PLOS Computational Biology

METHODS

Fast and interpretable quantification of biological shape heterogeneity via stratified Wasserstein kernel

Wenjun Zhao[1*], **Danica J. Sutherland**[2,3], **Khanh Dao Duc**[4]

1 Department of Mathematics, Wake Forest University, Winston-Salem, North Carolina, United States of America, **2** Department of Computer Science, University of British Columbia, Vancouver, Canada, **3** Alberta Machine Intelligence Institute (Amii), Edmonton, Canada, **4** Department of Mathematics, University of British Columbia, Vancouver, Canada

* zhaow@wfu.edu

OPEN ACCESS

Citation: Zhao W, Sutherland DJ, Dao Duc K (2026) Fast and interpretable quantification of biological shape heterogeneity via stratified Wasserstein kernel. PLoS Comput Biol 22(5): e1014254. https://doi.org/10.1371/journal.pcbi.1014254

Editor: Emma Lejeune, Boston University, UNITED STATES OF AMERICA

Received: December 3, 2025

Accepted: April 20, 2026

Published: May 7, 2026

Data availability statement: This study did not generate new data. Both processed (aligned and resampled) 2D cancer cell shape data are available at https://github.com/wxli0/dyn/tree/main%4092c7a58/dyn/datasets/breast_cancer. The 3D mitotic cell shape dataset is available at https://open.quiltdata.com/b/allencell/

Abstract

Modern imaging technologies produce vast collections of cellular and subcellular structures, calling for principled methods that enable shape comparison across individuals and populations. We introduce the stratified Wasserstein framework, which treats each shape as an unstructured point cloud and embeds it into Euclidean space via ranked local distance profiles. This embedding yields an isometry-invariant Euclidean distance and a positive-definite kernel for population analysis, with a consistent sample-based estimator that supports large datasets in near-quadratic time. By leveraging kernel methods, the framework enables statistically rigorous tasks such as nonparametric hypothesis testing, providing theoretical guarantees as well as interpretability. We demonstrate our framework's applicability to large-scale biological datasets. Analyzing 2D cancer cell contours, we quantify population-level discrepancies and identify representative cells contributing most strongly to the observed differences. Using 3D volumes of cell envelope and nucleus, we reveal progression patterns that capture morphological changes across cell populations both at the level of individual shapes. These results establish a simple and principled tool for population-level biological shape analysis, with potential impact across diverse domains of computational imaging and data science.

Author summary

Biological structures come in many shapes, from whole tissues to single cells, and even protein conformations. Modern imaging technologies now produce enormous collections of these shapes, giving us the opportunity to study how structure varies across conditions or evolves over time. However, it is still difficult to compare large numbers of complex shapes in a way that is both fast and interpretable. Many existing methods rely on hand-chosen features or landmarks, while others are too slow to apply to the large datasets now common in biology.

packages/aics/mitotic_annotation. A repository containing Python code for the 2D and 3D cell shapes for reproducing the results is available at https://github.com/WenjunZHAOwO/StratifiedShapes.

Funding: WZ was partially supported by a PIMS-Simons Postdoctoral fellowship and PIMS Kantorovich Initiative. DJS acknowledges support from the Canada CIFAR AI Chairs program. KDD was supported by a NSERC Discovery Grant RGPIN-2020-05348. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

In this work, we introduce a broadly applicable method that compares shapes directly from their raw geometry. Our approach summarizes each shape using its intrinsic distance, without relying on landmarks or alignment, and is efficient enough to handle large datasets. Because the method is built on solid statistical principles, it allows us to determine when two groups of shapes differ in meaningful ways and to identify the examples that contribute most to those differences. We also show that the method can reveal smooth morphological trends, allowing researchers to trace how shapes evolve along biological trajectories. Overall, our framework provides a general, scalable, and statistically grounded tool for analyzing shape variation across diverse biological systems.

Introduction

Recent advances in high-throughput imaging have generated massive collections of biological shapes across multiple scales, from whole tissues to single cells, subcellular structures, and protein conformations. These datasets enable population-level analyses of morphology across conditions, but they also expose a methodological gap: we lack tools that are both interpretable and computationally efficient for comparing shapes at the level of individual objects and at the level of distributions of objects. Questions such as how morphology evolves through the cell cycle [1] or how protein structures differ across species [2] call for comparisons that respect intrinsic geometry, scale to large datasets, and connect naturally to statistical analysis and machine learning.

Traditional approaches for shape comparison face well-known limitations. Many methods rely on reducing data to specific shape features, such as volumes, heights, persistent barcodes in topological data analysis [3], spherical harmonics [4], with features of interest being established *a priori*. Landmark-based analysis pipelines quantify the pairwise discrepancy through the correspondence of specific landmark points, which are manually placed points and can be labor-intensive, subjective, and difficult to reproduce [5] for large-scale datasets. Other classical parametric methods are tailored to particular domains [6,7], and/or require heavy pre-processing including down-sampling, alignment, or interpolation [8,9]. In contemporary imaging pipelines with thousands to millions of shapes, these constraints become bottlenecks.

While most existing work has also focused on pairwise distances or shape alignment at the individual level, population-level metrics between shape ensembles remain underexplored. Historically, this gap can be attributed to the relatively small size of shape datasets, where analysis was limited to tens or hundreds of samples. With modern imaging pipelines now producing thousands to millions of shapes, the need for scalable and statistically principled population-level distances has become pressing. A few population-level approaches have been proposed. For example, Fréchet mean and distances to all shapes [10], linear subspace embeddings and Kullback–Leibler divergence [11] have been considered in specific contexts, but they

are tied to specific assumptions, and none provide a general-purpose, theoretically grounded, and computationally efficient framework for comparing populations of arbitrary shapes.

In this context, optimal transport theory provides a principled way to compare shape objects given as unstructured point cloud or their probability histograms. In particular, Gromov–Wasserstein distances compare shapes through their internal pairwise distances and avoid explicit alignment [12]. Related lines of work summarize shapes by distributions of distances, either globally or around each point [12–15]. However, it is well-known that such computations can be intractable [16] and expensive [17], and consequently, they are hard to be applied to datasets at scale. Moreover, it is difficult to build positive definite kernels directly from these distances in dimensions greater than one, which limits downstream statistical tools such as kernel hypothesis tests [18–20] and representation learning [21,22]. Existing works [23,24] reconcile the issue by slicing the measure and reducing the transportation problem to one-dimensional, however, that requires sampling angles of slicing and may not fully utilize the information from the whole shape.

We address these challenges by introducing the *stratified Wasserstein distance*, a simple yet highly effective framework that embeds each unstructured point cloud into an Euclidean space. We show that the distance is Hermitian and produces a kernel that is positive definite and, under mild regularity conditions, characteristic. The construction is invariant to isometries by design and is injective up to isometry. Computationally, the method is nearly quadratic in number of points per shape and outperforms the Gromov–Wasserstein, which is cubic after acceleration via entropic regularization. We illustrate the utility of the framework across diverse biological settings on 2D and 3D shape datasets, focusing on both individual and population-level shape analysis.

Results

We begin by presenting an overview of the stratified Wasserstein framework, describing its construction and key properties. This provides the conceptual and theoretical foundation for the remainder of the paper. We then demonstrate its performance on two and three-dimensional biological imaging datasets, highlighting its ability to handle diverse shape types and to support population-level statistical analyses such as hypothesis testing and dependence detection.

Overview of the stratified Wasserstein framework

We propose stratified Wasserstein, a framework that embeds into Euclidean space each shape, represented as an unstructured point cloud, and facilitates kernel methods in that space for population-level quantification tasks. Compared to existing shape distances such as Gromov–Wasserstein and its lower bounds [12] as defined in Eqs M3, M4, and M5 (Methods), the proposed framework achieves similar discriminative power while being computationally more efficient, with complexity nearly quadratic in the number of points and empirical runtimes typically below 1% of those required by Gromov–Wasserstein methods. The induced kernel is characteristic, so population-level statistics via kernel methods retain their standard statistical guarantees, including consistency and power against alternatives [18,19,25]. Fig 1 provides a schematic overview of our procedure, that takes local distance distribution to produce shape embeddings and derive population-level statistics. Here we describe the main construction and summarize its key properties; detailed statements and proofs are deferred to section Methods and the Supplementary Information.

First, we represent each shape by a point cloud together with an intrinsic distance, such as the geodesic distance on the surface or the Euclidean distance in ambient space. For every point on the shape, we compute the distribution of distances to all other points, which captures its local geometric structure. We then rank the points according to a continuous functional of these local distance distributions, which induces a one-dimensional stratification of the shape. Within each stratum, we summarize the local geometry by recording a set of quantiles of the distance distributions. This defines an embedding of the shape into a two-dimensional function space indexed by the stratification variable and the quantile level. The stratified Wasserstein distance between two shapes is simply the L^p distance between their embeddings, computed over this two-dimensional domain.

