Convolutional neural networks predict mitochondrial structures from label-free microscopy images

Hsu, Chan-Min, Lee, Yi-Ju, Wei, An-Chi

Chan-Min Hsu, Yi-Ju Lee, An-Chi Wei, "Convolutional neural networks predict mitochondrial structures from label-free microscopy images," Proc. SPIE 11792, International Forum on Medical Imaging in Asia 2021, 117920G (20 April 2021); doi: 10.1117/12.2591089

SPIE.

Event: International Forum on Medical Imaging in Asia 2021, 2021, Taipei, Taiwan

Convolutional neural networks predict mitochondrial structures from label-free microscopy images

Chan-Min Hsu [a], Yi-Ju Lee [a], An-Chi Wei [a,*]
[a] Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan 106

ABSTRACT

Convolutional neural networks (CNNs) have shown significant success in image recognition and segmentation. Based on a CNN-like U-Net architecture, such a model can effectively predict subcellular structures from transmitted light (TL) images after learning the relationships between TL images and fluorescent-labeled images.
In this paper, we focused on building corresponding models of subcellular mitochondrial structures using the CNN method and compared the prediction results derived from confocal microscopic, Airyscan microscopic, z-stack, and time-series images. With multi-model combined prediction, it is possible to generate integrated images using only TL inputs, which reduces the time required for sample preparation and increases the temporal resolution. This enables visualization, measurement, and understanding of the morphology and dynamics of mitochondria and mitochondrial DNA.

Keywords: convolutional neural networks, label-free imaging, confocal microscopy, mitochondria

1. INTRODUCTION

Mitochondria are dynamic organelles that are able to continuously alter their morphologies and internal cristae structures to adjust energy output in response to different physiological conditions and bioenergetic needs[1]. Mitochondrial architecture influences the efficiency of mitochondrial energy production depending on the cell type, tissue, and physiological and pathological conditions[2]. Understanding mitochondria and mitochondrial DNA (mtDNA) structure and dynamics is the first step in understanding mitochondrial functions[3].

The technique of fluorescence microscopy has been adopted in live-cell imaging to study mitochondrial structures and dynamics in recent decades[4–6]. Fluorescence microscopy resolves subcellular structure in living cells through specific labeling but requires technical instrumentation and time-consuming sample preparation. In addition, the challenges of photobleaching and phototoxicity have emerged in z-stack and time-lapse imaging, creating a tradeoff between the quality (image spatial resolution) and quantity (temporal resolution) achievable for live-cell imaging[7]. In contrast, transmitted light

* acwei86@ntu.edu.tw; phone +886-233668612; fax +886-233663754

International Forum on Medical Imaging in Asia 2021, edited by Ruey-Feng Chang, Proc. of SPIE
Vol. 11792, 117920G · © 2021 SPIE · CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2591089

(TL) microscopy does not require labeling, which can significantly reduce phototoxicity, sample preparation complexity, and experimental cost[8].

Advances in deep learning have recently revealed its potential to achieve significant success in image processing. Convolutional neural networks (CNNs), one branch of deep learning, can learn nonlinear relationships between source and target images, resulting in considerably improved performance in computer vision tasks, such as classification and segmentation[9]. A method combining relatively low-cost TL images with clear fluorescent images would be a useful tool for observing subcellular structures in a quick and efficient way[10]. To implement such tasks, two research groups have previously constructed CNNs to successfully predict subcellular structures from TL images[11,12]. In their work, they trained the CNN models with unlabeled TL images (target) and fluorescence-labeled images (ground truth) and predicted staining patterns of fluorescence images from unlabeled TL images. Although the prediction studies previously done have addressed 3D cell imaging, their models did not perform well for mitochondrial structure prediction in AC16 cardiac-derived cell line images under our own experimental settings. This may be because few of those previous studies focused on mitochondria imaging, especially using time-series images that can capture mitochondrial dynamics.

To gain insights into improved tools for the study of the morphology and dynamics of mitochondria, we adopted and modified the label-free U-Net method[11,13] to train and predict z-stack and time-series fluorescence images of AC16 cell TL images. Using both confocal and Airyscan technology[14], we further improved the overall performance of CNN models in predicting mitochondrial structures.

