

A CONTINUOUS AND INTERPRETABLE MORPHOMETRIC FOR ROBUST QUANTIFICATION OF DYNAMIC BIOLOGICAL SHAPES

Roua Rouatbi^{1,2,3,4}, Juan-Esteban Suarez Cardona^{6,7}, Alba Villaronga-Luque², Jesse V. Veenliet^{2,3,5}, Ivo F. Sbalzarini^{1,2,3,4,5}

1. Dresden University of Technology, Faculty of Computer Science
2. Max Planck Institute of Molecular Cell Biology and Genetics
3. Center for Systems Biology Dresden
4. Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI) Dresden/Leipzig
5. Cluster of Excellence Physics of Life, Dresden University of Technology
6. Ludwig-Maximilians-Universität München
7. Munich Center for Machine Learning (MCML)

ABSTRACT

We introduce the Push-Forward Signed Distance Morphometric (PF-SDM) for shape quantification in biomedical imaging. The PF-SDM compactly encodes geometric and topological properties of closed shapes, including their skeleton and symmetries. This provides robust and interpretable features for shape comparison and machine learning. The PF-SDM is mathematically smooth, providing access to gradients and differential-geometric quantities. It also extends to temporal dynamics and allows fusing spatial intensity distributions, such as genetic markers, with shape dynamics. We present the PF-SDM theory, benchmark it on synthetic data, and apply it to predicting body-axis formation in mouse gastruloids, outperforming a CNN baseline in both accuracy and speed.

Index Terms— geometric shape analysis, level-set methods, shape quantification, morphometric, vector embedding

1. INTRODUCTION

The human visual system excels at grouping similar-looking objects based on their shape, forming equivalence classes that remain invariant under shape-preserving transformations—rotation, translation, reflection, and scaling. Capturing this in a quantitative metric, however, remains a challenge.

Shape metrics, often called *morphometrics* in biomedicine, have been extensively investigated to quantify differences in shapes. Classic morphometrics, such as Generalized Procrustes Analysis (GPA) [1] and Elliptical Fourier Analysis (EFA) [2], are widely used but have important limitations: GPA depends on landmark selection, while EFA is sensitive to the initial tracing point. Topological approaches like the Euler Characteristic Transform (ECT) and Persistent Homology Transform (PHT) [3] capture global multi-scale features but lack smooth geometric information.

Approaches based on Signed Distance Functions (SDF) [4] provide geometric representations from which interpretable shape descriptors can be derived. The SDF measures the signed orthogonal distance from any point in the domain to the boundary of a closed shape. The singularities in the SDF are the topological skeleton of the shape, which is important for shape similarity [5]. However, most SDF methods are limited to discrete representations of the SDF, and the SDF is not invariant to shape-preserving transformations.

Here, we provide a shape-preserving and continuous SDF variant, the Push-Forward Signed Distance Function (PF-SDF). It is mathematically smooth and relies on a common reference domain for shape invariance. From the Fourier transform of the PF-SDF, we derive the Push-Forward Signed Distance Morphometric (PF-SDM) to capture topological and geometric shape features. We show that this morphometric yields robust and interpretable embeddings that extend to temporal dynamics and can be fused with spatial distributions of, e.g., genetic markers. We showcase this by applying PF-SDM to predicting body-axis formation in mouse gastruloids. Formulating the resulting machine-learning problem in PF-SDF space, instead of pixel space, improves both accuracy and computational cost over a CNN baseline.

2. METHOD

Once a geometric model of a biological object is available, e.g., from image segmentation, the PF-SDM is computed on a per-object basis in three stages: SDF computation, PF-SDF mapping, and morphometric evaluation.

