# Learning *Drosophila* ventral furrow formation with graph neural networks

**Shaoxun Huang**[1]  
huangsx@mit.edu

**Haiqian Yang**[1] *  
hqyang@mit.edu

**Hyuntae Jeong**[2]  
hyeontae_jeong@brown.edu

**Xinran O. Zhao**[3]  
xinzh765@mit.edu

**Ian Y. Wong**[2]  
ian_wong@brown.edu

**Markus J. Buehler**[1]  
mbuehler@mit.edu

**Adam C. Martin**[3]  
acmartin@mit.edu

**Ming Guo**[1] *  
guom@mit.edu

[1]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA  
[2]School of Engineering, Brown University, Providence, RI  
[3]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA

## Abstract

During early embryonic development, a relatively homogeneous population of cells collectively and spontaneously self-organize into complex structure. Understanding how tissue architecture guides these coordinated cell behaviors is essential for explaining how embryo reliably build structures that ultimately form the adult organism. Therefore, decoding the spatiotemporal information of tissue structure and dynamics is a central question in developmental biology. A paradigm process is the *Drosophila* gastrulation, during which the apical constriction gives rise to the formation of a **ventral furrow**, a critical step towards forming complex tissue structures. In this process, tissue morphologies and cell dynamics are closely correlated. Many prior biophysical models employ a graph representation of tissue structure, where the cells evolve following the gradient of a pre-defined free energy functional. Here we instead explore using graph neural networks to uncover the relation between tissue structures and dynamics from data, during the dynamical process of *Drosophila* ventral furrow formation. To do so, we build an experimental video dataset consisting of 320 fully segmented live-imaging movies, capturing the formation of ventral furrow over time. On this dataset, we use a graph neural network to learn individual cell speed from tissue structures, achieving good performance. Our results, along with the dataset, set the stage for future exploration of using graph neural networks for predictive modeling of embryonic development.

## 1 Introduction

The ability of a sheet of epithelial cells to spontaneously fold into specific structures is a fundamental process in development [1–3]. An important example is ventral furrow formation in *Drosophila* [4, 5]. During early stages of *Drosophila* embryonic development, the apical constriction of ventral cells gives rise to the formation of a furrow. This process is a key mechanism for the embryo to form

---

*Corresponding authors.

complex tissue structures. The *Drosophila* ventral furrow formation provides a powerful system to study the interaction between physical and biological factors that regulate tissue remodeling [5, 6]. Importantly, the physical properties and mechanical interactions between cells are shown to be critical factors regulating tissue folding processes, and recent studies reveal the complex interplay between the geometries and dynamics of living tissues [7–13].

Many recent modeling works in the context of biophysics and active matter seek to elucidate such structure-dynamics interplay through vertex models, where a tissue is paramterized by vertex cells, and it evolves by the gradient flow of pre-defined free energy functional, typically in a quadratic form of cell area and perimeter [14–18]. While these models successfully describe certain physical properties of tissues, they often fail to make predictions at single-cell resolution.

Nevertheless, a large number of prior biophysical models suggest a graph representation of the tissue structure (e.g. Voronoi cells, Delaunay triangles), making graph neural networks (GNNs) a promising candidate to model relation between structure and dynamics from data. Indeed, recent studies demonstrate GNNs as a powerful method to study cell and tissue dynamics [19–24]. However, to our knowledge, there has not been study focused on modeling the structure-dynamics interplay during **ventral furrow formation**, a paradigm process of early embryonic development, and this is particularly limited by the amount of publicly available video data.

