Foreground-aware Virtual Staining for Accurate 3D Cell Morphological Profiling

Alexandr A. Kalinin [1 2] **Paula Llanos** [1] **Theresa Maria Sommer** [3] **Giovanni Sestini** [3] **Xinhai Hou** [2] **Jonathan Z. Sexton** [4 5] **Xiang Wan** [6] **Ivo Dinov** [2 7 8] **Brian D. Athey** [2] **Nicolas Rivron** [3] **Anne E. Carpenter** [1] **Beth Cimini** [1] **Shantanu Singh** [1] **Matthew J. O'Meara** [2 5]

Abstract

Microscopy enables direct observation of cellular morphology in 3D, with transmitted-light methods offering low-cost, minimally invasive imaging and fluorescence microscopy providing specificity and contrast. Virtual staining combines these strengths by using machine learning to predict fluorescence images from label-free inputs. However, training of existing methods typically relies on loss functions that treat all pixels equally, thus reproducing background noise and artifacts instead of focusing on biologically meaningful signals. We introduce Spotlight, a simple yet powerful virtual staining approach that guides the model to focus on relevant cellular structures. Spotlight uses histogram-based foreground estimation to mask pixel-wise loss and to calculate a Dice loss on soft-thresholded predictions for shape-aware learning. Applied to a 3D benchmark dataset, Spotlight improves morphological representation while preserving pixel-level accuracy, resulting in virtual stains better suited for downstream tasks such as segmentation and profiling.

[1]Imaging Platform, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA [2]Gilbert S. Omenn Department of Computational Medicine and Bioinformatics, University of Michigan Medical School, Ann Arbor, MI 48109, USA [3]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna BioCenter (VBC), 1030 Vienna, Austria [4]Department of Internal Medicine – Gastroenterology, University of Michigan Medical School, Ann Arbor, MI 48109, USA [5]Department of Medicinal Chemistry, College of Pharmacy, University of Michigan, Ann Arbor, MI 48109, USA [6]Shenzhen Research Institute of Big Data, Chinese University of Hong Kong–Shenzhen, Shenzhen 518172, Guangdong, China [7]Statistics Online Computational Resource (SOCR), University of Michigan, Ann Arbor, MI 48109, USA [8]Health Behavior and Biological Sciences, University of Michigan School of Nursing, Ann Arbor, MI 48109, USA. Correspondence to: Alexandr A. Kalinin <akalinin@broadinstitute.org>, Matthew J. O'Meara <momeara@umich.edu>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025. Copyright 2025 by the author(s).

1. Introduction

Microscopy is an essential tool for studying cellular states, enabling direct observation of cell morphology and dynamics with high spatial and temporal resolution. Transmitted-light microscopy has low cost, straightforward implementation, and minimal invasiveness, which make it ideal for long-term and live-cell imaging. However, it is limited by the inherently low specificity and contrast. Fluorescence microscopy dramatically improves contrast and enables multiplexed imaging of several biological structures simultaneously by utilizing fluorescent probes that selectively label cellular targets. However, fluorescence microscopy requires more complex and costly instrumentation and sample preparation, frequently induces phototoxicity and photobleaching, and usually requires samples to be fixed, precluding studies of living samples over time. Moreover, the limited availability of fluorescence spectra restricts its utility for multiplexing. These limitations motivated the development of computational methods collectively termed "virtual staining" or "in silico labeling" (Ounkomol et al., 2018; Christiansen et al., 2018). Virtual staining leverages machine learning (ML) algorithms to predict fluorescent signals from transmitted-light microscopy images, aiming to integrate the simplicity, minimal invasiveness, and affordability of transmitted-light methods with the specificity and contrast of fluorescence imaging. Thus, virtual staining approaches offer a promising strategy to enhance throughput, minimize experimental variability, and facilitate longer-term, minimally invasive observation of live cells, significantly expanding the applicability and value of microscopy-based cellular studies.