2. METHODOLOGY

2.1 Cell culture and labeling

The AC16 human cardiomyocyte cell line (AC16) was used and seeded on a 30 mm glass-bottom plate (DMEM/F12 with 12.5% FBS and 1% antibiotic–antimycotic). The cell density was 250 to 500 thousand cells per plate. To stain the cells, AC16 cells were incubated in imaging media with 100 nM tetramethylrhodamine, methyl ester (TMRM) for 15 min, followed by incubation with 2000× to 5000× SYBR Gold™ (Thermo Fisher Scientific Inc.)[15] for 30 min.

2.2 Image acquisitions

Cells were imaged on a Zeiss microscope LSM800 with a Plan-Apochromat 1.40-NA, 63× objective, and Zeiss Zen Blue 2.6 software was used to acquire images. Three channels were used to acquire TL, SYBR Gold-labeled (nuclear and mitochondrial DNA), and TMRM-labeled (mitochondria) images. The z-stack and time-series image acquisition setting and resolution are listed in Table 1.

2.3 Model architecture

The modified U-Net model[13] comprises of different layers that perform convolution functions: convolutional layers with a stride of 2 pixels in a contracting path, convolutional layers with a stride of 1 pixel, and deconvolutional layers (transposed convolution) with a stride of 2 pixels in an expansive path, each with batch normalization and ReLU function (Fig. 1a). In the last layer of the model, there is no ReLU function or batch normalization. The number of output channels in each layer is illustrated in Fig. 1a. For comparison, a different network, global voxel transformer networks (GVTNets)[16], was implemented and tested. GVTNets introduced an attention operator, which can compute each output unit as a weighted sum of all input units, into the original U-Net model. In addition, the weights computed by this operator are input-dependent.

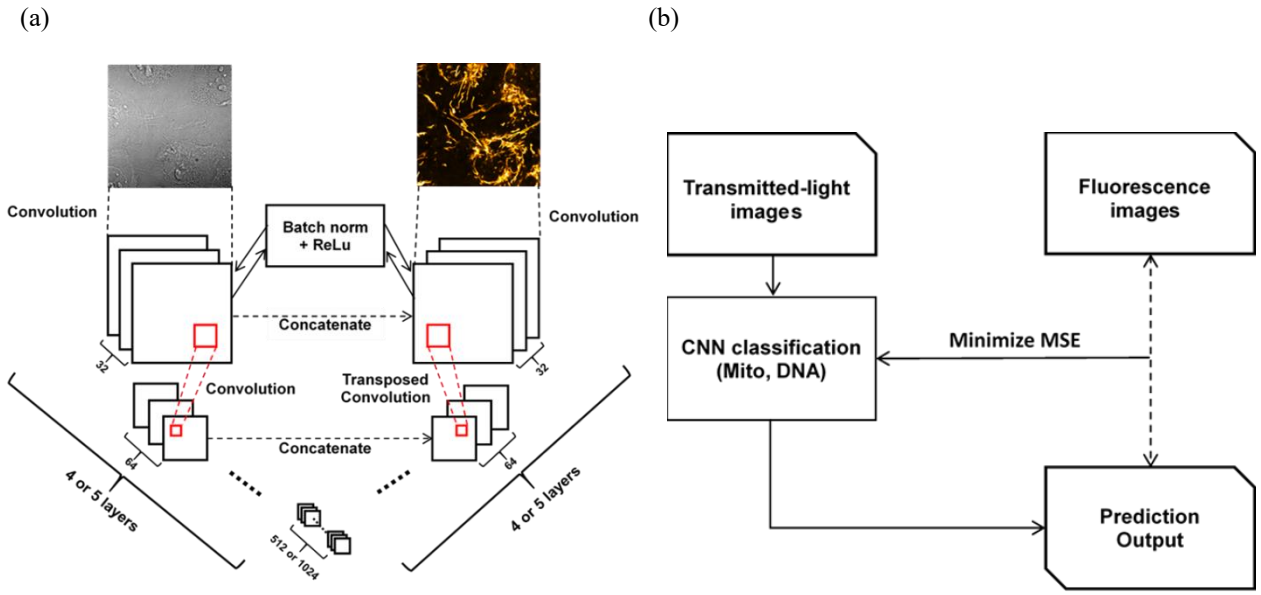

Figure 1. (a) The U-net model architecture. (b) Flow chart of the subcellular structure prediction of AC16 cell using convolutional neural networks. The models predict mitochondria and DNA (including mitochondrial DNA and nuclear DNA) staining patterns of fluorescence images from unlabeled TL images.