In this work, we explore using GNNs to model the structure-dynamics relation during tissue development. Specifically, we organize a dataset consisting of 320 optical live-imaging videos of *Drosophila* ventral furrow formation. We segment all the movies and tracked the cells over time. On this fully segmented dataset, we use GNNs for a regression task of predicting cell movement speed from static tissue structural configurations. We further quantify the contribution of different geometric features. Our work demonstrates that GNN is a suitable approach to model tissue morphodynamical processes.
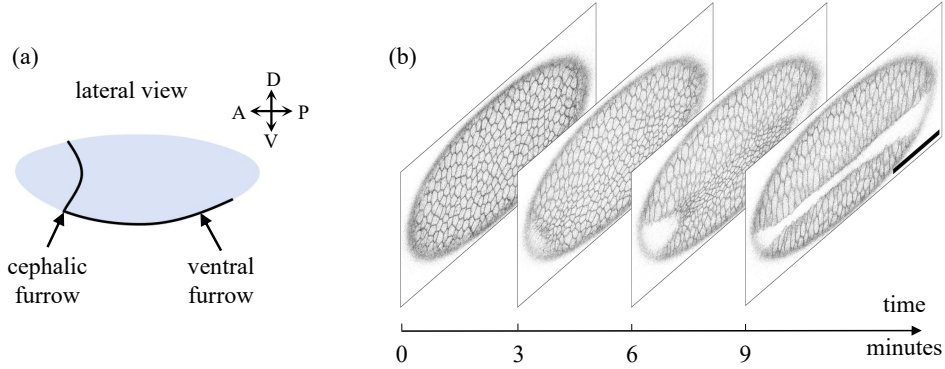


Figure 1: *Drosophila* ventral furrow formation. (a) Schematic showing the lateral view of a *Drosophila* embryo. The ventral furrow is highlighted. A, anterior side, P, posterior side, D, dorsal side, V, ventral side. (b) Example frames of a time-lapse optical live-imaging movie capturing the ventral furrow formation, where the cells in the middle of the field of view fold inside the embryo, forming a *furrow*. Scale bar, 50 μm.

## 2  *Drosophila* ventral furrow formation video dataset

The videos were collected in one of the author's laboratory in the past years, and we curated the dataset by manually screening for imaging quality, field of view consistency, and complete coverage of the ventral furrow formation process. As a result, we obtain a dataset of 320 organized optical live-imaging videos. Time-lapse videos are acquired using fluorescence microscopy to capture the ventral furrow formation, with cell boundaries fluorescently labeled (Fig. 1b). The collection contains both wild type embryos and embryos in which specific genes are perturbed by RNA interference. This combination of normal and perturbed conditions provides a rich and diverse dataset that captures both typical morphogenetic dynamics and a range of variant phenotypes, making it well suited for training and evaluating machine learning models.

2

## 3 Methods

**Cell segmentation and tracking.** We segment the videos using CellPose with its embedded 'cyto3' network [25] to obtain cell masks in each frame. From cell masks, we extract cell centroids and track their trajectories by nearest-neighbor matching between consecutive frames using a customized script. The cell speed is calculated by $\|\frac{\mathbf{r}(t)-\mathbf{r}(t+\tau)}{\tau}\|_2$ where $\mathbf{r}(t)$ is the centroid coordinate of the cell at time $t$, and $\tau$ is the lag time.

**Graph construction.** Based on the segmented cell masks, we represent each cell as a node and define edges between adjacent cell pairs. To construct features, we fit polygons to the cell masks, and by default we compute the area, perimeter, and relative coordinates of the polygons as the node features, and compute the length of the shared junction and the cell-cell distance as edge attributes. To further explore position-aware information beyond local geometric descriptors, we add positional embeddings by computing Laplacian eigenvector positional encodings [26]. The resulting positional encodings provide a structural coordinate system that captures the global topology of the cell adjacency graph, and are concatenated to the original node features.

**Model architecture.** The GNN models are implemented in PyTorch Geometric [27]. The models are composed of multi-layer graph convolution, with normalization layers in between. We experiment with different types of graph convolution, including Principal Neighbourhood Aggregation (PNA) [28], Graph Transformer (GT) [29], and DeeperGCN [30].

