Image denoising for fluorescence microscopy by self-supervised transfer learning

Yina Wang[1], Henry Pinkard[2], Shuqin Zhou[1,3], Laura Waller[4,5], Bo Huang[1,5,6]

[1] Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA 94143, USA
[2] UC Berkeley - UCSF Joint Graduate Program in Bioengineering, University of California, Berkeley, Berkeley, CA 94720, USA
[3] School of Pharmacy, Tsinghua University, Beijing, China
[4] Department of Electric Engineering Computer Sciences, University of California, Berkeley, Berkeley, CA 94720, USA
[5] Chan Zuckerberg Biohub, San Francisco, CA 94158, USA
[6] Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143, USA
*Email: bo.huang@ucsf.edu

Abstract

When using fluorescent microscopy to study cellular dynamics, trade-off typically has to be made between light exposure and quality of recorded image to balance phototoxicity and image signal-to-noise ratio. Image denoising is an important tool for retrieving information from dim live cell images. Recently, deep learning based image denoising is becoming the leading method because of its promising denoising performance achieved by leveraging available prior knowledge about noise model and samples at hand. However, the practical application of this method has seen challenges because of the requirement of task relevant big training data. In this work, we show the approach of combining self-supervised learning with transfer learning to address the above challenge. We demonstrate the application of it in subcellular fluorescent imaging, where the light exposure dose can be significantly reduced and the spatial resolution is well restored.

Introduction

Fluorescent microscopy is an indispensable technique in studying biological dynamics for detecting and quantifying molecules in subcellular compartments. Recent advances such as light-sheet microscopy [1, 2] and super-resolution microscopy [3] have enable subcellular fluorescent imaging at a high temporal and/or spatial resolution. On one hand, cells are sensitive to the excitation light. Photobleaching of the fluorophores limits the total amount of signal that can be extracted from a biological sample. For live imaging, phototoxicity, such as failure or delay of cell division or perturbation of biological processes, can occur well before substantial photobleaching is observed [4, 5]. Therefore, trade-off has to be made between light exposure and quality of recorded image. In many cases, e.g. when tracking a very rapid process over a long period of time or when the abundance of the target molecule is low, the resulted images could become extremely noisy. In these cases, image denoising is important to enable extracting useful information from the data [6-8].

Recently, deep learning based image denoising is becoming the leading method because of its promising performance achieved by leveraging available prior knowledge about noise model and the sample at hand, such as the Content-Aware Restoration (CARE) method [8]. In this method, typically a supervised convolutional neural network (CNN) is trained with a large number of noisy and clean image pairs. With this data-driven prior knowledge, and the network learns to statistically transform noisy pixels to clean ones. Still, this strategy poses two practical challenges. First, the performance of deep learning system greatly depends on the amount and quality of the training dataset [6]. Typically, hundreds to thousands of noisy and clean image pairs relevant to the task are needed for fluorescence microscopy denoising. Acquiring such training data sets require intensive effort and often dedicated experiments [8]. Moreover, it is not always possible to acquire clean images due to intrinsic constraints of certain samples. Second, supervised neural networks often have trouble in generalization for images not present or adequately represented in the training data set. They can easily memorize the training data and become prior content-aware [6, 8]. However, when applied to really images acquired under different conditions or from other types of cellular structures, hallucination artifacts can occur, producing results that appear real but are incorrect. This issue not only signifies the burden of high quality, matched training data acquisition, but also casting doubts on the ability for denoising by supervised learning to discover previously unknown biological phenomena.

More recently, self-supervised deep learning image denoising methods have been developed to address these challenges [9, 10]. These methods utilize the independence of noise among noisy images of the same sample or pixels across the same image, so that only noisy images are needed for training. This approach eliminates the need to acquire clean training datasets and can even rely solely on the images to be denoised as the training data. However, compared to supervised learning using matching training data sets, the denoising performance of these self-supervised learning methods scarifies due to the absent of prior knowledge [9].

Here, we present a denoising method that leverages transfer learning to take advantages of both supervised and self-supervised learning. The method first pre-train a deep neural network with generic and/or synthetic noisy/clean cellular image pairs using supervised learning. It then re-trains

this network with the noisy images from a specific task using self-supervised learning. We have demonstrated that this scheme solves challenge of requiring task-specific clean training data while transferring prior knowledge of fluorescence microscopy, particularly image resolution, to improve the denoising performance over self-supervised learning alone.