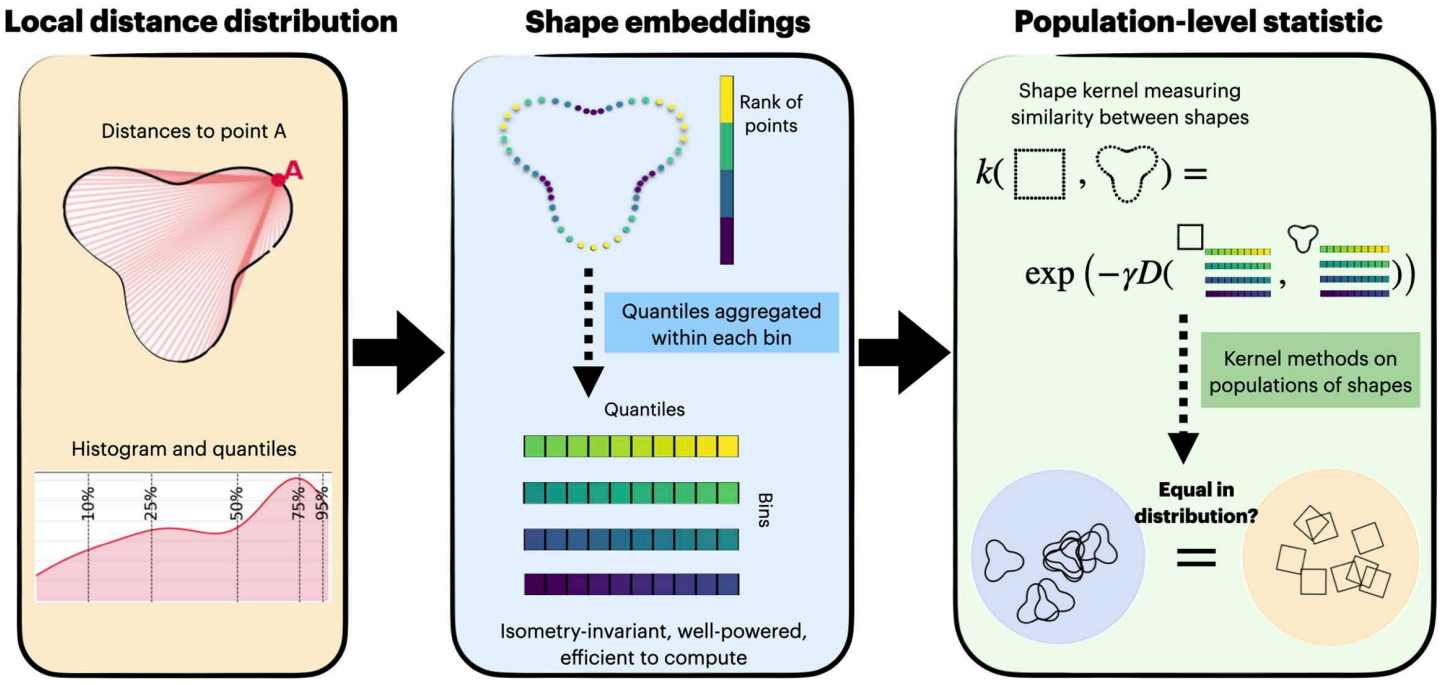

Fig 1. A summary of our procedure. Given unstructured point cloud in 2D or 3D representing a shape of a wide range of types, we first compute the intrinsic distances on each shape and localized for each point. Using these local distances, a wide range of downstream tasks can be performed given a large population of shapes, such as dimension reduction, clustering, hypothesis testing, and feature selection. The detailed methodology is presented in subsequent sections, with rigorous theoretical foundation discussed.

https://doi.org/10.1371/journal.pcbi.1014254.g001

In this work, we show that the stratified Wasserstein distance inherits many of the desirable properties of Gromov–Wasserstein while significantly improving computational efficiency. It is naturally invariant to isometric transformations because of the invariance property of local distance distributions. Under mild regularity conditions on the ranking functional used for disintegrating the shape, the distance is injective up to isometry: two shapes have zero distance if and only if they are isometric. Because the distance is induced by an L^p norm in the embedded space, standard kernels such as the Gaussian and Laplace built from it are positive definite. When the distance is injective, these kernels are also characteristic, meaning they can distinguish any two distributions of shapes. The estimator of the distance based on finite point clouds is statistically consistent: as the number of points per shape grows, with discretization refined accordingly, the estimated distance converges to the population distance. The computational cost scales nearly quadratically with the number of points per shape, which is far more efficient than Gromov–Wasserstein methods.

Finally, because the kernel is characteristic, standard population-level statistics built from it inherit strong theoretical guarantees. For example, maximum mean discrepancy (MMD) [18] measures any form of discrepancy between two populations, while the Hilbert–Schmidt Independence Criterion (HSIC) [19] measures the strength of dependence between variables. Both statistics are consistent and have power against all fixed alternatives, which enables rigorous nonparametric two-sample testing and dependence detection between shape distributions and external covariates.

We provide Table 1 to compare our proposed stratified distance between shapes against other relative alternatives. Compared to the other distances, stratified Wasserstein is Hilbertian, can be consistently estimated from finite samples, and achieves the lowest asymptotic computational complexity. It is conditionally injective and yields a characteristic kernel if the oracle functional for sorting and binning is chosen carefully. The balance between discriminative power and computational complexity is further demonstrated in numerical examples on real world biological shape

Table 1. Comparison between distances with shapes represented as point clouds in 2D/3D. For Gromov–Wasserstein and Wasserstein between local distances, computational cost for both exact computation (left) and entropic approximation (right) are both shown. N = points per shape. * Exact GW is widely believed to be NP-hard [16]; † Entropic GW complexity depends on λ (regularization) and ϵ (tolerance) [26], both are small, positive values for accurate approximation; ‡ Systematic studies on computational costs can be found in [27–29]. § Our framework has a conditionally injective and characteristic kernel under Hypotheses (H1) and (H2) with details in Supplementary Information.

point-cloud based distance	injective	Hilbertian	characteristic kernel	consistent	computational cost
Gromov–Wasserstein	✓ [12]	×[12]	×	✓ [12]	intractable* / $\mathcal{O}(N^3/\sqrt{\lambda\epsilon})^\dagger$
Wasserstein (local distances)	✓ [14]	×[30]	×	✓ [31]	$\mathcal{O}(N^3)$ / $\mathcal{O}(N^2/\lambda\epsilon)^\ddagger$
Wasserstein (global distances)	× [13]	✓[24]	×	✓ [15]	$\mathcal{O}(N^2 \log N)$
Stratified Wasserstein **(this work)**§	✓	✓	✓	✓	$\mathcal{O}(N^2 \log N)$

https://doi.org/10.1371/journal.pcbi.1014254.t001

datasets as discussed in the next section. Empirical comparison on these distances on simple synthetic shapes can be found in Figs C and D in S3 Text.

Breast cancer cell contour shapes from fluorescence microscopy

We apply our framework to 2D cancer cell shapes obtained from fluorescence microscopy [32,33]. Cell images are binarized, and their boundaries are extracted to form discrete curves given by the 2D coordinates of the cell contours. The dataset includes cancer cell shapes from three different cell lines for breast cancer, corresponding to: (1) MCF10A (228 cells): non-tumorigenic human breast mammary gland epithelial cell line, which is a classic model for normal cells, (2) MCF7 (225 cells): breast cancer line with relatively low metastatic potential, and (3) MDA-MB-231 (abbreviated MDA, 224 cells): a triple-negative breast cancer cell line that is highly invasive and commonly used as a model for metastatic progression. Given populations of these cell contours, the goal is to test whether the underlying shape distributions differ significantly across groups defined by cell lines. Understanding these shape differences may help reveal whether and how cell morphology encodes functional behaviors for different cancer types. An example of the raw images, as well as random samples of cells from each cell line are visualized in Fig 2A and 2B, and more images can be found in S1 Fig. Implementation details can be found in Methods and Table 2.

Relative MMD test reveals similarity between triple negative cancer cells and non-cancerous cells: Upon testing relative MMD between three populations, we observed that: Despite being a cancer line, MDA resembles MCF10A (normal cells) more than MCF7 (low-metastasis cancer). The empirical estimate of squared MMD reveals that MMD^2(MCF10A,MDA)=0.0189, much smaller than the estimate of MMD^2(MCF7,MDA)=0.1409. To test whether the difference is statistically significant, we perform a relative MMD test for the following:

$$H_0 : \mathrm{MMD}^2(\mathrm{MCF7}, \mathrm{MDA}) = \mathrm{MMD}^2(\mathrm{MCF10A}, \mathrm{MDA}), \quad H_1 : \mathrm{MMD}^2(\mathrm{MCF7}, \mathrm{MDA}) \neq \mathrm{MMD}^2(\mathrm{MCF10A}, \mathrm{MDA}).$$

The p-value is computed using a permutation test that pools MCF7 and MCF10A, under the null hypothesis that the two populations are equally distant from MDA. Specifically, we permute the MCF7/MCF10A labels and recompute the difference in squared MMD to MDA for each shuffle. With 1000 permutations, this test results in a p-value of 0.001 (S1C Fig), suggesting that the MDA line (high metastatic cancer) is morphologically closer to MCF10A (normal cell line) than to MCF7 (low metastatic cancer). This observation is consistent to findings from gene expression studies [34], which reported that both MCF10A and MDA share a basal-like subtype, while MCF7 belongs to the luminal class. Our shape-based results suggest that cell morphology is more closely aligned with molecular subtype (basal versus luminal) than with cancer status.

