameters optimization approach. They share a common architecture but differ in the dataset (denoted DS_1 and DS_2) used to train them. The first model is trained from DS_1 while the second is trained from DS_2 as detailed below. To constitute DS_1 and DS_2, 28 fields of view were captured on blood smears belonging to the same 2 patients as those already used in section 3.2 (14 FOV per patient). The z of the microscope objective lens is optimized from visual observation of the red blood cells sharpness in the central region of the FOV. After reconstruction, each FOV contains images of 7168x5376 pixels. Each of these large images was then split into 48 patches of size 896x896 pixels. Thus 1344 patches are obtained from all the captured fields of view. Dataset DS_1 is made of all these patches. Dataset DS_2 contains the same images whose quality has been further enhanced after their processing with the U-Net introduced in section 3. Data augmentation consisted of 3 rotations ($\pi/2$, π, $3\pi/2$) and led to a total amount of 5376 patches for DS_1 and 5376 patches for DS_2. All patches have been labeled by an expert. The label consists in the delimitation of the boxes enclosing each red blood cells. A class indicating the presence or not of a parasite is also defined for each box. For each dataset, 3840 patches were used for the training, 768 for the test and 768 for the validation. The distribution of healthy and infected red blood cells contained over each split is detailed in Table 2.

Disclosures. The authors declare that there are no conflicts of interest related to this article.

Data availability. Data underlying the results presented in this paper are not publicly available at this time but may be obtained from the authors upon reasonable request.

References

1. OT4D-ANR project, (French national research agency), dedicated to the development of innovative optical microscopy and artificial intelligence algorithms for medical diagnosis (2021-2025).
2. Z. Bian, C. Guo, S. Jiang, J. Zhu, R. Wang, P. Song, Z. Zhang, K. Hoshino, and G. Zheng, "Autofocusing technologies for whole slide imaging and automated microscopy," J. Biophotonics **13**(12), 1 (2020).
3. Q. Li, X. Liu, J. Jiang, C. Guo, X. Ji, and X. Wu, "Rapid whole slide imaging via dual-shot deep autofocusing," IEEE Trans. Comput. Imaging **7**, 124–136 (2021).
4. J. Park, D. J. Brady, G. Zheng, L. Tian, and L. Gao, "Review of bio-optical imaging systems with a high space-bandwidth product," Adv. Photonics **3**(04), 1 (2021).
5. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**(9), 739–745 (2013).
6. J. Li, Q. Chen, J. Zhang, Y. Zhang, L. Lu, and C. Zuo, "Efficient quantitative phase microscopy using programmable annular led illumination," Biomed. Opt. Express **8**(10), 4687–4705 (2017).
7. J. W. Goodman and P. Sutton, "Introduction to Fourier optics," Quantum and Semiclassical Optics-Journal of the European Optical Society Part B **8**, 1095 (1996).
8. L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for Fourier ptychography with an led array microscope," Biomed. Opt. Express **5**(7), 2376–2389 (2014).
9. X. Ou, G. Zheng, and C. Yang, "Embedded pupil function recovery for Fourier ptychographic microscopy," Opt. Express **22**(5), 4960–4972 (2014).
10. L. Tian and L. Waller, "3d intensity and phase imaging from light field measurements in an led array microscope," Optica **2**(2), 104–111 (2015).
11. R. Horstmeyer, J. Chung, X. Ou, G. Zheng, and C. Yang, "Diffraction tomography with Fourier ptychography," Optica **3**(8), 827–835 (2016).
12. W. Pierré, L. Hervé, C. Paviolo, O. Mandula, V. Remondiere, S. Morales, S. Grudinin, P. F. Ray, M. Dhellemmes, C. Arnoult, and C. Allier, "3d time-lapse imaging of a mouse embryo using intensity diffractiontomography embedded inside a deep learning framework," Appl. Opt. **61**(12), 3337–3348 (2022).
13. M. Liang, C. Bernadt, S. B. J. Wong, C. Choi, R. Cote, and C. Yang, "All-in-focus fine needle aspiration biopsy imaging based on Fourier ptychographic microscopy," J. Pathol. Informatics **13**, 100119 (2022).
14. S. Zhang, G. Zhou, C. Zheng, T. Li, Y. Hu, and Q. Hao, "Fast digital refocusing and depth of field extended Fourier ptychography microscopy," Biomed. Opt. Express **12**(9), 5544–5558 (2021).
15. Y. Rivenson, Y. Zhang, H. Günaydın, D. Teng, and A. Ozcan, "Phase recovery and holographic image reconstruction using deep learning in neural networks," Light: Sci. Appl. **7**(2), 17141 (2017).
16. C. B. Delahunt, M. S. Jaiswal, and M. P. Horning, *et al.*, "Fully-automated patient-level malaria assessment on field-prepared thin blood film microscopy images," in *2019 IEEE Global Humanitarian Technology Conference (GHTC)*, (2019), pp. 1–8.
17. A. Köhler, "New method of illumination for photomicrographical purposes," J. R. Microsc. Soc. **14**, 261–262 (1894).
18. G. W. Gill, "Köhler illumination," in *Cytopreparation*, (Springer, 2013), pp. 309–323.

Biomedical Optics EXPRESS

19. A. Maiden, D. Johnson, and P. Li, "Further improvements to the ptychographical iterative engine," Optica **4**(7), 736–745 (2017).
20. J. R. Fienup, "Phase retrieval algorithms: a comparison," Appl. Opt. **21**(15), 2758–2769 (1982).
21. M. Guizar-Sicairos and J. R. Fienup, "Phase retrieval with transverse translation diversity: a nonlinear optimization approach," Opt. Express **16**(10), 7264–7278 (2008).
22. C. Mehanian, M. Jaiswal, and C. Delahunt, *et al.*, "Computer-automated malaria diagnosis and quantitation using convolutional neural networks," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*, (2017).
23. S. Cho, S. Kim, Y. Kim, and Y. Park, "Optical imaging techniques for the study of malaria," Trends Biotechnol. **30**(2), 71–79 (2012).
24. A. Pan, Y. Zhang, K. Wen, M. Zhou, J. Min, M. Lei, and B. Yao, "Subwavelength resolution Fourier ptychography with hemispherical digital condensers," Opt. Express **26**(18), 23119–23131 (2018).
25. O. Bunk, M. Dierolf, S. Kynde, I. Johnson, O. Marti, and F. Pfeiffer, "Influence of the overlap parameter on the convergence of the ptychographical iterative engine," Ultramicroscopy **108**(5), 481–487 (2008).
26. J. Sun, C. Zuo, J. Zhang, Y. Fan, and Q. Chen, "High-speed Fourier ptychographic microscopy based on programmable annular illuminations," Sci. Rep. **8**(1), 7669 (2018).
27. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT press, 2016).
28. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds. (Springer International Publishing, 2015), pp. 234–241.
29. A. Lucas, M. Iliadis, R. Molina, and A. K. Katsaggelos, "Using deep neural networks for inverse problems in imaging: Beyond analytical methods," IEEE Signal Process. Mag. **35**(1), 20–36 (2018).
30. S. Jiang, K. Guo, J. Liao, and G. Zheng, "Solving Fourier ptychographic imaging problems via neural network modeling and tensorflow," Biomed. Opt. Express **9**(7), 3306–3319 (2018).
31. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2016), pp. 779–788.
32. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, vol. 28 C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, eds. (Curran Associates, Inc., 2015).
33. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, eds. (Springer International Publishing, Cham, 2016), pp. 21–37.
34. J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," arXiv arXiv:1804.02767 (2018).