2.4 Model implementation

The model and codes were adapted from the paper published by Ounkomol et al. (2018)[11]. Briefly, the TL and fluorescence image pairs were used as inputs from different types of microscopy data, such as short-interval time-lapse, long-interval time-lapse, z-stacks at low and high resolutions (512×512 vs. 1834×1834 pixels), confocal microscopy and an Airyscan detector module (Table 1). Both time-series data and z-stacks were treated as 3D data with the third dimension in time or z-axis (XYT or XYZ). The image data were split into two sets: 75% training and 25% testing. The training data were then split again into two sets: 90% training and 10% validation (Fig. 1b). All patches were randomly subsampled across all training images. The training procedure updated its parameters via the stochastic gradient descent to minimize the mean squared error between the fluorescence ground truth and model output. Adam optimizer was used as an optimization

method with a learning rate of 0.001 for different batch iterations (depending on the size of training data). The optimized parameters can be obtained after 10,000 – 20,000 iterations.

The training pipelines were implemented using Python and the PyTorch package. The model was trained on a GeForce GTX 1080Ti with 12 GB RAM with a patch size of 32×64×64 pixels. As the patch size increased, we trained the model using data parallelism on multiple Tesla V100s with 32 GB RAM provided by Taiwan Computing Cloud (TWCC; https://www.twcc.ai/).

Table 1. The live-cell imaging data used in this research.

Structure (acquisition)	Model name	Third dimension	Number of data set (train/test)	Resolution	Number of slices (frames) per set
Mitochondria (Airyscan)	A_1	Time series (15 min)	90/30	512×512 (default)	64
DNA (Airyscan)	A_2	Time series (15 min)	90/30	512×512 (default)	64
Mitochondria (confocal)	B_1	Time series (5 s)	150/50	512×512 (default)	64
DNA (confocal)	B_2	Time series (5 s)	150/50	512×512 (default)	64
Mitochondria (Airyscan)	C_1	Time series (1 min)	47/20	1834×1834(optimal)	32
DNA (Airyscan)	C_2	Time series (1 min)	47/20	1834×1834 (optimal)	32
Mitochondria (confocal)	D_1	Time series (30 s)	60/15	917×917	32
DNA (confocal)	D_2	Time series (30 s)	60/15	917×917	32
Mitochondria (confocal)	E_1	Z-stack (0.100 μm)	36/14	512×512 (default)	64
DNA (confocal)	E_2	Z-stack (0.100 μm)	36/14	512×512 (default)	64
Mitochondria (Airyscan)	F_1	Z-stack (0.150 μm)	47/18	1834×1834 (optimal)	32
DNA (Airyscan)	F_2	Z-stack (0.150 μm)	47/18	1834×1834 (optimal)	32
Mitochondria (Airyscan)	G_1	Z-stack (0.150 μm)	47/18	917×917 (downscale from F_1)	32
DNA (Airyscan)	G_2	Z-stack (0.150 μm)	47/18	917×917 (downscale from F_2)	32
Mitochondria (confocal)	H_1	Z-stack (0.150 μm)	46/17	917×917	32
DNA (confocal)	H_2	Z-stack (0.150 μm)	46/17	917×917	32

2.5 Model performance analysis

The Pearson correlation coefficient (r) was used to quantify accuracy,

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}} \quad (1)$$

where y stands for the pixel intensities of the model's prediction (output), x stands for the ground truth test images, and $\bar{x}$

and $\bar{y}$ are the mean values. The closer the model prediction is to the ground truth, the closer the r value will be to one.

The structural similarity index (SSIM)[16] was also used to quantify the performance,

$$\mathrm{SSIM} = \frac{(2\bar{x}\bar{y}+c_1)(2\sigma_{xy}+c_2)}{(\bar{x}^2+\bar{y}^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \tag{2}$$

Where σ_x is the variance of x, σ_y is the variance of y, σ_{xy} is the covariance of x and y, and $c_1 = (0.01L)^2$, and $c_2 = (0.03L)^2$ are two constant parameters of SSIM. Note that L represents the range of the intensity values (which is set to 1).

3. RESULTS

This study evaluates the performance of predictions made from TL images using different microscopy data and CNN models. The study compared (i) the time-series predictions of mitochondria and DNA with long- and short-interval acquisitions, (ii) high image resolution (1834×1834 pixels) predictions with low image resolution (512×512 pixels) predictions in both time-series and z-stacks, (iii) the performance of the models trained by Airyscan microscopy images and confocal microscopy images in both time-series and z-stacks, and (iv) the performance between different network structures in both time-series and z-stacks. Fig. 3 summarizes the overall performance metrics for the models mentioned in this paper. The performance metrics were composed of the Pearson correlation coefficient (r) for each subcellular structure, mitochondria, and DNA. The model trained on images acquired using high-resolution confocal microscopy had superior performance than that trained on low-resolution images in both time series ($D_{1,2}$) and z-stack ($H_{1,2}$). Interestingly, the model trained on images using the confocal method predicted mitochondria and DNA structures with higher accuracy. Since the images acquired using confocal microscopy provided high contrast and relatively high resolution without post image processing, such methods ensured a highly relevant relationship between TL input images and fluorescence target images.

