Proceedings Track

Self-supervised cell instance segmentation through transformation-equivariant representation learning

Editors: List of editors' names

Abstract

Cell segmentation in microscopy images is a fundamental step in quantitative biological analysis, supporting tasks such as cell counting, morphological characterization, and downstream computational studies. While recent supervised deep learning methods have achieved remarkable segmentation performance, their reliance on large-scale manual annotations limits scalability across diverse imaging conditions. In this work, we present a self-supervised framework for cell instance segmentation that leverages transformation-equivariance as an inductive prior. Our method trains a neural network to produce vector fields that remain consistent under random geometric transformations. This equivariant behavior encourages structural alignment with cellular morphology, thereby enabling precise segmentation without manual labels. We demonstrate that our self-supervised method matches or exceeds pretrained supervised baselines on the LIVECell dataset and achieves strong qualitative results on additional datasets. These results highlight the potential of equivariance-driven self-supervision for label-efficient, morphology-aware segmentation. Code will be released upon acceptance.

Keywords: transformation equivariance, cell instance segmentation, microscopy images

1. Introduction

Accurate segmentation of individual cells in microscopy images is essential for quantitative analysis such as cell counting, morphological profiling, and tracking (Toyoshima et al., 2016; Caicedo et al., 2019; Franzesi et al., 2024). Supervised deep learning has achieved strong segmentation accuracy across diverse modalities and cell types by training on large curated datasets (Schmidt et al., 2018; Hollandi et al., 2020). Among existing tools, Cellpose (Stringer et al., 2021; Stringer and Pachitariu, 2025) has become the *de facto* generalist solution, performs reliably across heterogeneous imaging conditions and is widely adopted in biological and biomedical imaging. Despite this success, supervised models depend on annotated training data, and their reliability degrades when imaging conditions shift beyond the training distribution (Li et al., 2023; Han et al., 2024). Acquiring new annotations is costly and time-consuming, and this effort must be repeated whenever acquisition protocols or sample preparations change.

Self-supervised methods have been explored to reduce the need for labels, but they remain less robust in practice. Approaches based on handcrafted priors assume properties such as approximately uniform cell size or strong foreground-to-background contrast, which often fail under variable morphologies or low-contrast conditions (Wolf et al., 2023; Kochetov et al., 2024). Methods that generate pseudo-labels or rely on iterative post-processing can accumulate bias from early decisions and typically require dataset-specific hyperparameters (Miyaki et al., 2024; Baracaldo et al., 2025). As a result, their segmentation accuracy generally lags behind supervised baselines on the same datasets, limiting routine adoption.

A promising direction is to exploit geometric transformation consistency as a source of label-free supervision. The identity and morphology of a cell are preserved under rotations, reflections, and translations; therefore, a geometry-aware representation should transform predictably with the image (i.e., exhibit transformation equivariance). Since this property is tied to the underlying biological structure rather than characteristics of a specific dataset or imaging setup, it offers a general supervision signal that is robust to acquisition variations.

Building on this idea, we propose a framework that learns transformation-equivariant vector fields aligned with cellular morphology. For each input image, we generate multiple geometrically transformed variants, apply a network to each, invert the predictions back to the original coordinates, and average the results. The averaged field, which preserves features that are equivariant across transformations, serves as a self-supervised target. The network is then trained to reproduce this field from the original input, enforcing equivariance by construction. The learned vector field drives pixel aggregation to reconstruct instance masks, while an auxiliary foreground probability map suppresses spurious flows in background regions. This procedure requires no manual annotations and avoids dataset-specific assumptions beyond geometric invariances.

We evaluate the framework on diverse microscopy datasets and compare it against supervised baselines. Without labeled masks, our method produces accurate instance segmentations under high cell density, heterogeneous morphology, and low contrast. Our method matches or surpasses the performance of supervised models trained on external annotations, demonstrating that state-of-the-art accuracy can be achieved in a fully label-free setting.

2. Methods

2.1. Equivariance of untrained CNNs under transformation averaging

We begin with the observation that untrained convolutional neural networks (CNNs), when applied to randomly transformed versions of an input image, and whose outputs are inverse-transformed back and then averaged, produce structured vector fields aligned with cellular morphology in the original image. In particular, for microscopy images containing repetitive and locally symmetric objects, the averaged CNN output exhibits consistent center-directed flows over cell-like regions.

The emergence of these patterns can be explained by the equivariance properties of the transformation-averaged output. In the presence of local geometric symmetries in the input, equivariance imposes strong structural constraints on the vector field, restricting its direction to align with the center of symmetry. This interaction between input symmetry and induced equivariance accounts for the observed radial flows.