We define a shape $S \subseteq \Omega$ as a closed smooth manifold of co-dimension one [6], embedded in a domain $\Omega := (-1, 1)^2$, and $\mathbb{S}_M := \{S_i\}_{i=1}^N$ a set of $N \in \mathbb{N}$ shapes. We consider a data set $I_N \subseteq \mathbb{R}^{p \times p}$ of 2D $p \times p$ images. Each image is preprocessed by segmenting the shapes within it and scaling

each shape it to fit within the unit circle, yielding a set of points $\mathbf{x}_{S_i} := \{x_{l,i}\}_{l=1}^{N_p} \subseteq S_i$ sampled on each shape S_i .

2.1. Signed Distance Function computation

We first compute the classic Signed Distance Function (SDF) $\phi_S : \Omega \mapsto \mathbb{R}$ for a given shape S by numerically solving the viscous Eikonal equation [7]. The adjustable viscosity term $\mu \in \mathbb{R}_+$ ensures numerical stability by regularizing the SDF around singularities. Specifically, we solve

$$\begin{cases} \|\nabla\phi(x, y)\|_2^2 + \mu\Delta\phi(x, y) = 1 & (x, y) \in \Omega, \\ \phi(x, y) = 0 & (x, y) \in S, \end{cases} \quad (1)$$

where $\|\cdot\|_2$ denotes the ℓ^2 -norm of a vector. The boundary condition for Eq. (1) is sampled at the points $\mathbf{x}_S \subseteq S$ on the shape. Due to the ℓ^2 -term, Eq. (1) is nonlinear in ϕ . We solve it numerically using polynomial surrogate models (PSM) and Sobolev cubatures [8]. This results in a smooth polynomial approximation of the SDF.

2.2. Push-Forward Signed Distance Function

The SDF ϕ_S provides a smooth representation of a shape S . While SDFs are widely used in shape analysis, they are not invariant to shape-preserving transformations [9]. To address this limitation, we introduce the *push-forward signed distance function* (PF-SDF) $\phi_{S^*} : \Omega_r \mapsto \mathbb{R}$. This achieves scale and translation invariance by mapping to a reference domain $\Omega_r \subseteq \mathbb{R}^2$, deforming the original SDF ϕ_S such that its zero level set aligns with the reference shape $S_r := \partial\Omega_r$, i.e., $S_r = \{x \in \mathbb{R}^2 : \phi_{S^*}(x) = 0\}$.

In the following, we consider Ω_r to be the unit disk $\Omega_r = (0, 2\pi) \times (0, 1) := \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$ with boundary $S_r = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$. The *push-forward map* of the SDF ϕ_S onto S_r is represented as a deformation map $\Psi_\zeta : S \mapsto \mathbb{R}^2$, deforming S to S_r as $\Psi_\zeta(x, y) := (x + \nu_{\zeta_x}(x, y), y + \nu_{\zeta_y}(x, y))$. Here $\nu_{\zeta_x}, \nu_{\zeta_y} \in \Pi_m(S)$ are degree- m polynomials with Chebyshev coefficients $\zeta_x, \zeta_y \in \mathbb{R}^{(n+1)^2}$, determined by solving

$$\min_{\zeta_x, \zeta_y \in \mathbb{R}^{(n+1)^2}} \int_S (\Psi_\zeta - \mathbf{x}_r)^2 ds \quad \text{with} \quad (2)$$

$$\mathbf{x}_r(x, y) := \arg \min_{x_r, y_r \in S_r} (\|x_r - x\|_2 + \|y_r - y\|_2).$$

This determines Ψ_ζ on the subset $S \subseteq \Omega$, as the objective functional is integrated over S rather than the entire domain Ω . To obtain a global deformation map $\bar{\Psi}_\zeta : \Omega \rightarrow \mathbb{R}^2$, we construct a harmonic extension of Ψ_ζ [10]. In polar coordinates (θ, r) , this yields:

$$\phi_{S^*}(\theta, r) = \phi_S(\bar{\Psi}_{\zeta_x}(x(\theta, r), y(\theta, r)), \bar{\Psi}_{\zeta_y}(x(\theta, r), y(\theta, r))) \quad (3)$$

with $x(\theta, r) = r \cos(\theta)$, $y(\theta, r) = r \sin(\theta)$.

