Cell Identification from Partial Observation Using Spatiotemporal Attention in *Caenorhabditis elegans* Embryogenesis*

Henry Xue[1], Dali Wang[2,3]**, Erin Haus[4], Anthony Santella[4], and Zhirong Bao[4]

[1] Science Research Program, Horace Greeley High School, Chappaqua, NY, USA
[2] Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA
[3] Biological and Environmental Systems Science, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[4] Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

Abstract. Automated cell identification enables systematic phenotyping in developing embryos. Existing methods typically require nearly complete cellular configurations, but experimental constraints including phototoxicity and tissue-specific labeling often limit observation to cell subsets. We present a method that identifies *Caenorhabditis elegans* embryonic cells from partial observations containing 5–20 nuclei. Each cell is represented by geometric features computed from its local neighborhood, including pairwise relational statistics derived from the distance matrix. A joint attention encoder learns context-dependent embeddings, and cells are identified by nearest-neighbor lookup in a manifold built from training data. The model is trained on simulated embryos and fine-tuned on real data following the Twin Attention transfer learning approach. The method achieves 88.4% top-1 accuracy on held-out simulated embryos and 80.9% on real embryos after fine-tuning, with top-5 accuracy reaching 99.0% and 98.1% respectively. When predictions are filtered by confidence margin, accuracy reaches 95.8% for the top half of predictions. Most errors occur among spatially adjacent sister cells, consistent with biological expectations. These results demonstrate that partial cellular neighborhoods contain sufficient positional information for reliable identity assignment, enabling rapid cell identification in biomarker experiments and live imaging scenarios where complete observation is impractical.

Keywords: Cell Identification · Partial Observations · *C. elegans* · Transformer Networks · Point Cloud Matching · Developmental Biology

* This study was partly supported by NIH grants 1R01GM152927 to the Memorial Sloan Kettering Cancer Center and the University of Tennessee, Knoxville.
** Corresponding author. Email: dwang7@utk.edu

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

1 Introduction

The *Caenorhabditis elegans* embryo develops through a fully documented cell lineage: 671 cells are generated during embryogenesis, of which 113 undergo programmed death, leaving 558 cells traceable to the single-cell zygote [1]. Automated lineage tracing [4,5] now routinely generates complete records, enabling quantitative phenotype analysis at single-cell resolution [2,3]. However, identifying which cell is which remains a bottleneck: manual annotation requires hundreds of hours per genetic screen [6].

Experimental constraints frequently limit observable cells. Tissue-specific promoters label only cell subsets [7], while ubiquitous markers require intense illumination causing phototoxicity [3,8]. Even with ubiquitous markers, automated segmentation achieves 80–99% recall [4,5]. In both scenarios, the observer sees a subset and must determine which cells they are.

Recent deep learning methods address cell correspondence but assume complete configurations. Twin Attention [9] achieves ∼90% accuracy using a Transformer on pairs of complete embryonic point clouds, outperforming previous methods at ∼60%. fDNC [10] achieved 79.1% tracking neurons via semi-synthetic training. Both assume complete or nearly complete observations.

The partial observation setting introduces challenges beyond fewer cells. Complete embryos provide global spatial anchors—the anterior-posterior axis and stereotyped founder cell positions—that constrain matching. With only 5–20 cells visible, these anchors are often absent, and standard methods assuming complete point clouds (ICP [11], CPD [12], graph matching [13]) or independent point encoding (PointNet [14]) are ill-suited, as we discuss in Section 4.

We hypothesized that partial observations of 5–20 cells contain sufficient positional information for reliable identification at >85% accuracy, drawing on the biological logic that cell identity is determined by lineage and position [1], and local spatial relationships are constrained by physical packing within the eggshell. Our method extends the Twin Attention framework to partial observations through three modifications: (1) five geometric features characterizing each cell by relationships to other observed cells; (2) a joint attention encoder where each cell's embedding depends on neighborhood context; and (3) simulation-based pretraining followed by fine-tuning on real embryo data.