Let $\Omega \subset \mathbb{R}^2$ be the image domain. An image is $I : \Omega \to \mathbb{R}$ and a vector field is $V : \Omega \to \mathbb{R}^2$, and $I[x]$, $V[x]$ denote pointwise evaluation of I and V at coordinate x, respectively. Let $G \subset \mathrm{E}(2)$ be a finite subgroup of planar isometries. Each $g \in G$ has a coordinate map $\phi_g(x) = R_g x + t_g$ with $R_g \in O(2)$. Applying g to an image or a vector field is defined by $(g \cdot I)[x] := I[\phi_g^{-1}(x)]$, $(g \cdot V)[x] := R_g V[\phi_g^{-1}(x)]$. Let $f : \mathbb{R}^{H \times W} \to \mathbb{R}^{H \times W \times 2}$ be a CNN mapping images to vector fields. The transformation-averaged field is

$$T(I) := \frac{1}{|G|} \sum_{g \in G} (g^{-1} \cdot f(g \cdot I)), \quad T(I)[x] \in \mathbb{R}^2.$$

Proposition 1 (Transformation-equivariance of the averaged field) *For any transformation $h \in G$ and any image I, $h \cdot T(I) = T(h \cdot I)$.*

Proof By linearity,

$$h \cdot T(I) = \frac{1}{|G|} \sum_{g \in G} h \cdot (g^{-1} \cdot f(g \cdot I)) = \frac{1}{|G|} \sum_{g \in G} (h \cdot g^{-1}) \cdot f(g \cdot I).$$

Reindex with $k := g \cdot h^{-1}$ (so $g = k \cdot h$). Right-multiplication by h^{-1} is a bijection on G, hence $\{k : g \in G\} = G$. Using associativity, closure, and inverses,

$$h \cdot T(I) = \frac{1}{|G|} \sum_{k \in G} (h \cdot (k \cdot h)^{-1}) \cdot f((k \cdot h) \cdot I) = \frac{1}{|G|} \sum_{k \in G} k^{-1} \cdot f(k \cdot (h \cdot I)) = T(h \cdot I).$$

■

Proposition 2 (Local symmetry forces radial alignment) *Let $H \subset G$ contain rotations about a point x_c and reflections across lines through x_c. If the image is invariant under H (i.e., $h \cdot I = I$ for all $h \in H$), then for $V := T(I)$,*

$$V[x] \in \mathit{span}(x_c - x) \text{ for all } x \neq x_c.$$

Proof By Proposition 1 and $h \cdot I = I$, we have $h \cdot V = V$ for all $h \in H$.

Let ρ be any rotation about x_c and write $x_\rho := \phi_\rho(x) = x_c + R_\rho(x - x_c)$. Then,

$$V[x_\rho] = (\rho \cdot V)[x_\rho] = R_\rho V[x], \quad x_c - x_\rho = R_\rho(x_c - x).$$

Rotations preserve oriented angles, so

$$\angle(V[x_\rho], x_c - x_\rho) = \angle(R_\rho V[x], R_\rho(x_c - x)) = \angle(V[x], x_c - x).$$

Thus, the angle to the radial direction is constant along the rotation orbit of x.

Let σ be the reflection across the line through x_c. Since reflections reverse orientation,

$$\angle(V[\phi_\sigma(x)], x_c - \phi_\sigma(x)) = \angle(R_\sigma V[x], R_\sigma(x_c - x)) = -\angle(V[x], x_c - x).$$

Since rotational symmetry enforces a constant angle, while reflection symmetry negates it, the only consistent angle is 0 or π. Therefore, $V[x]$ is collinear with $x_c - x$ for all $x \neq x_c$.

■

As shown in Figures 1 and S3, when applied to a sample microscopy image, the averaged vector field $\bar{V}(x)$ exhibits smooth, center-directed patterns that align closely with individual cells. This phenomenon is consistent across different architectures (e.g., sequential CNN, U-Net (Ronneberger et al., 2015), and MAnet (Li et al., 2021)) and becomes more pronounced as the number of sampled transformations increases.

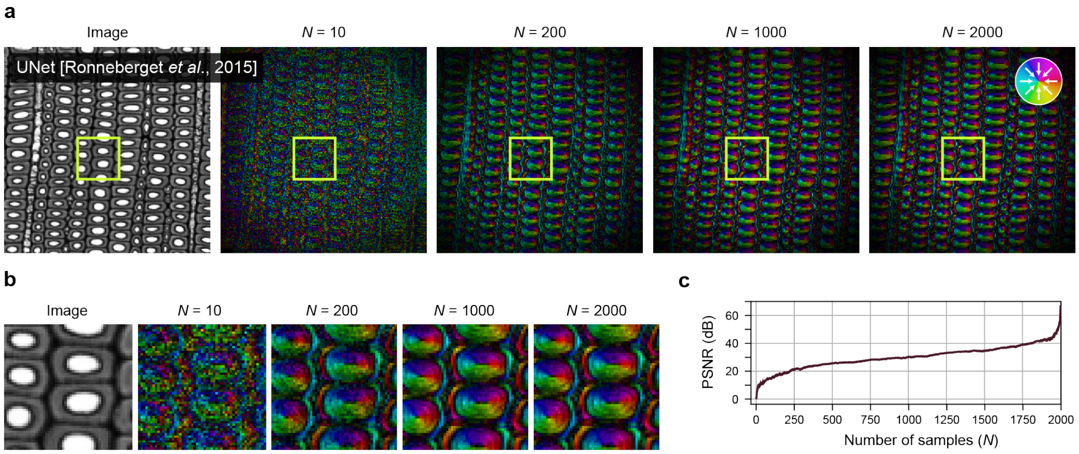

Figure 1: Averaged outputs of untrained CNNs under random rotations. (a) As the number of transformed samples N increases, the averaged vector fields converge to center-seeking patterns. (b) Zoomed-in views of the region in (a), showing progressive refinement of cell-like, center-directed structures. (c) PSNR curves computed against the result at $N = 2000$, showing convergence as N increases.

