Deep convolutional neural networks and digital holographic microscopy for in-focus depth estimation of microscopic objects

Tomi Pitkäaho,[1] Aki Manninen,[2] and Thomas J. Naughton[1]

[1] *Department of Computer Science, Maynooth University–National University of Ireland Maynooth, Maynooth, County Kildare, Ireland*

[2] *Biocenter Oulu, Oulu Center for Cell-Matrix Research, Faculty of Biochemistry and Molecular Medicine, University of Oulu, Finland*

Abstract

Deep artificial neural network learning is an emerging tool in image analysis. We demonstrate its potential in the field of digital holographic microscopy by addressing the challenging problem of determining the in-focus reconstruction depth of an arbitrary epithelial cell cluster encoded in a digital hologram. A deep convolutional neural network learns the in-focus depths from half a million hologram amplitude images. The trained network correctly determines the in-focus depth of new holograms with high probability, without performing numerical propagation. To our knowledge, this is the first application of deep learning in the field of digital holographic microscopy.

Keywords: Imaging, Digital Holographic Microscopy, Autofocusing, Deep Learning, Machine Learning

1 Introduction

Deep learning [LeCun et al., 2015] is a technique for solving hitherto open problems in image analysis and other fields that is starting to have an impact in the field of biomedical optics, for example OCT [Prentaši´c et al., 2016, Abdolmanafi et al., 2017, Karri et al., 2017] and other forms of microscopy [Cire¸san et al., 2012, Wang et al., 2014, Rezaeilouyeh et al., 2016, Gopakumar et al., 2017]. In the context of this work, deep learning is defined to be a deep convolutional neural network with multiple hidden layers that perform convolution operations on its multi-dimensional input. Image-based applications of deep learning [Russakovsky et al., 2015] are characterised by neural networks with at least eight hidden layers, at least tens of thousands of images, at least hundreds of images per class, at least millions of learned parameters, and training times of at least weeks if run on a single-processor personal computer. This type of network has been used successfully in various different visual object recognition and object detection applications [Krizhevsky et al., 2012, Simonyan and Zisserman, 2014].

A digital hologram is an efficient encoding of a diffraction volume. An obvious requirement in digital holography would be the possibility to edit the hologram directly, in order to effect some semantic change in the diffraction volume, or even more simply, analyse the hologram directly in order to construct an understanding of the 3D scene. Unfortunately, in the general case this has eluded researchers in digital holography, including the authors. Researchers are limited to sampling the reconstruction volume (i.e. using numerical propagation to reconstruct from the digital hologram at a plurality of depths) before they can understand the encoded 3D scene. A handful of notable exceptions exist, such as the landmark papers by Vikram and Billet [Vikram and Billet, 1984] and Onural and Özgen [Onural and Özgen, 1992], and subsequently others over the past decade [Soulez et al., 2007, Cheong et al., 2010, Yevick et al., 2014, Schneider et al., 2016] whose work allows one to determine the size and position of individual particles based on an analysis of the

 ISBN 978-0-9934207-2-6

hologram directly. However, these approaches are limited to the special case of idealized spherical particles. Here, we consider significantly more complicated multi-cellular partially-transparent objects.

Holography has a history as an enabling technology for artificial neural networks [Psaltis and Farhat, 1985], and conventional artificial neural networks have been applied before in the fields of digital holo-graphic microscopy [Kamilov et al., 2015, Schneider et al., 2016] and more generally digital holography [Frauel and Javidi, 2001, Shortt et al., 2006].

In this paper, we demonstrate that it is possible to design a deep convolutional neural network to predict the in-focus distance of a living cell cluster from the digital hologram plane amplitude only. With deep learning, we propose that digital holographic microscopy (DHM) researchers now have a tool at their disposal that is a major step towards removing the need to perform any propagation steps in order to determine the in-focus distance.

2 Deep learning

Deep convolutional neural networks (CNN), that are one form of deep learning, were discovered by LeCun et al. [LeCun et al., 1989]. CNN is an artifical neural network where some of the layers perform convolution operations on their multi-dimensional input. An image convolution is an operation where values of a vector (or pixels of an image) are multiplied by a weights of a sliding vector (or 2D matrix with 2D data such as an image) called a kernel. In convolutional neural network these weights are learned during the training. Four main ideas behind convolutional neural networks are: local connections, shared weights, pooling and the use of many layers [LeCun et al., 2015]. The basic principle of convolutional neural networks with images is that lower levels detect coarse features like edges that are combined by higher levels to parts and further to objects.

