

---

# Conditional Diffusion Models for Uncertainty Estimation in Super Resolution Microscopy

---

Anonymous Author(s)

Affiliation

Address

email

## Abstract

1 The field of deep generative modeling has experienced a spike in research in recent  
2 years, with various tools developed to model previously intractable distributions  
3 over high dimensional datasets. Yet, many powerful regressive models continue to  
4 be implemented, in spite of the fact that they are unable model distributions over  
5 their outputs. In particular, this mode of deep learning has attracted attention from  
6 researchers in the natural sciences, particularly microscopists, for fast extraction of  
7 physically relevant information from images. As powerful as they may be, simple  
8 and interpretable uncertainty quantification is a necessary modeling component  
9 in high-risk applications and in the sciences. In order to quantify uncertainty in  
10 otherwise deterministic image translation models, we propose a hybrid generative  
11 modeling framework based on denoising diffusion probabilistic models (DDPMs).  
12 Specifically, our model learns a distribution on a true image latent in the input  
13 conditioned on the network output, in order to represent the posterior on recon-  
14 structions. We apply this framework to the task of single molecule localization in  
15 fluorescence microscopy, and demonstrate that blending the DeepSTORM archi-  
16 tecture with a DDPM permits uncertainty quantification of kernel density estimates  
17 (KDEs) regressed by DeepSTORM. Our results suggest the proposed solution is an  
18 interesting addition to the modeling toolkit for fluorescence microscopists and the  
19 field of deep image translation in general.

## 20 1 Introduction

21 Deep learning has attracted tremendous attention from researchers in the natural sciences, with  
22 several foundational applications arising in microscopy, e.g., (Weigert 2018; Falk 2019). Recently,  
23 the application of deep image translation in single-molecule localization microscopy (SMLM) has  
24 received considerable interest (Ouyang 2018; Nehme 2020; Speiser 2021). SMLM techniques  
25 are a mainstay of fluorescence microscopy and can be used to produce a pointillist representation  
26 of biomolecules in the cell at diffraction-unlimited precision (Rust 2006; Betzig 2006). As this  
27 technology enables increasingly precise measurements of the cellular environment, there is an  
28 increasing need for machine learning methods to report uncertainty for quality control.

29 In previous applications of deep models to localization microscopy, super-resolution images can be  
30 recovered from a sparse set of localizations with conditional generative adversarial networks (Ouyang  
31 2018) or kernel density estimation can be performed using convolutional networks (Nehme 2020;  
32 Speiser 2021). Here, we focus on the latter class of models which perform single molecule localization  
33 using neural networks. In this approach, one estimates molecular coordinates by predicting kernel  
34 density estimates (KDEs)  $y$ , which are latent in the raw data  $x$ , using a convolutional neural network.  
35 Importantly, inferences in SMLM are often necessarily made on a single measurement, thus common  
36 measures of model performance are based on localization errors computed over ensembles of



Figure 1: Generative model of single molecule localization microscopy images

37 simulated images. However, this choice precludes computation of aleatoric uncertainty at test time  
 38 under a fixed model, and may result in the application of models to out of distribution datasets.

39 Bayesian probability theory offers us mathematically grounded tools to reason about model uncer-  
 40 tainty, but these usually come with a prohibitive computational cost (Gal 2022). A few approaches  
 41 to avoiding this intractability in deep models have been deterministic uncertainty quantification  
 42 (Amersfoort 2020), ensembling (Lakshminarayanan et al., 2017) or Monte Carlo dropout (Gal and  
 43 Ghahramani, 2016). Here, we report a method which models estimates uncertainty in KDE pre-  
 44 dictions by learning a distribution on the true image latent in the input conditioned on the network  
 45 output, in order to represent the posterior on reconstructions. Our approach preserves image structure  
 46 and produces pixel-wise uncertainties, which can be used for out of distribution sample detection or  
 47 filtering. We choose to model this distribution using a denoising diffusion probabilistic model (DDPM)  
 48 (Ho 2020), referred to here as simply “diffusion model”. Such models are well suited conditional  
 49 image generation tasks, demonstrating promising results in detail reconstruction, while directly  
 50 providing uncertainties in model predictions (Saharia 2021). Our approach could be readily integrated  
 51 with existing localization performance measures to address both model accuracy on training data and  
 52 precision on datasets produced by experiments.