2.2. Self-supervised learning of transformation-equivariant vector fields

The observation in Section 2.1 motivates our overall strategy: rather than relying on a fixed, untrained CNN, we use the transformation-averaged output as a dynamic self-supervised target, and train a CNN to predict such fields from a single image. Specifically, the network outputs a normalized vector $V[x] \in \mathbb{R}^2$ at each pixel, constrained to unit norm. This normalization resolves the scale ambiguity of the transformation-averaged target, which constrains only the direction of the vectors. The training objective enforces consistency between these predicted vectors and the network's own transformation-averaged output, leading to center-directed and transformation-consistent predictions.

However, since the vector field encodes only direction, background regions may produce arbitrary or noisy vectors that interfere with downstream grouping. To mitigate this, we introduce a foreground probability mask M, indicating where the directional signals are reliable. This leads to a joint prediction $f_\phi(I) = (V_{\phi_1}(I), M_{\phi_2}(I))$, where $\phi = (\phi_1, \phi_2)$ denotes the trainable parameters. Here, $V_{\phi_1}(I) \in \mathbb{R}^{H\times W\times 2}$ is the unit vector field (two channels), and $M_{\phi_2}(I) \in \mathbb{R}^{H\times W\times 1}$ is the foreground probability map (one channel), obtained with a sigmoid activation.

Transformation consistency is enforced on both outputs using a set of N randomly sampled transformations $\{g_1, \ldots, g_N\} \subset G$. The self-supervised targets are computed as empirical averages over the inverse-aligned predictions (Figure 2a):

$$\bar{V}(I) := \frac{1}{N}\sum_{i=1}^{N} g_i^{-1} \cdot V_{\phi_1}(g_i \cdot I), \quad \bar{M}(I) := \frac{1}{N}\sum_{i=1}^{N} g_i^{-1} \cdot M_{\phi_2}(g_i \cdot I).$$

The equivariance-based losses are as follows, with gradients stopped through the targets:

$$\mathcal{L}_{\text{vf}} = ||V_{\phi_1}(I) - \bar{V}||_2^2, \quad \mathcal{L}_{\text{mask}} = ||M_{\phi_2}(I) - \bar{M}||_2^2.$$

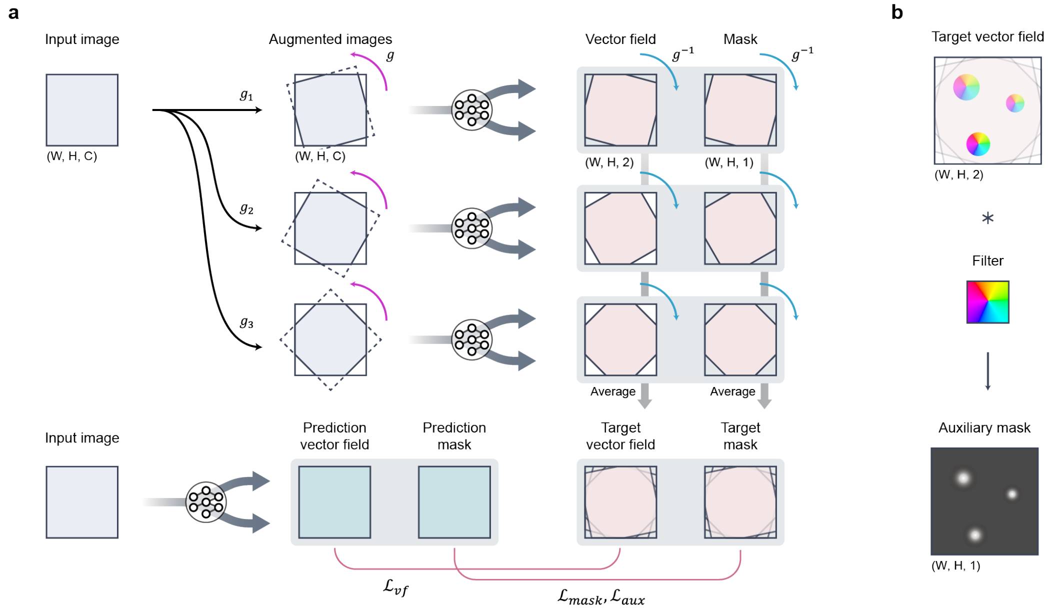

Figure 2: Overview of the proposed self-supervised training framework. (a) Multiple transformed views of the input image are passed through a shared network. The predicted vector fields and masks are inverse-transformed and averaged to form self-supervised targets. The model is trained to match both targets using equivariance-based losses. (b) The target vector field is convolved with a fixed filter to extract an auxiliary mask used for training.

While $\bar{M}$ provides weak supervision for the mask, it can be ambiguous in early training or in low-signal regions. To reinforce structural consistency, we derive a binary pseudo-label from the vector field as an additional training signal. Specifically, the averaged vector field $\bar{V}$ is convolved with a fixed center-directing filter K, $\bar{M}_{\text{aux}}(x) := K * \bar{V}(x)$ (Figure 2b).

