Automated detection of macrophages in quantitative phase images by deep learning using a Mask Region-based Convolutional Neural Network

Kai Eder[1], Tobias Kutscher[2], Anne Marzi[1], Álvaro Barroso[1], Jürgen Schnekenburger[1], Björn Kemper[1]

[1]Biomedical Technology Center of the Medical Faculty, University of Muenster D-48149 Muenster, Germany

[2]Nova Information Management School (NOVA IMS), Universidade Nova de Lisboa Lisboa, 1070-312, Portugal

ABSTRACT

We explored a Mask Region-based Convolutional Neural Network (Mask R-CNN) to detect macrophages in quantitative phase images, which were acquired by digital holographic microscopy (DHM), an interferometry-based variant of quantitative phase imaging (QPI). The Mask R-CNN deep learning architecture is capable to detect and segment single macrophage cells in quantitative phase images and allows to perform both tasks in a multi-stage process. Our results show that the combined detection and segmentation of cells through Mask R-CNN-based automated evaluation prospects a fast and robust screening in label-free high throughput microscopy.

Keywords: Digital holographic microscopy, quantitative phase imaging, deep learning, artificial intelligence, object detection, macrophages

1. INTRODUCTION

Quantitative phase imaging (QPI) has been demonstrated to be a versatile tool for minimally invasive label-free imaging of biological specimens and time-resolved cellular analysis [1-8]. For extraction of quantitative single cell data like dry mass and morphology from QPI images and their further evaluation, e.g., with machine learning algorithms, segmentation methods are crucial [9-14]. Cell segmentation commonly is conducted by thresholding and watershed-based methods [15]. These methods are widely used, as they can be implemented with moderate efforts and are computational efficient. However, their application for segmentation of complex cell structures or noisy images is challenging and critical, as segmentation errors can bias downstream analyses of quantitative single cell data [16]. New evolving deep learning approaches promise to improve segmentation [15, 17, 18]. We thus explored in this study their application for the analysis of QPI images of RAW 264.7 mouse macrophages that were observed by digital holographic microscopy (DHM) [19], an interferometry-based variant of (QPI), during toxicological studies and cellular growth experiments. Based on the acquired DHM QPI image data set we subsequently performed the detection and segmentation task utilizing a Convolutional Neuronal Network (CNN) with Mask region-based CNN (Mask R-CNN). The advantages of the Mask R-CNN are it's capability to segment individual objects in images and to distinguish between several classes of objects in a single trainable end-to-end manner. No further postprocessing tasks of the prediction results are needed [20]. The Mask R-CNN architecture was proven to deliver desirable results for biomedical data and in phase microscopy images [21, 22]. In earlier studies the ability of the Mask R-CNN to detect and segment objects was demonstrated, even when cells were overlapping or clustering [23]. In this study, we evaluated the performance of the Mask R-CNN for the segmentation and classification of macrophages in QPI images regarding two different Feature Pyramid Networks (FPN) and three learning rate settings.

[1]kai.eder@uni-muenster.de

Label-free Biomedical Imaging and Sensing (LBIS) 2021, edited by Natan T. Shaked, Oliver Hayden, Proc. of SPIE Vol. 11655, 116551K · © 2021 SPIE
CCC code: 1605-7422/21/$21 · doi: 10.1117/12.2577232

2. MATERIALS AND METHODS

2.1 Generation of quantitative phase images from RAW 264.7 macrophages

An inverted Nikon Ts2R microscope equipped with an automated microscope stage and an attached DHM module, based on concepts described previously [24], was applied for quantitative phase contrast imaging of living RAW 264.7 macrophages. The RAW 264.7 mouse macrophage cell line used for this study was obtained from ATCC (American Type Culture Collection, Manassas, Virginia) and subcultured routinely twice a week in standard cell culture as described in [25]. DHM imaging experiments were performed on subconfluent cells in µ-well plates with glass lids (ibidi GmbH, Munich, Germany). An incubator chamber (ibidi GmbH, Munich, Germany) allowed investigations of living cells in physiological environment (5% CO_2 atmosphere, T = 37°C. The coherent light source for the recording of digital holograms was a fiber coupled solid state laser (Cobolt 06-DPL, λ=532 nm, Cobolt AB, Sweden). Digital off-axis holograms of the specimens were recorded with a complementary metal-oxide-semiconductor (CMOS) sensor (UI-3260CP-M-GL Rev.2, IDS, Germany) using a 20x microscope lens (Nikon Plan 20x/0.4, Nikon, Japan). The reconstruction of quantitative phase images from digitally captured holograms was performed numerically as described with details in [26].

