
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

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

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

2 Background

2.1 Image Likelihood and Localization Error

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

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

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

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

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

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

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

2.2 Kernel density estimation with deep networks

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

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

3 Diffusion Model for SMLM

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 \mathbf{y}_i , and KDE estimates $\hat{\mathbf{y}}_i = \phi(\mathbf{x}_i)$. Observations \mathbf{x}_i are simulated under the likelihood (2). We aim to develop a framework for sampling from $p(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{y})$.

4 Conditional Diffusion Model

Point estimates $\hat{\mathbf{y}}_i$ produced by the DeepSTORM architecture lack uncertainty quantification. To address this, we propose a DDPM to model the conditional distribution $p(\hat{\mathbf{y}}|\mathbf{x}, \mathbf{y})$. 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

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

Evidently, the DDPM Ψ can be trained on pairs $(\mathbf{y}_i, \hat{\mathbf{y}}_i)_{i=1}^N$. The conditional DDPM generates a target 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 refines the KDE through successive iterations according to learned conditional transition distributions $p(\mathbf{y}_{t-1}|\mathbf{y}_t, \cdot)$ such that $\mathbf{y}_0 \sim p(\mathbf{y}|\hat{\mathbf{y}})$.

4.1 Gaussian Diffusion

Diffusion models (Sohl-Dickstein 2015; Ho 2020) are a class of generative models inspired by nonequilibrium statistical physics, which slowly destroy structure in a data distribution $p(\mathbf{y}_0|\mathbf{x})$ via a fixed Markov chain referred to as the *forward process*. In the present context, the forward process gradually adds Gaussian noise to the KDE \mathbf{y} according to a variance schedule $\beta_{0:T}$.

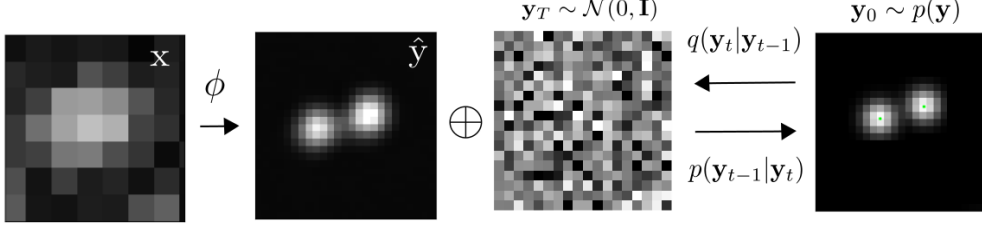


Figure 2: Conditional diffusion model for sampling kernel density estimates

$$q(\mathbf{y}_t | \mathbf{y}_0) = \prod_{t=1}^T q(\mathbf{y}_t | \mathbf{y}_{t-1}) \quad q(\mathbf{y}_t | \mathbf{y}_{t-1}) = \mathcal{N}(\sqrt{1 - \beta_t} \mathbf{y}_{t-1}, \beta_t \mathbf{I}) \quad (4)$$

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

$$q(\mathbf{y}_t | \mathbf{y}_0) = \mathcal{N}(\sqrt{\gamma_t} \mathbf{y}_0, (1 - \gamma_t) \mathbf{I}) \quad (5)$$

114 The usual procedure is then to learn a parametric representation of the *reverse process*, and therefore
 115 generate samples from $p(\mathbf{y}_0)$, starting from noise. Formally, $p_\theta(\mathbf{y}_0 | \hat{\mathbf{y}}) = \int p_\theta(\mathbf{y}_{0:T} | \hat{\mathbf{y}}) d\hat{\mathbf{y}}_{1:T}$ where
 116 \mathbf{y}_t is a latent representation with the same dimensionality of the data. $p_\theta(\mathbf{y}_{0:T} | \hat{\mathbf{y}})$ is a Markov process,
 117 starting from a noise sample $p_\theta(\mathbf{y}_T) = \mathcal{N}(0, \mathbf{I})$.

$$p_\theta(\mathbf{y}_{0:T}) = p_\theta(\mathbf{y}_T) \prod_{t=1}^T p_\theta(\mathbf{y}_{t-1} | \mathbf{y}_t) \quad p_\theta(\mathbf{y}_{t-1} | \mathbf{y}_t) = \mathcal{N}(\mu_\theta(\mathbf{y}_t), \beta_t \mathbf{I}) \quad (6)$$

118 where we reuse the variance schedule of the forward process (Ho 2020). We seek to learn a denoising
 119 model μ_θ which computes the mean of the Gaussian transition density at each time step t . For all
 120 $t > 0$, the mean of the transition density is computed as

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

121 where f_θ is a neural network. Only at $t = 0$ is this mean directly a function of \mathbf{x} .

