NO2SSM: A Discretization-Invariant Neural-Operator Approach to Statistical Shape Modeling

Abu Zahid Bin Aziz[1,2] and Shireen Elhabian[1,2]

[1] Kahlert School of Computing, University of Utah, Salt Lake City, UT 84112, USA
[2] Scientific Computing & Imaging Institute, University of Utah, Salt Lake City, UT 84112, USA
{zahid.aziz,shireen}@sci.utah.edu

Abstract. Statistical shape modeling (SSM) from point clouds is increasingly practical because learned methods can amortize correspondence prediction, but current point-cloud SSM networks remain tied to the input sampling regime used during training. In multi-site anatomical studies, point clouds may come from segmentation or meshing pipelines with different vertex densities, sparse intraoperative scans, or dense research-grade reconstructions; a model trained at one input resolution should therefore produce a consistent ordered point distribution model when the same anatomy is sampled differently. This resolution-generalization gap, rather than native-resolution accuracy alone, is the focus of this work. We introduce NO2SSM, a neural-operator formulation of self-supervised point-cloud SSM that decouples correspondence prediction from the input point count. NO2SSM uses a Geometry-Informed Neural Operator (GINO)-style encoder: radius-based graph integral operators transfer features between the irregular input point cloud and a fixed latent grid, and Fourier Neural Operator layers perform global spectral mixing on that grid. Unlike generic neural operators, which predict fields at query locations, SSM requires a fixed-cardinality, ordered, population-consistent point set. We therefore couple the operator encoder to a Point2SSM-style attention correspondence head that maps fixed landmark identities to probability distributions over the input samples, producing surface-anchored correspondence points as convex combinations of the observed geometry. Across pancreas, liver, and left-atrium datasets, NO2SSM matches state of the art baselines at the native training resolution while degrading substantially less when the input point count and sampling density vary at inference without retraining. This input-resolution robustness enables correspondence upsampling from sparse point-cloud evidence and provides a direct path toward reproducible SSM construction across heterogeneous acquisition and preprocessing settings. Our results suggest that the main value of neural operators for SSM is not simply replacing a point encoder, but addressing a structural resolution-generalization limitation of fixed-discretization point-cloud networks.

Keywords: Statistical Shape Modeling · Neural Operators · Resolution Generalization · Point Clouds.

1 Introduction

Statistical shape modeling (SSM) quantifies population-level variation in anatomical morphology and underpins a wide range of clinical and biomechanical analyses, including neuroscience [11], orthopedics [5, 7], cardiology, implant design [13], and growth modeling [22, 4]. A widely used SSM representation is the correspondence-based point distribution model (PDM), in which every shape in a cohort is described by a dense set of ordered landmarks whose indices have consistent anatomical meaning across subjects [9, 8, 10, 20]. Establishing these correspondences is the central difficulty. Traditional construction methods, such as particle-based optimization [8], can produce high-quality models but typically require clean segmentations or surface meshes, rely on templates or regularization assumptions, and incur group-wise optimization costs that must be repeated or updated when new subjects are added.

Learning-based methods reduce this cost by amortizing SSM prediction. DeepSSM-type methods regress PDM coordinates or loadings from images after correspondences have already been established by an external SSM pipeline [6]. They therefore learn an image-to-PDM shortcut, which is valuable but distinct from discovering correspondences directly from geometry. Point2SSM instead learns ordered correspondence points directly from anatomical point clouds using a self-supervised reconstruction objective and an attention map from fixed landmark identities to input samples [1]. After training, a forward pass yields a PDM for an unseen subject in seconds, making point-cloud SSM construction far more practical at scale.

A remaining limitation is input-resolution generalization. Point-cloud SSM networks are usually trained and evaluated at a prescribed input cardinality, and their learned features are computed over the input sampling itself: PointNet-like models aggregate over a finite set of samples [18, 19], while DGCNN-like models rebuild local neighborhoods from the sampled points [21]. Although such networks may accept different point counts syntactically, their learned neighborhood statistics and attention maps are not operators over a continuous surface. When the same anatomy is represented by a sparser, denser, or differently distributed point set, the effective local scale and feature distribution seen by the network change. This issue is clinically relevant: multi-site cohorts often combine segmentations produced by different preprocessing pipelines, mesh vertex density can vary with scanner resolution or meshing choices, and sparse intraoperative or portable-device scans may need to be compared with dense research-grade reconstructions.

Point2SSM++ addresses an important but distinct robustness problem. It extends Point2SSM with normalization and consistency losses that encourage sampling invariance and rotation equivariance [2]. However, its sampling-consistency objective compares different sampled views at a fixed prescribed cardinality, rather than making the encoder a discretization-independent map that can be evaluated across substantially different input resolutions at inference. Our question is therefore complementary: if the correspondence head, reconstruction loss, and correspondence regularization are held fixed, can replacing a fixed-discretization point encoder with a neural operator improve generalization across input point count and sampling density without retraining?

