
Uncertainty-Aware Localization Microscopy by Variational Diffusion

Anonymous Author(s)

Affiliation

Address

email

Abstract

1 Fast extraction of physically relevant information from images using deep neu-
2 ral networks has led to significant advances in fluorescence microscopy and its
3 application to the study of biological systems. For example, the application of
4 deep networks for kernel density (KD) estimation in single molecule localization
5 microscopy (SMLM) has accelerated super-resolution imaging of densely-labeled
6 structures in the cell. However, simple and interpretable uncertainty quantification
7 is lacking in these applications, and remains a necessary modeling component in
8 high-risk research. We propose a generative modeling framework for KD estima-
9 tion in SMLM based on variational diffusion. This approach allows us to probe the
10 structure of the posterior on KD estimates, creating an additional avenue toward
11 quality control. We demonstrate that data augmentation with traditional SMLM
12 architectures followed by a diffusion process permits simultaneous high-fidelity
13 super-resolution with uncertainty estimation of regressed KDEs.

14 1 Introduction

15 Deep models have attracted tremendous attention from researchers in the natural sciences, with several
16 foundational applications arising in microscopy [Weigert et al., 2018, Falk et al., 2019]. Recently,
17 the application of deep image translation in single-molecule localization microscopy (SMLM) has
18 received considerable interest. SMLM techniques are a mainstay of fluorescence microscopy, which
19 localize “blinking” fluorescent molecules to produce a pointillist representation of biomolecules in
20 the cell at diffraction-unlimited precision [Rust et al., 2006, Betzig et al., 2006]. Recently, the use of
21 deep models to perform localization has been proposed as an alternative to traditional localization
22 algorithms, in order to increase imaging speed and labeling density. In previous applications of deep
23 models to localization microscopy, super-resolution images have been recovered from a sparse set of
24 localizations with conditional generative adversarial networks [Ouyang et al., 2018] or localization
25 itself can be performed using traditional convolutional networks [Nehme et al., 2020, Speiser et al.,
26 2021]. Here, we train a deep model to perform localization indirectly by predicting kernel density
27 (KD) estimates of a population of fluorescent molecules.

28 Kernel density estimation in SMLM is necessarily performed using a single low-resolution image,
29 and thus common measures of model performance are based on localization errors computed over
30 ensembles of simulated images. Unfortunately, this choice precludes computation of uncertainty at
31 test time under a fixed model. Bayesian probability theory is therefore an attractive alternative, which
32 offers us mathematically grounded tools to reason about uncertainty.

33 Here, we address aleatoric uncertainty in KD estimation, by modeling a posterior on high-resolution
34 KD estimates conditioned on a low-resolution image. Our approach is based on a type of score based
35 generative model [Song et al., 2021], referred to as a denoising diffusion probabilistic model (DDPM)
36 in the literature [Ho et al., 2020, Song et al., 2021]. We find that this technique is complementary

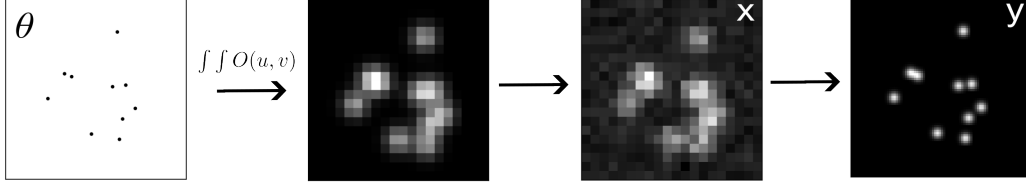


Figure 1: Generative model of single molecule localization microscopy images. Low resolution images \mathbf{x} are generated from coordinates θ by integration of the optical transfer function O and sampling from the likelihood (1): $\mathbf{x} \sim p(\mathbf{x}|\theta) = \prod_k p(\mathbf{x}_k|\theta)$.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

to relevant existing approaches to uncertainty estimation, which would primarily address epistemic sources of uncertainty, using techniques such as ensembling [Lakshminarayanan et al., 2017] or Monte Carlo dropout [Gal et al., 2022].

