Practical Fluorescence Reconstruction Microscopy for High-Content Imaging

Julienne LaChance[1], Daniel J. Cohen[1]*

[1]Department of Mechanical and Aerospace Engineering, Princeton University

Correspondence should be sent to:
Daniel J. Cohen
danielcohen@princeton.edu
111 Hoyt Laboratory
Princeton University
Princeton, NJ 08544

Abstract

Fluorescence reconstruction microscopy (FRM) is an approach where transmitted light images are passed into a convolutional neural network which then outputs predicted epifluorescence images. This approach enables many benefits including reduced phototoxicity, freeing up of fluorescence channels, simplified sample preparation, and the ability to re-process legacy data for new insights. However, current FRM benchmarks are single scores that are difficult to relate to how useful for trustworthy and FRM predicition is. Here, we relate the conventional benchmark to practical and familiar cell biology analyses to demonstrate that FRM should be judged in context. We further demonstrate that it performs remarkably well even with lower-magnification microscopy data, as are often collected in high content imaging. Specifically, we present promising results for nuclei, cell-cell junctions, and fine feature reconstruction; provide experimental design guidelines; and provide the code, sample data, and user manual to enable more widespread adoption of FRM.

Introduction

Deep learning holds enormous promise for biological microscopy data, and offers especially exciting opportunities for fluorescent feature reconstruction[1–5]**.** Here, fluorescence reconstruction microscopy (FRM) takes in a transmitted light image of a biological sample and outputs a series of reconstructed fluorescence images that predict what the sample would look like had it been labeled with a given series of dyes or fluorescently tagged proteins (Fig. 1A-C) [2,6–10]. FRM works by first training a convolutional neural network (e.g. U-Net) to relate a large set of transmitted light data to corresponding real fluorescence images (the ground truth) for given markers[11–13]. The network learns by comparing its fluorescence predictions to the ground truth fluorescence data and iterating until it reaches a cut off. Once trained, FRM can be performed on transmitted light data without requiring any additional fluorescence imaging. This is a powerful capability and allows FRM to: reduce phototoxicity; free up fluorescence channels for more complex markers; and enable re-processing of legacy transmitted light data to extract new information. In all cases, FRM data are directly compatible with any standard fluorescence analysis software or workflows (e.g. ImageJ plug-ins). Such capabilities are extremely useful, and FRM may eventually become a standard tool to augment quantitative biological imaging.

However, a number of challenges limit FRM accessibility to the larger biological community. Key among these is the difficulty in relating the abstract accuracy metrics used to score FRM to the practical value of FRM data for actual, quotidian biological analyses such as cell counting or morphological characterization. To better appreciate this, consider first that the quality of FRM is typically assessed using a single numerical metric (*P*) such as the Mean-Squared-Error or Pearson's Correlation Coefficient that typically range from (0,1) or (-1,1), and second that it is practically impossible to actually reach perfection ($P = 1$). *P* can be increased closer to 1 either by training with more images, or by using higher resolution magnification (e.g. 40X-100X) to capture finer details. However, increasing *P* also carries an intrinsic cost in increased wet-lab and computing time. That improving *P* is expensive and that *P* cannot be perfect beg the questions of how good is good enough, and good enough for what (Fig. 1D)? For instance, $P = 0.7$ lacks any practical context, and may be quite good enough for a given use case without requiring more work to raise the 'accuracy'. This is why context is extremely important for FRM and why the work we present here focuses on evaluating practical uses of FRM with respect to given *P* values.

Our goal here is not to improve FRM performance, but rather to provide a standardized implementation of it and demonstrate its practical performance for every-day tasks such as nuclear localization and tracking, characterizing cell morphology, cell-cell junction detection and analysis, and re-analyzing legacy data and data collected on different systems (Fig. 1). To further emphasize the use of FRM for routine tasks, we will exclusively focus on those lower magnifications (4X-20X) commonly used in high content imaging and cellular screening in contrast to the focus on higher magnifications in prior studies [9,10]. We hope that the included software we developed and the analyses and comparison data we present will help make FRM more approachable to the broader biological community.

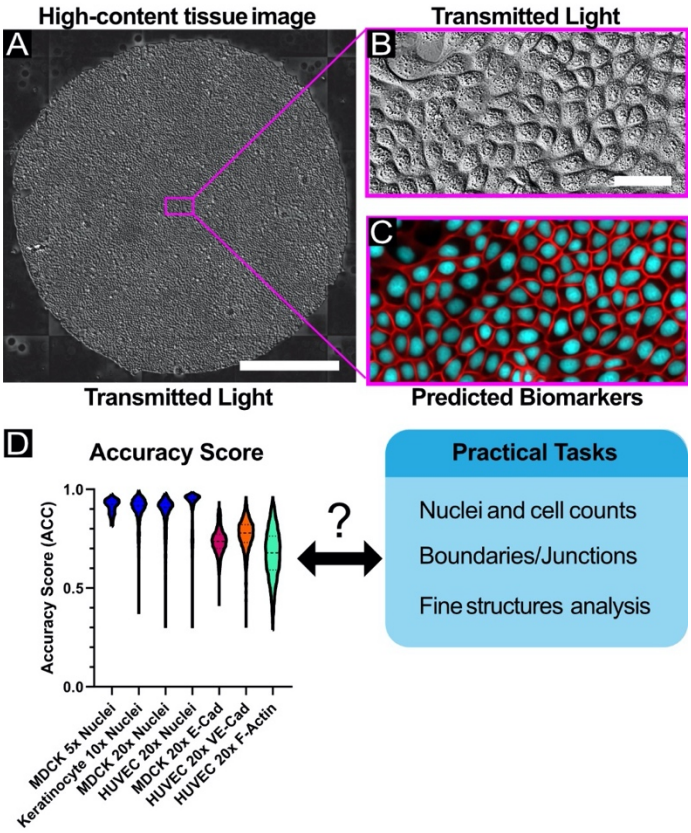


Fig 1. High-content, high-throughput labeling of fluorescent features. *(A) Sample large tissue of MDCK cells imaged via transmitted light (DIC). The scale bar represents 1 mm. A sub-region of the large tissue is enlarged in B. (B) A representative image which is given as input to the U-Net processing framework. The scale bar represents 50 µm. (C) The predicted fluorescent features (cell-cell junctions and nuclei) produced by the U-Net, corresponding to the same spatial region as in B. (D) Violin plot of accuracy score results from all experimental datasets. However, such accuracy metrics are not necessarily indicative of useful feature reconstructions for many practical applications.*

Adapting U-Nets for Low Magnification, High-Content FRM

While early FRM methods used computationally complex and expensive networks that relied on Z-stacks of images to capture 2D reconstruction[9], more recently this has been adapted to reconstruct 3D image stacks using a modified U-Net architecture[10]. The U-Net itself is commonly used in machine learning approaches because it is a lightweight convolutional neural network (CNN) which readily captures information at multiple spatial scales within an image, thereby preserving reconstruction accuracy while reducing the required number of training samples and training time. U-Nets, and related deep learning approaches, have found broad application to live-cell imaging tasks such as cell phenotype classification, feature segmentation[10,14–19], and histological stain analysis[20–23].

Our implementation here provides an archetypal U-Net and framework intended for the cell biology community. Briefly, our workflow is as follows. First, we collected multi-channel training images of cultured cells where each image comprised a transmitted light channel and associated fluorescence channels (labeled using genetically encoded reporters or chemical dyes; see Methods). These images were then broken into 256x256 pix^2 sub-images in ImageJ and then input into the network. Such image chopping is necessary for the average user to account for the average RAM and graphics cards available on standard workstations. These data are then passed through the U-Net network to generate trained weights—the pattern recognition side of the network. Here, the transmitted light images serve as input to the network, which is then optimized to minimize the difference between intensity values of the output

predicted images and the intensity values from the ground truth corresponding fluorescence images (e.g. Fig. 1). This process can be extended to full time-lapse video fluorescence reconstruction, making it well suited for high-content live imaging (see Movies S1-4). We have provided all of our code, sample raw data, and an extensive user manual (DataSpace, GitHub) to encourage exploration of FRM.

As our conventional performance metric, we selected the Pearson's Correlation Coefficient (PCC), which is commonly used in cell biology when comparing the co-localization of two or more proteins, and also used in computer vision to assess spatial-intensity when determining image similarity. However, we observed that naively applying the PCC across our whole dataset skewed the results due to the large number of images containing primarily background (common with high content imaging of oddly shaped or low density samples). This resulted in poor PCC scores as the network tried to reconstruct the pseudo-random background noise. To address this, we report a corrected accuracy score (*P*) representing the PCC of a large subset of images in a given dataset containing positive examples of the feature (nuclei, junctions, etc.) based on an intensity threshold (Fig. S1, see Methods). This approach will improve network performance for datasets containing large amounts of background signal.

To broadly explore the utility of FRM for high-content imaging applications, we captured transmitted light images using 4X, 10X, and 20X air objectives using either Phase Contrast or Differential Interference Contrast (DIC), and collected data across 3 different cell types—renal epithelial cells (MDCK), primary mouse skin keratinocytes (KC), and human umbilical vein endothelial cells (HUVEC). Variable training set sizes were tested to also explore the effect of data set size on 'accuracy'—a key practical aspect of designing an FRM study. The biomarkers we trained against comprised a nuclear dye (Hoechst 33342), an F-actin dye (SiR-Actin) and genetically encoded fluorescence reporters for E-cadherin and VE-cadherin. Traditional Accuracy Scores for each of these are summarized in Fig. 1D and Table S1, and we will next present case studies from each of these data sets before concluding with a discussion of how 'accuracy' relates to visual quality to help researchers design experiments for FRM.

Results.

Demonstration of FRM for low-magnification nuclear fluorescence reconstruction and analysis

One of the most common computational image processing needs for high content imaging is nuclei detection or segmentation, which enables cell counting, time-lapse tracking, and statistical analyses of ensemble distribution and geometry. While a variety of traditional image processing approaches exist to extract nuclei from phase or DIC images, the most reliable and standardized technique by far is using a vital dye (e.g. Hoechst 33342 or DRAQ) to stain the nuclei. However, Hoechst requires cytotoxic UV illumination while DRAQ (far-red fluorescence characteristics) has been linked to cell cycle alterations due to its chemistry [24–26]. Both dyes also exhibit loss of signal over extended time-lapse imaging. Alternately, genetic reporters such as H2B nuclear labels can be engineered into cells (e.g. transfection, viral

addition, etc.), but this adds more overhead, incurs phototoxicity, and still requires a dedicated fluorescence channel for a relatively simple structure (the nucleus) in lieu of a more complex or useful label. Hence, there is a clear practical benefit to fluorescence reconstruction of cell nuclei, especially in timelapse imaging where freeing up a channel and reducing phototoxicity are each quite valuable.

To validate low-magnification, high-accuracy nuclear FRM, we collected data in both MDCK renal epithelia cells (5X phase contrast) and primary skin keratinocytes (10X phase contrast) while using Hoechst to label nuclei and generate our ground truth training data. Fig. 2A displays the results of our U-Net predictions for these two cases, comparing the ground truth fluorescence (green) with the network predictions (red). These channels are overlaid for direct comparison where the perfect merger of red and green should appear yellow. The Accuracy Score (*P*) is included for context.

