Accelerated Sampling in Image Super-Resolution via Equilibrium Diffusion

Anonymous Author(s)
Affiliation
Address
email

Abstract

Deep learning has recently attracted considerable attention from researchers in the natural sciences, particularly microscopists, for fast extraction of physically relevant information from images. However, simple and interpretable uncertainty quantification is lacking in these applications, and remains a necessary modeling component in high-risk research. In order to quantify uncertainty in otherwise deterministic image translation architectures, we propose a hybrid generative modeling framework based on denoising diffusion probabilistic models (DDPMs). Specifically, our model combines a deterministic neural network with a DDPM, which can improve conditional synthesis speed and fidelity of the DDPM, while providing a natural mechanism for uncertainty estimation via Langevin dynamics. We apply our model to the task of single molecule localization in fluorescence microscopy, and demonstrate that blending the DeepSTORM architecture with a DDPM permits simultaneous high-fidelity super-resolution with uncertainty estimation of kernel density estimates (KDEs) regressed by DeepSTORM. Our results suggest the proposed solution is an interesting addition to the modeling toolkit for fluorescence microscopists and the field of deep image translation in general.

1 Introduction

Deep learning has attracted tremendous attention from researchers in the natural sciences, with several foundational applications arising in microscopy, e.g., (Weigert 2018; Falk 2019). Recently, the application of deep image translation in single-molecule localization microscopy (SMLM) has received considerable interest (Ouyang 2018; Nehme 2020; Speiser 2021). SMLM techniques are a mainstay of fluorescence microscopy and can be used to produce a pointillist representation of biomolecules in the cell at diffraction-unlimited precision (Rust 2006; Betzig 2006). In previous applications of deep models to localization microscopy, super-resolution images can be recovered from a sparse set of localizations with conditional generative adversarial networks (Ouyang 2018) or kernel density estimation can be performed using convolutional networks (Nehme 2020; Speiser 2021). Here, we focus on the latter class of models which perform single molecule localization using neural networks.

Inferences in SMLM, and other super-resolution image reconstruction tasks, are often made on a single measurement, and thus common measures of model performance are based on localization errors computed over ensembles of simulated images. Unfortunately, this choice precludes computation of uncertainty at test time under a fixed model. Yet, 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 intractiblity in deep models have been deterministic uncertainty quantification (Amersfoort 2020), ensembling (Lakshmi-

Submitted to 38th Conference on Neural Information Processing Systems (NeurIPS 2024). Do not distribute.

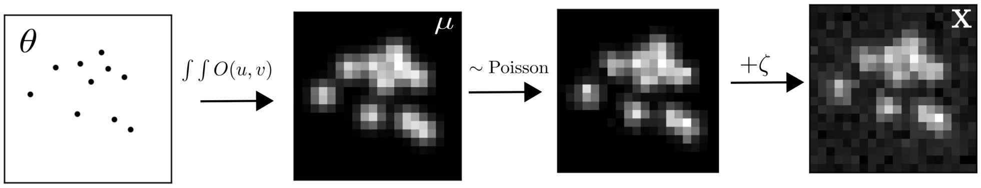

Figure 1: Generative model of single molecule localization microscopy images

narayanan et al., 2017) or Monte Carlo dropout (Gal and Ghahramani, 2016). Here, we choose to model a distribution on high-resolution KDE predictions conditioned on a low-resolution input using a denoising diffusion probabilistic model (DDPM) (Ho 2020; Song 2021), Such models are one class of *score based generative models* which implicitly compute the score of the data distribution at each noise scale starting from pure noise (Song 2021) .

In statistical physics, particularly sampling of complex molecular systems, integrating molecular dynamics does not begin with randomized configurations. Rather, sampling begins at a presumed global optimum, which is typically a configuration with minimal energy. This choice provides a reasonable way of integrating prior information about the configuration before sampling begins. Similarly, super-resolution tasks admit many neighboring reconstructions around the global optimum which are physically reasonable, while most others are not. In a similar vein, we propose sampling is preceded by an analagous procedure to energy-minimization, realized by a CNN, followed by Langevin-based sampling with a DDPM. Our approach admits faster sampling and produces pixel-wise uncertainties in model predictions with no modification to the existing architecture. This technique 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 $p(\zeta_k) = \mathcal{N}(o_k, \sigma_k^2)$