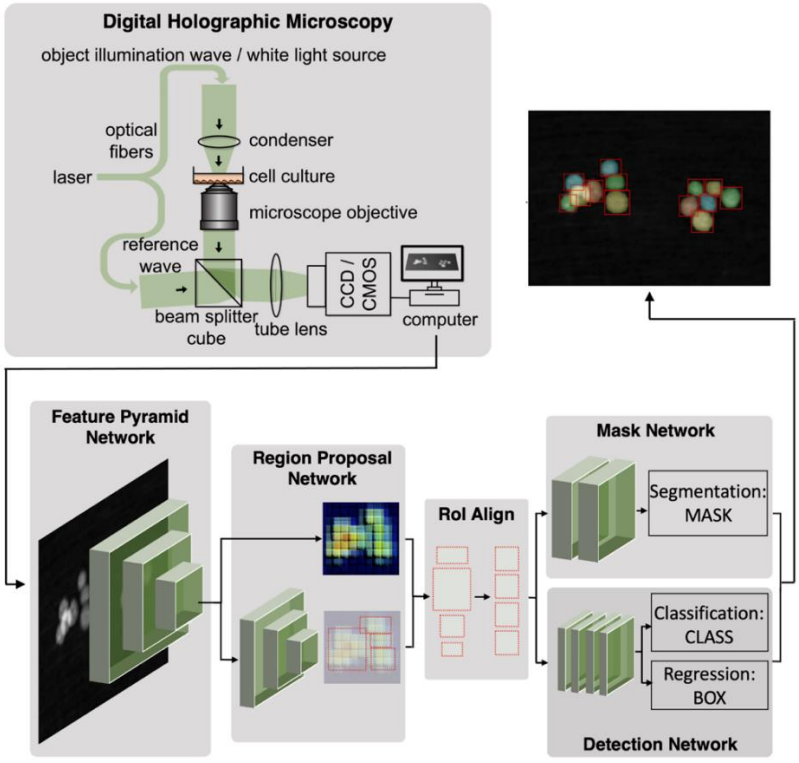

Figure 1: Illustration of the QPI image acquisition process with DHM and the subsequent training process of the proposed Mask R-CNN. Digital off-axis holograms of RAW 264.7 macrophages are recorded from which subsequently quantitative phase images are reconstructed. The QPI images are then transferred into the Mask R-CNN. The Feature Pyramid Network generates features from the inputted images, which are then used for learning the object shapes and the generation of object proposals with the Regional Proposal Network. The Region of Interest Align (RoI Align) function then converts all proposals to the same size by using bilinear interpolation of pixels. The equally sized grids are then inserted into the Mask Network for segmenting the pixels of an object and into the detection network, which classifies the objects and regresses the bounding box shape.

2.2 Dataset

In total 97 QPI images of RAW 264.7 macrophages from different experimental sequences of a toxicological *in vitro* experiment were manually labeled. Cells were segmented and classified into two phenotypes: either "normal" for round

shaped cells or "multi-shaped" for macrophages with elongated, different shapes. Both, train and test set were sampled from randomly selected QPI sequences in order to avoid biasing the testing process, as images from the same sequence can be highly similar. In analogy, the validation set was sampled from the same sequences as the train set. The train set consisted of 49 images including 3,855 normal cells and 638 multi-shaped cells, the validation set of 6 images with 264 and 106 normal and multi-shaped cells respectively (Table 1). The final test set consisted of 42 images with 2,332 normal and 425 multi-shaped cells and was found sufficiently large and diverse to evaluate the performance of the Mask R-CNN model.

Dataset	Number of images	Number of cells	Number of normal cells	Number of multi-shaped cells
Train	49	4,483	3,855	628
Validation	6	370	264	106
Test	42	2,747	2,332	415
Total	97	7,600	6,451	1,149

Table 1: Overview of the amount of annotated quantitative phase images of RAW 264.7 macrophages for training, validation and testing of the neural network.

