

PROCEEDINGS OF SPIE

[SPIDigitalLibrary.org/conference-proceedings-of-spie](https://www.spiedigitallibrary.org/conference-proceedings-of-spie)

Measuring localization confidence for quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy

Mazidi, Hesam, Ding, Tianben, Nehorai, Arye, Lew, Matthew

Hesam Mazidi, Tianben Ding, Arye Nehorai, Matthew D. Lew, "Measuring localization confidence for quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy," Proc. SPIE 11246, Single Molecule Spectroscopy and Superresolution Imaging XIII, 1124611 (13 February 2020); doi: 10.1117/12.2545033

SPIE.

Event: SPIE BiOS, 2020, San Francisco, California, United States

Measuring localization confidence for quantifying accuracy and heterogeneity in single-molecule super-resolution microscopy

Hesam Mazidi, Tianben Ding, Arye Nehorai, and Matthew D. Lew*

Department of Electrical and Systems Engineering
Washington University in St. Louis, MO, USA 63130

ABSTRACT

We present a computational method, termed Wasserstein-induced flux (WIF), to robustly quantify the accuracy of individual localizations within a single-molecule localization microscopy (SMLM) dataset without ground-truth knowledge of the sample. WIF relies on the observation that accurate localizations are stable with respect to an arbitrary computational perturbation. Inspired by optimal transport theory, we measure the stability of individual localizations and develop an efficient optimization algorithm to compute WIF. We demonstrate the advantage of WIF in accurately quantifying imaging artifacts in high-density reconstruction of a tubulin network. WIF represents an advance in quantifying systematic errors with unknown and complex distributions, which could improve a variety of downstream quantitative analyses that rely upon accurate and precise imaging. Furthermore, thanks to its formulation as layers of simple analytical operations, WIF can be used as a loss function for optimizing various computational imaging models and algorithms even without training data.

Keywords: localization accuracy, statistical confidence, localization software, model mismatch, Wasserstein distance, image quality, optimal transport

1. INTRODUCTION

Single-molecule localization microscopy (SMLM) relies on localizing individual molecules to reconstruct high-resolution images of cellular structures. These images, represented by points in a multi-dimensional space, contain nanoscale, quantitative information regarding the organization, interaction, and dynamics of molecular entities that need to be estimated robustly in the presence of noise. Existing metrics for assessing SMLM image quality can be categorized broadly into two classes: those that require knowledge of the ground-truth positions of fluorophores,¹ and those that operate directly on SMLM reconstructions alone, possibly incorporating information from other measurements (e.g., diffraction-limited imaging).^{2,3} One popular approach is the Jaccard index (JAC),^{1,4} which measures localization accuracy, but has limited applicability for SMLM experiments as it requires exact knowledge of ground-truth molecule positions. Methods that quantify performance by analyzing SMLM reconstructions exploit some aspect of prior knowledge of the target structure or SMLM data. While these methods are able to provide summary or aggregate measures of performance, none of them directly measure the accuracy of *individual* localizations. Such knowledge is critical for harnessing fully the power of SMLM for scientific discovery. In addition, these methods cannot be used to optimize computational imaging models and algorithms in an end-to-end fashion.

Here, we report our broadly-applicable method, termed Wasserstein-induced flux (WIF), to quantify the accuracy of individual localizations given a certain mathematical imaging model. Unlike methods that require ground-truth knowledge, such the exact position of each single molecule (SM), WIF only assumes a basic model of the imaging system such as its point-spread function (PSF). Using WIF, we accurately detect and quantify imaging artifacts arising from overlapping images of SMs in a high-density SMLM dataset of a tubulin network, which otherwise are not revealed by previous methods.

*mdlew@wustl.edu

2. PROPOSED METHOD

2.1 Mathematical modeling

We model estimated single molecules as collection of point sources located at specific positions, each having a positive brightness:

$$\hat{\mathcal{M}} = \sum_{i=1}^{\hat{N}} \hat{s}_i \delta(\mathbf{r} - \hat{\mathbf{r}}_i), \quad (1)$$

where $\hat{s}_i > 0$ and $\hat{\mathbf{r}}_i \in \mathbb{R}^2$ denote the i^{th} molecules' expected brightness (in photons) and position, respectively. Further, all valid $\hat{\mathcal{M}}$ are denoted by a set $\mathcal{M}$. The expected number of photons detected at j^{th} camera pixel can be written as

$$\mu_j = \sum_{i=1}^{\hat{N}} \{\hat{s}_i q_j(\hat{\mathbf{r}}_i)\} + b_j, \quad (2)$$

where $q_j(\hat{\mathbf{r}}_i)$ represents the integral of the PSF (for the i^{th} molecule) over the 2D area occupied by the j^{th} pixel and b_j denotes the expected number of background photons at the j^{th} pixel. Given a set of m pixel measurements as $\mathbf{g} \in \mathbb{R}^m$, the Poisson negative log likelihood $\mathcal{L}$ is then given by (neglecting constant terms)

$$\mathcal{L}(\hat{\mathcal{M}}; \mathbf{g}, \mathbf{b}) = \sum_{j=1}^m \{\mu_j - g_j \log(\mu_j)\}. \quad (3)$$