## 53 2 Background

### 54 2.1 Image Likelihood and Localization Error

55 The central objective of single molecule localization microscopy is to infer a set of molecular  
 56 coordinates  $\theta$  from measured low resolution images  $\mathbf{x}$ . The likelihood on a particular pixel  $k$ ,  
 57 i.e.,  $p(\mathbf{x}_k|\theta)$  is taken to be a convolution of Poisson and Gaussian distributions, due to shot noise  
 58  $p(s_k) = \text{Poisson}(\omega_k)$  and sensor readout noise  $p(\zeta_k) = \mathcal{N}(o_k, \sigma_k^2)$

$$p(\mathbf{x}_k|\theta) = A \sum_{q=0}^{\infty} \frac{1}{q!} e^{-\omega_k} \omega_k^q \frac{1}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\mathbf{x}_k - g_k q - o_k)^2}{2\sigma_k^2}} \quad (1)$$

59 where  $A$  is some normalization constant. In practice, (1) is difficult to work with, so we look for an  
 60 approximation. We will use a Poisson-Normal approximation for simplification, valid under a range  
 61 of experimental conditions (Huang 2013)

$$\mathbf{x}_k \sim \text{Poisson}(\omega'_k) \quad (2)$$

62 where  $\omega'_k = \omega_k + \sigma_k^2$ . This result can be seen from the fact the the convolution of two Poisson  
 63 distributions is also Poisson.

64 Reliable estimation of  $\theta$  from  $\mathbf{x}$ , for example by maximum likelihood estimation or deep models,  
 65 requires performance metrics for model selection. We use the Fisher information as an information  
 66 theoretic criteria to assess the model quality, with respect to the root mean squared error (RMSE) of  
 67 our predictions of  $\theta$  (Chao 2016). The Poisson log-likelihood  $\ell(\mathbf{x}|\theta)$  is also convenient for computing  
 68 the Fisher information matrix (Smith 2010) and thus the Cramer-Rao lower bound, which bounds the  
 69 variance of a statistical estimator of  $\theta$ , from below i.e.,  $\text{var}(\hat{\theta}) \geq I^{-1}(\theta)$ . The Fisher information is  
 70 straightforward to compute under the Poisson log-likelihood, which is detailed in the Appendix

$$\mathcal{I}_{ij}(\theta) = \mathbb{E}_{\theta} \left( \frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j} \right) = \sum_k \frac{1}{\omega'_k} \frac{\partial \omega'_k}{\partial \theta_i} \frac{\partial \omega'_k}{\partial \theta_j} \quad (3)$$

## 71 2.2 Kernel density estimation with deep networks

72 Direct optimization of the likelihood in (2) from observations  $\mathbf{x}$  alone is challenging when fluorescent  
 73 emitters are dense within the field of view and fluorescent signals significantly overlap. Convolutional  
 74 neural networks (CNN) have recently been used in fluorescence microscopy to extract parameters  
 75 describing fluorescent emitters such as color, emitter orientation,  $z$ -coordinate, and background signal  
 76 (Zhang 2018; Kim 2019; Zelger 2018). For localization tasks, CNNs typically employ upsampling  
 77 layers to reconstruct Bernoulli probabilities of emitter occupancy (Speiser 2021) or kernel density  
 78 estimates with higher resolution than experimental measurements (Nehme 2020). Kernel density  
 79 estimates are the most common data structure used in SMLM, and can be easlily generated from  
 80 molecular coordinates using well-understood models of the optical impulse response (Zhang 2007).  
 81 In addition, models of the optical impulse response can be combined with Poisson likelihood (2) to  
 82 generate ground-truth data for model training, which is often not available in experimental contexts.

83 The DeepSTORM CNN, initially proposed in (Nehme 2020) for 3D localization, can be viewed  
 84 as a deep kernel density estimator, reconstructing kernel density estimates  $\mathbf{y}$  from low-resolution  
 85 inputs  $\mathbf{x}$ . We utilize a simplified form of the original architecture for 2D localization, which we  
 86 denote  $\phi$  hereafter, which consists of three main modules: a multi-scale context aggregation module,  
 87 an upsampling module, and a prediction module. For context aggregation, the architecture utilizes  
 88 dilated convolutions to increase the receptive field of each layer. The upsampling module is then  
 89 composed of two consecutive 2x resize-convolutions, computed by nearest-neighbor interpolation,  
 90 to increase the lateral resolution by a factor of 4. For a common sCMOS camera, each pixel has a  
 91 lateral size of approximately 108 nanometers, giving approximately 27 nanometer pixels in the KDE.  
 92 The terminal prediction module contains three additional convolutional blocks for refinement of the  
 93 upsampled image, followed by an element-wise HardTanh.