Results

Limited improvement of supervised learning denoising with additional temporal information

In our initial efforts to improve image denoising performance of supervised deep learning, we took advantage of the temporal consistency of structures in living cells. To capture the redundant information across images in a sequence, we used the classical U-Net architecture [12] similar to that in CARE [8] and expanded the network to include time as an additional dimension. We referred to this architecture as timeUnet. (Supplementary Fig. 1. See Supplementary Note for details of the implementation). In timeUnet, a 2D image sequence is treated as a 3D data set, with the network predicting one denoised image from a sliding window of 11 images in the sequence. To benchmark the performance, we generated synthetic noisy image sequences by adding Poisson shot noise and calibrated sCMOS camera noise to experimental high signal-to-noise ratio (SNR) confocal movies of mitochondria in cells. Synthetic images with various SNR, achieved by linearly scaling the image peak intensity to a desired value, were used to test the denoising performance across a range of SNR. We have demonstrated that, rather unsurprisingly, timeUnet out-performed CARE, which denoises individual images (Supplementary Fig. 2). In the entire range of input SNR tested, timeUnet consistently produced lower Mean Averaged Error (MAE) and higher Structural Similarity Index Measurement (SSIM) [13] values when comparing the denoised images to the ground truths (Supplementary Fig. 3). In particular, the addition of time-domain information clearly helped reducing denoising artifacts, correctly recovering images of discrete mitochondria that were connected in the CARE result (Supplementary Fig. 2). Such artifacts could be highly detrimental in the study of certain biological processes such as mitochondria fission and fusion.

Despite the improved performance, timeUnet retains the drawbacks of supervised learning. A large library of paired noisy and clean movies is still needed. Moreover, this library must match the condition of the actual denoising task. For example, any difference between the SNR of training and test input data leads a degradation of denoising performance for both CARE and timeUnet (Supplementary Fig. 4a). Particularly, increasing the SNR of test input data actually produced quantitatively worse denoising results (Supplementary Fig. 4c), which is highly concerning for practical applications. In contrast, self-supervised learning using noise2self [9] performs more consistently over the SNR range of input data, although in the case when the SNR of training and test data sets are matched, supervised learning clearly yields better results (Supplementary Fig. 4c).

Image denoising by combining self-supervised learning and transfer learning

In order to combine the benefits of both supervised and self-supervised learning, we connected them using transfer learning. For this purpose, we took advantage of the fact that an identical network architecture can be used for supervised learning as well as self-supervised learning by noise2self. The only difference is that the input and output of supervised training are a pair of noisy and clean images, whereas those of noise2self come from the same noisy image split by a pixel mask. To enable transfer learning, we first trained a network by supervised learning using generic cellular microscopy images. Then, for each denoising task, we retrain this network by noise2self using the task image set (Fig. 1a). In comparison, previously self-supervised learning work only has the second stage, with the network parameters initialized randomly.

