Label-free prediction of three-dimensional fluorescence images from transmitted light microscopy

Chawin Ounkomol[1], Sharmishtaa Seshamani[2], Mary M. Maleckar[1], Forrest Collman[2] & Gregory R. Johnson[1]*

[1] *Allen Institute for Cell Science, 615 Westlake Ave N, Seattle, WA 98109*

[2] *Allen Institute for Brain Science, 615 Westlake Ave N, Seattle, WA 98109*

Understanding living cells as integrated systems, a challenge central to modern biology, is complicated by limitations of available imaging methods. While fluorescence microscopy can resolve subcellular structure in living cells, it is expensive, slow, and damaging to cells. Here, we present a label-free method for predicting 3D fluorescence directly from transmitted light images and demonstrate that it can be used to generate multi-structure, integrated images.

The various imaging methods currently used to capture details of cellular organization all present trade-offs with respect to expense, spatio-temporal resolution, and sample perturbation. Fluorescence microscopy permits imaging of specific proteins and structures of interest via labeling but requires expensive instrumentation and time consuming sample preparation. Critically, samples are subject to significant phototoxicity[1] and photobleaching[2], creating a tradeoff between the quality of data and timescales available to live cell imaging. Furthermore, the number of simultaneous fluorescent tags is restricted by both spectrum saturation and cell health, limiting the number of parallel labels which may be imaged together. Transmitted light microscopy, e.g., bright-field, phase, DIC, etc., in contrast, is a relatively low-cost, dye-free modality with greatly reduced phototoxicity and simplified sample preparation. Although valuable information about cellular organization is apparent in transmitted light images, they lack the specificity inherent in fluorescence labeling. A method which could combine the specificity of fluorescence microscopy with the simplicity, modest cost, and much lower toxicity of transmitted light techniques would present a potentially groundbreaking tool for biologists to garner insight into the integrated activities of subcellular structures.

Here, we present a convolutional neural network (CNN)-based tool (Fig. 1), employing a U-Net architecture[3] (methods, Supplementary Fig. 1) to model the relationships between 3D transmitted light (bright-field and DIC) and fluorescence images corresponding to several major subcellular structures (i.e., cell membrane, nuclear envelope, nucleoli, DNA, and mitochondria). We show that our method can train a model to learn this relationship for the structure of interest given only spatially registered pairs of images, even with a relatively small image set for training (30 image pairs per structure). The resultant model can, in turn, be used to predict a 3D fluorescence image from a new transmitted light input. Model predictions for a variety of subcellular structures can be combined, enabling multi-channel, integrated fluorescence imaging from a single transmitted light input (Fig. 1d, e).

While the biological detail that can be observed in predicted images varies by subcellular structure, many of the predicted images are quantitatively and qualitatively similar to ground truth fluorescence images in 3D. Nuclear structures are well-resolved: images produced by the DNA model (Fig. 1b) depict well-formed and separated nuclear regions as well as finer detail, including chromatin condensation just before and during mitosis, and the nuclear envelope model predictions (Supplementary Fig. 2) provide a high-resolution localization of its 3D morphology. The nucleoli model also resolves the precise location, number and morphology of individual nucleoli. Models for several other structures also perform favorably upon visual inspection as compared to ground-truth images. For example, the mitochondria model correctly identifies the regions of cells with high numbers of mitochondria as well as regions which are more sparsely populated. In many cases, individual mitochondria visible in the fluorescence data are also observable in predictions (Supplementary Fig. 2). While predictions for microtubules and the endoplasmic reticulum do not resolve individual filaments or detailed morphology, respectively, these models successfully capture broader 3D localization patterns of those structures. Given these promising results, we trained the DNA model with an extended training procedure (DNA+) to evaluate whether our results could be improved with additional training images and iterations. As expected, performance of the model improved (methods, Fig. 1c, Supplementary Fig. 2). Most critically, all of these details can be observed together in an integrated multi-channel prediction derived from a single transmitted light image (Fig. 1d, Supplementary Fig. 2 and Supplementary Video). Examples for all fourteen subcellular labeled structure models' test set predictions can be found in Supplementary Fig. 2.

The models' performance was quantified via the Pearson's correlation coefficient on new predicted and ground truth fluorescence image pairs (Fig. 1c) from an independent test set (methods). A theoretical upper bound was determined for ideal model performance based upon an estimate of the signal-to-noise ratio (SNR) of the fluorescence images used for training (methods). Individual structure model performance is well-bounded by this limit (Fig. 1c).

To assess whether structure models trained solely on static images may feasibly be used to predict temporal fluorescence image sequences from transmitted time-series imaging, we applied models for several structures to a single transmitted light 3D time-series (Fig. 1e, Supplemental Video 1). In addition to simultaneous visualization of several subcellular structures, characteristic dynamics of mitotic events, such as the reorganization of the nuclear envelope and cell membrane, are evident in the predicted multi-channel time-series (Fig. 1d). Due to the increased photoxicity which can occur in extended, live cell time-series fluorescence imaging, this information would otherwise be difficult to obtain, particularly in 3D. This result indicates that the models may be sufficiently robust for time-series predictions for which no fluorescence imaging ground truth is available, potentially greatly increasing the timescales over which cellular processes can be visualized and measured.

