

---

# Diffusion Probabilistic Models for Super Resolution Microscopy

---

Anonymous Author(s)

Affiliation

Address

email

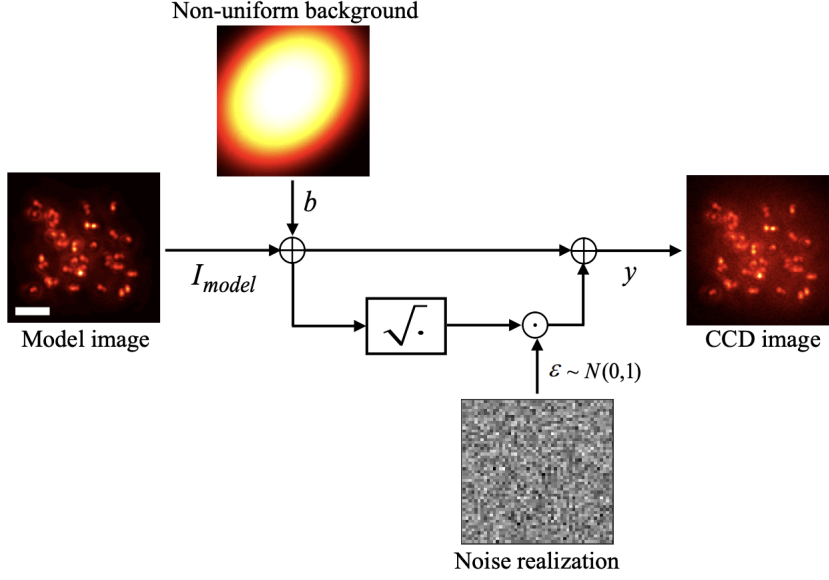
## Abstract

1 Single-molecule localization microscopy (SMLM) techniques are a mainstay of  
2 fluorescence microscopy and can be used to produce a pointillist representation  
3 of living cells at diffraction-unlimited precision. Classical SMLM approaches  
4 leverage the deactivation of fluorescent tags, followed by spontaneous or pho-  
5 toinduced reactivation, which can be used to estimate of the density of a tagged  
6 biomolecule in cellular compartments. Standard SMLM localization algorithms  
7 based on maximum likelihood estimators or least squares optimization require  
8 tight control of activation and reactivation to maintain sparse emitters, present-  
9 ing a tradeoff between imaging speed and labeling density. Deep models have  
10 generalized SMLM to densely labeled structures, yet uncertainty quantification  
11 is still lacking. Recently, denoising diffusion probabilistic models (DDPMs) have  
12 been adapted conditional super resolution tasks, demonstrating promising results  
13 in detail reconstruction, while directly providing uncertainties in model predictions.  
14 Here, we adapt DDPM to the task of single molecule localization, and demonstrate  
15 that DDPM approaches the Cramer-Rao lower bound on localization uncertainty  
16 over a wide range of experimental conditions.

## 17 1 Introduction

18 Single molecule localization microscopy (SMLM) relies on the temporal resolution of fluorophores  
19 whose spatially overlapping point spread functions would otherwise render them unresolvable at the  
20 detector. Common strategies for the temporal separation of molecules involve transient intramolecular  
21 rearrangements to switch from dark to fluorescent states or the exploitation of non-emitting molecular  
22 radicals. Estimation of molecular coordinates in SMLM is achieved by modeling the optical impulse  
23 response of the imaging system. However, dense localization suffers from the curse of dimensionality  
24 - the parameter space volume grows exponentially with the number of molecules, which is often  
25 unknown a priori. Exploration of this high dimensional parameter space in dense SMLM is often  
26 intractable.

27 Previous approaches to this issue has been to predict super-resolution images from a sparse set of  
28 localizations with conditional generative adversarial networks (Ouyang 2018) or direct prediction of  
29 coordinates using deep neural networks (Nehme 2020; Speiser 2021). However, diffusion models are  
30 an appealing alternative because they infer a distribution of deconvolved images that are compatible  
31 with an observation. Although conditional VAEs and conditional GANs can provide a distribution of  
32 deconvolved images, both are known to suffer from mode collapse and produce insufficient diversity  
33 in their outputs. Diffusion models are a recently developed alternative to VAEs and GANs that excel  
34 at producing diverse samples and have been successfully applied to solve inverse problems. Here,  
35 we present a novel diffusion model for deconvolution in single molecule localization microscopy.  
36 The first stage of our algorithm performs interpolation by computing second order coherence of pixel



37 pairs. Subsequent stages cast localization as a conditional image refinement task, realized by a U-Net  
 38 model trained on denoising at various noise levels.