2.3 Model architecture and training parameters

The entire model architecture is illustrated in Figure 1. Two different Deep Residual Learning architectures, also called ResNet architectures were tested to evaluate their performance as the FPN, namely ResNet50 and ResNet101 [27]. The ResNet50 and ResNet101 were pre-trained on the Common Objects in Context (COCO) dataset [28]. As macrophages appeared relatively small in the investigated field of view, the anchor generation sizes are reduced from 32, 64, 128, 256, 512 to 16, 32, 64 pixels. The corresponding aspect ratios are set to 0.5, 1, 2. The Non-maximum Suppression (NMS) threshold in the Region Proposal Network was set to 0.4. After filtering the proposals with NSM, only 4,000 top scoring anchor boxes are kept during training and 2,000 during testing, due to the high number of cells in the images. The RPN loss was weighted 1.5 times higher in the total loss function. This is done, because the main challenge of the algorithm is to localize and detect all cells in densely crowded areas, where cells can also significantly overlap. The intersection over union (IoU) for region of interest (ROI) boxes is set to 0.5. The batch size per image is set to 512 with a positive fraction ratio of 25%. Stochastic Gradient Descent (SGD) with momentum of 0.9 and weight decay of 0.0001 was used for the optimization process [29]. The algorithm was trained for 14,700 iterations, which is equal to 600 epochs. Three learning rates were tested during training, 0.01, 0.001 and 0.00025. Data Augmentation was used to enhance the variety of the dataset. In detail random brightness, saturation, contrast, rotation, flip and crop was used. The Mask R-CNN model was developed using PyTorch and the Detectron2 framework and trained with a Nvidia Tesla K80 GPU.

3. RESULTS

Six different configurations of the Mask R-CNN were tested and evaluated, namely two FPN's and three learning rates, which are listed in Table 2 that also gives an overview of the performance of all test results. It can be observed that the best performing model was trained with a ResNet101 backbone and a learning rate of 0.001.

Backbone FPN	Learning Rate	Box mAP	Segmentation mAP
ResNet50	0.01	80.08	80.15
	0.001	80.27	80.33
	0.00025	78.08	78.69
ResNet101	0.01	78.25	78.99
	0.001	**82.69**	**82.83**
	0.00025	80.83	80.85

Table 2: Overview of the analyzed models and their performances in terms of mAP for the bounding box and segmentation for different backbone FPN architectures and learning rates. mAP results are show with two decimals, for better understanding and observation of the model performance.

The PASCAL VOC mean average precision (mAP) for the test set is 82.69 and 82.83 for the bounding box and segmentation respectively [30]. Fig. 2 shows on the left the performances of the total training and validation loss and on the right the underlying sub-loss functions. The best performing model, which used a ResNet101 as FPN and is trained with a learning rate of 0.001, converges fast. The training total loss is steadily and slowly decreasing after a fast-decreasing process in the beginning of the training. The validation loss is evaluated every 500th iteration, has a similar behavior and does not decease significantly anymore after 12,000 iterations.

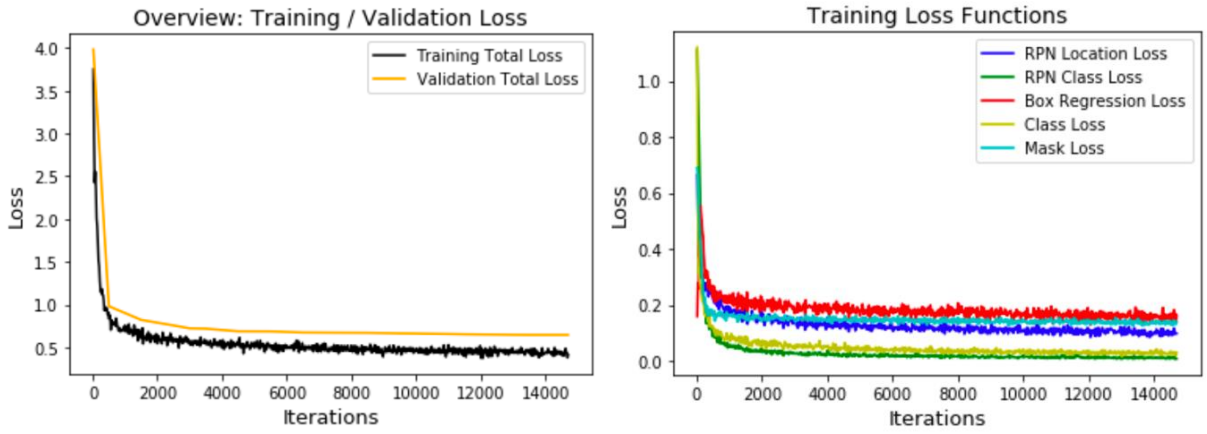

Figure 2: Visualization of the performances of the total training and validation loss (left) and the underlying sub-loss functions of the total loss function (right) during the Mask R-CNN training process. The loss curves visualize the history of improvement during the training process of the model. All loss functions are converging and improving.