The network performs visually well in both cases, with $P \sim 0.9$, but to represent what that means in practice, we quantified disparities in the predictions with respect to nuclear size for geometric accuracy (Fig. 2B) and centroid error to reflect positional accuracy (Fig. 2C). In both nuclear area cases, the U-Net slightly overpredicts area, likely due to slight noise in the predictions blurring the predicted nuclei and effectively increasing area. However, the distributions from the violin plots are quite similar in structure, and the predictions are well within the usable range for practical cell counting and segmentation. With respect to nuclear centroid localization, mean errors span 2 microns (5X MDCK) to 1 micron (10X KCs). The improvement from 5X to 10X can likely be attributed to the resolution increase in the magnification, but in both cases the errors are quite small and more than sufficient for standard cell counting, nuclei tracking, and neighbor distribution analyses. Whether a higher *P* would be beneficial would depend on the specific analysis in question—here, the accuracy is more than sufficient.

As a final demonstration of the utility of low-magnification reconstruction and nuclear tracking, we input legacy data from a 24 hr time-lapse experiment of the growth dynamics of large epithelia (2.5 mm^2, 5X) and the network output a reconstructed movie of nuclear dynamics (Fig. 2D, and Movie S1) compatible with standard nuclear tracking algorithms (e.g. Trackmate in FIJI). Images were captured every 10 minutes, and previous efforts to perform this experiment using fluorescent imaging of Hoechst resulted in pronounced phototoxicity, hence FRM proved highly effective both as an alternative nuclear labeling approach for large-scale, long-term imaging, and as a means to reprocess pre-existing datasets.

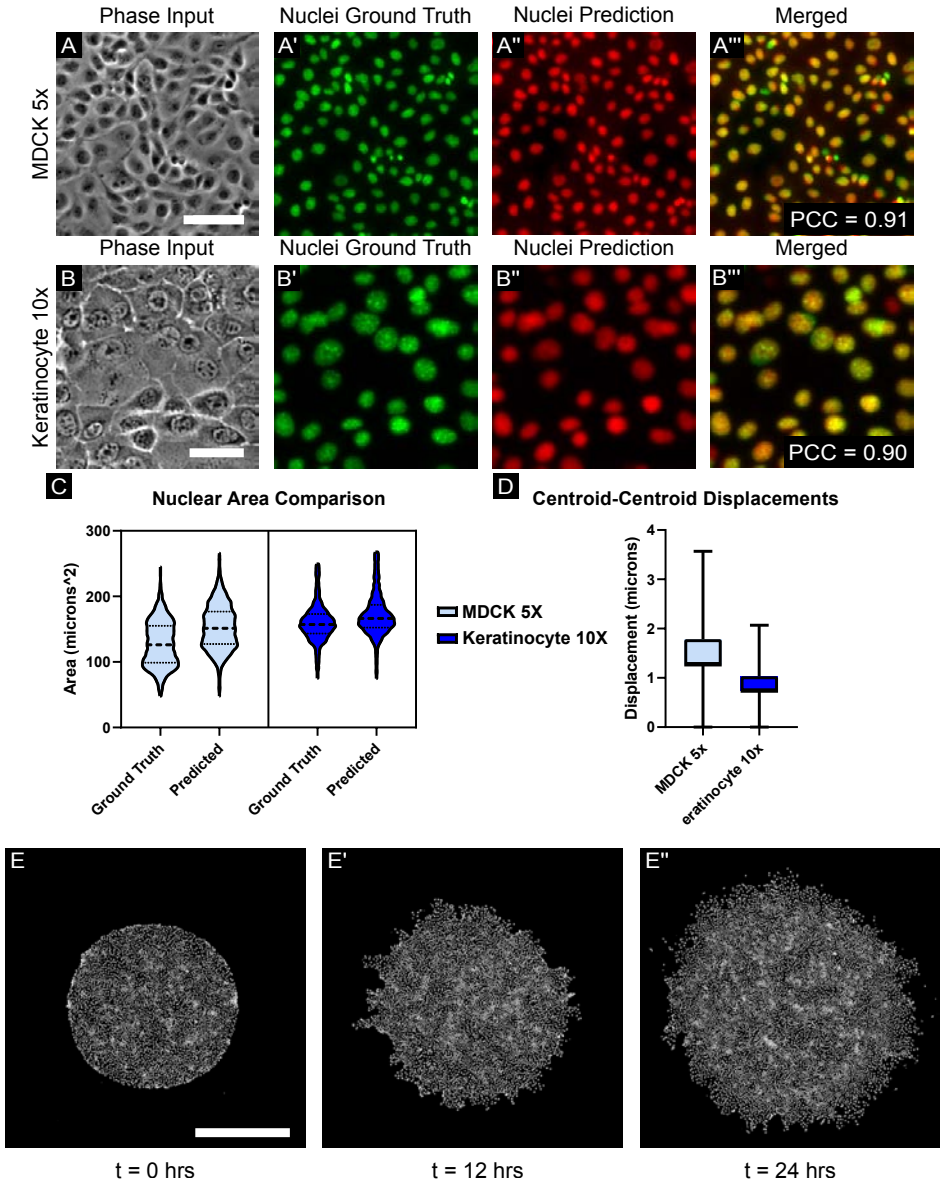


Fig 2. Low-magnification nuclei reconstruction.
(A) Representative transmitted-light image of MDCK cells at 5x magnification, with corresponding: (B) ground-truth nuclei, stained with Hoecscht 33342 and imaged with blue fluorescent light; (C) nuclear prediction produced by the network; and (D) the overlay of (B) and (C) displayed in red and green, respectively. The raw Pearson's correlation coefficient between (B) and (C) is given at right. The scale bar is 100 μm.
(E) Representative transmitted-light image of keratinocyte cells at 10x magnification, with corresponding (F, G, H) ground truth nuclei image, predicted nuclei, and overlay, respectively. The Pearson's correlation coefficient is given at right. The scale bar is 50 μm.
(I, J) A comparison of nuclear area estimations and centroid-centroid displacement estimations, respectively, for the two low-magnification datasets considered here. See *Methods*.
(K, K', K'') Sample nuclear predictions for the same growing tissue at 0, 12, and 24 hours of growth, respectively. Input data consists of MDCK WT cells imaged at 5x magnification and montaged; the U-Net was applied in a sliding-window fashion to predict small patches of the image in parallel. The scale bar is 1 mm.

Sharing FRM networks across laboratories and microscope hardware

In practice, it is far more convenient to deploy a pre-trained network rather than re-train a network. While this not always possible, if the same cell type and markers are desired at similar magnifications, it would facilitate improved sharing across different labs and further improve access to FRM. The problem here is that different labs typically have different optical platforms, meaning that the microscope, objective, illumination, and camera may all be different, which can prevent accurate reconstruction. However, it should be possible to correct for some of these differences in certain cases. As a simple demonstration, we collected and trained nuclear data using a Zeiss Observer microscope, a 5X phase contrast objective, and a camera with 42 μm^2 pixels and then tested that network against the sample imaged using a Nikon Ti2 with a 4X phase contrast objective and a camera with 53 μm^2 pixels. In addition to other optical differences, this reflects a 25% difference in magnification and a 26% difference in pixel area. Naïve testing of the network resulted in extremely poor performance by visual assessment (Fig. 3). In an attempt to compensate for this, we proposed an 'Inter-System Correction Factor' that basically attempts to correct for simple differences in optical hardware by rescaling the raw data into something closer to what the U-Net expects. This process flow is visually laid out in Fig. 3 and effectively comes down to calculating a scaling factor based on the ratio of the individual µm/pixel scales for each optical system. The result of this is shown in comparison in Fig. 3, where there is a clear performance improvement, suggesting some degree of cross-platform compatibility—perhaps enough to test out a network. That said, it is ideal to collect training data from the system where the primary data will be captured.

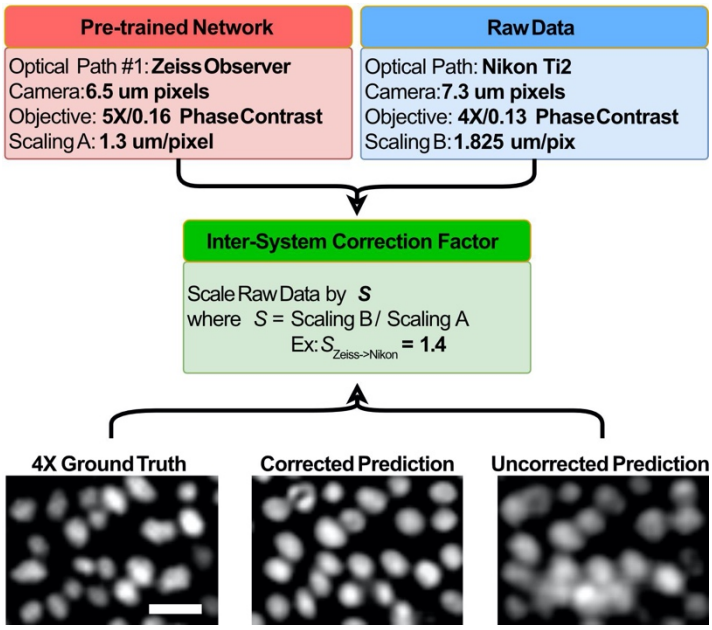


Fig 3. Cross-platform feature reconstruction is possible given an appropriate re-scaling of the input images.
A network is trained using data collected at low magnification (5x) on one system (Zeiss), and used to process new images collected on a different system (Nikon) at a similar magnification (4x), after the images have been appropriately rescaled as a function of the Inter-System Correction Factor. Representative ground truth nuclei from MDCK WT cells images on the Nikon system are shown next to the corresponding cross-platform predictions, resulting from both scaled and unscaled input images. Images were contrast adjusted for reproduction by normalizing the histograms and shifting the lower bound of the histogram up by 1/4 of the dynamic range. The scale bar represents 50 µm.

Reconstructing cell-cell junctions for segmentation and morphology applications

Cell-cell junctions and cellular boundaries in cellular ensembles have implications spanning the epithelial-mesenchyme-transition (EMT), tissue mechanics, and tissue maturation[27–29] and are of broad interest from cellular biophysics to high content screening. However, there are no vital dyes for junctional proteins (e.g. E-cadherin), necessitating either antibodies or genetic

reporters. In the absence of a specific marker, cell boundaries are relatively difficult to accurately segment, especially from DIC images (Fig. 4A), and proxy techniques such as Voronoi tessellation from nuclei data often fail to capture cell shape and organic features such as curved boundaries. Instead, junctions and boundary data most commonly come from biomarkers such as E-cadherin, so we trained our U-Net using MDCK cells stably expressing E-cadherin:RFP (Ecad:dsRed) and imaging with a 20X/0.75NA objective—a well-balanced objective favored for high-content imaging and immersion-free time-lapse imaging.

The U-Net was able to reconstruct E-cadherin junctions with high visual accuracy, as shown in the sequence from Figs. 4A-D. While $P = 0.74$, the reconstruction is quite spatially accurate, which is unexpected given how difficult it is for humans to detect cell-cell junctions by eye in a DIC image. To better highlight the accuracy and utility of junctional FRM, we explored how the network reconstructed a subtle 3D feature of epithelial junctions where a slanted junction is formed between two cells by one cell pushing slightly under another (the region enclosed in the dashed oval in Figs. 4B-D). Such slanted junctions may indicate a degree of fluidity or direction of migration and are also impossible to discern by eye. We quantified the accuracy of the FRM image by taking a line section perpendicular to this slanted junction (Figs. 4D,E) and comparing the profiles of the ground truth and the FRM image. In this line section, intensity values up to and including the peak value are similar, and intensity values exhibit a graded decay within the slanted junction, indicating that the FRM network is able to capture subtle 3D information from the 2D input image. To emphasize the 3D nature of this feature, a representative Z-section from an E-cadherin junction imaged by scanning confocal is shown in Figs. 4F,G. FRM can again be used for high-fidelity reconstruction during a timelapse, allowing both nuclei and junctions to be predicted throughout long acquisitions (see Movie S2). Overall, our network captures junctional intensity and geometry, both of which are invisible to the eye in the DIC input image.