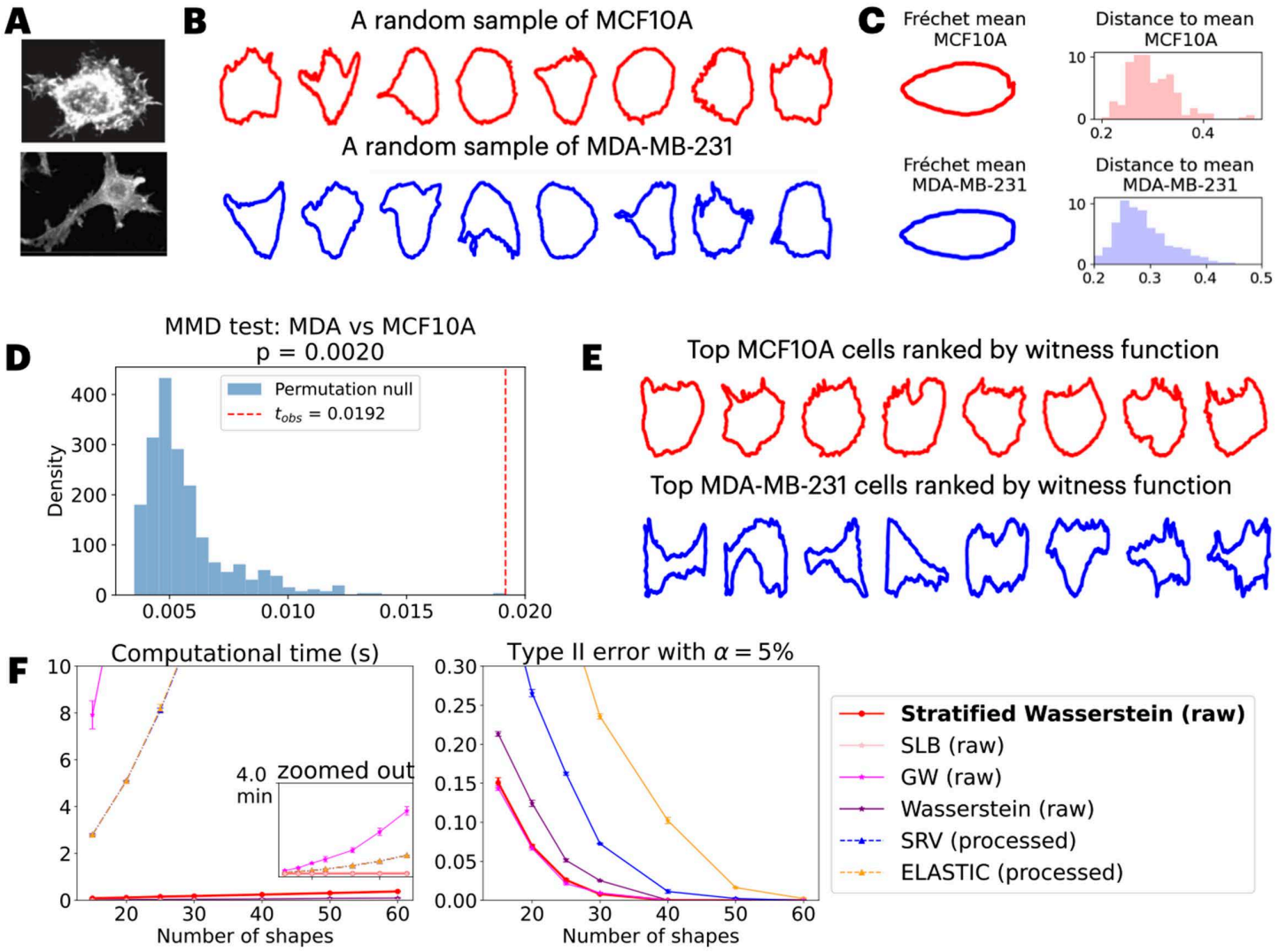

Fig 2. Case study on 2D cancer cell shape populations represented by contours. A: Example fluorescence microscopy images of cell lines curated by [32]. B: Cell contours after segmentation, with a random sample of 8 shapes from each population visualized. C: Inspection of first and second moments suggests similarity between the MCF10A and MDA populations, as shown via Fréchet mean shape and distances to the mean shape within each population under the SRV metric, computed using Geomstats [10]. D: To further probe differences between MCF10A and MDA, we compute the MMD-based test statistic and its p-value using our proposed stratified kernel, leveraging all available cell shapes. E: Cells with the top 8 witness scores in each population highlight those whose shapes differ most from the respective population majority. F: Empirical probability of Type II error and computational time under small (15) to moderate (60) sample sizes, benchmarked across various kernel-based tests, with results averaged over five independent trials (error bars: ± 1 SD).

https://doi.org/10.1371/journal.pcbi.1014254.g002

The result is also consistent with the prior approach [35] that computes Fréchet mean shape under the square root velocity (SRV) metric [6] through geomstats [10] package. The mean shapes of processed contours from all 3 populations are shown in S1B Fig, where MCF10A and MDA mean shapes are both elongated and almost coincide.

Absolute MMD test indicates discrepancy between populations on the tails: Given almost identical mean shapes and distribution of distance to mean shown in Fig 2C, it seems that the two shape populations of MDA and MCF10A agree up to mean and variance. A natural follow-up question is whether the generating distributions of the two shape populations

Table 2. Summary of experimental parameters and implementation details for 2D and 3D shape analyses.

Parameter	2D cancer cell shapes	3D mitotic cell shapes
Input data	Pre-segmented binary masks [32]	Binary cell and nucleus volumes [1]
Shape representation	2D contour point clouds	3D surface point clouds
Point extraction	Provided with dataset [47]	Binary erosion + boundary voxels
Intrinsic distance	Euclidean distance	Euclidean distance
Distance normalization	Median normalized to 1 (per shape)	Not normalized*
Points per shape (used)	All contour points (26–879)	200 (cell), 200 (nucleus)
Points per shape (GW)	100	100 (cell), 100 (nucleus)
Ranking functional $r_S(x)$	Mean of local distances	Mean of local distances
Number of bins	100	100
Quantiles per bin	100	100
Quantile computation	`numpy.quantile` (default)	`numpy.quantile` (default)
Embedding dimension	10,000	20,000 (10,000 cell + 10,000 nucleus)
Kernel on shapes	Laplace	Laplace
Kernel bandwidth	Median heuristic	Median heuristic
Permutation tests	1,000	1,000
GW solver	`gromov_wasserstein2` (POT [45])	`gromov_wasserstein2` (POT [45])
GW formulation	Non-entropic, squared loss	Non-entropic, squared loss
GW marginals	Uniform	Uniform
GW stopping criteria	Default POT settings	Default POT settings

*For the 3D mitotic cell data, distances were not normalized in order to preserve biologically meaningful size variation during mitosis across cell-cycle stages.

https://doi.org/10.1371/journal.pcbi.1014254.t002

are truly identical. To this end, we zoom in on the unexpected pair and perform an MMD test on those two populations, which detects arbitrary order of discrepancy in distribution that extends beyond second order:

$$H_0 : \mathrm{MMD}^2(\mathrm{MCF10A}, \mathrm{MDA}) = 0, \quad H_1 : \mathrm{MMD}^2(\mathrm{MCF10A}, \mathrm{MDA}) > 0.$$

With 1,000 permutations, it shows a p-value of 0.002 (Fig 2D), suggesting a significant difference in the two populations. To localize the morphological differences, we examine extreme shapes identified by the MMD witness function, which highlights regions where the shape distributions differ most. Top 8 cells in each population with the highest values of the MMD witness function are visualized in Fig 2E. The extremal subset of MDA cells exhibits highly irregular and protrusive morphologies, which are key contributors to the shape-based distinction from MCF10A. We hypothesize that such enriched shapes could correspond to invasive behaviors as known for MDA [36]. Notably, the statistical difference in MMD is not driven by the mean shape or the distances to mean shape (Fig 2C), but by this tail population in MDA, which can only be captured using nonparametric test like the kernel MMD.

The tests are computationally efficient and statistically well-powered: To assess the power of the absolute and relative tests, we perform randomized experiments to quantify the empirical Type I and II error rates. We compare our kernels to the framework suggested by Zhang et al [37] using square-root velocity (SRV) metric and its variant (Elastic) for MMD. To ensure a fair comparison with methods that require smoother, parameterized curves, the benchmark approaches involving SRV and elastic metrics are applied to preprocessed data, where each shape has been interpolated to have 2,000 points and aligned to the same reference shape beforehand. In contrast, in our approaches, we perform the test directly on raw data without any processing.

We perform the tests for sample sizes varying between 15 and 60 via MMD. To systematically quantify the error, for each sample size, we resample and repeat the experiments 1,000 times, compute the *p*-value via permutation test, and report the relative frequency of making an error in Fig 2F. For Type II error, we draw one sample from each group, perform the test, and compute the relative frequency that the equality of shape distributions cannot be rejected. Among our proposed distances, Gromov–Wasserstein consistently achieves the lowest Type II error, followed by stratified Wasserstein distances and the second lower bound (SLB) of Gromov–Wasserstein. All Wasserstein-type distances outperform SRV-based metrics at sample sizes under 40, which is consistent with the fact that Wasserstein-type distances operate on distributions of intrinsic distances and are less sensitive to curve parameterization and discretization effects. Similar experiments and observation on error rates for both Type I and II are done for relative test and shown in S1 Fig. We also note that the error rate is robust to the choices of hyperparameters, and sensitivity results are reported in S2 Fig.