Central to our method is the idea of transportation plan between two sets of localizations $\mathcal{M}_1 \in \mathcal{M}$ and $\mathcal{M}_2 \in \mathcal{M}$. Informally, a transportation plan π is a mapping that specifies how $\mathcal{M}_1 \in \mathcal{M}$ is mapped to $\mathcal{M}_2 \in \mathcal{M}$. We define the Wasserstein distance $\mathbb{W}_2$ between two sets of localizations $\mathcal{M}_1 \in \mathcal{M}$ and $\mathcal{M}_2 \in \mathcal{M}$, as the minimum cost of transporting one to the other among all valid transportation plans⁵ Π :

$$\mathbb{W}_2(\mathcal{M}_1, \mathcal{M}_2) = \sqrt{\min_{\pi \in \Pi} \left(\sum_{i=1}^{\hat{N}_1} \sum_{j=1}^{\hat{N}_2} \|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j\|_2^2 \pi(\hat{\mathbf{r}}_i, \hat{\mathbf{r}}_j) \right)}, \quad (4)$$

where $\pi(\hat{\mathbf{r}}_i, \hat{\mathbf{r}}_j)$ is the portion of photons from the molecule at position $\hat{\mathbf{r}}_i$ in $\mathcal{M}_1$ that is transported to position $\hat{\mathbf{r}}_j$ in $\mathcal{M}_2$.

Although molecules lie on a continuous domain, we seek to represent them using a finite number of parameters while closely approximating their continuous positions. To this end, we consider a set of $\mathcal{N}$ Cartesian grid points represented by $\mathcal{G} = \{\mathbf{r}_{\mathcal{G}_i}\}_{i=1:\mathcal{N}}$ for which the distance between any two adjacent grid points is given by 2ρ (Fig. 1a). In this way, a set of localizations can be uniquely represented via a discrete grid $\mathcal{G}$:

$$\hat{\mathcal{M}} = \sum_{i=1}^{\hat{N}} \hat{s}_{[i]} \delta(\mathbf{r} - \hat{\mathbf{r}}_{[i]}), \quad (5)$$

where $[i]$ represents a grid point index in $\{1, \dots, \mathcal{N}\}$, $\hat{\mathbf{r}}_{[i]} = \hat{\mathbf{r}}_{\mathcal{G}_{[i]}} + \Delta \hat{\mathbf{r}}_{[i]}$, $\hat{\mathbf{r}}_{\mathcal{G}_{[i]}}$ is the position of the closest grid point to the i^{th} molecule, and $\Delta \hat{\mathbf{r}}_{[i]}$ denotes a position offset relative to the position $\hat{\mathbf{r}}_{\mathcal{G}_{[i]}}$ of the closest grid point (Fig. 1a). Since this choice of model generally guarantees there exists more grid points than actual molecules within a single SMLM image, we denote the support of $\hat{\mathcal{M}}$, that is, all grid points that are associated with one molecule by $\text{Supp}(\hat{\mathcal{M}}) = \{i \in \{1, \dots, \mathcal{N}\} : \hat{s}_i > 0\}$.

Finally, we introduce our notion of source perturbation. Concretely, for each grid point $\hat{\mathbf{r}}_{\mathcal{G}_{[i]}}$ associated with a SM in $\hat{\mathcal{M}}$, we consider its 8 closest grid points denoted by $\hat{\mathbf{r}}_{\mathcal{G}_{[i,j]}}$, $j \in \{1, \dots, 8\}$. We consider *isotropic* perturbations of $\hat{\mathcal{M}}$ as follows:

$$\mathcal{M}_0 = \sum_{i=1}^{\hat{N}} \sum_{j=1}^8 \hat{s}_{[i,j]} \delta(\mathbf{r} - \hat{\mathbf{r}}_{\mathcal{G}_{[i,j]}}), \quad (6)$$