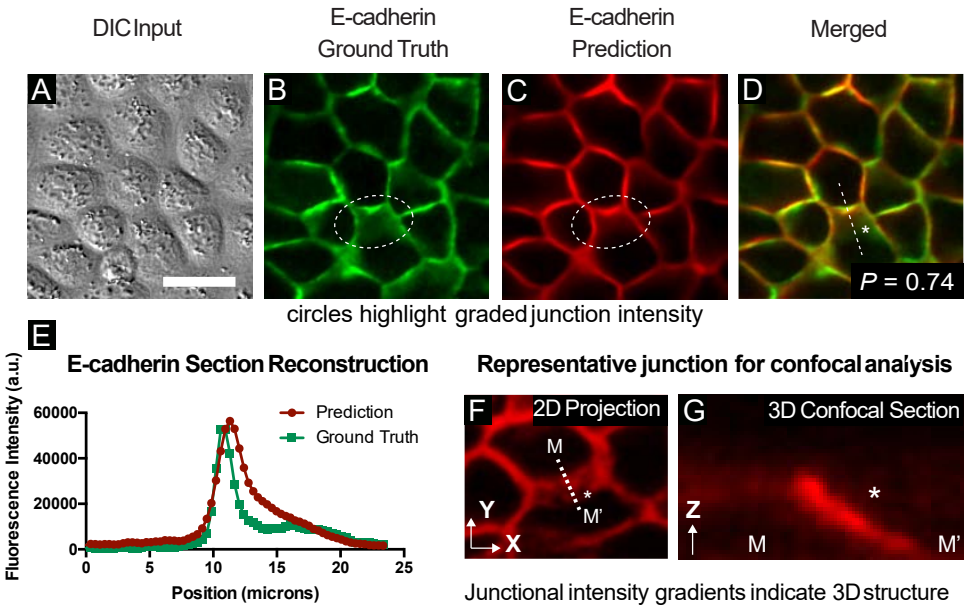


Fig 4. Cell-cell junction reconstruction captures otherwise invisible morphology
(A-D) Images of MDCK WT cells at 20x magnification were processed using a neural network trained to reconstruct cell-cell E-cadherin junctions. Representative ground truth features are shown alongside, and merged with, network predictions. The scale bar is 30 μm. (E) Line sections from identical spatial regions in (B) and (C) highlight the accuracy of predicted fluorescence intensity across the cell-cell junctions. From 2D transmitted light input (A), 3D structures may be approximated. (F, G) Representative cell-cell junction and corresponding confocal section, highlighting the relationship between 2D junction signal and 3D features. '*' in (G) is 7 μm above the basal plane.

Fine structure reconstruction

In practice, high content imaging is inherently a trade-off between throughput and resolution. The more detail we can extract from lower magnification images, the more efficient the imaging and analysis. Here, we demonstrate the practical performance of FRM and a 20X/0.8NA objective to reconstruct fluorescence signatures for several useful sub-cellular markers using HUVEC cells that stably expressed VE-Cadherin:YFP (mCitrine) and were labeled with Hoechst 33342 (live nuclear dye) and SiR Actin (infrared live actin dye). Processed timelapse data (see Movie S3) highlights the variation of these fluorescent features given the same input (DIC) image.

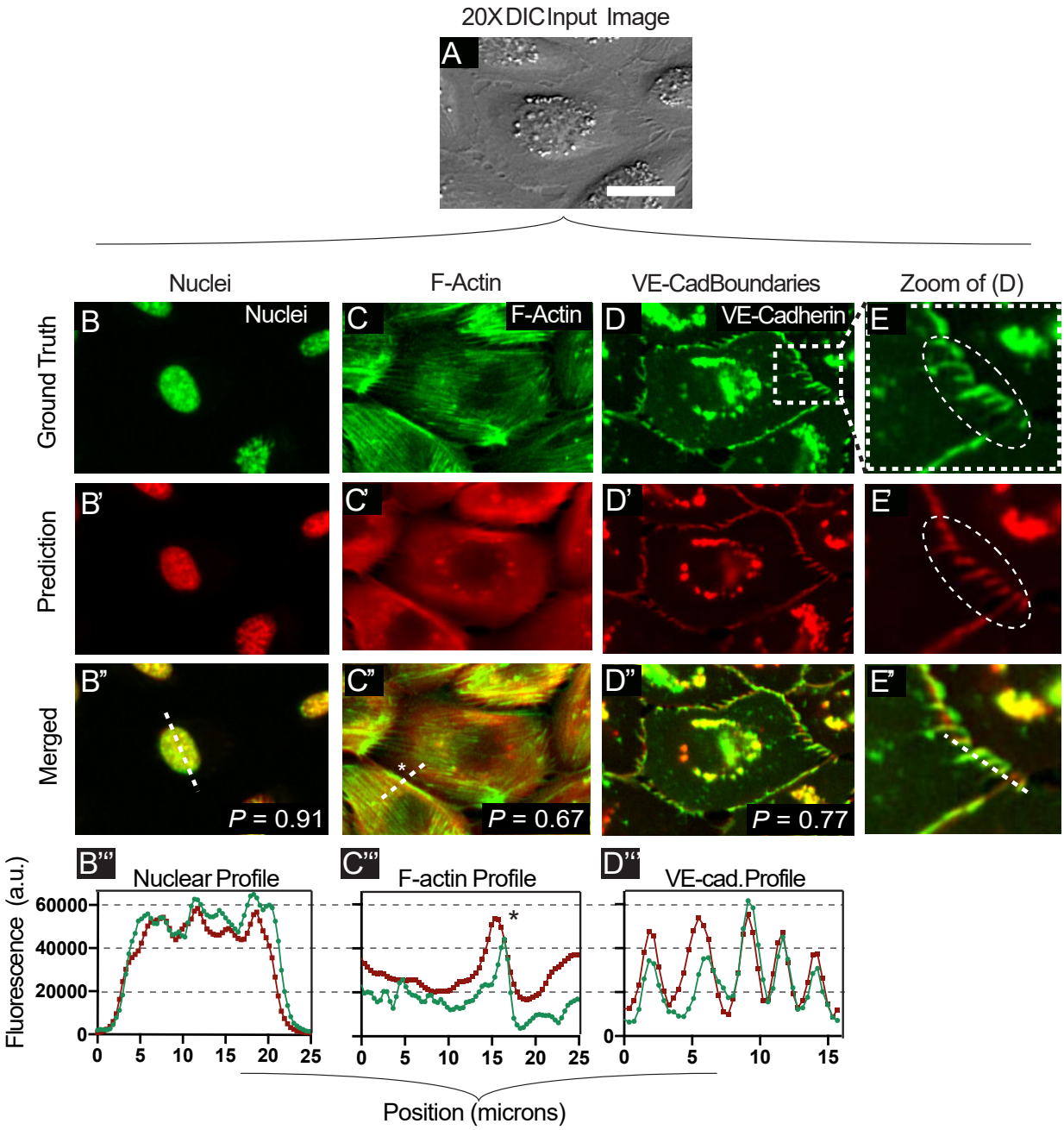
As a baseline, we characterized prediction accuracy for cell nuclei as the nucleus itself is relatively low resolution, but detection of sub-nuclear features requires higher accuracy. The Fig. B column demonstrates FRM performance for 20X nuclei including a line section through both the bulk structure and sub-nuclear granules. Visually, the FRM image is quite accurate, and $P = 0.91$ in this case. The line section easily captures the bulk form of the nucleus, but does not quite capture the texture inside the nucleus, although it does capture the rough form.

Next, we trained the network on identifying Actin after first staining HUVECs using the SiR-Actin live imaging dye. Here, the column in Fig. 5C shows significantly reduced performance as the fine F-actin filaments visible in the ground truth fail to be reconstructed in the predictions ($P = 0.67$) with the exception of some of the cortical filaments at the very edge of the cells (see the line profile). We hypothesize this is primarily due to fundamental limitations of DIC imaging and the lack of contrast for intracellular F-actin, but it may also be due to the network overprioritizing cortical filaments and the diffuse cytoplasmic signal. However, in practice we found that these FRM data were useful for general cell body detection and potential segmentation analyses due to the relatively homogeneous reconstructed fluorescence in the cytoplasmic space.

Finally, we trained the network with VE-Cadherin:YFP data in an attempt to reconstruct not only cellular borders, but also the well-characterized, nano-scale membrane fingers that develop in endothelial cell-cell junctions and indicate the direction of front-rear polarity in each cell [30]. In contrast to the actin performance, FRM proved far more capable here and readily detected both general VE-Cadherin boundaries (Fig. 5D column) and the membrane fingers (Fig. 5E column), although $P = 0.77$ still seems quite low and likely relates to the network attempting to reconstruct the more variable granules in the center of the cell, which are irrelevant for junctional analyses. While VE-Cadherin protrusions and boundaries are sometimes detectable by eye in DIC (as in Fig. 5A), they are still quite subtle in the best case, and developing a traditional computer vision process to detect and extract them has not been demonstrated, thereby again highlighting the practical value of FRM to reconstruct not only fluorescence, but also key morphological markers that are much easier to analyze in the FRM image than in the DIC image. Again, the value of FRM depends on the specific question and context, and the decision of whether it is ‘good enough’ at detecting fine structures rests with the end user. For reference, short movies of several markers are presented in Movies S3-5. Additionally, training the network with large and varied datasets enable it to begin to predict statistically rarer events, such as mitotic divisions, as shown in Movie S6.

268

269



270

Fig 5. Coarse-to-fine feature reconstruction.
(A) A representative transmitted-light image of HUVEC cells at 20x magnification with its corresponding structures of varying scale (B-D). The scale bar represents 30 μm. (B-D) the relatively large nuclei, the finer VE-Cadherin structures, and thin F-Actin filaments which are not readily resolved by the network. Ground truth fluorescent features are displayed alongside, and merged with, network predictions. (E) displays zoomed-in portions of images shown in (D). Line sections from (B", C", and D") are displayed graphically in (B''', C''', D'''), to enable intensity comparisons across the ground truth and predicted features.

Comparing FRM visual performance to P scores, training set size, and network

A key feature of FRM is that its performance can often be increased by collecting more training data, which in turn ought to improve *P*. However, *P* will never be perfect, nor is *P* necessarily the best metric to go by when determining if an FRM image is 'good enough', as clearly the context matters and the key question is 'good enough for what?' Hence, we sought to provide several examples of how the size of the training set affects both *P* and the actual visual accuracy or quality of the resulting FRM predictions.

To do this, we first swept through different sizes of training sets for many of the biomarkers presented earlier (see *Methods*). Briefly, we selected random subsets of very large datasets and trained the U-Net from scratch with these subsets. This process was repeated for different fractions of the complete dataset to capture the FRM performance versus training set size. The relation between *P* and training set size is shown in Fig. 6A to give a sense of how the quantitative accuracy progresses and eventually plateaus. In contrast, we also provide FRM results from the fractional training sets for the 20X HUVEC and MDCK data (Figs. 6B,C; respectively). Here, the input and ground truth data are presented alongside representative FRM predictions from networks trained with different numbers of training images (noted below each image). As a single image from the camera (2048x2048 pixels) is first split into 256x256 pixel sub-images for training, a fractional image (e.g. Fig. 6C, 1/16th column) implies that the network was trained on just a small crop from a single micrograph. Broadly speaking, and as expected, these data all indicate that FRM quality varies directly with the size of the training set, as expected.