## 94 3 Diffusion Model for SMLM

95 We consider datasets  $(\mathbf{x}_i, \mathbf{y}_i, \hat{\mathbf{y}}_i)_{i=1}^N$  of observed images  $\mathbf{x}_i$  true kernel density estimate (KDE) images  
 96  $\mathbf{y}_i$ , and KDE estimates  $\hat{\mathbf{y}}_i = \phi(\mathbf{x}_i)$ . Observations  $\mathbf{x}_i$  are simulated under the likelihood (2).

## 97 4 Conditional Diffusion Model

98 Point estimates  $\hat{\mathbf{y}}_i$  produced by the DeepSTORM architecture lack uncertainty quantification. Con-  
 99 sider this “triangular” Markov chain  $\mathbf{y} \rightarrow \mathbf{x} \rightarrow \hat{\mathbf{y}} \rightarrow \mathbf{y}$  for which we find  $p_{\phi}(\hat{\mathbf{y}}|\mathbf{x})p(\mathbf{y}) =$   
 100  $p_{\Psi}(\mathbf{y}|\hat{\mathbf{y}})p(\hat{\mathbf{y}})$ . Therefore the entropy is

$$H_{\Psi}(\mathbf{y}|\hat{\mathbf{y}}) = H_{\Phi}(\hat{\mathbf{y}}|\mathbf{x}) + \Delta H$$

101 where  $\Delta H = H(\mathbf{y}) - H(\hat{\mathbf{y}})$ . Evidently the entropy of the model prediction  $H_{\Phi}(\hat{\mathbf{y}}|\mathbf{x})$ , can be modeled  
 102 by the entropy of our Gaussian diffusion model  $H_{\Psi}(\mathbf{y}|\hat{\mathbf{y}})$ . After training  $\Psi$ , one can measure  $\Delta H$   
 103 by injecting the correct solution into the diffusion model modeling  $\Psi$  and measure  $\Delta H$ , and thus  
 104 compute  $H_{\Phi}(\hat{\mathbf{y}}|\mathbf{x})$ .

105 Evidently, a DDPM  $\Psi$  can be trained on pairs  $(\mathbf{y}_i, \hat{\mathbf{y}}_i)_{i=1}^N$ . The conditional DDPM generates a target  
 106 KDE  $\mathbf{y}_0$  in  $T$  refinement steps. Starting with a pure noise image  $\mathbf{y}_T \sim \mathcal{N}(0, \mathbf{I})$ , the model iteratively  
 107 refines the KDE through successive iterations according to learned conditional transition distributions  
 108  $p(\mathbf{y}_{t-1}|\mathbf{y}_t, \cdot)$  such that  $\mathbf{y}_0 \sim p(\mathbf{y}|\hat{\mathbf{y}})$

### 109 4.1 Gaussian Diffusion

110 Diffusion models (Sohl-Dickstein 2015; Ho 2020) are a class of generative models inspired by  
 111 nonequilibrium statistical physics, which slowly destroy structure in a data distribution  $p(\mathbf{y}_0|\mathbf{x})$  via a

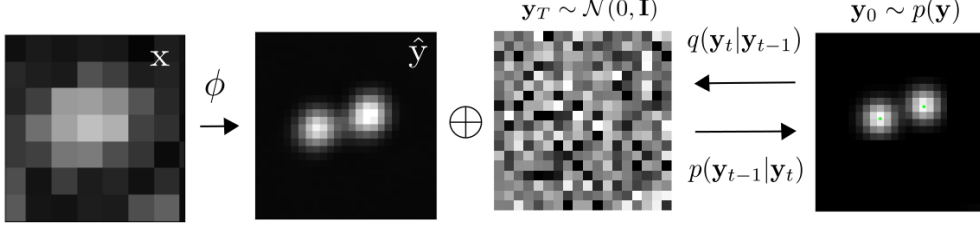


Figure 2: Conditional diffusion model for sampling kernel density estimates

fixed Markov chain referred to as the *forward process*. In the present context, the forward process gradually adds Gaussian noise to the KDE  $y$  according to a variance schedule  $\beta_{0:T}$

$$q(y_t|y_0) = \prod_{t=1}^T q(y_t|y_{t-1}) \quad q(y_t|y_{t-1}) = \mathcal{N}\left(\sqrt{1-\beta_t}y_{t-1}, \beta_t I\right) \quad (4)$$

