DEEPFOCUS: A FEW-SHOT MICROSCOPE SLIDE AUTO-FOCUS USING A SAMPLE INVARIANT CNN-BASED SHARPNESS FUNCTION

Adrian Shajkofci[1,2], *Michael Liebling*[1,3]

[1]Idiap Research Institute, CH-1920 Martigny, Switzerland
[2]École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland
[3]Electrical & Computer Engineering, University of California, Santa Barbara, CA 93106, USA

ABSTRACT

Autofocus (AF) methods are extensively used in biomicroscopy, for example to acquire timelapses, where the imaged objects tend to drift out of focus. AF algorithms determine an optimal distance by which to move the sample back into the focal plane. Current hardware-based methods require modifying the microscope and image-based algorithms either rely on many images to converge to the sharpest position or need training data and models specific to each instrument and imaging configuration. Here we propose DeepFocus, an AF method we implemented as a Micro-Manager plugin, and characterize its Convolutional Neural Network (CNN)-based sharpness function, which we observed to be depth co-variant and sample-invariant. Sample invariance allows our AF algorithm to converge to an optimal axial position within as few as three iterations using a model trained once for use with a wide range of optical microscopes and a single instrument-dependent calibration stack acquisition of a flat (but arbitrary) textured object. From experiments carried out both on synthetic and experimental data, we observed an average precision, given 3 measured images, of $0.30 \pm 0.16\,\mu\text{m}$ with a $10\times$, NA 0.3 objective. We foresee that this performance and low image number will help limit photodamage during acquisitions with light-sensitive samples.

Index Terms— Microscopy, autofocus, PSF estimation, convolutional neural networks, Micro-Manager

1. INTRODUCTION

Modern microscopy techniques rely on many components that are remotely controllable. This allows implementing control loops that limit the need for human-supervised operation. Auto-focusing systems, in particular, are used extensively in the acquisition of timelapses in developmental or cellular biology or to automatically image slides in a slide scanner. In the former application, imaged specimens tend to drift from the focal plane over time because of specimen growth, flow of the medium, or motion caused by temperature changes. In the latter case, variability in the mounting of the slides requires per-slide adjustment.

AF systems seek to determine the optimal shift by which to adjust the axial position to maximize image sharpness. AF solutions can be hardware-based (e.g. laser-based sensing of the sample drift [1] or phase detection by an auxiliary sensor [2]) or image-based, which does not require any modification of the optical path of the microscope as a focus score is retrieved from the image itself [3].

We can classify image-based AF algorithms into two categories. The first comprises AF methods that use iterative minimization of a one-dimensional objective function, the focus score, to move the object to the point at which it is sharpest. Because the output of the function is not predictable and depends on the sample, the AF has to acquire tens to hundreds of images at different axial positions in order to converge to a non-local optimum [3]. A high number of image acquisitions can be damaging for the sample, especially in fluorescence microscopy [4].

Additionally, existing objective functions only give a meaningful result in the neighborhood of the focal plane, and lose information (i.e. the gradient of the curve is zero) farther away from the focal plane. Furthermore, depending on the software implementation and the imaging modality, the acquisition of hundreds of images can take up to several minutes. The second category comprises single shot AF techniques (that need only one or a few images). Thanks to end-to-end CNNs, they take an image as input and directly deduce the optimal shift to be in focus ([5]–[7]). The drawback of these direct methods is that a long and computationally-intensive CNN training with a microscope objective-specific training data set, must be repeated whenever the optical system changes. Furthermore, these methods are not directly available in open microscope control software, such as μManager [8].

