Physics-Based Learned Diffuser for Single-shot 3D Imaging

Eric Markley[1*], **Fanglin Linda Liu**[2*], **Michael Kellman**[3], **Nick Antipa**[4], **Laura Waller**[1,2]
[1] UCB/UCSF Joint Graduate Program in Bioengineering, University of California, Berkeley
[2] Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
[3] Department of Pharmaceutical Chemistry, University of California, San Francisco
[4] Department of Electrical and Computer Engineering, University of California, San Diego
[*] Equal contribution
{emarkley, fanglin_liu, kellman, nick.antipa, waller}@berkeley.edu

Abstract

A diffuser in the Fourier space of an imaging system can encode 3D fluorescence intensity information in a single-shot 2D measurement, which is then recovered by a compressed sensing algorithm. Typically, the diffusers used in such systems are either off-the-shelf, heuristically designed, or merit function driven. In this work we use a differentiable forward model of single-shot 3D microscopy in conjunction with an invertible and differentiable reconstruction algorithm, ISTA-Net$^+$, to jointly optimize both the diffuser surface shape and the reconstruction parameters. By choosing a differentiable and invertible reconstruction method, we enable the use of memory-efficient backpropagation to trade off storage with a reasonable increase in compute time, in order to fit an unrolled network containing a large-scale 3D volume into a single GPU's memory. We validate our method on 2D and 3D single-shot imaging, where our learned diffuser demonstrates improved reconstruction quality compared to previous heuristic designs.

1 Introduction

Single-shot 3D imaging aims to use optical elements to encode 3D information into a 2D measurement, then recovers the 3D information computationally. Previously, in our Fourier DiffuserScope [1], we used random multi-focal lenslets (RML) as a diffuser in the Fourier space for single-shot 3D imaging. Here, we describe a physics-based learning method for designing diffusers that optimize performance across both hardware and software.

Our physics-based learning pipeline (Fig. 1) aims to jointly optimize the diffuser design (surface height of a refractive phase plate) and the reconstruction algorithm parameters. The optical simulator takes as input the current diffuser surface parameters, $\mathbf{\Theta}$, and outputs the microscope point spread function (PSF) corresponding to each depth plane. A noise-free 2D measurement is generated using each plane's PSF in a convolutional forward model. Noise is then added and the noisy, simulated measurement is fed into an ISTA-Net$^+$ [2, 3] reconstruction network with learnable parameters, $\mathbf{\Phi}$. A loss function is applied to the reconstructed volume, and optical system and reconstruction parameters ($\mathbf{\Theta}$ and $\mathbf{\Phi}$, respectively) are jointly updated using gradient-based updates in a memory-efficient manner.

2 Forward Model

Our RML diffuser (see Fig. 1) consists of a number of plano-convex lenslets of varying focal lengths. The learnable parameters are each lenslet's lateral position coordinates and radius, $\mathbf{\Theta}_i = \{x_i, y_i, r_i\}$. Compared to a parameterization using a pixel-wise surface height [4, 5], our diffuser model dramatically reduces the number of learnable parameters, thus preventing overfitting and decreasing the amount of data required for training. Compared to a Gaussian diffuser [6], contour-shape diffuser [7] or a Zernike polynomial based phase mask [8], our lenslets-based surface focuses light to sharp points, providing higher signal-to-noise ratio (SNR) and frequency coverage over a wide depth range [1, 9, 10].

Next, we describe our imaging model for encoding 3D information into a single 2D frame. Based on the Fourier DiffuserScope setup [1], we use a differentiable forward model followed by a differentiable noise model to generate noisy, simulated measurements. The configuration of the optical system is shown in the inset in Fig. 1. The system begins with an objective lens and a tube lens, as in a traditional fluorescence microscope. Because the Fourier plane of the objective resides inside the objective tube, we use a relay lens to form a $4f$ system together with the tube lens and place the diffuser at the relayed Fourier plane. The sensor is behind the diffuser at a distance equal to the average focal length of the lenslets. We model our optical system using wave-optics propagation, as described below.

