Research Article Vol. 13, No. 10 / 1 Oct 2022 / Biomedical Optics Express 5495
Biomedical Optics EXPRESS

Virtual labeling of mitochondria in living cells using correlative imaging and physics-guided deep learning

AYUSH SOMANI,[1,*] ARIF AHMED SEKH,[2] IDA S. OPSTAD,[3] ÅSA BIRNA BIRGISDOTTIR,[4] TRULS MYRMEL,[4] BALPREET SINGH AHLUWALIA,[3] ALEXANDER HORSCH,[1] KRISHNA AGARWAL,[3] AND DILIP K. PRASAD[1]

[1]*Bio-AI Lab, Department of Computer Science, UiT The Arctic University of Norway, Tromsø, 9037, Norway*
[2]*Computer Science and Engineering, XIM University, Bhubaneswar, 751002, India*
[3]*Department of Physics and Technology, UiT The Arctic University of Norway, Tromsø, 9037, Norway*
[4]*Cardiovascular group, Department of Clinical Medicine, UiT The Arctic University of Norway, Tromsø, 9037, Norway*
[*]*ayush.somani@uit.no*

Abstract: Mitochondria play a crucial role in cellular metabolism. This paper presents a novel method to visualize mitochondria in living cells without the use of fluorescent markers. We propose a physics-guided deep learning approach for obtaining virtually labeled micrographs of mitochondria from bright-field images. We integrate a microscope's point spread function in the learning of an adversarial neural network for improving virtual labeling. We show results (average Pearson correlation 0.86) significantly better than what was achieved by state-of-the-art (0.71) for virtual labeling of mitochondria. We also provide new insights into the virtual labeling problem and suggest additional metrics for quality assessment. The results show that our virtual labeling approach is a powerful way of segmenting and tracking individual mitochondria in bright-field images, results previously achievable only for fluorescently labeled mitochondria.

1. Introduction

Tracking movements and analyzing the activity of sub-cellular organelles like mitochondria create new opportunities for biological studies [1]. On one hand, to support such studies, computer vision based object tracking approaches [2,3] need to be adapted to relevant bio-image data. On the other hand, the bio-image data must fulfil certain requirements to be appropriate for the particular computer vision task at hand: (a) ability to exclusively capture the structures of interest with sufficient space and time resolution, (b) ability to record videos over a longer time duration, and (c) accomplishing a and b without significantly perturbing the sub-cellular system.

Label-free linear optical microscopy solutions, such as differential interference contrast, bright-field, and phase contrast microscopy, use the inherent optical contrast of the sub-cellular organelles relative to the cytosol present in cells to perform imaging [4]. Therefore, no dyes are needed to label the organelles, and both photo-bleaching and perturbation problems associated with labeling are avoided.

As we describe next, there is only one problem with label-free imaging that restricts its widespread use for organelle-specific sub-cellular studies. In label-free microscopy, everything with an optical contrast gets imaged simultaneously through light scattering, absorption, and interference phenomena. Therefore, it is difficult to either visually or computationally distinguish and isolate organelles of a specific type from the rest of the cell. This problem is illustrated in Fig. 1, where label-free (bright-field) with their corresponding fluorescence microscopy images

#464177 https://doi.org/10.1364/BOE.464177
Journal © 2022 Received 17 May 2022; revised 24 Aug 2022; accepted 7 Sep 2022; published 28 Sep 2022

Biomedical Optics EXPRESS

are provided. The composite images show that most structures visible in bright-field are not mitochondria, and that it is hard to identify which ones are.

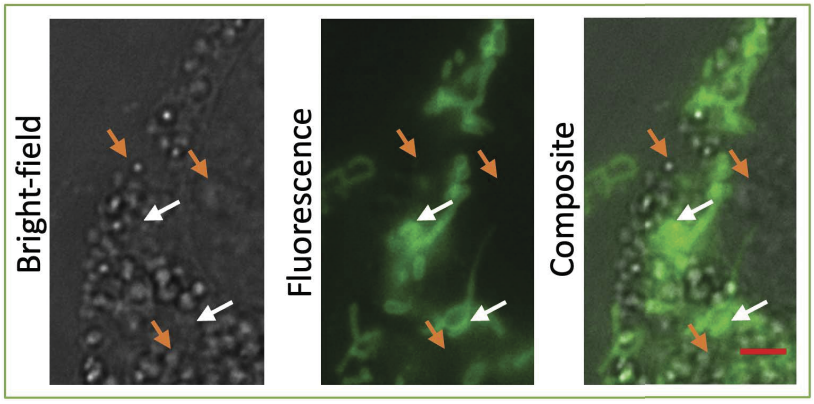

Fig. 1. Example of a bright-field image with the corresponding fluorescence image. The bright-field and fluorescence images are shown overlayed in the third panel (scale bar 2.5 μm). Note that the majority of structures visible in bright-field marked by the orange arrows are not mitochondria. Features indicated by the white arrows are mitochondria, but hardly visible in the bright-field image.

Here, we propose a method for tracking mitochondria in living cells imaged using (label-free) bright-field microscopy. We are unaware of previous works where mitochondria-specific tracking has been accomplished using solely label-free imaging data. Our approach consists of performing virtual fluorescent labeling of mitochondria, segmenting the virtually labeled mitochondria, and tracking the resulting segmentations. While we have used established techniques for segmentation [5] and tracking [3], our novel technical contribution lies in the task of performing virtual labeling. We consider this task as an image-to-image (I2I) translation problem that is exceptionally difficult due to the noise present in fluorescence image data. Therefore, we propose a physics-guided deep learning approach that incorporates the physics of the microscope into a custom-designed novel loss function such that the deep learning network automatically learns to deprioritize the background noise. This elegant and simple solution helps us improve the quality of virtual labeling beyond the state-of-the-art (SOTA) performance. Below, we present background work and our contributions is some detail.

Virtual labeling of label-free microscopy images using artificial intelligence (AI) is a recent and rising research endeavor. A correlative microscopy dataset comprising registered pairs of label-free and fluorescence microscopy images is used as a training dataset for a deep learning model. Labels can be histopathological dyes, fluorescence labels, or other cellular markers. Once trained, the model converts a previously unseen label-free image to a corresponding labeled image.

Bayramoglu et al. [6] used GAN for the labeling of lung histology images. Christiansen et al. [7] used a convolutional neural network (CNN) based transfer learning approach and applied it for images of human motor neurons, primary rat cortical cultures, and human breast cancer cell-line. Rivenson et al. [8] used upsampling and downsampling based CNN for virtual labeling of kidney, liver, and gland tissue. PhaseStain [4] uses CNN to transform the quantitative phase images of label-free tissue sections (kidney, liver, and skin tissue) into virtual pseudo-color images. An unsupervised learning approach is proposed in [9] and applied to immunohistochemistry images. Liu et al. [10] proposed a multi-scale input strategy for virtual labeling of cell nuclei in different cells and conditions. Cell nuclei, cell morphology, and golgi bodies have been addressed also by Cheng et al. [11], and Kandel et al. [12] using low numerical

Biomedical Optics EXPRESS

aperture microscopes targeting relative large structures inside cells. Kandel et al. [12] used quantitative phase image as the label-free modality, which is known to be closely related to the 3D morphology and optical contrast of the structures being imaged. 3D quantitative phase and/or refractive index maps were also considered for virtual labeling [13] and organelle segmentation [14]. Bright-field images on the other hand are related to these properties of the sample in significantly non-linear manner and therefore pose significant challenges. Nonetheless, ubiquity of bright-field microscopes attracts the users to consider virtual labeling of bright-field images. Three-dimensional fluorescence images from transmission light microscopy are used in [15]. The authors employed a multi-model architecture for predicting individual channels and validated their method on, e.g., nuclei, micro-tubules, and mitochondria.

The majority of the virtual labeling is performed for histopathology, cell nuclei, and cell-level labeling. Virtual labeling of nanoscale sub-cellular structures is complex and has received less attention, where Ounkomol et al. [15] and Guo et al. [16] are prominent contributions. Notably, deep learning solutions have been used in all the above works in a plug-and-play manner, with almost no customization of architecture or learning for the specific application. It is also notable that several SOTA works [7,13,15] use 3D or z-stacks of label-free and fluorescence images for learning the mapping. While Christiansen et al. [7] explain that the need of z-stack is related to the lack of information in single plane images, we differ in our opinion about the reason. In the conventional widefield fluorescence and label-free images, including from epifluorescence microscopes and label-free bright-field, transmitted light microscopes, etc, each imaging plane of course images structures within the plane as the primary foreground, but it also includes out-of-focus light leaking from structures in the other planes and the noise from various sources which severely afflict the background region. In the sense of out-of-focus light, we believe that the problem in using single planes for virtual labeling is not the lack of 3D information or 3D context, but the over-representation of the 3D information even in single planes which is quite difficult to isolate and/or map correctly. In the conventional plug-and-play approach, most of the efforts are used for learning the noisy background and out-of-focus light which present insignificant value for interpretation. Using the 3D data or the z-stack provides some overlapping 3D context for where the out-of-focus light in each plane is coming from and therefore helps in better learning. However, there are several disadvantages in using this approach. The foremost is the practicality of acquiring large datasets of well-registered 3D correlative image pairs and the need of extensive instrument calibration and imaging protocol optimization in order to achieve this target. In most cases, this requires instrument upgrades and demand additional costs and time. Further, even if in practice the users may be interested in single plane acquisitions only, they will be forced to take 3D images simply because the model was not designed to tackle out-of-focus light. This contributes issues in live cell imaging of moving entities and somewhat trumps the motivation of using label-free for being the mode with less light-dose. Here, we propose a physics-guided modification of the selected I2I approach that guides the deep learning model to focus on learning the foreground and deprioritize learning the noisy and out-of-focus background characteristics.

1.1. Performance evaluation of virtual labeling

Evaluation of the performance of virtual labeling using quantitative metrics is also an exciting aspect of the problem. The majority of the methods use Pearson correlation between the virtual and fluorescence images [9,15] as a preferred metric. A few of them use an intensity confusion matrix for the performance evaluation [7]. Other conventional image comparison metrics used in I2I problems [17], such as structural similarity index metric (SSIM), peak signal-to-noise ratio (PSNR), and intensity histogram comparison metrics such as Kullback–Leibler (KL) divergence, have not been used. We hypothesize that the reason is that the Pearson correlation coefficient and intensity confusion matrix are not sensitive to the actual intensity values. Therefore, structural differences, pixel-wise differences, and differences in the intensity histograms that arise from

Biomedical Optics EXPRESS

learning both the foreground and the noisy background pixels with equal weightage do not significantly affect comparisons based on these metrics. Nonetheless, different from previous approaches, we opted for the use of SSIM, PSNR, and KL divergence in addition to the Pearson correlation coefficient and intensity confusion matrix to evaluate the performance of virtual labeling. Our method shows consistency across all these wide varieties of performance metrics.