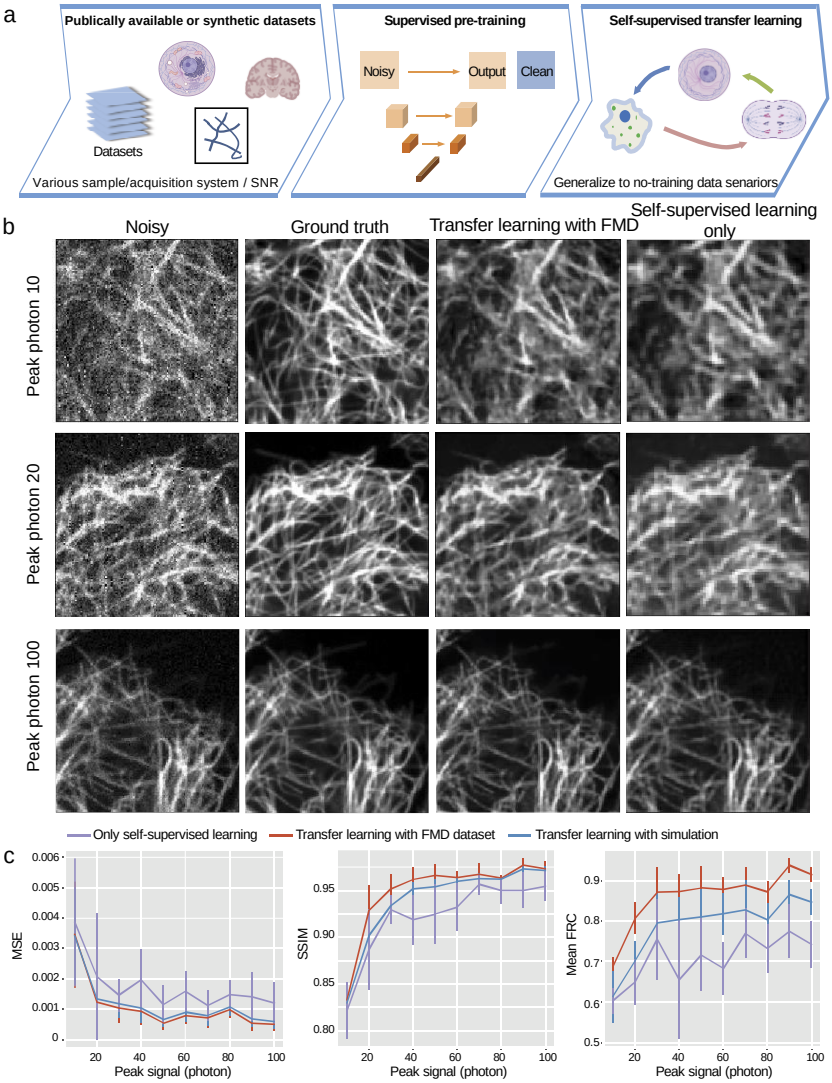

Fig. 1 Diagram (a) and performance (b-c) of self-supervised deep denoising with transfer learning. (b) Synthetic noisy image from microtubule confocal images and corresponding denoising images from self-supervised denoising with and without transfer learning. (c) The denoising performance, in terms of mean square error, structural similarity index measurement and mean Fourier Ring Correlation, as a function of the peak signal of synthetic noisy image. The cell and tissue cartoons in (a) were created at BioRender.com.

Specifically, we still used the U-Net architecture. For training at the supervised learning stage, we chose a publically available dataset, FMD [11], which contains pairs of noisy and clean images from various sample features (subcellular structures of nuclei, F actin, mitochondria; brain slices and zebrafish) and image acquisition settings (confocal, two-photon, wide-field) (Supplementary Fig. 5). The FMD dataset effectively contains 60,000 noisy image realizations from 240 field-of-view. For the self-supervised learning stage, we retrained the network by noise2self [9] on a test data set of 10 synthetic noisy images generated from high SNR experimental images in a similar way as described earlier. The retrained network generated denoised images for the same test data set (Fig. 1(b)). Compared to self-supervised learning alone (random initialization of parameters), transfer learning decreased the Mean Square Error (MSE) loss and provided a higher SSIM with the original high SNR images are the ground truth (Fig.1(c)). This indicates that prior information embedded in the pretrained network contributes to inferring lost information in the noisy measurements. The most prominent visual difference, though, is that denoising results by transfer learning clearly had much better spatial resolution (Fig. 1(b)), i.e., more recovery of high spatial frequency components. This visual impression was confirmed by quantifying the average Fourier Ring Correlation (FRC) between the denoised images and the ground truth (Fig. 1(c)). To further test the tolerance on pretraining data, we designed a set of simulated fluorescence microscopy images containing curve lines resembling microtubules (Supplementary Fig. 6). Using this simulated date set for pre-training before self-supervised retraining, the denoising performance was worse than FMD-pretrained model, but it still out-performed no-pretraining model in all three metrics (Fig. 1(c)).