This powerful method has inherent limitations and may not currently be well suited for all applications. Because models must learn a relationship between distinct but correlated imaging modes, predictive performance is contingent upon the existence of this association. In the case of desmosomes or actomyosin bundles, for example, model performance for the presented training protocol was comparatively poor, perhaps due to a weaker association between transmitted light and fluorescence images of these structures (Fig. 1c, Supplementary Fig. 2). Also, the quality and quantity of training data will likely influence accuracy of the model predictions, and we cannot assess a priori how models will perform in biological contexts for which there are very few or no examples in training or testing data. Specifically, models pre-trained using one cell type (e.g. hiPSC) may not perform as well when applied to inputs with drastically different cellular morphologies (e.g. cardiomyocytes). Furthermore, predictions from inputs acquired with imaging parameters identical to those used to compose a model training set are likely to provide the most accurate results as compared to ground truth data. For example, we successfully trained fluorescence models using both bright-field and DIC modalities. However, it would not be advisable to use a model on inputs from one modality, when that model was trained with a different modality. More subtle parameters can also matter, for example we observed a decrease in model accuracy when predicting fluorescence images from input transmitted light stacks acquired with shorter inter-slice intervals ($\sim 0.13\,\mathrm{s}$) than that in training data ($\sim 2.2\,\mathrm{s}$) (data not shown). Ultimately, when evaluating the utility of predicted images, the context for which those images will be used must be considered. For instance, DNA or nuclear membrane predictions may have sufficient accuracy for application to nuclear segmentation algorithms, but the microtubule predictions would not be effective for assaying rates of microtubule polymerization (Fig. 1e, Supplementary Fig. 2).

Transforming one imaging modality into another could be useful to a variety of imaging challenges, such as cross modal image registration, because it is challenging to automatically register two modalities that have drastically different contrast mechanisms. To demonstrate the utility of this method to solve such problems, we applied it to the registration of array tomography data[4] of ultrathin brain sections (Fig. 2). In this technique, electron micrographs (EM) and ten channels of immunofluorescence images, including myelin basic protein (MBP-IF), are obtained from the same sample but from two different microscopes and thus are not natively spatially registered. Analogous to the relationship between transmitted light and fluorescence, while myelin wraps are apparent in both modalities, the EM image lacks the specificity of the MBP-IF (Fig. 2). EM and

IF images can be registered by hand through identification of corresponding locations and fitting a similarity transformation, resulting in multi-channel conjugate EM images[4]. However, manual registration is a tedious, time-consuming process. While automation would speed data processing, there have been no successful attempts to date to automate this process via conventional statistical image registration techniques. We trained a 2D version of the label-free tool on manually registered pairs of EM and MBP-IF images and used hold-out EM images as input to predict the corresponding MBP-IF images (Fig. 2a); conventional intensity-based matching techniques (methods) were then used to register each MBP-IF prediction (and thus the EM image) to a target MBP-IF tile (Fig. 2b). Successful registration was performed in 86 of 90 image pairs, suggesting the tool's utility may be extended to different imaging modalities and additional downstream image processing tasks.

The label-free methodology presented here has wide potential for use in many biological imaging fields. Primarily, it may be possible to reduce or even eliminate routine capture of some images in existing imaging and analysis pipelines, permitting the same throughput in a far more efficient and cost-effective manner. Notably, data used for training requires little to no pre-processing and relatively small numbers of paired examples, drastically reducing the barrier to entry associated with some machine learning approaches. Areas where this approach may prove of particular value include image-based screens for cellular phenotypes[5] and pathology workflows requiring specialized staining[6]. The method is additionally promising in cases wherein generating a complete set of simultaneous ground-truth labels is not feasible, such as with the live cell time-series imaging example presented here. Finally, the tool permits the generation of integrated images by which interactions among cellular components can be investigated. This implies exciting potential for probing coordination of subcellular organization as cells grow, divide, and differentiate, and signifies a new opportunity for understanding structural phenotypes in the context of disease modeling and regenerative medicine. More broadly, the presented work may suggest an opportunity for a key new direction in biological imaging research: the exploitation of imaging modalities' indirect but learnable relationships to visualize biological features of interest with ease, low cost, and high fidelity.