This produces a smooth saliency map that highlights regions with convergent flow, indicative of cellular structures. After min-max normalization of $\bar{M}_{aux}$ to $[0, 1]$, the predicted mask is trained to match this auxiliary target using a binary cross-entropy loss:

$$\mathcal{L}_{\text{aux}} = \text{BCE}(M_{\phi_2}(I),\, \bar{M}_{\text{aux}}).$$

The final training objective is $\mathcal{L} = \mathcal{L}_{\text{vf}} + \lambda_{\text{mask}}\mathcal{L}_{\text{mask}} + \lambda_{\text{aux}}\mathcal{L}_{\text{aux}}$. This formulation enables the network to learn transformation-equivariant vector fields while reducing ambiguity by learning the mask on structurally convergent regions in a fully self-supervised manner.

2.3. Post-processing of vector field

To obtain discrete instance masks after training, we adopt a flow-based grouping strategy as in Cellpose (Stringer et al., 2021). For each pixel x with $M_{\phi_2}(x) > \tau$, we iteratively follow the direction of $V_{\phi_1}(x)$ for a fixed number of steps. Pixels that converge to the same spatial location are assigned the same instance label. This post-processing step converts the dense, continuous vector field and confidence mask into discrete cell instances. By reusing the established grouping scheme, we ensure that differences in performance can be attributed to our self-supervised representation rather than post-processing heuristics.

3. Experiments

Datasets. We evaluated our method on the LIVECell dataset (Edlund et al., 2021), the Cellpose dataset (Stringer et al., 2021), and a collection of real-world microscopy images from publicly available online sources. The LIVECell dataset contains phase-contrast images under diverse culture conditions, reflecting variation in shape and density. The Cellpose dataset includes fluorescence, brightfield, and phase-contrast images. We also tested on microscopy images from sources such as NCI Visuals Online[1], Science Photo Library[2], and biomedical outreach platforms[3]. These cover various cell types (e.g., HeLa, lung cancer, epithelial tissues) acquired under different imaging setups. We focus primarily on the LIVECell dataset, as it provides expert-validated ground truth with standardized train/test splits, while other datasets are used for qualitative comparison.

Baselines. We compared our method against three widely used supervised methods, Cellpose (Stringer et al., 2021), Stardist (Schmidt et al., 2018), and NucleAIzer (Hollandi et al., 2020). Cellpose was run with the official pretrained model (v0.6) and automatic scale estimation enabled. Stardist was performed using the official GitHub implementation with default settings and the pretrained 2D fluorescence nuclei model. NucleAIzer predictions were obtained from the authors' online service[4], using the "General - nuclei" preset.

Model and training configuration. We used a standard U-Net (Ronneberger et al., 2015) with depth of 4. The final layer branches into two heads: one outputs a 2-channel vector field, and the other outputs a 1-channel foreground mask with sigmoid activation to produce confidence scores.

Each model was trained on each image independently, without manual annotations. Input images were cropped into 512×512 patches. At each iteration, 16 random transformations were sampled to construct the transformation-averaged self-supervised targets as described in Section 2.2. Transformations included random rotations (uniformly sampled from $[-180°, 180°]$), horizontal/vertical flips, and translations (within $\pm 5\%$ of image size) (Jaderberg et al., 2015; Cho et al., 2023). The transformation-averaged self-supervised targets were detached from the network prior to loss calculation. Adam optimizer was used with a learning rate of 3×10^{-4} for 5,000 iterations. All experiments were conducted on a workstation with an Intel Xeon CPU and an NVIDIA RTX 3090 GPU.

4. Results

4.1. Quantitative and qualitative evaluation on LIVECell dataset

We first evaluated our method on the LIVECell dataset, which contains images with high variability in cell shape and density. Figure 3 presents representative segmentation results. Our model accurately captures cell boundaries and instance separation without access to ground-truth supervision during training. Compared to baselines, it produces more coherent boundaries in densely clustered regions. In contrast, Stardist frequently yielded near-empty

[1] https://visualsonline.cancer.gov/details.cfm?imageid=10655
[2] https://www.sciencephoto.com/media/464034/view/hela-cells-light-micrograph
[3] https://www.flickr.com/photos/146824358@N03/35649178794/
[4] http://www.nucleaizer.org/

predictions, while NucleAIzer missed many cells or produced merged instances. Cellpose performed reliably in sparse regions but often failed to detect individual cells in dense areas.

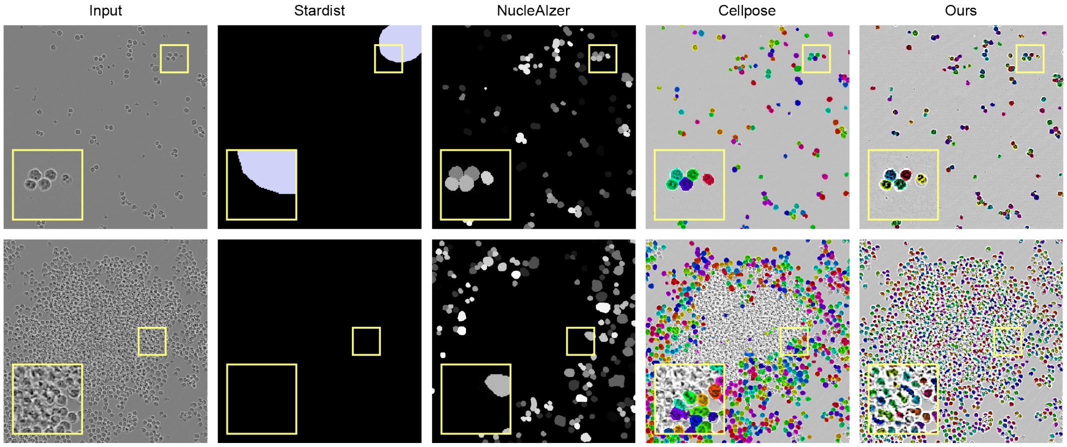

Figure 3: Qualitative segmentation results on LIVECell dataset. Each row shows a low-density image (top) and a high-density (bottom) image. From left to right: input image, Stardist, NucleAIzer, Cellpose, and ours.