2 Materials and Methods

2.1 Problem Setting and Dataset

Given a query subset $Q = \{(c_i, \mathbf{x}_i)\}_{i=1}^{n_q}$ containing $n_q \in [5, 20]$ cells with 3D coordinates $\mathbf{x}_i \in \mathbb{R}^3$ and unknown identities c_i, and a reference manifold $\mathcal{M}$ of cell embeddings with known identities, we assign identities to each query cell. The user provides query coordinates and an estimate of the total cell count (developmental stage), typically available from brightfield imaging. No complete embryo observation or tracking from earlier timepoints is required.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

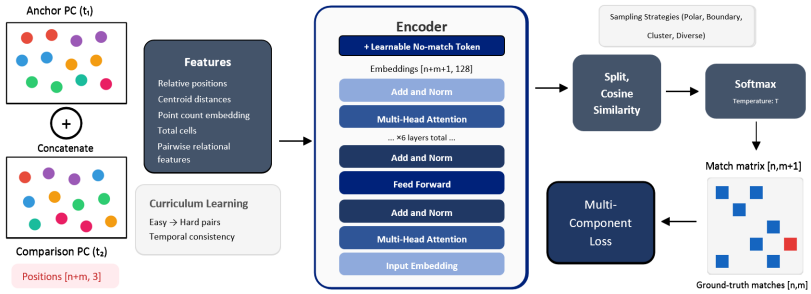

Fig. 1. Method overview. Pairs of cell subsets are sampled from simulated embryos. Five geometric features are computed per cell. Reference and query embeddings are concatenated with a no-match token and processed through a joint attention Transformer encoder. Cells are matched via cosine similarity.

Training data consisted of 300 simulated embryos (270 training, 30 validation) spanning 8–194 cell stages, generated using an agent-based model that draws from measured distributions of cell cycle length and migration paths [15]. For fine-tuning and evaluation, we used 50 real embryos from the Twin Attention dataset [16] (40 for fine-tuning, 10 held-out test), with manually curated positions established using AceTree [17].

2.2 Model Architecture

Our architecture follows the Transformer encoder [18] with modifications for partial point cloud processing (Figure 1). The model uses 6 layers, 8 attention heads, embedding dimension 128, feedforward dimension 512, and dropout 0.10, totaling 1,250,049 parameters.

2.3 Geometric Features

Rather than raw 3D coordinates, which depend on absolute embryo position, we compute five geometric features that remain informative under partial observation. The direct coordinate embedding from Twin Attention [9] achieved only 70.3% for subsets (Section 3.3), because a 10-cell subset's centroid and scale are determined by whichever cells happen to be sampled, unlike a complete embryo where normalization is stable.

The five feature groups, concatenated to form 128-dimensional input embeddings, are: (1) *Relative position to centroid* ($\mathbf{r}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, 20 dim.), removing absolute position dependence; (2) *Centroid distance* ($d_i = \|\mathbf{r}_i\|$, 16 dim.), encoding central vs. peripheral position; (3) *Subset size embedding* (16 dim.); (4) *Total cell count embedding* (24 dim.), providing developmental stage context; and (5) *Pairwise relational features* (52 dim.)—local density, mean distance, minimum neighbor distance, distance standard deviation, and centrality rank, computed from the pairwise distance matrix.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

2.4 Joint Attention Encoder and No-Match Token

Following Twin Attention's co-processing strategy [9], we concatenate reference and query embeddings as a single sequence: $[\mathbf{z}_1^{\text{ref}}, \ldots, \mathbf{z}_{n_r}^{\text{ref}}, \mathbf{z}_{\text{nm}}, \mathbf{z}_1^{\text{qry}}, \ldots, \mathbf{z}_{n_q}^{\text{qry}}]$, with token type embeddings distinguishing reference from query. The Transformer enables inter-sample attention, so each cell's embedding depends on cells in both sets. The learnable no-match token $\mathbf{z}_{\text{nm}} \in \mathbb{R}^D$ handles query cells with no corresponding reference cell; without it, the model forces incorrect assignments.