Virtual staining techniques originated from image-to-image translation research in computer vision, where models map paired image domains such as sketches to photographs, or day scenes to night scenes (Isola et al., 2017). Adopting this idea for virtual staining, Ounkomol et al. (2018) and Christiansen et al. (2018) introduced U-net–like models (Ronneberger et al., 2015) trained with pixel-wise losses for predicting fluorescence signals for multiple subcellular structures in 3D from brightfield or phase-contrast stacks, establishing the feasibility of the overall approach. Subsequent works have experimented with input modalities and custom imaging setups (Cheng et al., 2021; Guo et al., 2020)

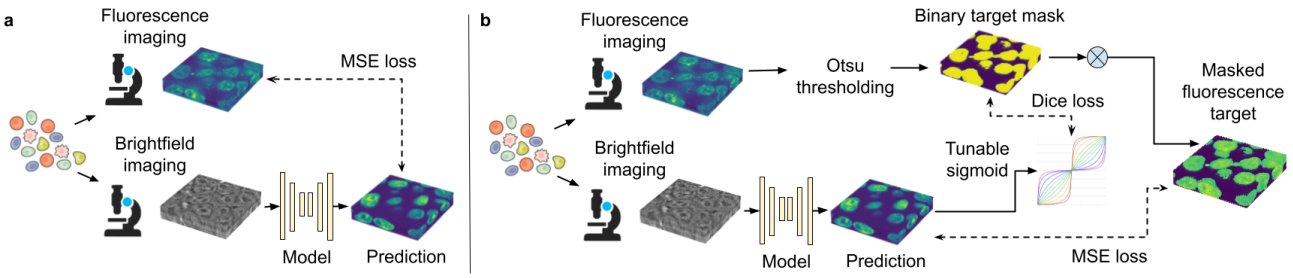

Figure 1. A schematic overview of Spotlight. a: Typical virtual staining models are trained using a pixel-wise loss such as mean squared error (MSE) on the whole image. b: Spotlight uses foreground estimation obtained by histogram thresholding to restrict pixel-wise loss to foreground areas and also employs soft-thresholding of the prediction to compute segmentation loss.

and enhanced architectures (Wang et al., 2021; Zhou et al., 2023), but the core training scheme remained supervised regression against real fluorescence images. The training objectives in most of these models are dominated by pixel-wise loss functions such as mean squared error (MSE) or mean absolute error (MAE), sometimes supplemented with perceptual or structural similarity terms (Liu et al., 2024). Adversarial losses have been introduced to improve realism, yet they are coupled with per-pixel MSE/MAE terms to stabilize training (Rivenson et al., 2019; Wang et al., 2023).

These approaches overlook a fundamental difference between biological microscopy images and natural images: translating transmitted-light microscopy into fluorescence images is not merely stylistic but must accurately capture underlying biological structures. Unlike natural images, where even background pixels typically convey semantic information such as representing the sky, trees, or buildings, microscopy images often have biologically irrelevant backgrounds. Because most existing virtual staining models are trained with pixel-wise losses (Figure 1a), they treat all pixels as equally important and learn to reproduce noise and artifacts from biologically irrelevant background regions of original fluorescence images, making downstream analysis challenging. This limitation becomes even more critical in 3D imaging, where axial aberrations and elongation distort cellular structures. As a result, predicted fluorescence images often inherit the segmentation difficulties of the original fluorescence data, complicating the segmentation of individual cells, which is a common downstream task.

Here, we introduce a new method, termed Spotlight, to address these challenges by explicitly guiding virtual staining models to focus on predicting cellular structures in the foreground (FG), while ignoring background (BG). Spotlight incorporates two key components (Figure 1b). First, we apply a simple histogram thresholding procedure to approximate the location of FG structures, using this to mask traditional pixel-wise loss functions and restrict supervision to informative regions of the image. Second, we encourage accurate modeling of cell shape by soft-thresholding the predicted fluorescence output and computing a segmentation loss with respect to the FG mask. The combined training objective balances precise FG intensity prediction with geometric fidelity, enabling the model to generate morphologically accurate virtual stains, thereby facilitating downstream tasks such as segmentation and profiling. We demonstrate Spotlight on a benchmark dataset and show that it improves 3D morphological representation compared to standard training approaches.

2. Spotlight

Training virtual staining models with pixel-wise loss functions is especially limiting in 3D microscopy, where the background dominates the voxel count, while containing noise and optical aberration artifacts without real biological signal. Our main insight is that even a weak approximation of FG is sufficient to prioritize training on meaningful content and suppress irrelevant noise from BG regions.