Our approach is inspired by recent variational perspectives on diffusion [Dirmeier et al., 2023, Ribeiro et al., 2024, Kingma et al., 2021, 2023]. A primary advantage of this approach is that it provides a mechanism for scalable variational inference, which can be trained using a variational bound written in terms of the signal-to-noise ratio of the diffused data, and a simple noise estimation loss. Indeed, recent efforts have shown that the variational bound can be reparameterized to give several more conventional diffusion losses [Kingma et al., 2021, 2023, Ribeiro et al., 2024].

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 KD estimation as an alternative to direct localization using low-resolution images, followed by demonstration of our variational diffusion model for measuring uncertainty KD estimation at scale.

2 Background

2.1 Image Likelihood and Localization Error

The central objective of SMLM is to infer a set of molecular coordinates $\theta = (\theta_u, \theta_v)$ 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 $p(\zeta_k) = \mathcal{N}(o_k, w_k^2)$

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

where A is some normalization constant. 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 et al., 2013]. After subtraction of a known offset o_k of the pixel array, which can be easily measured, we have $\omega'_k = \omega_k + w_k^2$. 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 k of width δ centered at $\xi_k = (u_k, v_k)$, we define, along the first image dimension

$$\Delta E_u := \int_{u_k - \delta/2}^{u_k + \delta/2} O(u; \theta_u) du = \frac{1}{2} \left(\text{erf} \left(\frac{u_k + \frac{\delta}{2} - \theta_u}{\sqrt{2}\sigma_{\mathbf{x}}} \right) - \text{erf} \left(\frac{u_k - \frac{\delta}{2} - \theta_u}{\sqrt{2}\sigma_{\mathbf{x}}} \right) \right) \quad (2)$$

The expected value at each pixel is then $\omega_k \propto \Delta E_u(u_k, \theta_u, \sigma_{\mathbf{x}}) \Delta E_v(v_k, \theta_v, \sigma_{\mathbf{x}})$. Using this, sampling from the convolution distribution in (1) can be carried out by $\mathbf{x}_k = s_k + \zeta_k$ for

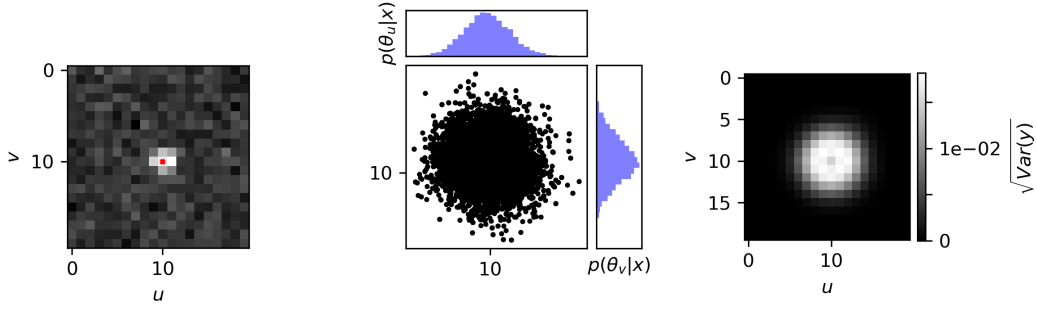


Figure 2: Estimation of the marginal variances $\sqrt{\text{Var}(\mathbf{y}_k)}$ for an isolated fluorescent emitter. MCMC sampling is carried out using the low-resolution image (left) to estimate the posterior $p(\theta|\mathbf{x})$ (middle) which is sampled $\theta \sim p(\theta|\mathbf{x})$ and combined with (3) to estimate marginal variances (right)

69 $s_k \sim \text{Poisson}(\omega_k)$, $\eta_k \sim \mathcal{N}(o_k, w_k^2)$. The complete generative process is depicted in (Figure 1). In
70 this simple case, an estimate of the posterior $p(\theta|\mathbf{x})$ can be found by Markov Chain Monte Carlo
71 (MCMC) sampling as shown in (Figure 2). Details of the MCMC algorithm used here are outlined in
72 the Appendix.

