



Published in final edited form as:

*Nat Methods*. 2018 November ; 15(11): 917–920. doi:10.1038/s41592-018-0111-2.

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

Chawin Ounkomol<sup>1</sup>, Sharmishta Seshamani<sup>2</sup>, Mary M. Maleckar<sup>1</sup>, Forrest Collman<sup>2</sup>, and Gregory R. Johnson<sup>1,\*</sup>

<sup>1</sup>Allen Institute for Cell Science, Seattle, WA 98109

<sup>2</sup>Allen Institute for Brain Science, Seattle, WA 98109

### Abstract

Understanding cells as integrated systems is a challenge central to modern biology. The different microscopy approaches used to probe biological organization each present limitations, ultimately restricting insight into unified cellular processes. Fluorescence microscopy can resolve subcellular structure in living cells, but is expensive, slow, and toxic. Here, we present a label-free method for predicting 3D fluorescence directly from transmitted light images and demonstrate its use to generate multi-structure, integrated images.

Imaging methods currently used to capture details of cellular organization all present restrictions with respect to expense, spatio-temporal resolution, and sample perturbation. Fluorescence microscopy permits imaging of structures of interest by specific labeling, but requires advanced instrumentation and time-consuming sample preparation. Significant phototoxicity and photobleaching perturb samples, creating a tradeoff between data quality and timescales available for live cell imaging<sup>1,2</sup>. Furthermore, the number of simultaneous fluorescent tags is restricted by both spectrum saturation and cell health, limiting the number of parallel labels for joint imaging. In contrast, transmitted light microscopy (TL), e.g., bright-field, phase, DIC, is relatively low-cost, and is label-free (greatly reduced phototoxicity<sup>3</sup> and simplified sample preparation). Although valuable information about cellular organization is apparent in TL images, these lack the clear contrast of fluorescence labeling, a limitation also present in other widely-used microscopy modalities. Electron micrographs also contain a rich set of biological detail about subcellular structure, but often require tedious expert interpretation. A method combining the clarity of fluorescence microscopy with the relative simplicity and modest cost of other imaging techniques would present a groundbreaking tool for biological insight into the integrated organization of subcellular structures.

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:[http://www.nature.com/authors/editorial\\_policies/license.html#terms](http://www.nature.com/authors/editorial_policies/license.html#terms)

\*Correspondence and requests for materials should be addressed to [gregj@alleninstitute.org](mailto:gregj@alleninstitute.org).

**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.

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

Convolutional neural networks (CNNs) capture non-linear relationships over large areas of images, resulting in vastly improved performance for image recognition tasks as compared to classical machine learning methods. Here, we present a CNN-based tool, employing a U-Net architecture<sup>4</sup> (Supplementary Fig. 1, Methods) to model relationships between distinct but correlated imaging modalities, and show the efficacy of this tool for predicting corresponding fluorescence images directly from both 3D TL live cell images and 2D electron micrographs, alone.

The label-free prediction tool learns each relationship between 3D TL and fluorescence live cell images for several major subcellular structures (Fig. 1a, b, c, e.g. cell membrane, DNA, etc.). A resultant model can then predict a 3D fluorescence image from a new TL input. A single TL input can be applied to multiple subcellular structure models, enabling multi-channel, integrated fluorescence imaging (Fig. 1d, e). The method can similarly be used to predict 2D immunofluorescence (IF) images directly from electron micrographs (EM) to highlight distinct subcellular structures and to register conjugate multi-channel fluorescence data with EM<sup>5</sup> (Fig. 2).

In our experiments, we used only spatially registered pairs of images from a relatively small set to train each structure model (30 pairs per structure for 3D TL-to-fluorescence, and 40 for 2D EM-to-fluorescence; Methods). Biological detail observed in predictions varies among subcellular structures modeled; however, in the case of the 3D TL-to-fluorescence models, predictions appear structurally similar to ground truth fluorescence. Nuclear structures are well-resolved, for example: 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 high-resolution localization and 3D morphology. The nucleoli model also resolves the precise location and morphology of individual nucleoli (Supplementary Fig. 2).

TL-to-fluorescence models' performance was quantified by Pearson correlation coefficient on 20 predicted and corresponding ground truth fluorescence image pairs (Fig. 1c) from independent test sets for each model (Methods). A theoretical upper bound for model performance based upon an estimate of signal-to-noise ratio (SNR) of fluorescence images used for training was determined (Methods). Model performance for each structure is well-bounded by this limit (Fig. 1c).