We performed similar tests on synthetic images based on confocal images of lysosome structures, which are also absent in the FMD dataset. our results showed the same trends as those from microtubule images, with transfer learning clearly outperforms self-supervised learning without pretraining in all three metrics (Fig. 2). It is also evident that, in the simulation pretraining case, self-supervise retraining on noisy lysosome images can allow the model to correctly restore the punctate appearance of lysosomes, despite that morphologically it is drastically different from the curved lines used in pretraining. Another benefit of pretraining is the stability of performance in repetitive tests on the same set of lysosome images but random noises. For transfer learning self-supervision, each self-supervised training rerun gave almost exact same output (indicated by the almost invisible error bar in Fig. 2(b)); meanwhile the output from no-pretrained model is not stable at all due to the random initialization. Thus, transfer learning also acts as a stabilization of the denoising performance.

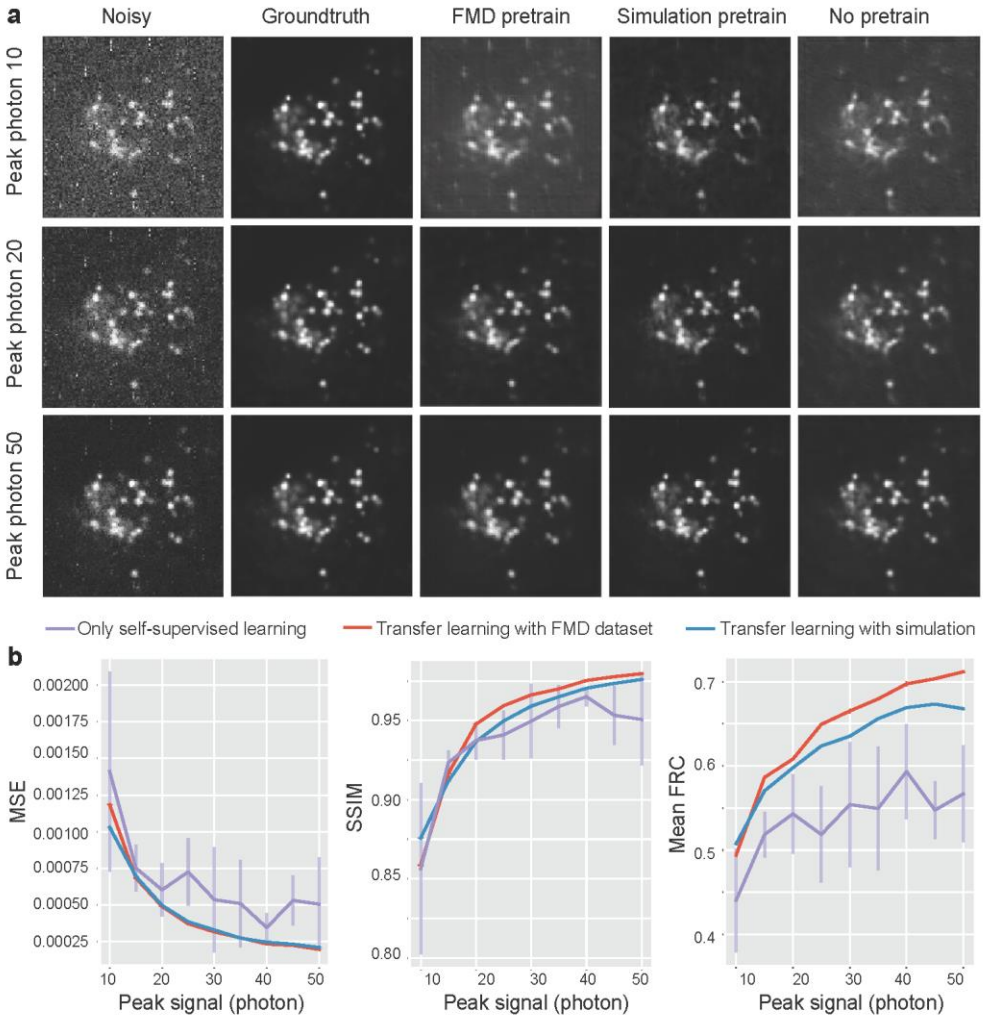

Fig.2 Performance of self-supervised deep denoising with transfer learning on lysosome images. (a) Synthetic noisy images from lysosome confocal images, ground truth confocal images, results from pretrained network using transfer learning from FMD dataset, simulated line dataset and only self-supervised learning (no-pretrain) respectively are shown in the five columns. Each row corresponds to various peak signal in photon of the noisy image. (b) The denoising performance, in terms of mean square error, structural similarity index measurement and mean Fourier Ring Correlation, as a function of the peak signal of synthetic noisy image.