For quantitative evaluation, we report average precision across cell-matching thresholds, the primary metric adopted in prior work (Stringer and Pachitariu, 2025), averaged over all test images. To assess robustness under different conditions, the test set was split into low- and high-density subsets, each containing 15 images. As shown in Figure 4 and Table 1, our method achieves stronger results than the supervised baselines in both regimes.

4.2. Qualitative evaluation on Cellpose dataset and in-the-wild images

We next evaluated our method qualitatively on the Cellpose dataset, which spans fluorescence, brightfield, and phase-contrast modalities. Figure 5a shows representative segmentation results. Despite large variations in intensity profiles and cell morphology, our method consistently separated individual cells without over-segmentation.

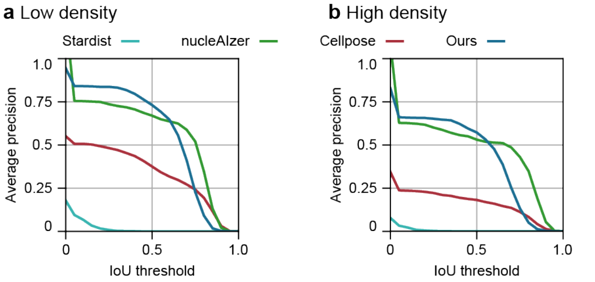

Figure 4: Segmentation accuracy across varying IoU thresholds on the low- (left) and high-density (right) subsets from LIVECell dataset.

	Low density	High density
Stardist	0.000	0.000
NucleAIzer	0.669	0.530
Cellpose	0.375	0.181
Ours	**0.731**	**0.572**

Table 1: Segmentation accuracy at an IoU threshold of 0.5 for each method.

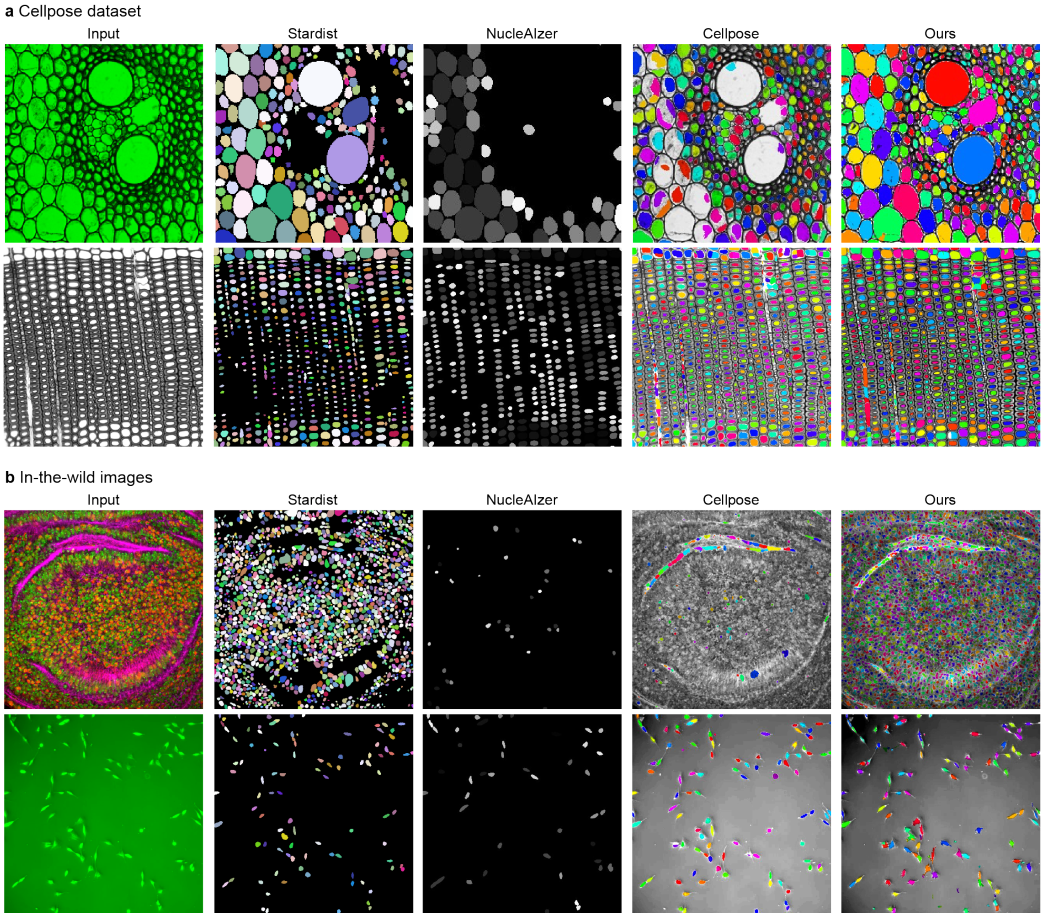

Figure 5: Qualitative segmentation results on (a) Cellpose and (b) in-the-wild images. From left to right: input image, Stardist, NucleAIzer, Cellpose, and ours.