73 2.2 Gaussian kernel density estimation

74 Direct optimization of the likelihood in (1) from observations \mathbf{x} alone is challenging when fluorescent
75 emitters are dense within the field of view and fluorescent signals significantly overlap. However,
76 convolutional neural networks (CNNs) have recently proven to be powerful tools fluorescence
77 microscopy to extract parameters describing fluorescent emitters such as color, emitter orientation,
78 z -coordinate, and background signal Zhang et al. [2018], Kim et al. [2019], Zelger et al. [2018]. For
79 localization tasks, CNNs typically employ upsampling layers to reconstruct Bernoulli probabilities of
80 emitter occupancy [Speiser et al., 2021] or KD estimates with higher resolution than experimental
81 measurements [Nehme et al., 2020]. We choose to use KD estimates in our model, denoted by \mathbf{y} ,
82 which are latent in the low-resolution data \mathbf{x} . KDEs are the most common data structure used in
83 SMLM, and can be easily generated from molecular coordinates, alongside observations \mathbf{x} , using
84 well-understood models of the optical impulse response [Zhang et al., 2007].

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

$$\mathbf{y}_k(\theta) = \sum_{n=1}^N \Delta E_u(u_k, \theta_u, \sigma_{\mathbf{y}}) \Delta E_v(v_k, \theta_v, \sigma_{\mathbf{y}}) \quad (3)$$

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

90 3 Uncertainty-Aware Localization Microscopy by Variational Diffusion

91 We now consider more realistic datasets $(\mathbf{x}_i, \mathbf{y}_{0,i}, \hat{\mathbf{y}}_i)_{i=1}^N$ of observed images \mathbf{x}_i true KD images $\mathbf{y}_{0,i}$,
92 and augmented low-resolution inputs $\hat{\mathbf{y}}_i = \phi(\mathbf{x}_i)$, where ϕ is a CNN. Observations \mathbf{x}_i are simulated
93 under the convolution distribution (1) and KDEs are generated by (4).

94 3.1 Problem Statement

95 Kernel density estimates produced by the traditional deep architectures for localization microscopy
96 produce strong results, but lack uncertainty quantification. Unfortunately, the posterior $p(\theta|\mathbf{x})$ has
97 no known analytical form and can be difficult to compute at test time, since (i) molecules cannot

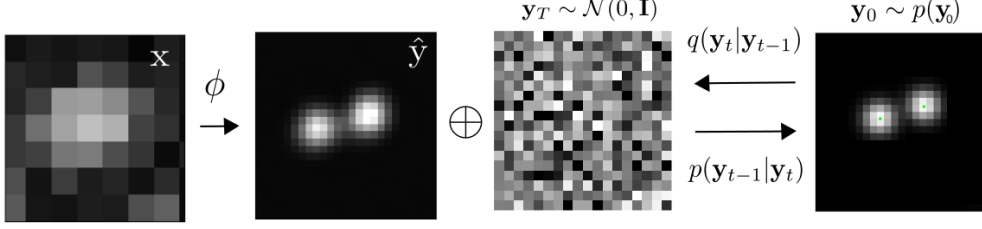


Figure 3: Conditional diffusion model for sampling kernel density estimates. An initial KDE estimate $\hat{\mathbf{y}}$ is found by the CNN ϕ , which is used as conditional input in the DDPM

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 et al., 2015, Ho et al., 2020, Song et al., 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 et al., 2021, Vahdat et al., 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. Therefore, we propose that sampling is preceded by an augmentation 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. Moreover, mixing can be difficult because of the need of traversing low density regions to transition between modes of the distribution [Song et al., 2019].

3.2 Variational Diffusion

Diffusion models [Sohl-Dickstein et al., 2015, Ho et al., 2020, Song et al., 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 the variational interpretation of this model class [Kingma et al., 2021, 2023] to approximate the posterior $p(\mathbf{y}|\mathbf{x})$.

Diffusion Model. We use a forward process which gradually adds Gaussian noise to the latent \mathbf{y}_0 in discrete time, according to a variance schedule β_t :

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

An important property of the forward process is that it admits sampling \mathbf{y}_t at an arbitrary timestep t in closed form [Ho et al., 2020]. Using the notation $\alpha_t := 1 - \beta_t$ and $\gamma_t := \prod_{s=1}^t \alpha_s$, we have $q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}(\sqrt{\gamma_t}\mathbf{y}_0, (1-\gamma_t)I)$ or $\mathbf{y}_t = \sqrt{\gamma_t}\mathbf{y}_0 + \sqrt{1-\gamma_t}\epsilon$ for $\epsilon \sim \mathcal{N}(0, I)$. The signal to noise ratio (SNR) as defined in [Kingma et al., 2023], at a time step t reads $\text{SNR}_t = \gamma_t/(1-\gamma_t)$.

