ENF-SSM: AMORTIZING ANATOMICAL STATISTICAL SHAPE MODELING WITH ELASTIC NEURAL FLOW

ABSTRACT

Riemannian statistical shape analysis repeatedly computes deformation paths to estimate templates, evolve initial fields, and transfer deformations. We introduce ENF-SSM, which amortizes these computations through a shared endpoint-conditioned neural path. Building on the Spectral Meets Spatial backbone, we use an elastic regularizer derived from established bending, scaling, and shearing decompositions, together with spatial smoothness. The learned path supports a finite-step logarithm surrogate, frozen-network endpoint inversion, recursive template estimation, and deformation-based path regression. We derive the conditional bounds that separate time-discretization error, reference-path error, inverse conditioning, and endpoint attachment. These results identify the assumptions needed for operator reuse. Existing experiments on pancreas, hippocampus, and heart meshes report competitive reconstruction and lower online runtimes than the evaluated iterative implementations. They provide evidence for the computational utility of the shared path and its potential values in anatomical shape analysis. The code is available at `https://anonymous.4open.science/r/ENF-SSM-4C98/`.

1 INTRODUCTION

Statistical shape models (SSMs) constrain shape variation to patterns observed in a training cohort and thereby give a compact, statistically usable account of anatomical geometry. Since the seminal works (Kendall, 1984; Cootes et al., 1995), they have been central to computational anatomy and biological morphology. The field has moved from linear models on landmark or mesh coordinates (Heimann & Meinzer, 2009) to analysis on non-Euclidean shape spaces (Lüthi et al., 2017) and, more recently, to deep networks that learn correspondences, atlases, or implicit surfaces (Sun et al., 2022; Iyer & Elhabian, 2023; Adams & Elhabian, 2024; El Amrani et al., 2024b; Jiao et al., 2024; 2026). Within this landscape, Riemannian statistical shape modeling (Rm-SSM) (Pennec et al., 2019; Guigui et al., 2023)—from classical computational anatomy and elastic shape analysis (Beg et al., 2005; Srivastava & Klassen, 2016) to learned diffeomorphic autoencoders (Dummer et al., 2024)—treats a cohort as points on a shape manifold and performs intrinsic statistics, providing a metric-based framework for cohort analysis.

Rm-SSM operators rarely admit closed forms on high-dimensional anatomical shape spaces, making repeated numerical path computation a shared bottleneck (Pennec et al., 2019; Myers et al., 2023). Neural interpolation offers an opportunity to amortize this computation. SmS learns correspondence and interpolation jointly (Cao et al., 2024), and FUSS adapts this approach to anatomical SSM construction (El Amrani et al., 2024a). *Our question is whether one trained path can also support several nonlinear statistical operations, with its approximation errors made explicit.*

ENF-SSM uses an elastic-regularized, endpoint-conditioned interpolant as a reusable computational primitive. Correspondence is estimated by the spectral module; subsequent operations use a fixed common mesh representation. We reuse the spatial module through direct evaluations and endpoint inversions. The contribution is this shared operator construction and its error analysis, rather than a new correspondence architecture or a new continuous elastic metric. Our contributions are:

Based on the motivation above, we propose ENF-SSM, a framework that uses an elastic neural interpolant as a reusable path primitive for the Rm-SSM pipeline. We therefore train a mesh interpolant (Cao et al., 2024) with general elastic shape metric (GESM)-based energy that separately scores bending, area change, and shearing. The trained network is an *energy-regularized path surrogate*: given two corresponding meshes it returns a discrete trajectory whose successive displacements are scored by that energy. We then recast logarithm, elastic deformation energy-based template estimation, initial value evolving, deformation transfer, and regression steps as network evaluations or small inversions that reuse the same primitive. The same regularized interpolant can stand in for

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

those operators at much lower cost than online geodesic optimization, and the resulting pipeline is empirically useful on anatomical meshes. Our contributions are:

Shared computation. We formulate template estimation, endpoint recovery, deformation transfer, and cross-sectional path fitting around one frozen interpolant, exposing which operations require only inference and which require numerical inversion.

Explicit mathematical conditions. We derive the finite-step logarithm error and local endpoint-inversion bounds. A conditional action-based estimate links a fixed-endpoint numerical reference to the learned initial field. These statements distinguish algebraic guarantees from geometric properties that require measurement.

Anatomical evaluation. We show on anatomical shape datasets that ENF-SSM is substantially more efficient than iterative elastic path solvers, remains reliable under matched correspondence, and achieves promising results on shape-statistics tasks.

2 Related Work

Statistical shape models: Statistical shape models (SSMs) evolved from classical linear statistical models (Cootes et al., 1995) along two perspectives: advanced shape representations and statistical techniques for shape space. Shape representations include explicit forms such as meshes and point clouds (Kurtek et al., 2011; Hartman et al., 2023; Adams & Elhabian, 2024), implicit forms such as signed distance functions (Sun et al., 2022; Jiao et al., 2024; Dummer et al., 2024; Lüdke et al., 2022; Park et al., 2019) and neural latent codes (Huang et al., 2021; Lee et al., 2022; Yang et al., 2023). Statistics on shape space use Riemannian methods (Pennec, 2006; Pennec et al., 2019), including the LDDMM family (Beg et al., 2005; Bône et al., 2018b;a; 2020) modeling deformation in ambient space, and statistics on intrinsic shape spaces (Srivastava & Klassen, 2016; Zhang et al., 2015; 2023; Rumpf & Wirth, 2015; Hartman et al., 2023). These methods are explainable and theoretically rigorous for clinical and scientific usage, but are usually optimization-based and computationally costly. Other methods include Gaussian process-based shape modeling (Lüthi et al., 2017; Fouefack et al., 2023), and linear statistics on the latent coding of the shape (Jiao et al., 2026; Dummer et al., 2024; Yang et al., 2022). NAISR (Jiao et al., 2024) and PRISM (Jiao et al., 2026) fit implicit atlases with optional demographic covariates. ENF-SSM assumes pre-extracted meshes and operates in a common triangulation after correspondence. Unlike approaches that learn an atlas and its latent distribution, our focus is reusing an endpoint-conditioned path for several statistical operations; the term surrogate distinguishes these constructions from exact intrinsic operators.

Neural shape interpolants: A neural shape interpolant learns to continuously deform a source shape to the target, implemented in explicit mesh forms (Cao et al., 2024; Eisenberger et al., 2021; Hartshorne et al., 2025), implicit representations (Sang et al., 2025a;b; Wang et al., 2025; Sun et al., 2022), or a neural latent space (Cosmo et al., 2020; Huang et al., 2021). To this end, various deformation measures have been used, including as-rigid-as-possible (ARAP) (Sorkine & Alexa, 2007), area-preserving energy (Bastian et al., 2024), conformal energy (Yoshiyasu et al., 2014), flip-preventing energy (Fu et al., 2015), and volume-preserving energy (Eisenberger & Cremers, 2020). Different regularizers encode different deformation preferences. FUSS already uses learned anatomical deformations to construct a statistical shape model, and learned Riemannian models such as RDA-INR connect neural representations to geometric statistics (El Amrani et al., 2024a; Dummer et al., 2024). Our specific focus is sharing one pairwise path across endpoint inversion, cohort estimation, and deformation transfer, while accounting for the approximation errors introduced by this reuse.

3 Problem Formulations

Domain and time convention. A shape is represented by vertex coordinates $\mathcal{X} \in \mathbb{R}^{V\times 3}$ and a fixed face set after correspondence. We work locally on the open set of nondegenerate meshes; no global injectivity of a deformation is implied by fixed connectivity. A vertex transfer $\mathbf{P}_{\mathcal{X}\mathcal{Y}} \in \mathbb{R}^{V_X \times V_Y}$ maps target coordinates to the source triangulation. All inversions and derivatives below hold this transfer fixed. Set $T = K + 1$ and $h = 1/T$: there are T steps, K interior frames, and $T + 1$ frames in total.

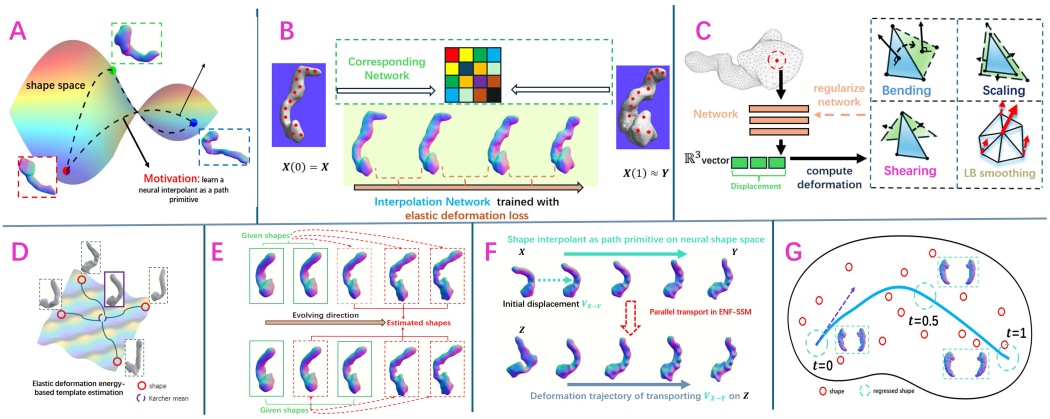


Figure 1: Overview of the ENF-SSM framework. A motivates replacing online geodesic solves with a learned energy-regularized interpolant. B summarizes the interpolant (Cao et al., 2024). C details the discrete elastic energy. D–G instantiate the Rm-SSM operators with that path primitive: Elastic deformation energy-based template estimation, initial-value evolution, Schild transport, and (cross-sectional) regression.

We use the Spectral Meets Spatial (SmS) network (Cao et al., 2024; El Amrani et al., 2024a) as the interpolant backbone, which jointly learns spectral correspondence and spatial interpolation. We retain this architecture and specify its losses and conventions in Appendix A.2, and emphasize the spatial module $\mathrm{Intp}(\mathcal{X}, \mathcal{Y}; K) = \{\mathcal{X}(k)\}_{k=0}^{K+1}$ for interpolation. In SmS, the total displacement between source shape $\mathcal{X}(0)$ and the k-th intermediate shape $\mathcal{X}(k)$ is:

$$\mathcal{X}(k) - \mathcal{X}(0) = (k/(K+1)) \cdot \Omega(\mathrm{concat}[\mathcal{X}; \mathbf{P}_{\mathcal{X}\mathcal{Y}}\mathcal{Y} - \mathcal{X}; (k/(K+1))\mathbf{1}]),\ k = 1, \ldots, K+1, \quad (1)$$

where Ω is the interpolant network, concat denotes feature concatenation, $\mathbf{P}_{\mathcal{X}\mathcal{Y}}$ corresponds $\mathcal{X}$ to $\mathcal{Y}$ with $\mathbf{P}_{\mathcal{X}\mathcal{Y}}\mathcal{Y} \approx \mathcal{X}(K+1)$, and $(k/(K+1))\mathbf{1}$ is the time stamp broadcast to all vertices.

Path surrogate. Write the continuous-time evaluation of the spatial network of Eqn.(1) as

$$\gamma_\theta(\mathcal{X}, \mathcal{Y}; t) = \mathcal{X} + t\,\Omega_\theta(\mathrm{concat}[\mathcal{X}; \mathbf{P}_{\mathcal{X}\mathcal{Y}}\mathcal{Y} - \mathcal{X}; t\mathbf{1}]), \quad t \in [0, 1]. \quad (2)$$

The discrete path is $\mathrm{Intp}(\mathcal{X}, \mathcal{Y}; K) = \{\gamma_\theta(\mathcal{X}, \mathcal{Y}; kh)\}_{k=0}^{T}$. Its increments are $\mathcal{D}(k) = \gamma_\theta((k+1)h) - \gamma_\theta(kh)$, with endpoints suppressed in this notation. The source is exact; the target is attached approximately, with residual $r_{\mathrm{end}} = \|\gamma_\theta(1) - \mathbf{P}_{\mathcal{X}\mathcal{Y}}\mathcal{Y}\|_F$.

The training energy controls the path's local deformation, while explicit endpoint and bidirectional consistency terms discourage endpoint mismatch (Appendix A.2), and we use this learned path to construct computational surrogates.

Operator tasks. The finite-step logarithm is $\widehat{\mathrm{Log}}^{h}_{\mathcal{X}}(\mathcal{Y}) = \mathcal{D}(0)/h$. Endpoint inversion fits a target parameter whose initial field matches a prescribed field. For a cohort, define the initial-field objective

$$\mathcal{J}(\mathcal{X}_0) = \sum_{i=1}^{N} E_{\mathrm{els}}\left(h\widehat{\mathrm{Log}}^{h}_{\mathcal{X}_0}(\mathcal{X}_i) \mid \mathcal{X}_0\right). \quad (3)$$

The recursive estimator below is a heuristic driven by these initial fields; it is not assumed to minimize $\mathcal{J}$. Deformation transfer transports a source field along a chosen path to a new base. Path regression fits endpoint parameters to observations $\{(\mathcal{X}_i, t_i)\}$ using rescaled time $s_i \in [0, 1]$. For cross-sectional observations, the fitted path describes an age-associated population trend. Predicting later visits of an individual is a distinct task requiring a longitudinal evaluation.

4 Methodology

Based on the problem formulation, this section details: **(1)** GESM-based elastic deformation energy used to train the neural interpolant; and **(2)** Rm-SSM operators within the ENF-SSM framework.