$$p(\mathbf{x}_k|\theta) = A\sum_{q=0}^{\infty}\frac{1}{q!}e^{-\omega_k}\omega_k^q\frac{1}{\sqrt{2\pi}\sigma_k}e^{-\frac{(\mathbf{x}_k-g_kq-o_k)}{2\sigma_k^2}} \approx \text{Poisson}(\omega_k') \tag{1}$$

where A is some normalization constant and $\omega_k' = \omega_k + \sigma_k^2$. For the sake of generality, we include a per-pixel gain factor g_k, which is often unity. In practice, the summation in (1) can be difficult to work with, and it is common to instead use a Poisson-Normal approximation for simplification, valid under a range of experimental conditions (Huang 2013). This result can be seen from the fact the the convolution of two Poisson distributions is also Poisson. The expectation of the Poisson process at each pixel of the image is computed from the optical transfer function $O(u, v)$, which is often a two-dimensional isotropic Gaussian.

$$\omega = i_0\iint O(u,v)dudv \tag{2}$$

The above integration can be carried out by computing differences of error functions, as detailed in Appendix A. The complete generative process is depicted in Figure 1.

Reliable estimation of θ from $\mathbf{x}$, for example by maximum likelihood estimation or with a deep model, requires performance metrics for model selection. We use the Fisher information as an information theoretic criteria to assess the quality of the model tested here, with respect to the root mean squared error (RMSE) of our predictions of θ (Chao 2016). The Poisson log-likelihood

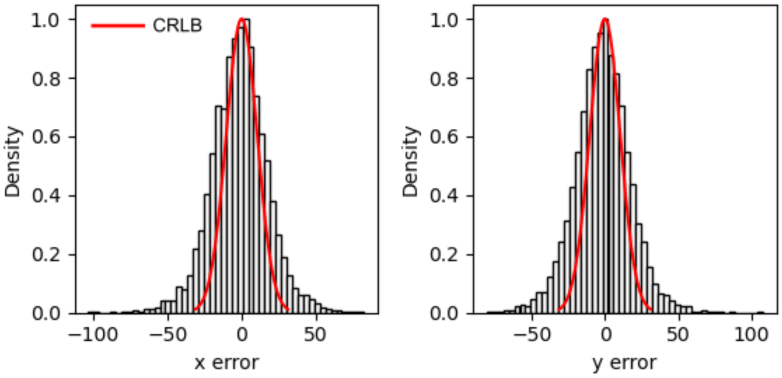

Figure 2: Localization errors of the trained model

$\ell(\mathbf{x}|\theta)$ 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., $\mathrm{var}(\hat{\theta}) \geq I^{-1}(\theta)$. The Fisher information is straightforward to compute under the Poisson log-likelihood, which is detailed in the Appendix

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

2.2 Kernel density estimation with deep networks

Direct optimization of the likelihood in (1) from observations $\mathbf{x}$ alone is challenging when fluorescent emitters are dense within the field of view and fluorescent signals significantly overlap. However, convolutional neural networks (CNN) have recently proven to be powerful tools 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). We choose to use kernel density estimates in our model, denoted by $\mathbf{y}$. KDEs are the most common data structure used in SMLM, and can be easily generated from molecular coordinates, alongside observations $\mathbf{x}$, using well-understood models of the optical impulse response (Zhang 2007).

3 Image Super-Resolution via Equilibrium Diffusion

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 Poisson likelihood (1) and KDEs are generated using (2) alone, followed by appropriate normalization.

3.1 Problem Statement

Point estimates $\hat{\mathbf{y}}_i$ produced by the traditional deep architectures for super resolution microscopy produce strong results, but lack uncertainty quantification. Recent advances in generative modeling, particularly DDPMs, therefore present a unique opportunity to integrate uncertainty awareness into the super-resolution microscopy toolkit. However, sampling from DDPMs is computationally expensive, given that generation amounts to solving a complex stochastic differential equation, effectively mapping a simple base distribution to the complex data distribution. The solution of such equations requires numerical integration with very small step sizes, resulting in thousands of neural network evaluations (Saharia 2021; Vahdat 2021). Furthermore, for conditional generation tasks in high-risk applications, generation complexity is further exacerbated by the need for the highest level of detail in generated samples.