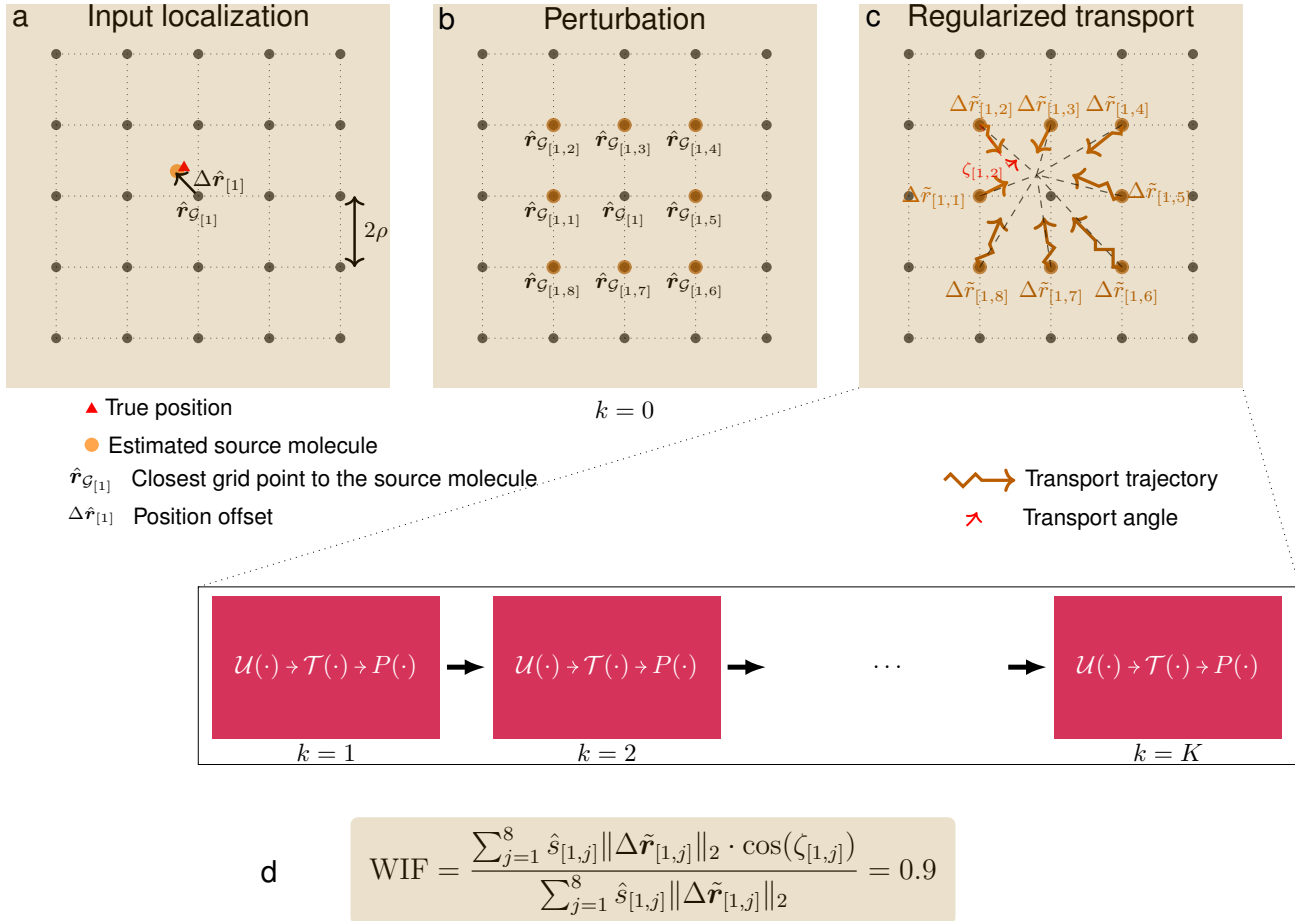


Figure 1. Computing Wasserstein-induced flux (WIF). (a) An input localization (orange circle) located at $\hat{\mathbf{r}}_{[1]}$ is mapped to its closest grid point $\hat{\mathbf{r}}_{g_{[1]}}$ and a position offset vector $\Delta \hat{\mathbf{r}}_{[1]}$, i.e., $\hat{\mathbf{r}}_{[1]} = \hat{\mathbf{r}}_{g_{[1]}} + \Delta \hat{\mathbf{r}}_{[1]}$. Note that 2ρ indicates the distance between any two adjacent points on the grid. (b) A perturbation redistributes a molecule's photons to its 8 closest grid points $\{\hat{\mathbf{r}}_{g_{[1,1]}}, \dots, \hat{\mathbf{r}}_{g_{[1,8]}}\}$. (c) Solving Eq. (14) amounts to finding a set of transport trajectories or displacements denoted by $\{\Delta \tilde{\mathbf{r}}_{[1,1]}, \dots, \Delta \tilde{\mathbf{r}}_{[1,8]}\}$. The transport angle $\zeta_{[1,j]}$ is defined as the angle between the displacement $\Delta \tilde{\mathbf{r}}_{[1,j]}$ and $(\hat{\mathbf{r}}_{[1]} - \hat{\mathbf{r}}_{g_{[1,j]}})$, the vector connecting the perturbed grid point to the estimated localization. Note that we can represent this step by K layers of operations as shown in Algorithm 1. (d) WIF is computed according to the recipe described in Eq. (9). Note that since the estimated localization (orange circle) is close to the true position (red triangle) in this example, we expect WIF to be close to 1, thereby indicating a high degree of confidence.

where $\sum_{j=1}^8 \hat{s}_{[i,j]} = \hat{s}_{[i]}$ and $\hat{s}_{[i,j]}$ is proportional to the distance between $\hat{\mathbf{r}}_{[i]}$ and $\hat{\mathbf{r}}_{g_{[i,j]}}$ (Fig. 1b).

2.2 Wasserstein-induced Flux

We denote $\mathcal{M}_0$ as a *local* perturbation of $\hat{\mathcal{M}}$ (see Eq. (6)). Intuitively, if the PSF model q and the localization algorithm used to recover $\hat{\mathcal{M}}$ are optimal, one expects that the localizations $\mathcal{M}_1$ obtained via

$$\mathcal{M}_1 = \arg \min_{\mathbb{W}_2(\mathcal{M}, \mathcal{M}_0) \leq \epsilon} \mathcal{L}(\mathcal{M}; \mathbf{g}, \mathbf{b}) \quad (7)$$

to be very “close” to $\hat{\mathcal{M}}$. Informally, this qualitative behavior follows from the convexity of the landscape of $\mathcal{L}$ in the vicinity of an optimal solution. Note that the constraint set $\{\mathcal{M} : \mathbb{W}_2(\mathcal{M}, \mathcal{M}_0) \leq \epsilon\}$ ensures that $\mathcal{M}_1$ remains within a small neighbourhood of $\hat{\mathcal{M}}$. Next, we quantify the “closeness” between $\mathcal{M}_1$ and $\hat{\mathcal{M}}$.