4.1 TRAINING ANATOMICAL SHAPE INTERPOLANT WITH ELASTIC DEFORMATION ENERGY

The motivation for an elastic deformation energy when training an anatomical interpolant is that as-rigid-as-possible (ARAP) energy in SmS and Chamfer distance (CD) between adjacent frames in the FUSS model are *ill-suited* for highly deformable anatomy, as follows:

ARAP penalizes deviation from local rigidity and matches near-isometric pose change on humans and animals, but anatomical trajectories are dominated by non-isometric thinning and twisting. Under ARAP those patterns are treated as artifacts, so intermediate frames remain locally rigid at the cost of implausible change along the path. In spatial module of SmS where correspondence between frames are determined, minimizing the Chamfer term between adjacent frames encourages to pull the consecutive meshes together and favors vanish steps.

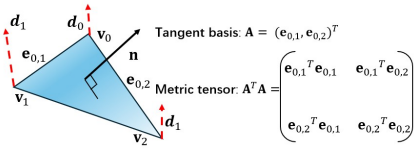


Figure 2: Discrete geometry on mesh.

Therefore, the deformation pattern of the energy used to train anatomical interpolants for our purpose should not be too constricted, $i.e.$ ARAP and Chamfer, and it can regularize the deformation path while retaining variation of the interpolated frames. This requirement motivates the application of General Elastic Shape Metric (GESM) to our cases.

GESM-based deformation energy: For a nondegenerate triangular face $\mathbf{f}$ shown in Fig. 2, its local deformation variation is characterized by the difference of displacement, *not* displacement itself. Let tangent basis $\mathbf{A} = (\mathbf{v}_1 - \mathbf{v}_0, \mathbf{v}_2 - \mathbf{v}_0) := (\mathbf{e}_{01}, \mathbf{e}_{02}) \in \mathbb{R}^{3\times 2}$ and displacement difference $D = (\mathbf{d}_1 - \mathbf{d}_0, \mathbf{d}_2 - \mathbf{d}_0) \in \mathbb{R}^{3\times 2}$, by defining a mapping $f_{\mathbf{A}}(\mathbf{X}) = \mathrm{Tr}(\mathbf{X}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{X}^T)$, its GESM-based deformation energy (Su et al., 2020) reads:

$$\begin{aligned} &\mathrm{E}_{\mathrm{bend}}(\mathbf{D}|\mathbf{f}) = f_{\mathbf{A}}(\mathbf{D}_\perp),\ \mathrm{E}_{\mathrm{shear}}(\mathbf{D}|\mathbf{f}) = f_{\mathbf{A}}(\mathbf{D}_m),\ \mathrm{E}_{\mathrm{scale}}(\mathbf{D}|\mathbf{f}) = f_{\mathbf{A}}(\mathbf{D}_+),\ \text{where} \\ &\mathbf{D}_\perp = \mathbf{D} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{D},\ \mathbf{D}_+ = \frac{1}{2}\mathrm{Tr}((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{D})\mathbf{A}, \\ &\mathbf{D}_m = \frac{1}{2}\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{D} + \mathbf{D}^T\mathbf{A}) - \frac{1}{2}\mathrm{Tr}((\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{D})\mathbf{A}, \end{aligned} \tag{4}$$

Here E_{bend} measures normal variation, E_{scale} the isotropic tangential component, and E_{shear} the trace-free symmetric tangential component. For a spatial smoothness term, let Δ_L denote a symmetric positive-semidefinite stiffness matrix, so that the Dirichlet quadratic form is nonnegative (Magnet et al., 2022):

$$E_{\mathrm{smth}}(\mathcal{D}|\mathcal{X}) = \mathrm{Tr}(\mathcal{D}^T\Delta_L\mathcal{D}). \tag{5}$$

To sum up, given displacement field $\mathcal{D} = \{\mathbf{d}_i\}_{i=1}^V$ with face-wise difference $\{\mathbf{D}_j\}_{j=1}^F$ and source shape with vertex set $\{\mathbf{v}\}_{i=1}^V$ and face set $\{\mathbf{f}_j\}_{j=1}^F$, the elastic deformation energy $E_{\mathrm{els}}(\mathcal{D}|\mathcal{X})$ is

$$E_{\mathrm{els}}(\mathcal{D}|\mathcal{X}) = \lambda_1 \sum_{j=1}^{F} \big(aE_{\mathrm{bend}}(\mathbf{D}_j|\mathbf{f}_j) + bE_{\mathrm{scale}}(\mathbf{D}_j|\mathbf{f}_j) + cE_{\mathrm{shear}}(\mathbf{D}_j|\mathbf{f}_j)\big)|\mathbf{f}_j| + \lambda_2 E_{\mathrm{smth}}(\mathcal{D}|\mathcal{X}), \tag{6}$$

$|\mathbf{f}|$ is the area of face $\mathbf{f}$. The global weights (a, b, c) are Softmax outputs and are fixed after training; $\lambda_1 = \lambda_2 = 10^2$. Deformation loss for training SmS is $\sum_{k=0}^{T-1} E_{\mathrm{els}}(\mathcal{D}(k) \mid \mathcal{X}(k))$. For comparisons across time resolutions, the corresponding discrete action is

$$\mathcal{A}_h = \sum_{k=0}^{T-1} h\, E_{\mathrm{els}}(\mathcal{D}(k)/h \mid \mathcal{X}(k)) = h^{-1} \sum_{k=0}^{T-1} E_{\mathrm{els}}(\mathcal{D}(k) \mid \mathcal{X}(k)). \tag{7}$$

At the fixed training resolution this is a constant rescaling of the regularizer, with a corresponding rescaling of its loss weight.

4.2 RM-SSM OPERATORS WITH LEARNED INTERPOLANT

Finite-step logarithm. The initial field used by all operators is

$$\widehat{\mathrm{Log}}^h_{\mathcal{X}}(\mathcal{Y}) := \frac{\gamma_\theta(\mathcal{X}, \mathcal{Y}; h) - \mathcal{X}}{h} = \Omega_\theta(\mathrm{concat}[\mathcal{X}; \mathbf{P}_{\mathcal{X}\mathcal{Y}}\mathcal{Y} - \mathcal{X}; h\mathbf{1}]). \tag{8}$$

It is a finite difference of a learned path. The following statement separates that discretization from approximation to a specified geometry.

Proposition 1 (Initial-field error)**.** *Suppose γ_θ is twice differentiable on $[0,h]$ and $\sup_t \|\ddot{\gamma}_\theta(t)\|_F \le M$. In a common vertex chart, for any differentiable reference path γ_* from $\mathcal{X}$ to $\mathcal{Y}$, if $\|\dot{\gamma}_\theta(0) - \dot{\gamma}_*(0)\|_F \le \delta$, then*

$$\|\widehat{\mathrm{Log}}^h_{\mathcal{X}}(\mathcal{Y}) - \dot{\gamma}_*(0)\|_F \le Mh/2 + \delta. \tag{9}$$

For a unit-time minimizing geodesic in a normal neighbourhood, $\dot{\gamma}_*(0) = \mathrm{Log}^g_{\mathcal{X}}(\mathcal{Y})$.

The proof is the integral Taylor remainder (Appendix A.3). The assumptions are local: low training energy alone does not bound δ. A conditional fixed-endpoint action certificate is given in Proposition 3.

Recursive template estimation. Algorithm 1 replaces costly reference logarithms by Eq. (8) in a recursive averaging rule motivated by Ho et al. (2013); Hartman et al. (2023). We use the resulting shape as a cohort template. For learned logarithms, this update is not generally the gradient of Eq. (3); its exact differential is derived in Appendix A.5. We therefore assess geometric representativeness separately from minimization of that objective.

Endpoint inversion and initial fields. After transferring targets to the source triangulation, let $F_{\mathcal{X},h}(\mathcal{Y}) = \widehat{\mathrm{Log}}^h_{\mathcal{X}}(\mathcal{Y})$, with $\mathbf{P} = \mathbf{I}$ in this fixed coordinate chart. Define an endpoint parameter, distinct from a path or its realised terminal frame, by

$$\widehat{\mathrm{End}}_{\mathcal{X}}(v) := \arg\min_{\mathcal{Y}} \|F_{\mathcal{X},h}(\mathcal{Y}) - v\|_F^2, \quad \mathrm{IV}(\mathcal{X}, \widehat{\mathcal{D}}; K) := \mathrm{Intp}(\mathcal{X}, \widehat{\mathrm{End}}_{\mathcal{X}}(\widehat{\mathcal{D}}/h); K). \tag{10}$$

The frozen-network optimization is solved with L-BFGS (Liu & Nocedal, 1989); an approximate minimizer is understood. For an observation at any $0 < c \le 1$, define

$$\widehat{\mathrm{End}}^c_{\mathcal{X}}(d) := \arg\min_{\mathcal{Y}} \|d - [\gamma_\theta(\mathcal{X}, \mathcal{Y}; c) - \mathcal{X}]\|_F^2, \quad \mathrm{CV}(\mathcal{X}, d, c; K) := \mathrm{Intp}(\mathcal{X}, \widehat{\mathrm{End}}^c_{\mathcal{X}}(d); K). \tag{11}$$

The case $c = 1$ tests attachment inversion.

Proposition 2 (Local inverse stability)**.** *On a neighbourhood U, suppose $\|F_{\mathcal{X},h}(Y) - F_{\mathcal{X},h}(Z)\|_F \ge m\|Y - Z\|_F$ for $m > 0$. Let $Y_* \in U$ satisfy $\|F_{\mathcal{X},h}(Y_*) - v_*\|_F \le \delta_F$. If the prescribed field is $v_* + \eta$ and an inversion $\widehat{Y} \in U$ has residual at most ρ, then*

$$\|\widehat{Y} - Y_*\|_F \le (\rho + \|\eta\|_F + \delta_F)/m. \tag{12}$$

For the realised terminal frame, add its attachment error $\|\gamma_\theta(\mathcal{X}, \widehat{Y}; 1) - \widehat{Y}\|_F$ to this bound.

Appendix A.3 proves the bound and gives a sufficient Jacobian condition for the local assumption.

Deformation transfer. Schild's ladder motivates transferring an initial field through local midpoint and reflection constructions which is an efficient first-order approximation of parallel transport in general curved space of complicated geometry (Lorenzi & Pennec, 2014; Hartman et al., 2023). A consistent construction using the learned midpoint and the $c = 1/2$ inverse is specified in Algorithm 2. For a transferred field $v_{\mathcal{B}}$ and endpoint parameter $Q = \widehat{\mathrm{End}}_{\mathcal{B}}(v_{\mathcal{B}})$, we define the displacement-scale residual

$$\tau_{\mathcal{B}}(\mathcal{X} \to \mathcal{Y}; K) := hF_{\mathcal{B},h}(Q), \qquad \mathrm{PT}(\mathcal{X}, \mathcal{Y}, \mathcal{B}; K) := \mathrm{Intp}(\mathcal{B}, Q; K). \tag{13}$$

Even if $\mathcal{B}$ lies on the original path, transporting the full initial vector is not the same as recomputing the logarithm to $\mathcal{Y}$. Appendix A.6 proves exact recovery of Euclidean vector transport for the linear-path special case, as a toy case for better understanding the transfer scheme.

Path regression and empirical rectification. Let $s_i = s(t_i)$ and order observations by time, with $s_N = 1$. Fixing $\mathcal{X}_0$ to the youngest observation, fit

$$\widehat{\mathcal{X}}_t := \arg\min_{\mathcal{Y}} \sum_i \|\mathcal{X}_i - \gamma_\theta(\mathcal{X}_0, \mathcal{Y}; s_i)\|_F^2 + \|\mathcal{Y} - \mathcal{X}_N\|_F^2. \tag{14}$$

With $\widehat{\mathcal{X}}_i = \gamma_\theta(\mathcal{X}_0, \widehat{\mathcal{X}}_t; s_i)$, the reported rectification rule aggregates transferred residuals,

$$g_0 = \sum_i \tau_{\mathcal{X}_0}(\widehat{\mathcal{X}}_i \to \mathcal{X}_i; K), \qquad g_D = \sum_i \tau_{\widehat{\mathcal{X}}_t}(\widehat{\mathcal{X}}_i \to \mathcal{X}_i; K), \tag{15}$$

and updates the endpoint parameters as $\mathcal{X}_0^+ = \widehat{\mathrm{End}}_{\mathcal{X}_0}(\eta_0 g_0/h)$ and $\widehat{\mathcal{X}}_t^+ = \widehat{\mathrm{End}}_{\widehat{\mathcal{X}}_t}(\eta_D g_D/h)$, with $\eta_0 = \eta_D = 10^{-4}$. This optional one-step correction is an empirical update, not a derived gradient or a guaranteed descent step. Appendix A.5 gives the actual objective gradients and the Euclidean least-squares reference. Although the actual objective gradients are intractable in high-dimensional shape space and Euclidean case is too simplified, they can give the intuition to better understand the operators. Algorithm 3 summarizes this procedure.

5 Experiments

We evaluate reconstruction, local path descriptors, cohort templates, endpoint inversion, deformation transfer, and cross-sectional fitting. These tasks address different claims: reconstruction tests endpoint fidelity; geometric descriptors characterize local changes; inversion tests recoverability; downstream tasks assess practical utility.

5.1 Shape reconstruction

Two geometric metrics in reconstruction/registration scenarios are used (Hartman et al., 2024; Iyer & Elhabian, 2023; Lüdke et al., 2022): **Chamfer Distance** (**CD**) and **Surface-to-Surface Distance** (**S2S**). SSM metrics for population-level statistics (Munsell et al., 2008), including generality (Gen), specificity (Spec) w.r.t. CD, and compactness (Com), are used with their area under the curve (AUC) in Tab. 1. The description of these metrics is in Appendix A.9 and Appendix A.11. The methods used for comparison include FlowSSM (Lüdke et al., 2022), SmS (Cao et al., 2024)[1], FUSS (El Amrani et al., 2024a), FUSS trained without CD loss (FUSS w/o CD), ESA (Hartman et al., 2023) and Point2SSM++ (Adams et al., 2026). Except for optimization-based ESA, the methods are split into 80% training, 10% validation, and 10% test data for model training and performance comparisons. Table 1 reports the lowest mean CD and S2S for ENF-SSM on each dataset. Relative to FUSS w/o CD, the CD differences are approximately 12.0%, 5.5%, and 4.1% for pancreas, hippocampus, and heart. Also, the significantly better performance of FUSS w/o CD against FUSS and SmS supports that CD and ARAP are *ill-suited* for anatomical shape modeling, as analyzed in Sec. 4.1, and ENF-SSM's better performance than FUSS w/o CD supports the used GESM-based elastic deformation energy over simple area preserving energy. ENF-SSM outperforms the optimization-based counterpart ESA with a similar elastic deformation energy, supporting the plausibility of using neural representations.