The percentage of correctly predicted objects out of all predicted objects (detection accuracy) is 93.5% for the best performing model, based on the mAP. However, the classes are imbalanced so the detection accuracy might be biased towards the more present class.

Measure	**Performance**
Bounding Box mAP	82.69
Segmentation mAP	82.83
Detection Accuracy	0.935
MAE – Total	1.6
MAPE – Total	4.4 %
Relative Difference – Total	1.1%
MAE – Normal cells	1.7
MAPE – Normal cells	6.1 %
Relative Difference – Normal cells	0.04%
MAE – Multi-shaped cells	1.2
MAPE – Multi-shaped cells	18.9%
Relative Difference – Multi-shaped cells	7.5%

Table 3: Overview of all measured performance metrics for the best performing model (based on the mAP)

The mean absolute percentage error (MAPE) of the ground truth number and the predicted number of cells in one image is 4.4%, for normal cells and 6.1% and 18.9% for multi-shaped cells as shown in Table 3. In contrast, the mean absolute error (MAE) is 1.6 in total, for normal cells it is 1.7 and 1.2 for multi-shaped macrophages. On the whole dataset perspective error rates are lower, as over- and underpredictions in the images cancel each other out. In total, the test set contains 2,747 cells and 2,777 cells are predicted by the model, so the relative difference is 1.1%. Moreover, the test set contains 2,332 normal cells and the model predicts 2,331. The relative difference amounts to 0.04%. The actual number of multi-shaped cells is 415, but the model predicts 446, with results in a relative difference is 7.47%. Note that the

difference between the ground truth number of cells with the predicted number of cells is a trivial metric and does not consider false predictions. Considering all performance metrics from Table 3, the mAP is the most reliable, as it is less biased towards class imbalance. To validate the quantitative evaluation, a visual evaluation by an expert was conducted. Figure 3 shows the performance of the best performing model for a single exemplary DHM QPI image and illustrates the precision of the detection and segmentation of the best performing model.