To assess computational efficiency, we record the wall time required to compute all pairwise distances between shapes in each subsample, averaged over five independent runs. These results are shown in Fig 2. While GW offers the best statistical power, its computational cost is substantial. Our stratified Wasserstein achieves a favorable balance: while offering second best Type II error control, slightly worse than Gromov–Wasserstein, it is significantly more efficient than Gromov–Wasserstein and its second lower bound, and is second only to the pooled Wasserstein method, which simply aggregates and sorts all pairwise distances.

In sum, our nonparametric, distribution-level approach operates directly on raw contour shapes, detects tail effects, and localizes the drivers of discrepancy while remaining computationally practical.

Allen institute 3D cell and nucleus shapes

In this example, we demonstrate that quantiles of distance matrices encode meaningful morphological information, using 3D cell imaging data from Viana et al. [1]. We analyze a subset of 5,764 cells used by the original authors to train a classifier, including annotations for cell types reflecting six mitotic stages (M0, M1M2, M3, M4M5, M6M7 early, and M6M7 half) as well as three outlier types (blob, dead, wrong). Implementation details on our method can be found in Methods and Table 2. Existing dimension reduction methods rely either on hand-crafted features (e.g., volume and height of the cell and nucleus) [1] or on learned latent representations from deep models such as variational autoencoders [38], followed by UMAP. However, these methods are either highly specialized for the application or require significant training effort.

Dimension reduction for shapes reveals relevant morphological variations. We pre-process each binary image by eroding it to extract surface points, followed by down-sampling to retain 200 points per cell. For each cell, we divide it into 100 bins, compute 100 quantiles for both the cell shape and the nucleus shape in each bin, concatenate them, and apply UMAP to the resulting 20,000-dimensional vector. Fig 3A shows 2D UMAP embeddings obtained using stratified Wasserstein distance. To benchmark our result, we performed dimension reduction using 4 additional methods and visualized them in Fig 3B that are based on (1) Euclidean distance in raw binary images, (2) dominant features through volumes of cell and nucleus, (3) the intermediate latent layer using a pre-trained PointNet model [39], and (4) Gromov-Wasserstein distance with each cell downsampled to 100 points due to prohibitive computational cost. For a fair comparison, we also applied the proposed stratified Wasserstein embedding to shapes downsampled to 100 points. As shown in S4 Fig, the resulting embedding is qualitatively similar to that obtained with 200 points. Our dimension reduction best preserves the cyclic and continuous nature of the shape dynamics, consistent with the feature-based and PointNet embeddings, but with a more smooth trajectory that reflects the remark on the ambiguity of manual annotation by Viana et al. [1].

This embedding provides biological insight into morphological changes during the mitotic cycle. In Fig 3C, we color cells by morphological features to show how these evolve along the cycle. Starting from the M1M2 phase, both cell and nucleus volumes increase significantly, followed by a marked rise in nucleus height during M3 and M4M5. These trends

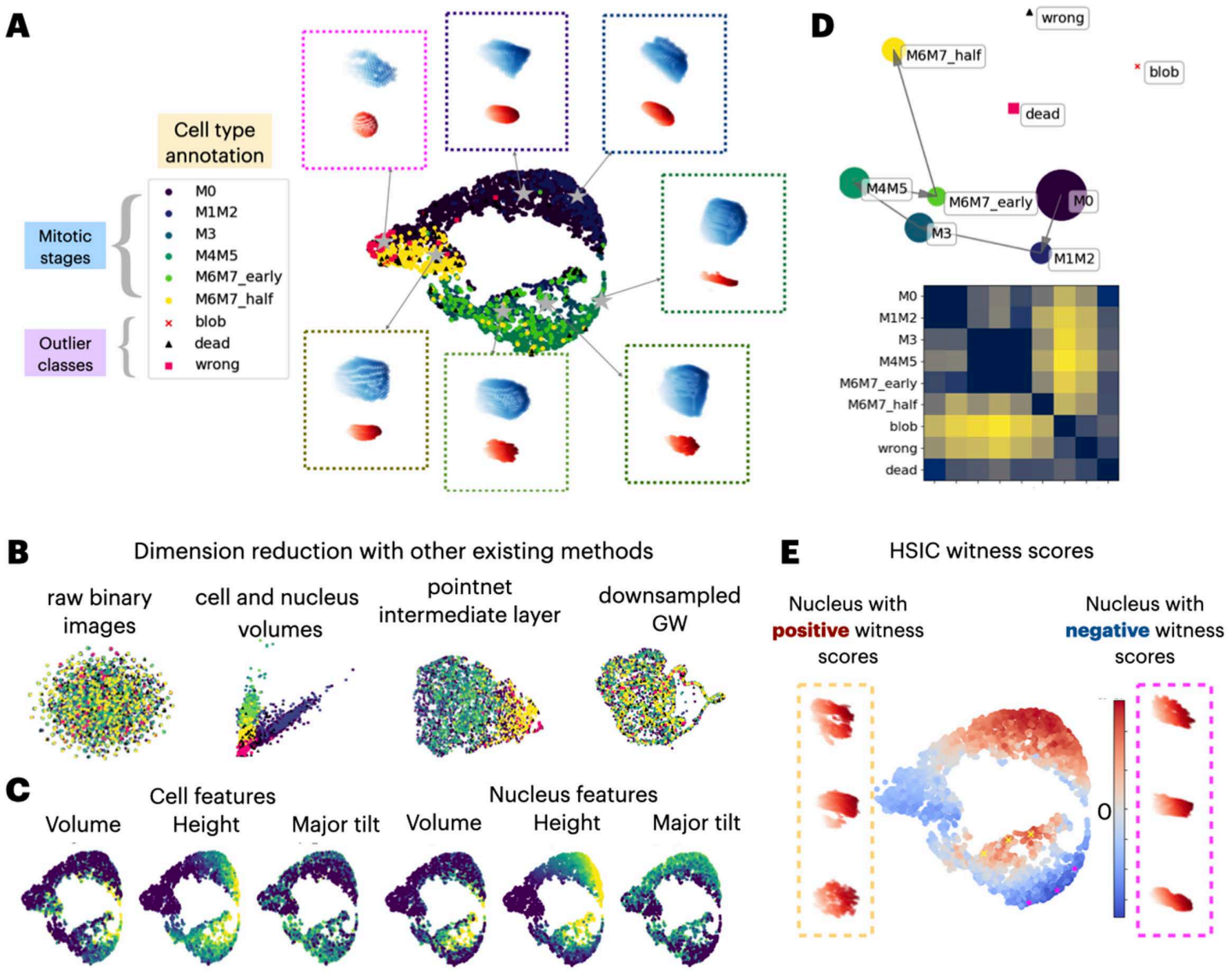

Fig 3. Case study on 3D cell and nucleus shape populations. A: UMAP embedding of the annotated mitotic dataset from Viana et al. [1] through stratified local distance distributions, and representative cell (blue) and nucleus (red) shapes along the cycle. B: Cell-level embeddings from standard methods fail to reveal dynamics. Euclidean distances between raw binary images and Gromov–Wasserstein distances on downsampled point clouds show no meaningful pattern, while feature-based embeddings (volumes or neural networks) capture heterogeneity but not the cyclic continuum of mitosis. C: Cell features displayed against UMAP allows for interpretation of shape changes within each stage. D: Population-level embedding using proposed MMD reveals cyclic process consistent to A. E: HSIC witness scores for individual cells, revealing regions with mixed positive and negative values near late mitotic stages. We selected 3 cells with high positive witness scores (yellow crosses) and 3 with high negative witness scores (magenta stars), and display the nucleus shapes on the sides.

https://doi.org/10.1371/journal.pcbi.1014254.g003

align with the known stages of mitosis, where volume increase reflects DNA and organelle replication, and structural transitions (e.g., chromatin condensation and nuclear envelope breakdown) are hallmarks of the later phases [40]. The outlier cells have particularly smaller volumes and heights in both cell and nucleus shapes, likely reflecting incomplete mitosis, abnormal states, or segmentation artifacts.

Embeddings of populations respects mitotic cycle progression. To better summarize discrepancies between shape populations corresponding to different mitotic stages, we embedded the cell populations using MMD as a distance metric between groups. Fig 3D displays the resulting distance matrix, which exhibits a banded structure: populations from adjacent stages (e.g., M0 and M1M2) have lower distances, while outlier cell types are farther from the mitotic populations. Since MMD defines a valid population-level distance, we further applied multidimensional scaling (MDS) to embed the populations in 2D space, revealing a clear progression through mitotic stages. We note that two alternative population metrics, energy distance (and Wasserstein distance), do not perform as nicely as MMD-based embedding, likely due to the fact that the kernel methods are more resistant to noise through smoothing, which is crucial for real imaging data. Comparison to other candidates of population-level distances can be found in S3 Fig.

HSIC witness scores locate representative shapes for mitotic progression. HSIC witness scores, much like the witness function used in MMD-based testing, provide a way to interpret how individual samples contribute to the overall dependence between variables. We exclude the outlier cell types, and used continuous values of mitotic stages to better capture the periodic nature by assigning equally spaced values from 0 to 2π, with von Mises kernel $k^{\gamma,\text{periodic}}(y_1, y_2) = \exp(\gamma \cos(y_1 - y_2))$. When applied to cell and nuclear shapes across the cell cycle, these scores highlight specific regions where the shape is particularly informative or uninformative of mitotic stage (Fig 3E). To better understand these regions, we visualized representative cells with the strongest positive and negative witness scores. Interestingly, although all selected cells belong to late mitotic stages, their nuclear morphologies differ substantially. Cells with positive witness scores (yellow crosses) tend to show bi-lobed or segmented nuclear shapes, consistent with active division, while those with negative scores (magenta stars) exhibit smooth, undivided nuclei. This suggests that the HSIC witness score captures subtle morphological differences that reflect how well shape aligns with mitotic progression, helping to identify stage-aware prototypes with distinct developmental states.

