Diffusion Probabilistic Models for Super Resolution Microscopy

Anonymous Author(s)
Affiliation
Address
email

## Abstract

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

1   Introduction

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 transient intramolecular
rearrangements to switch from dark to fluorescent states or the exploitation of non-emitting molecular
radicals. Estimation of molecular coordinates in SMLM is acheived by modeling the optical impulse
response of the imaging system. However, dense localization suffers from the curse of dimensionality
- the parameter space volume grows exponentially with the number of molecules, which is often
unknown a priori. Exploration of this high dimensional parameter space in dense SMLM is often
intractable.

Previous approaches to this issue has been to predict super-resolution images from a sparse set of
localizations with conditional generative adversarial networks (Ouyang 2018) or direct prediction of
coordinates using deep neural networks (Nehme 2020; Speiser 2021). However, diffusion models are
an appealing alternative because they infer a distribution of deconvolved images that are compatible
with an observation. Although conditional VAEs and conditional GANs can provide a distribution of
deconvolved images, both are known to suffer from mode collapse and produce insufficient diversity
in their outputs. Diffusion models are a recently developed alternative to VAEs and GANs that excel
at producing diverse samples and have been successfully applied to solve inverse problems. Here,
we present a novel diffusion model for deconvolution in single molecule localization microscopy.
The first stage of our algorithm performs interpolation by computing second order coherence of pixel
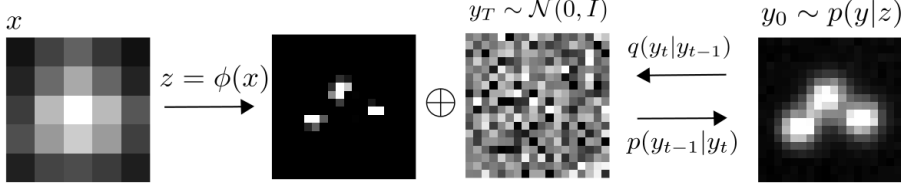
37 pairs. Subsequent stages cast localization as a conditional image refinement task, realized by a U-Net
38 model trained on denoising at various noise levels.

39 This is followed by coordinate refinement by a gradient-based Markov Chain Monte Carlo (MCMC)
40 scheme, known as Langevin dynamics.

2 Background

2.1 Degradation Model

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

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

$$\omega = i_0 \int O(u)du \int O(v)dv \tag{1}$$

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

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

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

51 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
52 number of photoelectrons at a pixel k, $\mathbf{s}_k$, is multiplied by a gain factor g_k [ADU/e^-], which is often
53 unity. The readout noise per pixel ζ_k can be Gaussian with some pixel-specific offset o_k and variance
54 σ_k^2. Ultimately, we have a Poisson component of the signal, which scales with N_0 and may have
55 Gaussian component, which does not. Therefore, in a single exposure, we measure:

$$\mathbf{x}_t = \mathbf{s}_t + \zeta \tag{3}$$

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

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

58 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
59 practice, (4) is difficult to work with, so we look for an approximation. We will use a Poisson-Normal
60 approximation for simplification. Consider,

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

2

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

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

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).

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

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

where the log-likelihood is $\ell(\mathbf{x}_t | \theta)$. We presume a strong model Θ and thus our generative model $P(\mathbf{y}|\mathbf{z}) \approx P(\mathbf{y}|\mathbf{x})$. The CRLB can be used to determine the performance of our model, and thus the quality of the full model (Θ, Φ).

4 Denoising Diffusion Probabilistic Model

Denoising diffusion probabilistic models (DDPM) have emerged as powerful generative models, exceeding GANs and VAEs in a variety of generative modeling tasks. Nevertheless, learning diffusion models directly in data space can limit expressivity of the model (Vahdat 2021). Following this line of work, we build on previous approaches by using a CNN to first compute a localization map, or latent representation $\mathbf{z} = \text{CNN}_\Theta(\mathbf{x})$. A denoising diffusion probabilistic model (DDPM) is then used to model the distribution $P_\Phi(\mathbf{y}|\mathbf{z})$.

Let $\mathbf{y}_0 = \sum_{i=1}^n \omega_n(\sigma)$ be a density estimate of the molecular distribution. The *forward* process is the joint distribution $p_\theta(\mathbf{y}_{0:T})$, which is Markovian.

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

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

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

98 where $\gamma_t = \prod_{i=1}^{t} \alpha_t$. Note that the model θ is not a function of t. The mean of the transition density
99 reads

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

5   Experiments