In this paper, we propose a local, CNN-based focus scoring function that remains nearly invariant when imaging different types of samples or modalities on any given microscope. We developed a correlation-based AF algorithm that takes advantage of the broad shape and unimodal minimum of this function, which helps to speed up convergence and remaining effective even when the imaged object is far from the focal (several times the depth-of-field (DOF), see Fig. 1). Since our
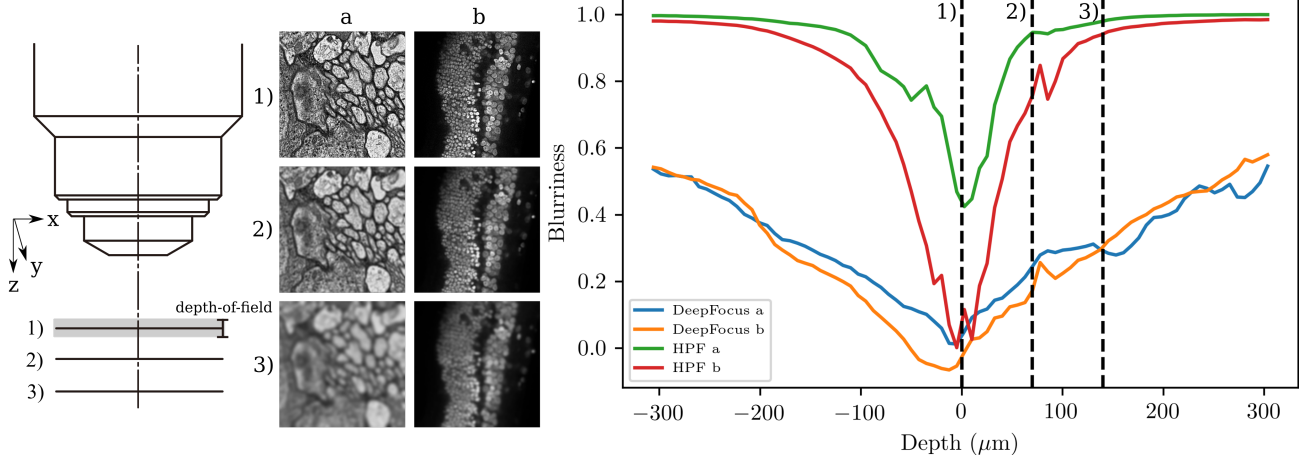
Fig. 1: The object may be outside of the DOF and appear blurry. Here we quantified the blur $b(z)$ using DeepFocus and an high-pass filter (HPF) for two different images. Using HPF, $b(z)$ changes shape and slope when different objects are presented under the microscope, and there is a lack of depth information for $|z| > 150\,\mu\text{m}$. Using DeepFocus, the slopes for both images are similar in shape and retain information about depth in the whole $[-300; 300]\,\mu\text{m}$ region.

CNN method does not require a microscope-specific data set for training besides a single stack of an arbitrary object, it is *plug-and-play*.

This paper is organized as follows. In Section 2, we present the blurriness scoring function, the calibration process, and the AF algorithm. In Section 3, we experimentally verify the scoring function's assumed invariance to a variety of samples and characterize performance with respect to the number of images and in comparison to common AF scoring functions, using both simulated and experimentally acquired data. We discuss our findings and conclude in Section 4.

2. METHODS

2.1. Problem statement

We consider a specimen, modeled as 2D manifold in 3D space (such as a thin microscopy slide), that we wish to image with a widefield microscope in bright field, fluorescence, or phase contrast. The entire specimen or some regions in the field-of-view (FOV) can be out of focus and outside of the DOF (see Fig. 1). We assume the microscope has a motorized stage for adjusting the focus. We aim at finding the optimal axial shift Δz by which to adjust the sample position such that it is in focus. We seek a solution that (i) does not require a manually selected reference image to be matched (such that the method can be used both for maintaining focus in live timelapses but also for imaging collections of fixed samples) (ii) requires a minimal number of images (to limit photodamage) (iii) shall not require imaging calibration specimens (PSF measurement beads, etc.) or large-scale, microscope-specific training.

2.2. Method description

The principle behind our proposed algorithm is to measure a blurriness score $b(z_i)$ for a few (M) images acquired at

different focus positions z_i, $i = 1, \ldots, M$, resulting in a set of pairs $\{(z_i, b(z_i)) | i = 1, \ldots, M\}$ and to determine the necessary focal shift Δz such that $\{(z_i - \Delta z, b(z_i)) | i = 1, \ldots, M\}$ matches a microscope objective-specific, sample-invariant, depth-blurriness response curve $b_{\text{calib}}(z)$ using cross-correlation. The curve invariance assumption has been similarly used by the model-based curve fitting approach of [9].

For this approach to work, we need a focus estimation function that is invariant to the sample shape or texture (sample-invariance) but co-variant with the sample's axial position and sufficiently informative beyond the immediate vicinity of the focal plane. To this end, we chose an estimator of the local optical properties of the microscope objective [10]. Briefly, it relies on a trained CNN to regress the parameters of a Zernike polynomial Point Spread Function (PSF) model [11], given a blurry image patch as an input. Here, we use the estimated Zernike coefficient corresponding to focus as a blurriness score, which provides, given an image as input, a local blurriness score $b[x, y, z]$ for the indicated position depth z.