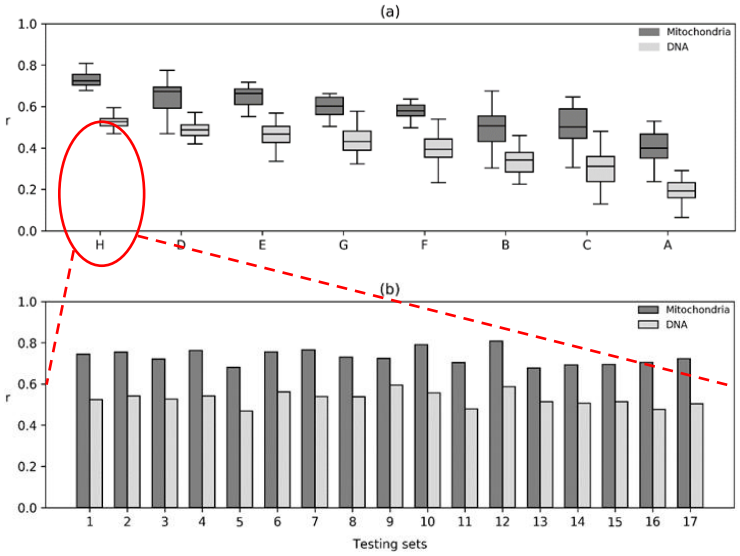

Figure 3. Summary of method performance. (a) Performance comparison between models. (b) Individual testing set performance from the model with best accuracy. The abbreviation of each group is shown in Table 1.

Many factors affected the model performance. The effect of the number of training images (input data) was first tested with 2 to 46 training image sets. When the number of training images increased, the overall accuracy improved until 32 image sets (Fig. 4a), indicating that increasing the dataset has limitations for improvement of the performance of the model. Next, different patch sizes of input data were tested (Fig. 4b). As the patch size increased, the model performance improved. Nevertheless, using a patch size of 128 did not improve the performance. Different network architectures also affect the performance (Table 2), indicating that novated network architectures do not always promise better performance. One should always choose an appropriate architecture that best fits the training data.

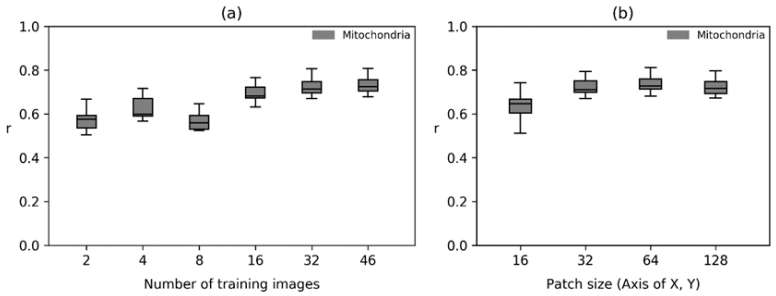

Figure 4. (a) Prediction performance across different numbers of training images. (b) Prediction performance across different patch sizes. All other hyperparameters remained the same. The dataset used in this experiment was the z-stack confocal images with a resolution of 917×917 pixels, as in H_1.

Table 2. The comparison of standard U-net method and GVTNet method in terms of PCC (r) and SSIM.

Model	**Z-stack (H_1)**	**Time-series (D_1)**
	PCC / SSIM	**PCC / SSIM**
U-Net	0.7312 / 0.5730	0.6431 / 0.5545
GVTNet	0.5226 / 0.4028	0.6845 / 0.6068

The time-series ($D_{1,2}$) and z-stack data ($H_{1,2}$) were used together to train a general model (combined model) to see if the combined model could improve the model performance for both time-series and z-stack prediction (Figs. 5 and 6a). The z-stack model and the general model demonstrated similar performance without an obvious difference in predicting ability ($r = 0.7203$ vs. $r = 0.7462$) (Fig. 6c). On the other hand, the difference in predictions between the time-series and general model was larger ($r = 0.6359$ vs. $r = 0.6779$), suggesting that the time-series-specific model is more suitable for the time-series prediction (Fig. 6b).