2.3. Morphometric evaluation

The PF-SDF maps any shape to the unit disk Ω_r , where $r = 1$ is the shape boundary and $r = 0$ its center, while retaining the geometric richness of traditional SDF formulations. In particular, morphometrics can be defined as shape similarity measures between the PF-SDFs $\phi_{S_i^*}, \phi_{S_j^*}$ of two shapes S_i, S_j . Here, we propose a morphometric based on the Fourier transform of the PF-SDF.

The Fourier transform is a natural choice because the PF-SDF for any radius r is 2π -periodic. Therefore, we consider

$$\hat{\phi}_{S^*}(\theta, r) = \sum_{n=0}^{N_F} c_n(r) e^{-in\theta}, \quad (4)$$

where $N_F \in \mathbb{N}$ is a fixed number of Fourier coefficients $c_n : (0, 1) \mapsto \mathbb{R}$. The coefficients are functions of r , given by

$$c_n(r) = \frac{1}{2\pi} \int_0^{2\pi} \phi_{S^*}(\theta, r) e^{-in\theta} d\theta. \quad (5)$$

This renders the PF-SDF invariant to rotation and reflection. The map $S \mapsto \mathbf{c}_{\phi_S}(r) := \{c_n(r)/\|c_n(r)\|_2\}_{n=1}^{N_F}$, $\mathbf{c}_{\phi_S} : (0, 1) \mapsto \mathbb{R}^{N_F}$, directly provides a vector embedding of the shape S .

The PF-SDF can optionally account for an intensity field $I_S : \Omega \mapsto \mathbb{R}$. This can be any scalar field in the image, for example the fluorescence intensity of a genetic marker. Using the deformation map from Eq. (2), the intensity signal is mapped to the same unit disk, yielding $I_{S^*} : \Omega_r \mapsto \mathbb{R}$ with normalized Fourier coefficients \mathbf{c}_{I_S} .

Together, this defines the intensity-aware PF-SDF distance between two shapes $S_1, S_2 \subseteq \Omega$,

$$\begin{aligned} d_\phi(S_1, S_2) &:= \int_0^1 (\mathbf{c}_{\phi_{S_1}}(r) - \mathbf{c}_{\phi_{S_2}}(r))^2 dr \\ &+ \int_0^1 (\mathbf{c}_{I_{S_1}}(r) - \mathbf{c}_{I_{S_2}}(r))^2 dr. \end{aligned} \quad (6)$$

The PF-SDF of any shape is invariant to shape-preserving transformations, and so is the above distance.

3. RESULTS

We validate the PF-SDF on a synthetic benchmark data set to confirm its invariance properties and compare it with previous morphometrics in terms of its robustness and interpretability. Then, we illustrate the application of the PF-SDF to a biological problem that includes temporal dynamics of shapes and of the intensity field of a genetic marker. Specifically, we consider the problem of predicting body-axis formation in mouse gastruloids from fluorescence microscopy images. In all experiments, we use $N_p = 100$ boundary points per shape.

3.1. Synthetic Validation Benchmark

We assess the robustness of the PF-SDM to shape-preserving transformations by considering a synthetic data set of five base shapes: a circle (S_1), an elongated shape (S_2), an asymmetrically folded shape (S_3), a rounded square (S_4), and a branched shape (S_5) (see examples in Fig. 1, bottom-right). Each base shape is transformed ten times by random translation, rotation, reflection, and isotropic scaling, resulting in 50 transformed shapes. We compute full pairwise distance matrices between these 50 shapes using three different morphometrics:

- **PF-SDM:** Shape distances are computed using Eq. (6) with $N_F = 10$ Fourier coefficients and no intensity field.
- **EFA:** We represent each shape using the first 10 elliptical Fourier harmonics, computed with the `spatial-efd` package [11]. The resulting coefficients are normalized, yielding one coefficient vector per shape.
- **GPA:** We sample equidistant boundary points from EFA. GPA is then used on these landmark points to compute pairwise distances between aligned shapes.