However, Fig. 6A demonstrates that the rate of change in quantitative quality (*P*) vs. training set size is neither linear nor is it uniform across different biomarkers. Further, the actual predicted images shown in Figs. 6B,C offer further nuance because they demonstrate that training the network against even a single image would be sufficient to capture nuclei for the purposes of tracking or segmentation, while just 6 images would be sufficient to capture cell shape and junctional geometry in epithelia assuming the researcher were willing to perform some simple manipulations such as background subtraction. There is an obvious performance increase for both cadherins when the training set comprises several hundred images, but it is difficult to visually detect a difference between nuclei reconstructed from 6 or 400 images.

An alternate way to improve FRM would be to alter the U-Net architecture. Here, we first compared the standard U-Net to a neural network architecture which was essentially two U-Nets stacked end-to-end with additional residual connections. Such an approach has been shown to improve network depth and performance in other applications[31–33]. Here, however, we observed no benefit to training a deeper network (see Fig. S2). Further, given the significant temporal and computational cost, we advise against its use for this kind of FRM. Alternately, we explored the role of the loss function, testing our Pearson's-based loss function against the traditional Mean-Squared-Error loss function and found no significant difference (Fig. S3; Methods). Hence, we conclude that our minimal U-Net implementation performs well as a foundation for a variety of daily analysis tasks without requiring significant fine tuning.

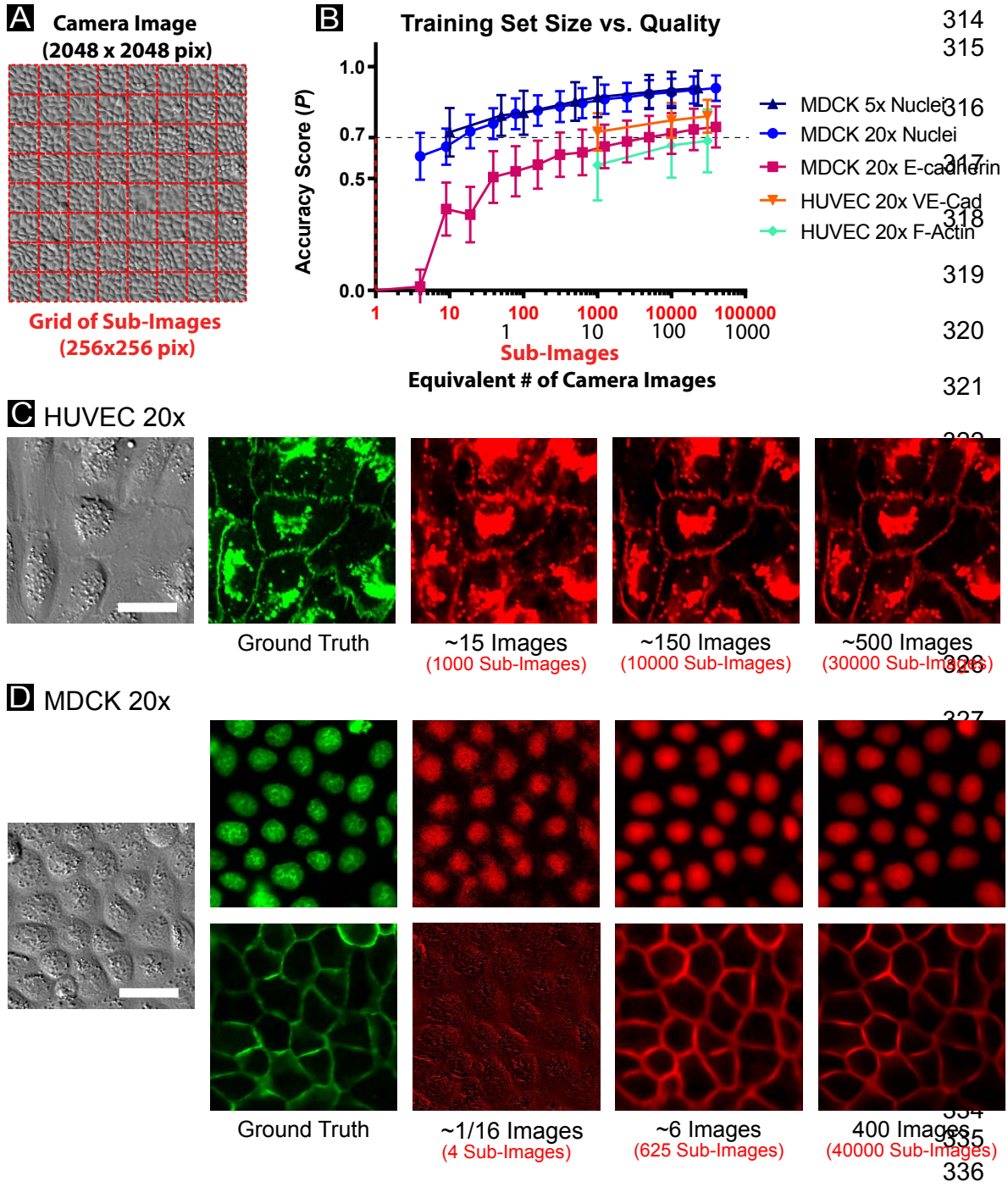


Fig 6. Impacts on prediction accuracy from smaller training sets.
(A) Cropping a sample image into 64 sub-images. (B) A comparison of network prediction accuracy as a function of training set size. The U-Net is trained with the complete dataset as described in Table S1 for each experimental condition. Then, random images representing a fraction of the total training set is used to train a new U-Net from scratch. (C, D) display representative images for the HUVEC 20x dataset and the MDCK 20x dataset, respectively, with predictions shown for various training set sizes. This type of analysis may assist users in collecting enough data for their task-specific quality requirements. All scale bars represent 30 μm.

Discussion

Limitations of existing accuracy metrics and the importance of context

Our data—collected from actual, real-world analyses, highlight the limitations of using traditional accuracy metrics from computer vision for biological image analysis. Specifically, while there is a general relation between an improved *P* Accuracy Score and FRM quality, it is not linear nor intuitive how to determine what is 'good enough' given only a *P* value devoid of context for a specific analysis. Further, and most critically, FRM does not reconstruct images according to human imperatives. The U-Net only optimizes via the specific loss function it has been given (e.g. Mean-Squared-Error or the Pearson's coefficient). What the computer considers 'good' need not match our own assessments of value and quality.

As a practical example, compare the FRM performance for E-cadherin ($P = 0.73$; Fig. 3) and F-actin ($P = 0.67$; Fig. 5). While the accuracy metrics differ by < 10%, the FRM of F-actin only detected peripheral actin cables, otherwise blurring all internal features into a homogeneous signal. Nonetheless, even this plainly 'inaccurate' signal could prove useful for cytoplasmic reconstruction and tracking. In stark contrast, the E-cadherin data was much more visually accurate and also captured key quantitative features of the ground truth such as junctional localization and intensity, and even the subtle intensity gradients representing 3D morphology despite having only a slight improvement in *P*-values. Yet despite that, a score of 0.73 is far enough from '1' that it is ambiguous in absence of a specific analysis, which is why FRM must be evaluated in the context of a given question or analysis.

Practical considerations for training on new, low-magnification data

We specifically targeted the lower-magnification end of the imaging spectrum to explore how well FRM performed at magnifications more commonly used for high content imaging applications such as timelapse studies of very large cellular colonies or massive screens using multiwell plates. Our data indicate that such magnifications can be effectively combined with FRM for applications spanning nuclear tracking, cell-cell junction analysis, and certain fine-structure reconstruction even at just 20X.

A particular concern for the average user of a complex machine learning process is the size of the dataset required as this can impose potentially strenuous experimental demands. However, our characterization of FRM performance vs. data set size again shows the importance of context as relatively few images are needed to get quite accurate nuclei reconstruction, while a greater number of images are needed for junction reconstruction (Fig. 6). However, we also note that our largest training set size comprised at most 500 camera images at 20X (approximately one six-well plate)—something easily obtained with a standard automated microscope, and still compatible with manual capture. Further, a very common approach in machine learning is to 'augment' an image dataset by performing reflections and rotations on images such that the network perceives each augmented image as a different datapoint, thereby virtually increasing the size of the dataset. We did not perform such augmentation here for the sake of simplicity and transparency, which suggests that significantly smaller datasets, if augmented, could still produce good results.

Finally, we provided a limited demonstration of a simple approach to apply a network trained on one optical system to data from a different system. While our scaling approach works quite well in our hands for simple nuclear tracking, we would still strongly recommend collecting training data on the primary system. However, sharing networks between labs can still prove useful, and if the optical systems are close enough, simple scaling can help at least validate the feasibility of FRM before a lab embarks on training their own network.

FRM versus machine-learning segmentation approaches

FRM is a quickly developing technology that exists alongside another popular approach where machine learning is used for feature segmentation[14,19,34]. In the latter case, the network is trained to specifically detect 'features' (e.g. nuclei) as binary objects, whereas FRM instead reconstructs the effective fluorescent image of what a fluorescent label against that structure might show. Both are useful techniques, and the best approach depends on the application. However, there are several unique advantages to FRM. First, reconstructing an effective fluorescent image from auto-annotated data (e.g. chemical dyes, antibodies, fluorescent proteins) obviates the need for any manual annotation or pre-processing—often quite time consuming and subjective. This means that an FRM image can be directly incorporated into any existing analysis pipeline intended for fluorescent images, including traditional threshold-based segmentation approaches. Further, more of the original data is preserved in an FRM image, allowing the capture of things such as fluorescence intensity gradients (e.g. Fig. 3), and features that might be lost during traditional binary segmentation.

Concluding remarks

Here, we characterize the value of fluorescence reconstruction microscopy (FRM) for everyday analysis tasks facing researchers working with cell biology. We specifically highlight the need for individual researchers to explore and evaluate FRM in the context of specific research questions rather than accuracy metrics. We also highlight the surprisingly good performance of FRM even with lower magnification imaging or relatively fine structures such as VE-cadherin fingers. Finally, we have made all of our tools and key training datasets publicly available to improve accessibility and provide a starting point for researchers new to FRM to easily explore it for themselves and to eventually build on and improve.

415 **Methods.**

416 ***Tissue culture***

417 MDCK-II (G-type) cells stably expressing E-cadherin:dsRed were cultured in low glucose
418 DMEM. The MDCK-II culture media was supplemented with 10% Fetal Bovine Serum (Atlanta
419 Biological) and penicillin/streptomycin. HUVEC endothelial cells stably expressing VE-
420 cadherin:mCitrine were cultured using the Lonza endothelial bullet kit with EGM2 media
421 according to the kit instructions. Primary murine keratinocytes were isolated from neonatal mice
422 (courtesy of the Devenport Laboratory, Princeton University) and cultured in custom media[35]. All
423 cell types in culture were maintained at 37°C and 5% CO_2 in air.

424 ***Preparation of training samples***

425 We collected training data using 3.5-cm glass-bottomed dishes coated with an appropriate
426 ECM. To coat with ECM, we incubated dishes with 50 µg/mL in PBS of either collagen-IV
427 (MDCK) or fibronectin (HUVEC, primary keratinocytes) for 30 min 37 °C before washing 3 times
428 with DI water and air drying the dishes.

