Uncertainty-Aware Localization Microscopy by Variational Diffusion

Clayton W. Seitz*
Department of Physics
Indiana University
Indianapolis, IN 46202
cwseitz@iu.edu

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. In particular, single molecule localization microscopy (SMLM) has benefited significantly from recent advances in deep image translation. 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 deep networks used for SMLM, we propose a generative modeling framework based on variational diffusion. This approach allows us to probe the structure of the posterior on reconstructions, creating an additional avenue toward quality control. Our model is tested on kernel density (KD) estimation in fluorescence microscopy, and we demonstrate that data augmentation with traditional SMLM architectures followed by diffusion permits simultaneous high-fidelity super-resolution with uncertainty estimation of regressed KDEs.

1 Introduction

Deep learning has attracted tremendous attention from researchers in the natural sciences, with several foundational applications arising in microscopy (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 localization itself can be performed using traditional convolutional networks (Nehme 2020; Speiser 2021). Here, we focus on the latter class of models which perform KD estimation using neural networks.

Inferences in SMLM 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 aleatoric 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 mostly epistemic uncertainty quantification such as ensembling (Lakshminarayanan et al., 2017) or Monte Carlo dropout (Gal and

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

Preprint. Under review.

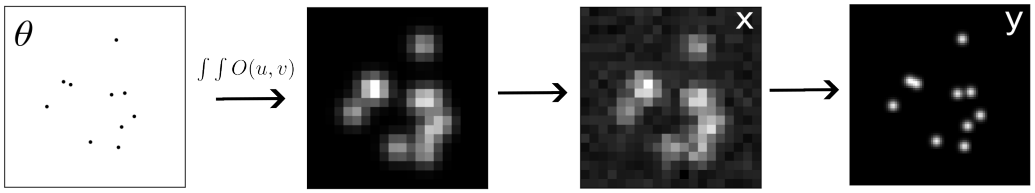

Figure 1: Generative model of single molecule localization microscopy images. Low resolution images $\mathbf{x}$ are generated from coordinates θ by integration of the optical transfer function O and sampling from the likelihood (1): $\mathbf{x} \sim p(\mathbf{x}|\theta) = \prod_k p(\mathbf{x}_k|\theta)$. A kernel density estimate $\mathbf{y}$ is inferred from $\mathbf{x}$

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).

Variational perspectives on diffusion have also been studied (Tzen 2019; Huang 2021; Vahdat 2021). A primary advantage of this approach is that it provides an intuitive expression of the variational lower bound in terms of the signal-to-noise ratio of the diffused data, leading to much simplified expressions of the loss. Moreover, recent efforts have shown that the variational bound can be reparameterized to conventional diffusion losses such as the noise estimation loss (Ho 2020), emphasizing the utility of diffusion models for variational inference. In the remainder of this paper, we first introduce the likelihood of low-resolution images in localization microscopy, and show uncertainty quantification in a rudimentary example scenario. Then, we introduce kernel density estimation as an alternative to direct localization using low-resolution images, and conclude by demonstrating our variational diffusion model for measuring uncertainty KD estimation at scale.

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 $p(\mathbf{x}_k|\theta)$ 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 $p(\zeta_k) = \mathcal{N}(o_k, w_k^2)$

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

where A is some normalization constant. After subtraction of a known offset o_k of the pixel array, which can be easily measured, we have $\omega_k' = \omega_k + w_k^2$. For the sake of generality, we include a per-pixel gain factor g_k, which is often unity. Sampling from $p(\mathbf{x}_k|\theta)$ is trivial; however, for computation of a lower bound on uncertainty in θ, the summation in (1) can be difficult to work with. Therefore, we choose to use a Poisson approximation for simplification, valid under a range of experimental conditions (Huang 2013). The expectation of the Poisson process ω_k at each pixel of the image must then be computed from the optical impulse response $O(u, v)$, which is taken to be an isotropic Gaussian in two-dimensions. Consider an idealized scenario where an isolated fluorescent molecule exists in the image plane. For a particular pixel of width δ centered at $\xi_k = (u_k, v_k)$ in the first dimension, we define

$$\Delta E_{\theta_i} := \int_{\xi_{k,i}-\delta/2}^{\xi_{k,i}+\delta/2} O(u; \theta_i) du = \frac{1}{2} \left(\text{erf}\left(\frac{\xi_{k,i} + \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) - \text{erf}\left(\frac{\xi_{k,i} - \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) \right) \tag{2}$$