2.5 Training Strategy

Training pairs are generated by sampling two subsets from the same embryo at different timepoints (50%) or from different embryos at similar stages (50%). Subsets of 5–20 cells use random sampling (60%) and biologically-informed strategies (farthest point, cluster, boundary, polar sampling, each 10%). Task difficulty increases progressively over 150 epochs via curriculum learning. The primary loss is cross-entropy over cell matches with learnable temperature τ:

$$\ell_i = -\log \frac{\exp(\mathbf{e}_i^{\text{qry}} \cdot \mathbf{e}_{y_i}^{\text{ref}}/\tau)}{\sum_j \exp(\mathbf{e}_i^{\text{qry}} \cdot \mathbf{e}_j^{\text{ref}}/\tau)} \tag{1}$$

Two auxiliary losses improve discrimination: hard negative mining upweights loss from geometrically nearby cells, and temporal smoothness encourages consistent embeddings across adjacent timepoints. We train with AdamW [19] (lr 2×10^{-4}, OneCycleLR), then fine-tune on 40 real embryos (lr 2×10^{-5}, early stopping) following Twin Attention's transfer learning methodology [9]. Training requires ∼24 hours on CPU. Code: https://github.com/daliwang/Cell_ID_Xue.

2.6 Deployment

A reference manifold is constructed offline by embedding each cell identity using multiple random reference subsets, aggregated via trimmed mean. Query cells are embedded jointly with a reference subset, then identified via weighted k-nearest-neighbor voting ($k = 30$) with developmental stage filtering. Inference requires <2 seconds per query on CPU.

3 Results

3.1 Overall Performance

The model achieved 88.4% top-1 accuracy on held-out simulated embryos (Table 1), exceeding our 85% hypothesis, with top-5 accuracy of 99.0%. After fine-tuning on 40 real embryos, accuracy reached 80.9% on 10 held-out real embryos [16] (top-5: 98.1%). We additionally compute a confidence margin—the gap between the top and second-highest vote scores. Filtering to predictions

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

whose margin exceeds the median yields 95.8% accuracy on the top half of predictions, providing a practical mechanism for flagging unreliable assignments for manual review.

Table 1. Cell Identification Accuracy.

Dataset	Top-1 (%)	Top-5 (%)	High-Conf. (%)
Simulated (held-out)	88.4	99.0	95.8
Real (fine-tuned, $n=10$)	80.9	98.1	94.1

3.2 Performance by Stage and Subset Size

Accuracy varied with developmental stage: early stages (5–50 cells) achieved the highest accuracy where each cell occupies a unique position, while late stages (101–194 cells) showed reduced accuracy reflecting increased cell density and more similar local neighborhoods. Accuracy was robust across the 5–20 cell query range, with 10–15 cells providing an optimal balance between accuracy and acquisition speed.

3.3 Ablation Studies

Ablation experiments on simulated data assessed the contribution of each component (Table 2). The geometric feature set proved essential: using raw coordinates instead of our engineered features reduced accuracy by 18.1 percentage points to 70.3%, confirming that relational features provide stronger identity cues than absolute position. This raw-coordinate configuration is analogous to a PointNet-like [14] independent encoding baseline and demonstrates the necessity of our geometric feature design. Hard negative mining contributed 7.3 pp by encouraging discrimination between nearby cells, the no-match token added 6.3 pp by handling missing correspondences, and curriculum learning contributed 5.0 pp through progressive difficulty increase.

Table 2. Ablation Study Results (Simulated Data).

Configuration	Accuracy (%)	Δ (pp)
Full model	88.4	—
Raw coordinates (no geometric features)	70.3	−18.1
Without hard negative mining	81.1	−7.3
Without no-match token	82.1	−6.3
Without curriculum learning	83.4	−5.0

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

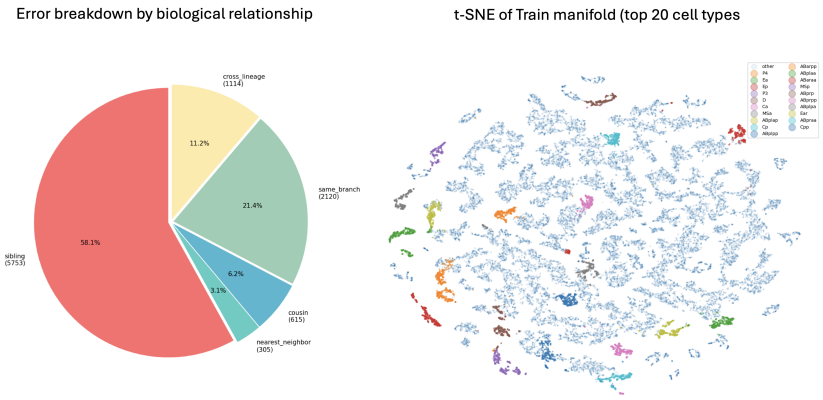

Fig. 2. Error analysis and Embedding Visualization. (Left) Distribution of misclassifications by biological relationship. Sibling confusions dominate (58.1%), consistent with biological expectations.(Right) t-SNE visualization of learned embeddings from simulated data, colored by founder lineage. (AB, MS, E, C, D, P). Cells cluster by lineage with related cells appearing nearby.