Methods

Data for modeling training and validation The 3D light microscopy data used to train and test the presented models consists of z-stacks of genome-edited human induced pluripotent stem cell (hiPSc) lines, each expressing a protein endogenously tagged with either mEGFP or mTagRFP that localizes to a particular subcellular structure[7]. The EGFP-tagged proteins and their corresponding structures are: alpha-tubulin (microtubules), beta-actin (actin filaments), desmoplakin (desmosomes), lamin B1 (nuclear envelope), fibrillarin (nucleoli), myosin IIB (actomyosin bundles), sec61B (endoplasmic reticulum), STGAL1 (Golgi apparatus), Tom20 (mitochondria) and ZO1 (tight junctions). The cell membrane was labelled by expressing RFP tagged with a CAAX motif. Samples were prepared by plating cells on 96-well plates and allowing them to propagate for four days. CellMask plasma membrane stain (ThermoFisher) and NucBlue DNA stain (ThermoFisher) were added to the wells to final concentrations of $3\times$ and $1\times$ respectively. Cells were incubated at $37\,^{\circ}\mathrm{C}$ and 5% CO_2 for $10\,\mathrm{min}$ and gently washed with pre-equilibrated phenol red-free mTeSR1. Cells were imaged immediately following the wash step for up to $2.5\,\mathrm{h}$ on a Zeiss spinning disk microscope. Images of cells with EGFP tags were acquired at $100\times$, with four data channels per image: transmitted light (either bright-field or DIC), cell membrane labeled with CellMask, DNA labeled with Hoechst, and EGFP-tagged cellular structure. Respectively, acquisition settings for each channel were: white LED, $50\,\mathrm{ms}$ exposure; $638\,\mathrm{nm}$ laser at $2.4\,\mathrm{mW}$, $200\,\mathrm{ms}$ exposure; $405\,\mathrm{nm}$ at $0.28\,\mathrm{mW}$, $250\,\mathrm{ms}$ exposure; $488\,\mathrm{nm}$ laser at $2.3\,\mathrm{mW}$, $200\,\mathrm{ms}$ exposure. The CAAX-RFP-based cell membrane images were taken with a $63\times$ objective, a $561\,\mathrm{nm}$ laser at $2.4\,\mathrm{mW}$, and a $200\,\mathrm{ms}$ exposure. We did not use the CellMask images in this report because the CAAX tagging provided higher quality cell membrane images. Time-series data were acquired using the same imaging protocol as for acquisition of training data but on unlabeled, wild-type hiPSCs and with all laser powers set to zero. The images were resized via cubic interpolation such that each voxel corresponded to $0.29\,\mu\mathrm{m} \times 0.29\,\mu\mathrm{m} \times 0.29\,\mu\mathrm{m}$. Pixel intensities of all input and target images were independently z-scored. We paired fluorescence and corresponding transmitted light channels, resulting in 13 image collections. For each collection, we allocated 30 image pairs to a training set and all the remaining image pairs to a test set. The training set for the DNA+ model was supplemented with additional images for a total of 540 image pairs.

For conjugate array tomography data [4], images of 50 ultra-thin sections were taken with a wide-field fluorescence microscope using 3 rounds of staining and imaging to obtain 10-channel immunofluorescence (IF) data (including myelin basic protein, MBP) at $100\,\mathrm{nm}$ per pixel. 5 small regions were then imaged with a field emission scanning electron microscope to obtain high resolution electron micrographs at $3\,\mathrm{nm}$ per pixel. Image processing steps independently stitched the

IF sections and one of the EM regions to create 2D montages in each modality. Each EM montage was then manually registered to the corresponding MBP channel montage with TrakEM2[8]. To create a training set, 40 pairs of these registered EM and MBP montages were resampled to $10\,\mathrm{nm}$ per pixel. For each montage pair, a central region of size $2544\,\mathrm{px} \times 2352\,\mathrm{px}$ was cut out and used for the resultant final training set. Pixel intensities of the images were z-scored.

Model architecture description and training procedure We employed a convolutional neural network (CNN) based on various U-Net/3D U-Net architectures[3,9] (Supplementary Fig. 1). In recent years, they have been used in biomedical imaging for a wide range of tasks, including image classification, object segmentation[10], and estimation of image transformations[11]. The model consists of layers that perform one of three convolution types, followed by a batch normalization and ReLU operation. The convolutions are either 3 pixel convolutions with a stride of 1-pixel on zero-padded input (such that input and output of that layer are the same spatial area), 2-pixel convolutions with a stride of 2 pixels (to halve the spatial area of the output), or 2-pixel transposed-convolutions with a stride of 2 (to double the spatial area of the output). There are no normalization or ReLU operations on the last layer of the network. The number of output channels per layer are shown in Supplementary Fig. 1. The 2D and 3D models use 2D or 3D convolutions, respectively.

Due to memory constraints associated with GPU computing, we trained the model on batches of either 3D patches ($32\,\mathrm{px} \times 64\,\mathrm{px} \times 64\,\mathrm{px}$, z-y-x) for light microscopy data or on 2D patches ($256\,\mathrm{px} \times 256\,\mathrm{px}$, y-x) for conjugate array tomography data, which were randomly subsampled uniformly both across all training images as well as spatially within an image. The training procedure took place in a typical forward-backward fashion, updating model parameters via stochastic gradient descent (backpropagation) to minimize the mean squared error between output and target images. All models presented here were trained using the Adam optimizer[12] with a learning rate of 0.001 and with betas 0.5 and 0.999 for 50,000 mini-batch iterations. We used a batch size of 24 for 3D models and of 32 for 2D models. Running on a Pascal Titan X, each model completed training in approximately 16 hours for 3D models and in 7 hours for 2D models. Training of the DNA+ model was extended to 616,880 mini-batch iterations. For prediction tasks, we minimally crop the input image such that its size in any dimension is a multiple of 16, to accommodate the multi-scale aspect of the CNN architecture. Prediction takes approximately 1 second for 3D images and 0.5 seconds for 2D images. Our model training pipeline was implemented in Python using the PyTorch package (`http://pytorch.org`).

3D light microscopy model results analysis and validation For 3D light microscopy applications, model accuracy was quantified by the Pearson's correlation coefficient between the model's output and independent, ground truth test images. For each model, a corresponding estimate of noise was developed based upon image stacks taken of unlabeled, wild-type hIPSC cells for which microscope settings were identical to those used during labeled acquisitions. For each image prediction, we calculated a theoretical upper bound of model performance, based upon the assumption that the variance of the unlabeled image stacks is a lower bound on the variance of uncorrelated, random fluctuations N in the ground truth images T, which should not be predictable, such that at each voxel location: $T_{x,y,z} = N_{x,y,z} + S_{x,y,z}$, where S is the predictable signal in the image. The best possible model would output exactly S, and thus the highest correlation between a model's output and the target is the correlation between T and S which can be calculated to be $C_{max} = \sqrt{\frac{SNR}{1+SNR}}$

where $SNR = \frac{<S^2>}{<N^2>}$.