Overall, our framework recovers the known mitotic trajectory while pinpointing rare, stage-specific morphologies, demonstrating the potential to reveal unknown heterogeneity within data.

Discussion

To provide tools for quantifying heterogeneity in shape data at both the individual and population levels, we proposed a stratified Wasserstein framework to embed shape data into a Euclidean space and utilize kernel methods therein. The construction is naturally invariant under isometric transformations and admits a consistent sample estimate for both shape–shape distances and between-population distances via maximum mean discrepancy (MMD). On 2D breast-cell contours and 3D mitotic cell and nuclear segmentations, the framework supports dimensionality reduction, clustering, and nonparametric hypothesis testing, matching or surpassing state-of-the-art methods while substantially reducing runtime. Although we focus on smooth planar curves and surfaces, the formulation extends to other metric–measure data types, e.g., neuron trees modeled as metric graphs, when equipped with a geodesic distance and mild regularity [14,41]. Because it operates directly on unstructured point clouds with variable sizes and requires no landmarks or global alignment, the approach offers a unified and scalable route to shape analysis across domains. More broadly, our framework has potential applications across domains that involve the quantification of shape/graph data, such as brain images, protein density maps, and social networks. Unlike deformation-based shape models, the stratified Wasserstein framework is not intended to provide an invertible parameterization of individual shapes or to define geodesic deformations between shapes, but rather to support statistically principled and computationally efficient comparison of shapes and shape populations.

Next, we outline several limitations and potential avenues for improvement. First, in practice we rank points by the mean of their local distance distributions; on near-symmetric or distance-degenerated shapes (Fig B in S3 Text), this statistic can be weakly discriminative. While the performance is comparable to Gromov–Wasserstein and second lower

bound for random polygons (Fig C in S3 Text), the shape distance under the stratified framework is significantly overestimated on T-shapes with symmetry (Fig D in S3 Text). A principled remedy is a lexicographic rank using the first k moments (or quantiles) of the local distance laws; we show such rankings become injective on finite samples as $k \to \infty$, but selecting a small, data-adaptive k and a corresponding ranking functional remains open. Second, the numbers of bins and quantiles are user-chosen without specific guidelines. Deriving finite-sample error bounds for the stratified estimator would enable data-driven defaults that balance discretization bias against bin variance. Third, the kernel bandwidth is currently set by the median-distance heuristic or ad hoc tuning; under low signal-to-noise, a common pitfall in current imaging, more principled rules (e.g., maximizing estimated MMD test power or kernel alignment, or noise-aware plug-ins) could improve performance. Addressing these issues would further strengthen robustness and ease of use for practitioners.

The proposed framework has the potential to facilitate new computational tools and biological discoveries in multiple ways. By connecting shape space to kernels, the framework unlocks a broad toolbox for inference and representation. (i) With multimodal measurements (e.g., shapes and transcriptomics), kernel-based conditional independence (KCI) [20] can yield p-values for identifying driver genes that explain morphology while controlling confounders, aiding inference on gene regulation underlying shape change. (ii) Coupled with kernel representation learning [22] and functional data analysis (e.g., kernel PCA or Gaussian processes on the embedding), one can learn low-dimensional surrogates that capture discrete phenotypes or continuous trajectories along which shapes vary most rapidly. (iii) Given its computational efficiency, the framework can be integrated into in vivo perturbation screens to prioritize conditions that induce the largest distributional shifts in shape. As a whole, these analyses turn the framework into a practical engine for hypothesis testing, causal inference, perturbation discovery, and more–at scale.

By addressing the challenges listed above and further extending the framework to other tasks, stratified Wasserstein has the potential to become an even more powerful tool, enabling more comprehensive insights into modeling the highly heterogeneous shape space across diverse biological contexts.

Methods

Shape representation and intrinsic distances

We represent each shape as an unstructured point cloud $\mathcal{S} = \{x_i\}_{i=1}^N \subset \mathbb{R}^D$ with $D \in \{2, 3\}$. Each shape is equipped with an intrinsic distance $d_{\mathcal{S}} : \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ and a uniform measure $\mu_{\mathcal{S}}$. Distances can be chosen flexibly, depending on the application, including Euclidean distance, geodesic distance on a k-nearest neighbor graph [41], or diffusion distance [42]. The intrinsic distance matrix is $C_{ij} = d_{\mathcal{S}}(x_i, x_j)$.

A shape statistic central to this work is the local distance distribution, defined as follows. For each point $x \in \mathcal{S}$, the local distance distribution is the pushforward measure

$$\nu_x^{\mathcal{S}} := (d_{\mathcal{S}}(x, \cdot))_{\#}\mu_{\mathcal{S}},$$

where $\mu_{\mathcal{S}}$ denotes the uniform probability measure on $\mathcal{S}$ and # denotes the pushforward operator. In the discrete setting of a point cloud $\mathcal{S} = \{x_i\}_{i=1}^N$, this reduces to the empirical measure

$$\nu_{x_i}^{\mathcal{S}} = \frac{1}{N-1} \sum_{j \neq i} \delta_{d_{\mathcal{S}}(x_i, x_j)}. \tag{M1}$$

For comparison, one may also define the global distance distribution of a shape $\mathcal{S}$ as the pushforward

$$\nu^{\mathcal{S}} := (d_{\mathcal{S}})_{\#}(\mu_{\mathcal{S}} \otimes \mu_{\mathcal{S}}), \tag{M2}$$

which captures the overall distribution of pairwise distances on the shape. In the discrete setting, this corresponds to the empirical distribution of all pairwise distances $\{d_{\mathcal{S}}(x_i, x_j)\}_{i \neq j}$. The local distance distributions $\{\nu_x^{\mathcal{S}}\}_{x \in \mathcal{S}}$ are known to have better discriminative power [12,14] compared to global distance distributions for their ability to retain pointwise geometric information and are not recoverable from $\nu^{\mathcal{S}}$ alone.

For scale invariance, shapes are rescaled so that the median pairwise distance equals one.

Gromov–Wasserstein and related distances

Gromov–Wasserstein distance. For two metric measure spaces $(\mathcal{S}_i, d_i, \mu_i)$, $i = 1,2$, the p–Gromov–Wasserstein (GW) distance [12,43] is defined by

$$GW_p(\mathcal{S}_1, \mathcal{S}_2) = \left(\inf_{\pi \in \Pi(\mu_1, \mu_2)} \iint_{\mathcal{S}_1^2 \times \mathcal{S}_2^2} \left| d_1(x, x') - d_2(y, y') \right|^p \, d\pi(x, y) \, d\pi(x', y') \right)^{1/p}. \tag{M3}$$

In the discrete setting, this corresponds to minimizing over coupling matrices between points. Exact computation is widely believed to be NP-hard [16], and entropic regularization is typically used to obtain approximate solutions with complexity $\mathcal{O}(N^3/\sqrt{\lambda\epsilon})$ [26]. In all numerical experiments, we used the implementation provided by Python Optimal Transport [44,45].

Global Wasserstein distance. A simpler lower bound compares the global distributions of pairwise distances [13,15,46]. Let F_1^{-1} and F_2^{-1} be the quantile functions of the empirical global distance distributions defined as in Eq (M2). The p–Wasserstein distance between these distributions is

$$d_p^W(\mathcal{S}_1, \mathcal{S}_2) = W_p(c_1, c_2) = \left(\int_0^1 |F_1^{-1}(q) - F_2^{-1}(q)|^p \, dq \right)^{1/p}. \tag{M4}$$

This distance can be computed efficiently in $\mathcal{O}(N^2 \log N)$, but it is not injective: different shapes may yield the same global distance distribution. An illustrating example [12] can be found in Fig A in S3 Text.

Second lower bound (SLB). A stronger metric compares local distance distributions [12,14]:

$$\mathrm{SLB}_p(\mathcal{S}_1, \mathcal{S}_2) = \left(\inf_{\gamma} \iint W_p^p(\nu_x^{\mathcal{S}_1}, \nu_y^{\mathcal{S}_2}) \, \gamma(x, y) \, dx \, dy \right)^{1/p}, \tag{M5}$$

where $\nu_x^{\mathcal{S}}$ is the local distance distribution at point x and γ is a coupling between points. SLB is strictly stronger than the global Wasserstein distance and is injective for smooth closed shapes under regularity assumptions [14], but remains computationally demanding and is not Hermitian.

Stratified Wasserstein distance

The proposed stratified Wasserstein distance combines the geometric discriminative power of GW-type distances with the efficiency of quantile-based embeddings. For each point $x \in \mathcal{S}$, we compute its local distance distribution $\nu_x^{\mathcal{S}}$, and further summarize it by a scalar ranking functional $r_{\mathcal{S}}(x)$, which represents the 1D distribution in a way that reflects the point's geometric context within the shape. In practice, we use the mean of local distance distributions, which tends to separate points near the shape barycenter from those near peripheral or protrusive regions.

