Supplementary Information

Mitigating Out-of-Focus Noises in Single-Molecule Localization via the Orientation-aware Deep Network

Qifeng Gao, Yutong Li, Zhao Xie, Fei Zhu, Yong Wang, Leiting Pan*, Yuping Duan†

Contents

*Corresponding author: plt@nankai.edu.cn
†Corresponding author: doveduan@gmail.com

1 Model Evaluation

1.1 Euler's elastica regularization

Let $\Omega \subset \mathbb{R}^2$ be an open connected domain. A set $\mathcal{C} \subset \Omega$ is defined as a simple, closed, oriented, smooth curve of class C^2 if it can be obtained as a regular level set of a function $u : \Omega \to \mathbb{R}$ in the following way. Let u be locally Lipschitz on Ω and suppose that there exists an open neighborhood $U \subset \Omega$ of $\mathcal{C}$ such that

$$u \in C^2(U), \qquad |\nabla u(x,y)| > 0 \ \text{ for all } (x,y) \in U,$$

and

$$\mathcal{C} = \big\{(x,y) \in \Omega \mid u(x,y) = 0\big\}. \tag{1}$$

The non-vanishing gradient guarantees that $\mathcal{C}$ is a regular C^2 curve, and the map

$$\mathbf{n}(x,y) = \frac{\nabla u(x,y)}{|\nabla u(x,y)|}, \qquad (x,y) \in U, \tag{2}$$

defines a unit normal vector field of class C^1 on U that coincides with the oriented normal of $\mathcal{C}$ on the curve itself. The (signed) curvature κ of $\mathcal{C}$ is defined pointwise as the Euclidean divergence of $\mathbf{n}$ restricted to the curve:

$$\kappa(p) = \big(\nabla \cdot \mathbf{n}\big)(p), \qquad p \in \mathcal{C}. \tag{3}$$

In terms of the defining function u, (3) reads

$$\kappa = \nabla \cdot \Big(\frac{\nabla u}{|\nabla u|}\Big) \quad \text{on } \mathcal{C}. \tag{4}$$

Because $\mathbf{n}$ is a unit field, its divergence on $\mathcal{C}$ coincides with the tangential derivative of the normal direction; this is precisely the classical notion of signed curvature. A direct computation of the divergence in (4) yields the explicit formula

$$\kappa = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}} \qquad \text{on } \mathcal{C}, \tag{5}$$

whose right-hand side is independent of the particular choice of the defining function u as long as $|\nabla u| > 0$ near $\mathcal{C}$.

Let $\alpha, \beta > 0$ be fixed constants. For a C^2 curve $\mathcal{C}$ as defined above, the Euler's elastica energy is the functional

$$E(\mathcal{C}) = \int_{\mathcal{C}} \big(\alpha + \beta\, \kappa^2\big)\, ds, \tag{6}$$

where ds denotes the arc-length element on $\mathcal{C}$. When $\mathcal{C}$ is represented implicitly via the function u satisfying (1), we may rewrite (6) as an integral over the whole domain Ω by means of the co-area formula. For a.e. $t \in \mathbb{R}$ and any continuous function f

$$\int_{\{u=t\}} f(x,y)\, ds = \int_{\Omega} f(x,y)\, |\nabla u(x,y)|\, \delta\big(u(x,y) - t\big)\, dx\, dy, \tag{7}$$

where δ is the one-dimensional Dirac distribution, i.e., $\delta(z)$ is defined by its action on test functions: $\int_{\mathbb{R}} \varphi(z)\delta(z)\, dz = \varphi(0)$. Taking $t = 0$ and $f(x,y) = \alpha + \beta\kappa^2$ in (7) gives the level-set formulation of the Euler's elastica energy:

$$E(u) = \int_{\Omega} \Big[\alpha + \beta\Big(\nabla \cdot \big(\tfrac{\nabla u}{|\nabla u|}\big)\Big)^2\Big]\, |\nabla u(x,y)|\, \delta\big(u(x,y)\big)\, dx\, dy, \tag{8}$$

where the divergence term $\nabla \cdot \big(\frac{\nabla u}{|\nabla u|}\big)$ is understood as the pointwise function on U defined by (4) and its restriction to $\mathcal{C}$ coincides with the geometric curvature κ.

1.2 Directional convolution

We introduce a learnable directional convolution layer to better capture the anisotropic spatial structure of point-spread functions (PSFs) in imaging data. The layer acquires adaptive directional sensitivity via parameter learning and enforced sparsity. Since the shape of a PSF often varies significantly along the optical axis, the learnable weights enable the filter to adjust its directional selectivity dynamically, thereby reflecting the PSF's anisotropic geometry more faithfully. Moreover, because real imaging data contain varying levels of out-of-focus noise, the proposed convolution learns robust feature representations across different noise conditions, which helps reduce localization errors in complex imaging scenarios.

Let $x : \mathbb{Z}^2 \to \mathbb{R}$ denote a discrete image. For a centre pixel $p_0 = (0, 0)$, its 3×3 neighbourhood is indexed by

$$\mathcal{N} = \{p_1, \ldots, p_8\} = \{(-1,-1), (0,-1), (1,-1), (-1,0), (1,0), (-1,1), (0,1), (1,1)\}.$$

For each neighbour $p_n \in \mathcal{N}$, we define the corresponding unit direction vector

$$\mathbf{v}_n = \frac{p_n - p_0}{\|p_n - p_0\|_2},$$

and the Euclidean distance $d_n = \|p_n - p_0\|_2$, which takes values in $\{1, \sqrt{2}\}$. A discrete approximation of the directional derivative along $\mathbf{v}_n$ at p_0 is given by the finite difference

$$g_n = \frac{x(p_n) - x(p_0)}{d_n}.$$

We construct a generalised gradient operator that combines these directional differences in a learnable way. We introduce per-direction parameters $a_n \in \mathbb{R}$ (amplitude) and $\delta_n \in \mathbb{R}^+$ (distance scaling), and define

$$y(p_0) = \sum_{n=1}^{8} \frac{a_n}{\delta_n} \left[x(p_n) - x(p_0)\right].$$

To guarantee a zero response in any constant region, i.e., whenever $x(p_n) \equiv x(p_0)$, we impose the zero-sum constraint

$$\sum_{n=1}^{8} \frac{a_n}{\delta_n} = 0,$$

which ensures translational invariance and prevents an output bias. The operation above can be rewritten as a standard convolution with a 3×3 kernel $\mathcal{F}$:

$$y(p_0) = \sum_{k \in \mathcal{N} \cup \{p_0\}} \mathcal{F}(k)\, x(p_0 + k),$$

where the kernel entries are

$$\mathcal{F}(k) = \begin{cases} \dfrac{a_n}{\delta_n}, & k = p_n \in \mathcal{N}, \\ -\displaystyle\sum_{n=1}^{8} \dfrac{a_n}{\delta_n}, & k = p_0. \end{cases}$$