5.2 Shape interpolation

Here we test whether adjacent frames change local geometry moderately as a valid surrogate of a path primitive. Concretely, with the interpolation trajectory $\{\mathcal{X}(k)\}_{k=0}^{K+1}$, using the face-wise incremental deformation gradient, we record four descriptors: logarithmic strain ε, area change α, conformal distortion κ, and dihedral change β. They isolate stretching, scaling, shearing and bending of the same step; the precise forms are given in Appendix A.12. Also, the average vertex-wise $L2$ distance along interpolation is used to measure the variation of adjacent frame. Tab. 2 reports the path averages $\bar{\varepsilon}$, $\bar{\alpha}$, $\bar{\kappa}$, $\bar{\beta}$ and $\overline{\mathrm{L2}}$. The intermediate frames of FUSS moves slightly, and its matching mainly depends on the non-degraded corresponding of SmS's spectral matching due to its CD-based reguarlization. ARAP in SmS contributes to small deformable variations, but also impedes its variations among the frames. FUSS w/o CD promotes area-preserving and has small $\bar{\varepsilon}$ and $\bar{\alpha}$ related to deformation energy. These supports the necessity of a general deformation-modeling energy. Noticeably, ENF-SSM also outperforms ESA, an optimization-based explicit modeling method. All these support the plausibility of learned interpolant as surrogates for the path in shape space.

[1]The SmS implementation here does not use CD-based test-time adaptation, as studied by FUSS.

Table 1: Geometric metrics and AUC of SSM metrics w.r.t. PCA embedding modes. Geometric metrics are reported as mean ± standard deviation. For the heart dataset, CD, S2S, Gen and Spec are scaled by 10^3 as noted in Appendix A.8. The best performance is **bold and underlined**, the second best is underlined. The AUCs of the SSM metrics are calculated from the curves in Fig. 3.

Dataset	Methods	Geometrical Metric		AUC of SSM metric		
		CD (mm) ↓	S2S (mm) ↓	Gen (CD) ↓	Spec (CD) ↓	Comp ↑
Pancreas	FlowSSM	2.32 ± 0.39	0.79 ± 0.18	2.04	3.21	0.817
	FUSS	2.97 ± 0.40	0.97 ± 0.19	2.73	3.38	0.788
	SmS	2.67 ± 0.31	0.82 ± 0.14	2.12	3.31	0.796
	FUSS w/o CD	1.84 ± 0.27	0.52 ± 0.11	1.75	**3.10**	0.816
	ESA	2.21 ± 0.34	0.69 ± 0.15	1.91	3.24	0.819
	Point2SSM++	2.25 ± 0.32	0.67 ± 0.14	1.88	3.14	**0.832**
	ENF-SSM	$\mathbf{1.62 \pm 0.46}$	$\mathbf{0.49 \pm 0.12}$	**1.64**	3.11	0.823
Hipp	FlowSSM	1.60 ± 0.64	0.62 ± 0.31	1.14	1.33	0.815
	FUSS	2.05 ± 0.75	0.75 ± 0.37	1.32	1.45	0.801
	SmS	1.67 ± 0.59	0.65 ± 0.28	1.22	1.39	0.809
	FUSS w/o CD	1.27 ± 0.52	0.49 ± 0.26	0.93	1.29	0.838
	ESA	1.31 ± 0.55	0.52 ± 0.27	1.03	1.29	0.831
	Point2SSM++	1.47 ± 0.63	0.59 ± 0.28	0.99	1.30	0.825
	ENF-SSM	$\mathbf{1.20 \pm 0.54}$	$\mathbf{0.45 \pm 0.27}$	**0.91**	**1.28**	**0.864**
Heart	FlowSSM	2.14 ± 0.20	0.71 ± 0.12	1.06	2.50	0.917
	FUSS	2.82 ± 0.28	1.03 ± 0.18	1.20	2.63	0.899
	SmS	2.08 ± 0.23	0.69 ± 0.15	1.18	2.56	0.923
	FUSS w/o CD	1.95 ± 0.23	0.65 ± 0.13	0.99	2.41	0.926
	ESA	1.97 ± 0.24	0.64 ± 0.16	1.11	2.51	0.936
	Point2SSM++	1.94 ± 0.22	0.63 ± 0.13	1.02	**2.36**	0.931
	ENF-SSM	$\mathbf{1.87 \pm 0.21}$	$\mathbf{0.60 \pm 0.13}$	0.96	2.38	**0.937**

Table 2: Geometric metrics on interpolation paths from the first pancreas shape to the remaining 272 shapes. Entries are mean ± standard deviation.

Method	$\bar{\varepsilon} \downarrow$	$\bar{\alpha} \downarrow$	$\bar{\kappa} \downarrow$	$\bar{\beta} \downarrow$	$\overline{\mathrm{L2}} \uparrow$
FUSS	0.010 ± 0.002	0.007 ± 0.001	0.008 ± 0.002	0.004 ± 0.001	0.32 ± 0.12
SmS	0.036 ± 0.015	0.024 ± 0.009	0.022 ± 0.010	0.006 ± 0.002	1.28 ± 0.86
FUSS w/o CD	0.102 ± 0.035	0.046 ± 0.014	0.102 ± 0.033	0.055 ± 0.012	1.48 ± 0.96
ESA	0.142 ± 0.123	0.084 ± 0.054	0.072 ± 0.049	0.041 ± 0.018	1.73 ± 1.28
ENF-SSM	0.094 ± 0.023	0.048 ± 0.016	0.038 ± 0.021	0.021 ± 0.010	1.54 ± 1.02

5.3 Anatomical template estimation from shape cohort

We compute a recursive path-based template with Algorithm 1 and compare with ImplicitAtlas (Yang et al., 2022), NDF (Sun et al., 2022), RDA-INR (Dummer et al., 2024), and ESA (Hartman et al., 2023). For the 273-shape pancreas cohort, Table 3 reports mean CD and squared varifold discrepancy to aligned cohort members. ENF-SSM has the smallest recorded means; the gaps to ESA are 0.54 in CD and 0.13 in the displayed varifold scale, and ENF-SSM avoids the non-mesh artifacts of the optimization-based ESA, as shown in Fig.4. For pancreas dataset, ENF-SSM takes four minutes, much faster than ESA's 220 minutes.

5.4 Shape evolution via initial value

The evaluation contains three parts: **(1) Numerical experiments:** On the pancreas dataset, using $K = 6$, $c = k/(K+1)$, $k = 1, \ldots, K$ and the intermediate frame $\mathcal{X}(c(K+1))$ to solve Eqn. (11), we compare the estimated $\hat{\mathcal{Y}}_c$ with the ground-truth $\mathcal{Y}$ by average L^2 distance, as shown in the left panel of Tab. 4 and Fig.5 of an example. The reconstruction errors decrease as c increases (approaching the target shape). **(2) Other numerical methods:** We use the initial-value solvers of ESA and LDDMM (with ENF-SSM's correspondence as landmarks) with the same number of steps for comparison. At shared time stamp "c=1/7", ENF-SSM achieves better than ESA and LDDMM. In terms of runtime, ENF-SSM takes less than five seconds and landmark LDDMM takes 30 seconds, while ESA takes more than 5 minutes. Related visualization results and analysis are in Appendix 5. **(3) Clinical applications:** On ADNI subjects with at least three visits, we use extended IV as *held-out later-visit prediction* as below. From the ADNI hippocampus dataset, patients with at least three

Table 3: Comparisons of template from different methods using Chamfer distance and varifold distance.

	ImplicitAtlas	NDF	RDA-INR	ESA	ENF-SSM
CD ↓	26.78 ± 4.82	25.10 ± 4.37	25.34 ± 4.58	24.66 ± 3.98	24.12 ± 3.65
Varifold (10^{-2}) ↓	7.76 ± 2.51	6.54 ± 2.16	6.75 ± 2.27	6.51 ± 2.08	6.38 ± 1.97

Table 4: Endpoint-inversion self-consistency (L^2, left) and recorded later-visit volume error (VD, right).

Method	c=1/7	c=2/7	c=3/7	c=4/7	c=5/7	c=6/7	c=1.0	VD ↓
NAISR	—	—	—	—	—	—	—	0.092±0.072
ENF-SSM	2.25±0.15	0.41 ± 0.10	0.11±0.01	0.09±0.01	0.07±0.01	0.06±0.01	0.05±0.01	0.095±0.061
ESA	3.35±0.48	—	—	—	—	—	—	—
LDDMM	2.53±0.16	—	—	—	—	—	—	—

Table 5: Deformation-transfer descriptors (left) and frozen attribute observer (right). Age/BMI ACC are the banded criteria of Appendix A.14; MAE is the primary continuous score.

Method	$d_R \downarrow$	$\Delta\varphi \downarrow$	Time (s)	sex ACC↑	age ACC ↑	BMI ACC ↑	age MAE↓	BMI MAE↓
LDDMM	8.69 ± 0.53	19.71 ± 7.32	40.5	0.97	0.58	0.67	5.6	2.1
ESA	7.56 ± 0.82	16.58 ± 6.47	753.6	0.98	0.62	0.69	4.9	2.2
ENF-SSM	5.61 ± 0.37	14.48 ± 7.58	2.1	0.98	0.81	0.89	2.9	1.3

imaging visits, ordered by chronological age $t_0 < t_1 < \cdots < t_n$, are used for evaluation. Let $\mathcal{X}^{(0)}, \mathcal{X}^{(1)}, \ldots, \mathcal{X}^{(n)}$ denote the aligned sequence and normalized time $s_i = (t_i - t_0)/(t_n - t_0)$. Take $c = s_1$ and estimate the target shape $\hat{\mathcal{Y}}_c$ using Eqn. (11); the estimated intermediate shape is $\hat{\mathcal{X}}(s_i) = \mathcal{X}^{(0)} + s_i \Omega(\text{concat}[\mathcal{X}^{(0)}; \hat{\mathcal{Y}}_c - \mathcal{X}^{(0)}; s_i \mathbf{1}])$. Following NAISR (Jiao et al., 2024), Volume Difference (VD) between the gold standard and the estimate is used for quantification. The results are in the rightmost panel of Tab. 4. ENF-SSM achieves results comparable to NAISR without using any time covariate for training.

5.5 Parallel transport for deformation transfer

On pancreas, we conduct numerical experiment using $N = 100$ random ordered triples $(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ and compares source and transferred deformation descriptors, d_R and $\Delta\varphi$ (Piras et al., 2024), against LDDMM and ESA. Table 5 reports smaller mean descriptor differences and lower online time for ENF-SSM, and Fig.6 qualitatively shows one sample of parallel transport-based deformation transfer results, where ENF-SSM shows more plausible result. To test whether the transferred shape reflects clinical attributes, we train an attribute-prediction network on the heart dataset, where the input is the heart shape and the output is its sex, age, and BMI. We then input the transferred shape and test whether the network outputs the desired attributes. The prediction network, evaluation metrics of the associated attributes and test samples are in Appendix A.14, and the performance comparison is in the right panel of Tab. 5. ENF-SSM yields transported shapes whose morphology better agrees with the specified sex, age, and BMI, supporting its use for demographic deformation transfer.

5.6 Cross-sectional path regression

On ADNI hippocampus meshes we fit a *cross-sectional* shape path to study the age trend. Forty surfaces are sampled uniformly along chronological age so that $s_i = (t_i - t_{\min})/(t_{\max} - t_{\min}) \in [0, 1]$. The youngest mesh is the intercept; all other surfaces are mapped onto its triangulation by the matching network. Deformetrica (Bône et al., 2018b) uses landmark L^2 attachment and a control-point LDDMM geodesic. Geomstats Fletcher regression on $\mathbb{R}^{3V}$ (Miolane et al., 2020; Thomas Fletcher, 2013). ENF-SSM **w/o** Rctf uses only Eqn. (14); the full variant adds the one-step rectification of Eqn. (15). Tab. 6 and Fig. 7 show fitting fidelity versus path regularity. The evaluation metrics include fidelity measured by the mean vertex-wise L^2 distance averaged over samples ($\overline{L^2}$), and the deformation metrics in Sec. 5.2 that quantify trajectory smoothness. With less time, ENF-SSM achieves better fitting fidelity and produces a much smoother deformable trajectory than LDDMM and Geomstats. ENF-SSM **w/o** rectification is much faster than the complete version, and its trajectory-level deformation is smaller. However, as shown in Fig. 7, ENF-SSM **w/o** rectification produces oversmoothed results and may overlook local deformable variations. Either setting can be chosen according to whether fidelity to local variation or a compact mean trend is preferred.

6 Ablation studies

Comparisons of different elastic deformation energies: We compare GESM-based deformation energy with another widely used deformation energy elastic thin-shell (TS) (Heeren et al., 2012;

Table 6: Cross-sectional path fitting on 40 hippocampus samples. Deformation metrics are reported as means, and the L^2 average distance is reported as mean $\pm$ standard deviation.

Method	$\bar{\varepsilon} \downarrow$	$\bar{\alpha} \downarrow$	$\bar{\kappa} \downarrow$	$\bar{\beta} \downarrow$	$\overline{\mathrm{L2}} \downarrow$	Time (s) $\downarrow$
LDDMM	0.13	0.027	0.026	0.018	2.96 ± 0.74	362.5
Geomstats	0.15	0.035	0.028	0.019	2.85 ± 0.76	482.5
ENF-SSM **w/o** Rctf	0.057	0.016	0.016	0.015	2.63 ± 0.81	96.4
ENF-SSM	0.067	0.021	0.018	0.019	2.32 ± 0.70	190.7

Table 7: Comparisons of the different elastic deformation energy and different component of the proposed elastic deformation energy with neural interpolant implementation.