Registration across imaging modalities Here we employed a 2D version of our tool trained on the montage pairs described above in 'Data for modeling training and validation'. For the test set, each of the individual EM images (without montaging) from all five regions (a total of 1500 images) was used as an input to directly register to the corresponding MBP image in which it lies. For this, each image was first downsampled to $10\,\mathrm{nm}$ per pixel without any transformations to generate a $1500\,\mathrm{px} \times 1500\,\mathrm{px}$ image. This was then reflection padded to $1504\,\mathrm{px} \times 1504\,\mathrm{px}$ as in[3], run through the trained model, and then cropped back to the original input size to generate an MBP prediction image. This MBP prediction image was first roughly registered to MBP IF images using cross-correlation-based template matching for a rigid transformation estimate. Next, the residual optical flow[13] between the predicted image transformed by the rigid estimate and the MBP IF image was calculated, which was then used to fit a similarity transformation that registers the two images (implemented using OpenCV[14]). 90 prediction images were randomly selected from the larger set, where more than 1% of the predicted image pixels were greater than 50% of the maximum intensity, to ensure that the images contained sufficient MBP content to drive registration. Ground truth transformation parameters were calculated by two independent authors on this subset of EM images by manual registration (3-4 minutes per pair) to the MBP IF images using TrakEM2. Differences in registrations (between authors and between the algorithm estimate and one of the authors) was calculated by the average difference in displacement across an image, as measured in pixels of the target IF image.

Software and Data Software for training the models is available at https://github.com/AllenCellModeling/pytorch_fnet. The data used to train the 3D models is available at http://www.allencell.org.

1. Magidson, V. & Khodjakov, A. Circumventing photodamage in live-cell microscopy. *Methods Cell Biol* **114**, 10.1016/B978–0–12–407761–4.00023–3 (2013). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3843244/. 23931522[pmid].

2. Dempsey, G. T., Vaughan, J. C., Chen, K. H., Bates, M. & Zhuang, X. Evaluation of fluorophores for optimal performance in localization-based super-resolution imaging. *Nat Methods* **8**, 1027–1036 (2011). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3272503/. 22056676[pmid].

3. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 234–241 (Springer, 2015).

4. Collman, F. *et al.* Mapping synapses by conjugate light-electron array tomography. *Journal of Neuroscience* **35**, 5792–5807 (2015).

5. Goshima, G. *et al.* Genes required for mitotic spindle assembly in drosophila s2 cells. *Science* **316** (2007).

6. Gurcan, M. N. *et al.* Histopathological image analysis: A review. *IEEE Rev Biomed Eng* **2**, 147–171 (2009). URL http://www.ncbi.nlm.nih.gov/pmc/articles/

PMC2910932/. 20671804[pmid].

7. Roberts, B. *et al.* Systematic gene tagging using crispr/cas9 in human stem cells to illuminate cell organization. *bioRxiv* 123042 (2017).

8. Cardona, A. *et al.* Trakem2 software for neural circuit reconstruction. *PloS one* **7**, –38011 (2012).

9. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. 3d u-net: learning dense volumetric segmentation from sparse annotation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 424–432 (Springer, 2016).

10. Shen, D., Wu, G. & Suk, H.-I. Deep learning in medical image analysis. *Annual Review of Biomedical Engineering* **19**, 221–248 (2017). URL https://doi.org/10.1146/annurev-bioeng-071516-044442. PMID: 28301734, https://doi.org/10.1146/annurev-bioeng-071516-044442.

11. Litjens, G. *et al.* A survey on deep learning in medical image analysis. *Medical image analysis* **42**, 60–88 (2017).

12. Kingma, D. P. & Ba, J. Adam: A Method for Stochastic Optimization. *arXiv.org* (2014). 1412.6980v9.

13. Farnebäck, G. Two-frame motion estimation based on polynomial expansion. In *Proceedings of the 13th Scandinavian Conference on Image Analysis*, SCIA'03, 363–370 (Springer-Verlag, Berlin, Heidelberg, 2003).

14. Itseez. Open source computer vision library. https://github.com/itseez/opencv (2015).

Acknowledgment We thank the members of the Allen Institute for Cell Science team, who generated and characterized the gene-edited hiPS cell lines, developed image-based assays, and recorded the high replicate data sets suitable for modeling and without whom this work would not have been possible. We especially thank the Allen Institute for Cell Science Microscopy and Pipeline teams for providing images of different transmitted-light imaging modalities, and the Allen Institute for Cell Science Assay Development team and particularly Melissa Hendershott and Nathalie Gaudreault for providing the additional time series and CAAX datasets used here. We would like to thank the Allen Institute for Cell Science Animated Cell team and Thao Do specifically for providing her expertise in figure preparation. We thank Daniel Fernandes for developing an early proof of concept 2D version of the model. We would like to thank members of the Allen Institute for Brain Science Synapse Biology department for preparing samples and providing images that were the basis for training the conjugate array tomography data. These contributions were absolutely critical for model development. We thank Google Accelerated Science for telling us about their studies of 2D deep learning in neurons before beginning this project. This work was supported by grants from NIH/NINDS (R01NS092474) and NIH/NIMH (R01MH104227).

We thank Paul G. Allen, founder of the Allen Institute for Cell Science, for his vision, encouragement and support.