**Training pipeline.** We filter the dataset by requiring each graph (i.e., each frame of a video) to contain more than 200 cells with the median of cell shape index ($perimeter/\sqrt{area}$) smaller than 4.0. This procedure removes only low-quality frames that fail to meet the criteria. All 320 videos remain represented in the final dataset. We split train/validation/test datasets by videos to avoid data leakage across sets. We train the model for 200 epochs using the L1 loss and the AdamW [31] optimizer with weight decay (1e-3). The learning rate is scheduled by a ReduceLROnPlateau strategy (monitoring the validation loss in min mode, factor 0.5, patience 15, minimum learning rate 1e-5). Gradient clipping (max norm 1.0) and automatic mixed precision training are employed for stability and efficiency. When node features include absolute cell positions, we apply rigid geometric data augmentations such as flips, mirror reflections, axis swaps, translations, and jitter noise to improve invariance to orientation and coordinate shifts. The model checkpoint with the lowest validation loss is selected for evaluation on the held-out test set. For reproducibility, we fix random seeds for Python, NumPy, and PyTorch. We repeat each experiment with three independent seeds and report the mean and standard deviation across runs.

## 4 Results
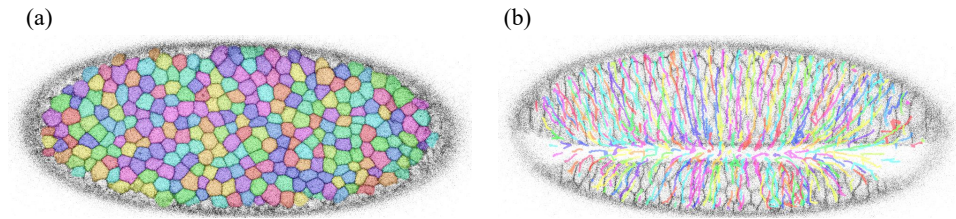
### 4.1 Cell graphs and trajectories.



Figure 2: (a) Example segmented cell masks, indicated by different colors. (b) Example cell trajectories, shown in different colors.

Example segmentation and tracking results are shown in Fig. 2, where panel (a) captures tissue structures through accurate detection of cell boundaries, and panel (b) demonstrates cell dynamics by reliably tracking individual cells across frames during ventral furrow formation. Below, we explore using a GNN model to establish a relation between structure and dynamics.

## 4.2 A regression task on cell speed from static graph configuration.

To study the interplay between tissue structures and cell dynamics, we use a GNN model, where we provide constructed graphs (cell as nodes and cell-cell adjacency as edges) as input and cell speed as output. As shown in Fig. 3, the model incorporating cell geometry features achieves good performance.

We notice that cell movement exhibits different distributions over different lag times. Over shorter lag times cell motion is more diffusive, while over longer lag times it becomes more persistent and ballistic. Therefore we test our model over different lag times. Though cell geometry features consistently improve model performance, the improvement is particularly pronounced at longer lag times. The gain from geometry features is smaller at shorter lag times, which we attribute to the predominance of diffusive cell movements in this regime. In contrast, when ballistic motion patterns become more prevalent, the tissue structure features are more informative for predicting cell speed.
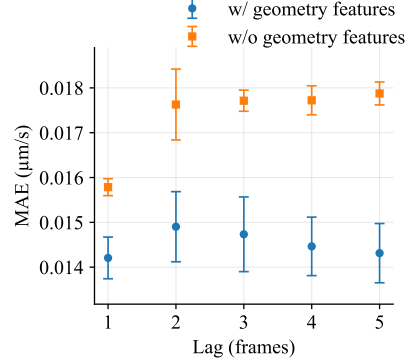


Figure 3: Mean absolute error ($\mu m/s$) on the test dataset, at different lag time, with and without geometry features. Results are averaged across three independent runs and the error bars report one standard deviation.