Each convolution operation produces a feature map that shares a same kernel, and different feature maps in a layer use a different kernel. The kernel is applied to the feature map of a previous layer, or at the beginning to the input image. Number of feature maps and therefore learned kernels is determined as part of a network architecture design. Output of a convolution is passed through an activation function such as rectified linear unit (ReLU). A standard principle is that higher convolution layers have higher depth (number of feature maps increases).

Pooling layers merge semantically similar features to one reducing dimensionality of feature maps. Pooling is based on small patches of a feature map of which for example a maximum or average is found/calculated and stored as a new value of a sub sampled feature map. Pooling increases the shift invariance [LeCun et al., 2015].

Deep convolutional neural networks that are one form of deep learning have been used successfully in various different visual object recognition and object detection applications [Krizhevsky et al., 2012, Simonyan and Zisserman, 2014]. Some of the layers in these networks perform convolution operations on its multi-dimensional input.

3 Digital holographic microscopy

DHM overcomes a problem present in optical microscopes of a shallow depth-of-field, allowing one to change the in-focus plane after hologram capture. A magnified digital hologram that is formed of reference, R, and object, *O*, waves as $H(x,y) = |R|^2 + |O|^2 + R^*O + RO^*$ can be propagated to any depth z using the Fresnel approximation [Goodman, 2005]

$$U(x,y,z) = \frac{-i}{\lambda z}\exp(\mathrm{i}kz)\, H(x,y) \otimes \exp\left(\mathrm{i}\pi \frac{x^2+y^2}{\lambda z}\right), \quad (1)$$

where λ is the wavelength of the light, ⊗ denotes a convolution operation and k=2π/λ. The amplitude component of the complex-valued reconstruction is defined as

$$A(x,y;z) = \{\mathrm{Re}[U(x,y;z)]^2 + \mathrm{Im}[U(x,y;z)]^2\}^{0,5}, \quad (2)$$

 ISBN 978-0-9934207-2-6

where Re and Im are extracting real and imaginary components, respectively. However, since each in-focus plane has a narrow depth of field, the object of interest is in focus only at a small range of reconstruction depths. The problem of determining the most appropriate in-focus depth is essential for applications such as autofocusing, extended focus imaging, and segmentation. The critical importance of this problem to digital holography researchers is evidenced by the regularity of newly proposed focus metrics to apply amplitude and phase reconstructions such as self-entropy [Gillespie and King, 1989], phase changes [Ferraro et al., 2003],wavelet analysis [Liebling and Unser, 2004], integrated amplitude modulus [Dubois et al., 2006], gray-level variance [McElhinney et al., 2007],power spectra [Langehanenberg et al., 2008], Tamura co-efficient [Memmolo et al., 2012], multi-wavelength Fourier phase [Dohet-Eraly et al., 2016], cosine score [He et al., 2017], structure tensor [Ren et al., 2017], and magnitude differential [Lyu et al., 2017] among others. However, each suffers from the same drawback: a stack of reconstructed images must be computed, and the focus metric must be applied to each reconstruction. This time-consuming drawback is compounded by the fact that the whole procedure must be applied to each new hologram. The greatest benefit of the deep learning method outlined in this paper is that after training, the in-focus depth can be obtained from the hologram plane intensity directly, and in constant time, without any numerical propagation.

4 Experimental results

We chose a convolutional neural network approach to tackle the autofocusing problem still existing in digital holographic microscopy of transparent samples. The architecture of the network is based on the AlexNet architecture that won the Large Scale Visual Recognition Challenge 2012 [Krizhevsky et al., 2012], and was used as a benchmark for subsequent network architectures. One inducement of its use, within the application reported here, is the fact that this paper and results thereof can be considered as a proof of concept with a well known network architecture. AlexNet has 5 convolution layers, 3 fully-connected layers, and uses convolutional filters up to 11×11 pixels in size (Fig. 1)

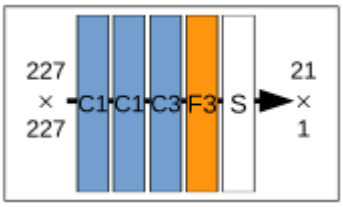

Figure 1: Network architecture. C, convolution block; F, fully-connected block; input size, 227×227 pixels. Numbers show amount of layers in each block. Each convolution block is followed by a maxpooling layer with kernel size of 3 and stride of 2.