The first two components of a Multi-Dimensional Scaling (MDS) [12] reduction of the resulting distance matrices are shown in Fig. 1.

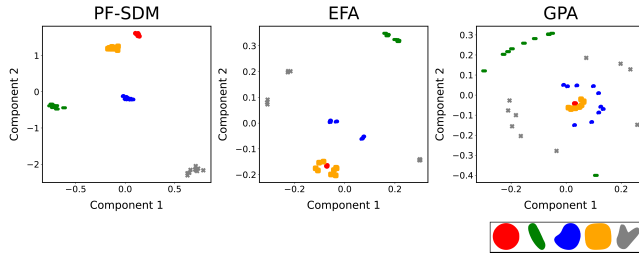


Fig. 1. First two MDS components of the shape distance matrices computed with PF-SDM (left), EFA (center), and GPA (right) over ten shape-preserving transformations of each of the five test shapes shown in the inset legend. Point colors match shape-legend colors.

The data shows that PF-SDM yields the most robust clustering under shape-preserving transformations. Randomly perturbed instances of the same shape remain more tightly grouped than in EFA and GPA. In addition, the PF-SDM uncovers a meaningful hierarchy of shape similarity: The circle and rounded square clusters are closest, followed by the folded shape, while the elongated and branched shapes are mapped the farthest apart. This demonstrates the robustness of PF-SDM to transformations, its geometric interpretability, and its superior performance over EFA and GPA.

Interestingly, the hierarchy revealed by PF-SDM—from spherical to folded to branched—mirrors the typical morphological progressions observed in developmental biological systems.

3.2. Application to Mouse Gastruloids

To illustrate the use of the PF-SDM in a challenging developmental biology application, we consider the developmental dynamics of mouse gastruloids. Gastruloids are an *in vitro* stem-cell-based embryo model recapitulating aspects of body patterning and morphogenesis [13]. Their generation induces expression of the transcription factor Brachyury, followed by symmetry breaking and body-axis formation [13, 14]. During this morphogenetic process, gastruloids transition from a spherical to an ovoid/teardrop and finally to an elongated shape [13, 14]. Unlike a natural embryo, however, gastruloids occasionally form multiple body axes [15]. We analyze time-lapse videos of gastruloids with an mCherry-labeled Brachyury reporter [16, 14], imaged in brightfield and mCherry channels every 45 min. The data set consists of 78 videos (25 time frames each) of gastruloids, of which 65 developed a single axis and 13 multiple axes.