To make this descriptor comparable across shapes and invariant to monotone transformations, we convert $r_{\mathcal{S}}(x)$ into a normalized rank

$$U_{\mathcal{S}}(x) := F_{\mathcal{S}}(r_{\mathcal{S}}(x)),$$

where $F_{\mathcal{S}}$ denotes the cumulative distribution function of $r_{\mathcal{S}}(X)$ for $X \sim \mu_{\mathcal{S}}$. This transformation maps points on the shape to the unit interval [0,1], inducing a one-dimensional stratification of the shape. The stratification preserves the ordering of points by geometric context while discarding absolute scale information, enabling consistent aggregation of local geometry across shapes.

The shape measure $\mu_{\mathcal{S}}$ is then disintegrated along $U^{\mathcal{S}}$,

$$\mu_{\mathcal{S}} = \int_0^1 \mu_{\mathcal{S}}^u \, du,$$

where $\mu_{\mathcal{S}}^u$ denotes the conditional law at stratum u. We define the function Q by evaluating the quantile function of ν at level $q \in [0, 1]$:

$$Q(\nu, q) \;:=\; F_\nu^{-1}(q),$$

where F_ν^{-1} denotes the (left-continuous) generalized inverse of F_ν. For each stratum $u \in [0, 1]$ and quantile level $q \in [0, 1]$, we compute

$$\Phi_{\mathcal{S}}(u, q) = \mathbb{E}_{x \sim \mu_{\mathcal{S}}^u}\left[Q(\nu_x^{\mathcal{S}}, q)\right].$$

Overall, we have defined an embedding $\Phi_{\mathcal{S}} \in L^2([0, 1]^2)$. The stratified Wasserstein distance between two shapes is the L^p distance between their embeddings:

$$D_p(\mathcal{S}_1, \mathcal{S}_2) = \left(\int_0^1 \int_0^1 |\Phi_{\mathcal{S}_1}(u, q) - \Phi_{\mathcal{S}_2}(u, q)|^p \, du \, dq \right)^{1/p}.$$

Under mild regularity conditions on the ranking functional, the distance is injective up to isometry (Theorem 2 in S1 Text). Since it is induced by an L^2 norm, Gaussian and Laplace kernels constructed from D_2 are positive definite, and if the embedding is injective, these kernels are characteristic. The empirical estimator of the distance, based on binning and quantiles, is statistically consistent in the joint regime where the number of sampled points per shape grows while the discretization is refined. Under regularity of the ranking functional, which ensures stability of ranks, and a bin-growth condition, which ensures sufficient samples per stratum, the estimated distance converges to the population distance (see S2 Text). Its computational complexity is nearly quadratic in the number of points, offering substantial savings over GW and SLB.

Detailed numerical experimental setups in our framework

Benchmarks used for comparison

Elastic metrics between 2D planar curves. In the task on 2D cancer contour shapes, for comparison, we used the elastic metric with two choices of parameters [47–49]. Let $\gamma : [0, 1] \to \mathbb{R}^d$ be a smooth parameterized curve, and h_1, h_2 be two tangent vector fields along γ. The *elastic metric* with parameters a, $b > 0$ is defined by

$$\langle h_1, h_2 \rangle_\gamma^{a,b} = \int_0^1 a^2 \, \langle D_s h_1(t), \, D_s h_2(t) \rangle \, ds \; + \; b^2 \, \langle h_1(t), \, h_2(t) \rangle \, ds,$$

where s denotes the arc-length parameter of γ, and

$$D_s h(t) = \frac{1}{\|\dot{\gamma}(t)\|}\frac{d}{dt}h(t)$$

is the derivative of h with respect to arc-length. The square root velocity has $a = 1/2$ and $b = 1$, while the elastic metric implemented by Geomstats [10] has default $a = 1$ and $b = 1/2$. For the 2D cell line dataset, we perform benchmarking via elastic and square root velocity metrics on curves post alignment and interpolated to consist 2,000 points each, which can be accessed from https://github.com/wxli0/dyn/tree/main%4092c7a58/dyn/datasets/breast_cancer/aligned/projection_rescale_rotation_reparameterization.

Data preprocessing for 3D images. Binary volumes were converted to surface point clouds by applying a single iteration of binary erosion using `scipy.ndimage.binary_erosion` with default settings, followed by extraction of boundary voxels. The resulting surface points were uniformly randomly downsampled to 200 points per cell for the stratified Wasserstein framework, and to 100 points per cell for the Gromov–Wasserstein benchmark.

Pointnet distance between 3D images. As a modern deep learning-based benchmark, we used the feature encoding layers of a pointnet++ model [39,50] that was pretrained on ModelNet40 classification task, accessed at https://guochengqian.github.io/PointNeXt/modelzoo/. Image of each cell and nucleus is first down-sampled to 2,048 points, and embedded into a feature space of dimension 1,024. We use this embedding for dimension reduction and Euclidean distance therein for MMD as benchmarks of our 3D cell image case study.

Features of 3D cell shapes. We computed features of 3D cell shapes for benchmarking and visualization purpose, specifically, for volume, height, and major tilt. Each image is centered at the centroid and applied principal component analysis to determine its dominant axes of variation. The volume is defined as the vertical extent (difference between maximum and minimum) along the third coordinate that has the minimum variance. The volume is computed for the convex hull of the points. The major tilt is computed by the angle between the dominant principal axis and the third, vertical axis.

Kernel methods on shape populations

Positive-definite kernels constructed from the stratified Wasserstein distance enable population-level statistical analysis using kernel methods. We focus on maximum mean discrepancy (MMD) and Hilbert–Schmidt independence criterion (HSIC), which are widely used nonparametric statistics with strong theoretical guarantees.

Maximum mean discrepancy (MMD). MMD [18] is the squared distance between kernel mean embeddings of two distributions P_A and P_B in the reproducing kernel Hilbert space (RKHS) associated with kernel k:

$$\mathrm{MMD}^2(P_A, P_B) = \mathbb{E}_{\mathcal{S},\mathcal{S}'\sim P_A}[\,k(\mathcal{S},\mathcal{S}')\,] + \mathbb{E}_{\mathcal{T},\mathcal{T}'\sim P_B}[\,k(\mathcal{T},\mathcal{T}')\,] - 2\,\mathbb{E}_{\mathcal{S}\sim P_A,\mathcal{T}\sim P_B}[\,k(\mathcal{S},\mathcal{T})\,].$$

Given samples $\{\mathcal{S}_i\}_{i=1}^{n_A} \sim P_A$ and $\{\mathcal{T}_j\}_{j=1}^{n_B} \sim P_B$, the (biased) empirical estimator is

$$\widehat{\mathrm{MMD}}^2 = \frac{1}{n_A^2}\sum_{i,i'} k(\mathcal{S}_i,\mathcal{S}_{i'}) + \frac{1}{n_B^2}\sum_{j,j'} k(\mathcal{T}_j,\mathcal{T}_{j'}) - \frac{2}{n_A n_B}\sum_{i,j} k(\mathcal{S}_i,\mathcal{T}_j).$$

Significance is assessed by permutation testing of group labels. When k is characteristic (as is the case for Gaussian or Laplace kernels on the stratified Wasserstein metric), MMD equals zero if and only if $P_A = P_B$. The test is consistent against all fixed alternatives, meaning its power converges to one as sample sizes grow [18,51].

MMD witness function. The MMD witness function [18]

$$\hat{h}(\mathcal{Z}) = \frac{1}{n_A}\sum_i k(\mathcal{S}_i^{(A)}, \mathcal{Z}) - \frac{1}{n_B}\sum_j k(\mathcal{S}_j^{(B)}, \mathcal{Z})$$

identifies regions of shape space where the two distributions differ most strongly. Evaluating $\hat{h}(\mathcal{Z})$ at sample shapes highlights representative shapes that contribute to the observed differences.

Hilbert–Schmidt independence criterion (HSIC). HSIC [19] quantifies dependence between random variables using kernels on each domain, and can be interpreted as the maximum mean discrepancy (MMD) between the joint distribution $P_{\mathcal{S},Y}$ and the product of its marginals $P_{\mathcal{S}} \otimes P_Y$. For paired data $\{(\mathcal{S}_i, Y_i)\}_{i=1}^n$, with characteristic kernels k on shapes and ℓ on covariates, HSIC is defined as the squared Hilbert–Schmidt norm of the cross-covariance operator in the associated RKHSs:

$$\mathrm{HSIC}(P_{\mathcal{S}Y}; \mathcal{H}, \mathcal{G}) = \|C_{\mathcal{S}Y}\|_{\mathrm{HS}}^2.$$

With Gram matrices $K_{ij} = k(\mathcal{S}_i, \mathcal{S}_j)$, $L_{ij} = \ell(Y_i, Y_j)$, and centering matrix $H = I_n - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$, the (biased) estimator is

$$\widehat{\mathrm{HSIC}} = \frac{1}{n^2}\mathrm{tr}(KHLH).$$

HSIC is zero if and only if $\mathcal{S}$ and Y are independent, provided both kernels are characteristic. It is consistent against all alternatives and detects arbitrary nonlinear dependencies [19,21].