We trained a model predicting DNA with an extended procedure ("DNA+"; Methods) to evaluate whether predictions improve with additional training images and iterations. Outcome improved as measured by an increase in Pearson correlation, and images qualitatively showed better clarity of sub-nuclear structure and precision of predictions around mitotic cells (Fig. 1c, Supplementary Fig. 2, Methods). Most critically, these details can be observed together in a 3D integrated multi-channel prediction derived from a single TL image (Fig. 1e and Supplementary Video). Examples for all fourteen labeled structure models' predictions on each model's test set can be found in Supplementary Fig. 2.

Transforming one imaging modality into another also can be useful in less direct ways: 2D IF images predicted from EM (Fig. 2) can be used to facilitate automatic registration of conjugate multi-channel fluorescence data with EM. Array tomography data<sup>5</sup> of ultrathin brain sections uses EM and ten channels of IF images (including the structure myelin basic protein, MBP-IF) from the same sample, but from two different microscopes. Thus EM and IF images are not natively spatially aligned. While EM and corresponding images from other modalities can be registered by hand, resulting in multi-channel conjugate EM images<sup>5,6,7</sup>, manual registration is tedious and time-consuming. We trained a 2D version of the label-free tool on manually registered pairs of EM and MBP-IF images and then used model predictions to register an EM image ( $15\ \mu\text{m} \times 15\ \mu\text{m}$ ) to a much larger target MBP-IF image ( $204.8\ \mu\text{m} \times 204.8\ \mu\text{m}$ ) (Fig. 2a). Test EM images were first input to the model to predict corresponding MBP-IF images (Fig. 2a) and conventional intensity-based matching techniques (Methods) were then used to register each MBP-IF prediction (and EM image) to the target MBP-IF image (Fig. 2b; successful convergence on 86 of 90 image pairs). The average distance between automated and ground truth registration was measured to be  $1.16 \pm 0.79$  px, (MBP-IF pixel data units). To the authors' knowledge, this is the first successful attempt to automate this registration process via a learning based technique, suggesting the label-free tool's utility can be extended to diverse imaging modalities and a variety of downstream image processing challenges.

We next determined that individual structure models, trained solely on static images, can be used to predict fluorescence time-lapse by applying several subcellular structure TL-to-fluorescence models to a single TL 3D time-series (covering 95 minutes at 5-minute intervals; Fig. 1e, Supplementary Video 1). In addition to simultaneous structure visualization, characteristic dynamics of mitotic events, including reorganization of the nuclear envelope and cell membrane, are evident in the predicted multi-channel time-series (Fig. 1e). Time-series acquired with a similar 5-minute acquisition interval but with three-channel spinning disk fluorescence reveal both obvious bleaching artifacts and changes in cellular morphology and health after 10–15 minutes (data not shown). The phototoxicity occurring in extended, multi-label live cell time-series fluorescence imaging on the hiPSCs used here evidences challenges in obtaining integrated structural information from fluorescence time lapse imaging. While many strategies exist to minimize this photodamage (i.e. oxygen scavenging, reduced laser power and exposure, advanced microscopy techniques<sup>2</sup>, machine learning driven denoising<sup>8</sup>), all present compromises with respect to ease, image quality, and fidelity. This method avoids trade-offs and directly produces time-series predictions for which no fluorescence imaging ground truth exists, greatly increasing the timescales over which some cellular processes can be visualized and measured.

Our method has inherent limitations. Models must learn a relationship between distinct but correlated imaging modes; predictive performance is contingent upon this association existing. In the case of desmosomes or actomyosin bundles, for example, model performance for the presented training protocol was poor, presumably due to a weaker association between TL and fluorescence images of these structures (Fig. 1c, Supplementary Fig. 2). Quality and quantity of training data also influences accuracy of model predictions, although this relationship is highly nonlinear in tested cases (for DNA model performance, we see diminishing returns between 30 and 60 images; Supplementary Fig. 3). Performance

between models varies with 2D and 3D information: using a 2D DNA model to predict z-slices selected from 3D images shows artifacts between predicted z-slices and decreased correlation between ground truth and predictions (Supplementary Fig. 4; Methods), suggesting that 3D interference patterns are valuable for predicting subcellular organization.

Additionally, we cannot assess *a priori* how models will perform in contexts for which there are very few or no examples in training or testing data. Models pre-trained using one cell type (i.e. hiPSC) do not perform as well with inputs of drastically different cellular morphologies (Supplementary Fig. 5). We compared predictions from the DNA+ model (trained on hiPSC images) to those from a model trained on images of DNA-labeled HEK-293 kidney-phenotype cells<sup>9</sup> (applied to both hiPSC and HEK-293 test images). While gross image features are comparable, prediction performance of morphological detail improves markedly when the model is trained on and applied to data of the same cell type. A similar reduction in predictive performance is evident when pre-trained models are used to predict a fluorescent DNA label in different cell types, like cardiomyocytes or HT-1080 fibroblast-phenotype cells<sup>10</sup> (Supplementary Fig. 5, Methods).