We enforce sparsity constraints on the pairs $\{a_n, \delta_n\}$ to enhance directional selectivity and reduce parameter redundancy. This yields two specialised kernels: $\mathcal{F}_x$, sensitive to horizontal and diagonal directions while

suppressing vertical neighbours, and $\mathcal{F}_y$, sensitive to vertical and diagonal directions while suppressing horizontal neighbours. The sparsity pattern is determined by the index sets

$$\mathcal{R}_x = \{p_1, p_3, p_4, p_5, p_6, p_8\} \quad \text{and} \quad \mathcal{R}_y = \{p_1, p_2, p_3, p_6, p_7, p_8\}.$$

The horizontal directional response is therefore

$$y_x(p_0) = \sum_{n \in \mathcal{R}_x} \frac{a_n}{\delta_n} \big[x(p_n) - x(p_0)\big],$$

which is implemented by the convolution

$$y_x(p_0) = \sum_{k \in \mathcal{N} \cup \{p_0\}} \mathcal{F}_x(k)\; x(p_0 + k),$$

with the kernel $\mathcal{F}_x$ given by

$$\mathcal{F}_x(k) = \begin{cases} \dfrac{a_n}{\delta_n}, & k = p_n \in \mathcal{R}_x, \\ 0, & k = p_n \notin \mathcal{R}_x, \\ -\displaystyle\sum_{n \in \mathcal{R}_x} \dfrac{a_n}{\delta_n}, & k = p_0. \end{cases}$$

Arranging these values on a 3×3 grid yields the kernel matrix

$$\mathcal{F}_x = \begin{pmatrix} \theta_1 & 0 & \theta_3 \\ \theta_4 & -\sum_{n \in \mathcal{R}_x} \theta_n & \theta_5 \\ \theta_6 & 0 & \theta_8 \end{pmatrix} \qquad \text{with } \theta_n := \frac{a_n}{\delta_n}.$$

Analogously, the vertical directional kernel is

$$\mathcal{F}_y = \begin{pmatrix} \theta_1 & \theta_2 & \theta_3 \\ 0 & -\sum_{n \in \mathcal{R}_y} \theta_n & 0 \\ \theta_6 & \theta_7 & \theta_8 \end{pmatrix}.$$

The proposed framework generalises several classical edge-detection operators as special cases, obtained by fixing the ratios $\theta_n = a_n/\delta_n$ appropriately. The specific choices that recover the Sobel, Prewitt, Frei–Chen, Roberts-cross, and Scharr operators are summarised in Table 1. Consequently, this learnable directional convolution not only generalises traditional fixed filters, but also provides a flexible mechanism for adapting the directional response to the particular anisotropy present in the PSF of a given imaging system.

1.3 Alternating update mechanism

The localization loss $\mathcal{L}_{\text{loc}}$ is formulated within a mixture-model framework and optimized via an alternating scheme that decouples emitter localization from uncertainty calibration. Let $\{\hat{p}_k, \hat{\mu}_k, \hat{\sigma}_k\}_{k=1}^K$ be the learnable mixture weights, predicted locations, and initial uncertainty estimates for K potential emitters. The ground-truth data consist of E emitter locations μ_e^* and associated uncertainties σ_e^*. The objective function is defined as a combination of a negative log-likelihood term for localization and a regularization term for uncertainty calibration:

$$\mathcal{L}_{\text{loc}} = -\frac{1}{E} \sum_{e=1}^{E} \left(\log\Big(\sum_{k=1}^{K} \pi_k \; \mathcal{P}\big(\mu_e^* \mid \hat{\mu}_k, \hat{\sigma}_k^2\big) \Big) + \lambda \log \sum_{k=1}^{K} \Big(\tilde{\pi}_k \; \mathcal{P}\big(\sigma_e^* \mid \hat{\sigma}_k, w_k^2\big) \Big) \right), \tag{1}$$

Table 1: Classical edge-detection operators as instances of the learnable directional convolution.

Operator	**Choice of** $\theta_n = a_n/\delta_n$
Sobel	$\theta_n = \begin{cases} 1, & p_n \in \{p_1, p_3, p_6, p_8\}, \\ 2, & p_n \in \{p_4, p_5\}, \\ 0, & \text{otherwise.} \end{cases}$
Prewitt	$\theta_n = \begin{cases} 1, & p_n \in \mathcal{R}_x \text{ (or } \mathcal{R}_y), \\ 0, & \text{otherwise.} \end{cases}$
Frei–Chen	$\theta_n = \dfrac{1}{d_n}, \quad a_n = 1,\ \delta_n = d_n.$
Roberts cross	$\theta_n = \begin{cases} 1, & p_n \in \{p_1, p_8\}, \\ 0, & \text{otherwise.} \end{cases}$
Scharr	$\theta_n = \begin{cases} 3, & p_n \in \{p_1, p_3, p_6, p_8\}, \\ 10, & p_n \in \{p_4, p_5\}, \\ 0, & \text{otherwise.} \end{cases}$

where $\pi_k = \dfrac{\hat{p}_k}{\sum_{j=1}^{K} \hat{p}_j}$ are the normalized mixture weights, $\tilde{\pi}_k = \dfrac{p_k}{\sum_{j=1}^{K} p_j}$ are the detached normalized weights with $p_k = \mathrm{detach}(\hat{p}_k)$, i.e., p_k carries the current value of $\hat{p}_k$ but does not propagate gradients, $\mathcal{P}(\cdot \mid \mu, \sigma^2)$ denotes the Gaussian probability density function, and $w_k > 0$ are auxiliary learnable scale parameters that model the deviation between predicted and true uncertainties. The optimization proceeds in two alternating stages, each focusing on a distinct component of the loss.