HSIC witness function. Similar to MMD, HSIC admits a witness function that localizes the contribution of individual shape–covariate pairs to the overall dependence. Given kernels k and ℓ centered in RKHS, the empirical HSIC witness is:

$$\hat{w}(s, y) = \frac{1}{n}\sum_{i=1}^{n} \tilde{k}(s, \mathcal{S}_i)\,\tilde{\ell}(y, Y_i),$$

where $\tilde{k}$ and $\tilde{\ell}$ are the centered kernels. Positive values indicate that the pair (s,y) supports the observed dependence, while negative values indicate opposition. It is precisely the MMD witness function applied to the joint distribution and the product of marginals. This function can be used to interpret the contributions of specific shapes or covariate values.

Statistical analyses

All population-level comparisons were performed using kernel methods derived from the stratified Wasserstein distance and other benchmarking distances.

Two-sample tests between shape populations were carried out using the maximum mean discrepancy (MMD) statistic [18]. For each test, we used a permutation test with 1,000 permutations to estimate the null distribution, and reported p-values based on the proportion of permuted statistics exceeding the observed value. The MMD witness function was used to localize representative shapes contributing to significant differences between groups.

To assess dependence between shape distributions and external covariates (e.g., developmental stage), we used the Hilbert–Schmidt independence criterion (HSIC) [19] with Laplace kernels on both shape and covariate domains, with bandwidth chosen via the median distance.

For clarity, all the tests in this section are performed using Laplace kernels between shapes with median bandwidth heuristic:

$$k_\gamma^{\mathrm{Lap}}(\mathcal{S}_1, \mathcal{S}_2) = \exp(-\gamma D(\mathcal{S}_1, \mathcal{S}_2)), \quad \gamma = \frac{1}{\mathrm{median}(D(\mathcal{S}_1, \mathcal{S}_2))}, \text{ median taken over all possible pairs of shapes,}$$

where $D(\mathcal{S}_1, \mathcal{S}_2)$ represents a distance in shape space, such as Gromov–Wasserstein and its alternatives.

For the 3D mitotic cell shape data, the mitotic stage variable consists of six ordered categories (`M0`, `M1M2`, `M3`, `M4M5`, `M6M7_early`, `M6M7_half`), which were encoded as integers $0, \ldots, 5$. The mitotic stage label was encoded as an angular variable $\theta = 2\pi\,\text{stage}/6$, and mapped to circular coordinates $[\cos(\theta), \sin(\theta)]$ on the unit circle. A Gaussian kernel applied to these circular embeddings,

$$k_\gamma^{\text{periodic}}(y_1, y_2) = \exp\left(-\gamma \left\|[\cos y_1, \sin y_1] - [\cos y_2, \sin y_2]\right\|^2\right) = \exp\left(2\gamma(\cos(y_1 - y_2) - 1)\right),$$

is equivalent (up to a constant factor) to a von Mises kernel (under a rescaling of γ)

$$k_\gamma^{\text{VM}}(y_1, y_2) = \exp\left(\gamma \cos(y_1 - y_2)\right).$$

The bandwidth parameter γ was selected using the median heuristic, that is, γ was set to the median of squared pairwise distances between stage embeddings on the unit circle. This encoding preserves the cyclic topology of the mitotic progression while allowing the stage variable to interact naturally with Euclidean distances between shapes.

Permutation testing over 1,000 trials of the covariate labels was used to assess significance. The HSIC witness function was used to visualize shape–covariate pairs with the strongest contribution to the dependence.

Computational details

All experiments were implemented in Python (v3.13.7). Numerical computations used NumPy (v2.3.4). Optimal transport computations were performed using the POT library (v0.9.6). Kernel-based statistical tests relied on scikit-learn (v1.7.2), dimensionality reduction used UMAP (v0.5.9), and elastic shape analysis used in benchmarks employed Geomstats (v2.8.0).

Timing experiments were conducted on a MacBook Pro (14-inch, 2024) equipped with an Apple M4 Pro chip and 48 GB unified memory, running macOS Sequoia (v15.7.3). All methods were evaluated under identical hardware and software conditions; reported runtimes are therefore intended for relative comparison rather than absolute benchmarking.

Supporting information

S1 Text. Detailed description and properties of stratified distance, including formal definitions of local distance distributions and stratified Wasserstein distances, and theorem for injectivity under assumptions.
(PDF)

S2 Text. Detailed properties of the stratified kernel and the population metric it induced, including consistency under growing sample size and computational complexity analyses.
(PDF)

S3 Text. Synthetic examples on 2D shapes, illustrating global and local distance metrics, failure cases of injectivity, and numerical stability (on random shapes) and instability results (on shapes with strong symmetry).
(PDF)

S1 Fig. Supplementary results for all 3 populations in 2D cancer cell contour datasets. (a) Random samples from MCF10A (red), MCF7 (yellow), and MDA (blue). (b) Mean shapes for each category. (c) test statistic for whether cancer cell populations are closer than cancer versus normal. (d) Type I and II errors for different kernels for the same hypothesis testing.
(TIF)

S2 Fig. Type II error with 25 cancer cell shapes in 2D when varying hyperparameters, randomized over 1000 trials. Lower (Darker) is better. Left: varied discretization. Right: varied ranking functional. The results are robust to the choices of hyperparameters when the discretization is refined enough, and lower order of moments is better at respecting geometry and achieves lower error.
(TIF)

S3 Fig. Comparison on population-level distances for mitotic dataset. Our framework shows a block-diagonal structure that sees similarity between consecutive stages, while both pointnet and stratified energy distance confuses M0 with intermediate stages such as M3 to M6M7 early.
(TIF)

S4 Fig. Embedding of 3D cell shapes using stratified quantiles downsampled to 100 points for each shape, matching with the benchmark of Gromov–Wasserstein.
(TIF)

Author contributions

Conceptualization: Wenjun Zhao, Danica J Sutherland, Khanh Dao Duc.

Formal analysis: Wenjun Zhao, Danica J Sutherland, Khanh Dao Duc.

Investigation: Wenjun Zhao.

Methodology: Wenjun Zhao, Danica J Sutherland.

Project administration: Wenjun Zhao, Khanh Dao Duc.

Resources: Khanh Dao Duc.

Software: Wenjun Zhao, Danica J Sutherland.

Supervision: Danica J Sutherland, Khanh Dao Duc.

Validation: Wenjun Zhao, Khanh Dao Duc.

Visualization: Wenjun Zhao.

Writing – original draft: Wenjun Zhao.

Writing – review & editing: Wenjun Zhao, Danica J Sutherland, Khanh Dao Duc.

References

1. Viana MP, Chen J, Knijnenburg TA, Vasan R, Yan C, Arakaki JE, et al. Integrated intracellular organization and its variations in human iPS cells. Nature. 2023;613(7943):345–54. http://dx.doi.org/10.1038/s41586-022-05563-7
2. Dao Duc K, Batra SS, Bhattacharya N, Cate JHD, Song YS. Differences in the path to exit the ribosome across the three domains of life. Nucleic Acids Research. 2019;47(8):4198–210. http://dx.doi.org/10.1093/nar/gkz106
3. Munch E. A User's Guide to Topological Data Analysis. Journal of Learning Analytics. 2017;4(2). http://dx.doi.org/10.18608/jla.2017.42.6
4. Kazhdan M, Funkhouser T, Rusinkiewicz S. Rotation invariant spherical harmonic representation of 3D shape descriptors. In: Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. SGP '03. Goslar, DEU: Eurographics Association; 2003. p. 156–64.
5. Ekvall M, Bergenstråhle L, Andersson A, Czarnewski P, Olegård J, Käll L, et al. Spatial landmark detection and tissue registration with deep learning. Nature Methods. 2024;21(4):673–9. http://dx.doi.org/10.1038/s41592-024-02199-5
6. Srivastava A, Klassen E, Joshi SH, Jermyn IH. Shape Analysis of Elastic Curves in Euclidean Spaces. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2011;33(7):1415–28. http://dx.doi.org/10.1109/TPAMI.2010.184
7. Dryden IL, Mardia KV. Statistical Shape Analysis, with Applications in R. Wiley; 2016. http://dx.doi.org/10.1002/9781119072492
8. Kendall DG. Shape Manifolds, Procrustean Metrics, and Complex Projective Spaces. Bulletin of the London Mathematical Society. 1984;16(2):81–121. http://dx.doi.org/10.1112/blms/16.2.81