Let us represent $\mathcal{M}_1$ via its grid representation as

$$\mathcal{M}_1 = \sum_{i=1}^{\tilde{N}} \sum_{j=1}^8 \hat{s}_{[i,j]} \delta(\mathbf{r} - \tilde{\mathbf{r}}_{[i,j]}). \quad (8)$$

We notice that $\mathcal{M}_1$ reveals transport trajectories $\Delta\tilde{\mathbf{r}}_{[i,j]} \triangleq \tilde{\mathbf{r}}_{[i,j]} - \hat{\mathbf{r}}_{\mathcal{G}_{[i,j]}}$ (from $\mathcal{M}_0$ to $\mathcal{M}_1$) for each source molecule in $\hat{\mathcal{M}}$, i.e., $\hat{\mathbf{r}}_{\mathcal{G}_{[i,j]}}$ are the set of perturbed molecule positions in $\mathcal{M}_0$ and $\tilde{\mathbf{r}}_{[i,j]}$ are the molecule positions in $\mathcal{M}_1$ (Fig. 1c). If $\mathcal{M}_1$ is close to $\hat{\mathcal{M}}$, we expect that all trajectories $\Delta\tilde{\mathbf{r}}_{[i,j]}$ converge to corresponding molecules in $\hat{\mathcal{M}}$ (Fig. 1c). Alternatively, we can interpret this behavior as the stability of $\hat{\mathcal{M}}$ upon a well-chosen local perturbation to $\mathcal{M}_0$.

Therefore, we define our confidence in the estimate of a source molecule, termed Wasserstein-induced Flux (WIF), as the portion of photon flux that returns toward said localization upon a local perturbation:

$$\text{WIF} \triangleq \frac{\sum_{j=1}^8 \hat{s}_{[i,j]} \Delta\tilde{\mathbf{r}}_{[i,j]} \cdot \Delta\hat{\mathbf{u}}_{[i,j]}}{\sum_{j=1}^8 \hat{s}_{[i,j]} \|\Delta\tilde{\mathbf{r}}_{[i,j]}\|}, \quad (9)$$

where WIF takes values in $[-1, 1]$ (1 represents a source molecule with the highest confidence). Note that $\Delta\hat{\mathbf{u}}_{[i,j]}$ is a unit vector pointing from the perturbed source position $\hat{\mathbf{r}}_{\mathcal{G}_{[i,j]}}$ to the original source position $\hat{\mathbf{r}}_{[i]}$.

2.3 Computing WIF

2.3.1 Relaxations and approximations

In this section, we discuss a strategy to compute WIF in Eq. (9). To begin, we need to solve Eq. (7), which is a nontrivial nonlinear optimization. One classic strategy to solve Eq. (7) is to reformulate it via Lagrange relaxation as

$$\mathcal{M}_1 = \arg \min_{\mathcal{M} \in \mathcal{M}} \{ \mathbb{W}_2^2(\mathcal{M}, \mathcal{M}_0) + \epsilon' \mathcal{L}(\mathcal{M}; \mathbf{g}, \mathbf{b}) \}, \quad (10)$$

where ϵ' is related to ϵ in Eq. (7). While Eq. (10) is a convex optimization, computing the Wasserstein distance itself requires solving an optimization problem. In this paper, we propose to regularize the transportation plans in Eq. (10), allowing us to minimize an upper bound of the objective function in Eq. (10), which can be solved efficiently via proximal algorithms.

One way to regularize the transportation in Eq. (10) is to consider a local constraint on $\mathcal{M}$. Informally, we enforce that the perturbed source molecules be transported along certain trajectories. To this end, we impose the following local constraint on $\mathcal{M}$ in Eq. (10):

$$\mathcal{C} \triangleq \{ \mathcal{M} : \|\Delta\mathbf{r}_i\|_2 \leq \rho \ \forall i \in \text{Supp}(\mathcal{M}_0) \}. \quad (11)$$

The constraint set $\mathcal{C}$ in Eq. (11) describes all transportation plans that move each perturbed source along a unique trajectory in the vicinity of the unperturbed source. Therefore, the regularized version of Eq. (10) can be written as

$$\mathcal{M}_1 = \arg \min_{\mathcal{M} \in \mathcal{M} \cap \mathcal{C}} \{ \mathbb{W}_2^2(\mathcal{M}, \mathcal{M}_0) + \epsilon' \mathcal{L}(\mathcal{M}; \mathbf{g}, \mathbf{b}) \}. \quad (12)$$