3.4 Error Analysis and Embedding Visualization

Analysis of misclassifications revealed biologically interpretable failure modes (Figure 2(left)). Sibling errors—confusion between cells differing only in their last division—constituted 58.1% of all errors. Same-branch errors (same founder lineage, more distant relationship) constituted 21.4%, cross-lineage errors 11.2%, cousin errors 6.2%, and nearest-neighbor errors 3.1%. This structure matches biological expectation: position-based identification fails most for cells that are genuinely closest in space. Immediately after division, sister cells occupy adjacent positions and share nearly identical local neighborhoods, posing a fundamental ceiling that morphological features (nuclear shape, size) or molecular markers may help address in future work.

Learned embeddings visualized via t-SNE show interpretable structure (Figure 2 (right)). Cells cluster by founder lineage (AB, MS, E, C, D, P), with related cells—siblings and cousins—appearing nearby in embedding space. This organization reflects both lineage relationships and spatial proximity in the embryo, consistent with the manifold properties that allow for reliable cell identification.

3.5 Robustness

Testing with added Gaussian positional noise (without retraining) yielded 86.2% at $\sigma = 0.1\times$ mean nearest-neighbor distance and 82.3% at $\sigma = 0.2\times$. Adding 1–3 spurious points yielded 79.7% on true cells, with the no-match token absorbing false detections.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

4 Discussion

We have demonstrated that cell identification from partial observations achieves 88.4% accuracy on simulated data and 80.9% on real embryos, with high-confidence filtering reaching 95.8%. These results confirm that partial cellular neighborhoods contain sufficient positional information for reliable identification.

Standard point cloud registration (ICP [11], CPD [12]) assumes both point sets represent the same underlying shape; when only 5–20 of 200 cells are observed, the query is a fundamentally different spatial object, not a corrupted version of the reference. Graph matching [13] does not adapt representations based on which cells co-occur, and PointNet-like architectures [14] encode each point cloud independently, lacking cross-attention. Our raw-coordinate ablation (70.3%, Table 2) serves as the closest independent-encoding baseline; the 18.1 pp gap to 88.4% quantifies the advantage of relational features with joint attention.

Sibling errors (58.1%) reflect genuine limits of position-based identification: sister cells share developmental history and occupy adjacent positions by necessity. For biomarker experiments, this is favorable—siblings typically share expression patterns from their mother cell, so misidentification between sisters has limited biological impact. Cross-lineage errors (11.2%), which would assign cells to entirely different developmental programs, are more consequential but rare, supporting reliability for biological applications.

The 80.9% real-embryo accuracy reflects a sim-to-real gap that additional annotated training data would help close. The architecture is organism-agnostic: zebrafish, *Drosophila*, ascidians, and early mouse embryos [20] face similar challenges. The relational features are invariant to absolute scale and orientation, generalizing across imaging platforms. Extension to mutant embryos is a key future direction for genetic screens at single-cell resolution.

Acknowledgments. This study was partly supported by NIH grant 1R01GM152927 to the Memorial Sloan Kettering Cancer Center and The University of Tennessee, Knoxville.