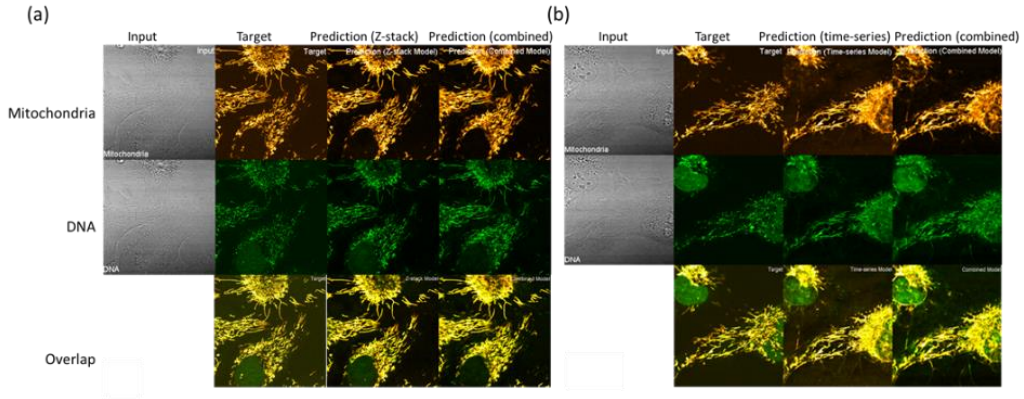

Figure 5. Comparison of mitochondria and DNA prediction from transmitted light images using z-stack and combined models (a), and time-series and combined models (b).

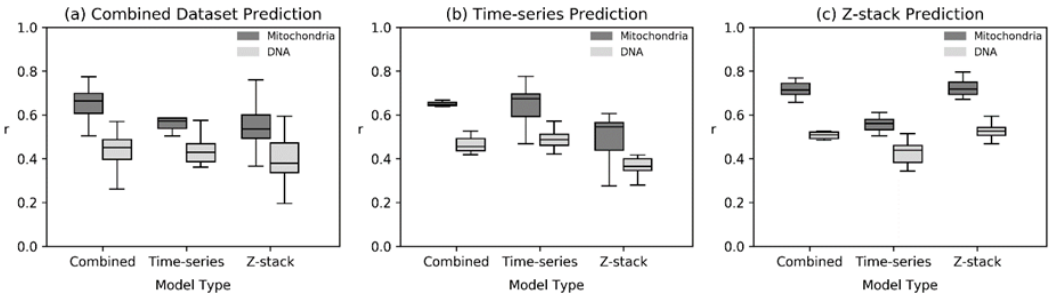

Figure 6. The overall performance of each model across different prediction tasks. (a) Combined dataset prediction performance across different model types. (b) Time-series prediction performance across different model types. (c) Z-stack prediction performance across different model types.

4. DISCUSSION

This research has resulted in the construction of specific models that provide time-series and z-stack predictions for mitochondria and mitochondrial DNA from TL images of AC16 cells without the need for specific fluorescent dye labeling. By optimizing labeling and image acquisition protocols, we obtained fine details of mitochondria images, allowing neural networks to learn relationships between TL and fluorescent labels. This approach enables us to study the morphology and dynamics of mitochondria without facing the problems of photobleaching and phototoxicity, and allows us to increase the duration of imaging and monitoring of mitochondrial structures and dynamics under more physiological conditions.

A compelling question is whether a z-stack-specific model could be used for predicting time-series mitochondrial structures or vice versa. Comparing the performance of testing time-series-specific, z-stack-specific, and combined models from

either time-series or z-stack inputs, a model trained for the specific input type will perform better. For example, the model trained with time-series images achieved better accuracy in predicting time-series images than that trained with the z-stack images. The combined model provided intermediate performance but offered a general purpose model with flexible inputs. Training and testing images sharing the same imaging parameters are recommended for better prediction results.

When acquiring input images in higher resolution, there are more total pixels for the network to train on. To efficiently extract the information from the input, the original U-Net model published by Ounkomol et al.[11] was modified with deeper layers to preserve extra features inside the images. To further improve the performance of high-resolution images, bigger patch size is needed. By adding more convolutional layers, a deeper neural network was created that can extract more features from input data with larger patch size. However, the resulting images were blurry because the background noise was also magnified under high-resolution TL images. Since deep neural network structure is susceptible to blur and noise distortion, such magnified noise will lead to poor performance of the model. It is intuitive that as the size of the signal increases, the size of random fluctuations will also increase, which consequently leads to degradation of model performance. For example, in some cases of the high-resolution imaging experiments, the model itself cannot classify between background noises and labeled targets. Hence, the result may be good as evaluated by the human observation but poor as evaluated by the Pearson correlation coefficient.