Furthermore, predictions from inputs acquired with imaging parameters identical to those used to compose models' training sets will provide the most accurate results versus ground truth data. For example we observed decreased model accuracy when predicting fluorescence images from input TL stacks acquired with a shorter inter-slice interval (~0.13 s) than that in training data (~2.2 s) (not shown). Ultimately, when evaluating the utility of predicted images, context of use must be considered. For instance, DNA or nuclear membrane predictions may have sufficient accuracy for application to downstream nuclear segmentation algorithms, but microtubule predictions would not be effective for assaying rates of microtubule polymerization (Fig. 1e, Supplementary Fig. 2). Finally, there may not be a direct quantitative link between the predicted intensity of a tagged structure and protein levels.

The presented methodology has wide potential use in many biological imaging fields. Primarily, it may reduce or even eliminate routine capture of some images in existing imaging and analysis pipelines, permitting similar throughput in a more efficient, cost-effective manner. Notably, training data requires no manual annotation, little to no pre-processing, and relatively small numbers of paired examples, drastically reducing the barrier to entry associated with some machine learning approaches. This approach may have particular value in image-based screens where cellular phenotypes can be detected via expressed fluorescent labels<sup>11</sup>, pathology workflows requiring specialized labels that identify specific tissues<sup>12</sup>, and long time-series observation of single cells<sup>1</sup>, tissues, or organism-level populations where more expensive instrumentation is not available<sup>2</sup>. Recent related work convincingly demonstrates that 2D whole-cell antibody stains can be predicted from TL<sup>13</sup>, supporting the conclusion that similar techniques can be applied to a wide variety of biological studies, as demonstrated here by automatic registration of conjugate multi-channel fluorescence data with EM. The method is additionally promising when generating a complete set of simultaneous ground-truth labels is infeasible, e.g. live cell time-series imaging. Finally, our tool permits the generation of integrated images by which multi-dimensional 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 suggests 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.

## Online Methods

### 3D Live Cell Microscopy

The 3D light microscopy data used to train and test the presented models consists of z-stacks of colonies of human embryonic kidney cells (HEK-293)<sup>9</sup>, human fibrosarcoma cells (HT-1080)<sup>10</sup>, genome-edited human induced pluripotent stem cell (hiPSC) lines<sup>10</sup> expressing a protein endogenously tagged with either mEGFP or mTagRFP that localizes to a particular subcellular structure<sup>14</sup>, and hiPSC-derived cardiomyocytes differentiated from the former. 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.

**Imaging**—All cell types were imaged for up to 2.5 h on a Zeiss spinning disk microscope with ZEN Blue 2.3 software and with a 1.25-NA, 100x objective (Zeiss C-Apochromat 100x/1.25 W Corr), with up to four, 16-bit 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 ms exposure; 638 nm laser at 2.4 mW, 200 ms exposure; 405 nm at 0.28 mW, 250 ms exposure; 488 nm laser at 2.3 mW, 200 ms exposure. The exception was CAAX-RFP-based cell membrane images, which were acquired with a 1.2-NA, 63x objective (Zeiss C-Apochromat 63x/1.2 W Corr), a 561 nm laser at 2.4 mW, and a 200 ms exposure. 100x-objective z-slice images were captured at a YX-resolution of 624 px × 924 px with a pixel scale of 0.108 μm/px, and 63x-objective z-slice images were captured at a YX-resolution of 1248 px × 1848 px with a pixel scale of 0.086 μm/px. All z-stacks were composed of 50 to 75 z slices with an inter-z-slice interval of 0.29 μm. Images of cardiomyocytes contained 1 to 5 cells per image whereas images of other cell types contained 10 to 30 cells per image. Time-series data were acquired using the same imaging protocol as for acquisition of training data but on unlabeled, wild-type hiPSCs at 5 minute intervals for 95 minutes, with all laser powers set to zero to reproduce the inter-z-slice timing of the training images.

**Tissue Culture**—hiPSCs, HEK-293 cells, or HT-1080 cells were seeded onto Matrigel-coated 96-well plates at densities specified below. The cells were stained on the days they were to be imaged, first by incubation in their imaging media with 1x NucBlue (Hoechst 33342, ThermoFisher) for 20 min. hiPSCs were then incubated in imaging media with 1x NucBlue and 3x CellMask (ThermoFisher) for 10 min, whereas HEK-293 and HT-1080

cells were then incubated in imaging media with 1x NucBlue and 0.5x CellMask for 5 min. The cells were washed with fresh imaging media before imaging.