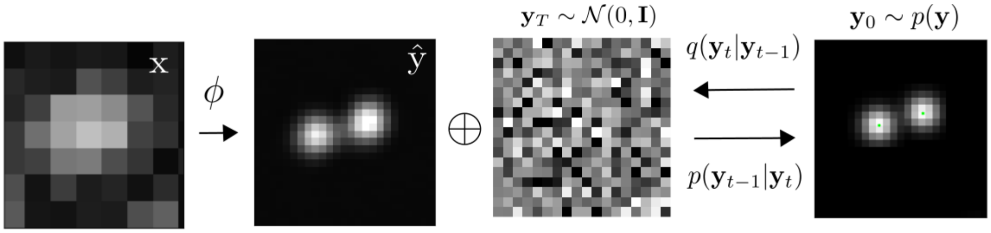
x
ϕ
$\hat{\mathrm{y}}$
$\oplus$
$\mathbf{y}_T \sim \mathcal{N}(0, \mathbf{I})$
$q(\mathbf{y}_t|\mathbf{y}_{t-1})$
$p(\mathbf{y}_{t-1}|\mathbf{y}_t)$
$\mathbf{y}_0 \sim p(\mathbf{y})$

Figure 3: Conditional diffusion model for sampling kernel density estimates

Conditional generation tasks, for example sampling from the distribution $p(\mathbf{y}|\mathbf{x})$, can benefit significantly from more suitable initial conditions. Therefore, we propose that DDPM sampling is preceded by a deterministic neural network ϕ, which seeds sampling in a target mode. Reasoning for this choice in the current application is two-fold:

Synthesis Speed. By training a preprocessor ϕ to obtain an approximate estimate of $\mathbf{y}$, we can reduce the number of iterations, since the DDPM only needs to model the remaining mismatch, resulting in a less complex model from which sampling becomes easier. Speed is critical in SMLM applications, which can produce large volumes of image data in a single experiment. Moreover, we note that this approach is analagous to preconditioned stochastic gradient langevin dynamics (Li 2016), wherein ϕ identifies the posterior mode followed by Langevin dynamics to sample from the posterior.

Sample Fidelity. Since Langevin dynamics will often be initialized in low-density regions of the data distribution, inaccurate score estimation in these regions will negatively affect the sampling process (Song 2019). Moreover, mixing can be difficult because of the need of traversing low density regions to transition between modes of the distribution. Preprocessing with a deterministic mapping ϕ can ameliorate this issue, by eliminating the need for score estimation in low density regions.

The preprocessor ϕ is realized by a CNN with upsampling layers. Consider the Markov chain wherein the KDE $\mathbf{y}$ is latent in and inferred from a noisy measurement $\mathbf{x}$, i.e., $\mathbf{x} \rightarrow \phi(\mathbf{x}) \rightarrow \hat{\mathbf{y}}$. By the data processing inequality the function ϕ can only destroy information in $\mathbf{x}$ pertaining to $\mathbf{y}$ i.e., $I(\mathbf{x};\mathbf{y}) \geq I(\phi(\mathbf{x});\mathbf{y})$ or $h(\mathbf{y}|\phi(\mathbf{x})) \geq h(\mathbf{y}|\mathbf{x})$ where I is the mutual information and h is the entropy. In other words, the function ϕ, while deterministic, can introduce additional uncertainty about $\mathbf{y}$ in downstream stochastic models by destroying information. Here, we are interested in measuring the upper bound $h(\mathbf{y}|\phi(\mathbf{x}))$, as this is the relevant quantity when a deterministic transformation ϕ is an unavoidable first step.

In practice, a 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 $p(\mathbf{y}_{t-1}|\mathbf{y}_t,)$ such that $\mathbf{y}_0 \sim p(\mathbf{y}|\hat{\mathbf{y}})$

3.2 Equilibrium Diffusion

Diffusion models (Sohl-Dickstein 2015; Ho 2020; Song 2021) 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, we apply leverage recent results from (Ho 2020; Song 2021; Saharia 2021) for applying this framework to sampling from $p(\mathbf{y}|\mathbf{x}, \hat{\mathbf{y}})$. The forward process gradually adds Gaussian noise to the KDE $\mathbf{y}$ according to a variance schedule $\beta_{0:T}$