Compared to a previous work done by Ounkomol et al.[11], our z-stack prediction focused on mitochondria and DNA at higher magnification, and provided more fine details of mitochondria structure. It is noted that under such magnification, noise and particles will affect performance significantly, resulting in larger statistical error. The limitations existing in our presented methodology notwithstanding, the present study suggests an alternative approach toward broader biological imaging areas where it may present an opportunity for a breakthrough.

ACKNOWLEDGMENT

This work was supported by the Ministry of Science and Technology of the Republic of China in Taiwan (MOST) (MOST-109-2636-B-002-001). We thank Venessa Yu from Zeiss for the technical support in microscopy imaging.

REFERENCES

[1] Galloway, C. A., Lee, H.., Yoon, Y., “Mitochondrial morphology-emerging role in bioenergetics.,” Free Radic. Biol. Med. **53**(12), 2218–2228 (2012).

[2] Mishra, P.., Chan, D. C., “Metabolic regulation of mitochondrial dynamics.,” J. Cell Biol. **212**(4), 379–387 (2016).

[3] Bulthuis, E. P., Adjobo-Hermans, M. J. W., Willems, P. H. G. M.., Koopman, W. J. H., “Mitochondrial morphofunction in mammalian cells.,” Antioxid. Redox Signal. **30**(18), 2066–2109 (2019).

[4] Lewis, S. C., Uchiyama, L. F.., Nunnari, J., “ER-mitochondria contacts couple mtDNA synthesis with

mitochondrial division in human cells.," Science **353**(6296), aaf5549 (2016).

[5] Opstad, I. S., Wolfson, D. L., Øie, C. I.., Ahluwalia, B. S., "Multi-color imaging of sub-mitochondrial structures in living cells using structured illumination microscopy," Nanophotonics **7**(5), 935–947 (2018).

[6] Liu, X., Yang, L., Long, Q., Weaver, D.., Hajnóczky, G., "Choosing proper fluorescent dyes, proteins, and imaging techniques to study mitochondrial dynamics in mammalian cells.," Biophys. Rep. **3**(4), 64–72 (2017).

[7] Skylaki, S., Hilsenbeck, O.., Schroeder, T., "Challenges in long-term imaging and quantification of single-cell dynamics.," Nat. Biotechnol. **34**(11), 1137–1144 (2016).

[8] Selinummi, J., Ruusuvuori, P., Podolsky, I., Ozinsky, A., Gold, E., Yli-Harja, O., Aderem, A.., Shmulevich, I., "Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images.," PLoS One **4**(10), e7497 (2009).

[9] Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M.., Van Valen, D., "Deep learning for cellular image analysis.," Nat. Methods **16**(12), 1233–1246 (2019).

[10] Vicar, T., Balvan, J., Jaros, J., Jug, F., Kolar, R., Masarik, M.., Gumulec, J., "Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison.," BMC Bioinformatics **20**(1), 360 (2019).

[11] Ounkomol, C., Seshamani, S., Maleckar, M. M., Collman, F.., Johnson, G. R., "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy.," Nat. Methods **15**(11), 917–920 (2018).

[12] Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'Neil, A., Shah, K., et al., "In silico labeling: predicting fluorescent labels in unlabeled images.," Cell **173**(3), 792–803.e19 (2018).

[13] Ronneberger, O., Fischer, P.., Brox, T., "U-Net: Convolutional Networks for Biomedical Image Segmentation," [Medical Image Computing and Computer-Assisted Intervention (MICCAI)], N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, eds., Springer International Publishing, Cham, 234–241 (2015).

[14] Huff, J., "The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution," Nat. Methods **12**(12), i–ii (2015).

[15] Jevtic, V., Kindle, P.., Avilov, S. V., "SYBR Gold dye enables preferential labelling of mitochondrial nucleoids and their time-lapse imaging by structured illumination microscopy.," PLoS One **13**(9), e0203956 (2018).

[16] Wang, Z., Xie, Y., Ji, S., "Global Voxel Transformer Networks for Augmented Microscopy." arXiv:2008.02340 (2020)