1.2. Beyond virtual labeling

In addition to virtual labeling, we performed segmentation and tracking of mitochondria in label-free bright-field images, with only very few precedents of such work on label-free data. Here, we briefly cover some relevant works in segmentation and tracking.

Segmentation of cells and cell-organelles has a long history in computer vision and biological research [18]. Morphological studies of subcellular structures like mitochondria attract many researchers [5,19,20]. While fluorescence images are popularly used for segmentation in the majority of cases [21,22], label-free segmentation is not intensely explored. Although label-free analysis of mitochondrial dynamics is proposed in Ref. [23,24], mitochondria segmentation and their stitching over time are not explicitly used to understand the dynamics at the level of individual mitochondria. Instead, a metric such as mean aspect ratio or connectedness is generated for each image containing several mitochondria, and the change in this metric over time is monitored. An underlying reason for the lack of individual mitochondrion segmentation and motion analysis in label-free data is that label-free segmentation is different and complex due to the visually suppressed nature of structure and lack of contrast [25]. Furthermore, the fact that it is challenging to isolate mitochondria in label-free images, such as those shown in Fig. 1, causes difficulties in generating segmented ground truth for training. Here, we propose to use virtually labeled images for segmenting label-free images. This also allows for using fluorescence images for training.

The body of work on cell tracking is quite large [26–29] and even includes label-free imaging [28,30]. However, tracking of sub-cellular, highly mobile structures is encountered quite rarely [3,5,31]. Sub-cellular structures such as mitochondria are considered hard to track [1], and recently some advanced image processing tools have been developed for the purpose of specifically analyzing mitochondria and its dynamics. Examples include Mitometer [32], MitoGraph [33], and MiNA [34], all of which use high resolution fluorescence images as the input data. Tracking such sub-cellular structures in label-free images poses an even more difficult problem. However, having virtually labeled the bright-field images, the SOTA segmentation and tracking methods in fluorescence microscopy can be directly employed, as we present in this work. However, the high quality and reliability of virtual labeling are imperative for the success of such advanced analysis. In this sense, the ability of the proposed physics-guided loss function to improve the performance of virtual labeling across all contemporary metrics provides confidence in taking such advanced analytical approaches designed originally to work with fluorescence only data.

1.3. Contributions of our work

Here, we present the contributions of our work explicitly for the convenience of readers.

Physics-guided loss function: As discussed above, fluorescence microscopy images have only a fraction of the pixels in the foreground. The remaining background pixels have non-zero pixel-wise independent intensity distribution arising from different noise sources during image acquisition. The conventional loss functions treat both the foreground and the background pixels equally. This makes the learning biased towards background pixels and creates a one-to-one match of the background pixel intensity. As the background intensity distribution is stochastic, it is impossible to derive a one-to-one match.

Our elegant physics-guided modification of the conventional loss function encourages the model to focus on learning the foreground structures in the labeled images while reducing the

Biomedical Optics EXPRESS

learning load to match the intensity at pixels in the noise-ridden background of microscopy images. Our loss function uses the microscope's fluorescent photon-to-pixel intensity mapping, called the point spread function (PSF), to design such characteristics inherently and in a soft manner, without resorting to the binary treatment of foreground and background pixels. This does not only improve the performance of learning but also eliminates the possibility of artifacts arising from binary segmentation of the foreground, which itself is usually not perfectly available [5]. Furthermore, our loss function is usable in most I2I translation networks and other similar deep learning approaches that involve fluorescence data.

Extensive performance analysis using various metrics: We present an extensive analysis of the virtual labeling performance using different metrics, such as Kullback–Leibler (KL) divergence based similarity, Pearson correlation, structural similarity, signal-to-noise ratio, and intensity confusion matrix. These provide a comprehensive array for analyzing the performance of virtual labeling until improved metrics are explicitly developed for this purpose. This is different from the conventional use of only the Pearson correlation coefficient or normalized intensity confusion matrix for virtual labeling.

Label-free segmentation and tracking: We extend the applicability of virtually labeled image sequences in segmentation and tracking for potential application task in quantifying mitochondrial dynamics. We essentially demonstrate that high-quality, virtually labeled data is directly amenable to post-processing pipelines originally used for fluorescence microscopy data.

2. Materials and methods

2.1. Cell culture and sample preparation

The rat cardiomyoblast cell-line H9c2 (cells derived from embryonic heart tissue; Sigma Aldrich) was genetically modified using a retrovirus to achieve a stable expression of tandem tagged (mCherry-EGFP) mitochondrial outer membrane protein 25 (OMP25)-transmembrane domain (TM). A uniform expression of fluorescence intensity in the cells was achieved through flow cytometry sorting. The stable H9c2 cells were cultured in high glucose (4.5 g/L) Dulbecco's Modified Eagle Medium (DMEM; [D5796, Sigma-Aldrich]) with 10% FBS, 1% streptomycin/penicillin and 1 µg/mL of puromycin (InvivoGen). For glucose deprivation and adaptation to galactose, the cells were grown in DMEM without glucose (11966-025, Gibco) supplemented with 2 mM l-glutamine, 1 mM sodium pyruvate, 10 mM galactose, 10% FBS, 1% streptomycin/penicillin and 1 µg/mL of puromycin (InvivoGen). The cells were adapted to galactose for a minimum of 7 days before experiments. The cells were seeded on MatTek dishes (P35G-1.5-14-C, MatTek Corporation) and imaged when they reached approximately 80% confluency.

2.2. Data acquisition

During image acquisition, the cells were kept at 37°C, 5% CO2, atmospheric oxygen, and a cell-culture medium (DMEM with 10% FBS). The time-lapse microscopy data were acquired using a DeltaVision OMX V4 Blaze imaging system (GE Healthcare Life Sciences, Marlborough, MA, USA) equipped with a 60X 1.42NA oil-immersion objective (Olympus) and three sCMOS cameras for rapid multi-channel imaging. For each cell, a correlated pair of images in bright-field and fluorescence modes were acquired.

The imaging mode was sequential, with acquisition order "All Z then Channel", meaning that different focal planes of one channel were acquired before imaging the same z-planes for the next channel/imaging mode. For the bright-field images, the DAPI channel (blue light) was used. For fluorescence, both FITC (green light) and A568 (red light) was used to capture the two different mitochondrial markers (from the tandem-tag acquisition, inferences of mitochondrial degradation can be made, although not relevant in to work). A total span along the z-axis of 8 µm were used, sampling every 125 nm (nanometer, giving a total of 65 sections for each channel).

Biomedical Optics EXPRESS

For each plane, the exposure times for both the bright-field and fluorescence acquisition were 10 ms. Each image has the size of 1024 × 1024 pixels, where each pixel corresponds to 80 nm in the cell sample. The different cameras and imaging modes were aligned using the imaging system's pre-calibration. From this, we used in total 3381 correlated pairs of images for training, 966 for validation, and 483 correlated pairs of images for testing.

2.3. Complete architecture for learning components

Our processing pipeline is depicted in Fig. 2(A). The first component is learning the I2I model for virtual labeling which translates the bright-field image to the corresponding fluorescence image. The second component applies the learned model to live-cell bright-field images/videos and creates equivalent fluorescence images/videos. The third component performs segmentation and tracking. This component may be another necessary image analysis or advanced processing task, or completely absent; for our problem tracking has been identified as relevant.

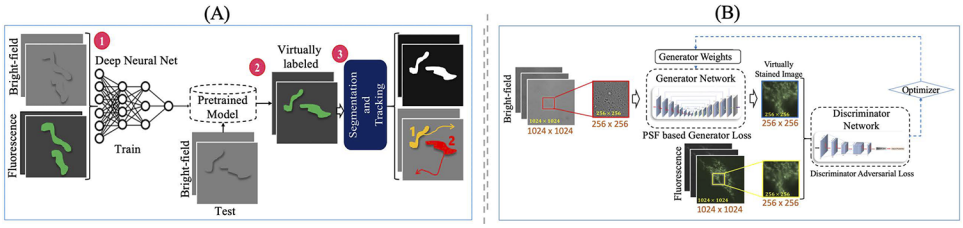

Fig. 2. (A) Outline of the proposed method. The red coloured numbers represent the principal modules and their sequence, and (B) conditional GAN of the employed architecture. The details of the architecture are presented in section 2.3.

Considerations about training datasets for virtual labeling: A microscope capable of correlative imaging of a sample in bright-field and fluorescence mode is used for collecting image pairs for supervised learning. Registration using the microscope's calibration software or fiducial markers may be needed. However, it is essential to ensure sufficient data diversity, and therefore, images acquired in different ROIs are preferred over videos. The input data comprises images in different planes. Our objective is to perform 2D I2I, i.e. mapping bright-field image of a single plane to the corresponding fluorescence image of the same plane without using other planes above and below it. Therefore, there are some planes with significant out-of-focus light and some other planes which are largely-in-focus and have relatively less out-of-focus light. When forming the training dataset, we incorporated both largely-in-focus planes and the other planes in a proportion for robust model training. Our empirical observation is that the inclusion of a small number of out-of-focus images for training gives better results and helps in avoiding over-training.

Each correlated image pair of 1024 × 1024 pixels was randomly cropped into 25 smaller patches (256 × 256 pixels). Random image cropping prevents the monotony of correlated image frames in model training. To help the model converge better experimentally, patches consisting mainly of background noise (i.e., the patches with less than 20% detected mitochondria) were discarded from the learning. The obtained 4830 training image pairs were proportionally split into 70-20-10% training, validation, and test sets. The validation and test result for other datasets is done similarly and mentioned in result section.

I2I network for virtual labeling: We have used the conditional GAN model, which comprises a generator network and a discriminator network working in tandem, as shown in Fig. 2(B). A variety of architectures can be considered to design generator and discriminator networks. We have used U-Net as the backbone, employed in a Pix2Pix architecture [35]. The flexible and adaptable nature of the models to a wide variety of tasks without explicitly defining the relationship is a major motivation for choosing this model.