Next, we bound $\mathbb{W}_2$ for all valid localizations in the constraint set $\mathcal{M} \cap \mathcal{C}$. It turns out that $\forall \mathcal{M} \in \mathcal{M} \cap \mathcal{C}$ we can bound $\mathbb{W}_2^2(\mathcal{M}, \mathcal{M}_0)$ from above with the group sparsity norm:

$$\mathcal{R}(\mathcal{M}) \triangleq \sum_{i=1}^{\mathcal{N}} \sqrt{s_i^2 + s_i^2 \|\Delta\mathbf{r}_i\|_2^2}, \quad (13)$$

so long as $\|\Delta\mathbf{r}_{[i,j]}\|_2^2 \leq \sqrt{1 + \|\Delta\mathbf{r}_{[i,j]}\|_2^2}$ (see Appendix A.1). We note that this assumption can be easily satisfied by appropriately scaling 2ρ , the separation between grid points, in the object model. Note that $\mathcal{R}$ defines a group-

sparsity norm for which it is assumed that $s_i = 0$ for grid indices i that do not contain perturbed sources, i.e., $\{s_i = 0 \mid i \notin \text{Supp}(\mathcal{M}_0)\}$.

We are now ready to state the main (convex) optimization problem that we use for computing WIF. By plugging in the upper bound of $\mathbb{W}_2$ derived in Eq. (22) we have

$$\mathcal{M}_1 = \arg \min_{\mathcal{M} \in \mathcal{M} \cap \mathcal{C}} \{\nu \mathcal{R}(\mathcal{M}) + \mathcal{L}(\mathcal{M}; \mathbf{g}, \mathbf{b})\}, \quad (14)$$

where $\nu > 0$ represents the regularizer strength.

2.3.2 Implementation

The non-smooth nature of the group-sparsity norm and the large size of signal, that is $\mathcal{N}$, makes the minimization in Eq. (14) challenging. Here, we present an accelerated proximal gradient algorithm⁶ to solve Eq. (14) efficiently. The inputs to the algorithm include the recorded image $\mathbf{g} \in \mathbb{R}^m$; a list of localizations $\mathbf{L}$ consisting of the estimated position and (expected) brightness of each SM; the estimated (expected) background $\mathbf{b} \in \mathbb{R}^m$; a PSF model $\mathbf{q} \in \mathbb{R}^m$ and its spatial gradients along x, y given by $\mathbf{q}_x, \mathbf{q}_y \in \mathbb{R}^m$, respectively; and the regularizer strength $\nu > 0$. We further represent $\mathcal{M}$ using its grid representation as $\sum_{i=1}^{\mathcal{N}} s_i \delta(\mathbf{r} - \mathbf{r}_{\mathcal{G}_i} - \Delta \mathbf{r}_i)$ and its equivalent vector representation

$$\boldsymbol{\gamma} = [s_1, \dots, s_{\mathcal{N}}, s_1 \Delta x_1, \dots, s_{\mathcal{N}} \Delta x_{\mathcal{N}}, s_1 \Delta y_1, \dots, s_{\mathcal{N}} \Delta y_{\mathcal{N}}]^T \quad (15)$$

$$= [\mathbf{s}^T, \mathbf{s}^T \odot \Delta \mathbf{x}^T, \mathbf{s}^T \odot \Delta \mathbf{y}^T]^T = [\mathbf{s}^T, \mathbf{p}_x^T, \mathbf{p}_y^T]^T \in \mathbb{R}^{3\mathcal{N}}, \quad (16)$$

where $\odot$ represents element-wise multiplication and T denotes the transpose operator. We note that such a representation in $\boldsymbol{\gamma}$ allows us to compute $q(\mathcal{M})$, that is the image of an arbitrary collection of SMs, using a first-order Taylor approximation of PSF q expanded around each grid point. Moreover, we can express the constraint set $\mathcal{C}$ in terms of $\boldsymbol{\gamma}$ as

$$\mathcal{C} \triangleq \{\boldsymbol{\gamma} : \|\mathbf{p}_{x,i}^2 + \mathbf{p}_{y,i}^2\|_2 \leq \rho s_i \forall i \in \text{Supp}(\boldsymbol{\gamma})\}. \quad (17)$$

A summary of variables and notations used in the algorithm is shown in Table 1.

At a high level, our accelerated proximal algorithm consists of K layers, each representing one iteration (Fig. 1c). At layer or iteration $k = 0$, perturbed sources as well as some initial parameters (such as step size) are computed. In the following layers ($k \geq 1$), the algorithm adaptively updates its hyper-parameters such as step size β (line 4 in Alg. 1); it locally “transports” perturbed SMs in $\boldsymbol{\gamma}$ (line 5 in Alg. 1); and it combines information from previous iterations to accelerate the convergence (line 6 in Alg. 1).