Method	$\bar{\varepsilon} \downarrow$	$\bar{\alpha} \downarrow$	$\bar{\kappa} \downarrow$	$\bar{\beta} \downarrow$	CD $\downarrow$
TS	0.106 ± 0.026	0.054 ± 0.017	0.022 ± 0.010	0.026 ± 0.008	1.91 ± 0.49
ENF-SSM	0.094 ± 0.023	0.048 ± 0.016	0.038 ± 0.021	0.021 ± 0.010	1.62 ± 0.46
w/ bend	0.147 ± 0.089	0.055 ± 0.027	0.088 ± 0.028	0.014 ± 0.008	1.82 ± 0.44
w/ scale	0.083 ± 0.021	0.041 ± 0.013	0.067 ± 0.031	0.052 ± 0.021	1.69 ± 0.47
w/ shear	0.138 ± 0.047	0.053 ± 0.028	0.029 ± 0.014	0.047 ± 0.019	1.91 ± 0.53

Zhang et al., 2015) for training a neural interpolant. Results on the pancreas dataset are reported in Tab. 7, using the reconstruction Chamfer distance from Sec. 5.1 and the deformation measures from Sec. 5.2. Drawing on the experimental results, we conclude that GESM is a good choice for training the SmS interpolant. Moreover, to show the necessity of the different deformation components in Eqn. (6), we conduct the experiment that only retains one deformation energy at each time, and compare with Eqn. (6) using the metrics above, as shown in Tab. 7. The full combination of deformation terms improves performance over using any single component.

7 Limitations and future work

The first limitation is that ENF-SSM is designed for topology-preserving anatomical shape analysis and may not be suitable for lesions such as perforation and fistula that introduce topological changes. The second limitation is that ENF-SSM uses the SmS/FUSS network architecture (Cao et al., 2024; El Amrani et al., 2024a) as its backbone, which is a mesh-based shape analysis framework with strict topological constraints. Other neural interpolants based on implicit surfaces or point clouds could be integrated with our Rm-SSM operators may be able to handle changing topology. Future work based on ENF-SSM is to extend the current framework with uncertainty-aware Rm-SSM, and to use the existing ENF-SSM toolkit as efficient basic solvers for a more comprehensive model (Bône et al., 2020).

8 Conclusions

In thiw paper, we propose an ENF-SSM framework to amortize the repeated path computations of Riemannian statistical shape modeling with one endpoint-conditioned interpolant. Trained on a Spectral Meets Spatial backbone with a GESM energy for bending, scaling, and shearing plus spatial smoothness, the frozen network supplies a finite-step logarithm and, through direct evaluation or short inversion, recursive templates, Schild-style transport, and cross-sectional regression. We derive the conditional bounds that separate time-discretization error, reference-path error, inverse conditioning, and endpoint attachment. On three anatomical datasets of pancreas, hippocampus, and heart meshes, ENF-SSM shows better or competitive performance with iterative and deep-learning baselines in reconstruction and shape-statistics tasks, at substantially lower online cost. The promising result indicates its potential as an effective, efficient and interpretable statistical shape modeling tool for anatomical analysis and scientific discovery.

AI use statement

In this work, we used generative AI tools for aid or polish writing. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ACKNOWLEDGMENTS

Pancreas meshes were obtained from the ShapeWorks resource (Cates et al., 2017); we thank the Scientific Computing and Imaging Institute at the University of Utah for distributing this correspondence software and the associated pancreas surface cohort. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Cardiac meshes were obtained from the representative models released with (Ugurlu et al., 2025) with the accompanying public archive.

REFERENCES

Jadie Adams and Shireen Elhabian. Point2SSM: Learning Morphological Variations of Anatomies from Point Clouds. In *The Twelfth International Conference on Learning Representations*, 2024.

Jadie Adams, Mokshagna Sai Teja Karanam, and Shireen Elhabian. Point2ssm++: Self-supervised learning of anatomical shape models from point clouds. *Medical Image Analysis*, 111:104073, 2026.

Lennart Bastian, Yizheng Xie, Nassir Navab, and Zorah Lähner. Hybrid functional maps for crease-aware non-isometric shape matching. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3313–3323. IEEE, 2024.

M Faisal Beg, Michael I Miller, Alain Trouvé, and Laurent Younes. Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International Journal of Computer Vision*, 61:139–157, 2005.

Alexandre Bône, Olivier Colliot, and Stanley Durrleman. Learning distributions of shape trajectories from longitudinal datasets: a hierarchical model on a manifold of diffeomorphisms. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9271–9280. IEEE, 2018a.

Alexandre Bône, Maxime Louis, Benoît Martin, and Stanley Durrleman. Deformetrica 4: an open-source software for statistical shape analysis. In *Shape in Medical Imaging Workshop, International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 3–13. Springer, 2018b.

Alexandre Bône, Olivier Colliot, Stanley Durrleman, and Alzheimer's Disease Neuroimaging Initiative. Learning the spatiotemporal variability in longitudinal shape data sets. *International Journal of Computer Vision*, 128(12):2873–2896, 2020.

Dongliang Cao, Marvin Eisenberger, Nafie El Amrani, Daniel Cremers, and Florian Bernard. Spectral meets spatial: Harmonising 3d shape matching and interpolation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3658–3668, 2024.

Joshua Cates, Shireen Elhabian, and Ross Whitaker. Shapeworks: particle-based shape correspondence and visualization software. In *Statistical shape and deformation analysis*, pp. 257–298. Elsevier, 2017.

Benjamin Charlier, Nicolas Charon, and Alain Trouvé. The fshape framework for the variability analysis of functional shapes. *Foundations of Computational Mathematics*, 17:287–357, 2017.

Nicolas Charon and Alain Trouvé. The varifold representation of nonoriented shapes for diffeomorphic registration. *SIAM Journal on Imaging Sciences*, 6(4):2547–2580, 2013.

Timothy F Cootes, Christopher J Taylor, David H Cooper, and Jim Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

Luca Cosmo, Antonio Norelli, Oshri Halimi, Ron Kimmel, and Emanuele Rodola. Limp: Learning latent shape representations with metric preservation priors. In *European conference on computer vision*, pp. 19–35. Springer, 2020.

Ian L Dryden, Alexey Koloydenko, and Diwei Zhou. Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging. *The Annals of Applied Statistics*, pp. 1102–1123, 2009.

Sven Dummer, Nicola Strisciuglio, and Christoph Brune. Rda-inr: Riemannian diffeomorphic autoencoding via implicit neural representations. *SIAM Journal on Imaging Sciences*, 17(4): 2302–2330, 2024.

Marvin Eisenberger and Daniel Cremers. Hamiltonian dynamics for real-world shape interpolation. In *European conference on computer vision*, pp. 179–196. Springer, 2020.

Marvin Eisenberger, David Novotny, Gael Kerchenbaum, Patrick Labatut, Natalia Neverova, Daniel Cremers, and Andrea Vedaldi. Neuromorph: Unsupervised shape interpolation and correspondence in one go. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7473–7483, 2021.

Nafie El Amrani, Dongliang Cao, and Florian Bernard. A universal and flexible framework for unsupervised statistical shape model learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 26–36. Springer, 2024a.

Nafie El Amrani, Dongliang Cao, and Florian Bernard. A universal and flexible framework for unsupervised statistical shape model learning. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, October 2024b.

Jean-Rassaire Fouefack, Bhushan Borotikar, Marcel Lüthi, Tania S Douglas, Valérie Burdin, and Tinashe EM Mutsvangwa. Dynamic multi feature-class gaussian process models. *Medical Image Analysis*, 85:102730, 2023.

Xiao-Ming Fu, Yang Liu, and Baining Guo. Computing locally injective mappings by advanced mips. *ACM Transactions on Graphics (TOG)*, 34(4):1–12, 2015.

Nicolas Guigui, Nina Miolane, Xavier Pennec, et al. Introduction to riemannian geometry and geometric statistics: from basic theory to implementation with geomstats. *Foundations and Trends® in Machine Learning*, 16(3):329–493, 2023.

Emmanuel Hartman, Yashil Sukurdeep, Eric Klassen, Nicolas Charon, and Martin Bauer. Elastic shape analysis of surfaces with second-order sobolev metrics: a comprehensive numerical framework. *International Journal of Computer Vision*, 131(5):1183–1209, 2023.

Emmanuel Hartman, Emery Pierson, Martin Bauer, Mohamed Daoudi, and Nicolas Charon. Basis restricted elastic shape analysis on the space of unregistered surfaces. *International Journal of Computer Vision*, pp. 1–26, 2024.

Adam Hartshorne, Allen Paul, Tony Shardlow, and Neill DF Campbell. Arc-flow: Articulated, resolution-agnostic, correspondence-free matching and interpolation of 3d shapes under flow fields. In *2025 International Conference on 3D Vision (3DV)*, pp. 11–33. IEEE, 2025.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Behrend Heeren, Martin Rumpf, Max Wardetzky, and Benedikt Wirth. Time-discrete geodesics in the space of shells. In *Computer Graphics Forum*, volume 31, pp. 1755–1764. Wiley Online Library, 2012.

Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: a review. *Medical Image Analysis*, 13(4):543–563, 2009.

Jeffrey Ho, Guang Cheng, Hesamoddin Salehian, and Baba Vemuri. Recursive karcher expectation estimators and geometric law of large numbers. In *Artificial Intelligence and Statistics*, pp. 325–332. PMLR, 2013.

Qixing Huang, Xiangru Huang, Bo Sun, Zaiwei Zhang, Junfeng Jiang, and Chandrajit Bajaj. Arapreg: An as-rigid-as possible regularization loss for learning deformable shape generators. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 5815–5825, 2021.

Krithika Iyer and Shireen Y Elhabian. Mesh2ssm: From surface meshes to statistical shape models of anatomy. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 615–625. Springer, 2023.

Yining Jiao, Carlton Jude ZDANSKI, Julia S Kimbell, Andrew Prince, Cameron P Worden, Samuel Kirse, Christopher Rutter, Benjamin Shields, William Alexander Dunn, Jisan Mahmud, and Marc Niethammer. NAISR: A 3d neural additive model for interpretable shape representation. In *The Twelfth International Conference on Learning Representations*, 2024.

Yining Jiao, Sreekalyani Bhamidi, Carlton Jude ZDANSKI, Julia S Kimbell, Andrew Prince, Cameron P Worden, Samuel Kirse, Christopher Rutter, Benjamin H Shields, William Alexander Dunn, Jisan Mahmud, and Marc Niethammer. $\texttt{PRISM}$:a 3d probabilistic neural representation for interpretable shape modeling. In *Forty-third International Conference on Machine Learning*, 2026.

David G Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London mathematical society*, 16(2):81–121, 1984.

Sebastian Kurtek, Eric Klassen, John C Gore, Zhaohua Ding, and Anuj Srivastava. Elastic geodesic paths in shape space of parameterized surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1717–1730, 2011.

Yonghyeon Lee, Seungyeon Kim, Jinwon Choi, and Frank Park. A statistical manifold framework for point cloud data. In *International Conference on Machine Learning*, pp. 12378–12402. PMLR, 2022.

Dong C Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, 1989.

William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 163–169. Association for Computing Machinery, 1987.

Marco Lorenzi and Xavier Pennec. Efficient parallel transport of deformations in time series of images: from schild's to pole ladder. *Journal of mathematical imaging and vision*, 50(1):5–17, 2014.

David Lüdke, Tamaz Amiranashvili, Felix Ambellan, Ivan Ezhov, Bjoern H Menze, and Stefan Zachow. Landmark-free statistical shape modeling via neural flow deformations. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 453–463. Springer, 2022.

Marcel Lüthi, Thomas Gerig, Christoph Jud, and Thomas Vetter. Gaussian process morphable models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1860–1873, 2017.

Robin Magnet et al. Smooth non-rigid shape matching via effective dirichlet energy optimization. In *2022 International Conference on 3D Vision*, pp. 495–504. IEEE, 2022.

Nina Miolane, Nicolas Guigui, Alice Le Brigant, Johan Mathe, Benjamin Hou, Yann Thanwerdas, Stefan Heyder, Olivier Peltre, Niklas Koep, Hadi Zaatiti, et al. Geomstats: A python package for riemannian geometry in machine learning. *Journal of Machine Learning Research*, 21(223):1–9, 2020.

Brent C Munsell, Pahal Dalal, and Song Wang. Evaluating shape correspondence for statistical shape analysis: A benchmark study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):2023–2039, 2008.

Adele Myers, Caitlin Taylor, Emily Jacobs, and Nina Miolane. Geodesic regression characterizes 3d shape changes in the female brain during menstruation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2542–2551, 2023.

Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 165–174, 2019.

Xavier Pennec. Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25:127–154, 2006.

Xavier Pennec, Stefan Sommer, and Tom Fletcher. *Riemannian Geometric Statistics in Medical Image Analysis*. Academic Press, 2019.

Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, Clifford R Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3): 201–209, 2010.

Steffen E Petersen, Nay Aung, Mihir M Sanghvi, Filip Zemrak, Kenneth Fung, Jose Miguel Paiva, Jane M Francis, Mohammed Y Khanji, Elena Lukaschuk, Aaron M Lee, et al. Reference ranges for cardiac structure and function using cardiovascular magnetic resonance (cmr) in caucasians from the uk biobank population cohort. *Journal of cardiovascular magnetic resonance*, 19(1):18, 2017.

Paolo Piras, Nicolas Guigui, and Valerio Varano. Comparison of different parallel transport methods for the study of deformations in 3d cardiac data. *Journal of Mathematical Imaging and Vision*, 66 (3):393–415, 2024.

Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

Zahra Raisi-Estabragh, Ahmed Salih, Polyxeni Gkontra, Angélica Atehortúa, Petia Radeva, Ilaria Boscolo Galazzo, Gloria Menegaz, Nicholas C Harvey, Karim Lekadir, and Steffen E Petersen. Estimation of biological heart age using cardiovascular magnetic resonance radiomics. *Scientific Reports*, 12(1):12805, 2022.

Martin Rumpf and Benedikt Wirth. Variational time discretization of geodesic calculus. *IMA Journal of Numerical Analysis*, 35(3):1011–1046, 2015.

Ahmed M Salih, Esmeralda Ruiz Pujadas, Víctor M Campello, Celeste McCracken, Nicholas C Harvey, Stefan Neubauer, Karim Lekadir, Thomas E Nichols, Steffen E Petersen, and Zahra Raisi-Estabragh. Image-based biological heart age estimation reveals differential aging patterns across cardiac chambers. *Journal of Magnetic Resonance Imaging*, 58(6):1797–1812, 2023.

Lu Sang, Zehranaz Canfes, Dongliang Cao, Florian Bernard, and Daniel Cremers. Implicit neural surface deformation with explicit velocity fields. *arXiv preprint arXiv:2501.14038*, 2025a.

Lu Sang, Zehranaz Canfes, Dongliang Cao, Riccardo Marin, Florian Bernard, and Daniel Cremers. 4deform: Neural surface deformation for robust shape interpolation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6542–6551. IEEE, 2025b.

Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov. Diffusionnet: Discretization agnostic learning on surfaces. *ACM Transactions on Graphics*, 41(3):1–16, 2022.

Olga Sorkine and Marc Alexa. As-rigid-as-possible surface modeling. In *Symposium on Geometry processing*, volume 4, pp. 109–116. The Eurographics Association, 2007.

Anuj Srivastava and Eric P Klassen. *Functional and Shape Data Analysis*, volume 1. Springer, 2016.

Zhe Su, Martin Bauer, Stephen C Preston, Hamid Laga, and Eric Klassen. Shape analysis of surfaces using general elastic metrics. *Journal of Mathematical Imaging and Vision*, 62:1087–1106, 2020.

Shanlin Sun, Kun Han, Deying Kong, Hao Tang, Xiangyi Yan, and Xiaohui Xie. Topology-preserving shape reconstruction and registration via neural diffeomorphic flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20845–20855, 2022.

P Thomas Fletcher. Geodesic regression and the theory of least squares on riemannian manifolds. *International journal of computer vision*, 105(2):171–185, 2013.

Devran Ugurlu, Shuang Qian, Elliot Fairweather, Charlene Mauger, Bram Ruijsink, Laura Dal Toso, Yu Deng, Marina Strocchi, Reza Razavi, Alistair Young, et al. Cardiac digital twins at scale from mri: open tools and representative models from˜ 55000 uk biobank participants. *Plos one*, 20(7): e0327158, 2025.

Miaowei Wang, Changjian Li, and Amir Vaxman. Canfields: Consolidating diffeomorphic flows for non-rigid 4d interpolation from arbitrary-length sequences. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 28587–28598, 2025.

Yue Wang, Yongbin Sun, Ziwei Liu, Sanjay E Sarma, Michael M Bronstein, and Justin M Solomon. Dynamic graph cnn for learning on point clouds. *ACM Transactions on Graphics*, 38(5):1–12, 2019.

Haitao Yang, Bo Sun, Liyan Chen, Amy Pavel, and Qixing Huang. Geolatent: A geometric approach to latent space design for deformable shape generators. *ACM Transactions on Graphics*, 42(6): 1–20, 2023.

Jiancheng Yang, Udaranga Wickramasinghe, Bingbing Ni, and Pascal Fua. Implicitatlas: learning deformable shape templates in medical imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15861–15871, 2022.

Yusuke Yoshiyasu, Wan-Chun Ma, Eiichi Yoshida, and Fumio Kanehiro. As-conformal-as-possible surface registration. *Computer Graphics Forum*, 33(5):257–267, 2014.

Chao Zhang, Behrend Heeren, Martin Rumpf, and William AP Smith. Shell pca: statistical shape modelling in shell space. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1671–1679, 2015.

Zhengwu Zhang, Yuexuan Wu, Di Xiong, Joseph G Ibrahim, Anuj Srivastava, and Hongtu Zhu. Lesa: Longitudinal elastic shape analysis of brain subcortical structures. *Journal of the American Statistical Association*, 118(541):3–17, 2023.

A Appendix

A.1 Notation summary

The conventions below distinguish a path, an endpoint parameter and an initial field. Norms on vertex matrices are Frobenius norms, equivalent to Euclidean norms after vectorization.

Table 8: Main notation and time conventions.

Symbol	Meaning
$\mathcal{X} \in \mathbb{R}^{V\times 3}$	Vertex coordinates in a fixed triangulation
$\mathcal{D} \in \mathbb{R}^{V\times 3}$	Vertex displacement field
$\mathbf{D} \in \mathbb{R}^{3\times 2}$	Facewise displacement differences
$\mathbf{A},\ g = \mathbf{A}^\top \mathbf{A}$	Face tangent basis; its 2×2 Gram matrix
$\mathbf{P}_{\mathcal{XY}}$	Transfer of $\mathcal{Y}$ coordinates to $\mathcal{X}$ vertices
$T = K+1,\ h = 1/T$	Steps; time interval between adjacent frames
$\gamma_\theta(\mathcal{X}, \mathcal{Y}; t)$	Time-t network evaluation, not an endpoint parameter
$\widehat{\mathrm{Log}}^h_{\mathcal{X}}(\mathcal{Y})$	Initial displacement divided by h
$\widehat{\mathrm{End}}_{\mathcal{X}}(v)$	Endpoint parameter obtained by initial-field inversion
$\mathrm{IV}(\mathcal{X}, d; K)$	Path with prescribed initial displacement d
$\tau_{\mathcal{B}}$	Transferred residual in displacement units

A.2 Supplementary description of Spectral Meets Spatial (SmS) Interpolant

Spectral correspondence (SC) module: Given two meshes $\mathcal{X}$ and $\mathcal{Y}$ with n_0 and n_1 vertices (w.l.o.g., $n_0 \le n_1$), the SC module first uses DiffusionNet (Sharp et al., 2022) to extract vertex-wise mesh features, denoted as $\mathbf{F}_{\mathcal{X}}$ and $\mathbf{F}_{\mathcal{Y}}$ respectively. Then soft correspondence $\mathbf{S}_{\mathcal{XY}} \in \mathbb{R}^{n_0\times n_1}$ in form of stochastic matrix is calculated as

$$\mathbf{S}_{\mathcal{XY}} := \mathrm{Softmax}(\mathbf{F}_{\mathcal{X}}\mathbf{F}_{\mathcal{Y}}^T), \tag{16}$$

where Softmax is applied in each row. Thus $\mathbf{S}_{\mathcal{XY}}$ pulls target vertex functions onto source vertices. Define $\mathbf{S}_{\mathcal{YX}}$ separately by row-normalizing the reverse feature product; it need not equal $\mathbf{S}_{\mathcal{XY}}^\top$. Let $\mathbf{\Phi}_{\mathcal{X}} \in \mathbb{R}^{n_0\times k}$ and $\mathbf{\Phi}_{\mathcal{Y}} \in \mathbb{R}^{n_1\times k}$ contain the spectral bases, and $\Lambda_{\mathcal{X}}, \Lambda_{\mathcal{Y}} \in \mathbb{R}^{k\times k}$ contain their eigenvalues. With mass-orthonormal bases, $\mathbf{\Phi}_{\mathcal{X}}^\dagger = \mathbf{\Phi}_{\mathcal{X}}^\top M_{\mathcal{X}}$. Here $\mathbf{C}_{\mathcal{XY}}$ maps coefficients of functions on $\mathcal{X}$ to coefficients on $\mathcal{Y}$.

With these dimensions and directions, a consistent descriptor and Laplacian-commutativity objective is

$$\mathbf{C}_{\mathcal{XY}} := \arg\min_{\mathbf{C}} \|\mathbf{C}\mathbf{\Phi}_{\mathcal{X}}^\dagger \mathbf{F}_{\mathcal{X}} - \mathbf{\Phi}_{\mathcal{Y}}^\dagger \mathbf{F}_{\mathcal{Y}}\|_F^2 + \alpha\|\mathbf{C}\Lambda_{\mathcal{X}} - \Lambda_{\mathcal{Y}}\mathbf{C}\|_F^2. \tag{17}$$

The reverse map is estimated analogously. Standard bijectivity and orthogonality penalties are

$$\begin{aligned} L_{\mathrm{bij}} &= \|\mathbf{C}_{\mathcal{XY}}\mathbf{C}_{\mathcal{YX}} - I\|_F^2 + \|\mathbf{C}_{\mathcal{YX}}\mathbf{C}_{\mathcal{XY}} - I\|_F^2, \\ L_{\mathrm{orth}} &= \|\mathbf{C}_{\mathcal{XY}}^\top\mathbf{C}_{\mathcal{XY}} - I\|_F^2 + \|\mathbf{C}_{\mathcal{YX}}^\top\mathbf{C}_{\mathcal{YX}} - I\|_F^2. \end{aligned} \tag{18}$$

Coupling the $\mathcal{X} \to \mathcal{Y}$ functional map to the matching vertex transfer requires the reverse pullback:

$$L_{\mathrm{coup}} = \|\mathbf{C}_{\mathcal{XY}} - \mathbf{\Phi}_{\mathcal{Y}}^\dagger \mathbf{S}_{\mathcal{YX}} \mathbf{\Phi}_{\mathcal{X}}\|_F^2. \tag{19}$$

The coordinate transfer $\mathbf{P}_{\mathcal{XY}} \in \{0,1\}^{n_0\times n_1}$ has one selected target vertex per source row. It is not a permutation matrix when it is rectangular, and rowwise selection alone does not guarantee a bijection or a valid embedded remeshing. The hard-assignment procedure must therefore be distinguished from the spectral bijectivity penalty.

Spatial interpolation (SI) module and joint training: The configuration of the SI module is important throughout the paper, as in Eqn. (1). It is implemented with five EdgeConv blocks (Wang et al., 2019) and residual connections (He et al., 2016).

We next introduce the joint training loss with the SC module. The SI losses are: **1)** bidirectional elastic loss $L_{\text{bi_els}}$ on every step; **2)** source-side correspondence alignment L_{align}; **3)** explicit endpoint reconstruction L_{end} that enforces $\mathcal{X}(K+1) \approx \mathbf{P}_{\mathcal{XY}}\mathcal{Y}$; **4)** path consistency L_{consist}, including the endpoint $k = K+1$. Their formulations are as follows:

$$L_{\text{bi_els}} := \sum\nolimits_{k=0}^{K} \Big(E_{\text{els}}(\mathcal{X}(k+1) - \mathcal{X}(k) | \mathcal{X}(k)) + E_{\text{els}}(\mathcal{Y}(k+1) - \mathcal{Y}(k) | \mathcal{Y}(k)) \Big), \tag{20a}$$

$$L_{\text{align}} := \|\mathcal{X}(0) - \mathbf{P}_{\mathcal{XY}}\mathcal{Y}\|_F^2 + \|\mathcal{Y}(0) - \mathbf{P}_{\mathcal{YX}}\mathcal{X}\|_F^2, \tag{20b}$$

$$L_{\text{end}} := \|\mathcal{X}(K+1) - \mathbf{P}_{\mathcal{XY}}\mathcal{Y}\|_F^2 + \|\mathcal{Y}(K+1) - \mathbf{P}_{\mathcal{YX}}\mathcal{X}\|_F^2, \tag{20c}$$

$$L_{\text{consist}} := \sum\nolimits_{k=1}^{K+1} \|\mathcal{X}(k) - \mathbf{P}_{\mathcal{XY}}\mathcal{Y}(K+1-k)\|_F^2. \tag{20d}$$

The total loss for training then reads as follows:

$$\begin{aligned} &L_{\text{total}} := \lambda_{\text{corrs}} \mathrm{L}_{\text{corrs}} + \lambda_{\text{bi_els}} L_{\text{bi_els}} + \lambda_{\text{align}} L_{\text{align}} + \lambda_{\text{end}} L_{\text{end}} + \lambda_{\text{consist}} L_{\text{consist}}, \\ &\mathrm{L}_{\text{corrs}} = \mathrm{L}_{\text{bij}} + \mathrm{L}_{\text{orth}} + \mathrm{L}_{\text{coup}}. \end{aligned} \tag{21}$$

where $\lambda_{(\cdot)}$ is the loss weight.

A.3 Proofs and scope of the initial-field and inverse bounds

Proof of Proposition 1. Twice integrating the derivative gives

$$\frac{\gamma_\theta(h) - \gamma_\theta(0)}{h} = \dot{\gamma}_\theta(0) + \frac{1}{h}\int_0^h (h-t)\ddot{\gamma}_\theta(t)\, dt. \tag{22}$$

The remainder norm is at most $h^{-1}\int_0^h (h-t)M\, dt = Mh/2$. Add and subtract $\dot{\gamma}_*(0)$ and apply the triangle inequality. The equality to a Riemannian logarithm requires the reference to be a unit-time minimizing geodesic in a normal neighbourhood. □

The same bound follows if $\dot{\gamma}_\theta$ is Lipschitz with constant M on $[0, h]$. EdgeConv neighbour changes or activation boundaries do not automatically satisfy this assumption. Alternatively, from $\gamma_\theta(t) = X + t\Omega_\theta(X, Y - X, t)$, continuity gives $\dot{\gamma}_\theta(0) = \Omega_\theta(X, Y - X, 0)$ when this derivative exists. If Ω_θ is L_t-Lipschitz in time on this interval, the finite difference differs from that initial value by at most $L_t h$. Neither regularity statement is supplied by endpoint reconstruction alone.