Biomedical Optics EXPRESS

The architecture and components of the generator network we used are described here. It has four downsampling convolution blocks in the following order: C32-C64-C128-C256. Here, C32 indicates a downsampling convolution block of 32 channels, and analogously for the remaining blocks. They are followed by a 4 upsampling blocks and skip connections across layers of the same level. Each residual block in each downsampling path consists of three convolution layers followed by three leaky rectified linear units (leaky ReLU) employed as an activation function. The blocks are connected by an average-pooling layer of stride 2 that downsamples the previous block's output by a factor of two (except for the initial block that increases from 1 input channel to 64 channels). The blocks in the upsampling layer are linked by the bilinear upsampling layer, which increases the size of the output by a factor of two in both lateral dimensions. To raise the number of channels from the previous block's output by two, a concatenation function, i.e., a skip network with the appropriate feature map from the same level's downsampling path, is utilised.

In the U-Net architecture, the modules have 3 types of configurations, namely in-scale, down-scale, and up-scale. In the in-scale configuration, the convolution kernel size $k = 7$ and stride $s = 1$ has been used. The downscale configuration has $k = 3$ and $s = 2$. In the up-scale configuration, $k = 3$, and $s = 2$ have been used and followed by max-pooling.

We experimented with different hyper-parameter combinations, as reported in Table 1. However, in each case we used a learning rate $\alpha = 2 \times 10^{-4}$ and exponential decay $\beta = 0.5$. Pool size is kept at zero. The convolution is not zero-padded rather uses reflection padding, thus "reflects" the row into the padding. This is valuable because it assures that the outputs transitions "smoothly" into the padding and the padded inputs are similar to the original data distribution [36]. Training was performed from trained from scratch for 400 epochs.

Table 1. Hyper-parameters used for training the proposed architecture for virtual labeling (– indicates that the parameter is the same as used in the method A). Learning rate α is set as 2×10^{-4} and exponential decay parameters β is set as 0.5.

Hyper-parameter	Method							
	A	B	C	D	E	F	G	H
Input channels	3	1	–	–	–	–	–	–
Batch size	1	–	16	16	–	16	–	–
Dropout	N.A.	0.5	0.5	0.5	–.	0.5	–	–
Optimizer	Adam	–	RMSProp	–	–	–	–	–
Generator loss	PSF	–	–	–	L1	L1	TN	Minimax
Activation	ReLU	–	Sigmoid	–	–	–	–	–

The architecture and components of the discriminator network we used are described here. The discriminator network comprises of 1 convolutional layer, 5 discriminator blocks, an average-pooling layer, and two fully connected layers. The first convolution layer receives the input which is either the virtually labeled image generated by the generator network or the target fluorescence image in the supervised data pair. This layer increases the number of channels to 64. The discriminator blocks consist of 2 convolutional layers each with the first layer maintaining the number of channels and the size of the feature map. In contrast, the second layer increases the number of channels by twofold and decreases the feature map's size by fourfold. The average-pooling layer has a filter size of 8×8 and results in a matrix with a size of $B \times 2048$, where B refers to the batch size. The output is then fed into 2 fully connected layers with the prior layer preserving the size of the feature map, and the later layer reducing the output channel to 1. This results in an output size of $(B, 1)$. The fully connected layer's output goes through a sigmoid function, indicating the probability that the channel discriminator input is a fluorescence image. All the convolutional layers and the fully connected layers are connected by leaky ReLU but with nonlinear activation functions for the discriminator network. The learnable parameters,

Biomedical Optics EXPRESS

including filters, weights, and biases in the convolutional layers and the fully connected layers in the discriminator network are updated using an adaptive moment estimation (Adam) optimizer with a learning rate of 2×10^{-5}.

The implementation details are presented here. The conditional GAN model was implemented using standard libraries and scripts that are publicly available, namely, Python (version 3.7.3), Pytorch (version 1.6.0) and Tensorflow (version 1.14.0). The other python libraries incorporated were Keras, OpenCV, scikit-learn, os, SciPy, NumPy, Matplotlib, time, the Python imaging library (PIL), visdom, dominate, tqdm. The model's training was majorly carried on a remote desktop computer with a Xenon Gold 5218 CPU at 2.30 GHz (Intel) and 768 GB RAM, running a Windows Server 2019 standard operating system. The network training, visualization, and testing were performed on 96 GB Nvidia Quadro RTX 8000 GPU.

2.4. Segmenting and tracking

Tracking is performed by the process of tracing individual mitochondria over time by preserving the identity. The process involves detection, identity assignment, and linking identity over time (see Fig. 3).

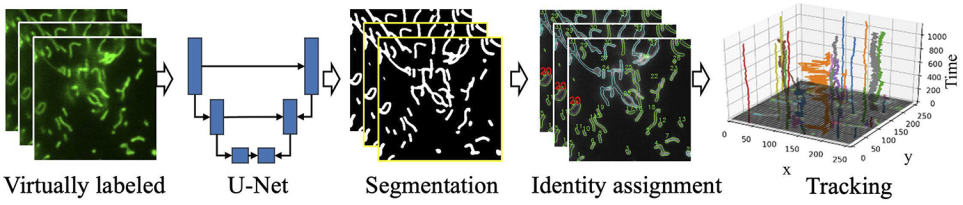

Fig. 3. The tracking module uses a U-Net based supervised architecture [37] to segment the mitochondria and a Kalman filter based tracker to estimate their motion.

The first step of detection is based on the segmentation of mitochondria. While the bright-field images are not suitable for segmentation directly, their virtually labeled versions emulate fluorescence images and are therefore amenable to segmentation. First, the U-Net network [37] is trained on a set of manually segmented images of fluorescently labeled mitochondria. Next, the virtually labeled images are used as the input for the segmentation.

After the segmentation step, a blob-based detection [38] is adopted to detect individuals or groups of mitochondria in each frame. The detection is represented by bounding boxes. Next, the bounding boxes are linked by preserving the identity (track number) in the temporal domain. A linear motion model is fit for each track. Equation (1) represents the state of a target, i.e. a bounding box during tracking:

$$x = [u, v, s, r, \dot{u}, \dot{v}, \dot{s}]^T, \tag{1}$$

where u and v are horizontal and vertical velocities, s and r represent the scale and the aspect ratio of the bounding box, respectively. The dot quantities represent the values after δt time instance. The target's state is updated using the initial detection. Kalman filter [39] is utilized to solve the velocity components. A linear motion model prediction is employed to update the missing frames. The Hungarian algorithm solves the track assignment problem among the Kalman filter prediction and newly arrived detection. We used the (squared) Mahalanobis distance to estimate the distance and optimum association of tracks. Finally, continuously detected bounding boxes with the preserved identity are included as the frames of a mitochondrion track.

3. Physics-guided loss function for virtual labeling

To the best of our knowledge, deep learning architectures and training approaches are often used in unmodified form, without adapting them for the specific problem at hand. This implies that if

Biomedical Optics EXPRESS

there is a well-posed mapping between the input and output spaces, the deficiency in learned models is mainly attributed to the architecture and learning process. Here, we assume that the data quality and span of the input and output spaces are not a problem. We opine that there is an opportunity for improvement if the process of learning can benefit from the application-specific a priori knowledge. For microscopy, this knowledge includes the optical PSF. Further, an important a priori information for fluorescence imaging is that fluorescence images consist of a large fraction of background pixels where mainly noise is measured and no structural details are to learn.

Choice or design of the generator loss function has a strong bearing on the performance of the model. Examples of conventional loss functions for the generator include the minmax [40] and the L1 norm based loss function [41]. Notably, these loss functions have successfully and routinely been used for consumer cameras and medical images where the noise is very low. However, we noted that their performance is poorer for fluorescence images of sub-cellular structures.

Because of the labeling of specific (and often sparse) structures, fluorescence microscopy images are largely empty with the fraction of foreground pixels being very small. In such a situation, the minmax and L1 loss tend to focus on learning background noise rather than the foreground structures since they consistently contribute to a mismatch in the background pixels due to the nature of noise. For example, Fig. 4(A) shows an actual fluorescence image and Fig. 4(B) a virtual candidate image created by the generator. Figure 4(C) shows the conventional L1 loss at each pixel. We observed that the sum of background loss can be higher compared to the foreground object due to the noisy background. If the generator minimizes the L1 loss, learning the noisy background is prioritized.

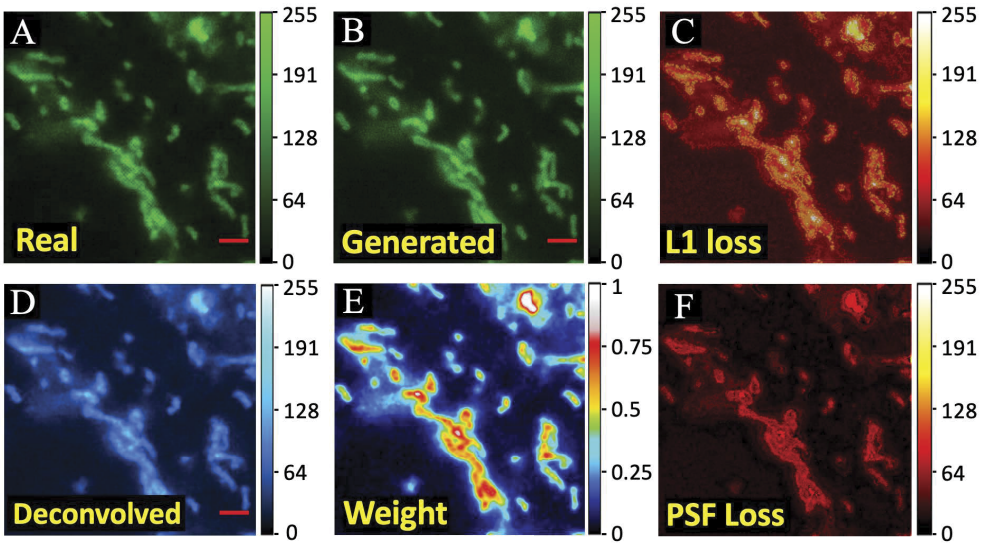

Fig. 4. Illustration of the proposed PSF-based loss function. (A) An example input data, (B) a candidate image generated by the generator during training, (C) pixel-by-pixel L1 loss map, (D) deconvolved image, (E) weight of each pixel for computing PSF-based loss function, and (F) PSF-based loss function map. It is noted that the proposed PSF-based loss function map under weighs the background pixels (scale bar 2.5 μm).