The 3D volume is divided into a stack of 2D slices of finite thickness in depth. Assuming our system is laterally shift-invariant, we only need one PSF from each depth layer to fully characterize the system response. From an on-axis point source at each depth, we calculate the spherical wavefront at the objective back focal plane, then multiply it by the apodization pupil function. The wavefront at the pupil is demagnified by the relay system and passes through the diffuser. The diffuser is modeled as a pure phase mask with phase delay $\boldsymbol{\phi} = \exp\left[i\frac{2\pi}{\lambda}(n_r - 1)\mathbf{T}\right]$, with refractive index, n_r, and surface thickness, $\mathbf{T}(\boldsymbol{\Theta})$, embedding $\boldsymbol{\Theta}$ in the forward model. The electric field after the diffuser is then digitally propagated to the sensor via angular spectrum method [11]. The intensity images at the sensor from all the depth layers form a PSF stack, $\mathbf{h}_{z=1,\ldots,Z}$, where $\mathbf{h}_z$ represents the simulated PSF at depth z and there are in total $Z = 11$ depth layers. The simulated measurement is modeled as the sum of all the lateral 2D convolutions (denoted by $\overset{[x,y]}{*}$) of object slices and PSFs, one for each depth:

$$\mathbf{y} \quad = \textstyle\sum_z \mathbf{h}_z \overset{[x,y]}{*} \mathbf{v}_z = \mathbf{A}\mathbf{v}, \tag{1}$$

where, $\mathbf{y}$ is the noise-free intensity image, $\mathbf{v}_z$ is the object intensity at depth z. $\mathbf{v}$ represents the entire 3D volume and $\mathbf{A}$ is a matrix with columns containing the PSF stack, used to write our forward model in compact matrix form.

To model noise, we approximate the expected light levels at 30k photons per fluorescent bead [4], with a Poisson distribution and negligible read noise. Since the sampling of the Poisson distribution is not differentiable with respect to its input, we use the Gaussian approximation of shot noise based on the Central Limit Theorem, giving a noise model that is differentiable with respect to the diffuser parameters. This noisy, simulated measurement, $\mathbf{y}_{\text{noisy}}$, is used in our reconstruction algorithm.

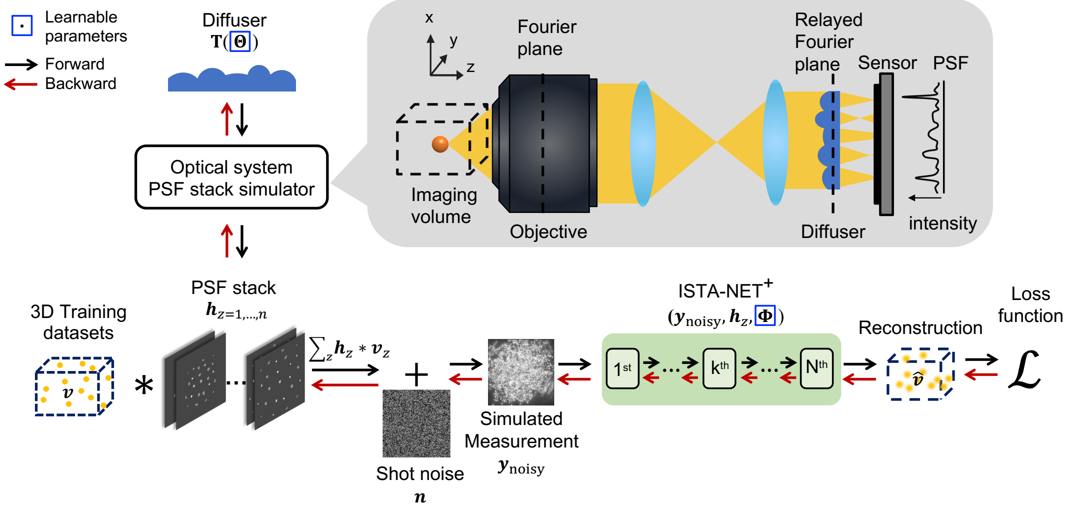

Figure 1: Overview of our physics-based learning pipeline. First, we simulate the point spread function (PSF) stack of the optical system with learnable diffuser parameters, $\boldsymbol{\Theta}$. The PSF stack is convolved with a 3D training volume and noise is added producing a noisy, simulated measurement. The ISTA-Net$^+$ reconstruction network with learnable parameters, $\boldsymbol{\Phi}$, takes in the PSF stack and measurement and outputs a reconstructed volume which is fed into a loss function. The loss is backpropagated through the pipeline to update both the diffuser's and the reconstruction network's sets of learnable parameters, $\boldsymbol{\Theta}$ and $\boldsymbol{\Phi}$. The inset depicts our diffuser-based single-shot 3D microscopy setup.