To identify the knowledge learned by the model during the pretraining and retraining stage, we applied the FMD- and simulated-pretrained model directly to the denoising of microtubule and lysosome images without self-supervised retraining. The denoised images clearly had numerous morphological artifacts (Supplementary Fig. 7): FMD-pretrained model output did not display well-defined structures, whereas simulation-pretrained model generates short segments of curved lines (despite that lysosomes should appear as small puncta). There artifacts are understandable because neither microtubule nor lysosome were present in the images we used from FMD for training, and the simulated pretraining data were curved lines themselves. Compare to the transfer learning results, the retraining using self-supervised learning effectively adapted the pretraining models to the morphology and noise statistics of unseen application data. On the other hand, the

major difference between self-supervised denoising with or without transfer learning is the image resolution as measured by FRC. It suggests that self-supervised training implicitly learned a lower image resolution than the actual image resolution because of the corruption of high-frequency information in the images by the noise, while supervised learning on clean images correctly learns the image resolution. This knowledge on image resolution can be effectively transferred to the re-trained model.

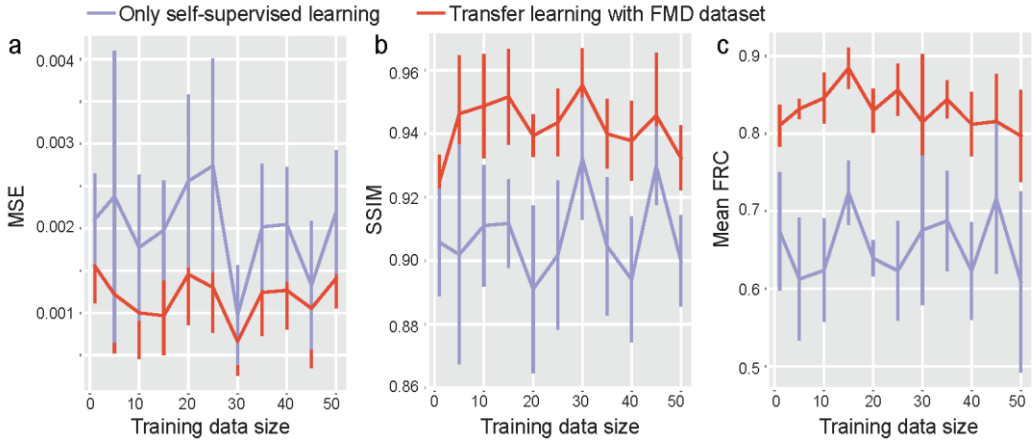

Fig.3 The MSE, SSIM and mean FRC denoising performance as a function of training data size used during the self-supervised training phase.

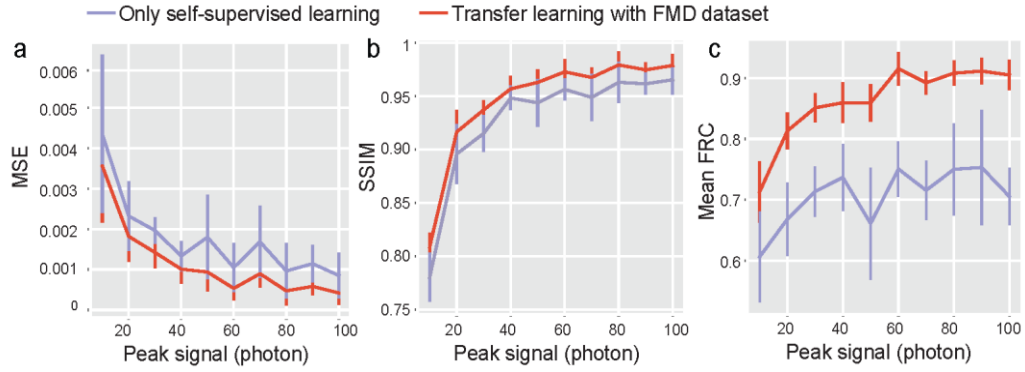

Fig.4 The MSE, SSIM and mean FRC denoising performance as a function of peak signal in synthetic noisy images when L1 is used as loss function instead of L2.

