Supplementary Note 2: Metrics of assessment

1 Notation

The notations $(\cdot)^s$ and $(\cdot)^t$ denote the software and ground truth quantities, respectively. Note we distinguish between molecule (*i.e.,* one physical fluorophore) and activation (*i.e.,* one emitting fluorophore during one frame acquisition).

1. x_i, y_i, z_i: Position of an activation i.
2. $x^{\mathrm{m}}, y^{\mathrm{m}}, z^{\mathrm{m}}$: True position of molecules.
3. Estimated activations (S): Set of estimated activations.
4. True activations (T): Set of true activations.
5. Γ denotes the cardinality of a set.

2 Statistics on localizations (frame-wise)

1. True positive (TP): $\Gamma(\mathrm{S} \cap \mathrm{T})$. Number of paired activations according to the protocol (*Methods*).
2. False positive (FP): $\Gamma(\mathrm{S} \setminus \mathrm{S} \cap \mathrm{T})$. Number of unpaired estimated activations.
3. False negative (FN): $\Gamma(\mathrm{T} \setminus \mathrm{S} \cap \mathrm{T})$. Number of unpaired true activations.
4. Jaccard index (JAC): TP/(TP + FP + FN). More robust than Recall.
5. Recall: TP/(TP + FN).
6. Precision: TP/(TP + FP).
7. RMSE volumetric (RMSE Volum.) [nm]: $\sqrt{\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}}(z_i^s - z_i^t)^2 + (x_i^s - x_i^t)^2 + (y_i^s - y_i^t)^2}$.
8. RMSE lateral (RMSE Lateral) [nm]: $\sqrt{\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}}(x_i^s - x_i^t)^2 + (y_i^s - y_i^t)^2}$.
9. RMSE axial (RMSE Axial) [nm]: $\sqrt{\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}}(z_i^s - z_i^t)^2}$.
10. Mean absolute deviation volumetric (MAD Volum.) [nm]: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} |x_i^s - x_i^t| + |y_i^s - y_i^t| + |z_i^s - z_i^t|$.
11. Mean absolute deviation lateral (MAD Lateral) [nm]: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} |x_i^s - x_i^t| + |y_i^s - y_i^t|$.
12. Mean absolute deviation axial (MAD Axial) [nm]: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} |z_i^s - z_i^t|$.
13. Δx: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} (x_i^s - x_i^t)$.
14. Δy: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} (y_i^s - y_i^t)$.
15. Δz: $\frac{1}{\mathrm{TP}}\sum_{i\in\mathrm{S}\cap\mathrm{T}} (z_i^s - z_i^t)$.

16. Photons correlation (Correl. photons): Pearson correlation $r = \frac{\sum_{i \in \mathrm{S} \cap \mathrm{T}} (p_i^s - \bar{p^s})(p_i^t - \bar{p^t})}{\sqrt{\sum_{i \in \mathrm{S} \cap \mathrm{T}} (p_i^s - \bar{p^s})^2} \sqrt{\sum_{i \in \mathrm{S} \cap \mathrm{T}} (p_i^t - \bar{p^t})^2}}$ with p the photon count and $\bar{\cdot}$ the averaged value.

17. Fourier shell correlation (FSC, Volum.) [nm]: Fourier Shell Correlation as defined in [1] for volume XYZ.

18. Fourier ring correlation (FRC yz) [nm]: Fourier Ring Correlation as defined in [1] for projected volume along the X axis (on plane YZ).

19. Fourier ring correlation (FRC xz) [nm]: Fourier Ring Correlation as defined in [1] for projected volume along the Y axis (on plane XZ).

20. Fourier ring correlation (FRC xy) [nm]: Fourier Ring Correlation as defined in [1] for projected volume along the Z axis (on lateral plane XY).