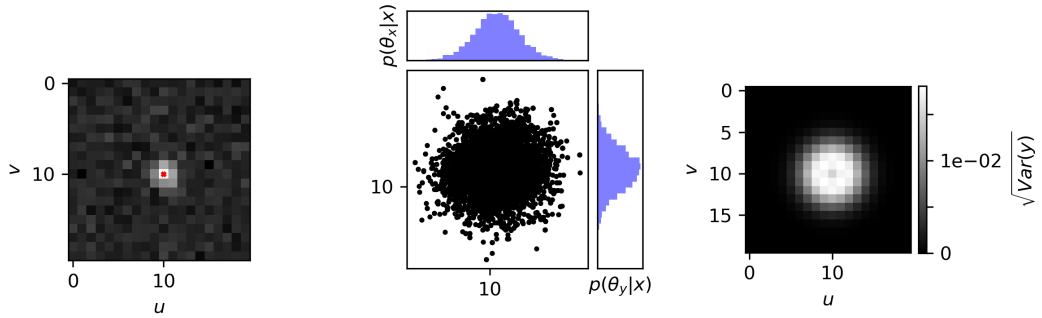

Figure 2: Monte Carlo estimation of the marginal variances $\sqrt{\mathrm{Var}(\mathbf{y}_k)}$ for an isolated fluorescent emitter. MCMC sampling is carried out using the low-resolution image (left) to estimate the posterior (middle) and generate an image of marginal variances (right)

The expected value at each pixel is then $\omega_k \propto \prod_{\theta_i \in \theta} \Delta E_{\theta_i}(\xi_{k,i}, \theta_i, \sigma_{\mathbf{x}})$. Using this, sampling from (1) is can be carried out by $\mathbf{x}_k = s_k + \zeta_k$ for $s_k \sim \mathrm{Poisson}(\omega_k), \eta_k \sim \mathcal{N}(o_k, w_k^2)$. The complete generative process is depicted in Figure 1.

2.2 Gaussian kernel density estimation

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 $\mathbf{y}$, which are latent in the low-resolution data $\mathbf{x}$. 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).

Similar to the generative process on low resolution images $\mathbf{x}$, we can generate KDEs $\mathbf{y}$ by repurposing the generative model (1) on an unsampled image without noise. In other words, we cast Gaussian kernel density estimation as a noiseless image generation process on the domain of $\mathbf{y}$. Under a fixed configuration of N particles θ, the value of an non-normalized KDE pixel $\mathbf{y}_k$ is given by

$$\mathbf{y}_k(\theta) = \sum_{n=1}^{N} \Delta E_{\theta_x}(u_k, \theta_{n,x}, \sigma_{\mathbf{y}}) \Delta E_{\theta_y}(v_k, \theta_{n,y}, \sigma_{\mathbf{y}}) \tag{3}$$

where the hyperparameter $\sigma_{\mathbf{y}}$ is the Gaussian kernel width.

3 Uncertainty-Aware Localization Microscopy by Variational Diffusion

We consider datasets $(\mathbf{x}_i, \mathbf{y}_{0,i}, \hat{\mathbf{y}}_i)_{i=1}^N$ of observed images $\mathbf{x}_i$ true kernel density estimate (KDE) images $\mathbf{y}_{0,i}$, and augmented low-resolution inputs $\hat{\mathbf{y}}_i = \phi(\mathbf{x}_i)$, where ϕ is a separate deep network. Observations $\mathbf{x}_i$ are simulated under the convolution distribution (1) and KDEs are generated by (4).