For hiPSCs, the culture media was mTeSR1 (Stem Cell Technologies) with 1% Pen-Strep. The imaging media was phenol-red-free mTeSR1 with 1% Pen-Strep. Cells were seeded at a density of ~2500 cells per well and were imaged 4 days after initial plating. For HEK-293 cells, the culture media was DMEM with GlutaMAX (ThermoFisher), 4.5 g/L D-Glucose, 10% FBS, and 1% antibiotic-antimycotic. The imaging media was phenol-red-free DMEM/F-12 with 10% FBS and 1% antibiotic-antimycotic. Cells were seeded at a density of 13 to 40 thousand cells per well and were imaged 1 to 2 days after initial plating. For HT-1080 cells, the culture media was DMEM with GlutaMAX, 15% FBS, and 1% Pen-Strep. The imaging media was phenol-red-free DMEM/F-12 with 10% FBS and 1% Pen-Strep. Cells were seeded at a density of 2.5 to 40 thousand cells per well and were imaged 4 days after initial plating.

CAAX-tagged hiPSCs were differentiated to cardiomyocyte phenotype by seeding onto Matrigel-coated six-well tissue culture plates at a density ranging from 0.15 to  $0.25 \times 10^6$  cells per well in mTeSR1 supplemented with 1% Pen-Strep, 10  $\mu$ M ROCK inhibitor (Stem Cell Technologies) (day -3). Cells were grown for 3 d with daily mTeSR1 changes. On day 0, we initiated differentiation by treating cultures with 7.5  $\mu$ M CHIR99021 (Cayman Chemical) in Roswell Park Memorial Institute (RPMI) media (Invitrogen) containing insulin-free B27 supplement (Invitrogen). After 2 d, cultures were treated with 7.5  $\mu$ M IWP2 (R&D Systems) in RPMI media with insulin-free B27 supplement. On day 4, cultures were treated with RPMI with insulin-free B27 supplement. From day 6 onward, media was replaced with RPMI media supplemented with B27 containing insulin (Invitrogen) every 2–3 d. Cardiomyocytes were re-plated at day 12 onto glass-bottom plates coated with PEI/laminin and were imaged on day 43 after initiation of differentiation. The imaging media was phenol-red-free RPMI with B27. Prior to imaging, cells were stained by incubation in imaging media with Nuclear Violet (AAT Bioquest) at a 1/7,500 dilution and  $\times 1$  CellMask for 1 min and then washed with fresh imaging media.

**Data for training and evaluation**—Supplementary Table 1 outlines the data used to train and evaluate the models based on 3D live cell z-stacks, including train-test data splits. All multi-channel z-stacks were obtained from a database of images produced by the Allen Institute for Cell Science’s microscopy pipeline (see <http://www.allencell.org>). For each of the 11 hiPSC cell lines, we randomly selected z-stacks from the database and paired the transmitted light channel with the EGFP/RFP channel to train and evaluate models (Fig. 1c) to predict the localization of the tagged subcellular structure. The transmitted light channel modality was bright-field for all but the DIC-to-nuclear envelope model. For the DNA model data, we randomly selected 50 z-stacks from the combined pool all bright-field-based z-stacks and paired the transmitted light channel with the Hoechst channel. The training set for the DNA+ model was further expanded to 540 z-stacks with additional images from the Allen Institute for Cell Science’s database. Note that while a CellMask channel was available for all z-stacks, we did not use this channel because the CAAX-membrane cell line provided higher quality images for training cell membrane models. A single z-stack time series of wild-type hiPSCs was used only for evaluation (Fig. 1e).



For experiments testing the effects of number of training images on model performance (Supplementary Fig. 3), we supplemented each model's training set with additional z-stacks from the database. Z-stacks of HEK-293 cells were used to train and evaluate DNA models whereas all z-stacks of cardiomyocytes and of HT-1080 cells were used only for evaluation (Supplementary Fig. 5). The 2D DNA model (Supplementary Fig. 4) used the same data as the DNA+ model.

All z-stacks were converted to floating-point and were resized via cubic interpolation such that each voxel corresponded to  $0.29 \mu\text{m} \times 0.29 \mu\text{m} \times 0.29 \mu\text{m}$ , and resulting images were  $244 \text{ px} \times 366 \text{ px}$  for 100x-objective images or  $304 \text{ px} \times 496 \text{ px}$  for 63x-objective images in Y and X respectively and between 50 and 75 pixels in Z. Pixel intensities of all input and target images were z-scored on a per-image basis to normalize any systematic differences in illumination intensity.