21. Signal-to-noise ratio volumetric (SNR Volum.) [dB]: SNR on the volume defined as $20 \log \left(\frac{I^t}{I^t - I^s} \right)$ with I the Gaussian rendered image ($\sigma = 5$nm). The volume has $1500 \times 1500 \times 750$ voxels.

22. Signal-to-noise ratio yz (SNR yz) [dB]: SNR on the plane YZ defined as $20 \log \left(\frac{I^t}{I^t - I^s} \right)$ with I the Gaussian rendered image ($\sigma = 5$nm). The 2D images I are the results of the projection of all localizations on the plane YZ.

23. Signal-to-noise ratio xz (SNR xz) [dB]: SNR on the plane XZ defined as $20 \log \left(\frac{I^t}{I^t - I^s} \right)$ with I the Gaussian rendered image ($\sigma = 5$nm). The 2D images I are the results of the projection of all localizations on the plane XZ.

24. Signal-to-noise ratio xy (SNR xy) [dB]: SNR on the plane XY defined as $20 \log \left(\frac{I^t}{I^t - I^s} \right)$ with I the Gaussian rendered image ($\sigma = 5$nm). The 2D images I are the results of the projection of all localizations on the plane XY.

3 Comparison molecule-wise

1. True molecules ($\mathrm{T_m}$): Set of true molecules.

2. PM: Set of paired molecules. Each molecule has a set of activations. We consider a molecule is paired if there is at least one paired activation from its set.

3. True positive molecule-wise (TP Mol.): $\Gamma(\mathrm{PM})$. Number of paired molecules.

4. $\overline{x^{\mathrm{m},s}}, \overline{y^{\mathrm{m},s}}, \overline{z^{\mathrm{m},s}}$: Position of a molecule estimated by a software. It corresponds to the average over all the paired activations of the considered molecule.

5. False negative molecule-wise (FN Mol.): Number of unpaired true molecules.

6. Recall molecule-wise (Recall Mol.): TP Mol./(TP Mol. + FN Mol.).

7. RMSE molecule-wise volumetric (RMSE Mol. Volum.) [nm]:
$$\sqrt{\frac{1}{\mathrm{TPMol.}} \sum_{i \in \mathrm{PM}} (\overline{z_i^{\mathrm{m},s}} - z_i^{\mathrm{m},t})^2 + (\overline{x_i^{\mathrm{m},s}} - x_i^{\mathrm{m},t})^2 + (\overline{y_i^{\mathrm{m},s}} - y_i^{\mathrm{m},t})^2}.$$

8. RMSE molecule-wise lateral (RMSE Mol. Lateral) [nm]:
$$\sqrt{\frac{1}{\mathrm{TPMol.}} \sum_{i \in \mathrm{PM}} (\overline{x_i^{\mathrm{m},s}} - x_i^{\mathrm{m},t})^2 + (\overline{y_i^{\mathrm{m},s}} - y_i^{\mathrm{m},t})^2}.$$

9. Mean absolute deviation molecule-wise volumetric (MAD Mol. Volum.) [nm]:
$$\frac{1}{\mathrm{TPMol.}} \sum_{i \in \mathrm{PM}} |\overline{z_i^{\mathrm{m},s}} - z_i^{\mathrm{m},t}| + |\overline{x_i^{\mathrm{m},s}} - x_i^{\mathrm{m},t}| + |\overline{y_i^{\mathrm{m},s}} - y_i^{\mathrm{m},t}|.$$

10. Mean absolute deviation molecule-wise lateral (MAD Mol. Lateral) [nm]:
$$\frac{1}{\mathrm{TPMol.}} \sum_{i \in \mathrm{PM}} +|\overline{x_i^{\mathrm{m},s}} - x_i^{\mathrm{m},t}| + |\overline{y_i^{\mathrm{m},s}} - y_i^{\mathrm{m},t}|.$$

11. Detection ratio molecule-wise (Detection ratio Mol.): Averaged ratio between paired activation and total number of activations for each molecule.