3 Physics-based Reconstruction

Our reconstruction algorithm aims to solve the following sparsity-constrained inverse problem:

$$\hat{\mathbf{v}} = \min_{\mathbf{v}} \|\mathbf{y} - \mathbf{A}\mathbf{v}\|_2^2 + \lambda \|G(\mathbf{v})\|_1, \tag{2}$$

where λ is a regularization parameter and $G(\cdot)$ is a transform that sparsifies the 3D volume. Traditional iterative optimization algorithms - for example, Fast Iterative Shrinkage-Threshold Algorithm (FISTA) [2] can be used to solve this problem, but suffer from slow computation time (due to the large number of iterations required) and require extensive hand-tuning of tuning variables and proximal operators. Deep network based reconstruction methods are significantly faster than iterative optimization, but lack an understanding of the physical system [3, 12]. Consequently, more training data is needed in order to achieve sufficient generalization at test time.

Here we apply a physics-based ISTA-Net^{+} [3] of 10 unrolls, each consisting of a gradient step followed by a proximal step, as shown in Fig. A.1 (a). The learnable parameter set, $\mathbf{\Phi}$, includes the regularization parameter λ, the sparsifying transform $G(\cdot)$ and its left inverse $\tilde{G}(\cdot)$, as in $\mathbf{\Phi} = \{\lambda, G, \tilde{G}\}$. The loss function driving the training of the pipeline is:

$$\mathcal{L} = \frac{1}{M}\sum_{m=1}^{M} \|\hat{\mathbf{v}}_m^{(N)} - \mathbf{v}_m\|_2^2 + \gamma\|\tilde{G}(G(\mathbf{v}_m)) - \mathbf{v}_m\|_2^2, \tag{3}$$

where M is the number of training examples per batch, γ is a tuning parameter, subscript *m* denotes the *m*-th training example, superscript *N* denotes the total number of unrolls, the first term penalizes the ℓ_2 reconstruction error between the ground truth and reconstruction and the second term encourages $\tilde{G}(\cdot)$ to be the left inverse of $G(\cdot)$.

By including the forward model in the reconstruction algorithm, physics-based approaches help minimize training data requirements and prevent overfitting. Additionally, due to the fact the forward model is a function of the diffuser's parameters, $\mathbf{\Theta}$, derivatives of the loss can be taken with respect to $\mathbf{\Theta}$. We treat both update steps of this algorithm as forward Euler steps, allowing for inversion through backward Euler steps with a fixed point method. This enables the use of memory-efficient learning techniques (see Sec. 3.1) [13–15].

3.1 Memory-efficient Backpropagation

Backpropagation is used to calculate the gradient of a single output with respect to multiple learnable variables of a system, with relatively low time complexity, by using a large amount of memory to store the full computational graph during the forward pass. Due to the 3D nature of our problem, GPU memory is at a premium. Therefore, we use memory-efficient backpropagation techniques to fit our problem in memory with a reasonable increase of compute time, allowing for calculating the necessary gradients from a series of unrolls while only having to fit a single unroll in memory. The primary memory-efficient backpropagation technique used is forward checkpointing based [16] but the model is also compatible with reverse checkpointing [14] as shown in Fig. A.1 (b).