The trained CNN [10] does not require re-training when used on different microscopes or different microscope objectives and produces a curve whose shape (up to an axial scaling) is invariant to the sample (an aspect that we verify experimentally in Section 3.1). In order to determine the axial scaling, which is instrument-dependent, we require a calibration step consisting in the acquisition of a full stack of an arbitrary planar and textured object. This yields a blurriness map $b_{\text{calib}}(z)$ that we center with its minimum at the origin.

We now describe our proposed AF, which follows the structure illustrated in Fig. 2 and is summarized in the steps:

1. Fit b_{calib} to a Moffat distribution [12] and extract its full width at half maximum (FWHM). Set $M = 3$, let z_1 be the initial focal plane position, and initialize a Golden Section Search (GSS) algorithm with the interval $[z_2, z_3] = [z_1 - 2\,\text{FWHM}, z_1 + 2\,\text{FWHM}]$. Acquire images at $z = z_1, z_2, z_3$ and compute, using the CNN, the blurriness scores
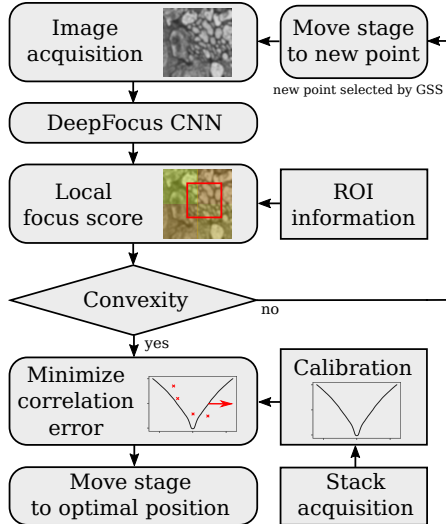
Fig. 2: Flowchart of the AF algorithm (see Section 2.2).

$b[z_1], b[z_2], b[z_3]$.

2. Check the convexity of $b[z_i]$, $i = 1, \ldots, M$ by fitting $b[z_i]$ to a quadratic polynomial. If the R^2 of the polynomial fit is higher than the R^2 of a linear fit, go to Step 6. Otherwise go to Step 3.

3. Increment $M \mathrel{+}= 1$. Update the GSS triplet to obtain and move to a new axial position z_M.

4. Acquire an image at the current axial position z_M.

5. Compute, using the CNN, the blurriness score $b[z_M]$ and go to Step 2.

6. Compute using cross-correlation the local optimal shift $\Delta z(x, y)$ minimizing the squared distance:

$$\Delta z(x, y) = \arg\min_{\Delta z} \sum_{i=1}^{M} \left(b[x, y, z_i] - b_{\text{calib}}(z_i - \Delta z) \right)^2.$$

7. Move the sample by Δz, averaged for the Region of Interest (ROI) in the (x, y) plane.

Table 1: Using the same experimental conditions as Fig. 3, we quantified the scoring function performance in terms of Standard Deviation (SD) in a $100 \, \mu\text{m}$ range (lower is better) and conditional entropy between the focus score and the distance (lower is better) in the whole $240 \, \mu\text{m}$ range. Using experimental acquisitions, DeepFocus outdo all other tested functions in terms of SD. Additionally, our method has for both modalities the lower conditional entropy and thus is more informative.

Focus score function	σ_{synth}	H_{synth}	σ_{exp}	H_{exp}
DeepFocus	0.03	**1.59**	**0.03**	**5.17**
HPF	0.06	6.38	0.08	44.31
Tenengrad [13]	0.10	4.17	0.06	21.77
LAPV	0.03	14.29	0.04	38.28
EWC [14]	**0.01**	3.47	0.16	8.21
SML	0.04	7.74	0.06	16.89
WS [15]	0.07	5.57	0.19	11.30

3. EXPERIMENTS

3.1. Characterization of regression invariance to image diversity

Since our AF algorithm relies on the invariance of $b_{\text{calib}}(z)$ to the type of imaged sample, we investigated whether our proposed CNN indeed satisfied this condition and whether other (existing) focus metrics could be substituted.