Stage 1: Localization update. This stage minimizes the negative log-likelihood term

$$\mathcal{L}_A = -\frac{1}{E} \sum_{e=1}^{E} \log\Big(\sum_{k=1}^{K} \pi_k \, \mathcal{P}\big(\mu_e^* \mid \hat{\mu}_k, \hat{\sigma}_k^2\big) \Big),$$

the parameters of which are updated via gradient descent:

$$\hat{p}_k^{(t+1)} = \hat{p}_k^{(t)} - \eta_A \frac{\partial \mathcal{L}_A}{\partial \hat{p}_k^{(t)}},$$
$$\hat{\mu}_k^{(t+1)} = \hat{\mu}_k^{(t)} - \eta_A \frac{\partial \mathcal{L}_A}{\partial \hat{\mu}_k^{(t)}},$$
$$\hat{\sigma}_k^{(t+1)} = \hat{\sigma}_k^{(t)} - \eta_A \frac{\partial \mathcal{L}_A}{\partial \hat{\sigma}_k^{(t)}}.$$

Maximizing the Gaussian likelihood at the ground-truth location μ_e^* encourages precise spatial predictions. A smaller variance $\hat{\sigma}_k^2$ produces a sharper density peak, which increases the likelihood contribution and thereby drives the model to refine both the location estimates $\hat{\mu}_k$ and the associated uncertainties $\hat{\sigma}_k$ for high-confidence emitters.

Stage 2: Uncertainty calibration. This stage minimizes the regularization term

$$\mathcal{L}_B = -\frac{\lambda}{E}\sum_{e=1}^{E}\log\sum_{k=1}^{K}\Big(\tilde{\pi}_k\,\mathcal{P}\big(\sigma_e^* \mid \hat{\sigma}_k, w_k^2\big)\Big),$$

which can be interpreted as performing maximum-likelihood estimation for the parameters $\hat{\sigma}_k$ and w_k under a Gaussian prior whose mean is the true uncertainty σ_e^*. Crucially, the mixture weights $\tilde{\pi}_k$ are detached from the computational graph; therefore, gradients of $\mathcal{L}_B$ do not propagate back to the original weights $\hat{p}_k$. The updates are:

$$\hat{\sigma}_k^{(t+1)} = \hat{\sigma}_k^{(t)} - \eta_B\,\frac{\partial\mathcal{L}_B}{\partial\hat{\sigma}_k^{(t)}},$$
$$w_k^{(t+1)} = w_k^{(t)} - \eta_B\,\frac{\partial\mathcal{L}_B}{\partial w_k^{(t)}},$$
$$p_k^{(t+1)} = \mathrm{detach}\big(\hat{p}_k^{(t+1)}\big).$$

This decoupling ensures that the calibration of uncertainty estimates does not interfere with the learned emitter-existence probabilities.

Iterative refinement. The alternating scheme establishes a mutually reinforcing feedback loop. The updated variance $\hat{\sigma}_k$ from Stage 1 is passed to Stage 2, where it is calibrated against the true uncertainty σ_e^*. The refined $\hat{\sigma}_k$ then re-enters Stage 1, leading to a more accurate probability density and consequently to improved estimates of $\hat{p}_k$ and $\hat{\mu}_k$. This iterative decoupling prevents conflicting gradient signals and jointly enhances localization accuracy and uncertainty reliability.

1.4 Differentiable non-maximum suppression

To enable end-to-end training of the single-molecule localization pipeline, we introduce a differentiable non-maximum suppression (NMS) algorithm designed to operate robustly across both low- and high-density imaging conditions. Instead of relying on fixed global thresholds, our method employs a lightweight post-processing network that predicts pixel-wise threshold maps $t_{\text{low}}, t_{\text{high}}, t_{\text{cum}} \in \mathbb{R}^{H\times W}$, each corresponding to a distinct detection criterion adapted to local signal density. In low-density regions, a relatively high threshold t_{low} is applied to select isolated, high-confidence candidates. A soft-selection mechanism is then used to suppress duplicate detections within local neighbourhoods. In high-density regions, where overlapping signals are frequent, a lower threshold t_{high} allows the detection of weaker emitters. An additional condition requires that the sum of responses from the centre pixel and its four-connected neighbours exceeds a cumulative threshold t_{cum} to reduce false positives. The entire procedure is made differentiable by replacing discrete operations (thresholding, local-max selection, summation) with continuous relaxations, thereby enabling gradient-based optimization through the NMS module.

Differentiable approximations. Hard thresholding is relaxed using a scaled sigmoid function

$$\mathrm{soft\_threshold}(p, t) = \sigma\big(s\cdot(p-t)\big),$$

where p is a pixel-wise probability (or intensity) value, t is the corresponding learned threshold, $\sigma(z) = (1+e^{-z})^{-1}$ is the logistic sigmoid, and $s > 0$ is a sharpness parameter. Larger s makes the transition steeper, approaching a step function.

Discrete local-maxima detection is replaced by a soft-attention mechanism over a 3×3 neighbourhood. Let $N_3(k)$ denote the set of pixel indices in the 3×3 window centred at pixel k. Define the local-max score as

$$\text{soft\_local\_max}(p)_k = \frac{\exp\big((p_k - \max_{i \in N_3(k)} p_i)/\tau\big)}{\sum_{j \in N_3(k)} \exp\big((p_j - \max_{i \in N_3(j)} p_i)/\tau\big)},$$

where $\tau > 0$ is a temperature parameter. As $\tau \to 0$ the distribution approaches a one-hot vector at the strict local maximum, while larger τ yields a smoother assignment of prominence. Neighbourhood aggregation for the high-density path is implemented via a fixed, normalized convolution with the four-connected stencil

$$K = \frac{1}{5} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Thus the aggregated response at pixel k is

$$\text{neighbour\_sum}(p)_k = \sum_{(i,j) \in N_3(k)} K_{i-k,\, j-k}\; p_{ij}.$$

Complete differentiable pipeline. The overall procedure, which combines the low-density and high-density detection paths, is summarised in Algorithm 1. The inputs are a batch of probability maps $p \in \mathbb{R}^{B \times H \times W}$ and the learned threshold maps $t_{\text{low}}, t_{\text{high}}, t_{\text{cum}} \in \mathbb{R}^{B \times H \times W}$. All operations are applied independently per batch element and per spatial location.