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

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

where we reuse the variance schedule of the forward process [Ho et al., 2020]. From (5) 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.

Learning the reverse process can be approached by either regressing noise ϵ from the forward process, or the true latent \mathbf{y}_0 , as there is a deterministic relationship between them. We adopt the former for consistency with other work, and define ψ as a neural denoising function which regresses the noise ϵ from a noisy \mathbf{y}_t . A relation between the noise estimate ϵ_ψ and μ_ψ is given in the Appendix, which gives an intuition for sampling. The proposed sampling scheme is depicted in (Figure 3).

Variational Objective. Following [Kingma et al., 2021], we interpret the reverse process as 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 variational bound

$$-\log p(\mathbf{y}_0) \leq -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \left(\frac{p_\psi(\mathbf{y}_{1:T}, \mathbf{y}_0)}{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \right) \quad (6)$$

$$= D_{KL}(q(\mathbf{y}_T|\mathbf{y}_0)||p(\mathbf{y}_T)) + \mathbb{E}_{q(\mathbf{y}_1|\mathbf{y}_0)} \log p(\mathbf{y}_0|\mathbf{y}_1) + \mathcal{L}_\psi \quad (7)$$

where we have omitted conditioning on the low-resolution \mathbf{x} to simplify the notation. Note that, this is similar to a hierarchical VAE, but in a diffusion model $q(\mathbf{y}_{1:T}|\mathbf{y}_0)$ is fixed by the forward process rather than learned. The so-called diffusion loss \mathcal{L}_ψ is shown in Appendix A, and is the term of interest as the first two terms do not contribute meaningfully to the loss [Ho et al., 2020]. Furthermore, it has become standard to use simplified forms of \mathcal{L}_ψ , such as a noise estimation loss, as this has shown superior performance. Importantly, \mathcal{L}_ψ is simply a reweighted variant of a family of diffusion objectives [Kingma et al., 2021, 2023]. We use the following Monte Carlo estimate of \mathcal{L}_ψ , which demonstrates that the variational bound can be written in terms of the common noise-estimation loss

$$\mathcal{L}_\psi = \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I), t \sim U(1, T)} \left[\left(\frac{\text{SNR}_{t-1}}{\text{SNR}_t} - 1 \right) \|\epsilon - \epsilon_\psi\|_2^2 \right] \quad (8)$$

A full derivation of this objective is outlined in the Appendix. Note that SNR_t is monotonically decreasing with t , and thus $\frac{\text{SNR}_{t-1}}{\text{SNR}_t} = \frac{\gamma_{t-1}}{\gamma_t} \frac{1-\gamma_t}{1-\gamma_{t-1}} \geq 1$, ensuring $\mathcal{L}_\psi \geq 0$. In this paper, we choose to use a uniformly weighted loss and leave the exploration of the weighted loss to future work.

4 Experiments

All training data consists 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 multiply ω_k by a constant $i_0 = 200$ for experiments for consistency with typical fluorophore emission rates. All KDEs have dimension 80×80 , are scaled between $[0, 1]$, 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 $\approx 27\text{nm}$. 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 $N = 1000$. Objects in the KDE $\hat{\mathbf{y}}_i$ are detected using the Laplacian of Gaussian (LoG) detection algorithm [Kong et al., 2013], which permits more direct comparison of model predictions to the Cramer-Rao lower bound on localization error, compared to other image similarity measures. Localization is carried out from scale-space maxima