429 In order to contain a variety of conditions within a single plate to ensure a broad training sample,
430 we placed silicone microwells into the dishes as described in [[36]] at densities from [1-2x10^6
431 cells/mL] which ultimately allowed for single cells, low density confluent monolayers, and high
432 density confluent monolayers to be captured. Silicone microwells consisted of 3x3 arrays of 9
433 mm^2 microwells into which we added 4 µL of suspended cells in media, allowed them to adhere
434 for 30 min in the incubator (6 hrs for keratinocytes), added media and returned them to the
435 incubator overnight prior to imaging. To further ensure variability, several dishes were also
436 randomly seeded with cells for each cell type.

437 ***Fluorescent labeling for ground truth data***

438 We used the live nuclear dye NucBlue (ThermoFisher; a Hoechst 33342 derivative) with a 1 hr
439 incubation for all nuclear labeling. We used SiR-Actin (Spirochrome) at 10 µM for live F-actin
440 labeling in HUVECs. All other labels were genetically encoded reporters as described.

441 ***Image Acquisition***

442 5X MDCK data was collected on a Zeiss (Observer Z1) inverted fluorescence microscope using
443 a 5X/0.16 phase-contrast objective, an sCMOS camera (Photometrics Prime) and controlled
444 using Slidebook (Intelligent Imaging Innovations, 3i). An automated XY stage, a DAPI filter set,
445 and a metal halide lamp (xCite 120, EXFO) allowed for multipoint phase contrast and
446 fluorescent imaging.

447 All epifluorescence imaging was performed using a Nikon Ti2 automated microscope equipped
448 with a 10X/0.3 phase objective, a 20X/0.75 DIC objective, and a Qi2 sCMOS camera (Nikon
449 Instruments). Time-lapse imaging effectively increased dataset size as long as sufficient time
450 was allowed between frames to avoid overfitting in the U-Net. MDCK data was collected at 20
451 min/frame, while HUVEC and keratinocytes were given 60 min/frame. Standard DAPI, CY5, and
452 YFP filters sets were used. Confocal sections of E-cadherin fluorescence in MDCK cells (Fig. 3)
453 were collected using a Leica SP8 scanning confocal tuned for dsRed excitation/emission.

All imaging was performed at 37 °C with 5% CO2 and humidity control. Exposures varied, but were tuned to balance histogram performance with phototoxic risk. Data with any visible sign of phototoxicity (blebbing, apoptosis, abnormal dynamics) were excluded entirely from training.

Data Pre-Processing and Training

Prior to input to the network, raw images were segmented into 256x256 pixel^2 sub-images, ensuring consistent slicing across the transmitted-light image and the corresponding fluorescent image. The images were then normalized by statistics collected across all images in each channel: that is, by subtracting from each image the mean and dividing by the standard deviation. A test-train split was applied, such that a random 20% of the total images were held out to comprise the test set. Additionally, 10% of the training data subset were held out for validation as is standard.

The U-Net style architecture shown in Figure 1 was trained using TensorFlow[37] and the ADADELTA optimizer[38]. In the standard training experiments, the mean squared error (MSE) loss function was applied across pixel intensity values in the predicted images compared to intensity values in the ground truth images. Results from the MSE were contrasted with results from two networks trained to maximize the Pearson's correlation coefficient (PCC). The PCC is commonly used in cell biology for evaluating the colocalization of two fluorescently labeled structures[39]. The PCC loss function was defined, for two intensity data sets R and G, as:

$$PCC = \frac{\sum_i (R_i - \bar{R}) \times (G_i - \bar{G})}{\sqrt{\sum_i (R_i - \bar{R})^2 \times \sum_i (G_i - \bar{G})^2}}$$

$$Loss = \frac{(1 - PCC)}{2}$$

Sample training loss plots are provided (Fig. S4), reflecting the use of early stopping during training. That is, when the validation loss did not decrease for 75 epochs, the training process terminated. The training and test set sizes and results are provided for all experimental conditions in Table S1.

For several of the experimental conditions, networks were trained using subsets of the original dataset. To do so, the network architecture was fixed, and networks were trained from scratch by using a random subset of matched input-output image pairs from the original training set. The training setup, including hyperparameters, was unmodified regardless of training set size. The original test set was used to compare results for all training set size tests relative to each experimental condition.

Data Testing and Image Processing

The Pearson's correlation coefficient (PCC) of each test set is determined by individually computing the PCC between each predicted image, as output by the network, and its corresponding ground truth image. Additionally, an accuracy score (*P)* based on the PCC was devised to more reliably represent the performance of the network. To determine P, we report the PCC on a subset of the test set which selects for only those test images containing positive examples of features (nuclei, junctions, etc.). We construct this subset by manually determining threshold values to distinguish image intensities indicating the presence of the relevant features versus background noise for each test set. That is, the histograms of a subset of the data containing positive examples of features are plotted, and an approximate lower bound on

intensity values is estimated to distinguish the features from the background. Then, the histograms of a subset of the data containing only background are plotted to ensure that the threshold value is adequate to label the images as background-only images. The MatLab function rmoutliers() was utilized to remove outliers when *P* is reported for each condition.

For the low-magnification experimental conditions, a nuclear area comparison was performed between corresponding ground-truth and predicted images. Initially, both pairs of output nuclear images were segmented independently using standard auto-thresholding, watershedding, and size exclusion (to exclude clusters) in ImageJ/FIJI, and then outliers were removed using the MatLab function rmoutliers(). We additionally report the centroid-centroid displacement values for the same segmented images. The ImageJ/FIJI plugin TrackMate was used to determine displacements between the ground truth and predicted images, as if they were two frames of a video. Standard TrackMate settings were used and outliers were removed using the MatLab function rmoutliers() for reporting.

When intensity plots for line slices are reported, a line is selected as an ROI in ImageJ/FIJI, and intensity values are exported for analysis.

New large transmitted-light images were processed using a sliding-window technique. We processed a large image by analyzing 256x256 pixel^2 patches of the input image with a stride of 64 pixels in each direction. Additionally, the border of each predicted patch was excluded in the sliding-window process, as features near the patch borders are likely to have lower accuracy (often as a function of cells being cut off). The sliding-window predictions at each pixel were then averaged to produce the final large predicted image. Timelapse movies can be processed on a frame-by-frame basis. If scaling was required as described in Fig. 3, the input was scaled in FIJI and then passed to the network for analysis.

Code and Dataset Availability

All code used for pre-processing data, training the network, testing a trained model, and applying the model to new images, along with an extensive user manual can be found at:

https://github.com/CohenLabPrinceton/Fluorescence-Reconstruction .

Additionally, sample datasets for testing purposes, along with pre-trained network weights, are available through our Dataspace server (see GitHub README file).

Acknowledgements

Special thanks to Gawoon Shim for assistance with HUVEC and keratinocyte data collection.

Supplementary Materials

Magnification	Cell type	Feature	Training Set Size (sub-images)	Test Set Size	Modified *P:* Reduced Test Set Size	PCC Mean	PCC St.Dev	Modifed *P* Mean	Modified *P* St.Dev
5x	MDCK	Nuclei	22835	5709	4443	0.72529	0.35884	0.90322	0.07801
10x	Keratinocyte	Nuclei	26214	6554	4156	0.57717	0.43833	0.9014	0.093155
20x	MDCK	Nuclei	40000	10000	4556	0.43844	0.43333	0.90327	0.05724
20X		E-cadherin	40000	10000	4539	0.37466	0.34172	0.73053	0.091688
20X	HUVEC	Nuclei	30720	7680	5533	0.68936	0.40728	0.93776	0.060886
20X		VE-cadherin	30720	7680	5666	0.60824	0.30398	0.77737	0.074176
20X		F-actin	30720	7680	5820	0.51247	0.30438	0.66808	0.14057

Table S1. Accuracy and training statistics for all experimental conditions.

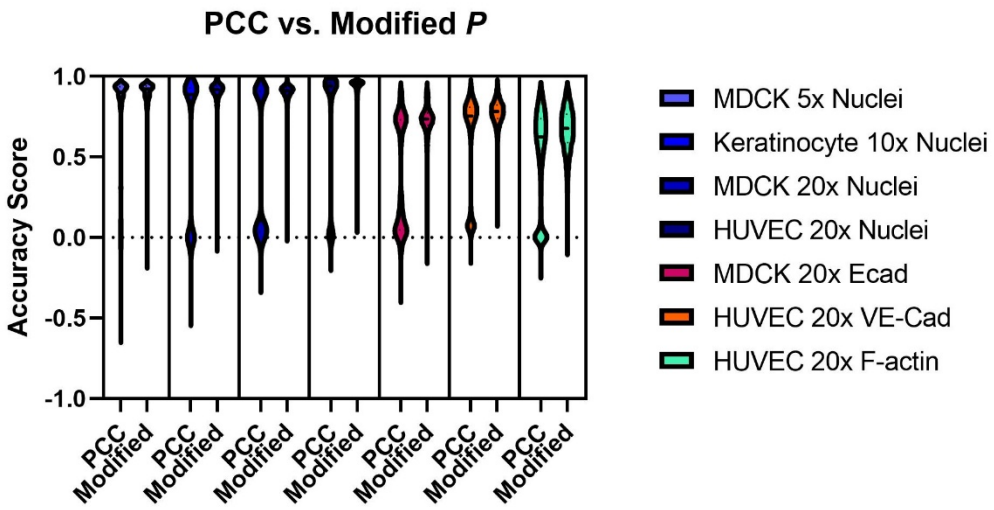


Figure S1. A comparison of experimental results in Pearson's Correlation Coefficient (PCC), versus a modified accuracy score *P* (see *Methods*). To ensure a fairer comparison, outliers were not removed; only intensity thresholding was performed to produce the modified *P* from the PCC. By filtering the PCC results by an intensity threshold in the fluorescent images, we remove low-scoring background images, which bias our accuracy score on the complete dataset. Visual inspection of the plot reveals the low-scoring images as "bumps" near 0.0.

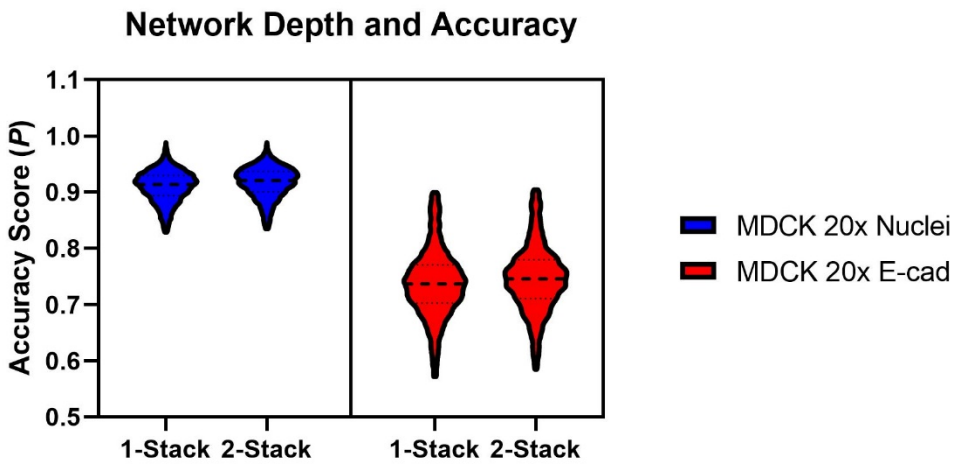


Figure S2. Representative accuracy results for a dataset trained on the standard (1-stack) U-Net, compared to a network comprised of two U-Nets stacked back-to-back, with residual connections (2-stack). Training conditions were otherwise unchanged. Accuracy scores, as reported in terms of the modified *P* (see *Methods*) were comparable.