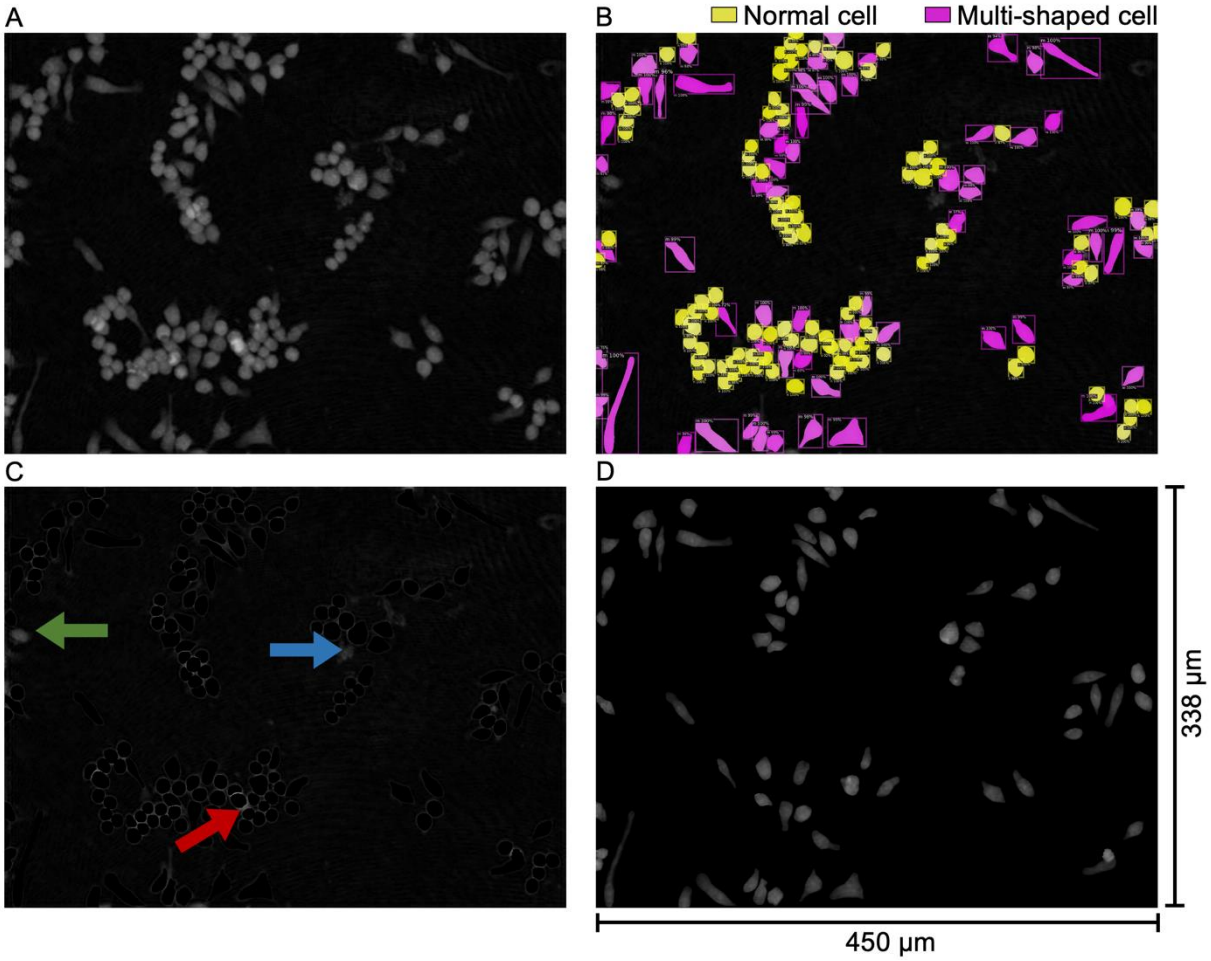

Figure 3: (**A**): QPI image of macrophages generated by DHM as illustrated in Fig. 1. (**B**): Illustration of the masks for normal cells (yellow) and multi-shaped cells (purple) predicted by the best performing model. (**C**): QPI image in (A) after removing all pixels of the detected masks. Arrows indicated segmentation errors (red), undetected cells (green), and cell debris that is correctly not detected as a living cell (blue). (**D**): QPI image in (A) in which only pixels of multi-shaped predicted cells are displayed.

4. CONCLUSIONS

Our results show that a Mask R-CNN can successfully be used efficiently to detect, segment, and classify RAW 264.7 mouse macrophages in DHM QPI images. A training data set of 49 images was sufficient to achieve a mAP of 82.69 and 82.83 for the bounding box and segmentation, respectively, on the test set consisting of 42 images. A detection accuracy of 93.5% was achieved. Mask R-CNN applied with the described parameters for automated cell detection and segmentation allows the extraction of single-cell quantitative physical data from QPI images of cell populations. In summary, the investigated deep learning approach for automated detection and segmentation of cells in QPI images prospects to be a valuable extension to the methods for the reliable extraction of cell type specific single cell information like dry mass from label-free acquired high-throughput quantitative phase microscopy *in vitro* image data.

Acknowledgements

This work was supported by funding from the EU HORIZON 2020 Project Regulatory Science Framework for Nano(bio)material-based Medical Products and Devices (REFINE, Grant agreement ID: 761104) and German Federal Ministry of Education and Research (BMBF) Project NanoBioQuant (FKZ 03XP0213D).

5. REFERENCES

[1] B. Kemper, and G. von Bally, "Digital holographic microscopy for live cell applications and technical inspection," Appl Opt, 47(4), A52-61 (2008).

[2] M. Kim, "Principles and techniques of digital holographic microscopy," SPIE Reviews, 1(1), 018005 (2010).

[3] B. Kemper, A. Bauwens, D. Bettenworth *et al.*, "Label-Free Quantitative In Vitro Live Cell Imaging with Digital Holographic Microscopy" Springer International Publishing, Cham(2019).

[4] K. Lee, K. Kim, J. Jung *et al.*, "Quantitative phase imaging techniques for the study of cell pathophysiology: from principles to applications," Sensors (Basel), 13(4), 4170-91 (2013).

[5] P. Marquet, C. Depeursinge, and P. J. Magistretti, "Review of quantitative phase-digital holographic microscopy: promising novel imaging technique to resolve neuronal network activity and identify cellular biomarkers of psychiatric disorders," Neurophotonics, 1(2), 020901 (2014).

[6] B. Kemper, A. Bauwens, A. Vollmer *et al.*, "Label-free quantitative cell division monitoring of endothelial cells by digital holographic microscopy," Journal of Biomedical Optics, 15(3), 036009 (2010).

[7] Y. Park, C. Depeursinge, and G. Popescu, "Quantitative phase imaging in biomedicine," Nature Photonics, 12(10), 578-589 (2018).