4.1 Training

A total of 494 holograms of semitransparent Madin-Darby canine kidney (MDCK) epithelial cell clusters were captured using an off-axis Mach Zehnder digital holographic microscope (Lyncée Tec T1000, Lyncée Tec SA, Lausanne, Switzerland). The microscope comprises a 660 nm laser source, a 1024×1024 pixel CCD camera with 6.45µm square pixels, and a 40X microscope objective with 0.7 numerical aperture (Leica HCX PL Fluotar). To obtain the ground truth data, one of the authors (T.P.) manually determined for each hologram the z (at 1 mm resolution) in Eq. 1 that brings the middle region of each cell cluster into focus. The middle region was considered to be in-focus when edges of cell cluster were estimated to display the lowest diffraction (caused by the top and bottom halves of the sample).

The holograms were used to generate a database of images as follows. An amplitude reconstruction was obtained from each hologram at each of 21 depths distributed equally over the range ± 100mm centred on the in-focus plane (see examples in Fig. 2). Through different combinations of rescaling and cropping, each re-

 ISBN 978-0-9934207-2-6

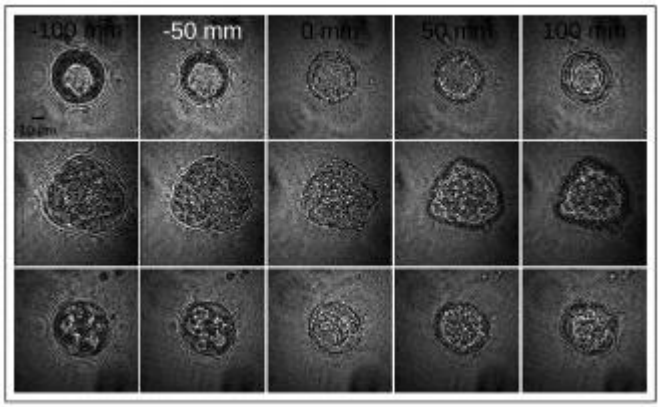

Figure 2: Example training images: each row shows amplitude reconstructions from one hologram (at the in-focus plane, and at the distances ± 100 mm, ± 50 mm from the in-focus plane).

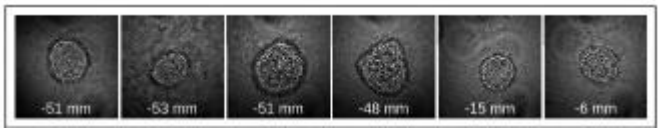

Figure 3: Example amplitudes of the twelve dc- and twin-free holograms used for testing (with ground truth in-focus distance shown).

construction was used to generate six similar but distinct 227×227 pixel rescaled and cropped images. This size was chosen as it is the size the network was originally designed for. Each such image was rotated through each of three distinct 90 deg. rotations, and each resulting image was further augmented through horizontal mirroring. This formed a database of 497 952 images. From this database, all augmented images from the twelve hand-picked holograms (comprising 12 096 images, 2.4% of the set) were set aside as test data (examples shown in Fig. 3). The remaining images were partitioned randomly into training (87.8%, 437 271) and validation (9.8%, 48 585) data. Finally, a mean image (calculated from the training data only) was subtracted from each image.

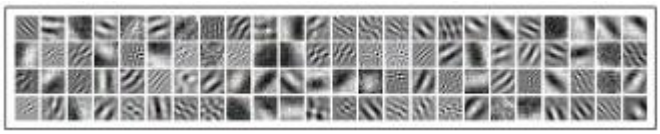

Figure 4: The learned filters from the first convolution layer: 96 11×11 pixel filters

The actual training used Nvidia's Deep Learning GPU Training System (DIGITS) software with Nvidia Titan X ('Maxwell') graphics card. Learning rate was fixed at 0.001. The network was trained for 100 epochs (16 hours) with a stochastic gradient descent solver. The loss function was categorical cross entropy and the network output a 21 element vector containing class probabilities.

The network was trained for 60 epochs with a stochastic gradient descent solver. The minibatch size was set to 100. When the training was finished, training loss was2.497×10^{-2}, validation loss 0.129, and validation accuracy 95.6%. The learned filters from the first convolution layers are shown in the Fig. 4, allowing one to infer the basic features that the network learned to extract from an image for analysis in subsequent layers.