3.1 Problem Statement

Point estimates $\hat{\mathbf{y}}_i$ produced by the traditional deep architectures for localization microscopy produce strong results, but lack uncertainty quantification. In principle, distribution on $\mathbf{y}_k$ is given directly from $p(\theta|\mathbf{x})$, which can be estimated by Metropolis-Hastings MCMC (Figure 2). Unfortunately, the posterior $p(\theta|\mathbf{x})$ has no known analytical form and can be difficult to compute at test time, since (i)

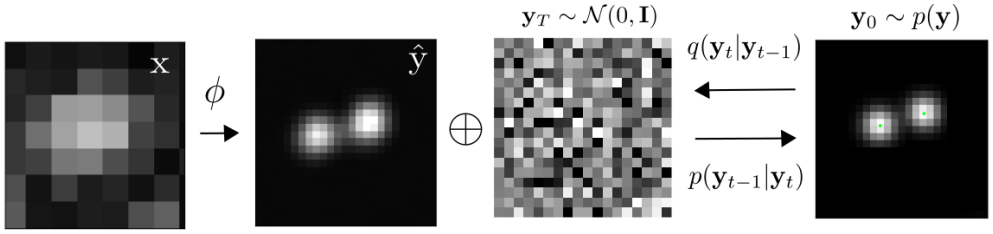

Figure 3: Conditional diffusion model for sampling kernel density estimates. An initial KDE estimate $\hat{\mathbf{y}}$ is found by the CNN ϕ, which is used as conditional input in the DDPM

molecules cannot be easily resolved and therefore θ is of unknown dimension and (ii) θ can be high dimensional and efficient exploration of the parameter space is challenging. The central goal of this paper is to instead model a conditional distribution on the latent $\mathbf{y}$: $p(\mathbf{y}|\mathbf{x})$, where $\mathbf{y}$ is of known dimensionality. We choose to model $p(\mathbf{y}|\mathbf{x})$ with a diffusion model, given that the distribution $p(\mathbf{y}|\mathbf{x})$ is expensive to compute, even if $p(\theta|\mathbf{x})$ were known.

Recent advances in generative modeling, particularly diffusion models (Sohl-Dickstein 2015; Ho 2020; Song 2021) present a unique opportunity to integrate uncertainty awareness into the localization microscopy toolkit. However, sampling from diffusion models can be 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). 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. Moreover, in the present application, modeling the posterior is far more important than diversity in generated samples. Therefore, we propose that sampling is preceded by an augementation network ϕ, which in essence generates an initial estimate to guide the diffusion process. Reasoning for this choice in our application is two-fold:

Synthesis Speed. By training the augmentation network ϕ to obtain an approximate estimate of $\mathbf{y}_0$, we can reduce the number of iterations, since the diffusion model 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 thousands of images in a single experiment.

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 neural network ϕ can ameliorate this issue, by aiding score estimation in low density regions.

3.2 Variational Diffusion

Diffusion models (Sohl-Dickstein 2015; Ho 2020; Song 2021) are a class of generative models originally inspired by nonequilibrium statistical physics, which slowly destroy structure in a data distribution via a fixed Markov chain referred to as the *forward process*. In the present context, we leverage a novel interpretation of this model class as likelihood-based models (Kingma 2021; Kingma 2023) to approximate the posterior $p(\mathbf{y}|\mathbf{x})$.

Diffusion Model. The forward process gradually adds Gaussian noise to the latent KDE $\mathbf{y}_0$ 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 $p(\mathbf{y}_0|\hat{\mathbf{y}})$, starting from noise. Formally, $p_\psi(\mathbf{y}_0|\hat{\mathbf{y}}) = \int p_\psi(\mathbf{y}_{0:T}|\hat{\mathbf{y}}) d\hat{\mathbf{y}}_{1:T}$

where $\mathbf{y}_t$ is a latent representation with the same dimensionality of the data and $p_\psi(\mathbf{y}_{0:T}|\hat{\mathbf{y}})$ is a Markov process, starting from a noise sample $p_\psi(\mathbf{y}_T) = \mathcal{N}(0, I)$.