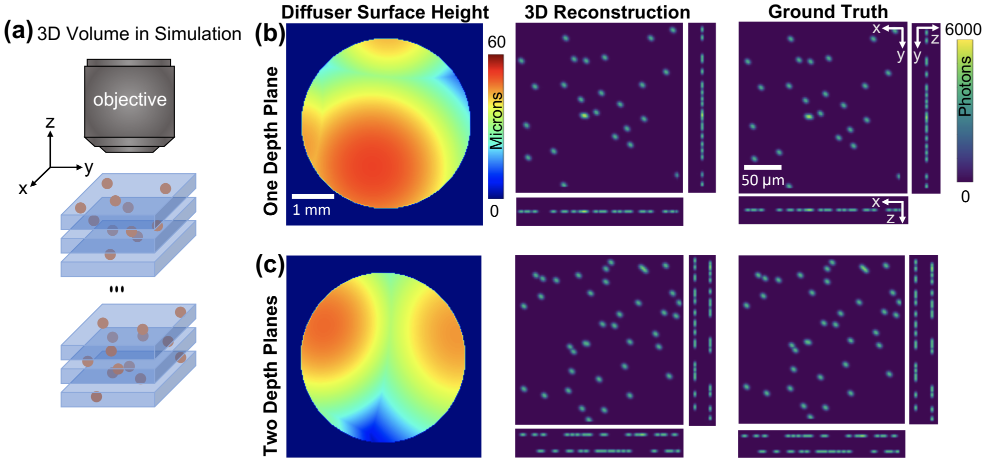

Figure 2: **(a)** Simulated 3D training volume with randomly placed fluorescent beads. **(b)** When the beads are constrained to a single depth plane, the learned random multi-focal lenslets (RML) consists of one dominating lens, as expected. **(c)** When the beads are constrained to two depth planes, the learned RML contains two lenses, each focusing at one depth plane. Maximum intensity projections show the reconstructed test volumes match well with ground truth.

4 Results

The optical system parameters are in Appendix B. Training was performed on 200 simulated volumes (see Appendix C). Testing was performed on 40 volumes generated in the same fashion as the training set. The model and training pipeline are written in Pytorch and training was performed on a NVIDIA RTX 3090 GPU.

We first perform two 'sanity checks' to show that the pipeline functions as expected, by optimizing the RML for two imaging scenarios where a reasonable guess of the ideal RML is known. We use a small volume at 1/15 the size of our experimental system to speed up computation time. In the first scenario, all beads are constrained to the center depth plane of the volume. This simplifies the problem to 2D imaging; hence, a single lenslet focused at that depth should be ideal. As seen in Fig. 2 (b), the optimized RML does contain a single, dominant lenslet, though there are additional features, likely due to the non-convexity of the problem. In the second scenario, the beads are constrained to 2 depth planes at opposite ends of the volume. In this scenario, a RML with two lenslets, each focusing at one of the depth planes, is expected. As seen in Fig. 2 (c), the result does show two lenslets that dominate the RML. In both scenarios, the learned RML matches our intuition, and the reconstruction of the testing volume matches well with the ground truth. We conclude that the pipeline is functioning sufficiently and move to more complex imaging scenarios.

Next, we compare our learned RML and algorithm to a heuristically-designed RML with unlearned algorithms in Fig. 3. In our training data, we use beads of $1\,\mu\text{m}$ and $2\,\mu\text{m}$ diameter, randomly spaced in a 3D volume of $500 \times 500 \times 50\,\mu\text{m}^3$ with $5\,\mu\text{m}$ axial steps. To achieve $1\,\mu\text{m}$ lateral resolution over a $50\,\mu\text{m}$ depth range, from first principles derivation [1], we need 3.2 lenslets in each direction, the square of which rounds up to 11 lenslets. Hence, the heuristic designed diffuser contains 11 randomly-located lenslets of varying focal length, focusing at uniformly-spaced depth planes. For the learned design, since we cannot take derivatives of the loss function with respect to the number of lenslets, we instead optimize the number of lenslets by starting with more than necessary and allowing them to merge and exit the pupil during the learning process. In experiments, the number of lenslets in the learned RML is consistent for each imaging scenario regardless of the initialized number, demonstrating robustness of the pipeline. The learned surface in Fig. 3 contains 8 lenslets; the lower number of lenslets increases the numerical aperture of each lenslet, collecting more photons per focus point and boosting SNR under our Gaussian-approximated shot noise model. For the case of the unlearned algorithm, we used an L1 proximal step with ISTA reconstructions ran until convergence. This runs $\sim$30x slower than our proposed reconstruction network while achieving worse normalized mean square error (NMSE). Note that to reduce memory requirements and training time, the scale of this simulated system is 1/4 of our experimental system.