We gathered $N_{\text{synth}} = 1000$ images from the evaluation dataset of [10] and blurred them with Gaussian PSFs mimicking a $10\times$, NA 0.3 objective for $M = 132$ points in the depth range $-120 \, \mu\text{m} \leq z \leq 120 \, \mu\text{m}$. In addition, we acquired $N_{\text{exp}} = 1000$ stacks of fixed rat brain slices tagged with three fluorescent stains using a widefield transmission light microscope with a $10\times$, NA 0.3 objective in a depth range of $-60 \, \mu\text{m} \leq z \leq 60 \, \mu\text{m}$. We then computed $b_{\text{calib}}(z)$ using DeepFocus and other methods, including HPF, LAPV, SML [16], Tenengrad [13], EWC [14], and WS [15], which cover a broad range of focus measures, as reviewed in [3], [17]–[19].

In Table 1, we reported the average SD of $b_{\text{calib}}(z)$ over all input images. Using the experimental dataset, our method had an average SD of $\sigma = 0.03$ (normalization scale with 1 and 0 the blurriest and sharpest values, respectively). We noticed, as illustrated in Fig. 3 (a) and (b), that DeepFocus' SD increased when $|z|$ increases (i.e when the acquired pictures contain a medium-to-high blur). A low SD implies that $b_{\text{calib}}(z)$ is similar with different types of imaged specimens. Other methods had a SD of $0.04 < \sigma < 0.19$, and hence confirmed the variance of these focus metrics with image diversity.

3.2. Characterization of information measure of the scoring function

We next investigated how robustly our proposed DeepFocus measure can report (de)focus information as the distance from focus is increased up to 10 times the DOF. We observed (Fig. 3) that focus metrics other than ours were unable to give any information about z from $b_{\text{calib}}(z)$ whenever $|z|$ is higher than $60 \, \mu\text{m}$, as they reach a value that does no longer vary as the position is increased further. Since the gradient in such plateau regions is small, minimization algorithms could not converge quickly. To quantify these visual observations regarding the uncertainty of recovering z from any given $b_{\text{calib}}(z)$, we computed the conditional entropy:

$$H(\mathbf{B}|\mathbf{Z}) = - \sum_{b_{\text{calib}} \in \mathcal{B}, z \in \mathcal{Z}} p(b_{\text{calib}}|z) \log \left(\frac{p(b_{\text{calib}}|z)}{\sum_{z_i} p(b_{\text{calib}}|z_i)} \right),$$

where $\mathbf{B}$ and $\mathbf{Z}$ are random variables representing the calibration blurriness score and the axial distances, $\mathcal{B}$ and $\mathcal{Z}$ their support sets, and $p(b_{\text{calib}}|z)$ the probability of a score b_{calib}, given the distance z. A high conditional entropy value implies a high uncertainty of detecting the right z position for a given b_{calib}. The results, compiled in Table 1, reveal that DeepFocus had a conditional entropy of $H_{\text{synth}} = 1.59$, a value smaller than that obtained when using any of the other
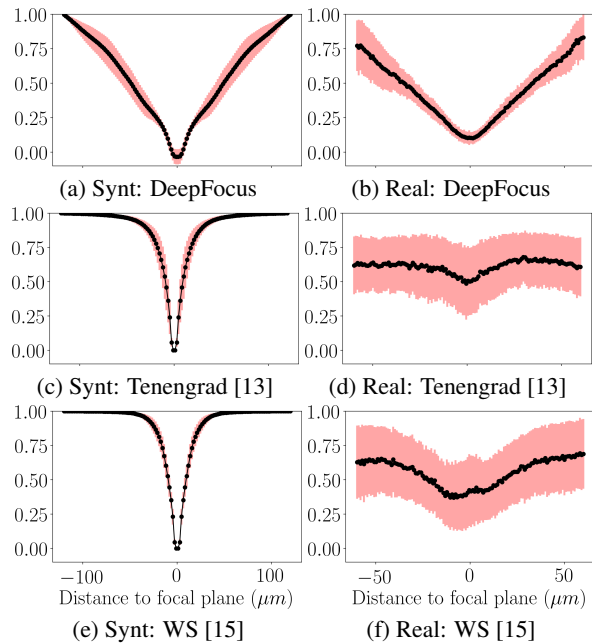
(a) Synt: DeepFocus (b) Real: DeepFocus

(c) Synt: Tenengrad [13] (d) Real: Tenengrad [13]

(e) Synt: WS [15] (f) Real: WS [15]