Two components of Spotlight incorporates use the FG approximation to guide the training process (Figure 1b). The first component applies FG mask to a traditional pixel-wise loss function (e.g., MSE), thereby restricting supervision to pixels likely to correspond to cellular structures. The second soft-thresholds the model's predicted output and computes Dice loss (Milletari et al., 2016) against the binary FG mask, encouraging preservation of shape and structure in the predicted fluorescence. The masked pixel loss enables the model to ignore uninformative BG regions, while the Dice loss promotes spatial coherence and reduces noise and artifacts in predictions. Together, these losses create a training signal that emphasizes morphological fidelity without compromising pixel-level accuracy.

Foreground estimation: To restrict supervision to biologically informative regions, Spotlight uses a FG mask

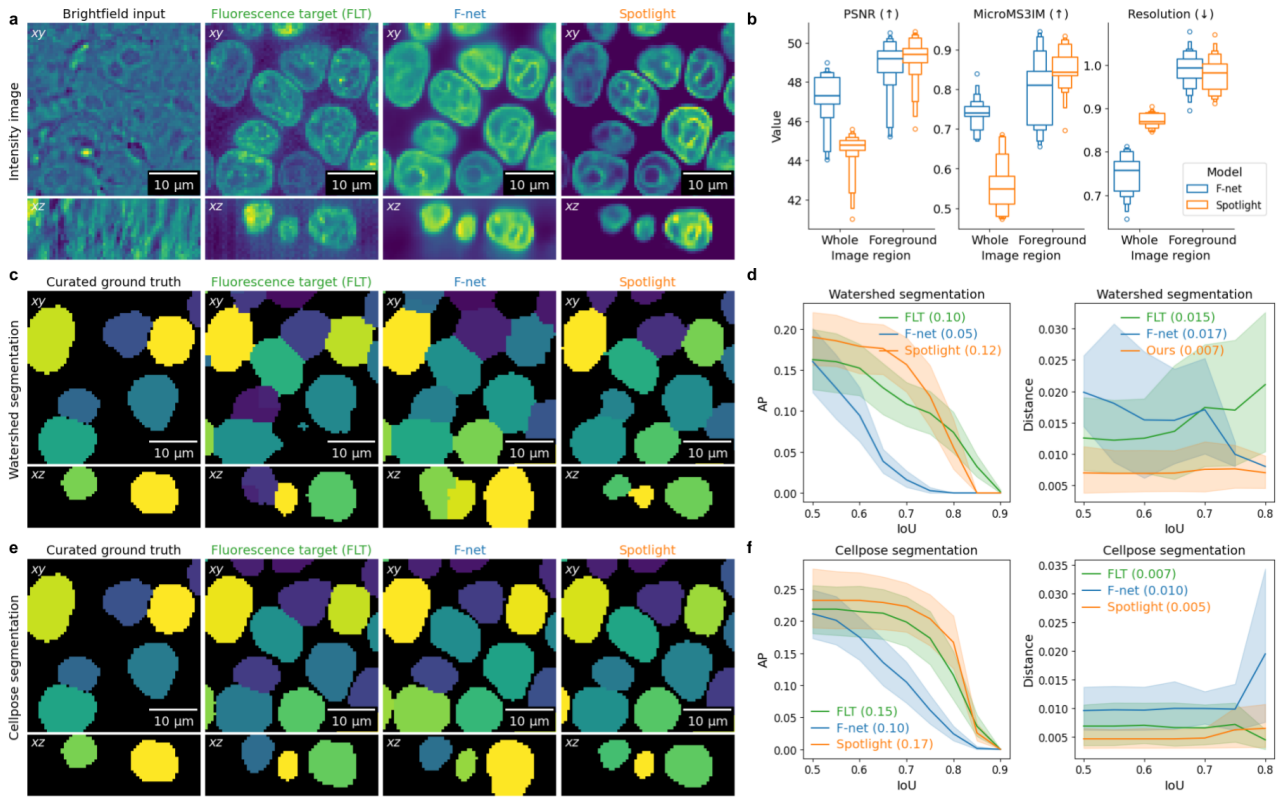

Figure 2. Performance evaluation on the Allen Cell Label-free Imaging dataset. a: Visualizations of F-net and Spotlight predictions demonstrate differences in BG prediction. b: Image quality metrics indicate that Spotlight preserves the same level of detail in FG intensities. c,e: Visualizations of wathershed and Cellpose segmentation results show artifacts when segmenting F-net predictions. d,f: Segmentation and feature-level metrics indicate that less noisy predictions by Spotlight produce more accurate morphologies.

derived from the target fluorescence image using Otsu's method (Otsu, 1979).