We propose to use a PSF-based loss function that is effective in mitigating the above-mentioned problem. The deconvolution of fluorescence microscopy images with the microscope PSF can be used to enhance the image contrast [42]. We computed a deconvolution kernel (square kernel of

Biomedical Optics EXPRESS

the size of approximately 1 PSF region) generated using the 2D Gibson-Lanni [43] PSF model. This means that we computed the point spread function using the Gibson-Lanni model over a region of 5×5 pixels and supplied it as the approximate PSF to the deconvolution function which internally performed its inversion to compute the deconvolution kernel. The deconvolution kernel is computed using the Lucy-Richardson algorithm [44]. Then, the deconvolution kernel is convolved over the fluorescence image to obtain a high contrast deconvolved image, see Fig. 4(D). This deconvolved image is normalized by the maximum value as shown in Fig. 4(E) to obtain a weight matrix which is multiplied by the L1 loss map (such as shown in Fig. 4(C) to obtain the proposed PSF based loss function map, shown in Fig. 4(F). In comparison to the original L1 loss map, this new PSF-based L1 loss map has small values in the background regions and emphasizes the foreground regions with a large mismatch. Since the value of the PSF-based loss is not strictly zero in the background, the learning of the low-intensity features, such as contributed by out-of-focus light, are still learned albeit with lesser priority. We highlight here that the target image is still the original widefield fluorescence image and the deconvolved image is being used to compute only the weight map in the loss function. The function of this intermediate weight map is that it should guide the model to prioritize what is physically important to learn (the structures largely-in-focus in the plane) and deprioritize what is physically either too random to learn (like noise) or too difficult and uncontextual to learn (out-of-focus light). The deconvolution does not improve the quality of target image since the target image itself remains unprocessed, it rather improves the ability of model in learning how to mimic the target image for the portion that is important to mimic.

The PSF-based loss is defined in Eq. (3), where y_o is the original pixel and y_g is the generated pixel, computed for all n pixels in the input image. Figure 4(F) shows the PSF loss computed for the real and generated image. Our observations confirmed that the proposed loss function successfully suppressed the background noise and the loss was computed only for the foreground objects.

$$I_{weight} = |I_f \otimes^{-1} I_{psf}|_{[0,1]}, \tag{2}$$

$$L_{L1} = \sum_{i=1}^{n} |I^i_{weight}(y_o - y_g)|. \tag{3}$$

4. Results

We first present extensive results on the core technical scope of the article, namely virtual labeling. Then, we provide example results for segmentation and tracking of mitochondria using bright-field images through virtual labeling.

4.1. Virtual labeling

Hyper-parameters of I2I: We have used six different sets of parameters (sets A to H) for the ablation study and found a suitable set of parameters for the problem. The parameters are summarized in Table 1. The networks for all the combinations were trained from scratch for 200 epochs, and the results were compared using SSIM and Pearson correlation (see Fig. 5). The sets A-D (green color in Table 1 and Fig. 5)) use the proposed PSF-based loss function while the sets E-F (red color) use the L1 loss function. The set G in yellow represents the thresholding-normalized (TN) loss defined in 2020's paper by Somani et al. [45]. The set H used the modified minmax loss defined in [46].

As can be observed from Fig. 5, the PSF-based loss functions (A-D) perform better than all other loss functions tested (E-H). Among these, set A performs the best. We have fine-tuned convolution blocks by varying hyper-parameters to improve the prediction accuracy. We found the best combination to be a filter size of 32, optimizer Adam, $\beta_1 = 0.5$, learning rate = 0.0002, and a batch size of 1. The pool size for all the network variants is kept at zero. The convolution

Biomedical Optics EXPRESS

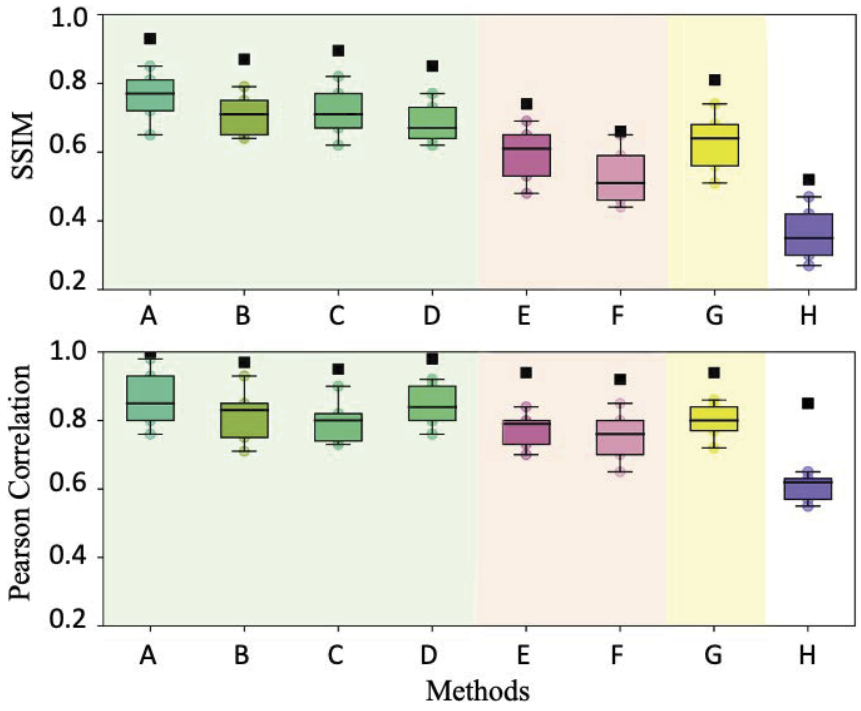

Fig. 5. Comparison of the performance of the methods (variety of learning setups) listed in Table 1 in terms of (a) SSIM and (b) Pearson correlation. The data consists of 10^{th}, 25^{th}, 50^{th}, 75^{th}, and 95^{th} quartile. The black squares represent the maximum values. The methods that use PSF-based loss functions (A-D) outperform the rest.

is not zero-padded but uses reflection padding, thus “reflects" the row into the padding. This assures the output to transition “smoothly" into the padding [36]. In the following, we further analyze the results obtained by this optimally curated and trained network.

Comparison of different weight maps: We explored and compared different candidates for weight maps. For performing these experiments, we used the same setting as the method A in Table 1, but used different candidates for weight maps. The results are shown in Fig. 6, where the last candidate (PSF deconvolution) is the proposed solution as is the same as method A in Fig. 5 and used throughout the manuscript unless stated otherwise.We have broken the argument of ‘physics-guided learning by prioritizing foreground’ through weighted L1 loss into following components:

1. If the foreground needs to be prioritized, one may simply use conventional binary segmentation approaches to obtain a binary map that is used as a weight image. This does not require any physics-guided learning except the knowledge that foreground needs to be prioritized. However, it is easily appreciated that using a binary mask will lead to artefacts at the boundaries of binary segmented regions and therefore it is not advisable.

2. From the same logic, it is clear that the foreground and background need soft treatment in the fluorescence microscopy images. This is further understood from the fact that point spread function blurs the boundaries between the foreground and background in fluorescence image and this character needs to be retained in virtual labeling. Then, the physics is incorporated through not just prioritizing foreground learning but also soft treatment of foreground and background. However, this does not mandate the use of PSF.

Biomedical Optics EXPRESS

So, we consider two different approaches, where the weight maps are computed using logit-based pixel-wise independent weight value in one case and Gaussian kernel based blurring in the other case using a 5×5 kernel ($\sigma = 1$). The results of these weight maps are shown in Fig. 6. In comparison to the PSF deconvolution (the proposed solution), they perform quite poorly.

3. The logic of using a blurring or smoothing operation is that it can reduce the pixel-wise random nature of noise in the background. Unsurprisingly, Gaussian blur based weight map indeed performs better than Logit-based weight map. The question then is instead of non-PSF based treatment, could using PSF (2D Gibson Lanni model) as the blurring function help further. This is also tested and shown in Fig. 6. It works further better than Gaussian blurring since it incorporates the softening of the order actually present in the image, but still does not match the performance of using deconvoution with 2D PSF.

4. Lastly, we move to the full-scale physics-guided learning through deconvolved image as the weight map. This incorporates not just the foreground prioritization, soft treatment of foreground and background, but two more physical aspects specific to the problem. First, It suppresses the pixel-wise random background noise without changing its distribution (which gets altered by smoothing operation). This happens as a consequence of improved contrast of the foreground. Second, deconvolution using PSF suppresses the out-of-focus blur introduced in the image plane, thereby also introducing a systematic deprioritization of out-of-focus blur without needing z-stack images from the other planes.

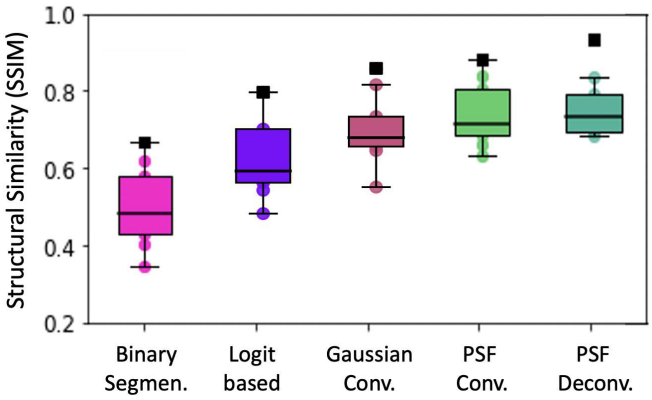

Fig. 6. Performance comparison of different weight map candidates is shown here. Hyper-parameters are the same as method in Table 1 except for the use of different weight maps for the weighted L1 loss. Conv. and Deconv. stand for convolution and deconvolution respectively. Further, logit refers to pixel-wise independent weight map generated based on classification strength of a pixel into background or foreground. Gauss indicates the use of a Gaussian kernel for convolution. PSF refers to the 2-dimensional point spread function computed using Gibson Lanni model. The right most candidate is the same as method A reported in Fig. 5.