Fig. 3: Comparison of the output of different sharpness scoring functions as a function of z, centered at the origin. We used as input synthetically blurred images (left) and stacks of fluorescent rat brain tissue with a $10\times$, NA 0.3 objective (right). Using DeepFocus, the SD of $b_{\text{calib}}(z)$ around the focal plane is lower than with the other scoring functions. Additionally, scoring functions other than DeepFocus do not infer depth information ($b_{\text{calib}}(z) = 1$ for all z) when $|z| > 50\mu$m.

scoring functions instead. In the case of experimental acquisitions, we observed again an improvement in terms of entropy ($H_{\text{exp}} = 5.17$), where other methods have values in the range $8.21 < H_{\text{exp}} < 44.31$. We further determined the threshold distance after which no distance information can be inferred from the image, i.e when the image is too blurry to make the AF converge. DeepFocus retained depth information for a range of 120 μm with a $10\times$, NA $= 0.3$ objective, which is equivalent, using the diffraction-limited DOF formula, to 11 times the DOF ($10.7\,\mu$m). In comparison, metrics like WS and SML achieved ranges of only 4 and 7 times the DOF, respectively.

3.3. Characterization of the AF error as a function of the number of acquisitions

We finally investigated how accurately DeepFocus could retrieve the focal distance as a function of the number of images acquired. We used 100 blurred images from the generated dataset in Section 3.1 with a known in-focus position and computed its distance to the output position of the AF. We also compared our method to other autofocus scoring functions (for which we used a bounded Brent's method as optimizer). The results are summarized in Fig. 4.

We observed that our proposed AF converged rapidly (3 iterations), while the two other focus functions needed more
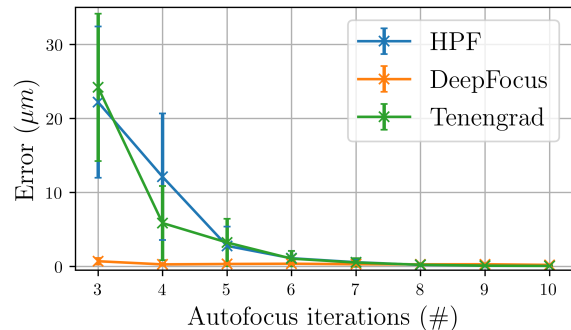

Fig. 4: Comparison of the AF error using 3 different AF scoring functions for 100 samples. We quantified the distance between the theoretical focus plane position and the AF output as a function of the number of AF iterations which represent additional input images. DeepFocus yields an error of $0.27 \pm 0.18\,\mu$m with 4 iterations. With 8 iterations or more, the others methods are on par or more accurate than ours.

than twice as many images to reach a similar focus accuracy. Using 8 iterations or more, we did not notice a better accuracy with our method compared to Tenengrad or HPF.

4. DISCUSSION AND CONCLUSION

In our experiments, we showed that the variance of b_{calib} over multiple images was usually lower using DeepFocus than when using other focus scoring functions, especially near the focal plane. Our explanation would be that the CNNs, already known to be translation-invariant [20], have been trained specifically for the recognition of the PSF parameters without discrimination on the input image type and position. By contrast, crafted features such as HPFs are computed from content-based calculations and differ from one image to another. When the image is acquired at a large distance from the focal plane, we noticed a loss of spatial features in the acquired image, due to the large FWHM of the PSF that degraded it. However, we have been able to retrieve depth information from the image up to 2.5 times farther away from the focal plane than with other methods. That could be mostly explained by the fact that DeepFocus computes features from a 128×128 px window, while Gradient-based methods use a much smaller window, such as 3×3 or 5×5.

In summary, we developed an AF method based on a combination of an CNN scoring function and optimization algorithms that are relying on the invariance of the scoring function. We showed that DeepFocus was robust to changes amongst samples, which enables the retrieval of the optimal axial shift using a correlation-based optimization process that needs as few as 3 images to converge. Our method is currently limited to imaging thin samples and further work will investigate the procedure for thicker objects. We implemented the calibration step and AF algorithm as two plugins (Java with a PyTorch [21] backend) for the μManager microscopy acquisition engine [8], which we will make available upon acceptance.

5. REFERENCES

[1] Y. Liron, Y. Paran, N. G. Zatorsky, B. Geiger, and Z. Kam, "Laser autofocusing system for high-resolution cell biological imaging," *J Microsc*, vol. 221, no. 2, pp. 145–151, 2006.