An important property of the forward process is that it admits sampling  $y_t$  at an arbitrary timestep  $t$  in closed form (Ho 2020). Using the notation  $\alpha_t := 1 - \beta_t$  and  $\gamma_t := \prod_{s=1}^t \alpha_s$ , we have

$$q(y_t|y_0) = \mathcal{N}(\sqrt{\gamma_t}y_0, (1 - \gamma_t)I) \quad (5)$$

The usual procedure is then to learn a parametric representation of the *reverse process*, and therefore generate samples from  $p(y_0)$ , starting from noise. Formally,  $p_\theta(y_0|\hat{y}) = \int p_\theta(y_{0:T}|\hat{y})d\hat{y}_{1:T}$  where  $y_t$  is a latent representation with the same dimensionality of the data.  $p_\theta(y_{0:T}|\hat{y})$  is a Markov process, starting from a noise sample  $p_\theta(y_T) = \mathcal{N}(0, I)$ .

$$p_\theta(y_{0:T}) = p_\theta(y_T) \prod_{t=1}^T p_\theta(y_{t-1}|y_t) \quad p_\theta(y_{t-1}|y_t) = \mathcal{N}(\mu_\theta(y_t), \beta_t I) \quad (6)$$

where we reuse the variance schedule of the forward process (Ho 2020). We seek to learn a denoising model  $\mu_\theta$  which computes the mean of the Gaussian transition density at each time step  $t$ . For all  $t > 0$ , the mean of the transition density is computed as

$$\mu_\theta(y_t, \hat{y}, \gamma_t) = \frac{1}{\sqrt{\alpha_t}} \left( y_t - \frac{(1 - \alpha_t)}{\sqrt{1 - \gamma_t}} f_\theta(y, \hat{y}, \gamma_t) \right) \quad (7)$$

where  $f_\theta$  is a neural network. Only at  $t = 0$  is this mean directly a function of  $x$ .

## 4.2 Optimization of the Denoising Model

To reverse the diffusion process, we optimize a neural denoising model  $f_\theta$  that takes as input  $\hat{y}$  and a noisy target image  $y_t \sim q(y_t|y_0)$ . That is, this noisy target image  $y_t$  is drawn from the marginal distribution of noisy images at a time step  $t$  of the forward diffusion process.

$$y_t = \sqrt{\gamma_t}y_0 + \sqrt{1 - \gamma_t}\epsilon, \quad \epsilon \sim \mathcal{N}(0, I) \quad (8)$$

In addition to a source image  $y_0$  and a noisy target image  $y_t$ , the denoising model  $f_\theta$  takes as input the sufficient statistics for the variance of the noise  $\gamma$ , and is trained to predict the noise vector  $\epsilon$ . We make the denoising model aware of the level of noise through conditioning on a scalar  $\gamma$ . The proposed objective function for training  $f_\theta$  is

$$\mathbb{E}_{(\hat{y}, y_0)} \mathbb{E}_{(\epsilon, \gamma)} \left[ f_\theta \left( x, \sqrt{\gamma}y_0 + \sqrt{1 - \gamma}\epsilon \mid y_t, \gamma \right) - \epsilon \right], \quad (9)$$

where  $(\hat{\mathbf{y}}, \mathbf{y}_0)$  is sampled from the training dataset and  $\gamma \sim p(\gamma)$ . The distribution of  $\gamma$  has a big impact on the quality of the model and the generated outputs. For our training noise schedule, we use a piecewise distribution for  $\gamma$ ,  $p(\gamma) = \frac{1}{T} \sum_{t=1}^T U(\gamma_{t-1}, \gamma_t)$  (Nanxin 2021). Specifically, during training, we first uniformly sample a time step  $t \sim \{0, \dots, T\}$  followed by sampling  $\gamma \sim U(\gamma_{t-1}, \gamma_t)$ . We set  $T = 100$  in all our experiments.

