ASCENT: Annotation-free Self-supervised Contrastive Embeddings for 3D Neuron Tracking in Fluorescence Microscopy

Supplementary Material

6.1. Hyperparameter Tuning on Embedding Size

To determine the optimal dimensions for the projection head and the final embedding vector, we performed a grid search over the linear layer dimension ($l \in \{256, 512, 1024, 2048\}$) and the final embedding size ($d \in \{64, 128, 256, 512\}$). We found the model's performance to be robust to these hyperparameters. The final tracking accuracies for all tested combinations fell within a narrow range of 94.90% (for $l = 256, d = 64$) to 95.10% (for $l = 1024, d = 512$). Based on this, we selected the configuration of $l = 1024$ and $d = 256$ used in the main paper as a representative choice that balances performance and dimensionality.

6.2. Implementation Details for Data Transformations

We implemented five 3D data augmentation classes adapted from standard `torchvision` (`tv`) transformations (see Fig. 7 for examples). The parameters were:

- **Random rotation:** Randomly rotates the input volume around the z-axis by an angle uniformly sampled from $[-180°, 180°]$. This operation follows the logic of `tv.transforms.RandomRotation`, applied consistently to each z-slice.
- **Random resized crop:** Performs a random crop on the XY-plane of the volume and resizes the cropped region to a fixed output size. We set the output size to $(448, 448)$, with a random crop scale between 0.6 and 1.0. This transformation extends `tv.transforms.RandomResizedCrop` to 3D data by applying the same crop and resize to all z-slices.
- **Random color jitter:** Randomly adjusts the brightness and contrast of the input. brightness=0.2 and contrast=0.2 parameters are used. This transformation is directly adapted from `tv.transforms.ColorJitter` to operate on multi-channel 3D images.
- **Random position jitter:** Applies a random translation to the neuron's coordinates within the volume to simulate localization uncertainty. The jitter ranges are set to 5.0 units in the XY-plane and 1.0 unit in the z-axis.
- **Random elastic deformation:** Elastic deformation is applied to the XY-plane of each 3D volume. A 3×3 grid of control points is generated, and a displacement field is computed by sampling displacements at each grid point from a normal distribution with mean 0 and $\sigma = 36$. The displacements are bilinearly interpolated across the image to produce a smooth deformation that warps the input.

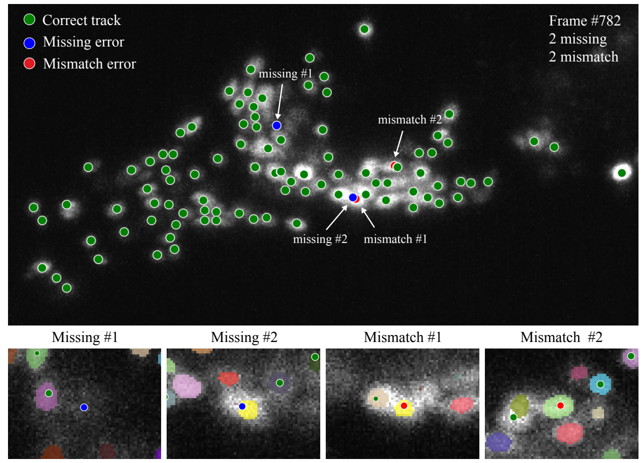

Figure 6. **Qualitative example of tracking error** The main panel shows a maximum intensity projection of an error-prone frame (#782), with ground-truth neurons colored by tracking status: correctly tracked (green), missing (blue), or mismatched (red). Bottom panels show zoomed-in views of individual errors at the relevant z-plane with the StarDist-generated segmentation mask overlaid. The analysis confirms most errors are attributable to the upstream detector, including failed detection (Missing #1), undersegmentation (Missing #2, Mismatch #1), and a single true linking failure (Mismatch #2).

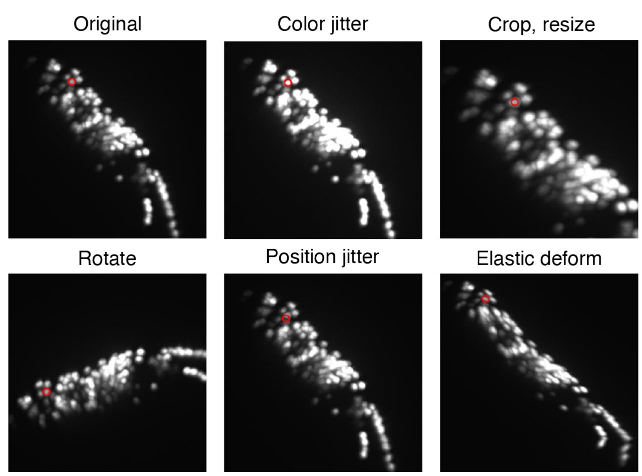

Figure 7. **Examples of random transformations.** The images are maximum intensity projections along the z-axis. The red circle marks the target object, illustrating the effect of the applied transformations.