Metrics for analysis of virtual labeling: The histograms of SSIM, PSNR, KL divergence, and correlation values obtained over the entire test datasets are shown in Fig. 7(A). The metrics are computed on crops of size 256×256 pixels that are present in the test dataset. Only the crops that had less than 20% of mitochondria pixels were removed from all the datasets. This means that crops with high density of mitochondria were also retained, and therefore the possibility of bias due to sparse regions can be excluded. The mean, median and Q75 values for the

Biomedical Optics EXPRESS

Pearson correlation coefficient are 0.86, 0.89, and 0.94, respectively. We note that our results are significantly better than the mean value of 0.71 for the Pearson correlation coefficient reported in [15] for mitochondria. We also note that our results are reported for 483 test images, in comparison to 20 z-stacks used in Ref. [15]. We present a disclaimer that we are not sure which 64×64×50 (or 75) pixels were used, or if multiple such stacks were sub-sampled from the original microscopy data.

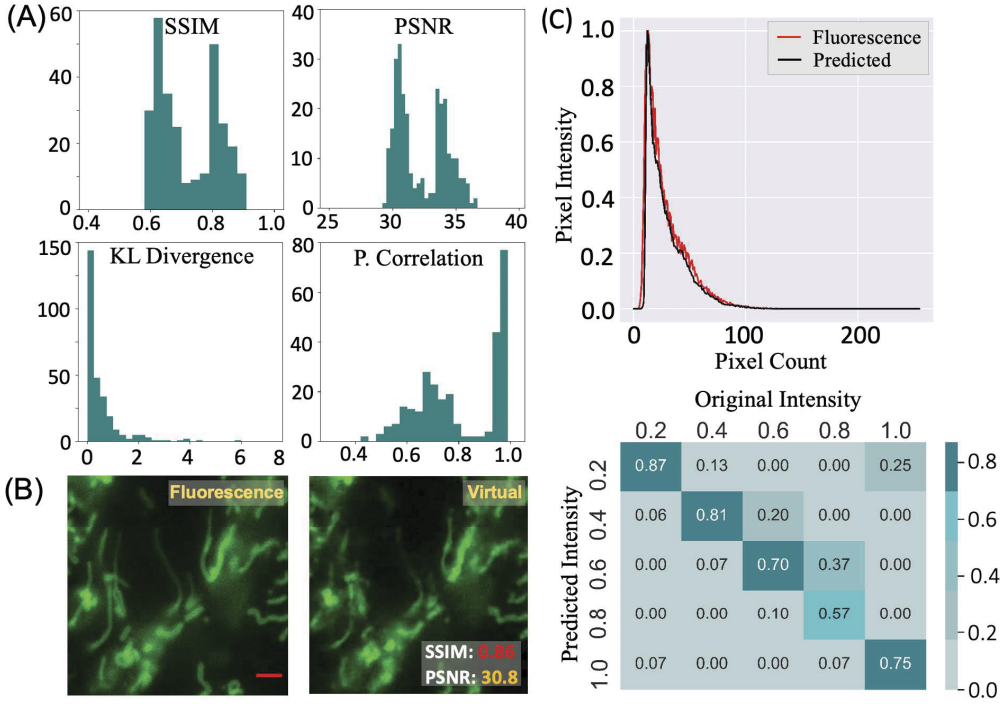

Fig. 7. Histograms of the quantitative metrics SSIM, PSNR, KL divergence, and Pearson correlation for the entire test dataset are shown in (A). An example of the original fluorescence and the predicted virtual image (scale bar 2.5 μm) contrast-enhanced with 0.2% pixel saturation for visualization purpose is shown in (B). Intensity distribution for the sample is shown in (C).

Examples of performance on challenging images: A visual comparison of the fluorescence images and corresponding predicted virtually labeled images are shown in Fig. 8. The first sample, case I, is comparatively easy because of the high visibility of mitochondria using both imaging modalities. It also contains less out-of-focus light compared to cases II and III. The SSIM (0.89) and PSNR (33.9) are therefore high. Cases II and III are more challenging samples due to the presence of (a) out-of-focus mitochondria, and (b) high visibility of organelles other than mitochondria. We can observe that the proposed method is still able to isolate mitochondrial structures and is successful in generating virtually labeled images with an SSIM of 0.87 and PSNR of 33.0 for case II. Case III has an additional challenge due to the unusual presence of a long tube-like structure in the bright-field image which are not mitochondria, resulting in slightly poorer performance (SSIM 0.86 and PSNR 30.6). Still, we can see that the proposed method successfully predicts both in-focus and out-of-focus regions in the qualitative sense. Since the proposed loss function tends to deprioritize learning the out-of-focus light, the SSIM value is poorer (0.86). The last two rows of Fig. 8 present the pixel-wise predicted histogram of the image and the intensity confusion matrices consisting of pixel-by-pixel true and predicted intensities.

Biomedical Optics EXPRESS

We note that the model produces highly correlated fluorescence images for label-free images given as input.

4.2. Comparison with the state-of-the-art virtual labeling

There is one challenging problem of sub-cellular virtual labeling that has recently been addressed successfully [15]. The authors trained a deep learning model to translate from z-stacks of label-free images to z-stacks of fluorescence images. The use of z-stacks makes the mapping better conditioned since what is out-of-focus in one plane may be in-focus in nearby planes. However, the approach needed a deep learning architecture capable of learning 3D mapping. No physics-guided loss function was used.

When evaluating the performance of virtual labeling, the authors used a modified version of Pearson correlation where the performance on background pixels is discounted. This means that the learning still treated the background and the foreground in a similar manner, but the performance for background pixels was not assessed irrespective of the quality of reconstruction. Nonetheless, a SOTA performance in terms of Pearson correlation coefficient was reported for challenging structures qualitatively using fluorescently labeled human induced pluripotent stem cells (hiPSCs) such as AICS-10 (endoplasmic reticulum) with max ~0.78, AICS-54 (cell membrane) ~0.7 and AICS-11 (mitochondria) ~0.71.

We propose a simpler architecture that only maps one label-free 2D image (x-y plane) to a fluorescence counterpart. Therefore, we deal with a problem that is not as well-conditioned, but using the PSF-based loss function. However, we compare how our model performs on the dataset of [15] by considering 2D image pairs for both training and testing from their dataset. Further, we assess the performance of virtual labeling for both the foreground and background pixels.

Quantitative results are shown in Fig. 9, for the endoplasmic reticulum, membrane, and mitochondria. Example qualitative results are shown in Fig. 10. We see that the results of virtual labeling are of high quality, both qualitatively and quantitatively. An intensity match is observed in the histograms for both the foreground and background pixels, which is also seen visually. Note that the Pearson correlation coefficients are consistently better than those reported in [15] with no explicitly imposed performance evaluation biasedness (like lower and average upper bound estimation by negating the variance of random fluctuations for background noise).

Through this experiment, we establish that (a) our method can be applied in a versatile manner across a variety of sub-cellular organelles, (b) our method learns the noise distribution of the background effectively even while prioritizing the learning of foreground pixels, and (c) the overall performance of our method is boosted owing to a better match across both the foreground and background pixels.

4.3. Failure analysis and discussion

Here, we present an example of a failure case in Fig. 11, which indicates an opportunity for insight and possible improvement in the performance of virtual labeling. This figure uses test examples from the dataset of [15]. The leftmost result has an SSIM of 0.16, significantly poorer than our generally observed results. Further results on the right show increasingly better SSIM. Below we present some insights and prospects for improvements.

Our observation across both our own dataset and the dataset of [15] is that the severity for the learning paradigm is not evident in the complexity or visual nature of bright-field images. However, the fluorescence images in the training datasets appear to have a strong bearing on learning effectiveness and the quality of virtual labeling. For instance, in the leftmost failure case, the fluorescence image contains a fairly complex structure lacking sharpness and exhibits high background due to the non-specific fluorescence labeling. These characteristics make learning structures quite challenging due to the minute pixel intensity differences. As we shift rightwards, the sharp increase in metric accuracy is supported by the fact that the crispness of the structure is