### 4.3 Optimization of the DeepSTORM architecture

A first pass at localization treats localization as a binary classification problem, such that 0 denotes a vacant pixel and 1 denotes an occupied pixel containing an emitter. Direct learning of pixel-wise classification with cross-entropy loss leads to an imbalance of occupied and unoccupied pixels in dense localization problems (Nehme 2020). CE loss is usually either weighted [51], replaced with a Focal loss [52], or applied to a "blobbed" version of the desired boolean volume e.g. by placing a disk around each GT position [53–55]. Alternative methods take a soft version of the binary classification problem. That is, by placing a small Gaussian around each GT position (e.g. with std of 1 pixel), and matching continuous heatmaps, backpropagation yields more meaningful gradients and eases the learning process convergence.

Localization heatmaps thus form a natural encoding for SMLM images, which can be input to our conditional diffusion model. Therefore, to encode raw data  $\mathbf{x}$  into a more tractable representation, we train the DeepSTORM architecture (Nehme 2020). Raw coordinates  $\theta$  are binned into an upsampled image  $\mathbf{z}$ .

$$\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}}) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$$

## 5 Experiments

All training data was simulated under the likelihood and impulse response (2,10), drawing coordinates uniformly over a disc with a radius of 7 pixels.

### 5.1 Model Precision on Simulated Ensembles

### 5.2 Model Uncertainty

We set  $T = 100$  for all experiments and treat forward process variances  $\beta_t$  as hyperparameters, with a linear schedule from  $\beta_0 = 10^{-4}$  to  $\beta_T = 10^{-2}$ . These constants were chosen to be small relative to data scaled to  $[-1, 1]$ , ensuring that reverse and forward processes have approximately the same functional form while keeping the signal-to-noise ratio at  $x_T$  as small as possible ( $L_T = D_{KL}(q(x_T|x_0) \parallel \mathcal{N}(0, I)) \approx 10^{-5}$  bits per dimension in our experiments).

To represent the reverse process, we used the DDPM architecture based on a U-Net backbone (Ho 2020). Parameters are shared across time, which is specified to the network using the Transformer sinusoidal position embedding ?. We use self-attention at the  $16 \times 16$  feature map resolution ?. Details are in Appendix A.

and the channel multipliers at different resolutions (see Appendix A for details). To condition the model on the input  $x$ , we up-sample the low-resolution image to the target resolution using bicubic interpolation. The result is concatenated with  $y_t$  along the channel dimension. We experimented with more sophisticated methods of conditioning, such as using, but we found that the simple concatenation yielded similar generation quality.

## 6 Related Work

### 6.1 Diffusion Models

Prior work of diffusion models ?? require 1-2k diffusion steps during inference, making generation slow for large target resolution tasks. We adapt techniques from ? to enable more efficient inference. Our model conditions on  $\gamma$  directly (vs  $t$  as in ?), which allows us flexibility in choosing the number of diffusion steps, and the noise schedule during inference. This has been demonstrated to work

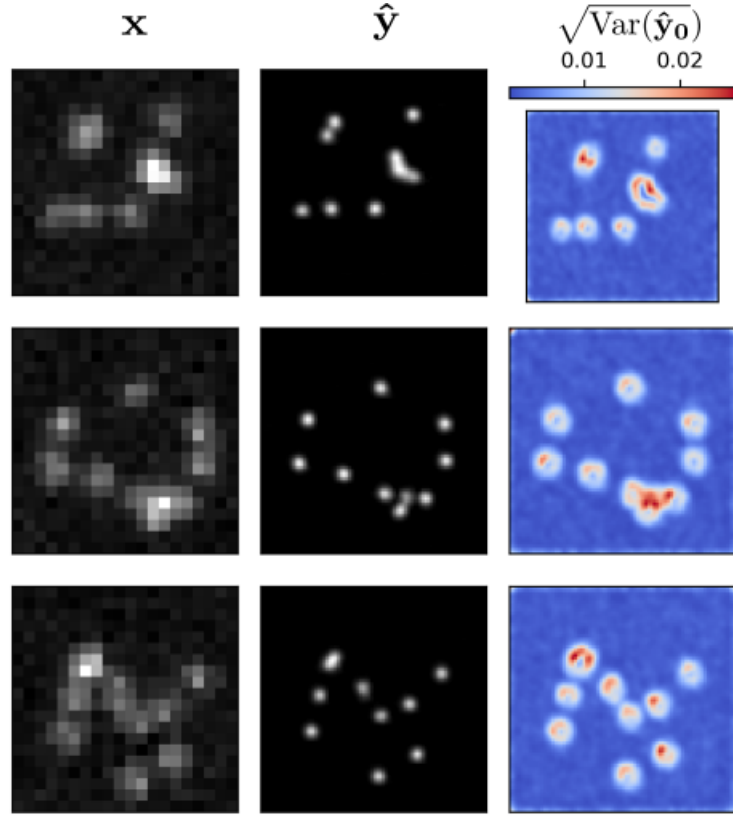


Figure 3: Kernel density estimates for various signal to noise ratios (SNR)