Proof of Proposition 2. The assumed lower Lipschitz bound and the triangle inequality give

$$\begin{aligned} m\|\widehat{Y} - Y_*\|_F &\le \|F(\widehat{Y}) - F(Y_*)\|_F && (23)\\ &\le \|F(\widehat{Y}) - (v_* + \eta)\|_F + \|\eta\|_F + \|v_* - F(Y_*)\|_F && (24)\\ &\le \rho + \|\eta\|_F + \delta_F. && (25) \end{aligned}$$

The terminal-frame statement is another triangle inequality. No property of a particular numerical optimizer is used. □

A sufficient local condition. Use equal-dimensional coordinates on an admissible mesh chart, fixing correspondence and any required gauges. Suppose F is continuously differentiable on a convex neighbourhood U of Y_0, its Jacobian $J_0 = DF(Y_0)$ has smallest singular value $\sigma > 0$, and $\sup_{Y \in U} \|DF(Y) - J_0\|_{\text{op}} \le \sigma/2$. Integration along the segment from Z to Y yields

$$F(Y) - F(Z) = J_0(Y - Z) + \int_0^1 [DF(Z + t(Y - Z)) - J_0](Y - Z)\, dt. \tag{26}$$

The first term has norm at least $\sigma\|Y - Z\|$, and the integral has norm at most $(\sigma/2)\|Y - Z\|$. Hence $m = \sigma/2$ is valid. The inverse function theorem supplies a local inverse near $F(Y_0)$ when J_0 is invertible. This does not imply that every prescribed field lies in its range, that the selected optimizer stays in U, or that L-BFGS finds it.

For intermediate-frame inversion, define $G_c(Y) = \gamma_\theta(X, Y; c) - X = c\Omega_\theta(X, Y - X, c)$. If the map $Y \mapsto \Omega_\theta(X, Y - X, c)$ has lower Lipschitz constant m_c^Ω, the same proof gives

$$\|\widehat{Y}_c - Y_*\|_F \le \frac{\rho_c + \|\eta_c\|_F + \delta_c}{c\, m_c^\Omega}. \tag{27}$$

Here η_c is noise in displacement units, and δ_c is the reference displacement mismatch. If the denominator's other factors remain comparable, smaller c amplifies displacement perturbations. Because the network itself depends on c, this inequality does not prove monotonic improvement with later frames; it identifies a conditioning mechanism to measure.

A.4 A CONDITIONAL CERTIFICATE USING A FIXED-ENDPOINT ACTION

Let $U = (X_1, \dots, X_{T-1})$ denote interior vertex coordinates, with fixed endpoints $X_0 = X$ and $X_T = Y$ in one common triangulation. Let $\mathcal{A}_h(U)$ be Eq. (7), evaluated with fixed learned energy weights.

Proposition 3 (Local action certificate)**.** *Suppose $\mathcal{A}_h$ is differentiable and μ-strongly convex on a convex neighbourhood containing a stationary reference U_* and a candidate U. Then*

$$\begin{aligned} \|U - U_*\| &\le \sqrt{\frac{2[\mathcal{A}_h(U) - \mathcal{A}_h(U_*)]}{\mu}}, \\ \|U - U_*\| &\le \frac{\|\nabla \mathcal{A}_h(U)\|}{\mu}. \end{aligned} \tag{28}$$

Their initial finite differences satisfy

$$\left\| \frac{X_1 - X}{h} - \frac{X_{*,1} - X}{h} \right\| \le \frac{1}{h} \min \left\{ \sqrt{\frac{2[\mathcal{A}_h(U) - \mathcal{A}_h(U_*)]}{\mu}}, \frac{\|\nabla \mathcal{A}_h(U)\|}{\mu} \right\}. \tag{29}$$

Proof. Strong convexity at U_*, where the gradient vanishes, gives $\mathcal{A}_h(U) - \mathcal{A}_h(U_*) \ge (\mu/2)\|U - U_*\|^2$. Strong monotonicity of the gradient gives $\mu\|U - U_*\|^2 \le \langle \nabla \mathcal{A}_h(U), U - U_* \rangle$. Apply Cauchy–Schwarz and divide by $\|U - U_*\|$ when nonzero; otherwise the claim is immediate. The norm of the first frame difference is bounded by the stacked-coordinate norm. Dividing by h proves the final statement. □

This is a conditional certificate, not a claim that anatomical path energies are globally or locally strongly convex in the tested regimes. Establishing a positive local μ requires additional analysis or computation. A numerical reference may be a local solution; it is not automatically a global geodesic. If a candidate has only approximate endpoint attachment, it does not meet the fixed-endpoint premise. A diagnostic can replace its final frame by the common target and recompute the action, reporting that replacement and the original attachment error separately. This operation is not assumed to have been applied in the existing tables.

The $1/h$ factor also explains why a small position/action error does not, by itself, give a uniformly accurate initial velocity as $h \to 0$. To connect this discrete certificate to a continuum logarithm, one must additionally control the time-discretization error of the reference path and the geometric validity of both paths.

A.5 TEMPLATE AND REGRESSION OBJECTIVE DIFFERENTIALS

Vectorize vertex coordinates within a local fixed-correspondence chart. Let $\ell_i(x) = \widehat{\mathrm{Log}}^h_x(X_i)$ and write $E_{\mathrm{els}}(d \mid x) = d^\top H(x) d$ with symmetric $H(x)$. Then $\mathcal{J}(x) = h^2 \sum_i \ell_i(x)^\top H(x) \ell_i(x)$. Wherever these maps are differentiable, its differential in direction u is

$$D\mathcal{J}(x)[u] = h^2 \sum_i \left[2(D\ell_i(x)[u])^\top H(x) \ell_i(x) + \ell_i(x)^\top (DH(x)[u]) \ell_i(x) \right]. \tag{30}$$

This follows directly from the product rule. It exhibits both the dependence of the learned initial field on the base and the dependence of the energy on that base. A sum of surrogate logarithms need not be a negative gradient of this objective. Differentiating through a changing hard assignment would require an additional treatment; Eq. (30) holds the assignment fixed.

There is a useful limiting check. Assume all samples already share a fixed vertex chart, so that the algorithm's alignment is the identity. If $\gamma(X, Y; t) = (1-t)X + tY$ exactly, then $\widehat{\mathrm{Log}}^h_X(Y) = Y - X$,

and the recursion in Algorithm 1 gives $\bar{X}_i = (i+1)^{-1} \sum_{j=0}^{i} X_{\sigma(j)}$ by induction. Thus it reduces to the average of the sampled sequence, including the initialization. For independent uniform sampling with replacement and finite variance, it converges to the cohort's coordinate mean by the law of large numbers. This special case does not establish convergence to the energy-based mean for a nonlinear learned path.

For regression, let $r_i = X_i - \gamma_\theta(X_0, Y; s_i)$ and $\mathcal{R}(X_0, Y) = \sum_i \|r_i\|^2 + \|Y - X_N\|^2$. With $J_{0,i} = D_{X_0}\gamma_\theta(X_0, Y; s_i)$ and $J_{Y,i} = D_Y\gamma_\theta(X_0, Y; s_i)$, the Euclidean coordinate gradients are

$$\nabla_{X_0}\mathcal{R} = -2\sum_i J_{0,i}^\top r_i, \qquad \nabla_Y \mathcal{R} = -2\sum_i J_{Y,i}^\top r_i + 2(Y - X_N). \tag{31}$$

The total derivative $J_{0,i}$ includes the occurrence of X_0 in the input difference $Y - X_0$. These gradients can be obtained by differentiation of the frozen network in a smooth local chart. They do not equal the unweighted transferred residual sums in Eq. (15) in general.

For linear interpolation, $J_{0,i} = (1 - s_i)I$ and $J_{Y,i} = s_i I$. The endpoint gradients therefore weight each residual by its time position. In the same Euclidean special case, exact transport leaves each residual unchanged; with equal rectification step sizes, Eq. (15) shifts both endpoints by the same vector and does not change their difference. This identifies the reported rectification as a heuristic rather than a general slope-gradient update.

For an unconstrained Euclidean regression $a + bt_i$, define $\bar{t} = N^{-1}\sum_i t_i$ and $\bar{X} = N^{-1}\sum_i X_i$. If $\sum_i (t_i - \bar{t})^2 > 0$, differentiating the least-squares objective gives the normal equations and the closed form

$$b = \frac{\sum_i (t_i - \bar{t})(X_i - \bar{X})}{\sum_i (t_i - \bar{t})^2}, \qquad a = \bar{X} - b\bar{t}. \tag{32}$$

For a fixed intercept X_0 and the additional endpoint penalty in Eq. (14), its matched linear-path solution is instead

$$Y = \frac{\sum_i s_i[X_i - (1 - s_i)X_0] + X_N}{\sum_i s_i^2 + 1}. \tag{33}$$

Both formulas apply coordinatewise and require only sums over the cohort. They provide distinct, explicitly specified reference objectives; neither requires iterative geodesic regression in Euclidean space.

A.6 Consistency check for the revised ladder and operator composition

Proposition 4 (Euclidean recovery)**.** *If $\gamma(X, Y; t) = (1 - t)X + tY$, all endpoint inversions are exact, and all shapes lie in the same Euclidean coordinate space, Algorithm 2 transports $v = Y - X$ unchanged to Z and returns the path $Z + t(Y - X)$.*

Proof. The assumptions give $\widehat{\mathrm{Log}}^h_X(Y) = Y - X$, $\widehat{\mathrm{End}}_X(v) = X + v$ and $\widehat{\mathrm{End}}^{1/2}_X(d) = X + 2d$. For one rung from B_j to B_{j+1},

$$S = B_j + hv, \qquad M = \tfrac{1}{2}(B_{j+1} + S), \qquad R = B_j + 2(M - B_j) = B_{j+1} + hv. \tag{34}$$

The updated vector is $\widehat{\mathrm{Log}}^h_{B_{j+1}}(R)/h = v$. Induction proves that every rung preserves v; the final endpoint is $Q = Z + v$ and the returned residual is hv. □

For curved geometry, this check is necessary but insufficient. The learned midpoint, inverse midpoint and initial-field map must approximate consistent reference operations. In particular, an arithmetic coordinate midpoint is not interchangeable with a geodesic midpoint without an error analysis. Nor does the fact that a new base lies on the old path justify replacing a transported full vector by the remaining logarithm to the old endpoint.

A general composition estimate makes the accumulation explicit. Let reference stage maps f_j be L_j-Lipschitz on a common admissible domain, and learned stages $\widehat{f}_j$ satisfy $\|\widehat{f}_j(z) - f_j(z)\| \le \epsilon_j$ there. For the respective composed states, the triangle inequality gives

$$e_j \le \epsilon_j + L_j e_{j-1}, \qquad e_n \le \left(\prod_{j=1}^{n} L_j\right) e_0 + \sum_{i=1}^{n} \epsilon_i \prod_{j=i+1}^{n} L_j. \tag{35}$$

The second inequality follows by recursively substituting the first; an empty product equals one. Inverse conditioning from Proposition 2 contributes to the stage constants. The assumptions require all intermediate states to remain in the domain. They are not established by the reported reconstruction table, so no convergence rate for the full learned ladder is claimed here.

A.7 Algorithm implementation of the main text

Algorithm 1 Recursive path-based template estimation.

Require: Shape set $\mathbb{X} := \{\mathcal{X}_1, ..., \mathcal{X}_N\}$; iteration number N_{iter}, initial guess $\bar{\mathcal{X}}(0) \leftarrow \mathcal{X}_{\sigma(0)}$.
Ensure: Cohort template $\bar{\mathcal{X}}$; minimization of $\mathcal{J}$ is not guaranteed.
1: **for** $i = 1, ..., N_{\text{iter}}$ **do**
2: Sample $\mathcal{X}_{\sigma(i)} \in \mathbb{X}$ with replacement and align its coordinates to the current template.
3: $\bar{\mathcal{X}}(i) \leftarrow \bar{\mathcal{X}}(i-1) + (i+1)^{-1} \widehat{\text{Log}}^{h}_{\bar{\mathcal{X}}(i-1)}(\mathcal{X}_{\sigma(i)})$.
4: **end for**
5: Return $\bar{\mathcal{X}} \leftarrow \bar{\mathcal{X}}(N_{\text{iter}})$.

Algorithm 2 Schild's ladder parallel transport with learned interpolant $\text{PT}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}; K)$.

Require: Corresponding $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$; $T = K + 1$, $h = 1/T$; frozen path and endpoint inversions.
Ensure: Transferred trajectory and residual in displacement units.
1: $B_j \leftarrow \gamma_\theta(\mathcal{X}, \mathcal{Z}; jh)$ for $0 \le j < T$; $B_T \leftarrow \mathcal{Z}$.
2: $v \leftarrow \widehat{\text{Log}}^{h}_{\mathcal{X}}(\mathcal{Y})$.
3: **for** $j = 0, \ldots, T-1$ **do**
4: $S \leftarrow \widehat{\text{End}}_{B_j}(hv)$.
5: $M \leftarrow \gamma_\theta(B_{j+1}, S; 1/2)$.
6: $R \leftarrow \widehat{\text{End}}^{1/2}_{B_j}(M - B_j)$.
7: $v \leftarrow \widehat{\text{Log}}^{h}_{B_{j+1}}(R)/h$.
8: **end for**
9: $Q \leftarrow \widehat{\text{End}}_{\mathcal{Z}}(v)$.
10: **return** $\text{Intp}(\mathcal{Z}, Q; K)$ and $\tau_{\mathcal{Z}} \leftarrow h\widehat{\text{Log}}^{h}_{\mathcal{Z}}(Q)$.