Competing Interests The authors declare that they have no competing financial interests.

Correspondence Correspondence and requests for materials should be addressed to gregj@alleninstitute.edu.

Author Contributions GRJ conceived the project. CO implemented the model for 2D and 3D images. MM provided guidance and support. CO, SS, FC, and GRJ designed computational experiments. CO, SS, MM, FC, and GRJ wrote the paper.

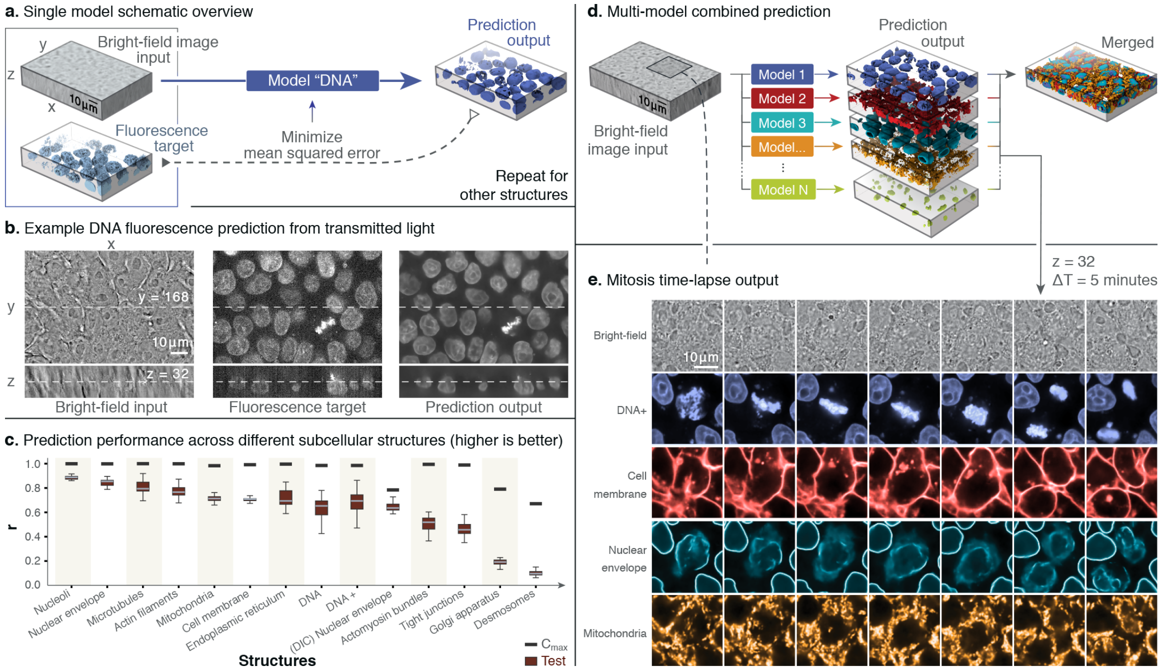


Figure 1: Label-free imaging tool pipeline. a) Given the input of transmitted light and fluorescence image pairs, the model is trained to minimize the mean squared error (MSE) between the fluorescence ground-truth and output of the model. b) Example of a 3D input transmitted light image, a ground-truth confocal DNA fluorescence image, and a tool prediction. c) Distributions of the image-wise correlation coefficient (r) between target and predicted test images from models trained on images for the indicated subcellular structure, plotted as a box across 25th, 50th and 75th percentile, with whiskers indicating the last data points within 1.5 × interquartile range of the lower and upper quartiles. For a complete description of structure labels, see Methods. Black bars indicate maximum correlation between the target image and a theoretical, noise-free image (C_{max}; details of metric in Methods). d) Different models applied to the same input and combined to predict multiple structures. e) Predicted localization of DNA (blue), cell membrane (red), nuclear envelope (cyan) and mitochondria (orange) of a sample using bright-field input images taken at 5-minute intervals. The center z-slice is shown. A mitotic event, along with stereotypical reorganization of subcellular structures, is clearly observed. All results are independent from training data except where explicitly labeled.