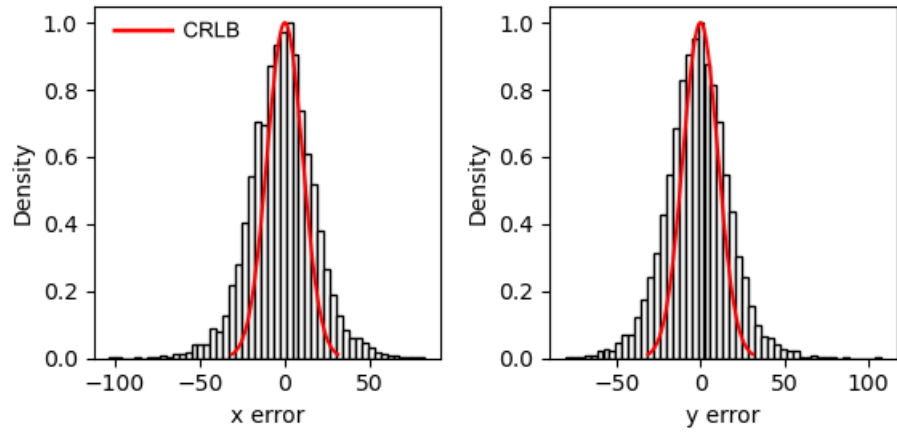


Figure 4: Localization errors of the trained model

well for speech synthesis ?, but has not been explored for images. For efficient inference, we set the maximum inference budget to 100 diffusion steps, and hyper-parameter search over the inference noise schedule. This search is inexpensive as we only need to train the model once ?. We use FID on held-out data to choose the best noise schedule, as we found PSNR did not correlate well with image quality.

## 7 Conclusion

## References

- [1] Nehme, E., et al. *DeepSTORM3D: dense 3D localization microscopy and PSF design by deep learning*. Nature Methods 17, 734–740 (2020).
- [2] Ouyang, W., et al. *Deep learning massively accelerates super-resolution localization microscopy*. Nature Biotechnology 36, 460–468 (2018).
- [3] Speiser, A., et al. *Deep learning enables fast and dense single-molecule localization with high accuracy*. Nature Methods 18, 1082–1090 (2021).
- [4] Sohl-Dickstein J., et al. *Deep unsupervised learning using nonequilibrium thermodynamics*. ICLR (2015).
- [5] Ho J., et al. *Denoising Diffusion Probabilistic Models*. Advances in Neural Information Processing Systems (2015).
- [6] Nanxin C., et al. *WaveGrad: Estimating Gradients for Waveform Generation*. ICLR (2021).
- [4] Chao, J., et al. *Fisher information theory for parameter estimation in single molecule microscopy: tutorial*. Journal of the Optical Society of America A 33, B36 (2016).
- [5] Schermelleh, L. et al. *Super-resolution microscopy demystified*. Nature Cell Biology vol. 21 72–84 (2019).
- [6] Zhang, B., et al. *Gaussian approximations of fluorescence microscope point-spread function models*. (2007).
- [7] Smith, C.S., *Fast, single-molecule localization that achieves theoretically minimum uncertainty*. Nature Methods 7, 373–375 (2010).
- [8] Nieuwenhuizen, R., et al. *Measuring image resolution in optical nanoscopy*. Nature Methods 10, 557–562 (2013).
- [9] Huang, F., et al. *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms*. Nat Methods 10, 653–658 (2013).
- [10] Rust, M., et al. *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM)*. Nat Methods 3, 793–796 (2006).
- [11] Betzig, E., et al. *Imaging intracellular fluorescent proteins at nanometer resolution*. Science 313, 1642–1645 (2006).
- [12] Weigert, M., et al. *Content-aware image restoration: pushing the limits of fluorescence microscopy*. Nat. Methods 15, 1090 (2018).
- [13] Falk, T., et al. *U-net: deep learning for cell counting, detection, and morphometry*. Nat. Methods 16, 67–70 (2019).
- [14] Boyd, N., et al. *DeepLoco: fast 3D localization microscopy using neural networks*. Preprint at bioRxiv <https://doi.org/10.1101/267096> (2018)
- [15] Zelger, P., et al. *Three-dimensional localization microscopy using deep learning*. Opt. Express 26, 33166–33179 (2018)
- [16] Zhang, P., et al. *Analyzing complex single-molecule emission patterns with deep learning*. Nat. Methods 15, 913 (2018)
- [17] Saharia, C., et al. *Image Super-Resolution via Iterative Refinement*. Preprint at arXiv <https://doi.org/10.48550/arXiv.2104.07636> (2021)
- [18] Kim, T., et al. *Information-rich localization microscopy through machine learning*. Nat Commun 10, 1996 (2019).