Algorithm 1 Regularized transport: an accelerated proximal gradient algorithm

- 1: **Input:** $\{\mathbf{g}, \mathbf{L}, \mathbf{b}, \nu, \mathbf{q}, \mathbf{q}_x, \mathbf{q}_y\}$
 - 2: **Step 0.** Compute $\boldsymbol{\gamma}_0 = \mathcal{P}(\mathbf{L})$, β_0 , and take $\mathbf{v}_1 = \boldsymbol{\gamma}_0, t_1 = 1$. $\triangleright \mathcal{P}$ perturbs $\boldsymbol{\gamma}$ according to Eq. (6).
 - 3: **Step k .** ($1 \leq k \leq K$)
 - 4: $\beta_k = \mathcal{U}(\beta_{k-1})$ $\triangleright \mathcal{U}(\cdot)$ outputs an appropriate step size (see Appendix A.2)
 - 5: $\boldsymbol{\gamma}_k = P_{\mathcal{C}} \left(\underbrace{\mathbf{v}_k - \beta_k [\nabla \mathcal{L}(\mathbf{v}_k) + \nabla w(\mathbf{v}_k)]}_{\text{transportation of } \mathbf{v}_k \triangleq \mathcal{T}(\mathbf{v}_k)} \right)$ $\triangleright P_{\mathcal{C}}(\cdot)$ denotes projection onto the set $\mathcal{C}$, regularizing the transportation (see Appendix A.4)
 - 6: $t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$, $\mathbf{v}_{k+1} = \boldsymbol{\gamma}_k + \frac{t_k - 1}{t_{k+1}}(\boldsymbol{\gamma}_k - \boldsymbol{\gamma}_{k-1})$
-

In the last module of our implementation, we use $\boldsymbol{\gamma}_K$ as an estimate of $\mathcal{M}_1$ to compute WIF according to Eq. (9).

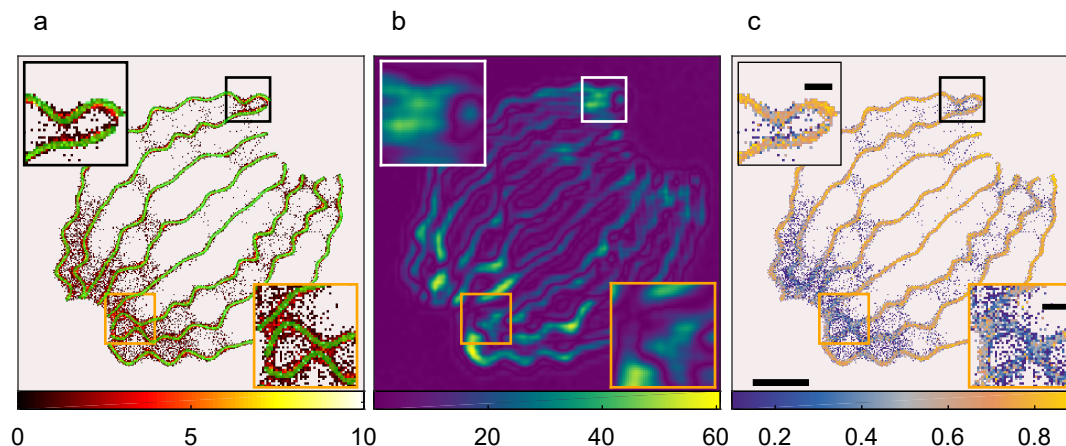


Figure 2. WIF confidence map reveals artifacts in recovering a tubulin network from high-density SMLM data. (a) Recovered structure (red) using FALCON overlaid with the ground truth (green). (b) Error map recovered by SQUIRREL (brighter colors correspond to larger errors). (c) WIF confidence map (brighter colors indicate higher confidence) obtained by averaging localization confidences in each pixel. Colorbars: (a) number of localizations, (b) error, and (c) confidence per $20 \times 20 \text{ nm}^2$. Scalebars: (c) $1 \mu\text{m}$, insets: 200 nm .

3. RESULTS

A key difficulty encountered in high-density (HD) localization, when images of molecules overlap on the camera, is image artifacts that distort SMLM reconstructions in a structured, or vectorial, manner. Constructing an SMLM error map using a reference image has been proposed to certify the reliability of an SMLM reconstruction, but such a map does not quantify the reliability of each individual localization within the image. Here, we illustrate the power of measuring SM confidence in quantifying and revealing artifacts in a challenging HD localization experiment.

We use FALCON,⁷ an HD localization algorithm, to reconstruct a simulated benchmark SMLM dataset⁴ consisting of 360 HD frames of a tubulin network (Fig. 2a). In regions where the tubules coalesce, corresponding to higher blinking densities, we see numerous inaccurate localizations (Fig. 2a, insets). In particular, we see fused and broadened tubules instead of thin and separate structures. A reliable error map should assign low confidence or high error to such regions while discriminating fine but accurate details of the structure.

Interestingly, we notice significant differences between an error map (Fig. 2b, obtained via SQUIRREL³) and the proposed confidence map (Fig. 2c). First, the error map appears to overestimate errors in regions with accurate localizations, while our confidence map exhibits low confidence for inaccurate localizations and assigns high confidence to neighboring, well-resolved parallel tubules (Fig. 2b,c, top insets). Second, the error map underestimates the error in the regions where tubules are apparently fused, whereas the confidence map assigns an overall low confidence to this region, suggesting potential artifacts (Fig. 2b,c, bottom insets). Overall, our WIF confidence map enables scientists to discriminate *specific* SM localizations that are trustworthy, while also assigning low confidence values to those that are not, thereby maximizing the utility of SMLM datasets without throwing away useful localizations.

4. CONCLUSION

In the present paper, we presented a computational imaging method, WIF, to quantify the reliability of individual localizations within an arbitrary SMLM dataset given a certain computational model. WIF capitalizes on the observation that accurate localizations exhibit a high degree of stability upon an arbitrary perturbation. We formulated a constrained optimization problem exploiting optimal transport theory to allow various perturbations. As an application of WIF, we showed that its confidence map enables detection and quantification of

imaging artifacts in recovering a tubulin network from a high-density SMLM dataset. From the computational prospective, WIF is the result of a cascade of simple, analytical operations, which makes it an attractive end-to-end loss function for training and evaluating computational models such as PSFs and algorithms such as deep neural networks directly on experimental data.