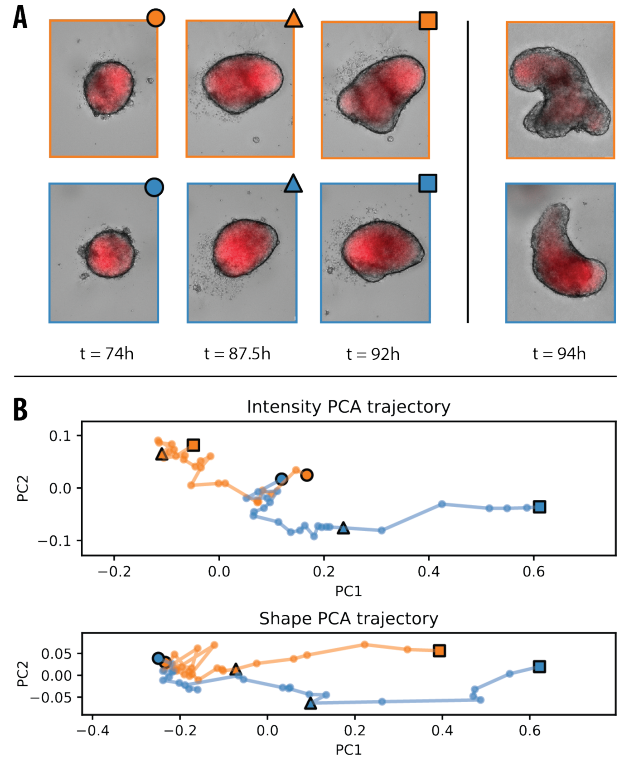


Fig. 2. Shape analysis of developing mouse gastruloids. (A) Representative gastruloid images from the single-axis (blue frame) and multi-axes (orange frame) classes at early (circle), intermediate (triangle), and late (square) times, along with a posterior reference at 94 h. Brachyury::mCherry intensity is shown in red. (B) PCA trajectories of $N_F = 25$ PF-SDM Fourier coefficients c for Brachyury intensity (top) and shape (bottom) for the examples from (A) across 25 time points.

We represent each gastruloid independently at every time point using the PF-SDF for both its shape and the Brachyury

intensity field. Figure 2A shows representative gastruloids from the two classes across time. Figure 2B shows the PCA trajectories of $N_F = 25$ PF-SDF Fourier coefficients for the shapes from (A). At early times (circle in Fig. 2B), gastruloids exhibit similar morphologies and intensity distributions. At intermediate stages (triangle), shape differences between the two classes start to be captured by the PF-SDM. The differences in shape persist until the final time (squares), whereas differences in shape-corrected intensity further amplify, indicating that molecular asymmetry precedes shape dynamics.

We next ask whether the eventual formation of multiple axes can be predicted from images at earlier times (74...92 h) when morphological differences between the classes are not yet obvious to the human eye (Fig. 2A). This makes the task intrinsically challenging but biologically meaningful, since it tests whether early cues in shape and gene expression can predict future developmental fate. Ground-truth labels (single-axis vs. multi-axes) for training were assigned retrospectively, based on morphological outcomes observed at later times (94...120 h), when the emergence of one or multiple axes becomes unambiguous.

For classification, the PF-SDF Fourier coefficients from all 25 time points are concatenated into one vector. We train logistic regression classifiers with balanced class weights separately for the shape and intensity features and compare classification performance with a convolutional neural network (CNN) baseline [14] trained directly on pixel segmentation masks and raw intensity images. The CNN is exposed to 1.55 million augmented images during training (62 videos \times 25 time points \times 1000 augmentations; affine transformations for shapes, affine and noise perturbations for intensities). All classifiers output separate class probabilities for the shape and intensity channels, p_{SDF} and p_I . Final predictions are computed by late fusion as $p_{\text{fusion}} = \alpha p_{\text{SDF}} + (1 - \alpha) p_I$, with $\alpha \in [0, 1]$ tuned for each classifier to maximize balanced accuracy on held-out test splits. Evaluation is done with repeated stratified shuffle splits ($k = 5$, test size 20%) [17].

The results in Fig. 3 show that logistic regression over PF-SDM features consistently outperforms the CNN baseline both in accuracy and balanced accuracy, despite the severe class imbalance. Moreover, classification based on PF-SDM required an order of magnitude less compute resources (4 min vs. 50 min for CNN training). The accuracy gains in PF-SDM are explained by the shape features, whereas the intensity features achieve the same performance as the CNN. Importantly, the fusion of shape and intensity improves the mean performance across test splits, suggesting complementary biological contributions of morphology and gene expression.