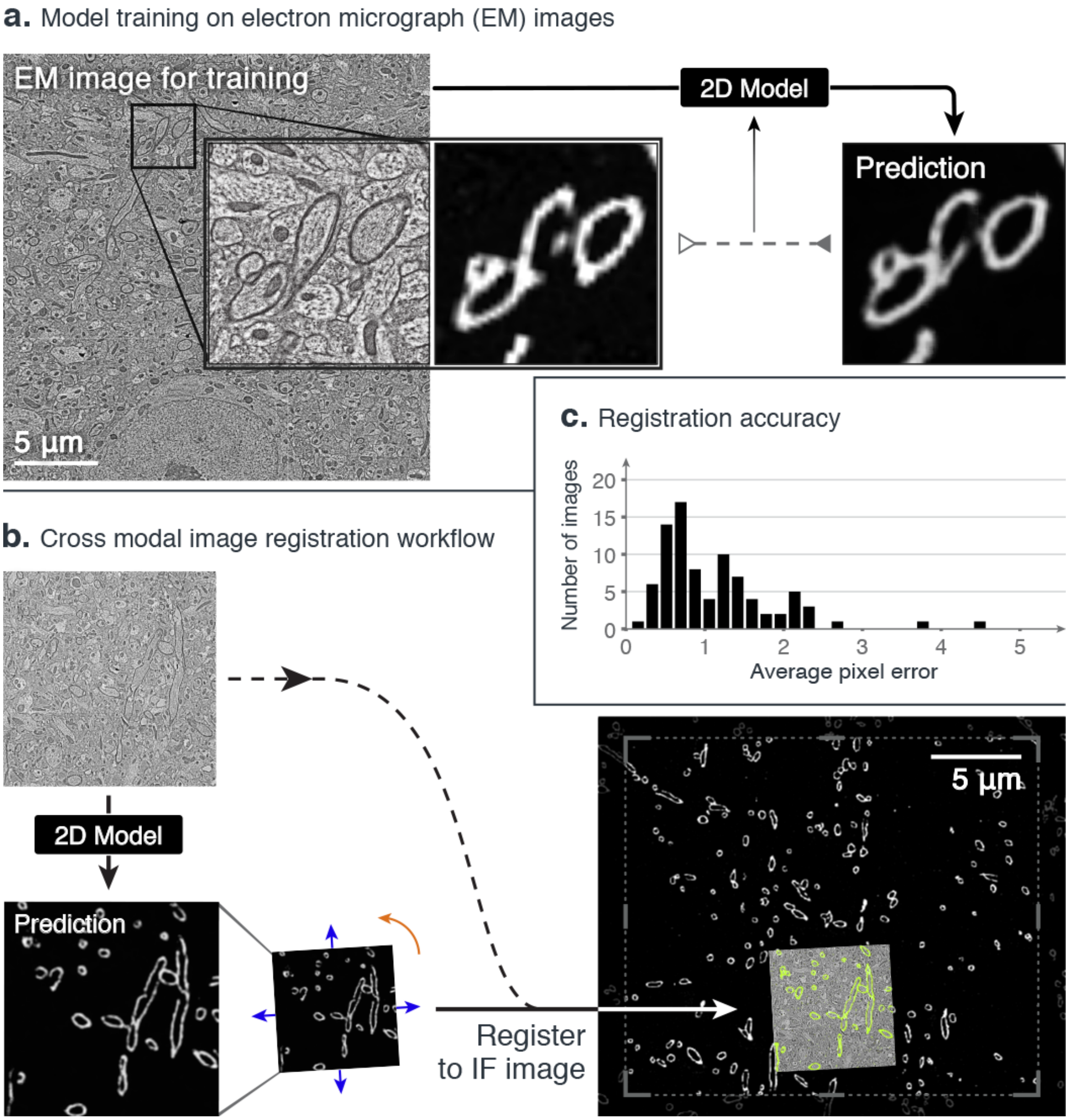
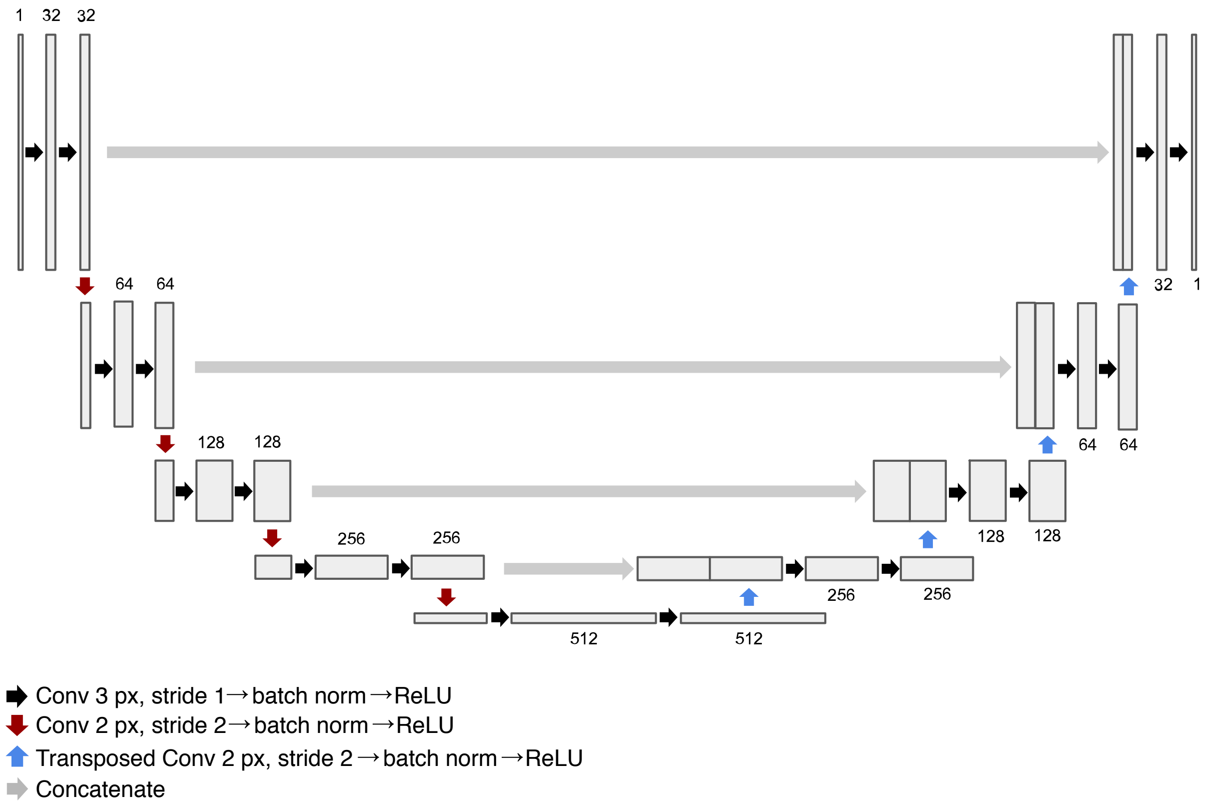
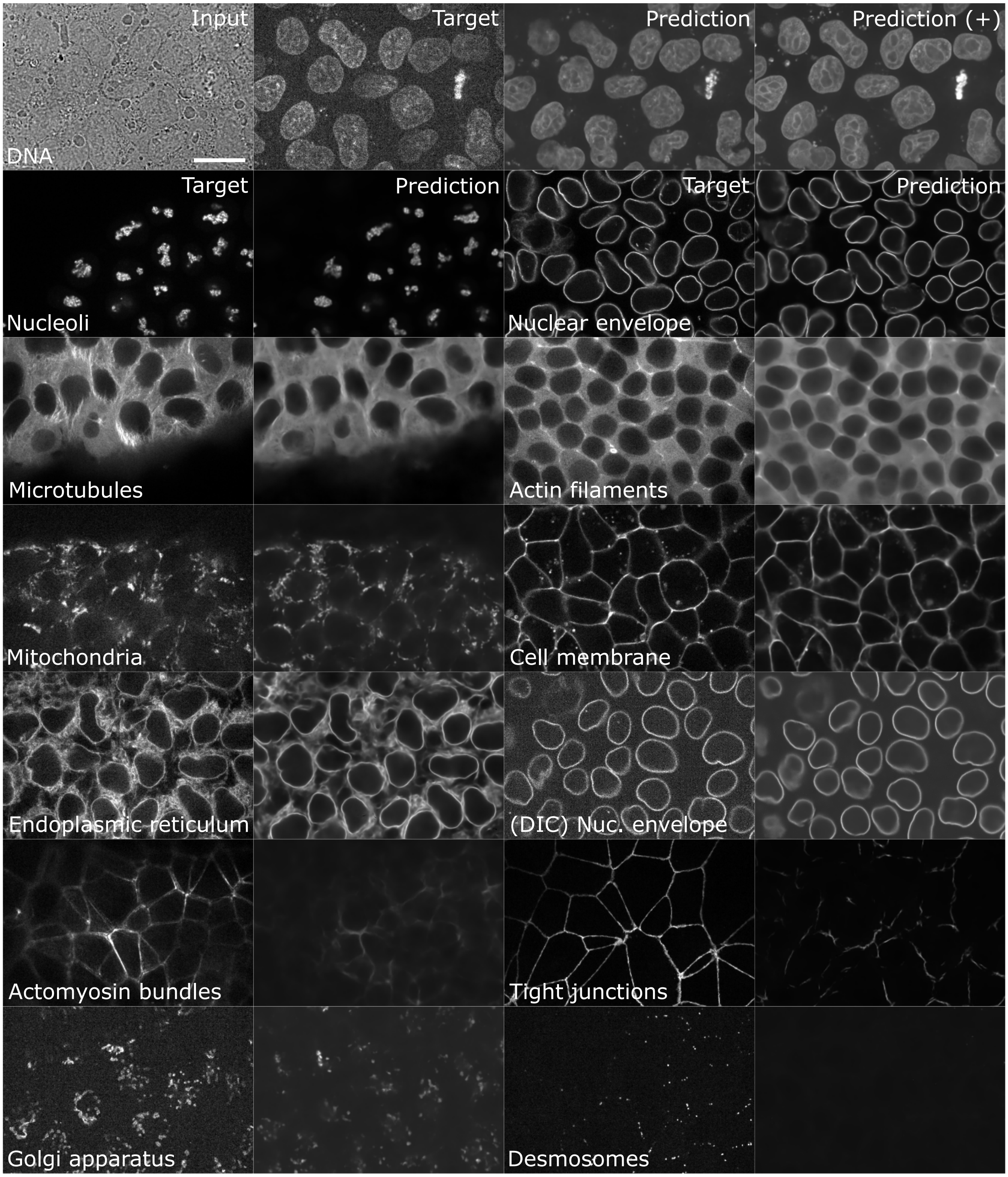