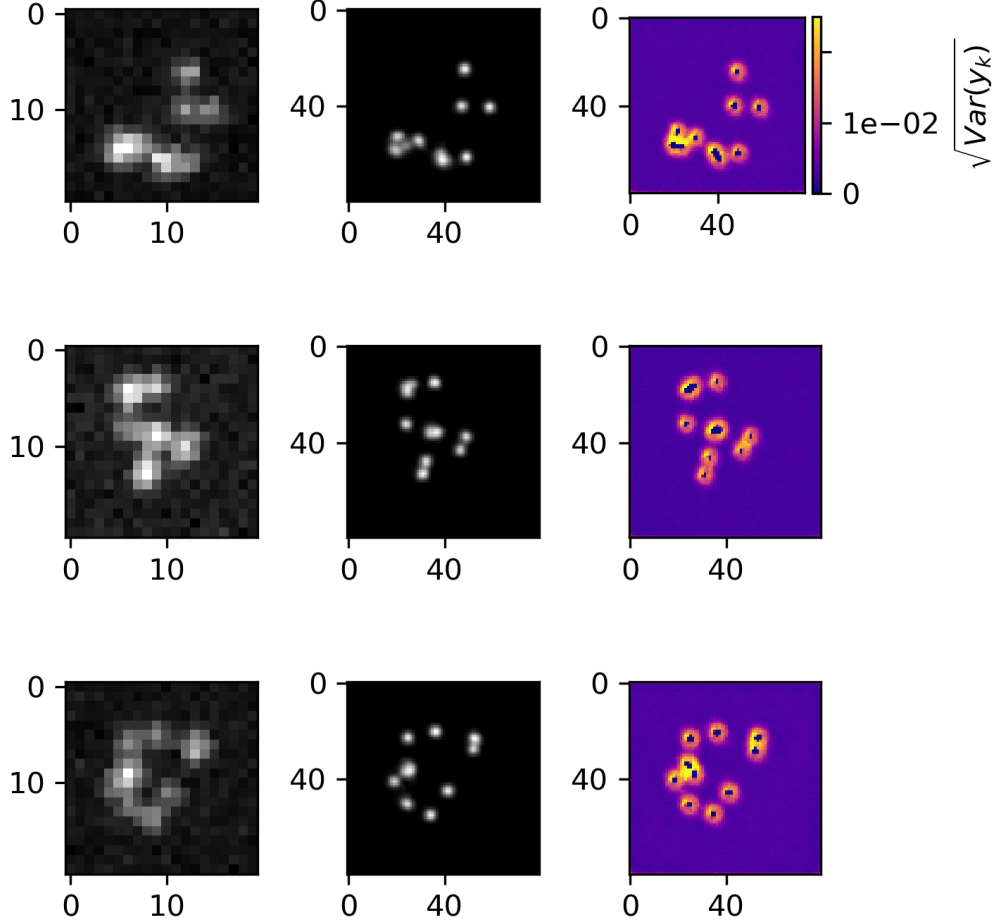


Figure 4: Non cherry-picked estimation of marginal variances. A low-resolution image \mathbf{x} (left column) is transformed by ϕ to produce a KDE estimate $\hat{\mathbf{y}}$ (middle column) and a DDPM ψ computes a map of marginal variances (right column)

174 directly in LoG, as opposed to fitting a model function to KDEs. A particular LoG localization in
 175 the KDE is paired to the nearest ground truth localization and is unpaired if a localization is not
 176 within 5 KDE pixels of any ground truth localization. In addition to localization error, we measure a
 177 precision $P = TP / (TP + FP) = 1.0$ and recall $R = TP / (TP + FN) = 0.85$, where TP denotes
 178 true positive localizations, FP denotes false positive localizations, and FN denotes false negative
 179 localizations.

180 **Variational Diffusion.** We set $T = 100$ for all experiments and treat forward process variances β_t
 181 as hyperparameters, with a linear schedule from $\beta_0 = 10^{-4}$ to $\beta_T = 10^{-2}$. These constants were
 182 chosen to be small relative to ground truth KDEs, which are scaled to $[-1, 1]$, ensuring that forward
 183 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
 184 $t = T$. Example KD estimates from low-resolution images and the marginal variances obtained from
 185 sampling $N = 100$ samples from $p_\psi(\mathbf{y}_0 | \mathbf{x})$ are shown in (Figure 4).

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 KD estimates from low-resolution inputs. This tractable posterior distribution is constructed by first augmenting low resolution inputs to a KD estimate using the DeepSTORM architecture with minor modifications [Nehme et al., 2020]. Conditioning a diffusion model on this initial estimate permits sampling with relatively fewer samples than most existing diffusion models 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 implications beyond its immediate application in SMLM. By leveraging diffusion models 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 interpretation 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 processing tasks highlights the potential to address uncertainty in a wide range of applications in bioimaging or medical imaging. 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.