Biomedical Optics EXPRESS

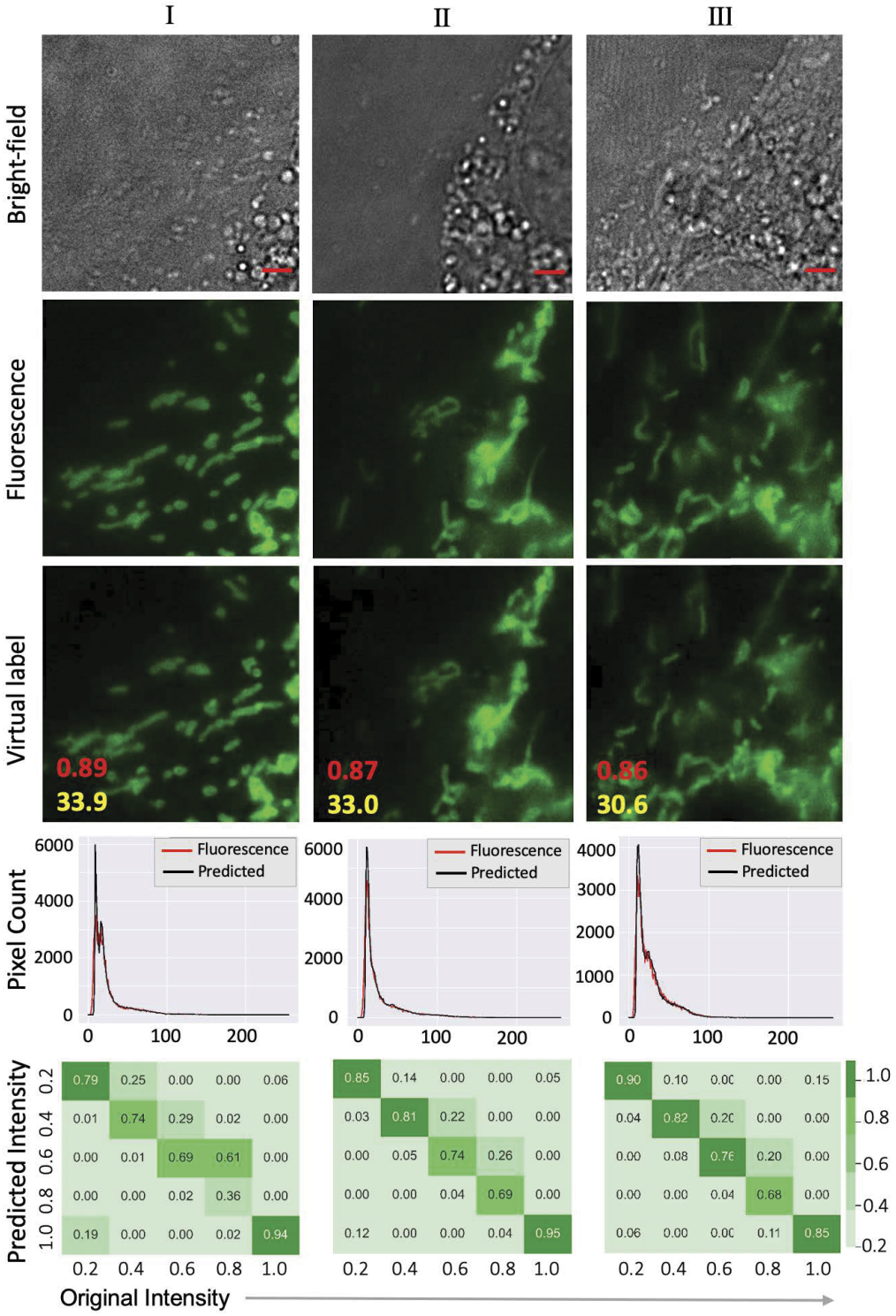

Fig. 8. Examples of bright-field, fluorescence, and virtually labeled (predicted) images (scale bar 2.5 μm). The bottom row represents intensity confusion matrix of the fluorescence and virtual images.

Biomedical Optics EXPRESS

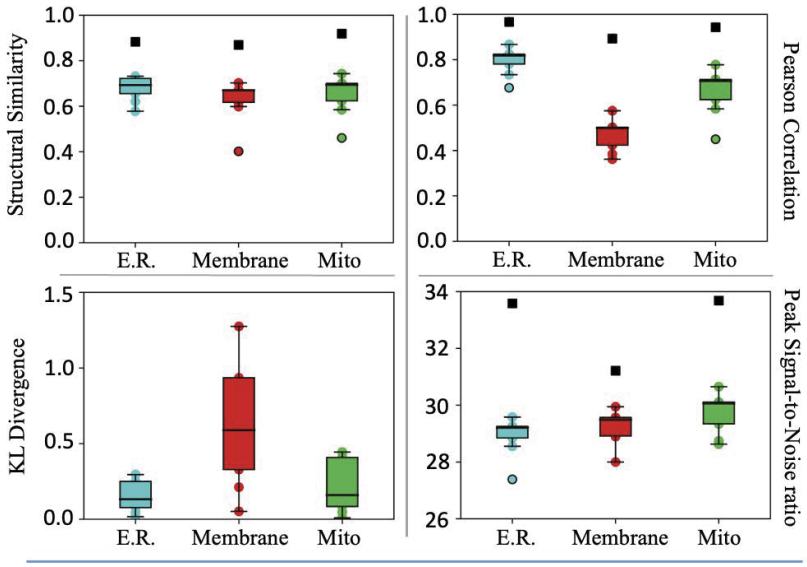

Fig. 9. Boxplot distribution of pairwise aligned dataset [15] performance measured on SSIM (marked in red), Pearson correlation (white), KL divergence, and PSNR (yellow) is illustrated on top. The comparison derived from the marked sub-cellular structure with max, mean, median, and interquartile range Q90, Q75, Q50, Q25, Q10 between the ground-truth and predicted test image shows more promising performance over the reported result [15]. The number of images used for testing were 650 for the endoplasmic reticulum, 494 for membrane and 620 for mitochondria.

higher, the contrast between the foreground and the background is higher, and this has likely helped achieve better accuracy for the respective image samples. We assume the reason for this behavior is that deep learning performs well with edge-based, high contrast and well-defined structures.

Potential solutions are as follows. Choosing a label with high structural specificity (as the genetically introduced mitochondrial tag used above) avoids both high background signals and learning the wrong structures. Furthermore, high-quality electronics and a longer exposure time can improve the signal-to-noise ratio and therefore present a good contrast in the fluorescence images. Some computational compensation for poor contrast images may also be designed in the future. However, the crispness of structures might be out of control of the experimental design. If a high quality virtual labeling is sought for non-sharp structures, more advanced loss-function design or modified architectures might be needed. Potentially a larger training dataset with more diversity of images might have alleviated the problem significantly. However, this hypothesis could not be tested within the regime of the limited data provided in Ref. [15].

4.4. Segmentation and tracking

Segmentation: The segmentation is performed using a U-Net based supervised deep neural network [37]. First, a set of 30 fluorescent images of 1024×1024 were manually segmented using the Sefexa image segmentation tool [47]. Next, randomly cropped ($\sim$ 2500 images of 256×256 pixels) fluorescence images and the corresponding manually segmented images were used to train the U-Net. The segmentation was validated on 500 images of 256×256 pixels. The predicted virtually labeled images were used as the input of the pre-trained segmentation module. Hence, errors in the virtual prediction model were also transferred to the segmentation module

Biomedical Optics EXPRESS

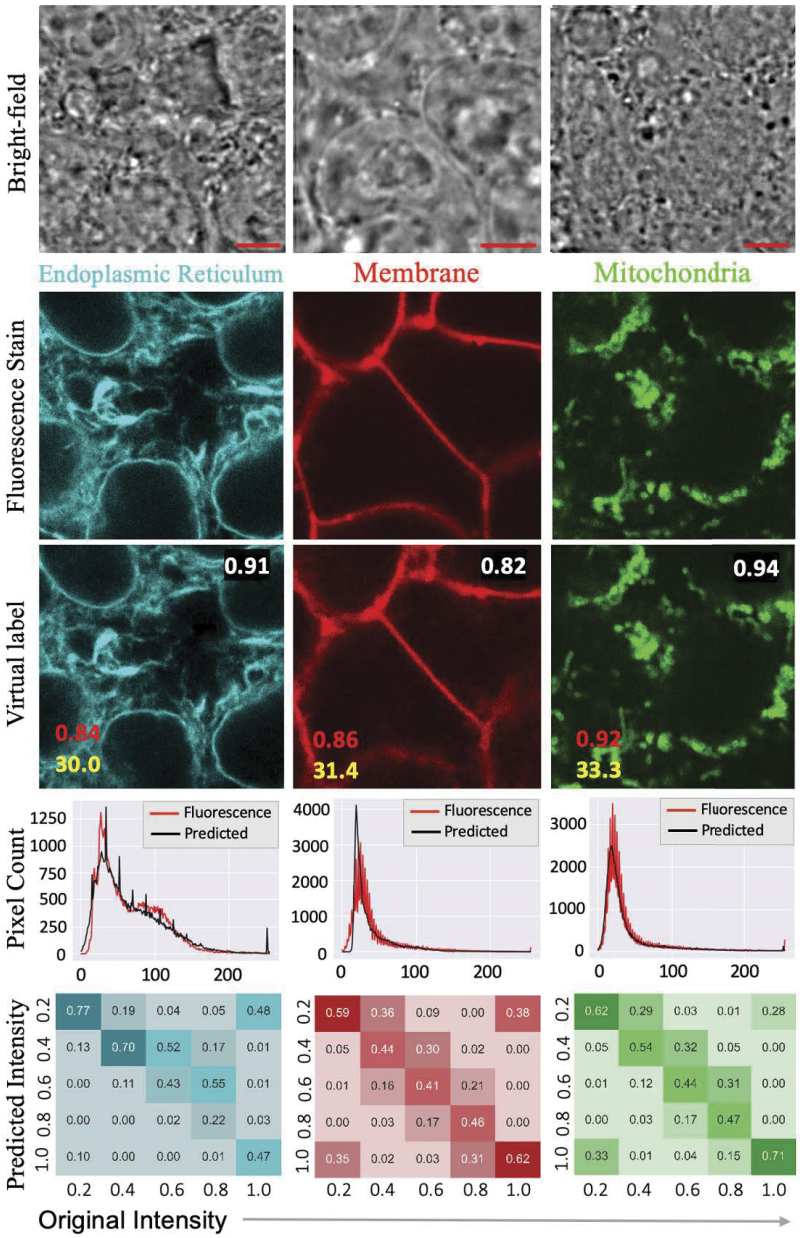

Fig. 10. A sample result each for all sub-cellular structures (scale bar 4 μm) obtained from test transmitted-light images. We include pixel count projection graph and intensity confusion metric, respectively.

Biomedical Optics EXPRESS

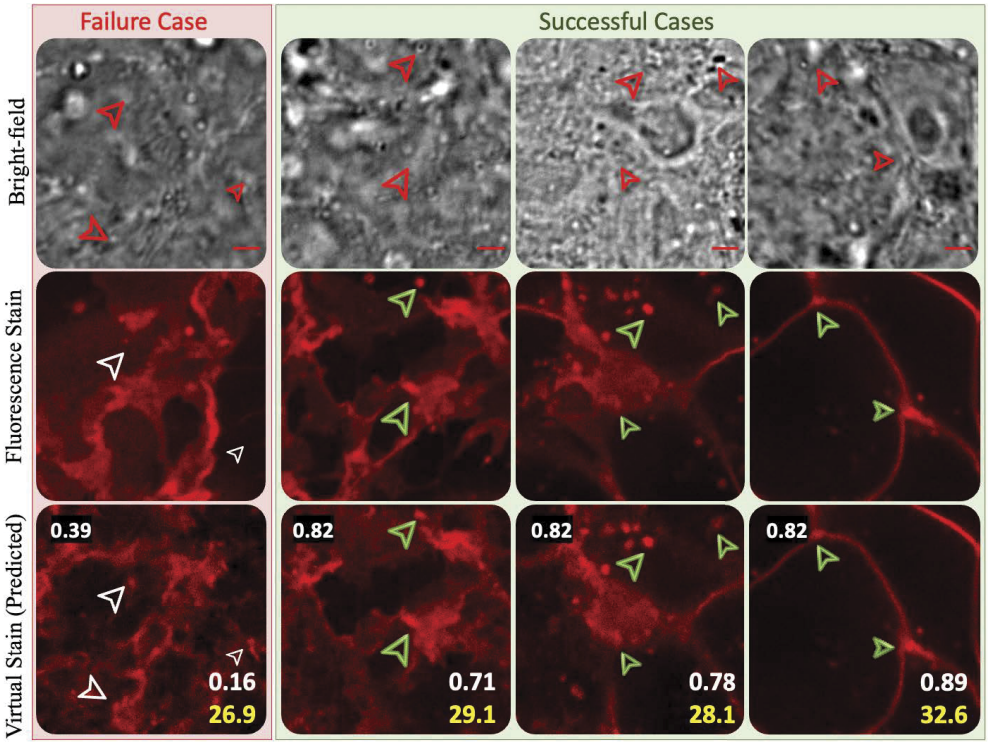

Fig. 11. Visualization of cell membrane prediction shown in red by our proposed loss. The first column is a randomly cropped label-free image (scale bar 2.5 μm) of 256 x 256 from the dataset [15]. The second row is the true fluorescence images for cell membrane, and the third depicts the predicted embodiment with quantitative metric - SSIM & PSNR. Left to right are the failure and successful cases for the dataset explained in detail in section 4.3.

and affected the accuracy. When applied to the test set, we achieved a mean intersection over union (mIoU) of 0.78.