[2] L. Silvestri, M. C. Müllenbroich, I. Costantini, A. P. Di Giovanna, L. Sacconi, and F. S. Pavone, "RAPID: Real-time image-based autofocus for all wide-field optical microscopy systems," Bioengineering, BioRxiv, 2017.

[3] Y. Sun, S. Duthaler, and B. J. Nelson, "Autofocusing in computer microscopy: Selecting the optimal focus algorithm," *Microscopy Research and Technique*, vol. 65, no. 3, pp. 139–149, 2004.

[4] V. Magidson and A. Khodjakov, "Circumventing photodamage in live-cell microscopy," *Methods Cell Biol*, vol. 114, pp. 545–560, 2013.

[5] L. Wei and E. Roberts, "Neural network control of focal position during time-lapse microscopy of cells," *Sci Rep*, vol. 8, no. 1, p. 7313, 2018.

[6] S. Jiang, J. Liao, Z. Bian, K. Guo, Y. Zhang, and G. Zheng, "Transform- and multi-domain deep learning for single-frame rapid autofocusing in whole slide imaging," *Biomed. Opt. Express*, vol. 9, no. 4, p. 1601, 2018.

[7] H. Pinkard, Z. Phillips, A. Babakhani, D. A. Fletcher, and L. Waller, "Deep learning for single-shot autofocus microscopy," *Optica*, vol. 6, no. 6, p. 794, 2019.

[8] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, "Computer control of microscopes using μManager," in *Current Protocols in Molecular Biology*, vol. 92, Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010, pp. 1–17.

[9] S. Yazdanfar, K. B. Kenny, K. Tasimi, A. D. Corwin, E. L. Dixon, and R. J. Filkins, "Simple and robust image-based autofocusing for digital microscopy," *Opt. Express*, vol. 16, no. 12, p. 8670, 2008.

[10] A. Shajkofci and M. Liebling, "Semi-blind spatially-variant deconvolution in optical microscopy with local point spread function estimation by use of convolutional neural networks," *IEEE ICIP*, pp. 3818–3822, 2018.

[11] F. von Zernike, "Beugungstheorie des schneidenverfahrens und seiner verbesserten form, der phasenkontrastmethode," *Physica*, vol. 1, no. 7, pp. 689–704, 1934.

[12] A. F. J. Moffat, "A Theoretical Investigation of Focal Stellar Images in the Photographic Emulsion and Application to Photographic Photometry," *Astronomy and Astrophysics*, vol. 3, p. 455, 1969.

[13] D. Sun, S. Roth, and M. J. Black, "Secrets of optical flow estimation and their principles," in *IEEE COMPUTER SOCIETY CONFERENCE on Computer Vision and Pattern Recognition Proceedings*, 2010, pp. 2432–2439.

[14] Hanghang Tong, Mingjing Li, Hongjiang Zhang, and Changshui Zhang, "Blur detection for digital images using wavelet transform," *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME)*, pp. 17–20, 2004.

[15] M. Liebling and M. Unser, "Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion," *J. Opt. Soc. Am. A*, vol. 21, no. 12, pp. 2424–2430, 2004.

[16] S. K. Nayar and Y. Nakagawa, "Shape from focus: An effective approach for rough surfaces," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1990, 218–225 vol.2.

[17] J. H. Price and D. A. Gough, "Comparison of phase-contrast and fluorescence digital autofocus for scanning microscopy," *Cytometry*, vol. 16, no. 4, pp. 283–297, 1994.

[18] J. M. Mateos-Pérez, R. Redondo, R. Nava, J. C. Valdiviezo, G. Cristóbal, B. Escalante-Ramírez, M. J. Ruiz-Serrano, J. Pascau, and M. Desco, "Comparative evaluation of autofocus algorithms for a real-time system for automatic detection of Mycobacterium tuberculosis," *Cytometry*, vol. 81A, no. 3, pp. 213–221, 2012.

[19] U. Ali and M. T. Mahmood, "Analysis of blur measure operators for single image blur segmentation," *Applied Sciences*, vol. 8, no. 5, p. 807, 2018.

[20] Y. LeCun, "Learning invariant feature hierarchies," in *Computer Vision – ECCV 2012. Workshops and Demonstrations*, A. Fusiello, V. Murino, and R. Cucchiara, Eds., 2012, pp. 496–505.

[21] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in PyTorch," in *31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017.