[8] D. Bettenworth, A. Bokemeyer, C. Poremba *et al.*, "Quantitative phase microscopy for evaluation of intestinal inflammation and wound healing utilizing label-free biophysical markers," Histol Histopathol, 33(5), 417-432 (2018).

[9] H. S. Park, M. T. Rinehart, K. A. Walzer *et al.*, "Automated Detection of P. falciparum Using Machine Learning Algorithms with Quantitative Phase Images of Unstained Cells," PLOS ONE, 11(9), e0163045 (2016).

[10] S. K. Mirsky, I. Barnea, M. Levi *et al.*, "Automated analysis of individual sperm cells using stain-free interferometric phase microscopy and machine learning," Cytometry Part A, 91(9), 893-900 (2017).

[11] J. Yoon, Y. Jo, M.-h. Kim *et al.*, "Identification of non-activated lymphocytes using three-dimensional refractive index tomography and machine learning," Scientific Reports, 7(1), 6654 (2017).

[12] T. H. Nguyen, S. Sridharan, V. Macias *et al.*, "Automatic Gleason grading of prostate cancer using quantitative phase imaging and machine learning," Journal of Biomedical Optics, 22(3), 036015 (2017).

[13] V. K. Lam, T. Nguyen, T. Phan *et al.*, "Machine Learning with Optical Phase Signatures for Phenotypic Profiling of Cell Lines," Cytometry A, 95(7), 757-768 (2019).

[14] V. Lam, T. Nguyen, V. Bui *et al.*, "Quantitative scoring of epithelial and mesenchymal qualities of cancer cells using machine learning and quantitative phase imaging," J Biomed Opt, 25(2), 1-17 (2020).

[15] S. Beucher, "The watershed transformation applied to image segmentation," Scanning Microscopy Supplement, 299-314 (1992).

[16] J. C. Caicedo, J. Roth, A. Goodman *et al.*, "Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images," bioRxiv, 335216 (2018).

[17] F. Milletari, N. Navab, and S. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation." 565-571.

[18] M. Z. Alom, M. Hasan, C. Yakopcic *et al.*, "Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation," (2018).

[19] B. Kemper, and G. von Bally, "Digital holographic microscopy for live cell applications and technical inspection," Applied Optics, 47(4), A52-A61 (2008).

[20] K. He, G. Gkioxari, P. Dollár *et al.*, "Mask R-CNN," 2017 IEEE International Conference on Computer Vision (ICCV), 2980-2988 (2017).
[21] J. C. Caicedo, A. Goodman, K. W. Karhohs *et al.*, "Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl," Nature Methods, 16(12), 1247-1253 (2019).
[22] P. Yuan, A. Rezvan, X. Li *et al.*, "Phasetime: Deep Learning Approach to Detect Nuclei in Time Lapse Phase Images," J Clin Med, 8(8), (2019).
[23] A. O. Vuola, S. U. Akram, and J. Kannala, "Mask-RCNN and U-Net Ensembled for Nuclei Segmentation," 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 208-212 (2019).
[24] P. Lenz, M. Bruckner, S. Ketelhut *et al.*, "Multimodal Quantitative Phase Imaging with Digital Holographic Microscopy Accurately Assesses Intestinal Inflammation and Epithelial Wound Healing," J Vis Exp(115), (2016).
[25] M. Baczewska, K. Eder, S. Ketelhut *et al.*, Holotomographic investigation of an influence of PFA cell fixation process on refractive index of cellular organelles in epithelial cells SPIE, PWB (2020).
[26] J. Min, B. Yao, V. Trendafilova *et al.*, "Quantitative phase imaging of cells in a flow cytometry arrangement utilizing Michelson interferometer-based off-axis digital holographic microscopy," Journal of Biophotonics, 12(9), e201900085 (2019).
[27] K. He, X. Zhang, S. Ren *et al.*, "Deep Residual Learning for Image Recognition," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770-778 (2016).
[28] T.-Y. Lin, M. Maire, S. J. Belongie *et al.*, "Microsoft COCO: Common Objects in Context," ArXiv, abs/1405.0312, (2014).
[29] J. Kiefer, and J. Wolfowitz, "Stochastic Estimation of the Maximum of a Regression Function," Ann. Math. Statist., 23(3), 462-466 (1952).
[30] M. Everingham, L. Van Gool, C. K. I. Williams *et al.*, "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision, 88(2), 303-338 (2010).