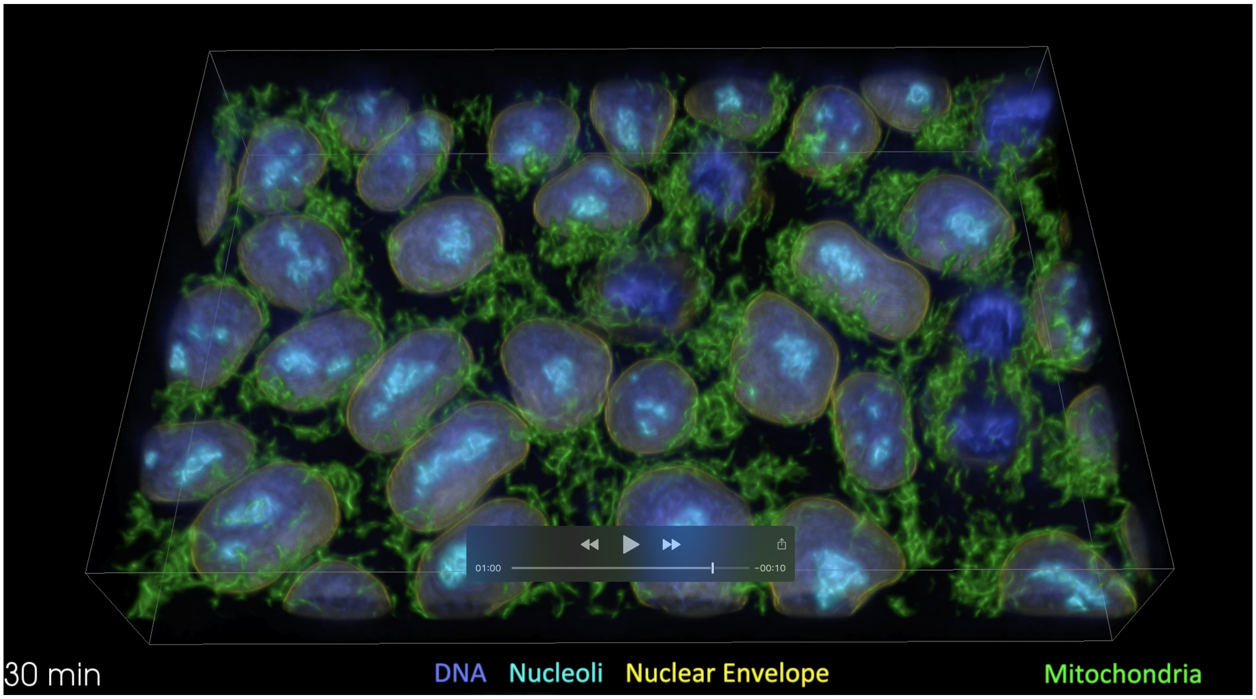
Figure 2: Automating registration across imaging modalities. a) Electron micrographs are manually registered to myelin basic protein immunofluorescence (MBF IF) images, to produce training data for a 2D model that can then predict MBP IF directly from electron micrographs. b) The trained 2D model was subsequently used in an automated registration workflow. Model predictions were registered via a similarity transformation to MBP IF images by searching with conventional automated computer vision techniques (see Methods for details). The figure shows only a 20 µm × 20 µm region from the 200 µm × 200 µm MBP-IF search image. c) Histogram of average distance between automated registration and manual registration as measured across 90 test images, in units of pixels of MBP IF data. This distribution has an average of 1.16 ± 0.79 px, where manual registrations between two independent annotators differed by 0.35 ± 0.2 px.