Algorithm 1 Differentiable non-maximum suppression

Require: Probability map $p \in \mathbb{R}^{B \times H \times W}$; learned thresholds $t_{\text{low}}, t_{\text{high}}, t_{\text{cum}} \in \mathbb{R}^{B \times H \times W}$; sharpness $s > 0$, temperature $\tau > 0$.

Low-density detection

$m_{\text{low}} \leftarrow \text{soft\_threshold}(p, t_{\text{low}})$ ▷ soft mask after high threshold

$a_{\text{low}} \leftarrow \text{soft\_local\_max}(p)$ ▷ soft local-max assignment

$p_{\text{low}} \leftarrow m_{\text{low}} \odot a_{\text{low}}$ ▷ low-path activation

High-density detection

$m_{\text{high}} \leftarrow \text{soft\_threshold}(p, t_{\text{high}})$ ▷ soft mask after low threshold

$s_{\text{sum}} \leftarrow \text{neighbour\_sum}(p)$ ▷ aggregated neighbourhood response

$m_{\text{cum}} \leftarrow \text{soft\_threshold}(s_{\text{sum}}, t_{\text{cum}})$ ▷ soft cumulative mask

$p_{\text{high}} \leftarrow m_{\text{high}} \odot m_{\text{cum}}$ ▷ high-path activation

Fusion

$p_{\text{out}} \leftarrow \max(p_{\text{low}}, p_{\text{high}})$ ▷ element-wise maximum

return $p_{\text{out}} \in \mathbb{R}^{B \times H \times W}$

The output p_{out} is a soft activation map that highlights pixel-wise detection confidences while being fully differentiable with respect to the input probability map p and the learned thresholds $t_{\text{low}}, t_{\text{high}}, t_{\text{cum}}$. This design allows the NMS module to be integrated seamlessly into an end-to-end trainable localization network.

2 Numerical Implementation

2.1 Simulation data generation

To quantitatively evaluate the performance of ORIENCODE on SMLM data with out-of-focus noise, we simulated two parallel fluorescent lines and a hollow rod, both randomly sampled within a three-dimensional space of $6.4 \times 6.4 \times 1.0\ \mu\text{m}^3$. The separation between the parallel lines was set to 100 nm, 80 nm, 60 nm, and 40 nm, corresponding to a pixel size of 100 nm. The axial (z-axis) sampling range for the main structures was defined as $[-500, 500]$ nm. The hollow rod had a radius of 60 nm and a length of 4400 nm. For each molecule, two out-of-focus emitters were introduced in the surrounding region. The in-plane (xy) coordinates of these out-of-focus emitters were randomly sampled within an annular ring located at 0.3–1 full width at half maximum (FWHM) from the central molecule, while their axial positions were selected from the union of the intervals $[-1500,\ -500] \cup [500,\ 1500]$ nm.

To emulate realistic data acquisition, we modelled the noise characteristics of an EMCCD camera. The imaging process involves several sources of noise that must be considered for accurate signal interpretation. The first originates from photon counting, which follows a Poisson process. The expected number of electrons generated in pixel k can be approximated as

$$\lambda_k = \lambda_{0,k} \cdot QE + b, \tag{2}$$

where $\lambda_{0,k}$ is the average number of photons arriving at the pixel, QE is the quantum efficiency, and b represents a background term such as dark current. The second noise source comes from the electron multiplication register, a key feature of EMCCD sensors. This amplification process introduces additional variability and can be approximated using a Gamma-like distribution:

$$\rho(x \mid s_k, \theta) \propto x^{s_k - 1} e^{-x/\theta}, \tag{3}$$

where s_k is the number of incoming electrons and θ is the effective gain parameter. Finally, read noise is added during signal digitization. This is typically modeled as a zero-mean Gaussian noise with a constant standard deviation σ, and can be written as

$$\rho(x) \sim \mathcal{N}(0, \sigma^2). \tag{4}$$

Before image analysis, the raw signal was calibrated by subtracting the baseline offset and normalizing by the total gain, which includes both the analog gain g and the electron multiplying gain θ.

2.2 Training data workflow

Stage 1: Emitter parameter initialization. First, in-focus emitters are generated. An emitter is considered in-focus when its axial displacement satisfies $|\Delta z_k| \leq 500$ nm. For such emitters we sample

$$\begin{aligned} (\Delta x_k, \Delta y_k) &\sim \mathrm{U}\big([-0.5,\ 0.5]^2\big) \quad \text{(pixel units)}, \\ \Delta z_k &\sim \mathrm{U}\big([-500,\ 500]\big) \ \text{nm}, \\ I_k &\sim \mathcal{N}(\mu_d, \sigma_d^2), \end{aligned}$$

where μ_d, σ_d^2 are fluorophore-specific intensity parameters. Next, out-of-focus emitters ($500\,\text{nm} < |\Delta z_k| < 1500\,\text{nm}$) are created as follows. Let $\rho \in [0, 1]$ denote the fraction of in-focus emitters that are selected as reference emitters. For each reference emitter we generate n out-of-focus companions. The lateral position of a companion is obtained by adding a random radial offset to the reference position:

$$\begin{pmatrix} \Delta x_k^{\text{out}} \\ \Delta y_k^{\text{out}} \end{pmatrix} = \begin{pmatrix} \Delta x_k^{\text{ref}} \\ \Delta y_k^{\text{ref}} \end{pmatrix} + R \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \qquad R \sim \mathrm{U}\big([\text{FWHM},\ 2\times\text{FWHM}]\big),\ \ \theta \sim \mathrm{U}\big([0,\ 2\pi]\big),$$

where FWHM is the full-width at half-maximum of the in-focus PSF (in pixel units). The axial displacement of the companion is sampled uniformly from the two out-of-focus intervals:

$$\Delta z_k \sim \mathrm{U}\big([-1500,\,-500] \cup [500,\,1500]\big)\ \text{nm},$$

and its intensity is drawn from the same fluorophore-specific distribution $I_k \sim \mathcal{N}(\mu_d, \sigma_d^2)$.