4.2 Testing

Testing was performed on a separate computer without a discrete graphics card to demonstrate the portability of deep learning. As an operating system this computer was running Ubuntu 14.04, had Intel Core i5 processor,

 ISBN 978-0-9934207-2-6

and 16 GB memory. The trained model (with size 227.4 MB) was imported into the Caffe [Jia et al., 2014] deep learning framework using the general-purpose Python programming language. The run time (mean of 200 holograms) was 247 ms. Using the same PC that was used for training (with a GPU support), run time was 4 ms. For comparison, a single Tamura coefficient calculation (including aberration removal, reconstruction, phase unwrapping and Tamura coefficient calculation) is 932 ms (aberration removal 380 ms, reconstruction 318 ms, phase unwrapping 231 ms, Tamura coefficient calculation 3 ms).

Testing was performed by classifying holograms that were not used in training or validation, as explained. Of the 12 096 test images, 99.9% were classified within one class of the ground truth depth (see Fig. 5). Al-though the depth classes are completely unrelated as far as the network is concerned, it is a remarkable indicator of robustness that where model incorrectly classifies an input, it invariably chose a neighbouring depth class instead.

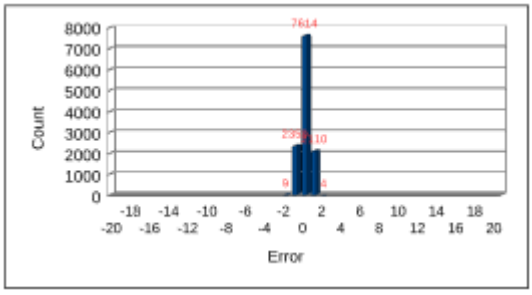

Figure 5: Classification errors with the testing data. In 62.9% of test cases the correct depth class was predicted. In 99.9% of test cases a correct depth is within one depth class.

Table 1: **Test results using the 12 holograms from Fig. 3, showing classification result from the network**

Groundtruth (mm)	Top two predictions (mm,mm)	Confidence values (%,%)
-51	-40,-50	100,0
-53	-50,-40	89,11
-51	-60,-50	99,1
-48	-40,-30	91,9
-15	-10,-20	100,0
-6	-10,0	100,0
1	0,10	100,0
7	0,10	100,0
0	0,-10	56,44
-2	0 -10	100,0
-74	-80,-70	100,0
-65	-60,-50	100,0

To examine how the network responded to holograms that may have an in-focus distance not a multiple of the 10 mm discretization used in training, the holograms from Fig. 3 were used directly (see Table 1). The two top predictions for the network typically straddle the correct answer. The network typically classifies with high confidence holograms with an in-focus distance close to a multiple of 10 mm. The mean absolute error over the 12 holograms was 5.01 mm.

Systematic testing was then performed with the holograms from Fig. 3 over the range±100 mm centred on the in-focus depth, but this time with a finer depth resolution of 1 mm. For a system to generalise well outside the discrete set of 21 in-focus depth classes with which it was trained, the shape of the scatter plot should exactly be a staircase with a linear trend. The network generalized well with each test hologram, and a typical example eis shown in Fig. 6(a). To push the network past its designed capabilities, the network was tested (over the same depth range and resolution) with a human cell line sample captured with the same DHM hardware. The

 ISBN 978-0-9934207-2-6

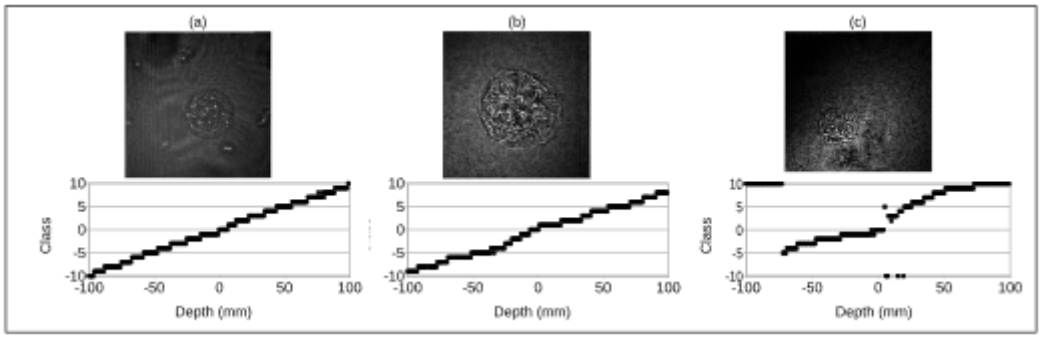

Figure 6: Fine (1 mm) depth-resolution test results (a linear staircase indicates perfect classification): (a) same DHM hardware and same sample type as in training, (b) same DHM hardware but different cell line, (c) different DHM illumination, magnification, and chemically fixed sample from a different cell line.