1
32
32
64
64
128
128
256
256
512
512
256
256
128
128
64
64
32
1
➡ Conv 3 px, stride 1→batch norm→ReLU
⬇ Conv 2 px, stride 2→batch norm→ReLU
⬆ Transposed Conv 2 px, stride 2→batch norm→ReLU
➡ Concatenate

Supplementary Figure 1: Diagram of CNN architecture underpinning presented tool. There are no batch normalization or ReLU layers on the last layer of the network, and the number of output channels per layer is shown above the box of each layer. Figured adapted from Ronneberger *et al.*


Input
Target
Prediction
Prediction (+)
DNA
Target
Prediction
Target
Prediction
Nucleoli
Nuclear envelope
Microtubules
Actin filaments
Mitochondria
Cell membrane
Endoplasmic reticulum
(DIC) Nuc. envelope
Actomyosin bundles
Tight junctions
Golgi apparatus
Desmosomes

Supplementary Figure 2: Additional labeled structure models and predictions for 3D light microscopy. The top row shows results for the tool trained to predict DNA fluorescence images (as further described in methods). From left, a single z-slice of a 3D transmitted light input image; a ground-truth ("target", observed) fluorescence image; an image predicted by the DNA model under standard training (as described in methods); and an image predicted by an extended version of the DNA model (DNA+). The following 6 rows below are divided into two columns, each with paired images of a correspondingly labeled structure. In each column, leftmost images show a single z-slice of a ground-truth ("target", observed) fluorescence image for the labeled structure, while images on right reveal an image predicted by the structure model under standard training (as described in methods). From 2nd row left, moving across columns and down rows, these structure models are presented by performance (as detailed in methods and as can be seen in Figure 1c): nucleoli, nuclear envelope, microtubules, actin filaments, mitochondria, cell membrane, endoplasmic reticulum, nuclear envelope (DIC), actomyosin bundles, tight junctions, Golgi apparatus, and desmosomes. All models trained on and used bright-field images as inputs (not shown), except where noted (nuclear envelope, DIC). Z-slices were chosen in a curated fashion so as to highlight the structure of interest associated with each model. Image-slice pairs were identically contrast stretched, such that black and white values corresponded to the 0.1 and 99.9th percentiles of the target image intensity, respectively. All images shown are independent from model training data. Scale bar is 20 µm.



Supplemental Video 1: 3D rendering of light microscopy prediction results. Movie illustrates the relationship between 3D time lapse transmitted light input and multiple prediction images. First, individual z-plane images from a 3D transmitted light are shown in succession. Next, individual predictions are shown overlaid in color in the following order: DNA (blue), nucleoli (cyan), nuclear envelope (yellow), cell membrane (magenta), and mitochondria (green). Next, a composite rendering of all channels is shown, followed by a time lapse of single plane from the dataset shown in Fig. 1e. Finally, a volumetric 3D rendering is shown and played through the individual timepoints 4 times, alternating between showing mitocondria and membrane, along with the nuclear related structures. The boxed outline depicts the extent of the field of view of this volume which encompasses 97 µm × 65 µm × 19 µm.