9. Styner M, Oguz I, Xu S, Brechbühler C, Pantazis D, Levitt JJ, et al. Framework for the statistical shape analysis of brain structures using SPHARM-PDM. Insight J. 2006;(1071):242–50.
10. Miolane N, Guigui N, Brigant AL, Mathe J, Hou B, Thanwerdas Y, et al. Geomstats: A Python Package for Riemannian Geometry in Machine Learning. Journal of Machine Learning Research. 2020;21(223):1–9. http://jmlr.org/papers/v21/19-027.html
11. Astore MA, Woollard G, Silva-Sánchez D, Zhou W, Kopylov M, Dao Duc K, et al. The Inaugural Flatiron Institute Cryo-EM Conformational Heterogeneity Challenge. bioRxiv. 2025. http://dx.doi.org/10.1101/2025.07.18.665582
12. Mémoli F. Gromov–Wasserstein Distances and the Metric Approach to Object Matching. Foundations of Computational Mathematics. 2011;11(4):417–87. http://dx.doi.org/10.1007/s10208-011-9093-5
13. Boutin M, Kemper G. On reconstructing n-point configurations from the distribution of distances or areas. Advances in Applied Mathematics. 2004;32(4):709–35. http://dx.doi.org/10.1016/S0196-8858(03)00101-5
14. Mémoli F, Needham T. Distance distributions and inverse problems for metric measure spaces. Studies in Applied Mathematics. 2022;149(4):943–1001. http://dx.doi.org/10.1111/sapm.12526
15. Fournier N, Guillin A. On the rate of convergence in Wasserstein distance of the empirical measure. Probability Theory and Related Fields. 2014;162(3–4):707–38. http://dx.doi.org/10.1007/s00440-014-0583-7
16. Kravtsova N. Note on computational complexity of the Gromov-Wasserstein distance. arXiv; 2024. https://arxiv.org/abs/2408.06525
17. Scetbon M, Peyré G, Cuturi M. Linear-Time Gromov Wasserstein Distances using Low Rank Couplings and Costs. In: Chaudhuri K, Jegelka S, Song L, Szepesvari C, Niu G, Sabato S, editors. Proceedings of the 39th International Conference on Machine Learning. vol. 162 of Proceedings of Machine Learning Research. PMLR; 2022. p. 19347–65. https://proceedings.mlr.press/v162/scetbon22b.html
18. Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A. A kernel two-sample test. J Mach Learn Res. 2012;13(null):723–73.
19. Gretton A, Fukumizu K, Teo C, Song L, Schölkopf B, Smola A. A Kernel Statistical Test of Independence. In: Platt J, Koller D, Singer Y, Roweis S, editors. Advances in Neural Information Processing Systems. vol. 20. Curran Associates, Inc.; 2007. https://proceedings.neurips.cc/paper_files/paper/2007/file/d5cfead94f5350c12c322b5b664544c1-Paper.pdf
20. Zhang K, Peters J, Janzing D, Schölkopf B. Kernel-based conditional independence test and application in causal discovery. In: Proceedings of the Twenty-Seventh Conference on Uncertainty in Artificial Intelligence. UAI'11. Arlington, Virginia, USA: AUAI Press; 2011. p. 804–13.
21. Song L, Smola A, Gretton A, Bedo J, Borgwardt K. Feature selection via dependence maximization. J Mach Learn Res. 2012;13(1):1393–434.
22. Pogodin R, Deka N, Li Y, Sutherland DJ, Veitch V, Gretton A. Efficient Conditionally Invariant Representation Learning. In: The Eleventh International Conference on Learning Representations; 2023. https://openreview.net/forum?id=dJruFeSRym1
23. Perez RC, da Veiga S, Garnier J, Staber B. Gaussian process regression with Sliced Wasserstein Weisfeiler-Lehman graph kernels. arXiv; 2024. Available from: https://arxiv.org/abs/2402.03838
24. Kolouri S, Zou Y, Rohde GK. Sliced Wasserstein Kernels for Probability Distributions. arXiv; 2015. https://arxiv.org/abs/1511.03198
25. Sriperumbudur BK, Gretton A, Fukumizu K, Schölkopf B, Lanckriet GRG. Hilbert space embeddings and metrics on probability measures. arXiv; 2009. https://arxiv.org/abs/0907.5309
26. Rioux G, Goldfeld Z, Kato K. Entropic Gromov-Wasserstein distances: stability and algorithms. J Mach Learn Res. 2024;25(1).
27. Cuturi M. Sinkhorn Distances: Lightspeed Computation of Optimal Transport. In: Burges CJ, Bottou L, Welling M, Ghahramani Z, Weinberger KQ, editors. Advances in Neural Information Processing Systems. vol. 26. Curran Associates, Inc.; 2013. https://proceedings.neurips.cc/paper_files/paper/2013/file/af21d0c97db2e27e13572cbf59eb343d-Paper.pdf
28. Altschuler J, Weed J, Rigollet P. Near-linear time approximation algorithms for optimal transport via Sinkhorn iteration. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 1961–71.
29. Dvurechensky P, Gasnikov A, Kroshnin A. Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn's Algorithm. In: Dy J, Krause A, editors. Proceedings of the 35th International Conference on Machine Learning. vol. 80 of Proceedings of Machine Learning Research. PMLR; 2018. p. 1367–76. https://proceedings.mlr.press/v80/dvurechensky18a.html
30. Gabriel P, Marco C. Computational Optimal Transport with Applications to Data Sciences. Foundations and Trends in Machine Learning. 2019;11(5–6):355–607. http://dx.doi.org/10.1561/2200000073
31. Catalano M, Lavenant H. Hierarchical integral probability metrics: a distance on random probability measures with low sample complexity. In: Proceedings of the 41st International Conference on Machine Learning. ICML'24. JMLR.org; 2024.
32. Alizadeh E, Castle J, Quirk A, Taylor CDL, Xu W, Prasad A. Cellular morphological features are predictive markers of cancer cell state. Computers in Biology and Medicine. 2020;126:104044. https://www.sciencedirect.com/science/article/pii/S0010482520303759
33. Alizadeh E, Xu W, Castle J, Foss J, Prasad A. TISMorph: A tool to quantify texture, irregularity and spreading of single cells. PLoS One. 2019;14(6):e0217346.
34. Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, et al. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell. 2006;10(6):515–27. http://dx.doi.org/10.1016/j.ccr.2006.10.008
35. Li W. breast_cancer_analysis.ipynb; 2024. Cited 2025-10-12; commit 92c7a58. GitHub. Available from: https://github.com/wxli0/dyn/blob/92c7a58/dyn/notebooks/breast_cancer_analysis.ipynb

36. Franchi M, Piperigkou Z, Karamanos KA, Franchi L, Masola V. Extracellular Matrix-Mediated Breast Cancer Cells Morphological Alterations, Invasiveness, and Microvesicles/Exosomes Release. Cells. 2020;9(9):2031. http://dx.doi.org/10.3390/cells9092031
37. Zhang R, Ogden RT, Picard M, Srivastava A. Nonparametric k-Sample Test on Shape Spaces with Applications to Mitochondrial Shape Analysis. Journal of the Royal Statistical Society Series C: Applied Statistics. 2022;71(1):51–69. http://dx.doi.org/10.1111/rssc.12521
38. Burgess J, Nirschl JJ, Zanellati MC, Lozano A, Cohen S, Yeung-Levy S. Orientation-invariant autoencoders learn robust representations for shape profiling of cells and organelles. Nature Communications. 2024;15(1). http://dx.doi.org/10.1038/s41467-024-45362-4
39. Qi CR, Su H, Mo K, Guibas LJ. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. arXiv preprint arXiv:161200593. 2016.
40. Walczak CE, Cai S, Khodjakov A. Mechanisms of chromosome behaviour during mitosis. Nature Reviews Molecular Cell Biology. 2010;11(2):91–102. http://dx.doi.org/10.1038/nrm2832
41. Govek KW, Nicodemus P, Lin Y, Crawford J, Saturnino AB, Cui H, et al. CAJAL enables analysis and integration of single-cell morphological data using metric geometry. Nature Communications. 2023;14(1). http://dx.doi.org/10.1038/s41467-023-39424-2
42. Liu YS, Li Q, Zheng GQ, Ramani K, Benjamin W. Using diffusion distances for flexible molecular shape comparison. BMC Bioinformatics. 2010;11(1). http://dx.doi.org/10.1186/1471-2105-11-480
43. Boyer DM, Lipman Y, St Clair E, Puente J, Patel BA, Funkhouser T, et al. Algorithms to automatically quantify the geometric similarity of anatomical surfaces. Proceedings of the National Academy of Sciences. 2011;108(45):18221–6. http://dx.doi.org/10.1073/pnas.1112822108
44. Flamary R, Courty N, Gramfort A, Alaya MZ, Boisbunon A, Chambon S, et al. POT: Python Optimal Transport. Journal of Machine Learning Research. 2021;22(78):1–8. http://jmlr.org/papers/v22/20-451.html
45. Flamary R, Vincent-Cuaz C, Courty N, Gramfort A, Kachaiev O, Quang Tran H, et al. POT Python Optimal Transport (version 0.9.5); 2024. Available from: https://github.com/PythonOT/POT
46. Osada R, Funkhouser T, Chazelle B, Dobkin D. Shape distributions. ACM Transactions on Graphics. 2002;21(4):807–32. http://dx.doi.org/10.1145/571647.571648
47. Li W, Prasad A, Miolane N, Dao Duc K. Unveiling cellular morphology: statistical analysis using a Riemannian elastic metric in cancer cell image datasets. Information Geometry. 2024;7(S2):845–59. http://dx.doi.org/10.1007/s41884-024-00145-0
48. Bauer M, Bruveris M, Marsland S, Michor PW. Constructing reparameterization invariant metrics on spaces of plane curves. Differential Geometry and its Applications. 2014;34:139–65. https://www.sciencedirect.com/science/article/pii/S092622451400062X
49. Needham T, Kurtek S. Simplifying Transforms for General Elastic Metrics on the Space of Plane Curves. SIAM Journal on Imaging Sciences. 2020;13(1):445–73. Available from: http://dx.doi.org/10.1137/19m1265132
50. Qi CR, Yi L, Su H, Guibas LJ. PointNet++: deep hierarchical feature learning on point sets in a metric space. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 5105–14.
51. Sutherland DJ, Deka N. Unbiased estimators for the variance of MMD estimators. arXiv; 2019. Available from: https://arxiv.org/abs/1906.02104