Tracking: Here, we present a case study of the proposed tracking method. A time-lapse sequence of 100 frames (label-free) is used as the input. In Fig. 12, we show an example of a mitochondrion (id:22) tracked over time. The temporal morphological changes are easily identified.

Segmentation results on Mitometer: Mitometer [32] is an open source tool for performing advance analysis of mitochondria including segmentation and tracking. It uses high resolution fluorescence images of mitochondria as input. We show the versatility of our virtual labeling approach by demonstrating that advanced fluorescence microscopy image analysis tools can be applied directly without any modification to bright-field images through our virtual labeling. We take different crops and perform segmentation using Mitometer on both fluorescence images and virtually labeled images and present the results in Fig. 13. It is seen that the segmentations are quite comparable, indicating good adaptation of Mitometer for mitochondria in our virtually labeled images.

Biomedical Optics EXPRESS

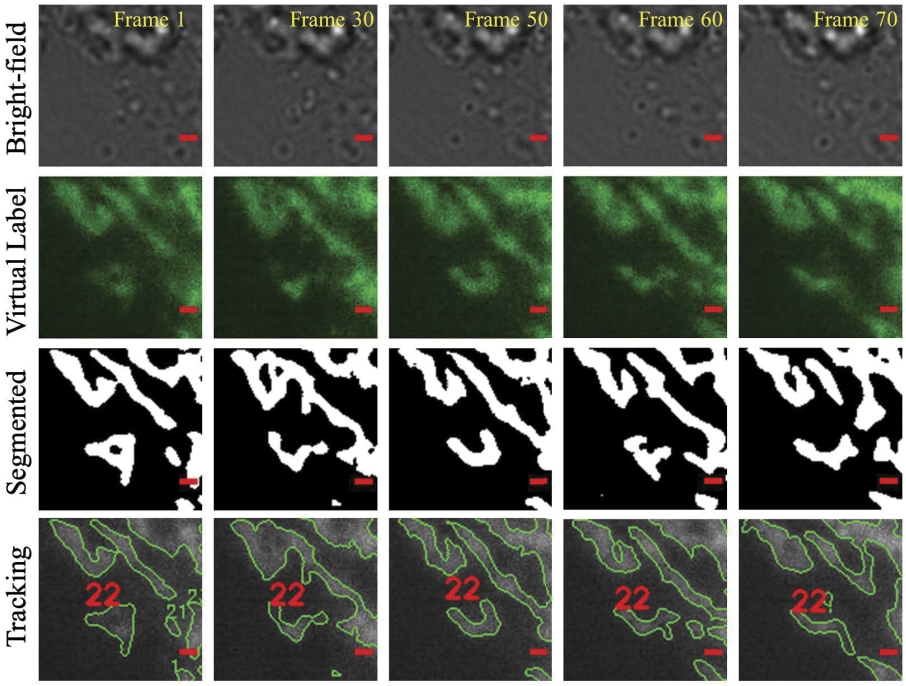

Fig. 12. Example of tracking of mitochondria in a living H9c2 cell. A mitochondrion (id: 22) is tracked across several frames. The scale bar is 1 µm.

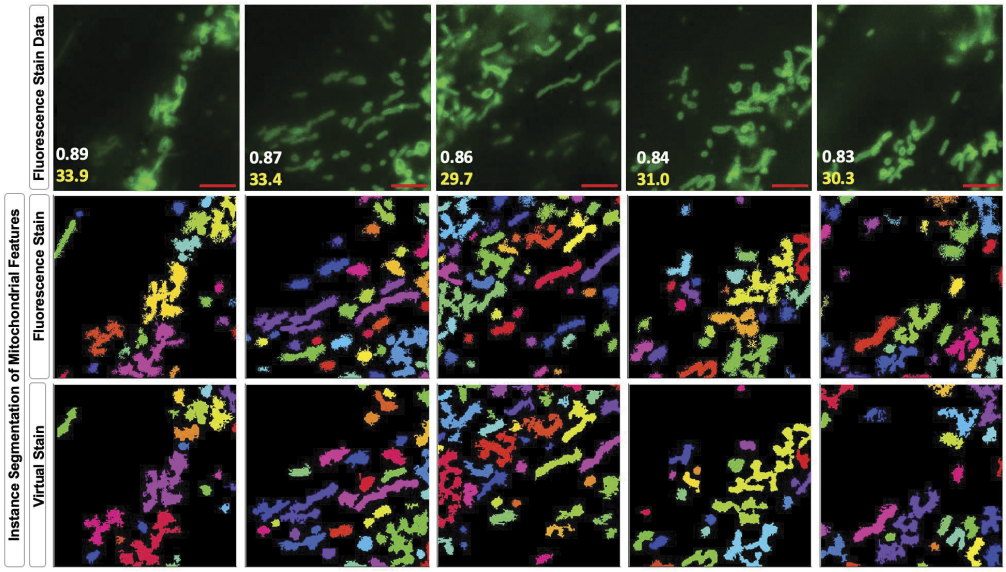

Fig. 13. Example of using mitochondria toolset Mitometer [32] without any modification on virtually labeled images of mitochondria. They indicate comparable performance for actual fluorescence and virtually labeled images.

Biomedical Optics EXPRESS

5. Discussion and conclusion

This work successfully integrated the microscope PSF in deep learning for obtaining SOTA virtual labeling of sub-cellular organelles such as mitochondria. The PSF-based deconvolved fluorescence ground truth image is used for weighing the L1 loss such that the learning of noisy background is down-weighed while the learning for the foreground features are prioritized. The concept is expected to pave the way for physics-guided artificial intelligence solutions for highly sparse fluorescence microscopy data of sub-cellular organelles. Furthermore, virtual labeling for very challenging situations is demonstrated successfully. Besides this, an important contribution is to provide additional insights by incorporating quantitative performance evaluation metrics conventionally used in computer vision, but not currently used in virtual labeling problems. Finally, through a good virtual labeling approach, tasks like segmentation and tracking of individual sub-cellular structures such as mitochondria using completely label-free strategies have been demonstrated to be feasible. We expect such tracking to be a powerful tool for studying organelles' short- and long-term dynamics in living cells due to the infinite photon-budget and low photo-toxicity of label-free microscopy.

We discuss some opportunities and future prospects of effective and accurate virtual labeling. We have taken the problem of virtual labeling mitochondria in this work for two reasons – the technical challenges of performing virtual labeling of mitochondria and our inherent interest in studying mitophagy in cardiac cells in the long term. It will be interesting in the future to plan more extensive virtual labeling of multiple organelles using single bright-field images in order to facilitate studies of molecular cell biology mechanisms involving multiple organelles. Specific studies of interests that benefit from such studies include autophagy, apoptosis, drug response among others.

It will also be interesting to design studies in which most morphologically specific organelles are virtually labeled and fluorescence labeling is used primarily for chemical specificity or other spatially distributed biophysical features such as diffusing population of molecules or vesicles, pH level markers, and other such properties. Such experiments will enable functional studies with maximal information collection supported by multi-color multi-modal microscopes. Further, as demonstrated, application of fluorescence image analysis tools on label-free images imply that quantitative analysis and other morphology-based studies are not rules out in such studies.

Lastly, we draw attention to some drawbacks and possible improvements. The first drawback is the need to exclude the crops that are primarily background. We need to ensure that each crop has certain minimum number of pixels in the foreground otherwise the crops that are primarily background bias the model to learn background noise, which may destablize the learning process or result into poor learning despite the use of PSF-guidance or both. We believe that this drawback cannot necessarily be worked around but sophisticated and optimized automated pipelines can be designed to perform data curation with good outcomes. Next, we note that our approach of PSF-guided virtual labeling is quite good for small structures such as mitochondria, but performs poorly for distributed structures such as membrane, as seen in Fig. 9 and Fig. 11. Such distributed structures may benefit from a weight map based on smoothing rather than deconvolution, however this is pending exploration in the near future. Another interesting possibility to explore in the future is whether the PSF-guided virtual labeling helps also to translate label-free widefield images to equivalent confocal fluorescence images. The point of interest here is that the label-free images widefield images suffer from out-of-focus light but the confocal microscopes suppress out-of-focus light with significantly better efficiency than widefield fluorescence microscopes. Lastly, we believe that the principle of physics-guided deep learning through PSF-guided weighted loss functions will present advantages for applications of deep learning other than virtual labeling in microscopy image processing and image translation problems.

Funding. Universitetet i Tromsø (2061348, Tematisk Satsinger project VirtualStain); Norges Forskningsråd (309802,

Biomedical Optics EXPRESS

INTPART Project); H2020 Future and Emerging Technologies (964800, FetOpen RIA); H2020 Excellent Science (804233, ERC Starting Grant).

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

Data availability. The data will be made available through UiT public repository for large datasets, namely DataverseNO. The code and dataset link to DataverseNO repository are available at VirtualStain_I2I project repository [48].