To further test robustness, we applied the model to microscopy images collected from heterogeneous, publicly accessible sources. As shown in Figure 5b, our method produced coherent instance masks even under large variations in contrast, resolution, and imaging modality. These results highlight the robustness of the proposed transformation-equivariant framework in diverse, unlabeled microscopy settings.

5. Discussion

We introduced a self-supervised framework for instance-level cell segmentation that leverages transformation-equivariant vector fields. By averaging CNN outputs over randomly transformed inputs, we observe consistent center-directed flows arising from the local symmetry of cells in microscopy images. This structural regularity was exploited as a self-supervision signal, allowing the model to learn instance representations without labeled data. Our approach achieves performance that is competitive with supervised baselines across diverse imaging conditions. These findings demonstrate that geometric symmetries inherent in cellular structures can serve as an effective alternative to explicit supervision.

References

Laura Baracaldo, Blythe King, Haoran Yan, Yizi Lin, Nina Miolane, and Mengyang Gu. Unsupervised cell segmentation by fast gaussian processes. *arXiv preprint arXiv:2505.18902*, 2025.

Juan C Caicedo, Allen Goodman, Kyle W Karhohs, Beth A Cimini, Jeanelle Ackerman, Marzieh Haghighi, CherKeng Heng, Tim Becker, Minh Doan, Claire McQuin, Mohammad Rohban, Shantanu Singh, and Anne E Carpenter. Nucleus segmentation across imaging experiments: the 2018 data science bowl. *Nature methods*, 16(12):1247–1253, 2019.

Junmo Cho, Seungjae Han, Eun-Seo Cho, Kijung Shin, and Young-Gyu Yoon. Robust and efficient alignment of calcium imaging data through simultaneous low rank and sparse decomposition. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1939–1948, 2023.

Christoffer Edlund, Timothy R Jackson, Nabeel Khalid, Nicola Bevan, Timothy Dale, Andreas Dengel, Sheraz Ahmed, Johan Trygg, and Rickard Sjögren. Livecell—a large-scale dataset for label-free live cell segmentation. *Nature methods*, 18(9):1038–1045, 2021.

Giovanni Talei Franzesi, Ishan Gupta, Ming Hu, Kiryl Piatkevich, Murat Yildirim, Jian-Ping Zhao, Minho Eom, Seungjae Han, Demian Park, Himashi Andaraarachchi, Zhaohan Li, Jesse Greenhagen, Amirul Muhammad Islam, Parth Vashishtha, Zahid Yaqoob, Nikita Pak, Alexander D. Wissner-Gross, Daniel A. Martin-Alarcon, Jonathan J. Veinot, Peter T. C. So, Uwe Kortshagen, Young-Gyu Yoon, Mriganka Sur, and Edward S. Boyden. In vivo optical clearing of mammalian brain. *bioRxiv*, pages 2024–09, 2024.

Seungjae Han, Joshua Yedam You, Minho Eom, Sungjin Ahn, Eun-Seo Cho, and Young-Gyu Yoon. From pixels to information: Artificial intelligence in fluorescence microscopy. *Advanced Photonics Research*, 5(9):2300308, 2024.

Reka Hollandi, Abel Szkalisity, Timea Toth, Ervin Tasnadi, Csaba Molnar, Botond Mathe, Istvan Grexa, Jozsef Molnar, Arpad Balind, Mate Gorbe, Maria Kovacs, Ede Migh, Allen Goodman, Tamas Balassa, Krisztian Koos, Wenyu Wang, Juan Carlos Caicedo, Norbert Bara, Ferenc Kovacs, Lassi Paavolainen, and Peter Horvath. nucleaizer: a parameter-free deep learning framework for nucleus segmentation using image style transfer. *Cell systems*, 10(5):453–458, 2020.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. *Advances in neural information processing systems*, 28, 2015.

Bogdan Kochetov, Phoenix D. Bell, Paulo S. Garcia, Akram S. Shalaby, Rebecca Raphael, Benjamin Raymond, Brian J. Leibowitz, Karen Schoedel, Rhonda M. Brand, Randall E. Brand, Jian Yu, Lin Zhang, Brenda Diergaarde, Robert E. Schoen, Aatur Singhi, and Shikhar Uttam. Unseg: unsupervised segmentation of cells and their nuclei in complex tissue samples. *Communications Biology*, 7(1):1062, 2024.

Rui Li, Shunyi Zheng, Ce Zhang, Chenxi Duan, Jianlin Su, Libo Wang, and Peter M Atkinson. Multiattention network for semantic segmentation of fine-resolution remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–13, 2021.

Xinyang Li, Yuanlong Zhang, Jiamin Wu, and Qionghai Dai. Challenges and opportunities in bioimage analysis. *Nature methods*, 20(7):958–961, 2023.

Shintaro Miyaki, Shori Nishimoto, Yuta Tokuoka, Takahiro G Yamada, Takashi Morikura, and Akira Funahashi. Cell segmentation without annotation by unsupervised domain adaptation based on cooperative self-learning. *bioRxiv*, pages 2024–07, 2024.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *International conference on medical image computing and computer-assisted intervention*, pages 265–273. Springer, 2018.