122 4.2 Optimization of the Denoising Model

123 To reverse the diffusion process, we optimize a neural denoising model f_θ that takes as input $\hat{\mathbf{y}}$ and a
 124 noisy target image $\mathbf{y}_t \sim q(\mathbf{y}_t | \mathbf{y}_0)$. That is, this noisy target image \mathbf{y}_t is drawn from the marginal
 125 distribution of noisy images at a time step t of the forward diffusion process.

$$\mathbf{y}_t = \sqrt{\gamma_t} \mathbf{y}_0 + \sqrt{1 - \gamma_t} \epsilon, \quad \epsilon \sim \mathcal{N}(0, \mathbf{I}) \quad (8)$$

126 In addition to a source image \mathbf{y}_0 and a noisy target image \mathbf{y}_t , the denoising model f_θ takes as input
 127 the sufficient statistics for the variance of the noise γ , and is trained to predict the noise vector ϵ .
 128 We make the denoising model aware of the level of noise through conditioning on a scalar γ . The
 129 proposed objective function for training f_θ is

$$\mathbb{E}_{(\hat{\mathbf{y}}, \mathbf{y}_0)} \mathbb{E}_{(\epsilon, \gamma)} \left[f_\theta \left(x, \sqrt{\gamma} \mathbf{y}_0 + \sqrt{1 - \gamma} \epsilon \mid \mathbf{y}_t, \gamma \right) - \epsilon \right], \quad (9)$$

130 where $(\hat{\mathbf{y}}, \mathbf{y}_0)$ is sampled from the training dataset and $\gamma \sim p(\gamma)$. The distribution of γ has a big
 131 impact on the quality of the model and the generated outputs. For our training noise schedule, we
 132 use a piecewise distribution for γ , $p(\gamma) = \frac{1}{T} \sum_{t=1}^T U(\gamma_{t-1}, \gamma_t)$ (Nanxin 2021). Specifically, during

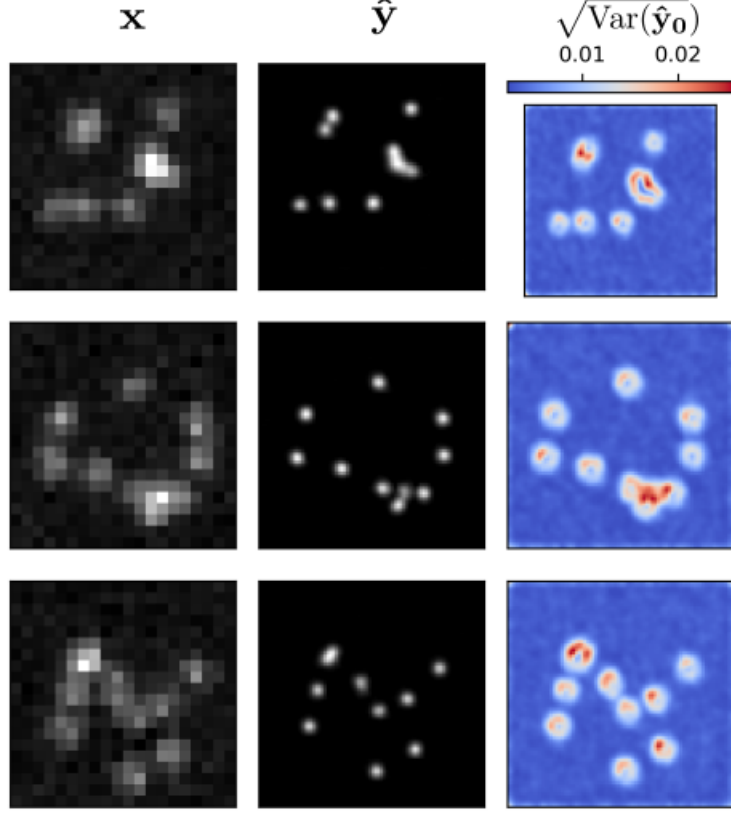


Figure 3: Kernel density estimates for various signal to noise ratios (SNR)

133 training, we first uniformly sample a time step $t \sim \{0, \dots, T\}$ followed by sampling $\gamma \sim U(\gamma_{t-1}, \gamma_t)$.
 134 We set $T = 100$ in all our experiments.

135 4.3 Optimization of the DeepSTORM architecture

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

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

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

149 5 Experiments

150 Training data was simulated under the image degradation model, drawing coordinates uniformly over
 151 a disc. We set $T = 100$ for all experiments and treat forward process variances β_t as hyperparameters,
 152 with a linear schedule from $\beta_0 = 10^{-4}$ to $\beta_T = 10^{-2}$. These constants were chosen to be small