network performed surprisingly well with this sample, however classification, while largely monotonic, is no longer linear [see Fig. 6(b)]. Later, this exact sample was chemically fixed and captured using a different digital holographic microscope with different illumination and a lower magnification 20X microscope objective with 0.5 numerical aperture (Leica HCX PL Fluotar 20x). The network fails at some depths which is not surprising given the enormous differences in scale and optical density from the training set, overall the network performs well with this data [see 6(c)].

5 Conclusions

In this paper, to our knowledge the first application of deep learning to digital holographic microscopy, we show that an artificial neural network can be designed to learn the appropriate in-focus depth of an arbitrary MDCK cell cluster encoded in a digital hologram. Its greatest benefit is that the in-focus depth can be obtained from the hologram plane intensity in constant time without any numerical propagation. It generalises well to in-focus depths different from its training set, and there is evidence that it will degrade gracefully with differences in cell line, in fixing conditions, and in DHM architecture, from that used in training. As has been discovered in recent years in other fields of microscopy, we believe that deep learning has the potential to become an important tool for DHM.

Acknowledgments

This publication has emanated from research conducted with the financial support of an Irish Research Council (IRC) Postgraduate Scholarship and of Science Foundation Ireland (SFI) under grant no. 13/CDA/2224.

References

[Abdolmanafi et al., 2017] Abdolmanafi, A., Duong, L., Dahdah, N., and Cheriet, F. (2017). Deep feature learning for automatic tissue classification of coronary artery using optical coherence tomography. *Biomed. Opt. Express*, 8(2):1203–1220.

[Cheong et al., 2010] Cheong, F. C., Krishnatreya, B. J., and Grier, D. G. (2010). Strategies for three-dimensional particle tracking with holographic video microscopy. *Opt. Express*, 18(13):13563–13573.

[Cire ̧san et al., 2012] Cire ̧san, D., Giusti, A., Gambardella, L. M., and Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Advances in Neural Information Processing Systems*, 25:2843–2851.

[Dohet-Eraly et al., 2016] Dohet-Eraly, J., Yourassowsky, C., and Dubois, F. (2016). Fast numerical autofocus of multispectral complex fields in digital holographic microscopy with a criterion based on the phase in the Fourier domain. *Opt. Lett*., 41(17):4071–4074.

[Dubois et al., 2006] Dubois, F., Schockaert, C., Callens, N., and Yourassowsky, C. (2006). Focus plane detection criteria in digital holography microscopy by amplitude analysis. *Opt. Express*, 14(13):5895–5908.

[Ferraro et al., 2003] Ferraro, P., Coppola, G., Nicola, S. D., Finizio, A., and Pierattini, G. (2003). Digital holographic microscope with automatic focus tracking by detecting sample displacement in real time. *Opt. Lett*., 28(14):1257–1259.

[Frauel and Javidi, 2001] Frauel, Y. and Javidi, B. (2001). Neural network for three-dimensional object recognition based on digital holography. *Opt. Lett*., 26(19):1478–1480.

[Gillespie and King, 1989] Gillespie, J. and King, R. A. (1989). The use of self-entropy as a focus measure in digital holography. *Pattern Recogn. Lett*., 9(1):19–25.

[Goodman, 2005] Goodman, J. W. (2005). *Introduction to Fourier Optics*. Roberts and Company Publishers.

[Gopakumar et al., 2017] Gopakumar, G., Babu, K. H., Mishra, D., Gorthi, S. S., and Subrahmanyam, G. R. K. S. (2017). Cytopathological image analysis using deep-learning networks in microfluidic microscopy. *J. Opt. Soc*. Am. A, 34(1):111–121.

[He et al., 2017] He, G., Xiao, W., and Pan, F. (2017). Automatic focus determination through cosine and modified cosine score in digital holography. *Opt. Eng*., 56(3):034103.

[Jia et al., 2014] Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., and Darrell, T. (2014). Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv*: 1408.5093.

[Kamilov et al., 2015] Kamilov, U. S., Papadopoulos, I. N., Shoreh, M. H., Goy, A., Vonesch, C., Unser, M., and Psaltis, D. (2015). Learning approach to optical tomography. *Optica*, 2(6):517–622.