References

- Eric Betzig et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313: 1642–1645, 2006.
- Jin Chao et al. Fisher information theory for parameter estimation in single molecule microscopy: tutorial. *Journal of the Optical Society of America A*, 2016.
- Simon Dirmeier et al. Diffusion models for probabilistic programming. In *Advances in Neural Information Processing Systems*, 2023.
- Thorsten Falk et al. U-net: deep learning for cell counting, detection, and morphometry. *Nature Methods*, 16:67–70, 2019.
- Yarin Gal et al. Bayesian uncertainty quantification for machine-learned models in physics. *Nature Physics*, 2022.
- Sarah Frisken Gibson and Frederick Lanni. Diffraction by a circular aperture as a model for three-dimensional optical microscopy. *J. Opt. Soc. Am. A*, 6:1357–1367, 1989.
- Jonathan Ho et al. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems*, 2020.
- Fang Huang et al. Video-rate nanoscopy using scmos camera-specific single-molecule localization algorithms. *Nature Methods*, 10:653–658, 2013.
- Tae Hyun Kim et al. Information-rich localization microscopy through machine learning. *Nature Communications*, 10:1996, 2019.

234 Diederik Kingma et al. Variational diffusion models. In *Advances in Neural Information Processing*
235 *Systems*, 2021.

236 Diederik Kingma et al. Understanding diffusion objectives as the elbo with simple data augmentation.
237 In *Advances in Neural Information Processing Systems*, 2023.

238 Hui Kong et al. A generalized laplacian of gaussian filter for blob detection and its applications.
239 *IEEE Transactions on Cybernetics*, 43:1719–1733, 2013.

240 Balaji Lakshminarayanan et al. Simple and scalable predictive uncertainty estimation using deep
241 ensembles, 2017.

242 Elias Nehme et al. Deepstorm3d: Dense 3d localization microscopy and psf design by deep learning.
243 *Nature Methods*, 17:734–740, 2020.

244 Wei Ouyang et al. Deep learning massively accelerates super-resolution localization microscopy.
245 *Nature Biotechnology*, 36:460–468, 2018.

246 Filipe Ribeiro et al. Demystifying variational diffusion models. 2024.

247 Richards and Wolf. Electromagnetic diffraction in optical systems. *Proceedings of the Royal Society*
248 *A*, 253:1358–379, 1959.

249 Michael Rust et al. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy
250 (storm). *Nature Methods*, 3:793–796, 2006.

251 Chinmay Saharia et al. Image super-resolution via iterative refinement. 2021.

252 Carlos Smith et al. Fast, single-molecule localization that achieves theoretically minimum uncertainty.
253 *Nature Methods*, 7:373–375, 2010.

254 Jascha Sohl-Dickstein et al. Deep unsupervised learning using nonequilibrium thermodynamics. In
255 *International Conference on Learning Representations*, 2015.

256 Yang Song et al. Generative modeling by estimating gradients of the data distribution. In *Advances*
257 *in Neural Information Processing Systems*, 2019.

258 Yang Song et al. Score-based generative model through stochastic differential equations. In *International*
259 *Conference on Learning Representations*, 2021.

260 Adam Speiser et al. Deep learning enables fast and dense single-molecule localization with high
261 accuracy. *Nature Methods*, 18:1082–1090, 2021.

262 Arash Vahdat et al. Score-based generative modeling in latent space. In *Advances in Neural*
263 *Information Processing Systems*, 2021.

264 Martin Weigert et al. Content-aware image restoration: pushing the limits of fluorescence microscopy.
265 *Nature Methods*, 15:1090, 2018.

266 Philipp Zelger et al. Three-dimensional localization microscopy using deep learning. *Opt. Express*,
267 26:33166–33179, 2018.

268 Bo Zhang et al. Gaussian approximations of fluorescence microscope point-spread function models.
269 *Applied Optics*, 46:1819–1819–1829, 2007.

270 Peng Zhang et al. Analyzing complex single-molecule emission patterns with deep learning. *Nature*
271 *Methods*, 15:913, 2018.

272 A Appendix

273 A.1 Sampling

274 Sampling from the reverse process $p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x})$ is achieved by estimation of the noise ϵ_ψ from \mathbf{y}_t
 275 by the denoising model ψ , and therefore estimation of \mathbf{y}_0

$$\hat{\mathbf{y}}_0 = \frac{1}{\sqrt{\gamma_t}}(\mathbf{y}_t - \sqrt{1 - \gamma_t}\epsilon_\psi) \quad (9)$$

276 followed by sampling from the forward process $\mathbf{y}_{t-1} \sim q(\mathbf{y}_{t-1}|\hat{\mathbf{y}}_0) = \mathcal{N}(\sqrt{\gamma_{t-1}}, (1 - \gamma_{t-1})I)$.