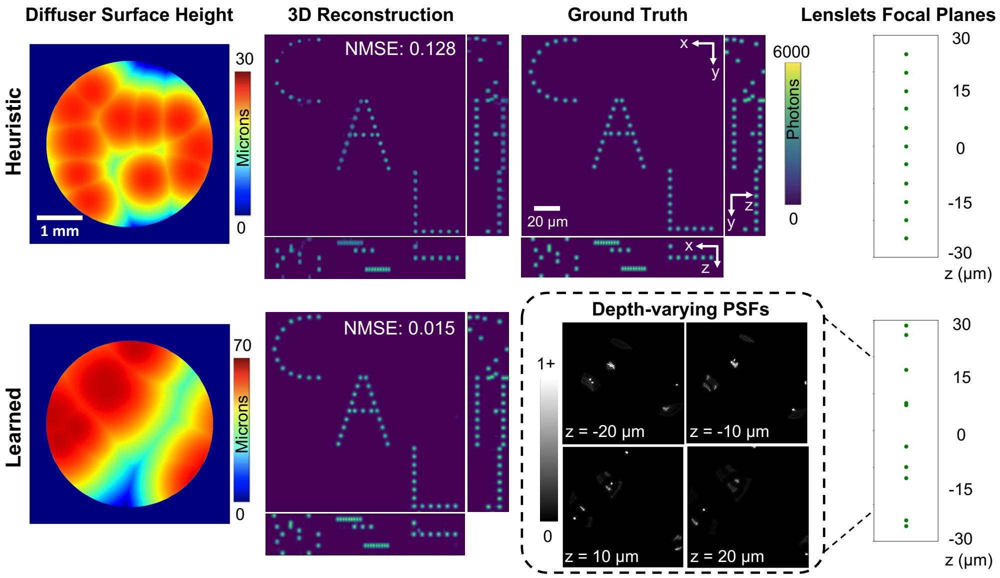

Figure 3: Comparing results of our learned random microlenslet (RML) design to that of an unlearned heuristic design. The first row labeled **Heuristic** depicts (from left to right) the heuristic RML surface height map containing 11 lenslets, the reconstructed volume using heuristic RML and ISTA ran to convergence ($\sim$3000 iterations), the ground truth volume, and the focal length distribution of the lenslets in the heuristic RML. The second row labeled **Learned** depicts (from left to right) the learned RML surface height map containing 8 lenslets, the reconstruction volume using the learned RML and learned reconstruction algorithm (see A.1), point spread functions at several depth planes, and the focal length distribution of the lenslets in the learned RML.

5 Conclusion

We demonstrated improved reconstruction speed and improved image reconstruction quality for Fourier DiffuserScope single-shot 3D microscopy by designing an optical element (a diffuser) that jointly optimizes the experimental setup and reconstruction algorithm via end-to-end learning. This data-driven approach directly optimizes the reconstruction loss and provides better insights into design for a non-traditional optical system where the first principles are limited.

References

[1] F. L. Liu, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Fourier DiffuserScope: single-shot 3D Fourier light field microscopy with a diffuser," *Opt. Express*, vol. 28, pp. 28969–28986, sep 2020.

[2] A. Beck and M. Teboulle, "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems," *Society for Industrial and Applied Mathematics Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[3] J. Zhang and B. Ghanem, "ISTA-Net: Interpretable Optimization-Inspired Deep Network for Image Compressive Sensing," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1828–1837, 2018.

[4] E. Nehme, D. Freedman, R. Gordon, B. Ferdman, L. E. Weiss, O. Alalouf, T. Naor, R. Orange, T. Michaeli, and Y. Shechtman, "DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning," *Nature Methods*, vol. 17, no. 7, pp. 734–740, 2020.

[5] C. A. Metzler, H. Ikoma, Y. Peng, and G. Wetzstein, "Deep Optics for Single-shot High-dynamic-range Imaging," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1372–1382, 2020.

[6] N. Antipa, G. Kuo, R. Heckel, B. Mildenhall, E. Bostan, R. Ng, and L. Waller, "DiffuserCam: lensless single-exposure 3D imaging," *Optica*, vol. 5, pp. 1–9, jan 2018.

[7] V. Boominathan, J. Adams, J. Robinson, and A. Veeraraghavan, "PhlatCam: Designed phase-mask based thin lensless camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[8] V. Sitzmann, S. Diamond, Y. Peng, X. Dun, S. Boyd, W. Heidrich, F. Heide, and G. Wetzstein, "End-to-end Optimization of Optics and Image Processing for Achromatic Extended Depth of Field and Super-resolution Imaging," *ACM Trans. Graph. (SIGGRAPH)*, 2018.

[9] G. Kuo, F. L. Liu, I. Grossrubatscher, R. Ng, and L. Waller, "On-chip fluorescence microscopy with a random microlens diffuser," *Optics Express*, vol. 28, pp. 8384–8399, mar 2020.