Stage 2: Optical system modeling. The optical image is formed by convolving the point-source distribution with depth-dependent PSF kernels. Let $\delta(x - x_k, y - y_k)$ denote a Dirac delta centred at the k-th emitter location, and let $\mathrm{PSF}(\cdot; z_k)$ be the point-spread function at depth z_k. The noise-free optical intensity at spatial coordinates (x, y) is

$$F_{\text{opt}}(x, y) = \sum_{k=1}^{K} I_k \left[\delta(x - x_k, y - y_k) * \mathrm{PSF}(x, y; z_k)\right],$$

where $*$ denotes two-dimensional convolution. In practice the PSF is represented by a discrete kernel whose shape varies smoothly with z_k.

Stage 3: Detector noise simulation. To emulate a real camera, we add photon-shot noise and readout noise. The expected photon count at pixel (i, j) is

$$\lambda_{ij} = F_{\text{opt}}(i, j) + \mu_{\text{bg}},$$

where μ_{bg} is the mean background intensity (dark current plus ambient light). Photon statistics are modelled by a Poisson distribution, and readout noise by additive Gaussian noise. Hence the final simulated pixel value is

$$F_{\text{final}}(i, j) \;=\; X_{ij} + \varepsilon_{ij}, \qquad X_{ij} \sim \mathrm{Poisson}(\lambda_{ij}),\;\; \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\text{read}}^2),$$

where X_{ij} and ε_{ij} are independent random variables. The standard deviation σ_{read} quantifies the readout-circuit noise.

2.3 Evaluation metrics

To evaluate the performance of ORIENCODE on simulated data containing out-of-focus noise, we employed two widely used metrics to assess detection accuracy and localization precision: the Jaccard Index (JI) and the Root Mean Square Error (RMSE).

The Jaccard Index (JI) measures detection accuracy by measuring the overlap between the set of detected localizations and the set of ground truth emitters. It is defined as:

$$\text{JI} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \tag{5}$$

where TP, FP, and FN denote the number of true positive, false positive, and false negative detections, respectively. For this classification, a detected localization is matched to a ground truth emitter and considered a true positive (TP) if its lateral displacement is less than 250 nm and its axial displacement is less than 500 nm.

In contrast, the Root Mean Square Error (RMSE) quantifies the localization precision for correctly identified emitters. It is calculated for all matched localizations (TPs) as the root mean square of the Euclidean distances between estimated and true positions:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\big((\hat{x}_i - x_i^{\text{GT}})^2 + (\hat{y}_i - y_i^{\text{GT}})^2 + (\hat{z}_i - z_i^{\text{GT}})^2\big)}, \tag{6}$$

where N is the number of true positive matches (TP count), $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$ are the estimated coordinates of the i-th matched emitter, and $(x_i^{\mathrm{GT}}, y_i^{\mathrm{GT}}, z_i^{\mathrm{GT}})$ are its corresponding ground truth coordinates.

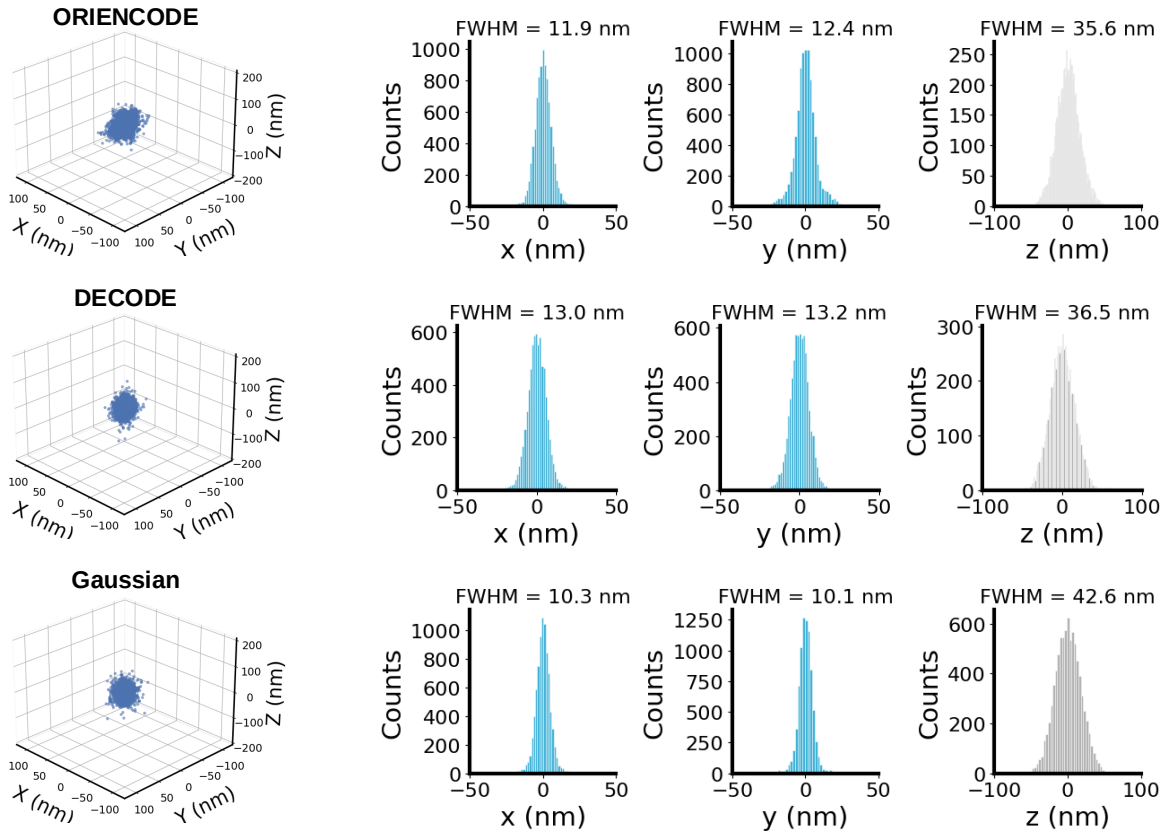
Furthermore, to provide a more detailed analysis, we decomposed the overall RMSE into its lateral and axial components. The lateral Root Mean Square Error ($\mathrm{RMSE}_{\mathrm{lat}}$) evaluates the localization precision in the xy-plane:

$$\mathrm{RMSE}_{\mathrm{lat}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left((\hat{x}_i - x_i^{\mathrm{GT}})^2 + (\hat{y}_i - y_i^{\mathrm{GT}})^2\right)}. \tag{7}$$