4 Axial (depth) assessment

In this section, we propose new metrics related to the axial (depth) performance of the software. These metrics are extracted from curves as shown inn Fig S12. From the curve showing a metric with respect to the axial position, we assess the depth performance of each software. We remind here that these metrics with respect to the axial position are "locally" computed for each axial position (i.e., small interval of axial positions). We fit a parabolic curve to each curve in order to reduce the effect of finite number per bin notably. We first assume a software is robust as long as the metric is above a certain threshold. We set it as a percentage of the maximal value the software reaches. We used two different thresholds T1 = 25% and T2 = 75%. We considered two metrics with respect to axial position: recall and Jaccard index.

1. R(z): Recall with respect to axial position.
2. Jac(z): Jaccard index with respect to axial position.
3. $R_p(z)$: Parabolic fit of R(z).
4. $JAC_p(z)$: Parabolic fit of Jac(z).
5. MinZ recall T1 [nm]: The lowest axial position for which $R_p(z)$ is below the threshold T1.
6. MaxZ recall T1 [nm]: The highest axial position for which $R_p(z)$ is below the threshold T1.
7. Mean RMSExy recall T1: Lateral RMSE averaged on the z-range [MinZ recall T1, MaxZ recall T1].
8. Mean RMSEz recall T1: Axial RMSE averaged on the z-range [MinZ recall T1, MaxZ recall T1].
9. MinZ recall T2 [nm]: The lowest axial position for which $R_p(z)$ is below the threshold T2.
10. MaxZ recall T2 [nm]: The highest axial position for which $R_p(z)$ is below the threshold T2.
11. Mean RMSExy recall T2: Lateral RMSE averaged on the z-range [MinZ recall T2, MaxZ recall T2].
12. Mean RMSEz recall T2: Axial RMSE averaged on the z-range [MinZ recall T2, MaxZ recall T2].
13. MinZ Jaccard T1 [nm]: The lowest axial position for which $JAC_p(z)$ is below the threshold T1.
14. MaxZ Jaccard T1 [nm]: The highest axial position for which $JAC_p(z)$ is below the threshold T1.
15. Mean RMSExy Jaccard T1: Lateral RMSE averaged on the z-range [MinZ Jaccard T1, MaxZ Jaccard T1].
16. Mean RMSEz Jaccard T1: Axial RMSE averaged on the z-range [MinZ Jaccard T1, MaxZ Jaccard T1].
17. MinZ Jaccard T2 [nm]: The lowest axial position for which $JAC_p(z)$ is below the threshold T2.
18. MaxZ Jaccard T2 [nm]: The highest axial position for which $JAC_p(z)$ is below the threshold T2.
19. Mean RMSExy Jaccard T2: Lateral RMSE averaged on the z-range [MinZ Jaccard T2, MaxZ Jaccard T2].
20. Mean RMSEz Jaccard T2: Axial RMSE averaged on the z-range [MinZ Jaccard T2, MaxZ Jaccard T2].
21. Max recall: Maximal recall of $R_p(z)$.
22. MinZ recall [nm]: The lowest axial position for which $R_p(z)$ is below the half of the maximum of $R_p(z)$.
23. MaxZ recall [nm]: The highest axial position for which $R_p(z)$ is below the half of the maximum of $R_p(z)$.
24. $FWHM_Z$ recall [nm]: Full-width half-maximum of $R_p(z)$, $FWHM_Z$ = MaxZ recall - MinZ recall.

To highlight the depth performance in one unified metric, we design a new metric from the product of $FWHM_Z$ with Max recall. We call this metric **Consolidated Z-range**.

References

[1] Robert PJ Nieuwenhuizen, Keith A Lidke, Mark Bates, Daniela Leyton Puig, David Grünwald, Sjoerd Stallinga, and Bernd Rieger. Measuring image resolution in optical nanoscopy. *Nature methods*, 10(6):557–562, 2013.