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

where we reuse the variance schedule of the forward process (Ho 2020). From (7) it can be seen that the learnable parameter in the reverse process is the expectation of the transition μ_ψ where ψ is a neural network. We omit conditioning on $\hat{\mathbf{y}}$ for each transition density $p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t)$, as this is only considered at the first step of the reverse process i.e., $p_\psi(\mathbf{y}_{T-1}|\mathbf{y}_T, \hat{\mathbf{y}})$. 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 $q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}\left(\sqrt{\gamma_t}\mathbf{y}_0, (1-\gamma_t)I\right)$ or $\mathbf{y}_t = \sqrt{\gamma_t}\mathbf{y}_0 + \sqrt{1-\gamma_t}\epsilon$ for $\epsilon \sim \mathcal{N}(0, I)$. Learning the reverse process can be approached by either regressing the starting noise ϵ or the true KDE $\mathbf{y}_0$ as there is a deterministic relationship between the two given by the forward process. We adopt the former for consistency with other work, and define s_ψ as a neural denoising function which regresses the noise ϵ from $\hat{\mathbf{y}}$. Importantly, the noise estimate ϵ_ψ and μ_ψ are related by (Saharia 2021)

$$\mu_\psi(\mathbf{y}_t, \gamma_t) = \frac{1}{\sqrt{1-\beta_t}} \left(\mathbf{y}_t - \frac{\beta_t}{\sqrt{1-\gamma_t}} \epsilon_\psi \right) \tag{6}$$

which forms the basic mechanism for transitions in the reverse process: $\mathbf{y}_{t-1} = \mu_\psi(\mathbf{y}_t, \gamma_t) + \sqrt{\beta_t}\xi_t$, for $\xi_t = \mathcal{N}(0, I)$.

Variational Objective. Following (Kingma 2021), we define our generative model by inverting a diffusion process, yielding a hierarchical generative model that samples a sequence of latents $\mathbf{y}_t$, with time running backward. Training of the model is achieved through the usual variational bound

$$-\log p(\mathbf{x}) \leq -\mathbb{E}_{q(\mathbf{y}_0|\mathbf{x})} \log p(\mathbf{x}|\mathbf{y}_0) + D_{KL}(q(\mathbf{y}_T|\mathbf{x})||p(\mathbf{y}_T)) + \mathcal{L}_\psi = \mathrm{VLB}(\mathbf{x}) \tag{7}$$

We focus on the diffusion loss $\mathcal{L}_\psi$, which is the only relevant term in the above objective (Rebeira 2024)

$$\mathcal{L}_\psi = \sum_{i=1}^{T} \mathbb{E}_{q_\psi(\mathbf{y}_t|\mathbf{x})}[D_{KL}(q_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x})||p(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}))] \tag{8}$$

In a diffusion model $p(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}) = q(\mathbf{y}_{t-1}|\mathbf{y}_t, \hat{\mathbf{x}}_\psi)$ which is modeled by $p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t) = \mathcal{N}\left(\mu_\psi(\mathbf{y}_t), \beta_t I\right)$. Importantly, $D_{KL}(q_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x})||p(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}))$ and therefore, the variational bound, are closely linked to the common noise estimation loss used in DDPMs (Ho 2020; Kingma 2021).

We use the following Monte Carlo estimate of $\mathcal{L}_\psi$, which demonstrates that the variational bound can be written in terms of the common noise-estimation loss

$$\mathcal{L}_\psi = \frac{T}{2} \mathbb{E}_{\epsilon \sim \mathcal{N}(0,I), t \sim U(1,T)}[\left(\frac{\mathrm{SNR}(t-1)}{\mathrm{SNR}(t)} - 1 \right) ||\epsilon - \epsilon_\psi(\mathbf{y}_t)||_2^2] \tag{9}$$

where $\mathrm{SNR}(t) = \gamma_t / \beta_t$. Full details of this derivation can be found in (Kingma 2021; Rebeira 2024) and are outlined in Appendix A.