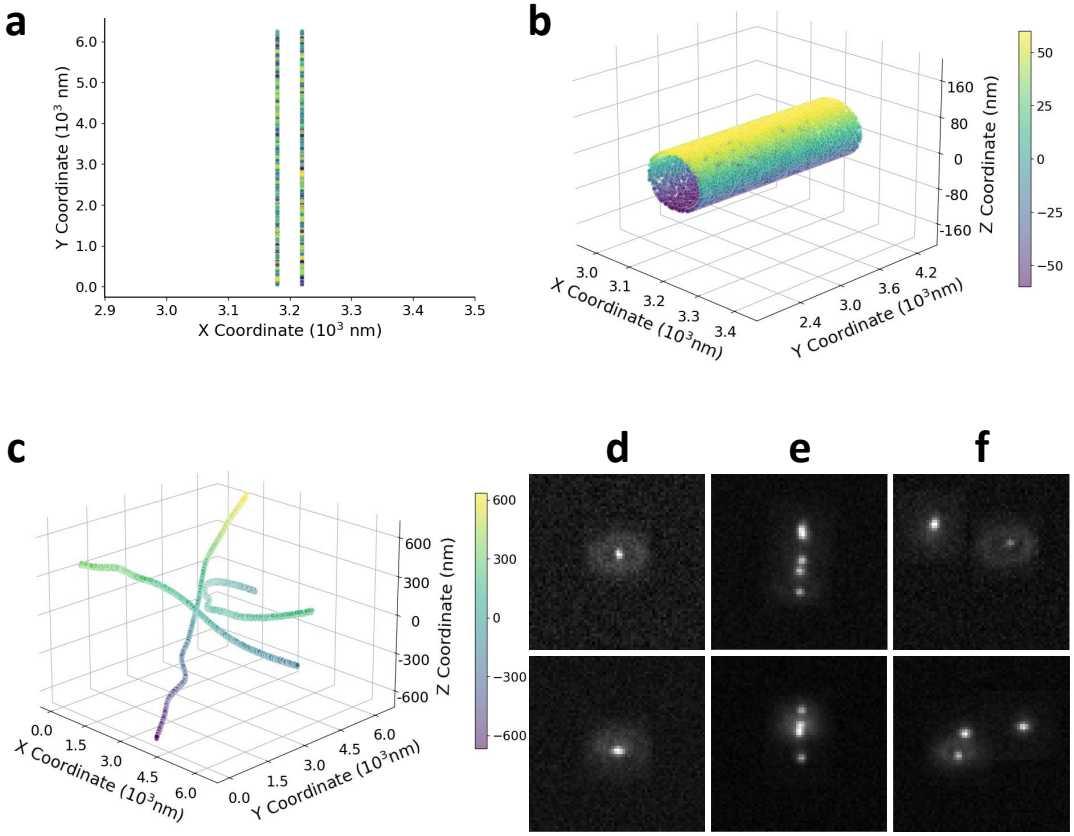
Similarly, the axial Root Mean Square Error ($\mathrm{RMSE}_{\mathrm{ax}}$) quantifies the precision along the optical axis (z-axis):

$$\mathrm{RMSE}_{\mathrm{ax}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{z}_i - z_i^{\mathrm{GT}}\right)^2}. \tag{8}$$