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

The usual procedure is then to learn a parametric representation of the *reverse process*, and therefore 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 $\mathbf{y}_t$ is a latent representation with the same dimensionality of the data. $p_\theta(\mathbf{y}_{0:T}|\hat{\mathbf{y}})$ is a Markov process, starting from a noise sample $p_\theta(\mathbf{y}_T) = \mathcal{N}(0, \mathbf{I})$.

$$p_\theta(\mathbf{y}_{0:T}) = p_\theta(\mathbf{y}_T)\prod_{t=1}^{T} p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) \quad p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) = \mathcal{N}\left(s_\theta(\mathbf{y}_t), \beta_t I\right) \tag{5}$$

where we reuse the variance schedule of the forward process (Ho 2020). We omit conditioning on $\hat{\mathbf{y}}$ for each transition density $p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t)$, as this is only considered at $t = 0$ i.e., $p_\theta(\mathbf{y}_1|\mathbf{y}_0, \hat{\mathbf{y}})$. An important property of the forward process is that it admits sampling $\mathbf{y}_t$ 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 $q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}\left(\sqrt{\gamma_t}\mathbf{y}_0, (1-\gamma_t)I\right)$.

$$\mathcal{L}(\theta) = \mathbb{E}\left[-\log p_\theta(\mathbf{y}_0|\mathbf{x})\right] \leq \mathbb{E}\left[-\log \frac{p_\theta(\mathbf{y}_{0:T}|\mathbf{x})}{q(\mathbf{y}_{1:T}|\mathbf{y}_0)}\right] \tag{6}$$

The objective in (6) can be expanded in terms of $D_{\mathrm{KL}}(p(\mathbf{y}_{t-1}|\mathbf{y}_t)||q(\mathbf{y}_t|\mathbf{y}_{t-1})$ as detailed in (Ho 2020). We choose to adopt the simplified from of the variational bound, which emphasizes that the DDPM estimates the score $\nabla_{\mathbf{y}} \log p(\mathbf{y}|\mathbf{x})$ at each noise level (Song 2021)

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \underset{(\hat{\mathbf{y}},\mathbf{y}_0)}{\mathbb{E}}\ \underset{(\epsilon,\gamma)}{\mathbb{E}} \left[s_\theta\left(x, \sqrt{\gamma}\mathbf{y}_0 + \sqrt{1-\gamma}\epsilon \,\middle|\, \mathbf{y}_t, \gamma\right) - \epsilon\right], \tag{7}$$

After training, samples can be generated by

$$\mathbf{y}_{t-1} = \frac{1}{\sqrt{1-\beta_i}}\left(\mathbf{y}_i + \beta_i s_\theta(\mathbf{y}_t)\right) + \sqrt{\beta_i}\xi \tag{8}$$

For many conditional generation tasks, estimation of the gradient s_θ in low-density regions in order to drive (8) toward high-density regions is unecessary and reduces performance of the sampler. Here, we instead propose an "energy-minimization" procedure preceding Langevin-based sampling in a DDPM, in order to speed up sampling from the equilibrium distribution

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\operatorname{argmin}} ||\mathbf{y} - \mathbf{y_0}||^2 \tag{9}$$

4 Experiments

All training data consits of low-resolution 20×20 images, simulated under the likelihood and impulse reponse (2,10), setting $\sigma = 0.92$ low-resolution pixels, for consistency with common experimental conditions with a 60X magnification objective lens and numerical aperture (NA) of 1.4. We choose $\omega_k = 200$ for experiments for consistency with typical bright fluorophore emission rates. All KDEs have dimension 80×80, are scaled between $[0, 1]$, and are generated using $\sigma = 3.0$ pixels in the upsampled image. For a typical CMOS camera, this results in KDE pixels with lateral dimension of $\approx 27\mathrm{nm}$. Initial coordinates θ were drawn uniformly over a two-dimensional disc with a radius of 7 low-resolution pixels.