277 A.2 Derivation of the variational bound

278 We now derive the so-called diffusion loss \mathcal{L}_ψ , written in (8) in the main text. Similar derivations can
 279 be found in [Kingma et al., 2021, Ribeiro et al., 2024], and we include it here only for completeness

$$\begin{aligned} -\log p(\mathbf{y}_0) &\leq -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \frac{p(\mathbf{y}_{0:T})}{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \\ &= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \frac{p(\mathbf{y}_T)p(\mathbf{y}_0|\mathbf{y}_1) \prod_{t=2}^T p(\mathbf{y}_{t-1}|\mathbf{y}_t)}{q(\mathbf{y}_T|\mathbf{y}_0) \prod_{t=2}^T q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)} \\ &= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \left[p(\mathbf{y}_0|\mathbf{y}_1) + \log \frac{p(\mathbf{y}_T)}{q(\mathbf{y}_T|\mathbf{y}_0)} + \sum_{t=2}^T \log \frac{p(\mathbf{y}_{t-1}|\mathbf{y}_t)}{q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)} \right] \\ &= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} [p(\mathbf{y}_0|\mathbf{y}_1)] + D_{KL}(q(\mathbf{y}_T|\mathbf{y}_0)||p(\mathbf{y}_T)) \\ &\quad + \sum_{t=2}^T \mathbb{E}_{q(\mathbf{y}_t|\mathbf{y}_0)} D_{KL}(q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)||p(\mathbf{y}_{t-1}|\mathbf{y}_t)) \end{aligned}$$

280 As before, we omit conditioning on \mathbf{x} to simplify the notation. The first term is typically ignored,
 281 as it does not contribute meaningfully to the loss [Ribeiro et al., 2024]. Furthermore, the second
 282 term is approximately zero by construction. Therefore we are left with the last term, called the
 283 diffusion loss \mathcal{L}_ψ . The KL-divergence of q and p is between two Gaussians with identical variances
 284 $\sigma^2 = \frac{(1-\gamma_{t-1})(1-\alpha_t)}{1-\gamma_t}$, and expectations

$$\mu = \frac{\sqrt{\gamma_{t-1}}(1-\alpha_t)}{1-\gamma_t}\mathbf{y}_0 + \frac{\sqrt{\alpha_t}(1-\gamma_{t-1})}{1-\gamma_t}\mathbf{y}_t \quad \mu_\psi = \frac{\sqrt{\gamma_{t-1}}(1-\alpha_t)}{1-\gamma_t}\hat{\mathbf{y}}_0 + \frac{\sqrt{\alpha_t}(1-\gamma_{t-1})}{1-\gamma_t}\mathbf{y}_t$$

285 for a fixed noise schedule [Saharia et al., 2021]. Therefore, we have

$$\begin{aligned} D_{KL}(q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)||p(\mathbf{y}_{t-1}|\mathbf{y}_t)) &= \frac{1}{2\sigma^2} \|\mu - \mu_\psi\|_2^2 \\ &= \frac{1}{2} \frac{\gamma_{t-1}(1-\alpha_t)}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\ &= \frac{1}{2} \frac{\gamma_{t-1}((1-\gamma_t) - \alpha_t(1-\gamma_{t-1}))}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\ &= \frac{1}{2} \frac{\gamma_{t-1}((1-\gamma_t) - \frac{\gamma_t}{\gamma_{t-1}}(1-\gamma_{t-1}))}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\ &= \frac{1}{2} \left(\frac{\gamma_{t-1}}{1-\gamma_{t-1}} - \frac{\gamma_t}{1-\gamma_t} \right) \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\ &= \frac{1}{2} (\text{SNR}_{t-1} - \text{SNR}_t) \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \end{aligned}$$