Algorithm 3 Cross-sectional path regression with one-step rectification

Require: Paired cohort $\{(\mathcal{X}_i, t_i)\}_{i=1}^{N}$; $h = 1/(K+1)$; trained interpolant $\text{Intp}(\cdot, \cdot; K)$; step sizes η_0, η_D.
Ensure: Intercept $\mathcal{X}_0$, endpoint $\mathcal{X}_t$, path $\text{Intp}(\mathcal{X}_0, \mathcal{X}_t; K)$.
1: Initialize $\mathcal{X}_0$ as the youngest mesh (or a separately specified cohort template).
2: Estimate $\hat{\mathcal{X}}_t$ by Eqn. (14) and $\hat{\mathcal{X}}_i$ by Eqn. (1).
3: Optional empirical rectification: $g_0 \leftarrow \sum_i \tau_{\mathcal{X}_0}(\hat{\mathcal{X}}_i \to \mathcal{X}_i; K)$, $g_D \leftarrow \sum_i \tau_{\hat{\mathcal{X}}_t}(\hat{\mathcal{X}}_i \to \mathcal{X}_i; K)$ using Eqn. (13).
4: $\mathcal{X}_0 \leftarrow \widehat{\text{End}}_{\mathcal{X}_0}(\eta_0 g_0/h)$, $\mathcal{X}_t \leftarrow \widehat{\text{End}}_{\hat{\mathcal{X}}_t}(\eta_D g_D/h)$.
5: Return $\mathcal{X}_0$, $\mathcal{X}_t$, and $\text{Intp}(\mathcal{X}_0, \mathcal{X}_t; K)$.

A.8 Datasets, training and configuration

Dataset: Details of the datasets used are as follows:

Pancreas (273 shapes) (Cates et al., 2017): Pancreas shapes are highly deformable across individuals, which makes the high degree of nonlinear variability challenging to model. Each shape has 2000 vertices.

ADNI hippocampus (1632 shapes) (Petersen et al., 2010): A multi-anatomy dataset because the hippocampus has left and right parts. Each shape has information on age, sex, and a binary AD tag. Each shape after Marching Cubes (Lorensen & Cline, 1987) on the binary mask has around 3200 vertices.

Heart (1432 shapes) (Ugurlu et al., 2025): Each shape has information on sex, body mass index (BMI) and age. Each shape is pre-processed to have around 5000 vertices.

Training configurations: All experiments run on an NVIDIA A100 GPU. During training, the number of epochs is 15, and the number of interior frames is $K = 6$ ($T = 7$ steps). The number of eigenfunctions is 40, and the optimizer for both the SC and SI modules is Adam with learning rate 10^{-3}. For the loss weights in Eqn. (21), $\lambda_{\text{bi_els}} = 10^2$, $\lambda_{\text{align}} = 10$, $\lambda_{\text{end}} = 10$, $\lambda_{\text{consist}} = 1.0$, and $\lambda_{\text{corrs}} = 1.0$; α in Eqn. (17) is 0.07. In Eqn. (6), $\lambda_1 = \lambda_2 = 10^2$. PyTorch `L-BFGS` is used with learning rate 0.5, at most 50 iterations, history size 20, and strong-Wolfe line search, for solving initial value evolving problem.

A.9 Point-based similarity measure

Given two mesh $\mathcal{X}$ and $\mathcal{Y}$ with vertex set $\{\mathbf{x}_i\}_{i=1}^m$ and $\{\mathbf{y}_j\}_{j=1}^n$, the Chamfer Distance (CD) and Surface-to-Surface Distance (S2S) used in Tab.1 are as follows:

$$\text{CD}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(\frac{1}{m}\sum\nolimits_{i=1}^{m} \min_j ||\mathbf{x}_i - \mathbf{y}_j||_2 + \frac{1}{n}\sum\nolimits_{j=1}^{n} \min_i ||\mathbf{x}_i - \mathbf{y}_j||_2), \tag{36a}$$

$$\text{S2S}(\mathcal{X}, \mathcal{Y}) = \frac{1}{2}(\frac{1}{m}\sum\nolimits_{i=1}^{m} \text{unsign_dist}(\mathbf{x}_i, \mathcal{Y}) + \frac{1}{n}\sum\nolimits_{j=1}^{n} \text{unsign_dist}(\mathbf{y}_j, \mathcal{X})), \tag{36b}$$

where $\text{unsign_dist}(\mathbf{x}, \mathcal{Y})$ is the unsigned distance function from point $\mathbf{x}$ to surface $\mathcal{Y}$. For the geometric metrics in Tab.1, they are averaged for all pair-to-pair similarities in the test sets with N samples, *i.e.*, for $\mathcal{X}_i, \mathcal{X}_j \in \text{test_set}, i \neq j$, denote $\hat{\mathcal{X}_{i\to j}}$ is the matched results from $\mathcal{X}_i$ to $\mathcal{X}_j$, then CD and S2S in Tab.1 is as follows:

$$\text{CD} = \frac{2}{N(N-1)}\sum\nolimits_{i\neq j} \text{CD}(\hat{\mathcal{X}}_{i\to j}, \mathcal{X}_j), \tag{37a}$$

$$\text{S2S} = \frac{2}{N(N-1)}\sum\nolimits_{i\neq j} \text{S2S}(\hat{\mathcal{X}}_{i\to j}, \mathcal{X}_j). \tag{37b}$$

A.10 Mesh-based varifold similarity measures

Varifold kernels compare two meshes without requiring vertex correspondence Charon & Trouvé (2013); Charlier et al. (2017). Write $\mathcal{X} = \{(\mathbf{c}_i, \mathbf{n}_i, s_i)\}_{i=1}^I$ and $\mathcal{Y} = \{(\mathbf{d}_j, \mathbf{m}_j, s_j)\}_{j=1}^J$ for face centres, unit normals, and areas. The scalar kernel is

$$K(\mathcal{X}, \mathcal{Y}) = \sum_{i=1}^{I}\sum_{j=1}^{J} \exp\Big(-\frac{\|\mathbf{c}_i - \mathbf{d}_j\|^2}{\sigma^2}\Big)(\langle \mathbf{n}_i, \mathbf{m}_j\rangle)^2 s_i s_j, \tag{38}$$

with $\sigma = 5$. The quantity reported as a distance in Tab. 3 is the *squared* kernel distance

$$d^2_{\text{var}}(\mathcal{X}, \mathcal{Y}) = K(\mathcal{X}, \mathcal{X}) + K(\mathcal{Y}, \mathcal{Y}) - 2K(\mathcal{X}, \mathcal{Y}), \tag{39}$$

not the raw cross inner product $K(\mathcal{X}, \mathcal{Y})$. Smaller d^2_{var} means closer shapes. Template rows average $d^2_{\text{var}}(\mathcal{X}_0, \mathcal{X}_i)$ over cohort members i.

A.11 Statistical shape model (SSM) evaluation metrics

For SSM metrics, we first pick the template shape that shares the minimal varifold distance Hartman et al. (2023) with the others, and then use linear principal component analysis to obtain a representation of the linear shape space Munsell et al. (2008). This PCA acts on aligned vertex coordinates and assesses a linear shape representation; it is not tangent PCA under E_{els}. The SSM metrics are summarized as follows:

- Compactness (**Comp**): Measures how efficiently the model represents the variance in the training set with few parameters/modes. A more compact model explains more variance using fewer modes. It is calculated as the cumulative function of eigenvalues;
- Generality (**Gen**): Measures the model's ability to represent (reconstruct) any shape from the training population (or unseen shapes from the same class). Lower reconstruction error indicates better generality. It is calculated by using the training sample to linearly approximate the aligned given sample;

- Specificity (**Spec**): Measures how well randomly generated instances from the model resemble valid shapes in the training class (i.e., avoids generating implausible shapes). Lower average distance to the nearest training shape indicates better specificity. It is calculated by generating 500 samples using linear model and then calculating the distance between the generated sample and its closest sample in the training set. In our cases, we choose the Chamfer distance to measure the performance in Generality and Specificity.

A.12 GEOMETRIC METRICS FOR PAIRWISE INTERPOLATION

This appendix gives the four descriptors used in Sec.5.2. All quantities are computed on a path $\{\mathcal{X}(k)\}_{k=0}^{K+1}$ whose frames share the triangulation of $\mathcal{X}(0)$. Indices $k = 1, \dots, K+1$ label steps. Incremental deformation gradient. For a face $\mathbf{f}$ with vertices $(\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2)$, write the tangent bases

$$\mathbf{A}_{\mathbf{f}}^{(k-1)} = \big(\mathbf{v}_1 - \mathbf{v}_0,\ \mathbf{v}_2 - \mathbf{v}_0\big)\big|_{\mathcal{X}(k-1)}, \qquad \mathbf{A}_{\mathbf{f}}^{(k)} = \big(\mathbf{v}_1 - \mathbf{v}_0,\ \mathbf{v}_2 - \mathbf{v}_0\big)\big|_{\mathcal{X}(k)} \in \mathbb{R}^{3\times 2}.$$

Take the thin QR factorization $\mathbf{A}_{\mathbf{f}}^{(k-1)} = Q_{\mathbf{f}} R_{\mathbf{f}}$, with $Q_{\mathbf{f}}^\top Q_{\mathbf{f}} = I_2$ and invertible $R_{\mathbf{f}} \in \mathbb{R}^{2\times 2}$. Coordinates in the reference tangent plane have edge matrix $R_{\mathbf{f}}$. Thus the differential from these orthonormal coordinates to the next embedded triangle is

$$F_{\mathbf{f}}^{(k)} = \mathbf{A}_{\mathbf{f}}^{(k)} R_{\mathbf{f}}^{-1} \in \mathbb{R}^{3\times 2}, \quad C_{\mathbf{f}}^{(k)} = (F_{\mathbf{f}}^{(k)})^\top F_{\mathbf{f}}^{(k)}, \quad U_{\mathbf{f}}^{(k)} = (C_{\mathbf{f}}^{(k)})^{1/2}. \tag{40}$$

For nondegenerate source and target faces, C, U are 2×2 positive-definite matrices. If $\sigma_1 \geq \sigma_2 > 0$ are the eigenvalues of U, then $\sigma_1\sigma_2$ equals the ratio of face areas, since $\det C = \det((\mathbf{A}^{(k)})^\top \mathbf{A}^{(k)}) / \det((\mathbf{A}^{(k-1)})^\top \mathbf{A}^{(k-1)})$. Let $\theta_{\mathbf{f}}^{(k)}$ be the dihedral angle with a consistently designated neighbouring face.

Face-wise incremental descriptors:

$$\varepsilon_{\mathbf{f}}^{(k)} = \big\|\log \mathbf{U}_{\mathbf{f}}^{(k)}\big\|_F, \quad \alpha_{\mathbf{f}}^{(k)} = \big|\sqrt{\det \mathbf{C}_{\mathbf{f}}{}^{(k)}} - 1\big|,$$
$$\kappa_{\mathbf{f}}^{(k)} = \frac{|\sigma_1 - \sigma_2|}{\sigma_1 + \sigma_2}, \qquad \beta_{\mathbf{f}}^{(k)} = \big|\theta_{\mathbf{f}}^{(k)} - \theta_{\mathbf{f}}^{(k-1)}\big|.$$

$\varepsilon_{\mathbf{f}}^{(k)}$ is the logarithmic strain of the step, $\alpha_{\mathbf{f}}^{(k)}$ the relative area change, $\kappa_{\mathbf{f}}^{(k)}$ the conformal (shear) distortion, and $\beta_{\mathbf{f}}^{(k)}$ the change in dihedral angle.

Path averages: With $T = K + 1$ steps and N faces,

$$\bar{\varepsilon} = \frac{1}{TN} \sum_{k=1}^{T} \sum_{\mathbf{f}} \varepsilon_{\mathbf{f}}^{(k)},$$

and $\bar{\alpha}$, $\bar{\kappa}$, $\bar{\beta}$ are defined analogously. The vertex-displacement descriptor is $\overline{L^2} = (TV)^{-1} \sum_{k=0}^{T-1} \sum_{v=1}^{V} \|\mathcal{X}_v(k+1) - \mathcal{X}_v(k)\|_2$. It measures movement without defining a predefined magnitude.

A.13 GEOMETRY-PRESERVING METRICS

Riemannian tensor distance d_R: For each source deformation and transported deformation, compute the 2×2 reference-tangent stretch U using Eq. (40). Comparing tensors at different reference meshes additionally requires an identification of their tangent coordinate frames. Let $O_{\mathbf{f}} \in O(2)$ map source reference coordinates to recipient reference coordinates, fixed independently of the compared deformations. Pull back the transported tensor as $\widetilde{U}_{\mathbf{f}}^{\mathrm{tr}} = O_{\mathbf{f}}^\top U_{\mathbf{f}}^{\mathrm{tr}} O_{\mathbf{f}}$. The positive-definite tensor distance (Dryden et al., 2009) is

$$d_R^{\mathbf{f}} = \Big\|\log\Big[(U_{\mathbf{f}}^{\mathrm{src}})^{-1/2} \widetilde{U}_{\mathbf{f}}^{\mathrm{tr}} (U_{\mathbf{f}}^{\mathrm{src}})^{-1/2}\Big]\Big\|_F, \qquad \bar{d}_R = \frac{1}{N_f} \sum_{\mathbf{f}} d_R^{\mathbf{f}}. \tag{41}$$

This is a tensor-comparison statistic after frame identification, not the Riemannian distance between the anatomical meshes.

Strain-energy variation $\Delta\varphi$: Following (Piras et al., 2024), $\Delta\varphi$ measures how much the local nonlinear strain energy changes after the deformation is transported. For each face, let $\varphi(\mathbf{U})$ denote the nonlinear strain energy of the right stretch tensor $\mathbf{U}$ obtained as above. If $\mathbf{U}^{\mathrm{src}}$ is the stretch of the source-to-target path and $\mathbf{U}^{\mathrm{tr}}$ is that of the transported path on the corresponding face, we set $\Delta\varphi^f = \left|\varphi(\mathbf{U}^{\mathrm{tr}}) - \varphi(\mathbf{U}^{\mathrm{src}})\right|$ and report the face-wise average $\Delta\varphi = \frac{1}{N}\sum_f \Delta\varphi^f$.