Robustness of transfer learning denoising

To characterize the robustness our approach, we tested the effect of self-supervised training data sizes on denoising performance of microtubule images (Fig. 3). Surprisingly the denoising performance does not show an obvious increase with the increase of training data size in the ranges of 1 to 50 test images, which indicates single-shot self-supervised denoising is possible. We also evaluated the denoising performance when L1 loss is used for the self-supervised training. Generally, L1 loss can better restore sharp features from noisy data compared to L2 loss. In this case, the denoising performance from transfer learning self-supervised model is still much better than no-pretraining model (Fig. 4), indicating that the source of performance improvement is not

in the loss function but because of transfer learning. To appreciate the limits of our proposed training strategy (Fig. 1(a)), we noted that our method has limited denoising performance for images with extremely low SNRs. At extremely low SNR conditions, very little useful information is in the image and there also lacks such adequate prior encoded in the pretrained network parameters. This problem potentially can be solved by including more representing low SNR images in the pretraining dataset.

Discussion

Deep learning, in particular convolutional neural network, is becoming a powerful and popular image reconstruction technique for microscopy [14-17]. Despite its successes, deep learning poses substantial challenges for discovery biological studies, including how to make the neural network generable to new tasks and how to get deep learning method into practical uses.

Firstly, current deep learning approaches typically do not generalize well to data obtained from different sample (Supplementary Fig. 5), different microscope or under different image acquisition conditions [6]. The solution of training the network with data that exceeds its memorization capability is impractical because adequate and big training dataset is typically unavailable for biological studies. On the other hand, the transfer learning approach [16, 18] can practically enable cross-study generalization and information sharing. Secondly, generating deteriorated image and ground truth image pairs for supervised training is not always practical. For example, for live sample imaging, it's impossible to acquire such image pairs without system hardware modifications. In this aspect, self-supervised learning becomes very attractive in eliminating the need for ground truth images. As a result, we took the approach of transfer learning to blend supervised training and self-supervised training to make our method practical and robust. We demonstrated the image high frequency components are transferrable. With these transferred information, we achieve blind denoising for new tasks just using a few snapshots of noisy images.

One other advantage of transfer learning is that parameters derived from supervised training becomes part of the framework to some extend. Even though the architecture or its hyper parameters cannot be tuned anymore, Unet [12] has been proved to be a robust network for multiple image reconstruction tasks, including image denoising, image segmentation and image super-resolution et.al. During self-supervised training, a few other hyper parameters, including learning rate and loss function, can be tweaked. Overall, the framework is easy to use and robust. As for the question that how the hallucination artifacts would affect new biological discover, we outlook that adding explicit physical models into the neural network architectures [19] would potentially help with reducing reconstruction artifacts.

Methods

Dataset for supervised pretraining

We used Fluorescence Microscopy Dataset (FMD) data set to perform the supervised pretraining. The dataset is downloaded from [11]. We choose this dataset because: (1) the dataset consists of representing images of multiple commonly imaged types of samples (cells, zebrafish, and mouse brain tissues) and multiple commonly used imaging modalities (commercial confocal, two-photon, and wide-field microscopes); (2) the dataset has multiple signal-to-noise ratio realizations of the same imaged scenes. The dataset composes of images from 240 field-of-views (FOV). For each FOV, 50 noisy camera frames were taken, and then image averaging was used to effectively generate the ground truth image and noisy images with various SNR.

In addition to the FMD dataset, we also generated simulation images of curved lines to test the capability of transferring sharp features of our framework. In the simulation, each image is generated by firstly convolving a normalized Gaussian point spread function (standard deviation of 1 pixel) with an object image that consists of random 10 curved lines with various lengths. Then the peak intensity in the image is linearly scaled to a desired value, and Poisson noise and calibrated sCMOS readout and gain noise is added to the image.