References

1. Sulston, J.E., Schierenberg, E., White, J.G., Thomson, J.N.: The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. **100**(1), 64–119 (1983). https://doi.org/10.1016/0012-1606(83)90201-4
2. Moore, J.L., Du, Z., Bao, Z.: Systematic quantification of developmental phenotypes at single-cell resolution during embryogenesis. Development **140**(15), 3266–3274 (2013). https://doi.org/10.1242/dev.096040
3. Du, Z., Santella, A., Fei, H., Tiongson, M., Bao, Z.: De novo inference of systems-level mechanistic models of development from live-imaging-based phenotype analysis. Cell **156**(1–2), 359–372 (2014). https://doi.org/10.1016/j.cell.2013.11.046
4. Bao, Z., Murray, J.I., Boyle, T., Ooi, S.L., Sandel, M.J., Waterston, R.H.: Automated cell lineage tracing in *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **103**(8), 2707–2712 (2006). https://doi.org/10.1073/pnas.0511111103

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16

5. Santella, A., Du, Z., Bao, Z.: A semi-local neighborhood-based framework for probabilistic cell lineage tracing. BMC Bioinformatics **15**, 217 (2014). https://doi.org/10.1186/1471-2105-15-217
6. Murray, J.I., Bao, Z., Boyle, T.J., Waterston, R.H.: The lineaging of fluorescently-labeled *Caenorhabditis elegans* embryos with StarryNite and AceTree. Nat. Protoc. **1**, 1468–1476 (2006). https://doi.org/10.1038/nprot.2006.222
7. Santella, A., Catena, R., Kovacevic, I., Shah, P., Yu, Z., Marquina-Solis, J., Bao, Z.: WormGUIDES: an interactive single cell developmental atlas and tool for collaborative multidimensional data exploration. BMC Bioinformatics **16**, 189 (2015). https://doi.org/10.1186/s12859-015-0627-8
8. Laissue, P.P., Alghamdi, R.A., Tomancak, P., Reynaud, E.G., Shroff, H.: Assessing phototoxicity in live fluorescence imaging. Nat. Methods **14**(7), 657–661 (2017). https://doi.org/10.1038/nmeth.4344
9. Haus, E., Santella, A., Xu, Y., Ren, R., Wang, D., Bao, Z.: A single-cell spatiotemporal manifold of tissue morphology and dynamics. bioRxiv (2025). https://www.biorxiv.org/content/early/2025/10/22/2025.10.22.683950
10. Yu, X., Creamer, M.S., Randi, F., Sharma, A.K., Linderman, S.W., Leifer, A.M.: Fast deep neural correspondence for tracking and identifying neurons in *C. elegans* using semi-synthetic training. eLife **10**, e66410 (2021). https://doi.org/10.7554/eLife.66410
11. Besl, P.J., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992). https://doi.org/10.1109/34.121791
12. Myronenko, A., Song, X.: Point set registration: coherent point drift. IEEE Trans. Pattern Anal. Mach. Intell. **32**(12), 2262–2275 (2010). https://doi.org/10.1109/TPAMI.2010.46
13. Gold, S., Rangarajan, A.: A graduated assignment algorithm for graph matching. IEEE Trans. Pattern Anal. Mach. Intell. **18**(4), 377–388 (1996). https://doi.org/10.1109/34.491619
14. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: deep learning on point sets for 3D classification and segmentation. In: Proc. IEEE CVPR, pp. 652–660 (2017). https://doi.org/10.1109/CVPR.2017.16
15. Wang, D., Wang, Z., Zhao, X., Xu, Y., Bao, Z.: An observation data driven simulation and analysis framework for early stage *C. elegans* embryogenesis. J. Biomed. Sci. Eng. **11**(08), 225–234 (2018). https://doi.org/10.4236/jbise.2018.118018
16. Santella, A., Bao, Z.: *C. elegans* embryo data used in Twin Attention developmental manifold project. Zenodo (2025). https://doi.org/10.5281/zenodo.16878148
17. Boyle, T.J., Bao, Z., Murray, J.I., Araya, C.L., Waterston, R.H.: AceTree: a tool for visual analysis of *C. elegans* embryogenesis. BMC Bioinformatics **7**, 275 (2006). https://doi.org/10.1186/1471-2105-7-275
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: Advances in Neural Information Processing Systems, vol. 30, pp. 5998–6008 (2017). http://arxiv.org/abs/1706.03762
19. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: Proc. ICLR (2019). https://arxiv.org/abs/1711.05101
20. McDole, K., et al.: In toto imaging and reconstruction of post-implantation mouse development at the single-cell level. Cell **175**(3), 859–876 (2018). https://doi.org/10.1016/j.cell.2018.09.031

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29912-3_16