References

1. A. Alsina, W. M. Lai, W. K. Wong, X. Qin, M. Zhang, and H. Park, "Real-time subpixel-accuracy tracking of single mitochondria in neurons reveals heterogeneous mitochondrial motion," Biochem. Biophys. Res. Commun. **493**(1), 776–782 (2017).
2. P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "Mots: Multi-object tracking and segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, (2019), pp. 7942–7951.
3. N. Chenouard, I. Smal, and F. De Chaumont, *et al.*, "Objective comparison of particle tracking methods," Nat. Methods **11**(3), 281–289 (2014).
4. Y. Rivenson, T. Liu, Z. Wei, Y. Zhang, K. de Haan, and A. Ozcan, "Phasestain: the digital staining of label-free quantitative phase microscopy images using deep learning," Light: Sci. Appl. **8**(1), 23 (2019).
5. A. A. Sekh, I. S. Opstad, G. Godtliebsen, Å. B. Birgisdottir, B. S. Ahluwalia, K. Agarwal, and D. K. Prasad, "Physics-based machine learning for subcellular segmentation in living cells," Nat. Mach. Intell. **3**(12), 1071–1080 (2021).
6. N. Bayramoglu, M. Kaakinen, L. Eklund, and J. Heikkila, "Towards virtual h&e staining of hyperspectral lung histology images using conditional generative adversarial networks," in *IEEE International Conference on Computer Vision Workshops*, (2017), pp. 64–71.
7. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, "In silico labeling: predicting fluorescent labels in unlabeled images," Cell **173**(3), 792–803.e19 (2018).
8. Y. Rivenson, H. Wang, Z. Wei, K. de Haan, Y. Zhang, Y. Wu, H. Günaydın, J. E. Zuckerman, T. Chong, A. E. Sisk, L. M. W. Westbrook, W. D. Wallace, and A. Ozcan, "Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning," Nat. Biomed. Eng. **3**(6), 466–477 (2019).
9. A. Lahiani, J. Gildenblat, I. Klaman, S. Albarqouni, N. Navab, and E. Klaiman, "Virtualization of tissue staining in digital pathology using an unsupervised deep learning approach," in *European Congress on Digital Pathology*, (Springer, 2019), pp. 47–55.
10. Y. Liu, H. Yuan, Z. Wang, and S. Ji, "Global pixel transformers for virtual staining of microscopy images," IEEE Trans. Med. Imaging **39**(6), 2256–2266 (2020).
11. S. Cheng, S. Fu, Y. M. Kim, W. Song, Y. Li, Y. Xue, J. Yi, and L. Tian, "Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy," Sci. Adv. **7**(3), eabe0431 (2021).
12. M. E. Kandel, Y. R. He, Y. J. Lee, T. H.-Y. Chen, K. M. Sullivan, O. Aydin, M. T. A. Saif, H. Kong, N. Sobh, and G. Popescu, "Phase imaging with computational specificity (pics) for measuring dry mass changes in sub-cellular compartments," Nat. Commun. **11**(1), 6256 (2020).
13. P. YongKeun, W. Park, Y. Jo, H. Min, and H. Cho, "Method and apparatus for generating 3-d molecular image based on label-free method using 3-d refractive index image and deep learning," (2021). US Patent App. 16/823, 453.
14. J. Choi, H.-J. Kim, G. Sim, S. Lee, W. S. Park, J. H. Park, H.-Y. Kang, M. Lee, W. Do Heo, J. Choo, H. Min, and Y. Park, "Label-free three-dimensional analyses of live cells with deep-learning-based segmentation exploiting refractive index distributions," bioRxiv (2021).
15. C. Ounkomol, S. Seshamani, M. M. Maleckar, F. Collman, and G. R. Johnson, "Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy," Nat. Methods **15**(11), 917–920 (2018).
16. S. Guo, Y. Ma, Y. Pan, Z. J. Smith, and K. Chu, "Organelle-specific phase contrast microscopy enables gentle monitoring and analysis of mitochondrial network dynamics," Biomed. Opt. Express **12**(7), 4363–4379 (2021).
17. L. Galteri, L. Seidenari, M. Bertini, and A. Del Bimbo, "Deep generative adversarial compression artifact removal," in *IEEE International Conference on Computer Vision*, (2017), pp. 4826–4835.
18. E. Meijering, "Cell segmentation: 50 years down the road [life sciences]," Image Vis. Comput. **29**(5), 140–145 (2012).
19. J. Nikolaisen, L. I. Nilsson, I. K. Pettersen, P. H. Willems, J. B. Lorens, W. J. Koopman, and K. J. Tronstad, "Automated quantification and integrative analysis of 2d and 3d mitochondrial shape and network properties," PLoS One **9**(7), e101365 (2014).
20. S. B. Ekanayake, A. M. El Zawily, G. Paszkiewicz, A. Rolland, and D. C. Logan, "Imaging and analysis of mitochondrial dynamics in living cells," in *Plant Mitochondria*, (Springer, 2015), pp. 223–240.
21. E. Lihavainen, J. Mäkelä, J. N. Spelbrink, and A. S. Ribeiro, "Mytoe: automatic analysis of mitochondrial dynamics," Bioinformatics **28**(7), 1050–1051 (2012).
22. A. Zahedi, V. On, R. Phandthong, A. Chaili, G. Remark, B. Bhanu, and P. Talbot, "Deep analysis of mitochondria and cell health using machine learning," Sci. Rep. **8**(1), 16354 (2018).

Biomedical Optics EXPRESS

23. M. Naser, R. S. Schloss, P. Berjaud, and N. N. Boustany, "Label-free dynamic segmentation and morphological analysis of subcellular optical scatterers," J. Biomed. Opt. **23**(09), 1 (2018).
24. A. Rohani, J. H. Moore, J. A. Kashatus, H. Sesaki, D. F. Kashatus, and N. S. Swami, "Label-free quantification of intracellular mitochondrial dynamics using dielectrophoresis," Anal. Chem. **89**(11), 5757–5764 (2017).
25. T. Vicar, J. Balvan, J. Jaros, F. Jug, R. Kolar, M. Masarik, and J. Gumulec, "Cell segmentation methods for label-free contrast microscopy: review and comprehensive comparison," BMC Bioinf. **20**(1), 360 (2019).
26. O. Hilsenbeck, M. Schwarzfischer, and S. Skylaki, *et al.*, "Software tools for single-cell tracking and quantification of cellular and molecular properties," Nat. Biotechnol. **34**(7), 703–706 (2016).
27. H. Mathys, C. Adaikkan, F. Gao, J. Z. Young, E. Manet, M. Hemberg, P. L. De Jager, R. M. Ransohoff, A. Regev, and L.-H. Tsai, "Temporal tracking of microglia activation in neurodegeneration at single-cell resolution," Cell Rep. **21**(2), 366–380 (2017).
28. T. He, H. Mao, J. Guo, and Z. Yi, "Cell tracking using deep neural networks with multi-task learning," Image and Vision Computing **60**, 142–153 (2017).
29. S. U. Akram, J. Kannala, L. Eklund, and J. Heikkilä, "Joint cell segmentation and tracking using cell proposals," in *IEEE International Symposium on Biomedical Imaging (ISBI)*, (IEEE, 2016), pp. 920–924.
30. J. Hayashida, K. Nishimura, and R. Bise, "Mpm: Joint representation of motion and position map for cell tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, (2020), pp. 3823–3832.
31. A. A. Sekh, I. S. Opstad, A. B. Birgisdottir, T. Myrmel, B. S. Ahluwalia, K. Agarwal, and D. K. Prasad, "Learning nanoscale motion patterns of vesicles in living cells," in *IEEE Conference on Computer Vision and Pattern Recognition*, (2020), pp. 14014–14023.
32. A. E. Lefebvre, D. Ma, K. Kessenbrock, D. A. Lawson, and M. A. Digman, "Automated segmentation and tracking of mitochondria in live-cell time-lapse images," Nat. Methods **18**(9), 1091–1102 (2021).
33. M. P. Viana, S. Lim, and S. M. Rafelski, "Quantifying mitochondrial content in living cells," in*Methods in Cell Biology*, vol. 125 (Elsevier, 2015), pp. 77–93.
34. A. J. Valente, L. A. Maddalena, E. L. Robb, F. Moradi, and J. A. Stuart, "A simple imagej macro tool for analyzing mitochondrial network morphology in mammalian cell culture," Acta Histochem. **119**(3), 315–326 (2017).
35. T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *IEEE Conference on Computer Vision and Pattern Recognition*, (2018).
36. G. Liu, K. J. Shih, T.-C. Wang, F. A. Reda, K. Sapra, Z. Yu, A. Tao, and B. Catanzaro, "Partial convolution based padding," arXiv preprint arXiv:1811.11718 (2018).
37. O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-assisted Intervention*, (Springer, 2015), pp. 234–241.
38. G. Wang, C. Lopez-Molina, and B. De Baets, "Automated blob detection using iterative laplacian of gaussian filtering and unilateral second-order gaussian kernels," Digit. Signal Process. **96**, 102592 (2020).
39. S.-K. Weng, C.-M. Kuo, and S.-K. Tu, "Video object tracking using adaptive kalman filter," J. Vis. Commun. Image Represent. **17**(6), 1190–1208 (2006).
40. K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "MedGAN: Medical image translation using gans," Comput. Med. Imaging Graph. **79**, 101684 (2020).
41. Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *IEEE Conference on Computer Vision and Pattern Recognition*, (2018), pp. 8789–8797.
42. D. Walter, A. Liu, E. Franklin, D. Macdonald, B. Mitchell, and T. Trupke, "Contrast enhancement of luminescence images via point-spread deconvolution," in *IEEE Photovoltaic Specialists Conference*, (IEEE, 2012), pp. 000307–000312.
43. S. F. Gibson and F. Lanni, "Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy," J. Opt. Soc. Am. A **9**(1), 154–166 (1992).
44. H. Yan, W.-J. Yan, and W.-W. Li, "Image restoration based on lucy-richardson algorithm," Computer Engineering **36**, 204–205 (2010).
45. A. Somani, A. A. Sekh, I. S. Opstad, Å. B. Birgisdottir, T. Myrmel, B. S. Ahluwalia, K. Agarwal, D. K. Prasad, and A. Horsch, "Digital staining of mitochondria in label-free live-cell microscopy," in *Bildverarbeitung für die Medizin 2021*, (Springer, 2021), pp. 235–240.
46. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," Advances in Neural Information Processing Systems **27**, 1 (2014).
47. "Sefexa Image Segmentation Tool," http://www.fexovi.com/sefexa.html (2021).
48. A. Somani, "Code for virtual labeling of mitochondria in living cells using correlative imaging and physics-guided deep learning," https://github.com/AyushSomani001/VirtualStain_I2I (2022).