Carsen Stringer and Marius Pachitariu. Cellpose3: one-click image restoration for improved cellular segmentation. *Nature methods*, 22(3):592–599, 2025.

Carsen Stringer, Tim Wang, Michalis Michaelos, and Marius Pachitariu. Cellpose: a generalist algorithm for cellular segmentation. *Nature methods*, 18(1):100–106, 2021.

Yu Toyoshima, Terumasa Tokunaga, Osamu Hirose, Manami Kanamori, Takayuki Teramoto, Moon Sun Jang, Sayuri Kuge, Takeshi Ishihara, Ryo Yoshida, and Yuichi Iino. Accurate automatic detection of densely distributed cell nuclei in 3d space. *PLoS computational biology*, 12(6):e1004970, 2016.

Steffen Wolf, Manan Lalit, Katie McDole, and Jan Funke. Unsupervised learning of object-centric embeddings for cell instance segmentation in microscopy images. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 21263–21272, 2023.

Appendix A. Illustrative example of transformation-equivariance

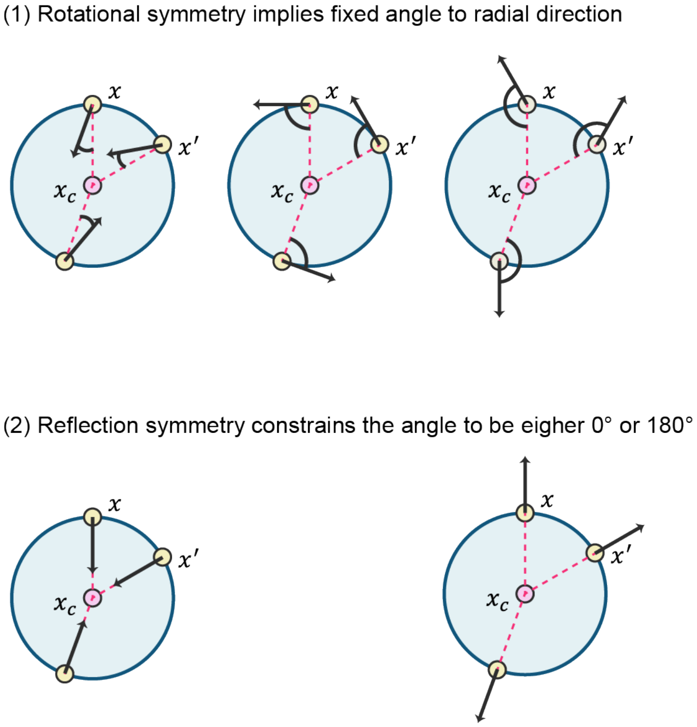
(1) Rotational symmetry implies fixed angle to radial direction
x
x'
x_c
(2) Reflection symmetry constrains the angle to be eigher 0° or 180°

Figure S1: **Radial alignment induced by rotational and reflectional symmetry.** Illustration corresponding to Section 2.1. When the image exhibits rotational symmetry about x_c, the angle between $[\bar{V}]_x$ and the radial vector $x_c - x$ remains constant across all x. When reflection symmetry is additionally present, this angle is restricted to either 0° or 180°.