Let Y denote the ground truth fluorescence image, $\hat{Y}$ the predicted image, and T the Otsu threshold computed from the histogram of Y as the intensity value that maximizes inter-class variance. We then define the binary foreground mask $M \in 0, 1^{H \times W \times D}$ by thresholding Y at T, such that $M_{i,j,k} = 1$ if $Y_{i,j,k} \geq T$, and 0 otherwise. Using this mask, we compute the masked mean squared error (MMSE) loss:

$$\mathcal{L}_{\text{MMSE}} = \frac{1}{\sum M} \sum_{i,j,k} M_{i,j,k} \left(Y_{i,j,k} - \hat{Y}_{i,j,k} \right)^2$$

Morphology-preserving component: To encourage morphological fidelity, we apply a differentiable soft-thresholding function to the prediction. We use a normalized tunable sigmoid (Emery, 2022) defined as:

$$\sigma_k(x) = \frac{x - kx}{k - 2k|x| + 1}$$

where k controls sharpness and nonlinearity.

Applying this to $\hat{Y}$, we compute the Dice loss against the binary foreground mask:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum \sigma_k(\hat{Y}) \cdot M}{\sum \sigma_k(\hat{Y}) + \sum M + \epsilon}$$

The total loss used for training is a weighted combination:

$$\mathcal{L}_{\text{Spotlight}} = \lambda \cdot \mathcal{L}_{\text{MMSE}} + (1 - \lambda) \cdot \mathcal{L}_{\text{Dice}}$$

This loss formulation emphasizes accurate reconstruction of informative regions and structurally coherent fluorescence predictions.

3. Experiments

3.1. Experimental setup

We used the "DNA" subset (Hoechst-lebeled nuclei) of the Allen Institute for Cell Science Label-free Determination dataset (Ounkomol et al., 2018). As a baseline, we re-trained the commonly used 3D F-net model proposed by Ounkomol

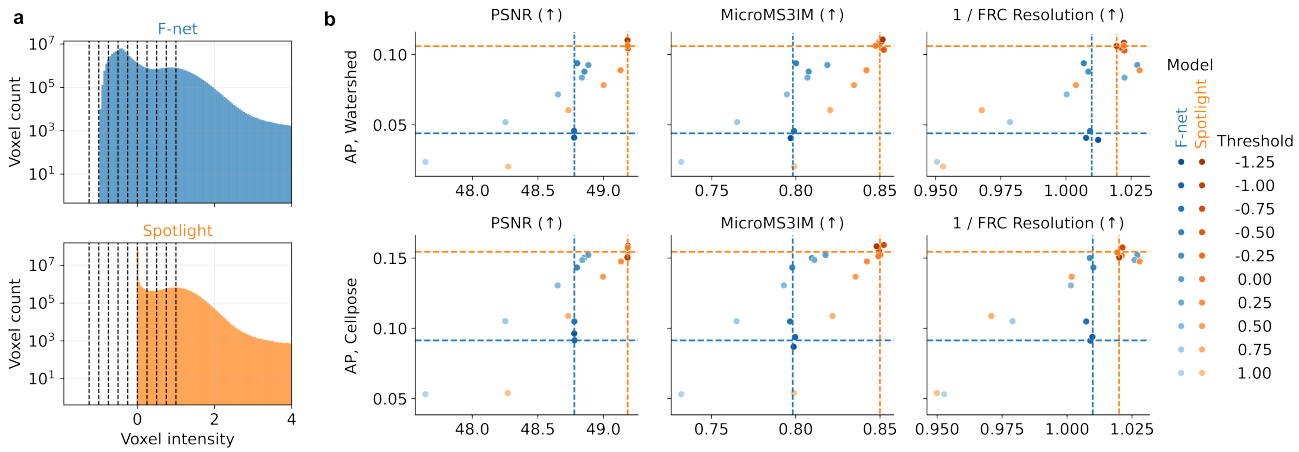

Figure 3. Performance comparison on thresholded predictions. a: Intensity histograms for F-net and Spotlight test predictions. (black dashed lines: tested post-processing thresholds). b: Segmentation and image-quality metrics calculated on thresholded F-net and Spotlight predictions. Colored dashed lines correspond to evaluation of F-net and Spotlight predictions without any thresholding.