4 Experiments

All training data consits of low-resolution 20×20 images, setting $\sigma_\mathbf{x} = 0.92$ in units of low-resolution pixels, for consistency with common experimental conditions with a 60X magnification objective lens and numerical aperture (NA) of 1.4. We choose $i_0 = 200$ for experiments for consistency with typical bright fluorophore emission rates. All KDEs have dimension 80×80, are scaled between $[0, 1]$,

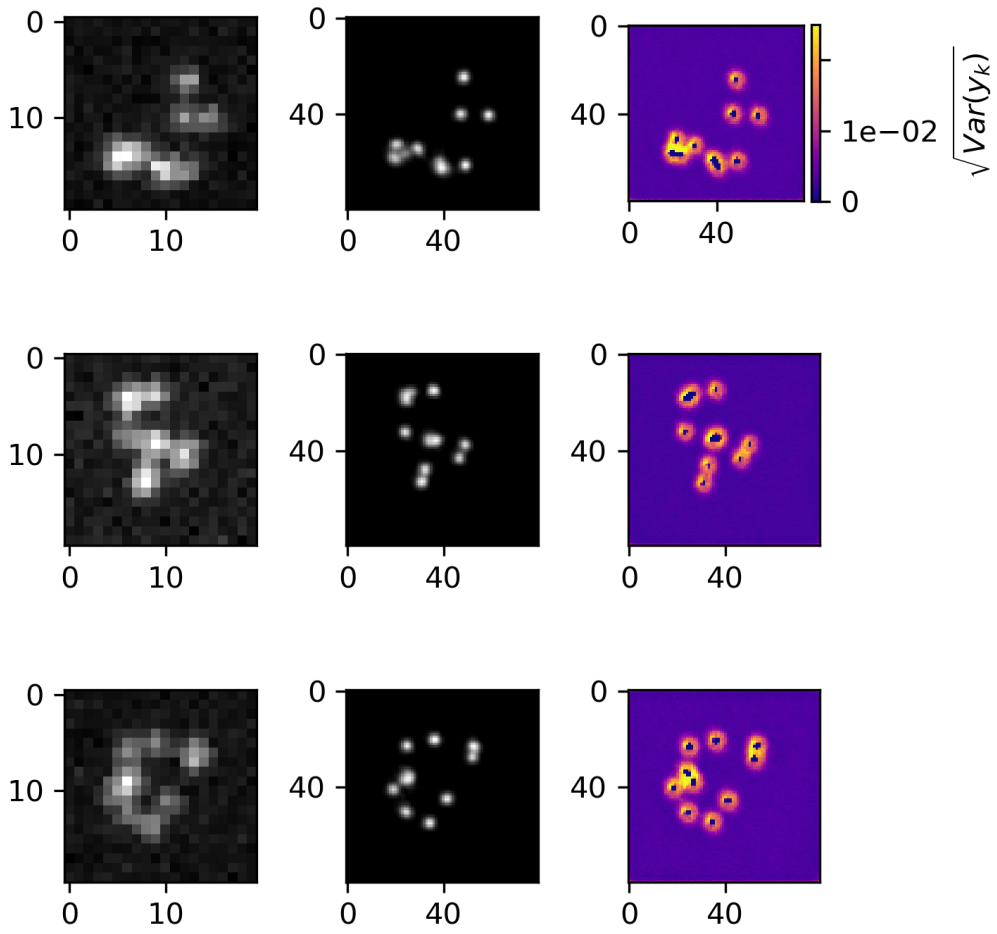

Figure 4: Non cherry-picked estimation of marginal variances. A low-resolution image $\mathbf{x}$ (left column) is transformed by ϕ to produce a KDE estimate $\hat{\mathbf{y}}$ (middle column) and a DDPM ψ computes a map of marginal variances (right column)

and are generated using $\sigma_{\mathbf{y}} = 3.0$ 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.

Localization RMSE. In order to verify the initial predictions made by the augmentation model ϕ, we simulated a dataset $(\mathbf{x}_i, \mathbf{y}_{0,i}, \hat{\mathbf{y}}_i)_{i=1}^N$ with $N = 1000$, and detect objects in the predicted $\hat{\mathbf{y}}_i$ using the Laplacian of Gaussian (LoG) detection algorithm (Lindeberg 2013). 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 measure 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.

Variational Diffusion. 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 $t = T$.