Generating synthetic subcellular structure images for self-supervised training and testing

In order to evaluate the performance of our framework, we first acquired high SNR confocal images of subcellular structures of mitochondria and lysosome. We seeded human HEK 293T cells on an 8-well glass bottom chamber (LabTek). In order to achieve better cell attachment, 8-well chamber was coated with Poly-L-Lysine (Sigma-Aldrich) for 20 mins before seeding cells. For microtubule staining, SiR-tubulin dye (Cytoskeleton) was added directly to the culture medium (100 nM final concentration) and incubate overnight before imaging. For lysosome staining, LysoTracker Blue DND-22 (Thermo Fisher Scientific) was added directly to the culture medium (50 nM final concentration) and incubate for 30 mins before imaging. Live-cell imaging was acquired on an inverted Nikon Ti-E microscope (UCSF Nikon Imaging Center), a Yokogawa CSU-W1 confocal scanner unit, a PlanApo VC 100x/1.4NA oil immersion objective, a stage incubator, an Andor Zyla 4.2 sCMOS camera and MicroManager software.

Then we synthesized noisy images from high SNR clean images to perform the self-supervised training and evaluate the results. The peak intensity in the image is again linearly scaled to a desired peak signal value, and Poisson shot noise and calibrated sCMOS readout and gain noise (based on the camera specifications) was added to the image using a Gaussian random number generator.

Neural network architecture and training

We used an UNet architecture implemented in noise2self paper [9, 12]. Each convolutional block consisted of two convolutional layers with 3x3 filters followed by an InstanceNorm. The number of channels was [32, 64, 128, 256]. Down-sampling used strided convolutions and up-sampling used transposed convolutions. The network was implemented in PyTorch. For supervised training, the loss is mean square error. Learning rate was set to 0.001. We trained with a batch size of 32 and 50 epochs.

We used the noise2self self-supervised training strategy. In practice, a masked image, in which a selected subset of pixels was set to zeros, was output to the neural network. The loss was only evaluated on the coordinates of that subset of pixels that are set to zeros. During training, the masked pixels were cycled to make sure a heterogeneous denoising performance over the whole image. In this way, the training process avoids identical map of the input and only relies on the independence between pixels. In this self-supervised training, the loss can be in a different form with supervised training if necessary. The learning rate was 0.0001, an order of magnitude smaller than supervised learning for network parameters fine-tuning. Because the self-supervised training was done use a few snapshots of noisy images alone, these images were processed in a single batch and early stopping with a patience number of 8 used.

Acknowledgements

We thank Joshua Batson and Loic Royer for inspirational discussions and help with the self-supervised learning code. B.H. is supported by the National Institutes of Health (R01GM131641 and R01GM124334), the UC Berkeley - UCSF Sackler Faculty Exchange program, and the UCSF Byers Award in Basic Science. S.Z was partly supported by a fellowship from Chinese Scholar Council. L.W and B.H. are Chan Zuckerberg Biohub Investigators.