et al. (2018). We evaluated predictions at the pixel, segmentation, and feature levels. At the pixel level, we compute prediction image quality metrics both on the whole image and only the foreground, including peak signal-to-noise ratio (PSNR), microscopy-specific 3D structural similarity (MicroMS3IM) (Ashesh et al., 2025), and the Fourier ring correlation-based (FRC) image resolution estimate (Koho et al., 2019). We tested two separate segmentation approaches: Otsu's thresholding (Otsu, 1979) followed by the watershed algorithm (Soille & Ansoult, 1990) and Cellpose with the pre-trained "nuclei" model (Stringer et al., 2021), and reported average precision (AP) calculated at different intersection over union (IoU) levels against the manually curated ground-truth segmentation masks provided with the dataset (Ounkomol et al., 2018). At the feature level, we report cosine distance between 3D measurement profiles extracted from the target fluorescence images and profiles extracted from the model predictions.

3.2. Results

Our results show that incorporating a foreground-aware loss substantially improves virtual staining of nuclei compared to the baseline F-net by reducing artifacts such as axial elongation (Figure 2). Trained with MSE loss, F-net predictions are smoothed compared to the original fluorescence images (Figure 2a). While this suppressed noise, cell edges remain blurred and the "halo" of axial elongation is clearly present. In contrast, Spotlight preserves nuclear boundaries and produces predictions free of these artifacts (Figure 2a). Quantitatively, image-quality metrics worsen when computed over the entire volume but are consistent when restricted to the FG mask (Figure 2b). This improved FG/BG separation translates directly into more accurate segmentations—using both watershed (Figure 2c) and Cellpose(Figure 2e)—and yields morphological measurements that align more closely with the ground truth (Figure 2d,f).

We also compared the voxel-intensity distributions of F-net and Spotlight predictions (Figure 3a). Whereas F-net produces a clear bimodal BG/FG histogram, Spotlight sharply collapses background variability toward zero, creating a tighter separation between background and foreground. To test how this affects segmentation, we thresholded each prediction volume at a range of intensity values spanning both BG/FG modes, and then ran two segmentation pipelines (watershed and Cellpose). For F-net, filtering out BG intensities improved both segmentation and image-quality scores—bringing them nearly up to Spotlight's levels—whereas Spotlight's metrics stayed consistent (Figure 3b). We show how learning from informative foreground regions in virtual staining acts allows to avoid complicated post-processing, to simplify segmentation and to improve its utility for 3D nuclear morphometry.

4. Conclusion

Less invasive, cost-effective computational imaging will be valuable for a wide range of applications in biomedicine, such as live-cell assays. Methods that do not compromise on morphological detail and preserve cellular geometry and structure can be helpful for downstream tasks such as segmentation, phenotyping, and profiling. As ML methods continue to shape biological imaging, it is essential to consider the statistical, structural, and physical properties of the data—not just to improve prediction accuracy, but to ensure that models yield biologically relevant results.

Acknowledgements

This work was pertially supported by Chinese Key-Area Research and Development Program of Guangdong Province (2020B0101350001). This work was partially supported by the Human Frontier Science Program (RGY0081/2019 to S.S.) and a grant from the National Institutes of Health NIGMS (R35 GM122547 to A.E.C.). Xin Rong of the University of Michigan donated NVIDIA TITAN X GPU used for this research, and the NVIDIA Corporation donated the TITAN Xp GPU used to execute the computationally intensive protocol.

Impact Statement

This work aims to reduce the cost, phototoxicity, and reagent use of fluorescence microscopy by generating virtual stains from label-free images. This can accelerate high-content screening and make advanced imaging more accessible in resource-limited labs. Because downstream biological conclusions may rely on the fidelity of virtual stains, inaccurate predictions could mislead research or clinical workflows if used uncritically. We therefore advocate reporting rigorous benchmarks so that users can assess reliability in their own settings before deployment.

References

Ashesh, A., Deschamps, J., and Jug, F. MicroSSIM: Improved structural similarity for comparing microscopy data. In Del Bue, A., Canton, C., Pont-Tuset, J., and Tommasi, T. (eds.), *Computer Vision – ECCV 2024 Workshops*, Lecture Notes in Computer Science. Springer Cham, July 2025.

Cheng, S., Fu, S., Kim, Y. M., Song, W., Li, Y., Xue, Y., Yi, J., and Tian, L. Single-cell cytometry via multiplexed fluorescence prediction by label-free reflectance microscopy. *Sci Adv*, 7(3), January 2021.