5 Conclusion

We proposed a variational diffusion model for uncertainty-aware localization microscopy. Our approach builds on recent advancements in conditional diffusion models, to model the posterior distribution on high-resolution kernel density estimates (KDE) from low-resolution inputs. This tractable posterior distribution is constructed by first augmenting low resolution inputs to a KDE estimate using a modified form of the DeepSTORM architecture (Nehme 2020). Conditioning a DDPM on this initial estimate permits sampling with relatively fewer samples than most existing DDPMs in similar applications, thereby making computation of marginal variances possible. Our approach made three core contributions: (i) we derived a relationship between the posterior on kernel density estimates with the posterior on molecular locations, and (ii) we demonstrated that a diffusion model can model a distribution on KDEs with qualitatively similar marginal variances expected from predictions made using MCMC. By using a recently discovered relationship of the variational lower bound to a traditional noise-estimation objective, we can confidently approximate the true posterior.

6 Broader Impact

The development of a method for uncertainty estimation in super-resolution imaging, as proposed here, holds profound implications beyond its immediate application in single-molecule localization microscopy (SMLM). By leveraging denoising diffusion probabilistic models (DDPMs) for uncertainty estimation, this approach not only enhances the reliability of super-resolution image reconstructions but also extends its utility to a diverse array of domains. The incorporation of a guided diffusion process facilitates efficient reconstruction while maintaining intepretation of the underlying uncertainty. Importantly, the principles underlying this method resonate across various fields, suggesting its potential applicability in domains beyond microscopy. For instance, the extension of similar techniques to general image resolution highlights the potential to address uncertainty in a wide range of bioimaging or medical imaging tasks. Moreover, the utilization of diffusion models for uncertainty estimation aligns with a broader trend in leveraging probabilistic frameworks for enhancing deep learning applications, with implications extending to fields such as natural language processing, computer vision, and autonomous systems. By bridging these interdisciplinary boundaries, this method not only addresses a critical need in localization microscopy but also contributes to the advancement of uncertainty-aware deep learning methodologies with far-reaching societal impacts. From improving medical diagnostics to enhancing environmental monitoring, the implications of this work reverberate across diverse applications, underscoring its potential to drive innovation and progress in numerous fields.

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).

[7] 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).

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

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

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

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

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

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

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

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

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

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

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

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

[20] Song, Y., et al. Score-based generative model through stochastic differential equations. ICLR (2021).

[21] Vahdat, A., et al. Score-based generative modeling in latent space. NeurIPS (2021).

[22] Kingma, D., et al. Auto encoding variational bayes. Preprint at arXiv https://doi.org/10.48550/arXiv.1312.6114 (2013).

[23] Song, Y., et al. Generative modeling by estimating gradients of the data distribution. NeurIPS (2019).

[24] Kong, X., et al. Information theoretic diffusion. ICLR (2023).

[25] Kong, L., et al. SDE-Net: Equipping Deep Neural Networks with Uncertainty Estimates. ICML (2020).

[26] van Amersfoort, J., et al. Uncertainty Estimation Using a Single Deep Deterministic Neural Network. Preprint at arXiv https://doi.org/10.48550/arXiv.2003.02037.

[27] Zhe Liu, J., et al. Simple and Principled Uncertainty Estimation with Deterministic Deep Learning via Distance Awareness. NeurIPS 2021.

[28] Gal Y., et al. Bayesian uncertainty quantification for machine-learned models in physics. Nature Physics 2022.

[29] Gal Y., et al. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. ICML 2016.

[30] Welling M., et al. Bayesian Learning via Stochastic Gradient Langevin Dynamics. ICML 2011.

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

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

[33] Nichol, A., et al. Improved denoising diffusion probabilistic models. Preprint at arXiv https://doi.org/10.48550/arXiv.2102.09672 (2021).

[34] Kingma, D., et al. Understanding Diffusion Objectives as the ELBO with Simple Data Augmentation. NeurIPS (2023).

[35] Kingma, D., et al. Variational Diffusion Models. NeurIPS (2021).