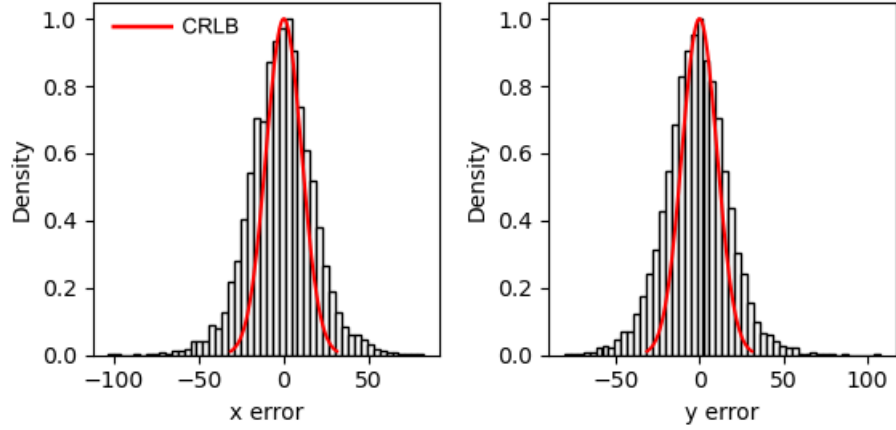


Figure 4: Localization errors of the trained model

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 ($L_T = D_{KL}(q(x_T|x_0)\|\mathcal{N}(0, I)) \approx 10^{-5}$ 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 16×16 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 γ 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 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).

182 [3] Speiser, A., et al. *Deep learning enables fast and dense single-molecule localization with high accuracy.*
183 Nature Methods 18, 1082–1090 (2021).

184 [4] Sohl-Dickstein J., et al. *Deep unsupervised learning using nonequilibrium thermodynamics.* ICLR (2015).

185 [5] Ho J., et al. *Denoising Diffusion Probabilistic Models.* Advances in Neural Information Processing Systems
186 (2015).

187 [6] Nanxin C., et al. *WaveGrad: Estimating Gradients for Waveform Generation .* ICLR (2021).

188 [4] Chao, J., et al. *Fisher information theory for parameter estimation in single molecule microscopy: tutorial.*
189 Journal of the Optical Society of America A 33, B36 (2016).

190 [5] Schermelleh, L. et al. *Super-resolution microscopy demystified.* Nature Cell Biology vol. 21 72–84 (2019).

191 [6] Zhang, B., et al. *Gaussian approximations of fluorescence microscope point-spread function models.* (2007).

192 [7] Smith, C.S., *Fast, single-molecule localization that achieves theoretically minimum uncertainty.* Nature
193 Methods 7, 373–375 (2010).

194 [8] Nieuwenhuizen, R., et al. *Measuring image resolution in optical nanoscopy.* Nature Methods 10. 557-562
195 (2013).

196 [9] Huang, F., et al. *Video-rate nanoscopy using sCMOS camera-specific single-molecule localization algorithms.*
197 Nat Methods 10, 653–658 (2013).

198 [10] Rust, M., et al. *Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM).*
199 Nat Methods 3, 793–796 (2006).

200 [11] Betzig, E., et al. *Imaging intracellular fluorescent proteins at nanometer resolution.* Science 313, 1642–1645
201 (2006).

202 [12] Weigert, M., et al. *Content-aware image restoration: pushing the limits of fluorescence microscopy.* Nat.
203 Methods 15, 1090 (2018).

204 [13] Falk, T., et al. *U-net: deep learning for cell counting, detection, and morphometry.* Nat. Methods 16, 67–70
205 (2019).

206 [14] Boyd, N., et al. *DeepLoco: fast 3D localization microscopy using neural networks.* Preprint at bioRxiv
207 <https://doi.org/10.1101/267096> (2018)

208 [15] Zelger, P., et al. *Three-dimensional localization microscopy using deep learning.* Opt. Express 26,
209 33166–33179 (2018)

210 [16] Zhang, P., et al. *Analyzing complex single-molecule emission patterns with deep learning.* Nat. Methods 15,
211 913 (2018)

212 [17] Saharia, C., et al. *Image Super-Resolution via Iterative Refinement.* Preprint at arXiv
213 <https://doi.org/10.48550/arXiv.2104.07636> (2021)

214 [18] Kim, T., et al. *Information-rich localization microscopy through machine learning.* Nat Commun 10, 1996
215 (2019).

216 A Appendix

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

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

231 of efforts to increase the density of fluorescent molecules imaged in a single frame while developing
 232 appropriate models for dense localization.

233 In fluorescence microscopy, each pixel is treated as a Poisson random variable (Smith 2010; Nehme
 234 2020; Chao 2016), with expected value

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

235 where $i_0 = \eta N_0 \Delta$. The scalar parameters η, Δ are the photon detection probability of the sensor and
 236 the exposure time, respectively. Without loss of generality, we assume $\eta = \Delta = 1$. Most importantly,
 237 N_0 represents the signal amplitude, which we assume maintains a fixed value. The optical impulse
 238 response $O(u, v)$ is often approximated as a 2D isotropic Gaussian with standard deviation σ (Zhang
 239 2007). This approximation has the convenient property, that the effects of pixelation can be expressed
 240 in terms of error functions. For example, given a fluorescent emitter located at $\theta = (u_0, v_0)$, we have
 241 that

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

242 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
 243 incorporates a normally distributed white noise per pixel ζ with offset o and variance σ^2 . Ultimately,
 244 we have a Poisson component of the signal, which scales with N_0 and a Gaussian component, which
 245 does not.

246 Consider,

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

247 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