[10] K. Yanny, N. Antipa, W. Liberti, S. Dehaeck, K. Monakhova, F. L. Liu, K. Shen, R. Ng, and L. Waller, "Miniscope3D: optimized single-shot miniature 3D fluorescence microscopy," *Light: Science Applications*, vol. 9, no. 1, p. 171, 2020.

[11] J. W. Goodman, *Introduction to Fourier optics*. Roberts and Company Publishers, 2005.

[12] K. Monakhova, J. Yurtsever, G. Kuo, N. Antipa, K. Yanny, and L. Waller, "Learned reconstructions for practical mask-based lensless imaging," *Opt. Express*, vol. 27, pp. 28075–28090, sep 2019.

[13] J. Behrmann, W. Grathwohl, R. T. Chen, D. Duvenaud, and J. H. Jacobsen, "Invertible Residual Networks," in *36th International Conference on Machine Learning, ICML 2019*, vol. 2019-June, pp. 894–910, 2019.

[14] M. Kellman, K. Zhang, E. Markley, J. Tamir, E. Bostan, M. Lustig, and L. Waller, "Memory-Efficient Learning for Large-Scale Computational Imaging," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1403–1414, 3 2020.

[15] M. Kellman, E. Bostan, M. Chen, and L. Waller, "Data-Driven Design for Fourier Ptychographic Microscopy," in *2019 IEEE International Conference on Computational Photography, ICCP 2019*, Institute of Electrical and Electronics Engineers Inc., 5 2019.

[16] A. Griewank and A. Walther, "Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation," *ACM Trans. Math. Softw.*, vol. 26, p. 19–45, Mar. 2000.

Appendix

A Physics-based network architecture

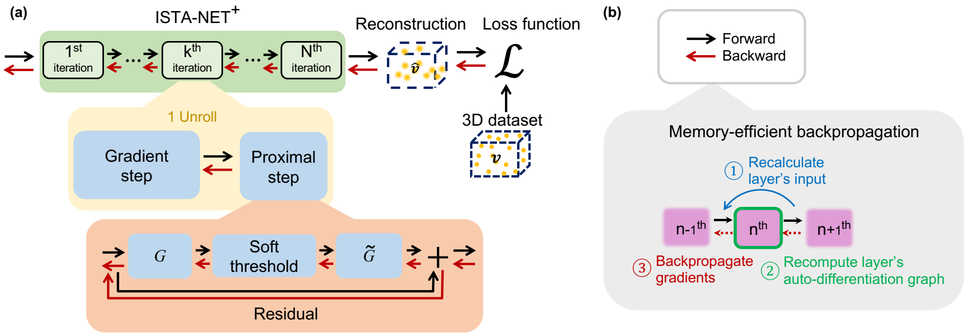

Figure A.1: Physics-based reconstruction algorithm. **(a)** Our reconstruction network uses ISTA-Net$^+$ with N unrolled iterations of ISTA, each including a gradient step followed by a proximal step. The proximal step consists of a learned nonlinear sparsifying transform G, soft thresholding, and a learned left inverse of the sparsifying transform $\tilde{G}$. **(b)** Reverse checkpointing based memory-efficient backpropagation calculates the nth layer's gradients in three steps: 1) recompute the layer's input from output. 2) recompute the layer's auto-differentiation graph. 3) recompute the gradients.

B Optical Setup

The optical system contains a $20\times$ 0.8 NA objective lens, a tube lens with focal length of $180\,\mathrm{mm}$, a relay lens with focal length of $48\,\mathrm{mm}$, and a RML with average focal length of $15.6\,\mathrm{mm}$, giving an overall system magnification of $6.5\times$.

C Training Data Generation

The size of the imaging volume is $500 \times 500 \times 50\,\mu\mathrm{m}^3$ consisting of 11 depth layers in $5\,\mu\mathrm{m}$ steps. The model was trained on simulated volumes consisting of 200 volumes of $750 \times 750 \times 11$ voxels containing beads of 1 and 2 micron diameters. The beads were simulated using Gaussians with full width at half maximum (FWHM) equal to the desired bead size. The distribution of the peak intensity of the beads is uniform from .8 to 1.2. The volumes are then multiplied by the desired photon emission level at training time to achieve the desired SNR. We also allow the density of the fluorescent beads to vary across training volumes.