[35] Ribeiro, F., et al. Demystifying Variational Diffusion Models. Preprint at arXiv https://doi.org/10.48550/arXiv.2401.06281 (2024).

A Appendix

A.1 Optical impulse response

It is common to describe the optical impulse response of a microscope as a two-dimensional isotropic Gaussian (Zhang 2007). This is an approximation to the more rigorous diffraction models given by Richards and Wolf (1959) or Gibson and Lanni (1989). Over a continuous domain, the impulse response reads

$$O(u,v) = \frac{1}{2\pi\sigma_{\mathbf{x}}^2} e^{-\frac{(u-\theta_u)^2+(v-\theta_v)^2}{2\sigma_{\mathbf{x}}^2}}$$

The above expression can be interpreted as a probability distribution over locations where a photon can be detected. Therefore, for discrete detectors, we discretize this expression by integrating over pixels. The number of photon arrivals will follow Poisson statistics, with expected value

$$\omega_k = i_0 \left(\int_{u_k-\delta/2}^{u_k+\delta/2} O(u;\theta_u)du \right) \left(\int_{v_k-\delta/2}^{v_k+\delta/2} O(v;\theta_v)dv \right)$$

The scalar quantity i_0 represents the amplitude of the signal, which is proportional the quantum efficiency of a pixel η, the duration of exposure, Δ, and the number of photons emitter by a fluorescent molecule N_0. With no loss of generality, $\Delta = \eta = 1$ and there is a single free parameter N_0. A simple change of variables $u' = u - \theta_u$ and $v' = v - \theta_v$ gives

$$\omega_k = i_0 \left(\int_{u_k-\delta/2-\theta_u}^{u_k+\delta/2-\theta_u} O(u)du \right) \left(\int_{v_k-\delta/2-\theta_v}^{v_k+\delta/2-\theta_v} O(v)dv \right)$$

One of these terms can be written as

$$\begin{aligned}\int_{u_k-\delta/2-\theta_u}^{u_k+\delta/2-\theta_u} O(u)du &= \int_0^{u_k+\delta/2-\theta_u} O(u)du - \int_0^{u_k-\delta/2-\theta_u} O(u)du \\ &= \frac{1}{2}\left(\operatorname{erf}\left(\frac{u_k + \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) - \operatorname{erf}\left(\frac{u_k - \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) \right)\end{aligned}$$

where we have used the common definition $\operatorname{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. To simplify the notation in the main text, we define a general expression

$$\Delta E_{\theta_i}(\xi_{k,i}, \theta_x, \sigma_{\mathbf{x}}) := \int_{u_k-\delta/2}^{u_k+\delta/2} O(u;\theta_i)du = \frac{1}{2}\left(\operatorname{erf}\left(\frac{\xi_{k,i} + \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) - \operatorname{erf}\left(\frac{\xi_{k,i} - \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) \right)$$

A.2 Metropolis-Hastings MCMC

To obtain numerical estimates of $p(\theta|\mathbf{x} \propto p(\mathbf{x}|\theta)p(\theta)$ and therefore $p(\mathbf{y}|\mathbf{x})$, for an isolated fluorescent molecule as shown in (Figure 2), we used Metropolis-Hastings Markov Chain Monte Carlo (MCMC) to estimate the posterior on coordinates. Under the Poisson approximation in (1), the model negative log-likelihood is

$$\ell(\mathbf{x}|\theta) = -\log \prod_k \frac{e^{-(\omega'_k)} (\omega'_k)^{n_k}}{n_k!} = \sum_k \log n_k! + \omega'_k - n_k \log(\omega'_k) \tag{10}$$

where n_k is the observed number events at a pixel. MCMC is asymptotically exact, which is not guaranteed by variational methods which may rely on a Laplace approximation around the MLE. We

choose a uniform prior $p(\theta)$, and Metropolis-Hastings is run for 10^4 iterations, the first 10^3 iterations are discarded as burn-in. A proposal $\theta' = \theta + \Delta\theta$ was generated with $\Delta\theta \sim \mathcal{N}(0, \sigma^2 I)$ where $\sigma^2 = 0.05$. The acceptance probability is

$$\alpha = e^{\beta(\ell(\theta) - \ell(\theta'))}$$

We choose $\beta = 0.2$ to achieve a target acceptance rate of 0.5.