## Electron and Immunofluorescence Microscopy

### Imaging

For conjugate array tomography data<sup>5</sup>, 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 nm per pixel. 5 small regions were then imaged with a field emission scanning electron microscope to obtain high resolution electron micrographs at 3 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<sup>15</sup>.

### Data Used for Training and Evaluation

40 pairs of registered EM and MBP montages were resampled to 10 nm per pixel. For each montage pair, a central region of size  $3280 \text{ px} \times 3214 \text{ px}$  was cut out and used for the resultant final training set. This corresponded to the central region of the montage which contained no unimaged regions across the sections used. Pixel intensities of the images were z-scored. For the registration task, a total of 1500 EM images (without montaging) were used as an input to directly register to the corresponding larger MBP image in which it lies. For this, each EM image was first downsampled to 10 nm per pixel without any transformations to generate a  $1500 \text{ px} \times 1500 \text{ px}$  image.

### Model Architecture Description and Training

We employed a convolutional neural network (CNN) based on the U-Net architecture<sup>4</sup> (Supplementary Fig. 1) due to its demonstrated performance in image segmentation and tracking tasks. In general, CNNs are uniquely powerful for image-related tasks (classification, segmentation, image-to-image regression) due to the fact that they are image-translation invariant, learn complex non-linear relationships across multiple spatial areas, circumvent the need to engineer data-specific feature extraction pipelines, and are straightforward to implement and train. CNNs have been shown to outperform other state-of-the-art models in basic image recognition<sup>16</sup> and have been used in biomedical imaging

for a wide range of tasks including image classification, object segmentation<sup>17</sup>, and estimation of image transformations<sup>18</sup>. Our U-Net variant 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 the 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 ( 64 px × 64 px × 32 px, YXZ) for light microscopy data or on 2D patches (256 px × 256 px) 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<sup>19</sup> with a learning rate of 0.001 and with beta values of 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 (205 hours for DNA+) 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 correlation coefficient,  $r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$ , between the pixel intensities of the model's output,  $y$ , and independent ground truth test images,  $xx$  (Fig. 1c, Supplemental Figures 3, 4b, 5b, 6). To estimate the theoretical upper bound on the performance of a model, we calculated the correlation between a theoretical model which is able to perfectly predict the spatial fluctuations of the signal but is unable to predict the random fluctuations in the target image that arise from fundamentally unpredictable phenomena (such as noise in the electronics of the camera or fluctuations in number of photons collected from a fluorescent molecule). Intuitively as the relative size of random fluctuations increases relative to the size of predictable signal, one would expect the performance of even a perfect model to degrade. The images in Figure 1 of DNA-labeled targets and predictions make this point, in so far as the model can not be expected to predict the background noise in the DNA-labeled imagery. Therefore, to estimate a lower bound on the amplitude of the random fluctuations we analyzed images of cells that were taken with identical imaging conditions but contained no fluorescent labels, for example, images taken with microscope settings designed to detect



Hoechst staining, but with cells for which there was no Hoechst dye applied, or images taken with microscope settings designed to detect GFP but with cells with no GFP present. We used the variance of pixel intensities across the image as an estimate of the variance of random fluctuations ( $N$ ), and then averaged that variance across control images in order to arrive at our final estimate. Calculating the correlation between a perfect model prediction  $S$  (equal to the predictable image) and an image  $T$  which is the combination of the predictable image and the random fluctuations ( $T_{x,y,z} = N_{x,y,z} + S_{x,y,z}$ ), is  $C_{max} = \sqrt{\frac{SNR}{1+SNR}}$  where  $SNR = \frac{\langle S^2 \rangle}{\langle N^2 \rangle}$ . If we assume the correlation between the predictable component and the random component is zero, then the variance of the predictable image ( $\langle S^2 \rangle$ ) can be calculated by taking the variance of the measured image ( $\langle T^2 \rangle$ ) and subtracting the variance of the random fluctuations ( $\langle N^2 \rangle$ ). The result is a formula for the theoretical upper bound of model performance which depends only on the lower bound estimate of the variance of the noise, and the variance of the measured image. We report the average value of  $C_{max}$  for all images in the collection as black tick marks in Figure 1c.

### Registration across imaging modalities

We employed a 2D version of our tool trained on the montage pairs described below. Electron microscopy (EM) images were reflection padded to 1504 px  $\times$  1504 px, passed through the trained model, and then predictions were cropped back to the original input size to generate a myelin basic protein (MBP) prediction image. This MBP prediction image was first roughly registered to the larger MBP IF images using cross-correlation-based template matching for a rigid transformation estimate. Next, the residual optical flow<sup>20</sup> 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 ([www.opencv.org](http://www.opencv.org)). 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. Since the images were registered using a similarity transformation where it is possible for the registration accuracy of the central pixels and those at the edges to be different, the registration errors were calculated by computing the average difference in displacement across an image, as measured in pixels of the target IF image. We report these results for registration differences (Fig. 2) between authors and between the algorithm estimate and one of the authors.