ACKNOWLEDGMENTS

Research reported in this publication was supported by the National Science Foundation under grant number ECCS-1653777 and by the National Institute of General Medical Sciences of the National Institutes of Health under grant number R35GM124858.

APPENDIX A. MATHEMATICAL DETAILS

A.1 Bounding Wasserstein distance from above

First, using our grid representation in Eq. (5), any measure $\mathcal{M} \in \mathcal{M} \cap \mathcal{C}$ can be represented as

$$\mathcal{M} = \sum_{i=1}^{\hat{N}} \left(\sum_{j=1}^8 s_{[i,j]} \delta(\mathbf{r} - (\mathbf{r}_{\mathcal{G}_{[i,j]}} + \Delta \mathbf{r}_{[i,j]})) \right) \quad (18)$$

such that $\sum_{i=1}^{\hat{N}} \sum_{j=1}^8 s_{[i,j]} = \hat{s}$, that is, detected photons are preserved. Notice that the constraint set $\mathcal{C}$ allows us to unambiguously represent the position of molecules in $\mathcal{M}$ as $\mathbf{r}_{\mathcal{G}_{[i,j]}} + \Delta \mathbf{r}_{[i,j]}$, $\|\Delta \mathbf{r}_{[i,j]}\|_2 \leq \rho$. Therefore,

$$\mathbb{W}_2^2(\mathcal{M}, \mathcal{M}_0) \leq \sum_{i=1}^{\hat{N}} \sum_{j=1}^8 s_{[i,j]} \|\Delta \mathbf{r}_{[i,j]}\|_2^2 \quad (19)$$

$$\leq \sum_{i=1}^{\hat{N}} \sum_{j=1}^8 s_{[i,j]} \sqrt{1 + \|\Delta \mathbf{r}_{[i,j]}\|_2^2} \quad (20)$$

$$= \sum_{i=1}^{\hat{N}} \sum_{j=1}^8 \sqrt{s_{[i,j]}^2 + s_{[i,j]}^2 \|\Delta \mathbf{r}_{[i,j]}\|_2^2}, \quad (21)$$

where the first inequality in Eq. (19) follows from the definition of Wasserstein distance and the second inequality in Eq. (20) is the consequence of our assumption. Notice that Eq. (21) may be recast as a group-sparsity norm:

$$\sum_{i=1}^{\hat{N}} \sum_{j=1}^8 \sqrt{s_{[i,j]}^2 + s_{[i,j]}^2 \|\Delta \mathbf{r}_{[i,j]}\|_2^2} = \sum_{i=1}^{\mathcal{N}} \sqrt{s_i^2 + s_i^2 \|\Delta \mathbf{r}_i\|_2^2} \triangleq \mathcal{R}(\mathcal{M}), \quad (22)$$

where it is assumed that $s_i = 0$ for grid index i that do not contain perturbed sources, i.e., $\{s_i = 0 \mid i \notin \text{Supp}(\mathcal{M}_0)\}$.

A.2 A simple backtracking algorithm

Here, we present a simple backtracking method for choosing an appropriate β , i.e., step size, which guarantees convergence and describes $\mathcal{U}$ in Alg. 1. Let $\eta > 0$ be a fixed number and consider β_{k-1} as previous step size. Then, the following algorithm, collectively denoted by $\mathcal{U}$, outputs β_k .

Algorithm 2 A simple backtracking algorithm

- 1: **Input:** $\{\beta_{k-1}, \mathbf{v}_k, \mathbf{g}, \mathbf{b}, \nu, \mathbf{q}, \mathbf{q}_x, \mathbf{q}_y\}$
 - 2: Find smallest integer i_k such that with $\bar{\beta} = \eta^{i_k} \beta_{k-1}$:
 - 3: $\bar{\mathbf{v}}_{k,\bar{\beta}} = P_{\mathcal{C}}(\mathbf{v}_k - \bar{\beta}[\nabla \mathcal{L}(\mathbf{v}_k) + \nabla w_{\nu}(\mathbf{v}_k)])$
 - 4: $\mathcal{L}(\bar{\mathbf{v}}_{k,\bar{\beta}}) \leq \mathcal{L}(\mathbf{v}_k) + [\nabla \mathcal{L}(\mathbf{v}_k) + \nabla w_{\nu}(\mathbf{v}_k)]^T (\bar{\mathbf{v}}_{k,\bar{\beta}} - \mathbf{v}_k) + 1/\bar{\beta} \left\| \bar{\mathbf{v}}_{k,\bar{\beta}} - \mathbf{v}_k \right\|_2^2$
 - 5: **Return** $\beta_k = \eta^{i_k} \beta_{k-1}$
-