[Karri et al., 2017] Karri, S. P. K., Chakraborty, D., and Chatterjee, J. (2017). Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration. *Biomed. Opt. Express*, 8(2):579–592.

[Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25:1097–1105.

[Langehanenberg et al., 2008] Langehanenberg, P., Kemper, B., Dirksen, D., and von Bally, G. (2008). Autofocusing in digital holographic phase contrast microscopy on pure phase objects for live cell imaging. *Appl. Opt.*, 47(19):D176–D182.

[LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

[LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

[Liebling and Unser, 2004] Liebling, M. and Unser, M. (2004). Autofocus for digital Fresnel holograms by use of a Fresnelet-sparsity criterion. *J. Opt. Soc*. Am. A, 21(12):2424–2430.

[Lyu et al., 2017] Lyu, M., Yuan, C., Li, D., and Situ, G. (2017). Fast autofocusing in digital holography using the magnitude differential. *Appl. Opt*., 56(13):F152–F157.

[McElhinney et al., 2007] McElhinney, C. P., McDonald, J. B., Castro, A., Frauel, Y., Javidi, B., and Naughton, T. J. (2007). Depth-independent segmentation of macroscopic three-dimensional objects encoded in single perspectives of digital holograms. *Opt. Lett.*, 32(10):1229–1231.

[Memmolo et al., 2012] Memmolo, P., Iannone, M., Ventre, M., Netti, P. A., Finizio, A., Paturzo, M., and Ferraro, P. (2012). On the holographic 3d tracking of in vitro cells characterized by a highly-morphological change. *Opt. Express*, 20(27):28485–28493.

[Onural and Özgen, 1992] Onural, L. and Özgen, M. T. (1992). Extraction of three-dimensional object-location information directly from in-line holograms using Wigner analysis. *J. Opt. Soc.* Am. A, 9(2):252–260.

[Prentaši´c et al., 2016] Prentaši´c, P., Heisler, M., Mammo, Z., Lee, S., Merkur, A., Navajas, E., Beg, M. F., Šaruni´c, M., and Lonˇcari´c, S. (2016). Segmentation of the foveal microvasculature using deep learning networks. *J. Biomed. Opt*., 21(7):075008.

[Psaltis and Farhat, 1985] Psaltis, D. and Farhat, N. (1985). Optical information processing based on an associative-memory model of neural nets with thresholding and feedback. *Opt. Lett*., 10(2):98–100.

[Ren et al., 2017] Ren, Z., Chen, N., and Lam, E. Y. (2017). Automatic focusing for multisectional objects in digital holography using the structure tensor. *Opt. Lett*., 42(9):1720–1723.

[Rezaeilouyeh et al., 2016] Rezaeilouyeh, H., Mollahosseini, A., and Mahoor, M. H. (2016). Microscopic medical image classification framework via deep learning and shearlet transform. *J. Med. Imaging*, 3(4):044501.

[Russakovsky et al., 2015] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., and Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *Int. J. Comput. Vision*, 115(3):211–252.

[Schneider et al., 2016] Schneider, B., Dambre, J., and Bienstman, P. (2016). Fast particle characterization using digital holography and neural networks. *Appl. Opt*., 55(1):133–139.

[Shortt et al., 2006] Shortt, A. E., Naughton, T. J., and Javidi, B. (2006). Compression of optically encrypted digital holograms using artificial neural networks. *J. Display Technol*., 2(4):401–410.

[Simonyan and Zisserman, 2014] Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv*:1409.1556.

[Soulez et al., 2007] Soulez, F., Denis, L., Fournier, C., Thiébaut, É., and Goepfert, C. (2007). Inverse-problem approach for particle digital holography: accurate location based on local optimization. *J. Opt. Soc. Am. A*, 24(4):1164–1171.

[Vikram and Billet, 1984] Vikram, C. S. and Billet, M. L. (1984). Far-field holography at non-image planes for size analysis of small particles. *Applied Physics B*, 33(3):149–153.

[Wang et al., 2014] Wang, H., Cruz-Roa, A., Basavanhally, A., Gilmore, H., Shih, N., Feldman, M., Tomaszewski, J., Gonzalez, F., and Madabhushi, A. (2014). Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *J. Med. Imaging*, 1(3):034003.

[Yevick et al., 2014] Yevick, A., Hannel, M., and Grier, D. G. (2014). Machine-learning approach to holographic particle characterization. *Opt. Express*, 22(22):26884–26890.

 ISBN 978-0-9934207-2-6