### 3D fluorescence image predictions from a 2D model

To compare performance between models trained on 2D and 3D data, we trained a 2D DNA model for evaluation against the DNA+ model. The 2D model, was trained on the same dataset with the same training parameters as the DNA+ with the exception that training patches of size 64 px  $\times$  64 px were sampled from random z-slices of the 3D training images. The model was trained for 250,000 mini-batch iterations with a batch size of 24 for a total training time of approximately 18 hours. After training, 3D predicted fluorescence images

were formed by inputting sequential 2D bright-field z-slices into the model and combining the outputs into 3D volumes (Supplementary Fig. 4).

### Life Sciences Reporting Summary

The Life Sciences Reporting Summary can be found along with the supplementary material.

**Software and Data**—Software for training models is available at [https://github.com/AllenCellModeling/pytorch\\_fnet/tree/release\\_1](https://github.com/AllenCellModeling/pytorch_fnet/tree/release_1). Data used to train the 3D models is available at <https://downloads.allencell.org/publication-data/label-free-prediction/index.html>.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

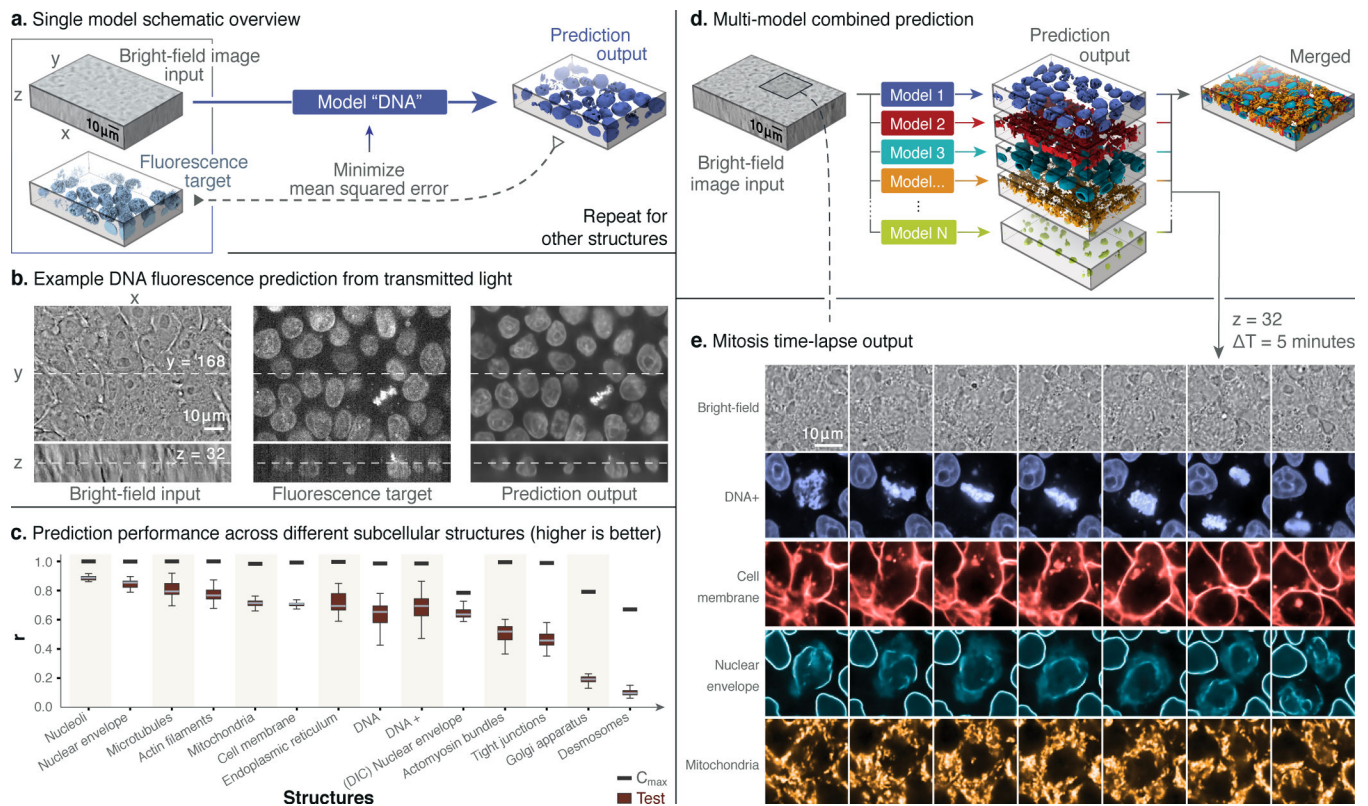
### Acknowledgments

We thank the entire 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 Gene Editing, Assay Development, Microscopy, and Pipeline teams for providing cell lines and images of different transmitted-light imaging modalities, and particularly Kaytlyn Gerbin, Angel Nelson, and Haseeb Malik for performing the cardiomyocyte differentiation and culture, Winnie Leung, Joyce Tang, Melissa Hendershott and Nathalie Gaudreault for gathering the additional time series, CAAX-labeled, cardiomyocyte, HEK-293, and HT-1080 data. 