## 4.3 Cell geometry embedding with different GNN models.

Geometric descriptors are critical for representing cell structures and interactions, and play an important role in the regression task. By default, we provide cell area, perimeter, and relative position coordinates as the node features, together with cell junction length and cell-cell distance as the edge features, which achieves a mean absolute error of $0.0145 \pm 0.0007$ ($\mu m/s$). Removing these geometric features damages the performance (Table 1). To quantify the contribution of each type of geometric feature, we add each feature back, using different variants of GNN. Using partial geometric features underperforms using the complete set. However, interestingly, different types of features yield similar performance (Table 1), indicating that these cell geometric features provide correlated information. Across different architectures, PNA consistently achieves the lowest error, while DeeperGCN and GT perform similarly. The inclusion of positional embeddings (PE) substantially improves performance in the absence of geometric features, while adding PE brings only marginal benefit when structural information is already included, and the benefit is relatively larger for the DeeperGCN architecture.

GNN models bridge cell geometric structures and tissue dynamics by capturing cell-cell interactions in the neighbourhood defined by the graph structure. Starting from this basic task, we can further explore the relationship between these geometric variables and how different combinations of geometric features contribute to tissue-level behaviors in the future.

Table 1: Mean absolute error ($\mu m/s$) on the test set across GNN variants under different configurations of input geometric features, with and without PE.

| Model | Without PE | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| | Full | N/A | Area | Perimeter | Coordinates | Junction length | Cell-cell distance |
| PNA | $0.0145 \pm 0.0007$ | $0.0177 \pm 0.0003$ | $0.0146 \pm 0.0005$ | $0.0147 \pm 0.0005$ | $0.0154 \pm 0.0014$ | $0.0156 \pm 0.0003$ | $0.0158 \pm 0.0017$ |
| GT | $0.0151 \pm 0.0005$ | $0.0180 \pm 0.0003$ | $0.0160 \pm 0.0005$ | $0.0164 \pm 0.0005$ | $0.0159 \pm 0.0010$ | $0.0179 \pm 0.0002$ | $0.0163 \pm 0.0007$ |
| DeeperGCN | $0.0150 \pm 0.0006$ | $0.0185 \pm 0.0001$ | $0.0161 \pm 0.0003$ | $0.0175 \pm 0.0005$ | $0.0187 \pm 0.0031$ | $0.0169 \pm 0.0004$ | $0.0174 \pm 0.0011$ |
| Model | With PE | | | | | | |
| | Full | N/A | Area | Perimeter | Coordinates | Junction length | Cell-cell distance |
| PNA | $0.0143 \pm 0.0008$ | $0.0159 \pm 0.0008$ | $0.0144 \pm 0.0007$ | $0.0147 \pm 0.0008$ | $0.0160 \pm 0.0011$ | $0.0152 \pm 0.0011$ | $0.0160 \pm 0.0010$ |
| GT | $0.0156 \pm 0.0008$ | $0.0161 \pm 0.0007$ | $0.0155 \pm 0.0007$ | $0.0156 \pm 0.0007$ | $0.0163 \pm 0.0008$ | $0.0164 \pm 0.0006$ | $0.0162 \pm 0.0006$ |
| DeeperGCN | $0.0151 \pm 0.0010$ | $0.0166 \pm 0.0013$ | $0.0157 \pm 0.0011$ | $0.0157 \pm 0.0010$ | $0.0165 \pm 0.0012$ | $0.0156 \pm 0.0009$ | $0.0163 \pm 0.0015$ |

## 5   Conclusion and Outlook

In this work we organize a dataset consisting of 320 optical live-imaging movies capturing the ventral furrow formation during the dynamic process of *Drosophila* gastrulation. On this dataset, we explore using GNN models on a regression task to learn individual cell speed from the multicellular tissue structure. We further use the task as a framework to investigate the influence of different cell geometric features. By providing a fully segmented, large-scale collection of ventral furrow movies, our work establishes a new benchmark for quantitatively evaluating structure–dynamics relation in developmental biology. Taken together, our study demonstrates that GNNs provide a useful strategy for modeling the structure–dynamics relation during embryonic development.