Christiansen, E. M., Yang, S. J., Ando, D. M., Javaherian, A., Skibinski, G., Lipnick, S., Mount, E., O'Neil, A., Shah, K., Lee, A. K., Goyal, P., Fedus, W., Poplin, R., Esteva, A., Berndl, M., Rubin, L. L., Nelson, P., and Finkbeiner, S. In silico labeling: Predicting fluorescent labels in unlabeled images. *Cell*, 173(3):792–803.e19, April 2018.

Emery, D. H. Normalized tunable sigmoid function, 2022. URL https://dhemery.github.io/DHE-Modules/technical/sigmoid/. DHE Modules.

Guo, S.-M., Yeh, L.-H., Folkesson, J., Ivanov, I. E., Krishnan, A. P., Keefe, M. G., Hashemi, E., Shin, D., Chhun, B. B., Cho, N. H., Leonetti, M. D., Han, M. H., Nowakowski, T. J., and Mehta, S. B. Revealing architectural order with quantitative label-free imaging and deep learning. *Elife*, 9, July 2020.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. Array programming with numpy. *Nature*, 585(7825):357–362, 2020.

Hunter, J. D. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, July 2017.

Kalinin, A. A., Hou, X., Ade, A. S., Fon, G.-V., Meixner, W., Higgins, G. A., Sexton, J. Z., Wan, X., Dinov, I. D., O'Meara, M. J., et al. Valproic acid-induced changes of 4d nuclear morphology in astrocyte cells. *Molecular Biology of the Cell*, 32(18):1624–1633, 2021.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015.

Koho, S., Tortarolo, G., Castello, M., Deguchi, T., Diaspro, A., and Vicidomini, G. Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nat Commun*, 10(1):3103, July 2019.

Liu, Z., Hirata-Miyasaki, E., Pradeep, S., Rahm, J., Foley, C., Chandler, T., Ivanov, I., Woosley, H., Lao, T., Balasubramanian, A., Marreiros, R., Liu, C., Leonetti, M., Aviner, R., Arias, C., Jacobo, A., and Mehta, S. B. Robust virtual staining of landmark organelles. June 2024.

Milletari, F., Navab, N., and Ahmadi, S.-A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, October 2016.

Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.*, 9(1):62–66, January 1979.

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

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang,

L., Bai, J., and Chintala, S. PyTorch: An imperative style, High-Performance deep learning library. In *Proceedings of the 33rd International Conference on Neural Information Processing Systems*, 2019.

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ter Haar Romeny, B., Zimmerman, J. B., and Zuiderveld, K. Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, 39(3):355–368, 1987.

RAPIDSAI contributors. cuCIM: RAPIDS GPU-accelerated image processing library. https://github.com/rapidsai/cucim, 2025. Version 25.06.0.

Rivenson, Y., Wang, H., Wei, Z., de Haan, K., Zhang, Y., Wu, Y., Günaydın, H., Zuckerman, J. E., Chong, T., Sisk, A. E., Westbrook, L. M., Wallace, W. D., and Ozcan, A. Virtual histological staining of unlabelled tissue-autofluorescence images via deep learning. *Nat Biomed Eng*, 3(6):466–477, June 2019.

Ronneberger, O., Fischer, P., and Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Lecture Notes in Computer Science*, Lecture notes in computer science, pp. 234–241. Springer International Publishing, Cham, 2015.

Soille, P. J. and Ansoult, M. M. Automated basin delineation from digital elevation models using mathematical morphology. *Signal Processing*, 20(2):171–182, June 1990.

Stirling, D. R., Swain-Bowden, M. J., Lucas, A. M., Carpenter, A. E., Cimini, B. A., and Goodman, A. Cellprofiler 4: improvements in speed, utility and usability. *BMC bioinformatics*, 22:1–11, 2021.

Stringer, C., Wang, T., Michaelos, M., and Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods*, 18(1):100–106, January 2021.

Van der Walt, S., Schönberger, J. L., Nunez-Iglesias, J., Boulogne, F., Warner, J. D., Yager, N., Gouillart, E., and Yu, T. scikit-image: image processing in python. *PeerJ*, 2:e453, 2014.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., et al. Scipy 1.0: fundamental algorithms for scientific computing in python. *Nature methods*, 17(3):261–272, 2020.

Wang, R., Butt, D., Cross, S., Verkade, P., and Achim, A. Bright-field to fluorescence microscopy image translation for cell nuclei health quantification. *Biol Imaging*, 3:e12, June 2023.