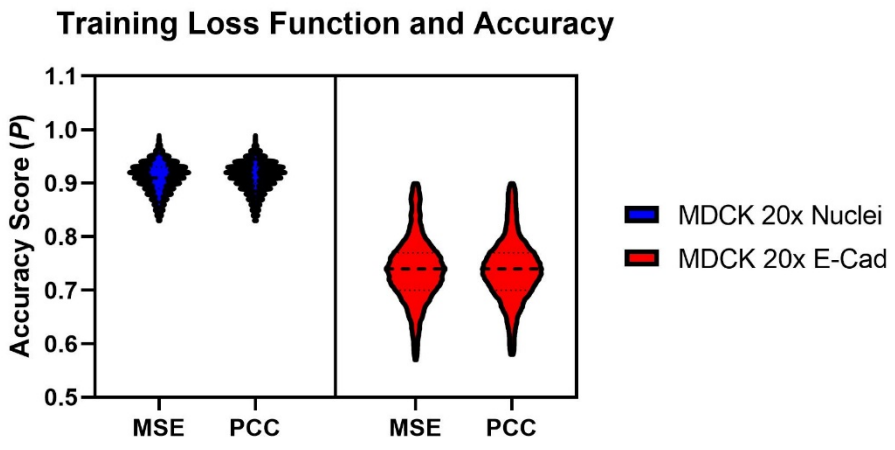


Figure S3. Representative accuracy results for networks trained using the Mean-Squared Error loss function (MSE) compared to the Pearson's Correlation Coefficient loss function (PCC). The neural network architecture and training conditions are the same, with the exception of the choice of loss function. Accuracy scores, as reported in terms of the modified *P* (see *Methods*) were comparable.

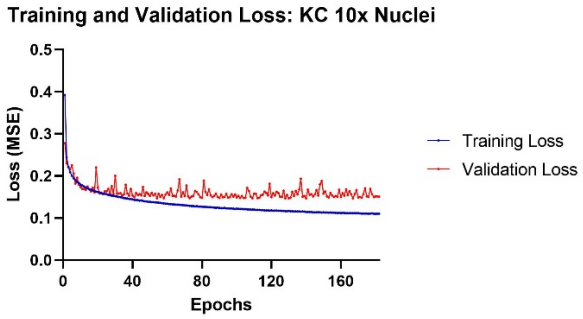


Figure S4. Representative loss functions from the standard U-Net training process. Early stopping was enabled, so that if the validation loss did not decrease within a set number of epochs, the training process terminated.

<u>Movie captions on the following page</u>

580 **Supplementary Movie Captions:**

581 **Movie S1. Fluorescence reconstruction microscopy (FRM) on timelapse data.**

582 A phase-contrast timelapse of MDCK cells, imaged at 5x magnification, is shown at left. The
583 center panel displays nuclear predictions produced by the trained U-Net, given individual frames
584 from the phase-contrast timelapse as input. The overlay of the phase-contrast movie and the
585 nuclear predictions is shown at right. Each panel is 0.31 cm x 0.31 cm, and time between
586 frames is 20 minutes. Video compressed for supplement, but raw data available on request.

587 **Movie S2. FRM for high-content screening.**

588 A DIC timelapse movie of MDCK cells, imaged at 20x magnification, is shown at left (top and
589 bottom). The top row displays E-cadherin junctions, while the bottom row displays nuclei.
590 Moving left from right, the second-from-left images are ground truth (actual) fluorescent images
591 of the junctions/nuclei in green, followed by the FRM predictions in red, and finally the merge of
592 the ground truth and predicted images. Predictions are produced by processing the DIC input on
593 the left through a neural network trained on a dataset of matched DIC and fluorescence image
594 pairs. Panel width is approximately 500 μm, and time between frames is 20 minutes.

595 **Movies S3-5. FRM for fine structures.**

596 A DIC timelapse movie of HUVEC cells, imaged at 20x magnification, is shown at left. Movie S3
597 shows VE-cadherin, Movie S4 shows Nuclei, and Movie S5 shows F-actin. All movies present
598 DIC/Ground Truth/Prediction/Merge from left to right. Individual panel width is 890 μm, with 20
599 minutes per movie frame.

600 **Movie S6. Mitotic division prediction.**

601 A neural network is able to capture rare events, such as cell divisions, when trained on a
602 sufficiently large and varied dataset. Left panel: a fluorescent timelapse of stained HUVEC
603 nuclei, imaged at 20x magnification. Center panel: the U-Net predictions from DIC images of the
604 same spatial region. Right panel: an overlay of the left and center panels for comparison. Time
605 between frames is 20 minutes.

606
607

References.

1. Caicedo, J. C. *et al.* Data-analysis strategies for image-based cell profiling. *Nat. Methods* **14**, 849–863 (2017).

2. Moen, E. *et al.* Deep learning for cellular image analysis. *Nature Methods* **16**, 1233–1246 (2019).

3. Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Mol. Syst. Biol.* **12**, 878 (2016).

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

5. Smith, K. *et al.* Phenotypic Image Analysis Software Tools for Exploring and Understanding Big Image Data from Cell-Based Assays. *Cell Systems* **6**, 636–653 (2018).

6. Belthangady, C. & Royer, L. A. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nat. Methods* **16**, 1215–1225 (2019).

7. Brent, R. & Boucheron, L. Deep learning to predict microscope images. *Nat. Methods* **15**, 868–870 (2018).

8. Weigert, M. *et al.* Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nat. Methods* **15**, 1090–1097 (2018).

9. Christiansen, E. M. *et al.* In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images. *Cell* **173**, 792-803.e19 (2018).

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

11. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2323 (1998).

12. Krizhevsky, A., Sutskever, I. & Hinton, G. E. *ImageNet Classification with Deep Convolutional Neural Networks*. (2012).

13. Ronneberger, O., Fischer, P. & Brox, T. U-net: Convolutional networks for biomedical image segmentation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9351**, 234–241 (Springer Verlag, 2015).

14. Van Valen, D. A. *et al.* Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments. *PLoS Comput. Biol.* **12**, (2016).

15. Falk, T. *et al.* U-Net: deep learning for cell counting, detection, and morphometry. *Nat. Methods* **16**, 67–70 (2019).

16. Gupta, A. *et al.* Deep Learning in Image Cytometry: A Review. *Cytometry Part A* **95**, 366–380 (2019).

17. Caicedo, J. C. *et al.* Evaluation of Deep Learning Strategies for Nucleus Segmentation in Fluorescence Images. *Cytom. Part A* **95**, 952–965 (2019).

18. Newby, J. M., Schaefer, A. M., Lee, P. T., Forest, M. G. & Lai, S. K. Convolutional neural networks automate detection for tracking of submicron-scale particles in 2D and 3D. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9026–9031 (2018).

19. Kim, D., Min, Y., Oh, J. M. & Cho, Y. K. AI-powered transmitted light microscopy for functional analysis of live cells. *Sci. Rep.* **9**, (2019).

20. Goodfellow, I. J. *et al.* Generative adversarial nets. in *Advances in Neural Information Processing Systems* **3**, 2672–2680 (2014).

21. Rivenson, Y. *et al.* PhaseStain: the digital staining of label-free quantitative phase microscopy images using deep learning. *Light Sci. Appl.* **8**, (2019).

22. Rivenson, Y. *et al.* Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**, 466–477 (2019).

23. De Haan, K., Rivenson, Y., Wu, Y. & Ozcan, A. Deep-Learning-Based Image Reconstruction and Enhancement in Optical Microscopy. *Proc. IEEE* **108**, 30–50 (2020).

24. Erba, E., Ubezio, P., Broggini, M., Ponti, M. & D'Incalci, M. DNA damage, cytotoxic effect and cell-cycle perturbation of Hoechst 33342 on L1210 cells in vitro. *Cytometry* **9**, 1–6 (1988).

25. Durand, R. E. & Olive, P. L. Cytotoxicity, mutagenicity and DNA damage by Hoechst 33342. *J. Histochem. Cytochem.* **30**, 111–116 (1982).

26. Wlodkowic, D. & Darzynkiewicz, Z. Please do not disturb: Destruction of chromatin structure by supravital nucleic acid probes revealed by a novel assay of DNA-histone interaction. *Cytom. Part A* **73**, 877–879 (2008).

27. Bray, M. A. *et al.* Cell Painting, a high-content image-based assay for morphological profiling using multiplexed fluorescent dyes. *Nat. Protoc.* **11**, 1757–1774 (2016).

28. Garcia, M. A., Nelson, W. J. & Chavez, N. Cell–cell junctions organize structural and signaling networks. *Cold Spring Harb. Perspect. Biol.* **10**, (2018).

29. Friedl, P. & Mayor, R. Tuning collective cell migration by cell-cell junction regulation. *Cold Spring Harb. Perspect. Biol.* **9**, 1–18 (2017).

30. Hayer, A. *et al.* Engulfed cadherin fingers are polarized junctional structures between collectively migrating endothelial cells. *Nat. Cell Biol.* **18**, 1311–1323 (2016).

31. He, K., Zhang, X., Ren, S. & Sun, J. *Deep Residual Learning for Image Recognition*.

(2016).

32. Newell, A., Yang, K. & Deng, J. Stacked hourglass networks for human pose estimation. in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9912 LNCS**, 483–499 (2016).

33. Szegedy, C. *et al. Going deeper with convolutions*. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* **07-12-June**, (2015).

34. Xing, F. & Yang, L. Robust nucleus/cell detection and segmentation in digital pathology and microscopy images: A comprehensive review. *IEEE Reviews in Biomedical Engineering* **9**, 234–263 (2016).

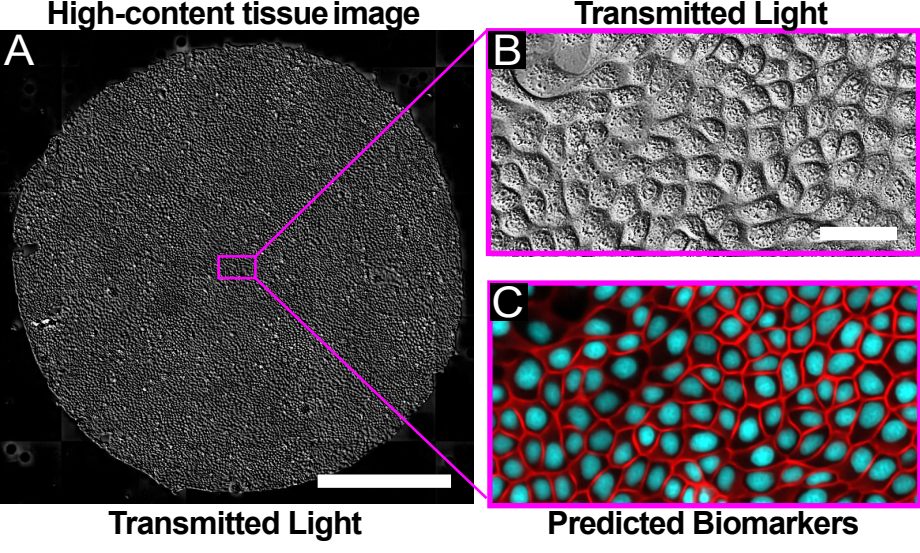
35. Nowak, J. A. & Fuchs, E. Isolation and culture of epithelial stem cells. *Methods Mol. Biol.* **482**, 215–232 (2009).