## A Appendix

Standard SMLM localization algorithms based on maximum likelihood estimators or least squares optimization require tight control of activation and reactivation to maintain sparse emitters, presenting a tradeoff between imaging speed and labeling density. Recently, deep models have generalized SMLM to densely labeled structures by predicting high-resolution kernel density estimates (KDEs) from low resolution images with convolutional networks. However, estimated KDEs may contain irregularities due to finite sample sizes and limited model capacity.

Single molecule localization microscopy (SMLM) relies on the temporal resolution of fluorophores whose spatially overlapping point spread functions would otherwise render them unresolvable at the detector. Common strategies for the temporal separation of molecules involve molecular photoswitching from dark to fluorescent states, permitting resolution of fluorophores beyond the diffraction limit. Estimation of molecular coordinates is typically carried out by modeling the optical impulse response of the imaging system and fitting model functions to the data. However, such models are only well-suited to isolated molecules, reducing the number of molecules in the field of view and limiting temporal resolution in super resolution microscopy. This issue has incited a series of efforts to increase the density of fluorescent molecules imaged in a single frame while developing appropriate models for dense localization.

In fluorescence microscopy, each pixel is treated as a Poisson random variable (Smith 2010; Nehme 2020; Chao 2016), with expected value

$$\omega = i_0 \int O(u) du \int O(v) dv \quad (10)$$

where  $i_0 = \eta N_0 \Delta$ . The scalar parameters  $\eta, \Delta$  are the photon detection probability of the sensor and the exposure time, respectively. Without loss of generality, we assume  $\eta = \Delta = 1$ . Most importantly,  $N_0$  represents the signal amplitude, which we assume maintains a fixed value. The optical impulse response  $O(u, v)$  is often approximated as a 2D isotropic Gaussian with standard deviation  $\sigma$  (Zhang 2007). This approximation has the convenient property, that the effects of pixelation can be expressed in terms of error functions. For example, given a fluorescent emitter located at  $\theta = (u_0, v_0)$ , we have that

$$\int O(u) du = \frac{1}{2} \left( \operatorname{erf} \left( \frac{u_k + \frac{1}{2} - u_0}{\sqrt{2}\sigma} \right) - \operatorname{erf} \left( \frac{u_k - \frac{1}{2} - u_0}{\sqrt{2}\sigma} \right) \right) \quad (11)$$

where we have used the common definition  $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt$ . Our generative model also incorporates a normally distributed white noise per pixel  $\zeta$  with offset  $o$  and variance  $\sigma^2$ . Ultimately, we have a Poisson component of the signal, which scales with  $N_0$  and a Gaussian component, which does not.

Consider,

$$\zeta_k - o_k + \sigma_k^2 \sim \mathcal{N}(\sigma_k^2, \sigma_k^2) \approx \text{Poisson}(\sigma_k^2) \quad (12)$$

Since  $\mathbf{x}_k = \mathbf{s}_k + \zeta_k$ , we transform  $\mathbf{x}'_k = \mathbf{x}_k - o_k + \sigma_k^2$ , which is distributed according to

Consider the factorization  $p(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{y})p(\mathbf{x}|\mathbf{y})p(\mathbf{y}) = p(\mathbf{x}|\mathbf{y}, \hat{\mathbf{y}})p(\mathbf{y}|\hat{\mathbf{y}})p(\hat{\mathbf{y}})$ . Given that  $\mathbf{x}$  is conditionally independent of  $\hat{\mathbf{y}}$ , we find

$$p_{\Psi}(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{y}) = p(\mathbf{y}|\hat{\mathbf{y}})$$