4.1 Localization RMSE

In order to verify the initial predictions made by the model ϕ, we simulated a dataset $(\mathbf{x}_i, \mathbf{y}_i, \hat{\mathbf{y}}_i)_{i=1}^{N}$ with $N = 1000$, and detect objects in the predicted KDE , $\hat{\mathbf{y}}_i$ using the Laplacian of Gaussian (LoG) detection algorithm (Lindeberg 2013). For simplicity, the localization is carried out from scale-space maxima directly in LoG, as opposed to fitting a model function to KDE predictions. A particular LoG localization in the KDE is paired to the nearest ground truth localization and is unpaired if a localization is not within 5 KDE pixels of any ground truth localization. In addition to localization error, we measured a precision $\mathrm{P} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FP}) = 1.0$ and recall $\mathrm{R} = \mathrm{TP}/(\mathrm{TP}+\mathrm{FN}) = 0.85$, where TP denotes true positive localizations, FP denotes false positive localizations, and FN denotes false negative localizations.

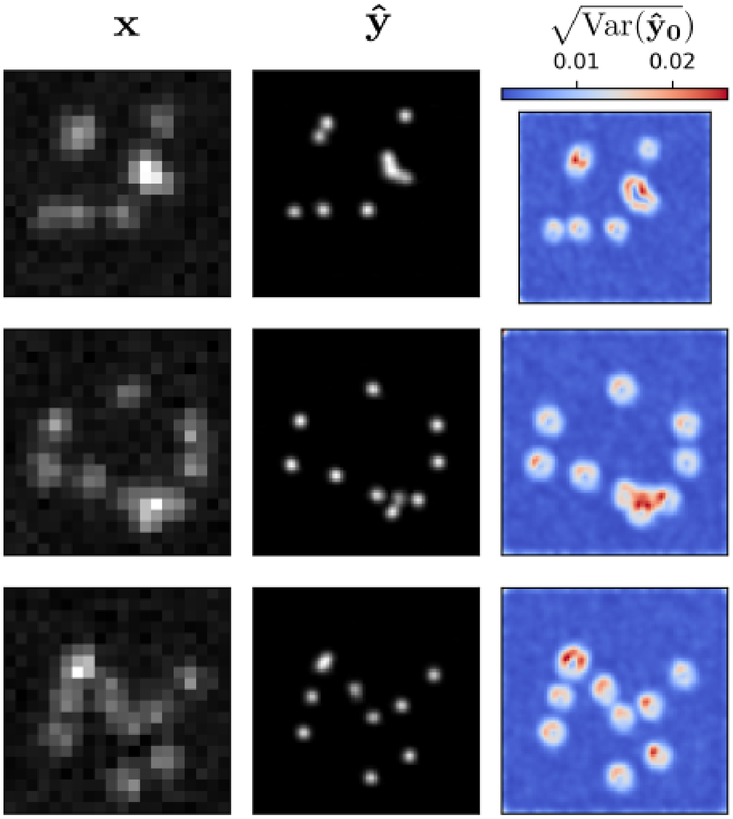

Figure 4: Kernel density estimates for various signal to noise ratios (SNR)

4.2 Model Uncertainty

We set $T = 100$ for all experiments and treat forward process variances β_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 ground truth KDEs, which are scaled to $[-1, 1]$, ensuring that forward process distribution $\mathbf{y}_T \sim q(\mathbf{y}_T|\mathbf{y}_0)$ approximately matches the reverse process $\mathbf{y}_T \sim \mathcal{N}(0, I)$ at $t = T$.

To represent the reverse process, we used a DDPM architecture based on a U-Net backbone proposed in (Saharia 2021). We chose a U-Net backbone with channel multipliers $[1, 2, 4, 8, 8]$ in the downsampling and upsampling paths of the architecture. 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. To condition the model on the input $\hat{\mathbf{y}}$, we concatenate the $\hat{\mathbf{y}}$ estimated by DeepSTORM along the channel dimension, which are scaled to $[0, 1]$, with $\mathbf{y}_T \sim \mathcal{N}(0, I)$. Others have experimented with more sophisticated methods of conditioning, but found that the simple concatenation yielded similar generation quality (Saharia 2021).

5 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.

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.

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 flurophores 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

$$\omega = i_0 \int O(u)du \int O(v)dv \tag{10}$$

where $i_0 = \eta N_0 \Delta$. The scalar parameters η, Δ 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 σ (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

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

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 ζ with offset o and variance σ^2. Ultimately, we have a Poisson component of the signal, which scales with N_0 and a Gaussian component, which does not.

Consider,

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

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

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