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. HEK-293 cells were provided via the Viral Technology Lab at the Allen Institute for Brain Science. Cardiomyocyte and hiPSC data in this publication were derived from cells in the Allen Cell Collection, a collection of fluorescently labeled hiPSCs derived from the parental WTC line provided by Bruce R. Conklin, at Gladstone Institutes. 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) (SS, FC) and NIH/NIMH (R01MH104227) (FC). We thank Paul G. Allen, founder of the Allen Institute for Cell Science, for his vision, encouragement and support.

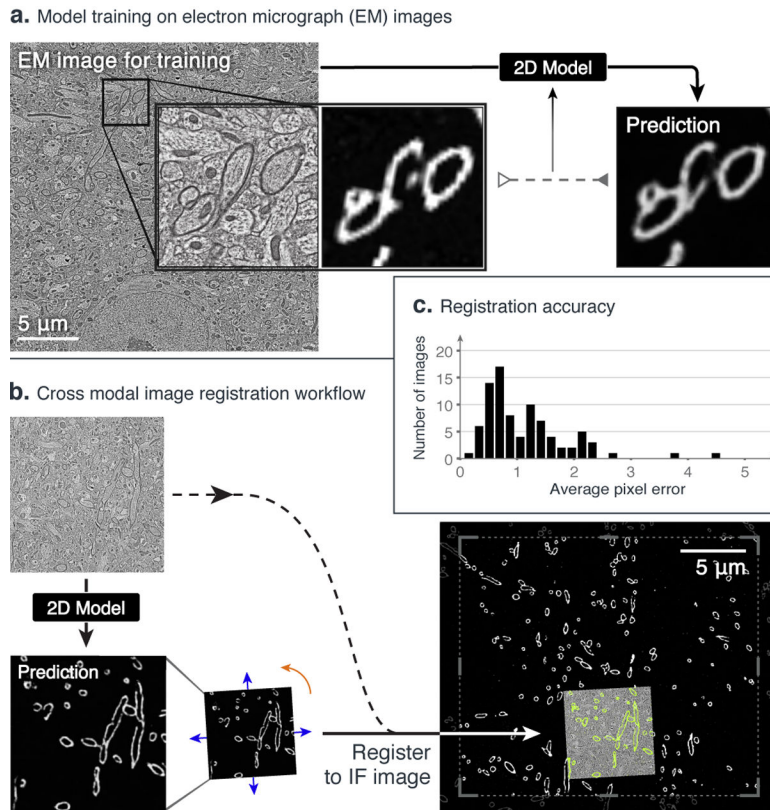
### References

1. Skylaki S, Hilsenbeck O & Schroeder T Challenges in long-term imaging and quantification of single-cell dynamics. *Nat. Biotechnol* 34, 1137–1144 (2016). [PubMed: 27824848]
2. Chen B-C et al. Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution. *Science* 346, 1257998 (2014). [PubMed: 25342811]
3. Selinummi J et al. Bright field microscopy as an alternative to whole cell fluorescence in automated analysis of macrophage images. *PLoS One* 4, e7497 (2009). [PubMed: 19847301]
4. Ronneberger O, Fischer P & Brox T U-Net: Convolutional Networks for Biomedical Image Segmentation. in *Lecture Notes in Computer Science* 234–241 (2015).
5. Collman F et al. Mapping Synapses by Conjugate Light-Electron Array Tomography. *Journal of Neuroscience* 35, 5792–5807 (2015). [PubMed: 25855189]
6. Russell MRG et al. 3D correlative light and electron microscopy of cultured cells using serial blockface scanning electron microscopy. *J. Cell Sci.* 130, 278–291 (2017). [PubMed: 27445312]
7. Kopek BG, Shtengel G, Grimm JB, Clayton DA & Hess HF Correlative photoactivated localization and scanning electron microscopy. *PLoS One* 8, e77209 (2013). [PubMed: 24204771]
8. Weigert M et al. Content-Aware Image Restoration: Pushing the Limits of Fluorescence Microscopy. bioRxiv 236463 (2018). doi:10.1101/236463