4. CONCLUSION

We presented the Push-Forward Signed Distance Morphometric (PF-SDM), a continuous geometric morphometric that is invariant to shape-preserving transformations. We provided

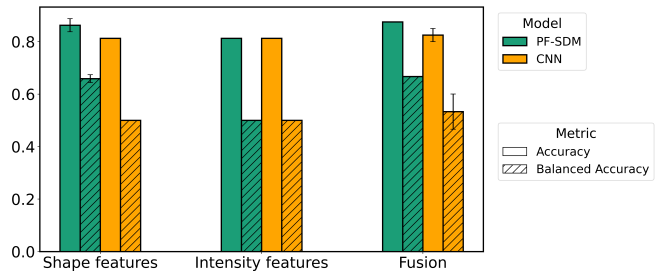


Fig. 3. Comparison of classification performance between logistic regression over PF-SDM features and the CNN baseline [14]. Accuracy (plain) and Balanced Accuracy (shaded) are shown for shape, intensity, and fusion (shape+intensity) features. Error bars: empirical standard deviation across five stratified test splits.

the theoretical foundation for PF-SDM and a practical algorithm for its computation (available in Python from <https://git.mpi-cbg.de/mosaic/software/machine-learning/pf-sdm> and as a *Napari* plugin from <https://git.mpi-cbg.de/mosaic/software/machine-learning/pf-sdm-napari-plugin>). Through experiments on synthetic shapes, we confirmed that PF-SDM generates robust (to shape transformations) and interpretable hierarchical embeddings of geometric objects based on their shapes, outperforming both Elliptical Fourier Analysis (EFA) and Generalized Procrustes Analysis (GPA). We then applied the proposed PF-SDM to the biological problem of predicting body-axis formation in mouse gastruloids from microscopy images. There, simple logistic regression over PF-SDM features outperformed a custom CNN trained on the raw images, while requiring ten-fold less computer time. This showcases the benefits of formulating a machine-learning question in a more appropriate space with the correct in- and equivariances. Indeed, the PF-SDM transforms spatiotemporal shape dynamics to mathematically smooth and geometrically interpretable compact features that can directly be fused with gene-expression patterns.

For the sake of clarity and simplicity, we considered only two-dimensional shapes and their temporal dynamics. Extending the PF-SDF framework to 3D+time is subject to ongoing work. All mathematical concepts do, however, generalize to higher dimensions. More difficult is the generalization from closed shapes to open or non-singly-connected manifolds, which would require tree or vector SDFs [18]. PF-SDM could also be generalized to Sobolev norms [19], which could help capture higher-order geometric features. Finally, future work could attempt to formally prove the differentiability of PF-SDM distances, which could be conjectured from their smoothness and continuity. This would open new avenues for incorporating PF-SDM terms in loss functions of machine-learning models to enhance shape representation [20].

5. ACKNOWLEDGMENTS

This work was supported by the Max Planck Society and the German Federal Ministry of Research, Technology and Space (BMFTR) as part of the Center for Scalable Data Analytics and Artificial Intelligence (ScaDS.AI). JVV acknowledges support from a European Innovation Council (EIC) Pathfinder grant under the Horizon Europe Research and Innovation Program (Horizon-EIC-2021-PathfinderChallenges-01 101071203, SUMO).