286 Reparameterizing the loss in terms of the noise, using $\|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 = \frac{1-\gamma_t}{\gamma_t} \|\epsilon_0 - \epsilon_\psi\|_2^2$ [Ribeiro et al.,
287 2024], we arrive at

$$\mathcal{L}_\psi = \frac{1}{2} \sum_{t=2}^T \mathbb{E}_{q(\mathbf{y}_t|\mathbf{y}_0)} \left(\frac{\text{SNR}_{t-1}}{\text{SNR}_t} - 1 \right) \|\epsilon - \epsilon_\psi\|_2^2$$

288 Using a Monte Carlo estimate of \mathcal{L}_ψ [Kingma et al., 2023] which optimizes random terms of the
289 summation to avoid calculating all terms simultaneously, we arrive at the objective written in the
290 main text (8)

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

291 A.3 Optical impulse response

292 It is common to describe the optical impulse response of a microscope as a two-dimensional isotropic
293 Gaussian [Zhang et al., 2007]. This is an approximation to the more rigorous diffraction models given
294 by [Richards and Wolf, 1959, Gibson and Lanni, 1989]. Over a continuous domain, the impulse
295 response reads

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

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

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

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

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

303 One of these terms can be written as

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

304 where we have used the common definition $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$.

305 A.4 Metropolis-Hastings MCMC

306 To obtain numerical estimates of $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$ and therefore $p(\mathbf{y}|\mathbf{x})$, for the isolated fluo-
307 rescent 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

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

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.5 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 et al., 2016]. The Poisson log-likelihood $\ell(\mathbf{x}|\theta)$ is also convenient for computing the Fisher information matrix [Smith et al., 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 in (1). In general, the Fisher information is given by the expression

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

For an arbitrary parameter, we find that, for a Poisson log-likelihood ℓ

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

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

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

A fundamental lower bound on the variance in our estimates of θ then is found from its inverse: $\text{CRLB} = I^{-1}(\theta)$. 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.6 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 [Nehme et al., 2020] 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

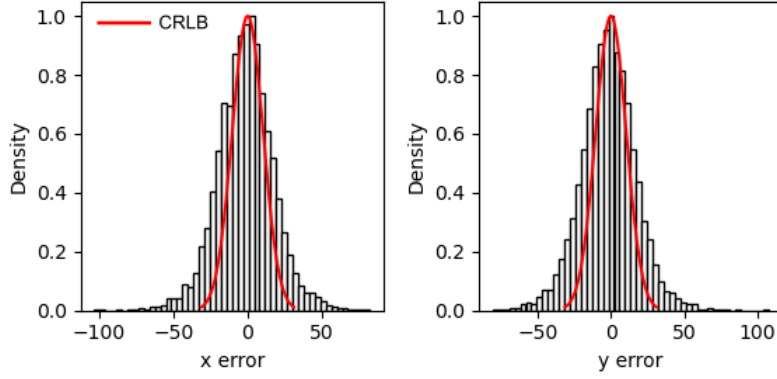


Figure 5: Localization errors of the trained model ϕ . The Cramer-Rao lower bound is shown in red, computing by taking the diagonal elements of $I^{-1}(\theta)$.

337 composed of two consecutive 2x resize-convolutions, computed by nearest-neighbor interpolation,
 338 to increase the lateral resolution by a factor of 4. Additional details regarding this architecture can
 339 be found in the original paper Nehme et al. [2020]. The terminal prediction module contains three
 340 additional convolutional blocks for refinement of the upsampled image, followed by an element-wise
 341 HardTanh. The architecture is trained using the objective $\mathcal{L}_\phi = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_{0,n} - \hat{\mathbf{y}}_n)^2$.

342 **DDPM** ψ . To represent the reverse process, we used a DDPM architecture originally proposed in
 343 [Saharia et al., 2021]. We chose the U-Net backbone to have channel multipliers $[1, 2, 4, 8, 8]$ in
 344 the downsampling and upsampling paths of the architecture. In this architecture, parameters are
 345 shared across time, which is specified to the network using the Transformer sinusoidal position
 346 embedding, and uses self-attention at the 16×16 feature map resolution. To condition the model on
 347 the input $\hat{\mathbf{y}}$, we concatenate the $\hat{\mathbf{y}}$ estimated by DeepSTORM along the channel dimension, which
 348 are scaled to $[0, 1]$, with $\mathbf{y}_T \sim \mathcal{N}(0, I)$. Others have experimented with more sophisticated methods
 349 of conditioning, but found that the simple concatenation yielded similar generation quality [Saharia
 350 et al., 2021].