A.3 Moreau envelope

The Moreau envelope of $\nu\mathcal{R}$, parameterized by a smoothing value $\mu > 0$, is a smooth, differentiable approximation to non-differentiable function $\nu\mathcal{R}(\gamma)$. In particular, its gradient w.r.t. γ is given by

$$\nabla w_{\gamma}(\mathbf{v}) = \frac{1}{\mu}(\mathbf{v} - \text{prox}_{\nu\mathcal{R}}(\mathbf{v})), \quad (23)$$

where $\text{prox}_{\nu\mathcal{R}}(\mathbf{v})$ is defined as

$$\text{prox}_{\nu\mathcal{R}}(\mathbf{v}) \triangleq \arg \min_{\gamma} \left(\nu\mathcal{R}(\gamma) + 1/2 \|\gamma - \mathbf{v}\|_2^2 \right). \quad (24)$$

A.4 Projection onto $\mathcal{C}$

Recall the constraint set $\mathcal{C}$ defined in terms of γ as

$$\mathcal{C} \triangleq \{\gamma : \|\mathbf{p}_{x,i}^2 + \mathbf{p}_{y,i}^2\|_2 \leq \rho s_i \ \forall i \in \text{Supp}(\gamma)\}. \quad (25)$$

It can be shown that the element-wise projection of γ onto $\mathcal{C}$ is given by ($\forall i \in \{1, \dots, \mathcal{N}\}$)

$$P_{\mathcal{C}}(\gamma_i) = \begin{cases} 0, & \text{if } \sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2} \leq -s_i/\rho \\ (s_i, s_i \Delta x_i, s_i \Delta y_i), & \text{if } \sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2} \leq s_i \rho \\ \frac{\rho(s_i + \rho \sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2})}{(1 + \rho^2) \sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2}} \left(\frac{\sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2}}{\rho}, s_i \Delta x_i, s_i \Delta y_i \right), & \text{if } \sqrt{(s_i \Delta x_i)^2 + (s_i \Delta y_i)^2} > s_i \rho \end{cases} \quad (26)$$

A.5 Table of mathematical notations

Table 1. Mathematical notations

Notation	Definition
$\mathcal{N} > 0$	number of grid points
$m > 0$	number of camera pixels in the area of interest
$\rho > 0$	half distance between two adjacent grid points
$\zeta \in \mathbb{N}$	upsampling factor between image space and object space
$\mathbf{g} \in \mathbb{R}^m$	measured image on the camera (a square area)
$\mathbf{b} \in \mathbb{R}^m$	estimated background
$\mathbf{s} \in \mathbb{R}^{\mathcal{N}}$	brightnesses at grid points
$\Delta \mathbf{x} \in \mathbb{R}^{\mathcal{N}}$	position offsets along x at grid points
$\Delta \mathbf{y} \in \mathbb{R}^{\mathcal{N}}$	position offsets along y at grid points
γ	all parameters at grid points (Eq. (16))
γ_{init}	an initial estimate
$\mathcal{R}(\gamma)$	group-sparsity norm of γ (Eq. (22))
$\mathcal{L}(\gamma)$	the negative log likelihood evaluated at γ
$\nabla \mathcal{L}_{\gamma}(\mathbf{v})$	the derivative of the negative log likelihood w.r.t. γ evaluated at $\mathbf{v}$
$\text{Supp}(\gamma)$	the set of grid points in γ that contain one molecule
$P_{\mathcal{C}}(\gamma)$	projection of γ onto $\mathcal{C}$
$\nu > 0$	regularizer strength
$\mu > 0$	smoothing parameter in approximating $\mathcal{R}$ with its Moreau envelope
$w(\gamma)$	Moreau envelope of $\nu \mathcal{R}(\gamma)$ (see Appendix A.3)
$\mathcal{U}$	backtracking algorithm for finding an appropriate step size
$\beta > 0$	step size
$K \in \mathbb{N}$	number of layers or iterations
k	iteration number
$\eta > 1$	backtracking parameter
T	transpose operator
$\odot$	element-wise product

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] Nieuwenhuizen, R. P. J., Lidke, K. A., Bates, M., Puig, D. L., Grünwald, D., Stallinga, S., and Rieger, B., “Measuring image resolution in optical nanoscopy,” *Nature Methods* **10**, 557–562 (2013).
- [3] Culley, S., Albrecht, D., Jacobs, C., Pereira, P. M., Leterrier, C., Mercer, J., and Henriques, R., “Quantitative mapping and minimization of super-resolution optical imaging artifacts,” *Nature Methods* **15**(4), 263–266 (2018).
- [4] Sage, D., Kirshner, H., Pengo, T., Stuurman, N., Min, J., Manley, S., and Unser, M., “Quantitative evaluation of software packages for single-molecule localization microscopy,” *Nature Methods* **12**, 717–724 (2015).
- [5] Peyré, G. and Cuturi, M., “Computational optimal transport,” *Foundations and Trends® in Machine Learning* **11**(5-6), 355–607 (2019).
- [6] Mazidi, H., Lu, J., Nehorai, A., and Lew, M. D., “Minimizing structural bias in single-molecule super-resolution microscopy,” *Scientific Reports* **8**, 13133 (2018).
- [7] Min, J., Vonesch, C., Kirshner, H., Carlini, L., Olivier, N., Holden, S., Manley, S., Ye, J. C., and Unser, M., “FALCON: fast and unbiased reconstruction of high-density super-resolution microscopy data,” *Scientific Reports* **4**, 4577 (2015).