6. REFERENCES

- [1] J. C. Gower, “Generalized Procrustes analysis,” *Psychometrika*, vol. 40, no. 1, pp. 33–51, Mar. 1975.
- [2] Frank P Kuhl and Charles R Giardina, “Elliptic Fourier features of a closed contour,” *Computer Graphics and Image Processing*, vol. 18, no. 3, pp. 236–258, 1982.
- [3] K. Turner, S. Mukherjee, and D. M. Boyer, “Persistent homology transform for modeling shapes and surfaces,” *Information and Inference*, vol. 3, no. 4, pp. 310–344, Dec. 2014.
- [4] Stanley Osher and Ronald Fedkiw, *Level set methods and dynamic implicit surfaces*, Applied mathematical sciences. Springer, Dec. 2003.
- [5] Andrea Torsello and Edwin R. Hancock, “A skeletal measure of 2D shape similarity,” *Computer Vision and Image Understanding*, vol. 95, no. 1, pp. 1–29, 2004.
- [6] Jeffrey M., *Manifolds and Differential Geometry*, Graduate studies in mathematics. American Mathematical Society, Providence, RI, Jan. 2010.
- [7] Lawrence C. Evans, *Partial differential equations*, American Mathematical Society, 2010.
- [8] Juan-Esteban Suarez Cardona, Phil-Alexander Hoffmann, and Michael Hecht, “Negative order Sobolev cubatures: preconditioners of partial differential equation learning tasks circumventing numerical stiffness,” *Machine Learning: Science and Technology*, vol. 5, no. 3, pp. 035029, July 2024.
- [9] M.C. Delfour and J.P. Zolesio, “Shape analysis via oriented distance functions,” *Journal of Functional Analysis*, vol. 123, no. 1, pp. 129–201, July 1994.
- [10] Zuoqiang Shi, Jian Sun, and Minghao Tian, “Harmonic extension,” <https://arxiv.org/abs/1509.06458>, 2015.
- [11] Stuart W D Grieve, “spatial-efd: A spatial-aware implementation of elliptical Fourier analysis,” *The Journal of Open Source Software*, vol. 2, no. 11, Mar. 2017.
- [12] Trevor Cox and Michael Cox, *Multidimensional Scaling*, Chapman and Hall/CRC, Sept. 2000.
- [13] Leonardo Beccari, Naomi Moris, Mehmet Girgin, David A Turner, Peter Baillie-Johnson, Anne-Catherine Cossy, Matthias P Lutolf, Denis Duboule, and Alfonso Martinez Arias, “Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids,” *Nature*, vol. 562, no. 7726, oct 2018.
- [14] Alba Villaronga-Luque, Ryan G Savill, Natalia López-Anguita, Adriano Bolondi, Sumit Garai, Seher Ipek Gassaloglu, Roua Rouatbi, Kathrin Schmeisser, Aayush Poddar, Lisa Bauer, Tiago Alves, Sofia Traikov, Jonathan Rodenfels, Triantafyllos Chavakis, Aydan Bulut-Karslıoglu, and Jesse V Veenvliet, “Integrated molecular-phenotypic profiling reveals metabolic control of morphological variation in a stem-cell-based embryo model,” *Cell Stem Cell*, vol. 32, no. 5, 2025.
- [15] Isma Bennabi, Pauline Hansen, Melody Merle, Judith Pineau, Lucille Lopez-Delisle, Dominique Kolly, Denis Duboule, Alexandre Mayran, and Thomas Gregor, “Size-dependent temporal decoupling of morphogenesis and transcriptional programs in gastruloids,” *BioRxiv*, 2024.
- [16] Jesse V Veenvliet, Adriano Bolondi, Helene Kretzmer, Leah Haut, Manuela Scholze-Wittler, Dennis Schifferl, Frederic Koch, Léo Guignard, Abhishek Sampath Kumar, Milena Pustet, Simon Heimann, René Buschow, Lars Wittler, Bernd Timmermann, Alexander Meissner, and Bernhard G Herrmann, “Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites,” *Science*, vol. 370, no. 6522, 2020.
- [17] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, “Scikit-learn: Machine learning in Python,” *J. Mach. Learn. Resch.*, vol. 12, no. 85, pp. 2825–2830, 2011.
- [18] Xun Xiao, Veikko F. Geyer, Hugo Bowne-Anderson, Jonathon Howard, and Ivo F. Sbalzarini, “Automatic optimal filament segmentation with sub-pixel accuracy using generalized linear models and B-spline level-sets,” *Medical Image Analysis*, vol. 32, pp. 157–172, 2016.
- [19] Raa and Jjff, “Preface,” in *Sobolev Spaces*, Pure and Applied Mathematics (Amsterdam). Elsevier, 2003.
- [20] Anna Foix Romero, Craig Russell, Alexander Krull, and Virginie Uhlmann, “ShapeEmbed: a self-supervised learning framework for 2D contour quantification,” 2025.