A.14 Attribute network for cardiac parallel-transport evaluation

Attribute network architecture and training: We use a separately trained, frozen point-cloud network to read sex, age and BMI from cardiac meshes. Its parameters do not enter ENF training. This separates the optimization objectives, but sample-level independence requires an explicit split audit (V9). Data and labels. We use the Heart cohort (1432 ICP-aligned surfaces, $\approx 5{,}000$ vertices). Sex, age and BMI are parsed from the mesh filename

$$\texttt{(female|male)_age_}\langle a\rangle\texttt{_bmi_}\langle b\rangle,$$

with female/male encoded as $0/1$ and a, b kept as scalars. Only vertices are used; faces are discarded. A mesh is represented by $N = 2048$ randomly sampled vertices. If a surface has fewer than N vertices, sampling is with replacement. Coordinates are centered and divided by the maximum radial norm. During training only, isotropic Gaussian noise with $\sigma = 0.002$ is added before normalization. The cohort is split by stratified sampling on $(\mathrm{sex}, \lfloor \mathrm{age}/5 \rfloor, \lfloor \mathrm{BMI}/3 \rfloor)$. About 8% of each stratum is held out for validation; the remainder is used to train the readout. Age and BMI are standardized with the training-set mean and standard deviation (μ_a, σ_a) and (μ_b, σ_b). These statistics are stored with the checkpoint and reused at test time. Architecture. The encoder is PointNet++ with single-scale grouping (Qi et al., 2017), without feature-propagation layers. Two set-abstraction (SA) blocks are followed by a global max-pooled MLP:

SA1 maps $2048 \to 512$ centroids with $k = 32$ neighbours. The input is relative XYZ. A shared MLP $3 \to 64 \to 64 \to 128$ is applied per neighbour and max-pooled in the neighbourhood, giving a 128-D feature per centroid. SA2 maps $512 \to 128$ centroids with $k = 16$. Relative XYZ is concatenated with the SA1 feature (131-D). The shared MLP is $131 \to 128 \to 128 \to 256$. A pointwise MLP $256 \to 256 \to 512$ is applied to the 128 remaining points and max-pooled to a 512-D global descriptor.

A shared head $512 \to 256$ (ReLU, dropout 0.3) feeds three linear scalars: a sex logit $\hat{s}$, a standardized age $\hat{a}$ and a standardized BMI $\hat{b}$. Farthest-point sampling and k-NN use the corresponding PyTorch3D operators. The network never sees faces, normals, or the interpolant parameters. Training. We minimize

$$\mathcal{L} = \mathrm{BCE}_{\mathrm{logits}}(\hat{s}, s; w_+) + \mathrm{SmoothL1}\Big(\hat{a}, \tfrac{a-\mu_a}{\sigma_a}\Big) + \mathrm{SmoothL1}\Big(\hat{b}, \tfrac{b-\mu_b}{\sigma_b}\Big),$$

where w_+ is the training negative-to-positive sex ratio. Optimization is AdamW (learning rate 10^{-3}, weight decay 10^{-4}) with cosine annealing over 100 epochs, batch size 8, and gradient-norm clipping at 1.0. After each epoch we report sex accuracy and the unstandardized MAEs

$$\mathrm{MAE}_a = \left|(\hat{a}\,\sigma_a + \mu_a) - a\right|, \qquad \mathrm{MAE}_b = \left|(\hat{b}\,\sigma_b + \mu_b) - b\right|$$

on the validation split. The checkpoint minimizing $\mathrm{MAE}_a + 0.5\,\mathrm{MAE}_b$ is kept and frozen for all subsequent tests. The recorded validation performance is sex accuracy 0.98, age MAE approximately 3.0 years, and BMI MAE approximately 1.5.

Evaluation metrics of heart attributes: We treat the demographic readout as a noisy observer rather than an exact labeler. On UK Biobank CMR, image-based heart-age estimators typically incur a mean absolute error (MAE) of about five years (Raisi-Estabragh et al., 2022; Salih et al., 2023), and population reference ranges are reported in decade-wide strata (Petersen et al., 2017). Our network's validation error is of the same order. We therefore report continuous MAE as the primary metric and, as an auxiliary banded accuracy, count a prediction as consistent when

$$|\hat{a} - a| \le 5 \text{ years} \quad \text{and} \quad |\hat{b} - b| \le 2.5\ \mathrm{kg/m^2},$$

i.e., roughly $1.5\times$ the readout MAE and well below a decade. Sex is evaluated as a binary match.

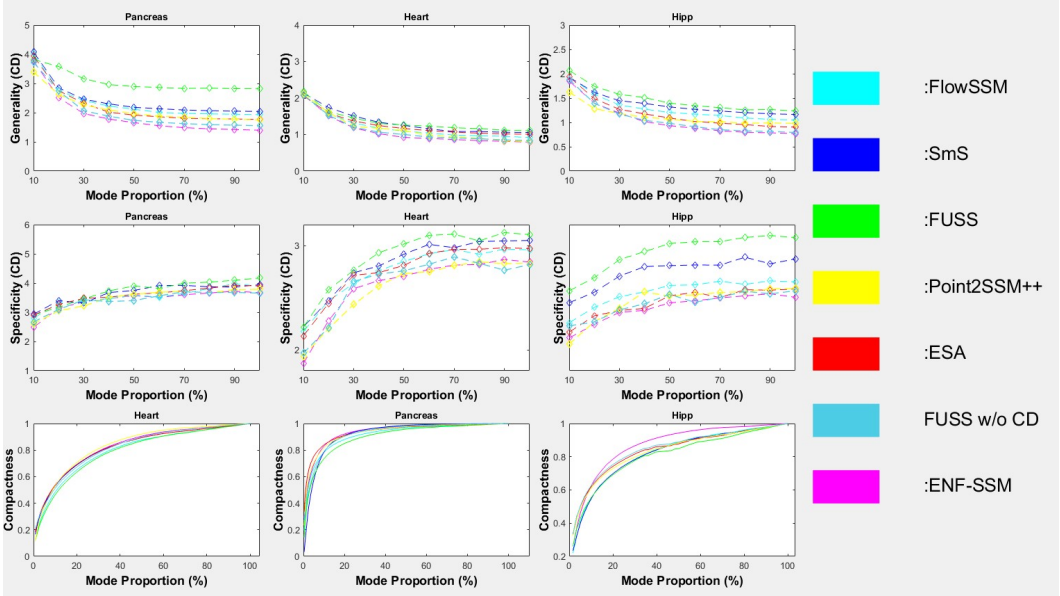


Figure 3: Quantitative plots of SSM metric (generality, specificity and compactness in the column view, respectively) of different methods on three datasets (in the row view). Horizontal axes show the proportions of PCA embedding modes. The vertical axes of generality and specificity show the corresponding chamfer distance (CD). The vertical axes of compactness mark the cumulated normalized PCA eigenvalues.

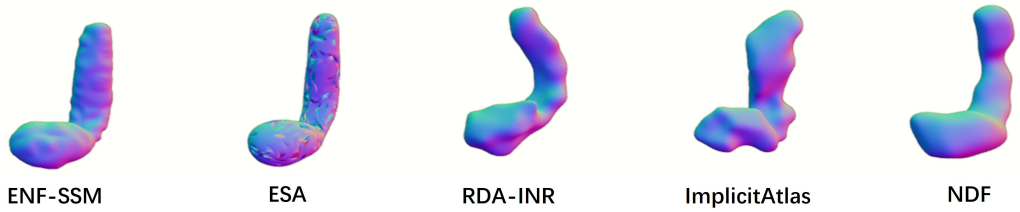
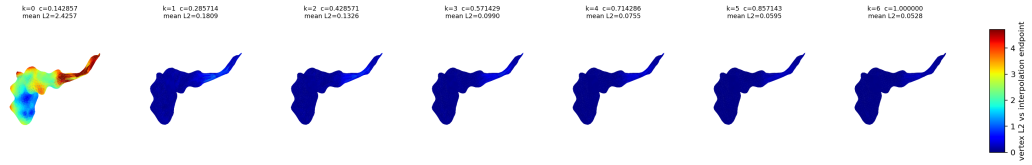
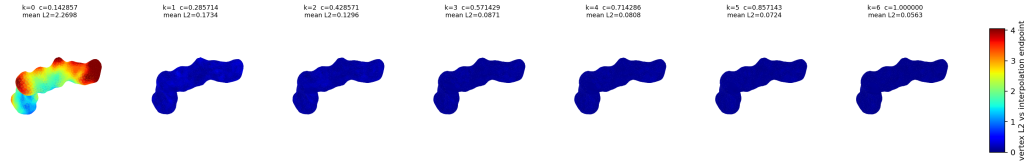


Figure 4: Visualization of template (mean shape) generated by different methods.



(a) Source $\mathcal{X} =$ `pancreas_001`, target `pancreas_249`.



(b) Source $\mathcal{X} =$ `pancreas_001`, target `pancreas_315`.

Figure 5: Extended initial-value evolution on two pancreas pairs. From the interpolation $\mathrm{Intp}(\mathcal{X}, \mathcal{Y}; K)$ with $K = 6$, the displacement at stamp $c_k = (k+1)/(K+1)$ is inverted by $\mathrm{CV}(\mathcal{X}, \hat{\mathcal{D}}_{c_k}, c_k; K)$ and shot to $t = 1$. Colour is the vertex-wise L^2 distance to the interpolation endpoint (same scale on each row). The leftmost panel ($k = 0$, $c = 1/7$) uses the first-step field; later inputs approach the terminal frame, and the $c = 1$ case has mean vertex error approximately 0.05. These panels illustrate inversion self-consistency.

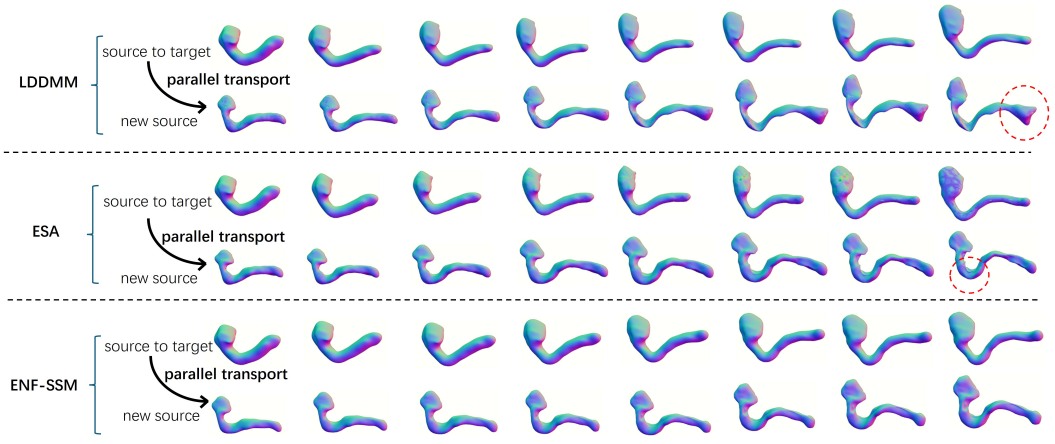


Figure 6: Visualization of parallel transport that transfers the deformation trajectory from `Pancreas_001` (source) to `Pancreas_175` (target) onto `Pancreas_315` (new source) of the pancreas dataset (Cates et al., 2017). Methods for comparison include the LDDMM implementation of Deformetrica (Bône et al., 2018b), ESA (Hartman et al., 2023), and ENF-SSM with Alg. 2. Red dashed circles highlight visual differences between these recorded examples.

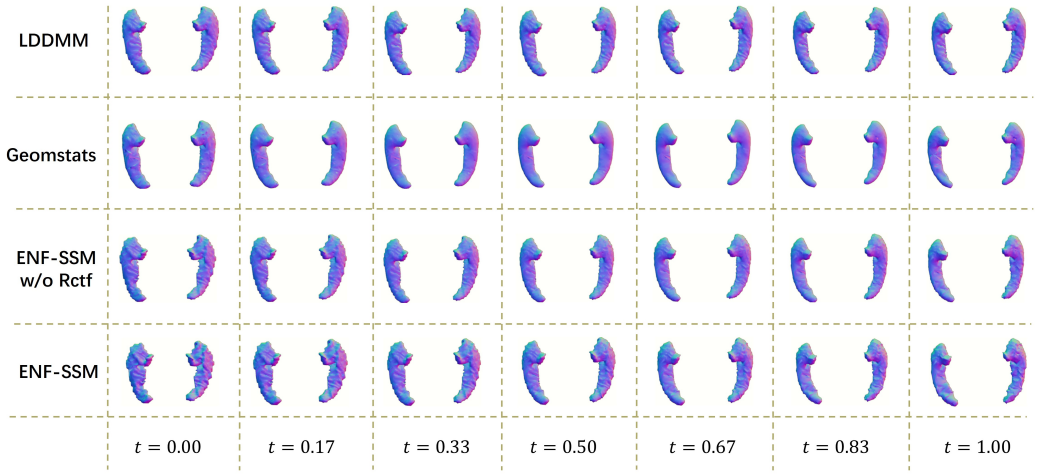


Figure 7: Visualization of shape-space path regression under the experimental schedule of Sec. 5.6. Fitted shapes at normalized times $t = 0.00, 0.17, 0.33, 0.50, 0.67, 0.83$, and 1.00 are shown. Geomstats produces over-smooth results without the wrinkles originally present on the hippocampus samples. LDDMM in the Deformetrica implementation and ENF-SSM **w/o** Rctf retain the necessary local geometry, and ENF-SSM (with rectification) has more details $w.r.t.$ local geometry and attains the best fidelity measures.