9. Graham FL, Smiley J, Russell WC & Nairn R Characteristics of a human cell line transformed by DNA from human adenovirus type 5. *J. Gen. Virol* 36, 59–74 (1977). [PubMed: 886304]
10. Rasheed S, Nelson-Rees WA, Toth EM, Arnstein P & Gardner MB Characterization of a newly derived human sarcoma cell line (HT-1080). *Cancer* 33, 1027–1033 (1974). [PubMed: 4132053]
11. Goshima G et al. Genes required for mitotic spindle assembly in *Drosophila* S2 cells. *Science* 316, 417–421 (2007). [PubMed: 17412918]
12. Gurcan MN et al. Histopathological Image Analysis: A Review. *IEEE Rev. Biomed. Eng* 2, 147–171 (2009). [PubMed: 20671804]
13. Christiansen EM et al. In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* 173, 792–803.e19 (2018). [PubMed: 29656897]
14. Roberts B et al. Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. *Mol. Biol. Cell* 28, 2854–2874 (2017). [PubMed: 28814507]
15. Cardona A et al. TrakEM2 software for neural circuit reconstruction. *PLoS One* 7, e38011 (2012). [PubMed: 22723842]
16. Krizhevsky A, Sutskever I & Hinton GE ImageNet classification with deep convolutional neural networks. *Commun. ACM* 60, 84–90 (2017).
17. Kevin Zhou S, Greenspan H & Shen D *Deep Learning for Medical Image Analysis*. (Academic Press, 2017).
18. Litjens G et al. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88 (2017). [PubMed: 28778026]
19. Kingma DP & Ba J Adam: A Method for Stochastic Optimization. *arXiv [cs.LG]* (2014).
20. Farnebäck G Two-Frame Motion Estimation Based on Polynomial Expansion. in *Lecture Notes in Computer Science* 363–370 (2003).

**Figure 1:**

Label-free imaging tool pipeline and application using 3D transmitted light-to-fluorescence models. a) Given transmitted light and fluorescence image pairs as input, the model is trained to minimize the mean squared error (MSE) between the fluorescence ground truth and output of the model. b) Left to right, an 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 Pearson correlation coefficient ( $r$ ) between ground truth (target) and predicted test images derived from the indicated subcellular structure models. Each target/predicted image pair in the test set is a point in the resultant  $r$  distribution; the 25th, 50th and 75th percentile image pairs are spanned by the box for each indicated structure, with whiskers indicating the last data points within the 1.5x interquartile range of the lower and upper quartiles. The number of images ( $n$ ) was 18 for the cell membrane, 10 for the DIC nuclear envelope, and 20 for all other distributions. For a complete description of the structure labels, see Methods. Black bars indicate maximum correlation between the target image and a theoretical, noise-free image ( $C_{\max}$ ; for details see Methods). d) Individual subcellular structure models are applied to the same input and combined to predict multiple structures. e) Localization of DNA (blue), cell membrane (red), nuclear envelope (cyan) and mitochondria (orange) as predicted for time lapse transmitted light (bright-field) input images taken at 5-minute intervals (center z-slice shown); a mitotic event with stereotypical reorganization of subcellular structures is clearly evident. Similar results were observed for two independent time-series input image sets. All results shown here are obtained from new transmitted light images not used during model training.

**Figure 2:**

Label-free imaging tool facilitates 2D automated registration across imaging modalities. We first train a model to predict a 2D myelin basic protein immunofluorescence image (MBP-IF) from a 2D electron micrograph (EM) and then register this prediction to automate cross-modal registration. a) An example EM image with a highlighted subregion (left), the MBP-IF image corresponding to the same subregion (middle), and the label-free imaging tool prediction of the same subregion given only the EM image as input (right). b) The EM image of the subregion to be registered (top left) is passed through the trained 2D model to obtain a prediction for the subregion (bottom left), which is then registered to MBP-IF images within a larger field of view (bottom right) (see Methods for details). Only a  $20\ \mu\text{m} \times 20\ \mu\text{m}$  region from the  $204.8\ \mu\text{m} \times 204.8\ \mu\text{m}$  MBP-IF search image is shown; predicted and registered MBP-IF are overlaid (in green) together with the EM 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  $1.16 \pm 0.79\ \text{px}$ , where manual registrations between two independent annotators differed by  $0.35 \pm 0.2\ \text{px}$ .