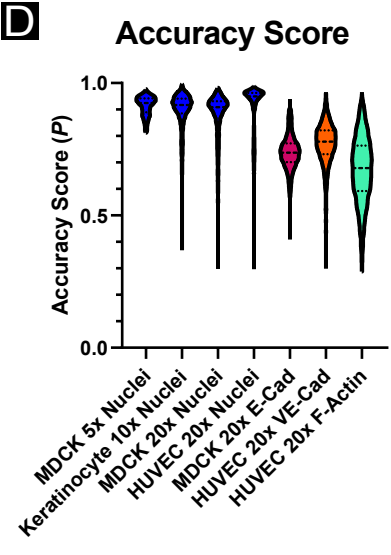
36. Cohen, D. J., Gloerich, M. & Nelson, W. J. Epithelial self-healing is recapitulated by a 3D biomimetic E-cadherin junction. (2016). doi:10.1073/pnas.1612208113

37. Weinman, J. J., Lidaka, A. & Aggarwal, S. TensorFlow: Large-scale machine learning. *GPU Comput. Gems Emerald Ed.* 277–291 (2011). doi:1603.04467

38. Zeiler, M. D. ADADELTA: An Adaptive Learning Rate Method. (2012).

39. Dunn, K. W., Kamocka, M. M. & McDonald, J. H. A practical guide to evaluating colocalization in biological microscopy. *American Journal of Physiology - Cell Physiology* **300**, (2011).

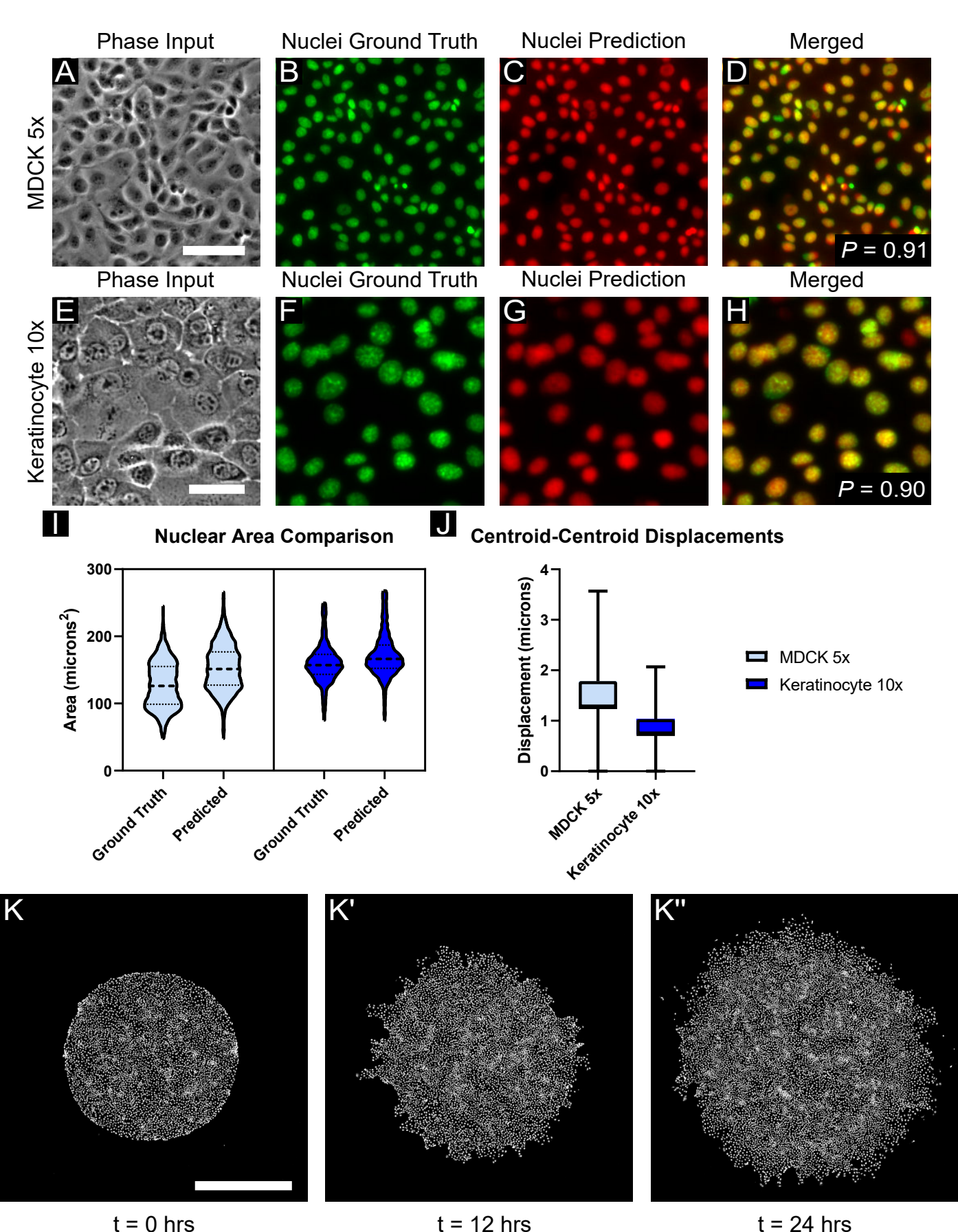


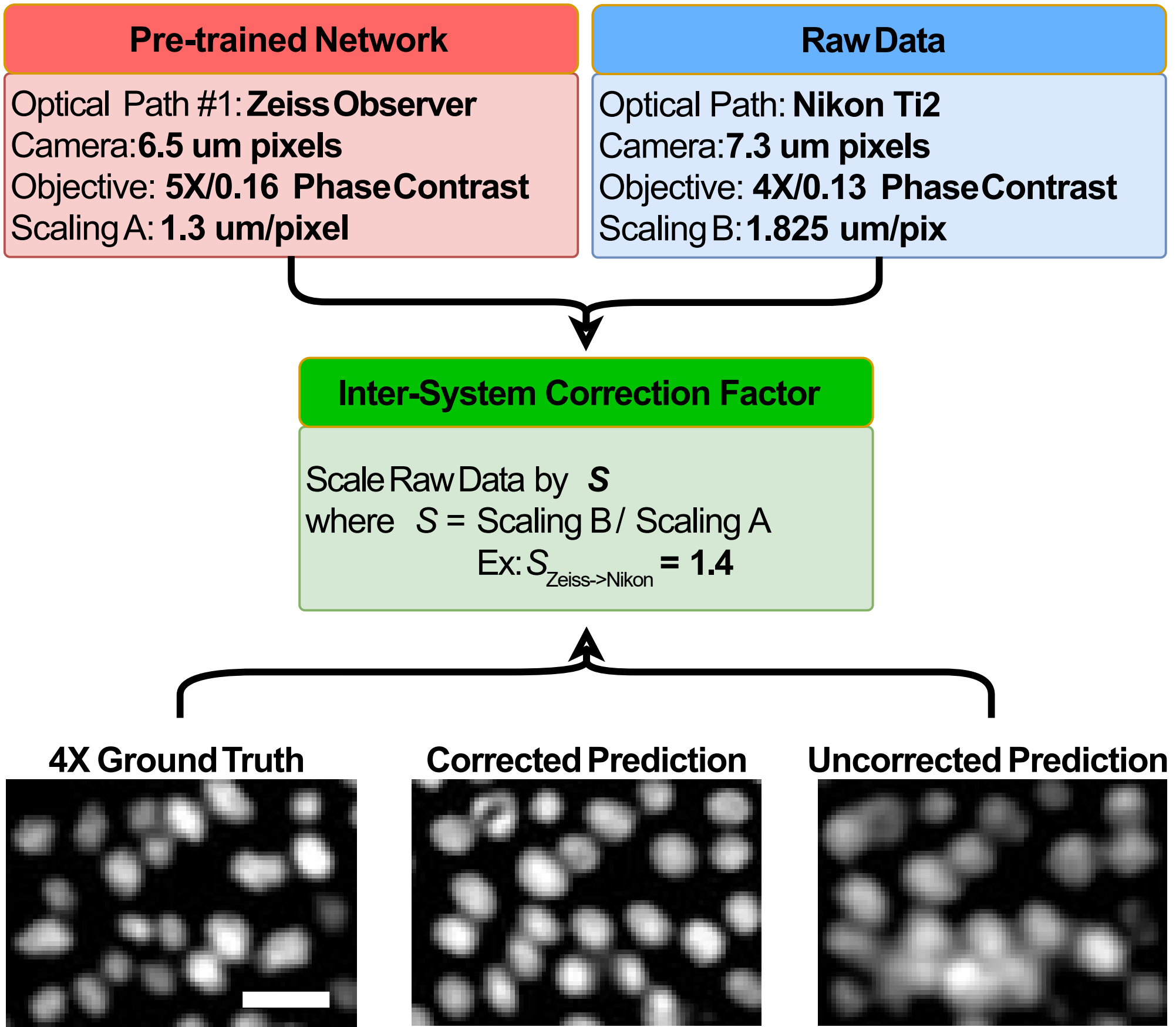


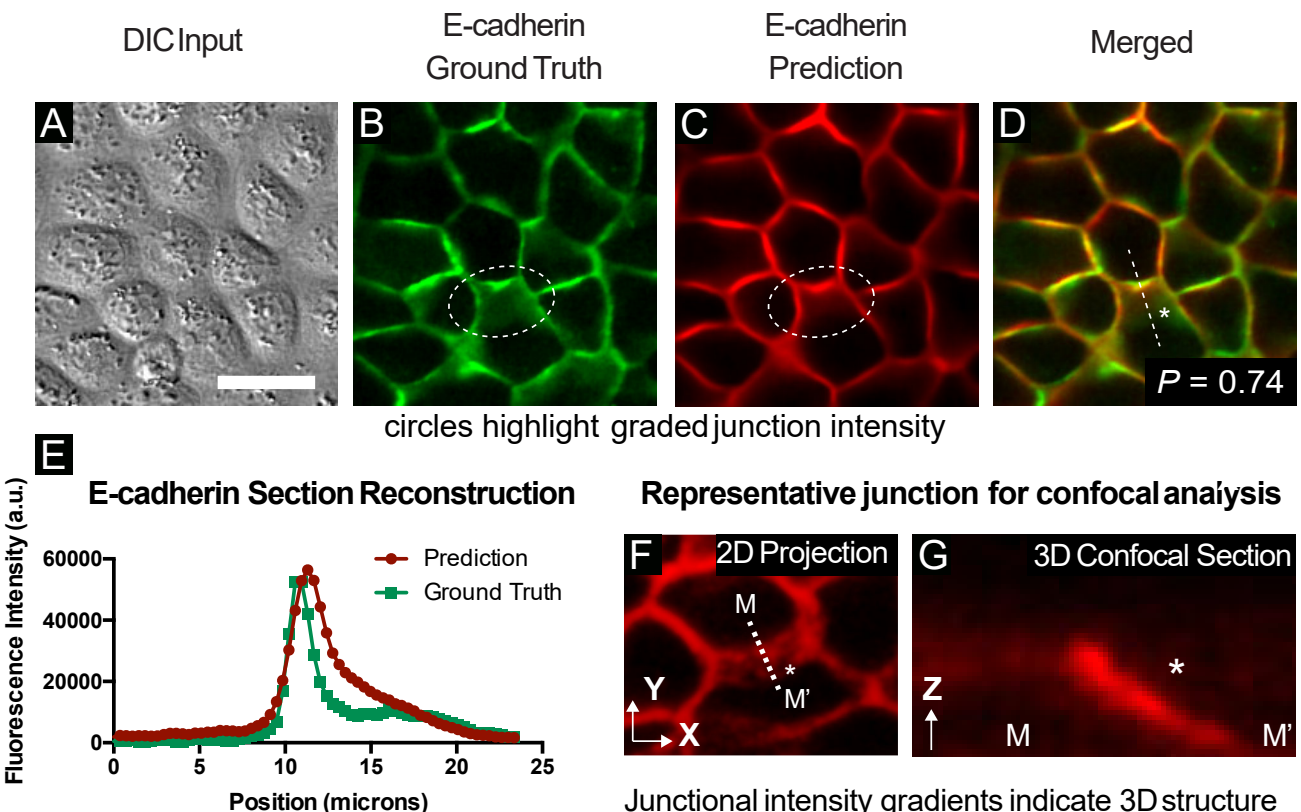
?