39 Denoising diffusion probabilistic models (DDPM) have emerged as powerful generative models,  
 40 exceeding GANs and VAEs in a variety of generative modeling tasks. Nevertheless, learning diffusion  
 41 models directly in data space can limit expressivity of the model (Vahdat 2021). Therefore, we build  
 42 on previous approaches by using a CNN to compute a latent representation  $\mathbf{z}_i$ . A denoising diffusion  
 43 probabilistic model (DDPM) is then used to model the distribution  $P_{\Phi}(\mathbf{y}|\mathbf{z})$ .

## 44 2 Denoising Diffusion Probabilistic Model for SMLM

45 We consider datasets  $(\theta_i, \mathbf{x}_i, \mathbf{y}_i)_{i=1}^N$  of observed images  $\mathbf{x}_i$  and kernel density estimate (KDE)  
 46 images  $\mathbf{y}_i$ , given an underlying set of object coordinates  $\theta_i$ . Observations  $\mathbf{x}_i$  are generated from  
 47  $\theta_i = (r_1, \dots, r_N)$  under an image degradation model  $F$ . We aim to develop a framework for  
 48 sampling from  $p(\mathbf{y}_i|\mathbf{x}_i)$  and inference of  $\theta_i$ , while fulfilling a resolution criterion under the condition  
 49  $|r_i - r_j| \geq \epsilon; \forall(i, j)$ .

### 50 2.1 Degradation Model

51 The central objective of single molecule localization microscopy is to infer a set of molecular  
 52 coordinates  $\theta$  from noisy, low resolution images  $\mathbf{x}$ . We define an abstract image stochastic degradation  
 53 function  $F$  such that  $\mathbf{x} = F(\theta)$ . In the following paragraphs, we define such a function  $F$ .

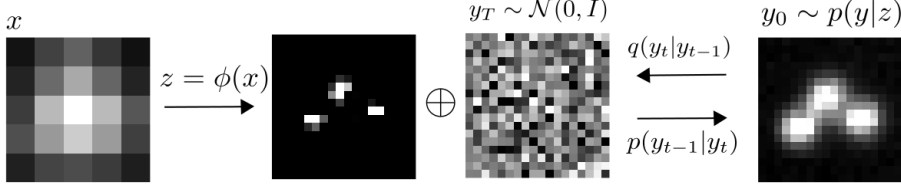
54 In fluorescence microscopy, each pixel follows Poisson statistics, with expected value

$$\omega = i_0 \int O(u) du \int O(v) dv \quad (1)$$

55 where  $i_0 = \eta N_0 \Delta$ . The optical impulse response  $O(u, v)$  is often approximated as a 2D isotropic  
 56 Gaussian with standard deviation  $\sigma$  (Zhang 2007). The parameter  $\eta$  is the photon detection probability  
 57 of the sensor and  $\Delta$  is the exposure time.  $N_0$  represents the number of photons emitted.

58 For a fluorescent emitter located at  $\theta = (u_0, v_0)$ , we have that

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



where we have used the common definition  $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^t e^{-t^2} dt$ . For the sake of generality, the number of photoelectrons at a pixel  $k$ ,  $\mathbf{s}_k$ , is multiplied by a gain factor  $g_k$  [ADU/ $e^-$ ], which is often unity. The readout noise per pixel  $\zeta_k$  can be Gaussian with some pixel-specific offset  $o_k$  and variance  $\sigma_k^2$ . Ultimately, we have a Poisson component of the signal, which scales with  $N_0$  and may have Gaussian component, which does not. Therefore, in a single exposure, we measure:

$$\mathbf{x}_t = \mathbf{s}_t + \zeta \quad (3)$$

What we are after is the likelihood  $p(\mathbf{x}_t | \theta)$  where  $\theta$  are the molecular coordinates. Fundamental probability theory states that the distribution of  $\mathbf{x}_k$  is the convolution of the distributions of  $\mathbf{s}_k$  and  $\zeta_k$ ,

$$p(\mathbf{x}_t | \theta) = A \sum_{q=0}^{\infty} \frac{1}{q!} e^{-\omega_k} \omega_k^q \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(\mathbf{x}_k - g_k \omega_k q - o_k)^2}{2\sigma_k^2}} \quad (4)$$

where  $P(\zeta_k) = \mathcal{N}(o_k, \sigma_k^2)$  and  $P(S_k) = \text{Poisson}(g_k \omega_k)$ ,  $A$  is some normalization constant. In practice, (4) is difficult to work with, so we look for an approximation. We will use a Poisson-Normal approximation for simplification. Consider,

$$\zeta_k - o_k + \sigma_k^2 \sim \mathcal{N}(\sigma_k^2, \sigma_k^2) \approx \text{Poisson}(\sigma_k^2) \quad (5)$$

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

$$\mathbf{x}'_k \sim \text{Poisson}(\omega'_k) \quad (6)$$

where  $\omega'_k = g_k \omega_k + \sigma_k^2$ . This result can be seen from the fact the the convolution of two Poisson distributions is also Poisson. The quality of this approximation will degrade with decreasing signal level, since the Poisson distribution does not retain its Gaussian shape at low expected counts. Nevertheless, the quality of the approximation can be predicted by the Komogonov distance between the convolution distribution (4).