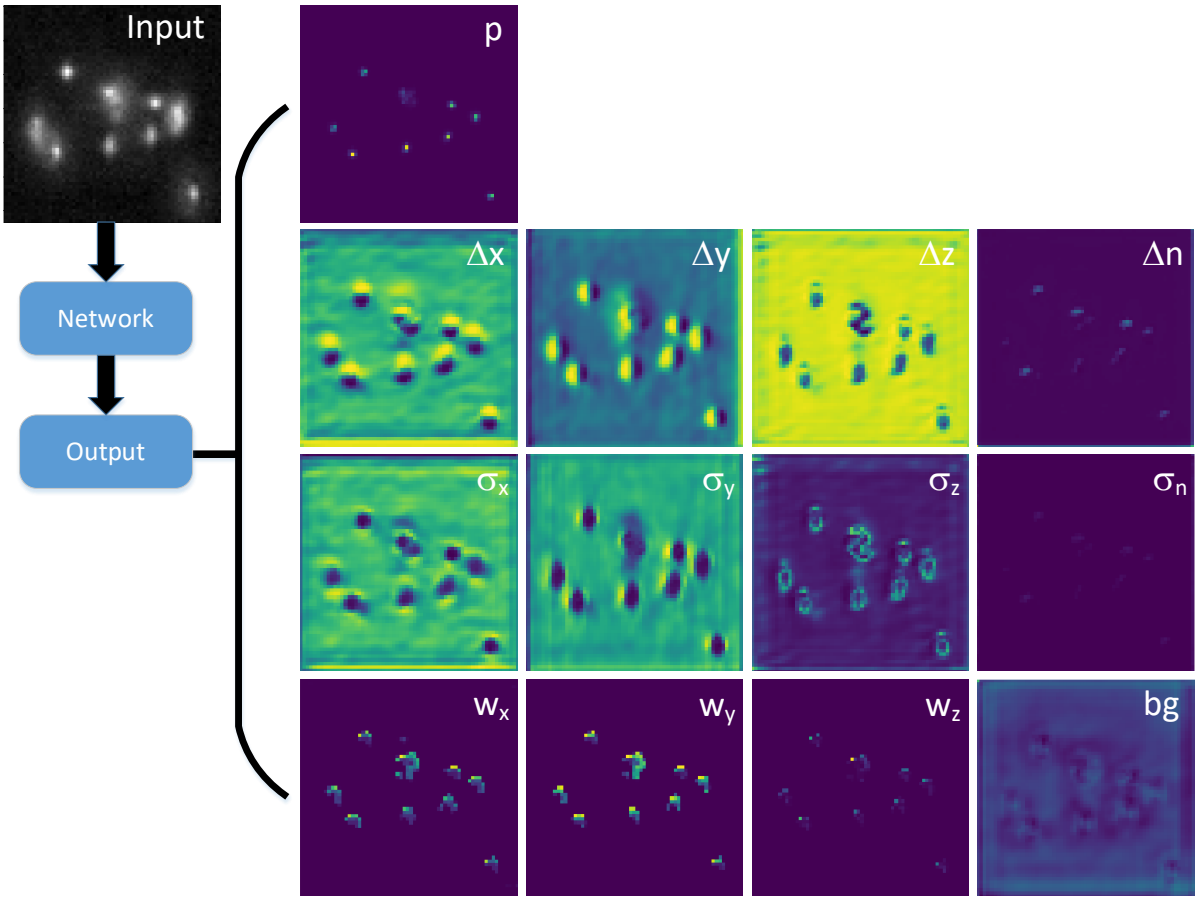
3 Supplementary Results



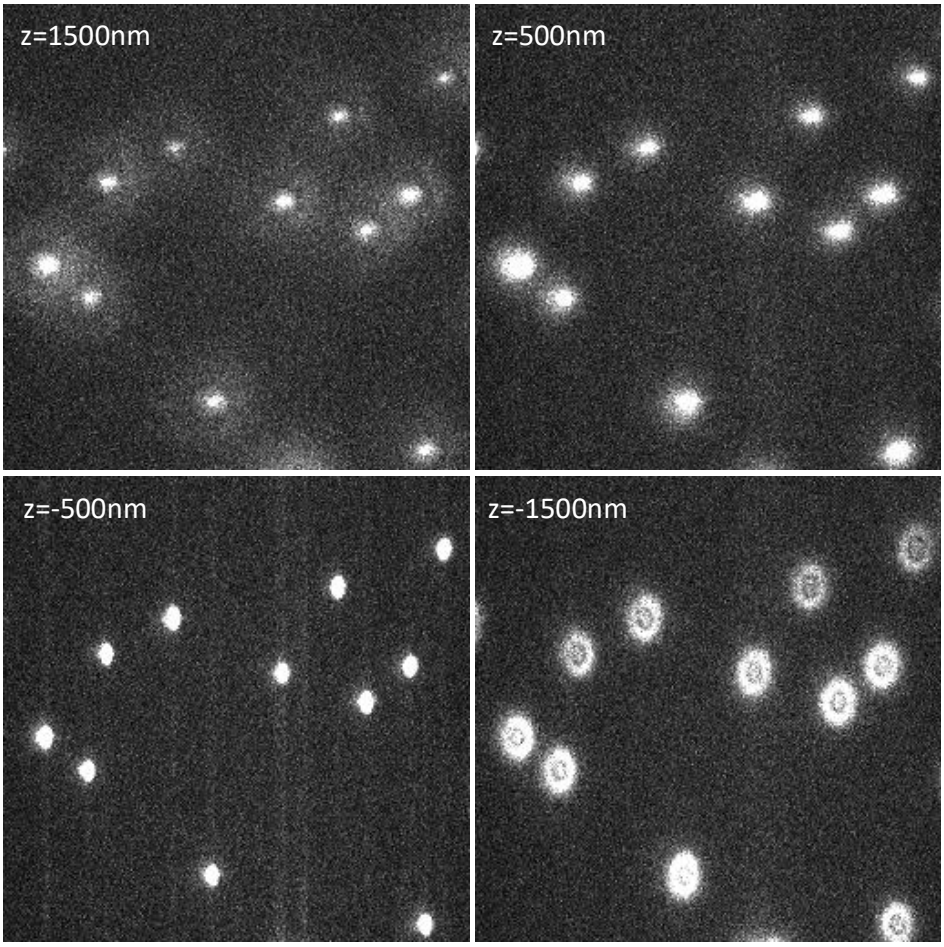
Supplementary Figure 1: **Quantification of intrinsic localization precision using simulated secondary antibody data under noise-free conditions.** The panels show aggregated 3D scatter plots (left) and position histograms in x, y, z (right) for 1,000 aligned emitters at the focal plane ($z = 0$). Rows from top to bottom display results obtained by ORIENCODE, DECODE, and Gaussian fitting, respectively.



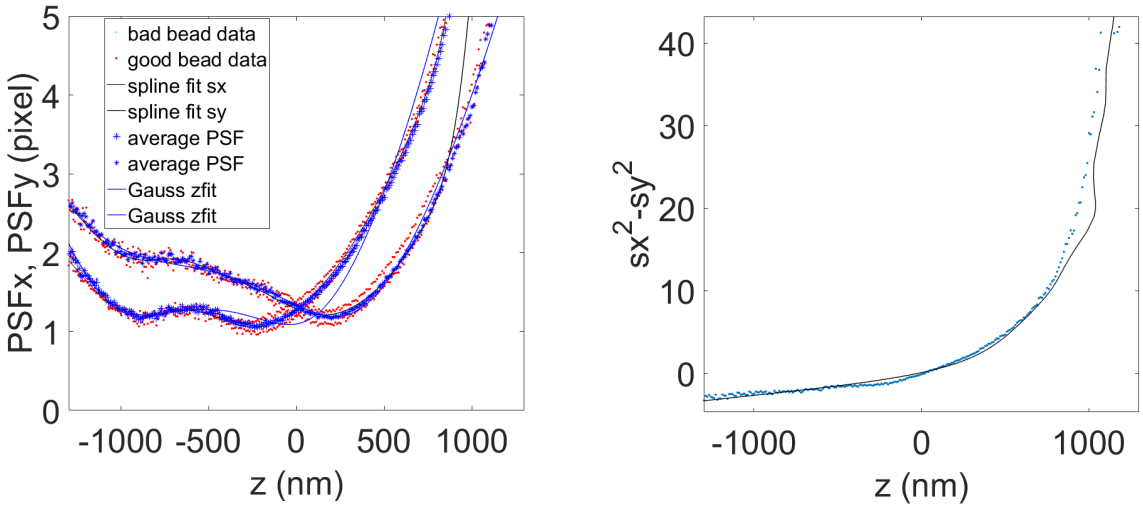
Supplementary Figure 2: **Structure and sample images of simulated data.** **a**, Schematic diagram of two simulated fluorescent lines with a spacing of 40nm and a length of 6400nm. **b**, Schematic diagram of a simulated hollow rod with a radius of 120 nm and a length of 5600 nm. **c**, SMLM challenge training dataset MT0 [1]. **d**–**f**, Visualizations of random frames in **a**–**c**, respectively.