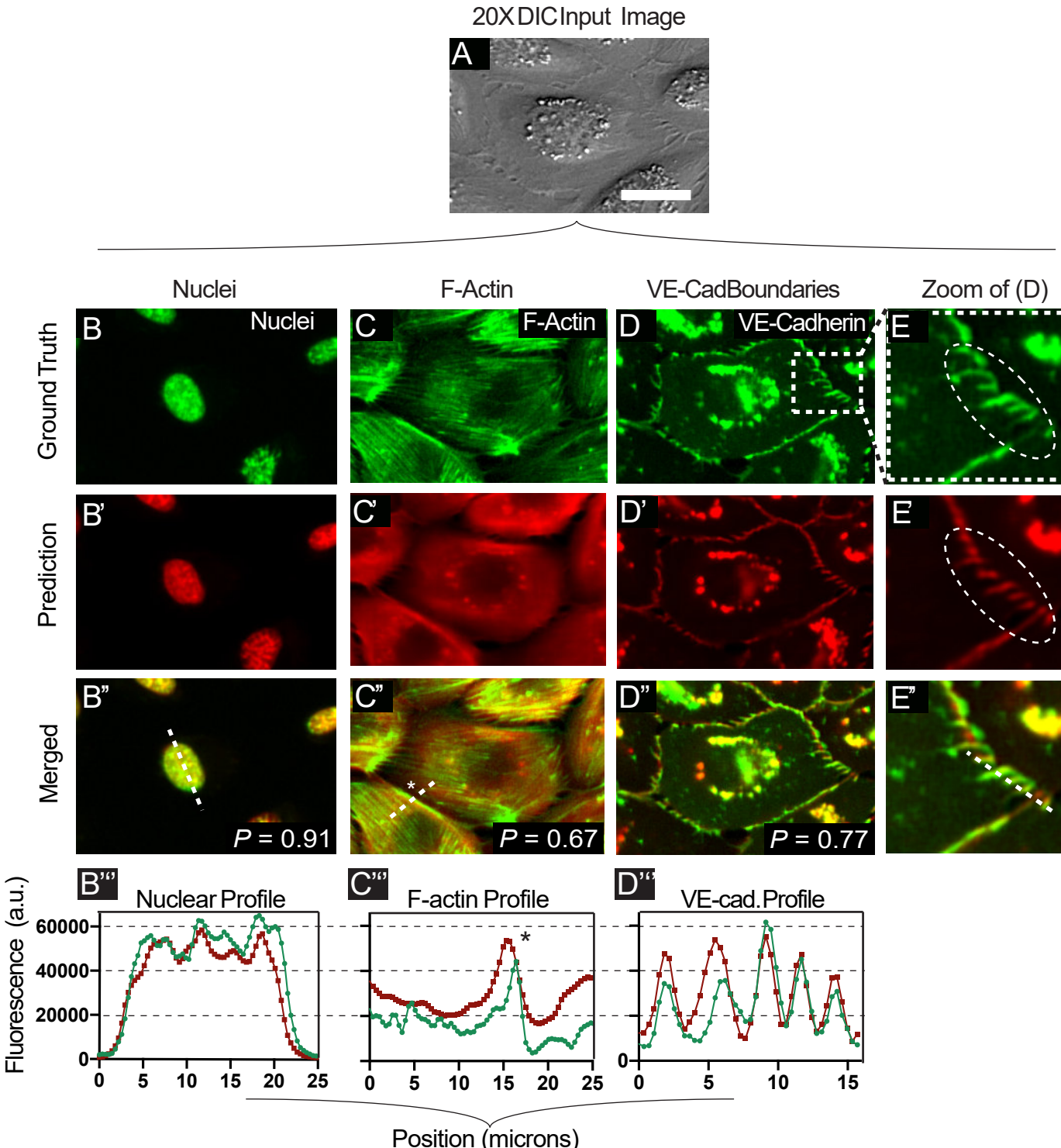
Practical Tasks

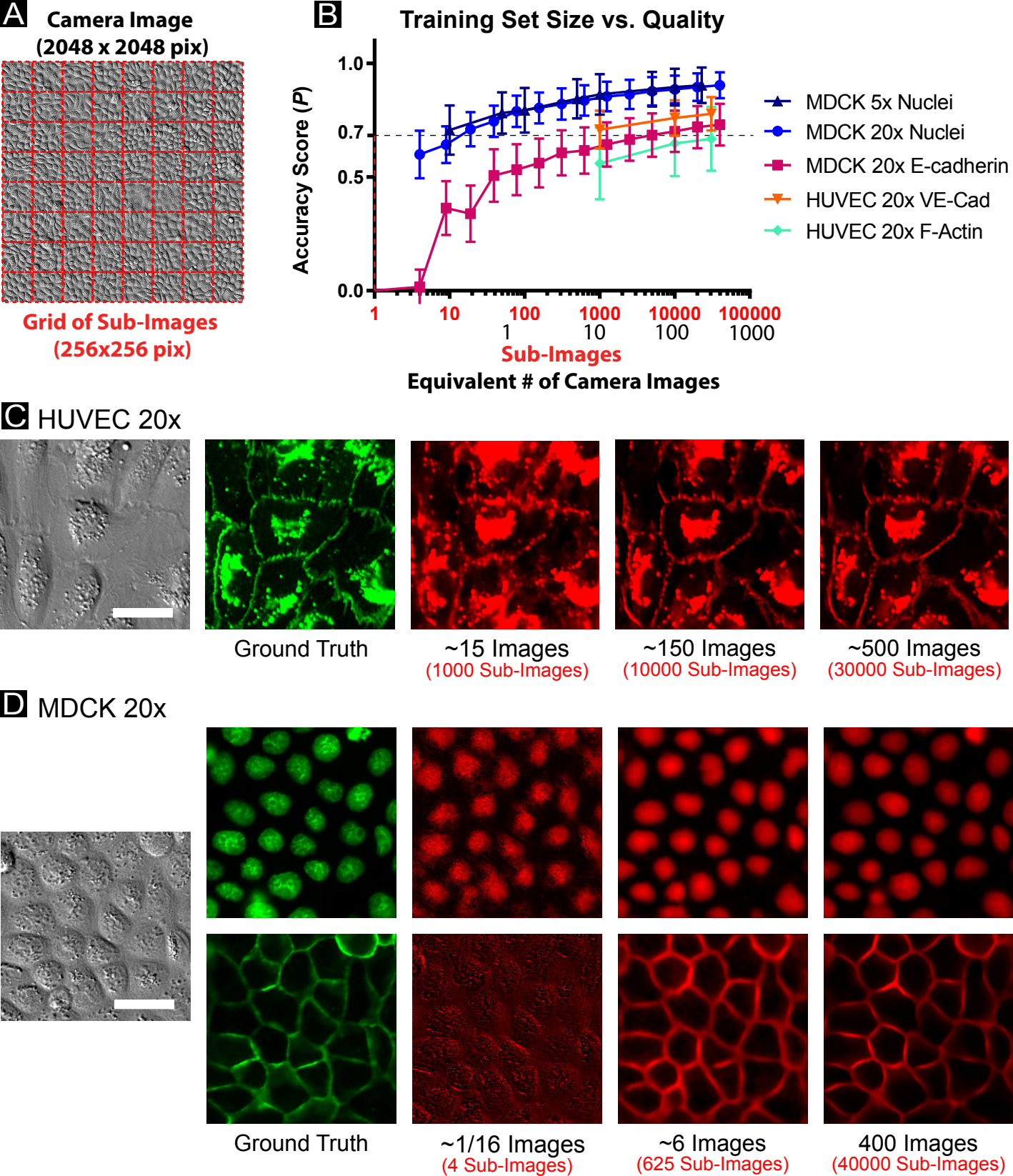
- Nuclei and cell counts
- Boundaries/Junctions
- Fine structures analysis


Phase Input
Nuclei Ground Truth
Nuclei Prediction
Merged
MDCK 5x
A
B
C
D
P = 0.91
Phase Input
Nuclei Ground Truth
Nuclei Prediction
Merged
Keratinocyte 10x
E
F
G
H
P = 0.90
I
Nuclear Area Comparison
Area (microns²)
300
200
100
0
Ground Truth
Predicted
Ground Truth
Predicted
J
Centroid-Centroid Displacements
Displacement (microns)
4
3
2
1
0
MDCK 5x
Keratinocyte 10x
MDCK 5x
Keratinocyte 10x
K
K'
K''
t = 0 hrs
t = 12 hrs
t = 24 hrs


Pre-trained Network
Optical Path #1: Zeiss Observer
Camera: 6.5 um pixels
Objective: 5X/0.16 Phase Contrast
Scaling A: 1.3 um/pixel
Raw Data
Optical Path: Nikon Ti2
Camera: 7.3 um pixels
Objective: 4X/0.13 Phase Contrast
Scaling B: 1.825 um/pix
Inter-System Correction Factor
Scale Raw Data by S
where S = Scaling B / Scaling A
Ex: S Zeiss->Nikon = 1.4
4X Ground Truth
Corrected Prediction
Uncorrected Prediction


DIC Input
E-cadherin Ground Truth
E-cadherin Prediction
Merged
A
B
C
D
*
P = 0.74
circles highlight graded junction intensity
E
E-cadherin Section Reconstruction
Fluorescence Intensity (a.u.)
60000
40000
20000
0
0
5
10
15
20
25
Position (microns)
Prediction
Ground Truth
Representative junction for confocal analysis
F
2D Projection
M
*
M'
Y
X
G
3D Confocal Section
*
Z
M
M'
Junctional intensity gradients indicate 3D structure


20X DIC Input Image
A
Nuclei
F-Actin
VE-Cad Boundaries
Zoom of (D)
Ground Truth
Prediction
Merged
B
Nuclei
C
F-Actin
D
VE-Cadherin
E
B'
C'
D'
E'
B"
C"
D"
E"
P = 0.91
P = 0.67
P = 0.77
B'''
Nuclear Profile
C'''
F-actin Profile
D'''
VE-cad. Profile
Fluorescence (a.u.)
60000
40000
20000
0
0
5
10
15
20
25
Position (microns)


A
Camera Image
(2048 x 2048 pix)
Grid of Sub-Images
(256x256 pix)
B
Training Set Size vs. Quality
Accuracy Score (P)
1.0
0.7
0.5
0.0
1
10
100
1000
10000
100000
1
10
100
1000
Sub-Images
Equivalent # of Camera Images
MDCK 5x Nuclei
MDCK 20x Nuclei
MDCK 20x E-cadherin
HUVEC 20x VE-Cad
HUVEC 20x F-Actin
C HUVEC 20x
Ground Truth
~15 Images
(1000 Sub-Images)
~150 Images
(10000 Sub-Images)
~500 Images
(30000 Sub-Images)
D MDCK 20x
Ground Truth
~1/16 Images
(4 Sub-Images)
~6 Images
(625 Sub-Images)
400 Images
(40000 Sub-Images)