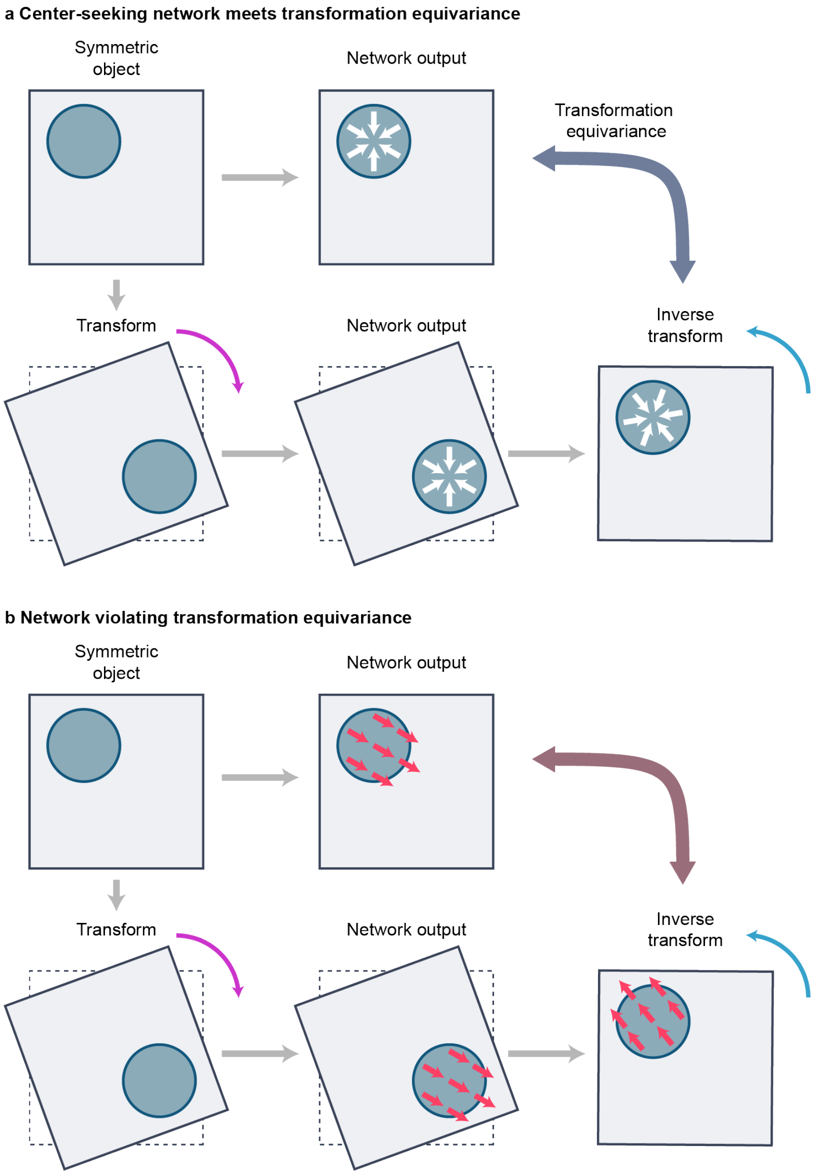

Figure S2: **Illustration of transformation equivariance in vector field prediction**. **a** A center-seeking network meets transformation equivariance. **b** If the model produces a vector field pointing to a non-center region, the inverse-transformed outputs from a transformed input become inconsistent, violating equivariance. The rotation case is shown; the same applies to flips and translations.

Appendix B. Averaged outputs of untrained CNNs under random rotations

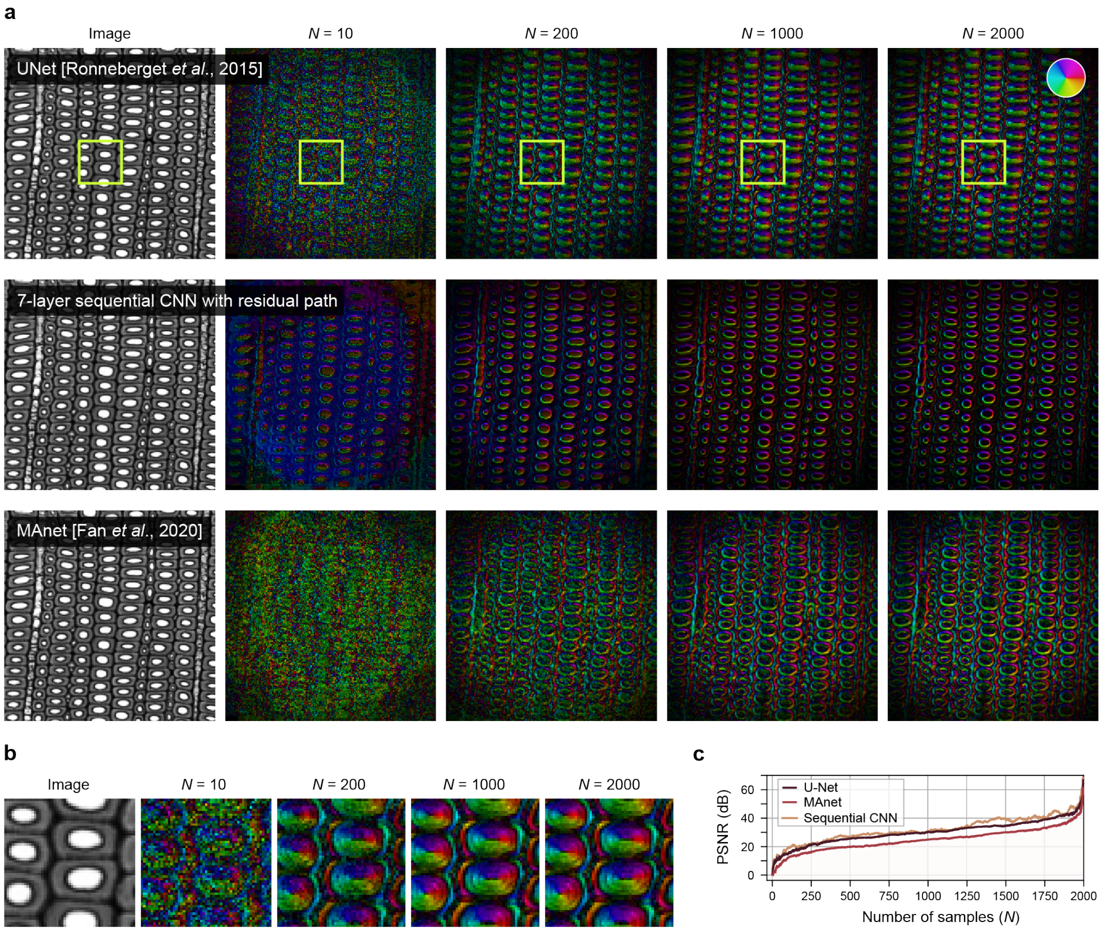

Figure S3: **Averaged outputs of untrained CNNs under random rotations.** (a) As the number of transformed samples N increases, the averaged vector fields converge to center-seeking patterns. (b) Zoomed-in views of the region in (a), showing progressive refinement of cell-like, center-directed structures. (c) PSNR curves computed against each architecture's result at $N = 2000$, showing convergence as N increases.