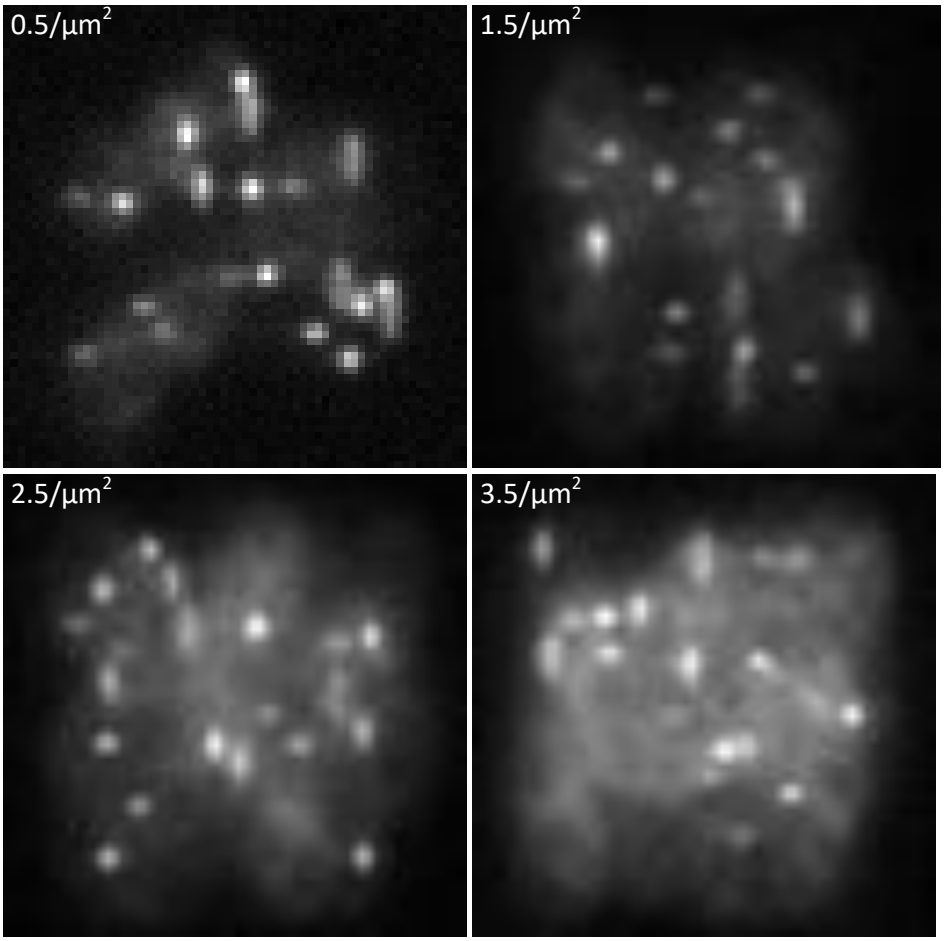
Supplementary Figure 3: **Visualization of the 13 channels predicted by the ORIENCODE.** The network's output is a 13-channel feature map providing a per-pixel description. The channels encode key physical parameters, including: the molecular presence probability (p); sub-pixel offsets for 3D position and intensity ($\Delta x, \ldots, \Delta I$); their associated uncertainties ($\sigma_x, \ldots, \sigma_I$); the predicted localization error bounds (w_x, w_y, w_z) based on the CRLB; and background (bg).



Supplementary Figure 4: **Visualization of raw data from calibration beads.** Samples were prepared by sparsely immobilizing fluorophore-labeled secondary antibodies on a glass coverslip, followed by data acquisition under standard STORM [2] imaging conditions. It is evident that out-of-focus PSFs at large axial distances (e.g., at $z = \pm 1500$ nm) appear highly diffuse and hazy. In contrast, this effect is absent for PSFs within the effective imaging depth (e.g., at $z = \pm 500$ nm), where they maintain a distinct shape.



Supplementary Figure 5: **Calibration of the astigmatic PSF.** To calibrate the astigmatic point spread function, a z-stack of 301 images of a sparsely sampled emitter was recorded. The images were acquired with a 10 nm axial step size over a ± 1.5 μm range around the focal plane. The central region (± 0.5 μm) was used to model the in-focus PSF, while the remaining images in the outer ranges were used for the out-of-focus PSF. **a**, Comparison of Gaussian and spline fitting [3] results for the experimental PSF obtained from calibration beads. **b**, The ellipticity and orientation of the fitted PSF are plotted as a function of axial position, showing their characteristic variation through focus.



Supplementary Figure 6: **Visualization of simulated data under different levels of out-of-focus noise.** The above visualizations display simulated data with a fixed number of 20 in-focus emitters. The density of out-of-focus emitters is progressively increased, with values of 0.5, 1.5, 2.5, 3.5 $/\mu\text{m}^2$. The specific method for generating the simulated data refers to Supplementary Note 2.2.

References

[1] Sage, D., Pham, T.-A., Babcock, H., Lukes, T., Pengo, T., Chao, J., Velmurugan, R., Herbert, A., Agrawal, A., Colabrese, S., *et al.*: Super-resolution fight club: assessment of 2d and 3d single-molecule localization microscopy software. Nature Methods **16**(5), 387–395 (2019)

[2] Huang, B., Wang, W., Bates, M., Zhuang, X.: Three-dimensional super-resolution imaging by stochastic optical reconstruction microscopy. Science **319**(5864), 810–813 (2008)

[3] Li, Y., Mund, M., Hoess, P., Deschamps, J., Matti, U., Nijmeijer, B., Sabinina, V.J., Ellenberg, J., Schoen, I., Ries, J.: Real-time 3d single-molecule localization using experimental point spread functions. Nature Methods **15**(5), 367–369 (2018)