Wang, Z., Xie, Y., and Ji, S. Global voxel transformer networks for augmented microscopy. *Nat. Mach. Intell.*, 3(2):161–171, January 2021.

Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60):3021, 2021. doi: 10.21105/joss.03021. URL https://doi.org/10.21105/joss.03021.

Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman (eds.), *Proceedings of the 9th Python in Science Conference*, pp. 56 – 61, 2010. doi: 10.25080/Majora-92bf1922-00a.

Zhou, D., Gu, C., Xu, J., Liu, F., Wang, Q., Chen, G., and Heng, P.-A. RepMode: Learning to re-parameterize diverse experts for subcellular structure prediction. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2023.

A. Appendix

A.1. Dataset details

We used the "DNA" subset (Hoechst-lebeled nuclei) of the Allen Institute for Cell Science Label-free Determination dataset (Ounkomol et al., 2018). This dataset includes manually curated ground truth segmentation masks, which we use to evaluate the segmentation quality of nuclei. Masks that did not pass quality control were removed, which led to some cells missing the ground truth segmentation masks (see Figure 2c, "Curated ground truth").

A.2. Baseline details

F-net: F-net (Ounkomol et al., 2018) is a fully convolutional U-Net variant that consists of layers that perform one of three convolution types, followed by a batch normalization and rectified linear unit (ReLU) operation. The convolutions are either 3-pixel convolutions with a stride of 1 pixel on zero-padded input (such that 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 pixels (to double the spatial area of the output). There are no normalization or ReLU operations on the last layer of the network.

A.3. Training and evaluation

Training protocol: We followed the original data split, preprocessing, and training protocols from Ounkomol et al. (2018). Specifically, we resized all z-stacks using cubic interpolation to achieve isotropic voxel dimensions of $0.29\times0.29\times0.29\mu m^3$. We trained models on batches of 3D patches ($32\times64\times64$ px^3, ZYX), which were randomly subsampled uniformly both across all training images and spatially within an image. All models were trained using the Adam optimizer (Kingma & Ba, 2015) with a learning rate of 0.001, a batch size of 24, and with beta values of 0.5 and 0.999 for 50,000 mini-batch iterations. All models were implemented in Python using the PyTorch package (Paszke et al., 2019).

Evaluation protocol: Evaluation followed Ounkomol et al. (2018), including cropping the input image such that its size in any dimension is a multiple of 16.

When calculating pixel-level image quality metrics, predicted intensities were rescaled into the original fluorescence image intensity range using maximum and minimum values of the combined training set. The FRC resolution was calculated using the miplib package (Koho et al., 2019) with image zero-padding and parameter bin_delta=5.

Segmentation was performed on images downscaled by half in all dimensions. Hyperparameters for both segmentation pipelines were chosen to optimize segmentation results on the real fluorescence images. Watershed segmentation roughly followed the CellProfiler's *3d_monolayer* tutorial (Stirling et al., 2021), with the additional step of contrast-limiting adaptive histogram equalization (Pizer et al., 1987). Segmentation post-processing included removing nuclei at image edges, filling holes, and filtering out nuclei by size at the empirically chosen cutoffs.

Morphological feature extraction was performed using a combination of `regionprops` method from scikit-image (Van der Walt et al., 2014) and mesh properties (Kalinin et al., 2021).

Implementation of the evaluation protocol relied on popular Python libraries, including numpy (Harris et al., 2020), scipy (Virtanen et al., 2020), pandas (Wes McKinney, 2010), and RAPIDS cuCIM (RAPIDSAI contributors, 2025). Figures were made using matplotlib (Hunter, 2007) and seaborn (Waskom, 2021).

A.4. Spotlight details

Spotlight uses the same model architecture as F-net. The only differences are introduced in data preprocessing, patch selection, and the loss function.

Preprocessing: While F-net training standardizes 3D patches using mean and standard deviation, we replaced mean with the threshold value calculated using Otsu's method to simplify soft-thresholding of predictions in the calculation of the Dice loss.

Patch selection: In order to facilitate foreground-aware training, we also implemented patch selection during training such that only 3D patches containing at least 0.1% of foreground voxels were used for training.

Loss function: The overall loss function design is described in Figure 1 and Section 2. We empirically set sharpness parameter k in the normalized tunable sigmoid function equal to -0.95 and weight λ equal to 0.5.