References

1. B.-C. Chen, W. R. Legant, K. Wang, L. Shao, D. E. Milkie, M. W. Davidson, C. Janetopoulos, X. S. Wu, J. A. Hammer, Z. Liu, B. P. English, Y. Mimori-Kiyosue, D. P. Romero, A. T. Ritter, J. Lippincott-Schwartz, L. Fritz-Laylin, R. D. Mullins, D. M. Mitchell, J. N. Bembenek, A.-C. Reymann, R. Böhme, S. W. Grill, J. T. Wang, G. Seydoux, U. S. Tulu, D. P. Kiehart, and E. Betzig, "Lattice light-sheet microscopy: Imaging molecules to embryos at high spatiotemporal resolution," Science **346** (6208), 1257998 (2014).
2. B. Yang, X. Chen, Y. Wang, S. Feng, V. Pessino, N. Stuurman, N. H. Cho, K. W. Cheng, S. J. Lord, L. Xu, D. Xie, R. D. Mullins, M. D. Leonetti, and B. Huang, "Epi-illumination SPIM for volumetric imaging with high spatial-temporal resolution," Nature Methods **16** (6), 501-504 (2019).
3. K. C. Gwosch, J. K. Pape, F. Balzarotti, P. Hoess, J. Ellenberg, J. Ries, and S. W. Hell, "MINFLUX nanoscopy delivers 3D multicolor nanometer resolution in cells," Nature Methods **17** (2), 217-224 (2020).
4. J. Icha, M. Weber, J. C. Waters, and C. Norden, "Phototoxicity in live fluorescence microscopy, and how to avoid it," BioEssays **39** (8), 1700003 (2017).
5. P. P. Laissue, R. A. Alghamdi, P. Tomancak, E. G. Reynaud, and H. Shroff, "Assessing phototoxicity in live fluorescence imaging," Nature Methods **14** (7), 657-661 (2017).
6. C. Belthangady and L. A. Royer, "Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction," Nature Methods **16** (12), 1215-1225 (2019).
7. P. M. Carlton, J. Boulanger, C. Kervrann, J.-B. Sibarita, J. Salamero, S. Gordon-Messer, D. Bressan, J. E. Haber, S. Haase, L. Shao, L. Winoto, A. Matsuda, P. Kner, S. Uzawa, M. Gustafsson, Z. Kam, D. A. Agard, and J. W. Sedat, "Fast live simultaneous multiwavelength four-dimensional optical microscopy," Proceedings of the National Academy of Sciences **107** (37), 16016-16022 (2010).
8. M. Weigert, U. Schmidt, T. Boothe, A. Müller, A. Dibrov, A. Jain, B. Wilhelm, D. Schmidt, C. Broaddus, S. Culley, M. Rocha-Martins, F. Segovia-Miranda, C. Norden, R. Henriques, M. Zerial, M. Solimena, J. Rink, P. Tomancak, L. Royer, F. Jug, and E. W. Myers, "Content-aware image restoration: pushing the limits of fluorescence microscopy," Nature Methods **15** (12), 1090-1097 (2018).

9. J. Batson and L. Royer, "Noise2Self: blind denoising by self-supervision," Preprint at https://arxiv.org/abs/1901.11365 (2019).
10. A. Krull, T. O. Buchholz, and F. Jug, "Noise2Void—learning denoising from single noisy images," Preprint at https://arxiv.org/abs/1811.10980 (2018).
11. Z. Yide, Z. Yinhao, N. Evan, W. Qingfei, Z. Siyuan, S. Cody, and H. Scott, "A Poisson-Gaussian Denoising Dataset with Real Fluorescence Microscopy Images," Preprint at https://arxiv.org/abs/1812.10366 (2018).
12. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 (Springer International Publishing, 2015), 234-241.
13. W. Zhou, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing **13** (4), 600-612 (2004).
14. E. M. Christiansen, S. J. Yang, D. M. Ando, A. Javaherian, G. Skibinski, S. Lipnick, E. Mount, A. O'Neil, K. Shah, A. K. Lee, P. Goyal, W. Fedus, R. Poplin, A. Esteva, M. Berndl, L. L. Rubin, P. Nelson, and S. Finkbeiner, "In Silico Labeling: Predicting Fluorescent Labels in Unlabeled Images," Cell **173** (3), 792-803.e719 (2018).
15. E. Nehme, L. E. Weiss, T. Michaeli, and Y. Shechtman, "Deep-STORM: super-resolution single-molecule microscopy by deep learning," Optica **5** (4), 458-464 (2018).
16. H. Wang, Y. Rivenson, Y. Jin, Z. Wei, R. Gao, H. Günaydın, L. A. Bentolila, C. Kural, and A. Ozcan, "Deep learning enables cross-modality super-resolution in fluorescence microscopy," Nature Methods **16** (1), 103-110 (2019).
17. Y. Wu, Y. Rivenson, H. Wang, Y. Luo, E. Ben-David, L. A. Bentolila, C. Pritz, and A. Ozcan, "Three-dimensional virtual refocusing of fluorescence microscopy images using deep learning," Nature Methods **16** (12), 1323-1331 (2019).
18. J. Wang, D. Agarwal, M. Huang, G. Hu, Z. Zhou, C. Ye, and N. R. Zhang, "Data denoising with transfer learning in single-cell transcriptomics," Nature Methods **16** (9), 875-878 (2019).
19. G. Barbastathis, A. Ozcan, and G. Situ, "On the use of deep learning for computational imaging," Optica **6** (8), 921-943 (2019).