For future studies, there are several interesting directions. First, incorporating equivariant architectures could allow the model to better respect symmetries and potentially improve generalization. Second, extending the current regression formulation to temporal or generative models, such as autoregressive or diffusion models, would allow capturing full morphogenetic trajectories. Third, integrating physically informed constraints could ensure the model to be consistent with known tissue mechanical principles. Finally, expanding the dataset to include a broader range of perturbations, developmental stages and modalities would further strengthen the model's ability to study complex tissue morphodynamical processes.

## References

[1] Yue Liu, Xufeng Xue, Shiyu Sun, Norio Kobayashi, Yung Su Kim, and Jianping Fu. Morphogenesis beyond in vivo. *Nature Reviews Physics*, 6(1):28–44, 2024.

[2] Sandra B Lemke and Celeste M Nelson. Dynamic changes in epithelial cell packing during tissue morphogenesis. *Current Biology*, 31(18):R1098–R1110, 2021.

[3] Yu-Chiun Wang, Zia Khan, Matthias Kaschube, and Eric F Wieschaus. Differential positioning of adherens junctions is associated with initiation of epithelial folding. *Nature*, 484(7394): 390–393, 2012.

[4] Dari Sweeton, Suki Parks, Michael Costa, and Eric Wieschaus. Gastrulation in drosophila: the formation of the ventral furrow and posterior midgut invaginations. *Development*, 112(3): 775–789, 1991.

[5] Adam C Martin, Matthias Kaschube, and Eric F Wieschaus. Pulsed contractions of an actin–myosin network drive apical constriction. *Nature*, 457(7228):495–499, 2009.

[6] Lior Atia, Dapeng Bi, Yasha Sharma, Jennifer A Mitchel, Bomi Gweon, Stephan A. Koehler, Stephen J DeCamp, Bo Lan, Jae Hun Kim, Rebecca Hirsch, et al. Geometric constraints during epithelial jamming. *Nature physics*, 14(6):613–620, 2018.

[7] Tomer Stern, Stanislav Y Shvartsman, and Eric F Wieschaus. Deconstructing gastrulation at single-cell resolution. *Current Biology*, 32(8):1861–1868, 2022.

[8] Xun Wang, Matthias Merkel, Leo B Sutter, Gonca Erdemci-Tandogan, M Lisa Manning, and Karen E Kasza. Anisotropy links cell shapes to tissue flow during convergent extension. *Proceedings of the National Academy of Sciences*, 117(24):13541–13551, 2020.

[9] Christian Cupo, Cole Allan, Vikram Ailiani, and Karen E Kasza. Signatures of structural disorder in the developing drosophila germband epithelium. *PRX Life*, 2(4):043004, 2024.

[10] G Wayne Brodland, Vito Conte, P Graham Cranston, Jim Veldhuis, Sriram Narasimhan, M Shane Hutson, Antonio Jacinto, Florian Ulrich, Buzz Baum, and Mark Miodownik. Video force microscopy reveals the mechanics of ventral furrow invagination in drosophila. *Proceedings of the National Academy of Sciences*, 107(51):22111–22116, 2010.

[11] Vito Conte, José J Munoz, Buzz Baum, and Mark Miodownik. Robust mechanisms of ventral furrow invagination require the combination of cellular shape changes. *Physical biology*, 6(1): 016010, 2009.

[12] Nikolas H Claussen, Fridtjof Brauns, and Boris I Shraiman. A geometric-tension-dynamics model of epithelial convergent extension. *Proceedings of the National Academy of Sciences*, 121(40):e2321928121, 2024.

[13] Fridtjof Brauns, Nikolas H Claussen, Matthew F Lefebvre, Eric F Wieschaus, and Boris I Shraiman. The geometric basis of epithelial convergent extension. *eLife*, 13:RP95521, 2024.