A.3 Cramer-Rao Lower Bound

Reliable inference of θ from $\mathbf{x}$ in general requires performance metrics for model selection. We use the Fisher information as an information theoretic criteria to assess the quality of the data augmentation model ϕ tested here, 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., $\mathrm{var}(\hat{\theta}) \geq I^{-1}(\theta)$. The Fisher information is straightforward to compute under the Poisson log-likelihood in (1). In general, the Fisher information is given by the expression

$$I_{ij}(\theta) = \mathop{\mathbb{E}}_{\theta}\left(\frac{\partial \ell}{\partial \theta_i}\frac{\partial \ell}{\partial \theta_j}\right) \tag{11}$$

For an arbitrary parameter, such the θ_x or θ_y coordinate, we find that, for a Poisson log-likelihood ℓ

$$\begin{aligned}\frac{\partial \ell}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i}\sum_k \log n_k! + \omega'_k - n_k \log\left(\omega'_k\right) \\ &= \sum_k \frac{\partial \omega'_k}{\partial \theta_i}\left(\frac{\omega'_k - n_k}{\omega'_k}\right)\end{aligned}$$

Using this result, we can compute the Fisher information matrix $I(\theta)$

$$I_{ij}(\theta) = \mathop{\mathbb{E}}_{\theta}\left(\sum_k \frac{\partial \omega'_k}{\partial \theta_i}\frac{\partial \omega'_k}{\partial \theta_j}\left(\frac{\omega'_k - n_k}{\omega'_k}\right)^2\right) = \sum_k \frac{1}{\omega'_k}\frac{\partial \omega'_k}{\partial \theta_i}\frac{\partial \omega'_k}{\partial \theta_j}$$

A fundamental lower bound on the variance in our estimates of θ then is found from its inverse: $\mathrm{CRLB} = I^{-1}(\theta)$. This result is used to show in (Figure 5), that the data augmentation model ϕ efficiently estimates molecular coordinates under the experimental conditions tested here.

A.4 Neural Networks ψ, ϕ

DeepSTORM CNN ϕ. The DeepSTORM CNN, 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 in this paper, 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. Additional details regarding this architecture can be found in the original paper (Nehme 2020). The terminal prediction module contains three additional convolutional blocks for refinement of the upsampled image, followed by an element-wise HardTanh. The architecture is trained using the objective $\mathcal{L}_\phi = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{y}_{0,n} - \hat{\mathbf{y}}_n)^2$.

DDPM ψ. To represent the reverse process, we used a DDPM architecture originally proposed in (Saharia 2021) where the full model is described. We chose a U-Net backbone to have channel multipliers $[1, 2, 4, 8, 8]$ in the downsampling and upsampling paths of the architecture. Parameters

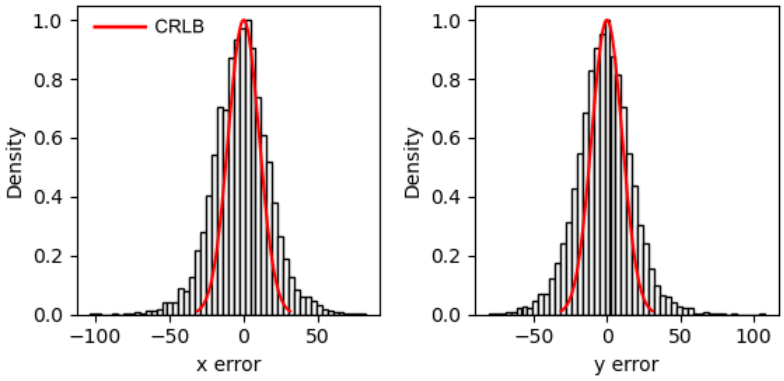

Figure 5: Localization errors of the trained model ϕ. The Cramer-Rao lower bound is shown in red, computing by taking the diagonal elements of $I^{-1}(\theta)$.

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).