## 2.2 The Information Bottleneck for Localization

Inversion of the degradation function  $F$  is generally intractable, particularly when fluorescent molecules are dense within the field of view. This difficulty arises because the parameter  $\theta$  is typically of large and unknown dimension, rendering maximum likelihood estimation or Markov Chain Monte Carlo sampling computationally difficult. Previous solutions to this problem leverage convolutional neural networks (CNNs) to infer coordinates directly by learning a deterministic image transformation  $F^{-1}$ , which we refer to as a "localization map" (Nehme 2021). Such methods faithfully capture the information content in degraded images; however, such methods apply arbitrary thresholding to the CNN localization map, potentially creating erroneous localizations, and do not permit sampling.

We seek a generative approach, which casts localization as an image restoration problem, where a high resolution kernel density estimate  $\mathbf{y}$  is reconstructed from a low resolution image  $\mathbf{x}$ . Building on previous efforts, we utilize a CNN learns a representation which compresses  $\mathbf{x}$  while preserving the relevant information to the prediction of  $\mathbf{y}$ . We use the Fisher information as the information theoretic criteria (Chao 2016). The generative model (6) is also convenient for computing the Fisher

90 information matrix (Smith 2010) and thus the Cramer-Rao lower bound, which bounds the variance  
 91 of a statistical estimator of  $\theta$ , from below. The Fisher information is

$$\mathcal{I}_{ij}(\theta) = \mathbb{E} \left( \frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j} \right) = \sum_k \frac{1}{\omega'_k} \frac{\partial \omega'_k}{\partial \theta_i} \frac{\partial \omega'_k}{\partial \theta_j} \quad (7)$$

92 where the log-likelihood is  $\ell(\mathbf{x}_t|\theta)$ .

### 93 **3 Image Restoration Model**

#### 94 **3.1 The Encoder Network**

#### 95 **3.2 Optimization of the Encoder Network**

#### 96 **3.3 Conditional Denoising Diffusion Model**

97 Given datasets  $(\theta_i, \mathbf{x}_i, \mathbf{y}_i)_{i=1}^N$  which represent samples drawn from an unknown conditional distribu-  
 98 tion  $p(\mathbf{y}|\mathbf{x})$ . This is a one-to-many mapping in which many target images may be consistent with  
 99 an input image. The conditional DDPM model generates a target image  $y_0$  in  $T$  refinement steps.  
 100 Starting with a pure noise image  $y_T \sim \mathcal{N}(0, I)$ , the model iteratively refines the image through  
 101 successive iterations according to learned conditional transition distributions  $p(y_{t-1}|y_t, x)$  such that  
 102  $y_0 \sim p(\mathbf{y}|\mathbf{x})$

#### 103 **3.4 Gaussian Diffusion Model**

104 The *forward* process is the joint distribution  $p_\theta(\mathbf{y}_{0:T})$ , which is Markovian.

$$q(\mathbf{y}_t|\mathbf{y}_0) = \prod_{t=1}^T q(\mathbf{y}_t|\mathbf{y}_{t-1}) \quad q(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}(\mathbf{y}_{t-1}, \sqrt{\alpha_t}\mathbf{y}_{t-1}, (1 - \alpha_t)I) \quad (8)$$

105 We optimize a denoising model  $f_\theta$  which takes as input an interpolated low-resolution input  $\mathbf{y}$  and a  
 106 noisy input  $\mathbf{y}_T$ .

$$p_\theta(\mathbf{y}_{0:T}) = p_\theta(\mathbf{y}_T) \prod_{t=1}^T p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) \quad p_\theta(\mathbf{y}_{t-1}|\mathbf{y}_t) = \mathcal{N}(\mathbf{y}_{t-1}, \mu_\theta(\mathbf{y}_t, \gamma_t), \sigma_t^2 I) \quad (9)$$

107 where  $\gamma_t = \prod_{i=1}^t \alpha_i$ . Note that the model  $\theta$  is not a function of  $t$ . The mean of the transition density  
 108 reads

$$\mu_\theta(\mathbf{x}_t, \mathbf{y}, \gamma_t) = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{y}_t - \frac{1 - \alpha_t}{\sqrt{1 - \gamma_t}} f_\theta(\mathbf{x}_t, \gamma_t) \right) \quad (10)$$

#### 109 **3.5 Optimization of the Denoising Model**

### 110 **4 Experiments**