[14] Philipp Spahn and Rolf Reuter. A vertex model of drosophila ventral furrow formation. *PloS one*, 8(9):e75051, 2013.

[15] Alexander G Fletcher, Miriam Osterfield, Ruth E Baker, and Stanislav Y Shvartsman. Vertex models of epithelial morphogenesis. *Biophysical journal*, 106(11):2291–2304, 2014.

[16] Steve Runser, Roman Vetter, and Dagmar Iber. Simucell3d: three-dimensional simulation of tissue mechanics with cell polarization. *Nature Computational Science*, 4(4):299–309, 2024.

[17] Dapeng Bi, JH Lopez, Jennifer M Schwarz, and M Lisa Manning. A density-independent rigidity transition in biological tissues. *Nature Physics*, 11(12):1074–1079, 2015.

[18] Tatsuzo Nagai and Hisao Honda. A dynamic cell model for the formation of epithelial tissues. *Philosophical Magazine B*, 81(7):699–719, 2001.

[19] Haiqian Yang, Florian Meyer, Shaoxun Huang, Liu Yang, Cristiana Lungu, Monilola A Olayioye, Markus J Buehler, and Ming Guo. Learning collective cell migratory dynamics from a static snapshot with graph neural networks. *PRX Life*, 2(4):043010, 2024.

[20] Haiqian Yang, Anh Q Nguyen, Dapeng Bi, Markus J Buehler, and Ming Guo. Multicell-fold: geometric learning in folding multicellular life. *ArXiv*, pages arXiv–2407, 2024.

[21] Quirine JS Braat, Giulia Janzen, Bas C Jansen, Vincent E Debets, Simone Ciarella, and Liesbeth MC Janssen. Shape matters: inferring the motility of confluent cells from static images. *Soft Matter*, 2025.

[22] Koen Minartz, Tim d'Hondt, Leon Hillmann, Jörn Starruß, Lutz Brusch, and Vlado Menkovski. Deep neural cellular potts models. *arXiv preprint arXiv:2502.02129*, 2025.

[23] Jesús Pineda, Benjamin Midtvedt, Harshith Bachimanchi, Sergio Noé, Daniel Midtvedt, Giovanni Volpe, and Carlo Manzo. Geometric deep learning reveals the spatiotemporal features of microscopic motion. *Nature Machine Intelligence*, 5(1):71–82, 2023.

[24] Takaki Yamamoto, Katie Cockburn, Valentina Greco, and Kyogo Kawaguchi. Probing the rules of cell coordination in live tissues by interpretable machine learning based on graph neural networks. *PLOS Computational Biology*, 18(9):e1010477, 2022.

[25] Carsen Stringer and Marius Pachitariu. Cellpose3: One-click image restoration for improved cellular segmentation. *Nature Methods*, 22(3):592–599, March 2025. ISSN 1548-7105. doi: 10.1038/s41592-025-02595-5.

[26] Vijay Prakash Dwivedi, Chaitanya K. Joshi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks, December 2022.

[27] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with pytorch geometric. *arXiv preprint arXiv:1903.02428*, 2019.

[28] Gabriele Corso, Luca Cavalleri, Dominique Beaini, Pietro Liò, and Petar Veličković. Principal neighbourhood aggregation for graph nets. *Advances in neural information processing systems*, 33:13260–13271, 2020.

[29] Vijay Prakash Dwivedi and Xavier Bresson. A generalization of transformer networks to graphs. *arXiv preprint arXiv:2012.09699*, 2021. URL https://arxiv.org/abs/2012.09699.

[30] Guohao Li, Chenxin Xiong, Ali Thabet, and Bernard Ghanem. DeeperGCN: All you need to train deeper GCNs. *arXiv preprint arXiv:2006.07739*, 2020. URL https://arxiv.org/abs/2006.07739.

[31] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.