Neural operators provide a natural modeling framework for this question because they learn mappings between function spaces rather than between fixed-dimensional vectors [12]. DeepONet established the operator-learning paradigm through a branch–trunk architecture [17], while the Fourier Neural Operator (FNO) parameterizes integral kernels in the spectral domain, yielding global receptive fields and zero-shot super-resolution on regular grids [15]. Because FNO assumes grid-structured data, the Geometry-Informed Neural Operator (GINO) bridges irregular geometry to the FNO using graph kernel-integral operators that transfer signals between a point cloud and a fixed latent grid [16]. This design is especially appealing for anatomy, where the underlying surface is continuous but the observed point cloud is only one discretization of that surface.

Applying neural operators to SSM is not a direct use of standard operator learning. Generic neural operators usually regress a field at user-specified query locations; SSM requires a fixed-cardinality, ordered set of correspondence points whose indices remain meaningful across subjects. NO2SSM addresses this operator-to-PDM mismatch by using the neural operator as the geometry encoder, not as an unconstrained coordinate regressor. Specifically, a GINO-style encoder lifts irregular point-cloud samples to a fixed latent grid using radius-based graph integral operators, performs global spectral reasoning with an FNO, and decodes the resulting feature field back to the input samples. A Point2SSM-style attention head then predicts a row-stochastic map from M fixed landmark identities to the N input samples, and each correspondence point is computed as a convex combination of the observed points. This design preserves the ordered PDM output structure while giving the encoder a continuum-style inductive bias for variable input sampling.

The value of this formulation should be evaluated by resolution generalization, not only by in-distribution reconstruction at the training cardinality. We evaluate NO2SSM on pancreas, liver, and left-atrium datasets from the Medical Segmentation Decathlon [3], and compare against Point2SSM and Point2SSM++ under a controlled setup that isolates the encoder contribution. In addition to native-resolution Chamfer distance, point-to-face distance, and standard SSM metrics, we report resolution sweeps, measuring whether predictions for the same subject remain stable as N changes.
Our contributions are as follows:

- We identify input-resolution generalization as a central gap in learned point-cloud SSM and evaluate it explicitly through resolution and sampling-density sweeps.
- We introduce NO2SSM, a neural-operator formulation of point-cloud SSM that replaces the fixed-discretization point encoder with a GINO-style operator encoder while retaining the ordered PDM output structure.
- We define an operator-to-PDM coupling: the operator encoder provides input-sampling robustness, while a Point2SSM-style attention head maps fixed landmark identities to surface-anchored correspondences.

– We compare against Point2SSM and Point2SSM++ under controlled native-resolution, sparse-input, and variable-resolution settings, showing that the operator encoder provides substantially stronger resolution generalization without retraining.

2 Methods

2.1 Problem Formulation

Let $\mathcal{S} = \{S^b\}_{b=1}^{B_{\text{data}}}$ denote a cohort of pre-aligned anatomical shapes, where each full target point cloud $S^b = \{s_i^b\}_{i=1}^{P_b}$, $s_i^b \in \mathbb{R}^3$, samples the surface of subject b. During training, an input point cloud $X^b = \{x_i^b\}_{i=1}^{N_b} \subseteq S^b$ is sampled from this full target. The goal is to predict an ordered set of M correspondence points

$$C^b = \{c_m^b\}_{m=1}^{M}, \qquad c_m^b \in \mathbb{R}^3,$$

such that C^b reconstructs the target anatomy and the point index m has consistent anatomical meaning across subjects.

We learn a mapping

$$\mathcal{F}_\theta : X^b \mapsto C^b, \qquad C^b \in \mathbb{R}^{M \times 3}, \tag{1}$$

in a self-supervised manner. The full point cloud S^b is used only as a reconstruction target; no manually annotated landmarks are required. Following Point2SSM [1], NO2SSM does not directly regress free landmark coordinates. Instead, the network predicts a row-stochastic probability map

$$A^b = \mathcal{A}_\theta(X^b) \in [0,1]^{M \times N_b}, \qquad \sum_{i=1}^{N_b} A_{mi}^b = 1, \tag{2}$$

and computes each correspondence point as a convex combination of the observed input samples:

$$c_m^b = \sum_{i=1}^{N_b} A_{mi}^b x_i^b, \qquad C^b = A^b X^b. \tag{3}$$

The fixed row index m defines the correspondence ordering across subjects, whereas the softmax-normalized weights allow each correspondence point to adapt to subject-specific geometry. Because each output is a convex combination of input samples, the prediction is anchored to the observed geometry and lies in the convex hull of the input point cloud rather than being an unconstrained coordinate regression.

At training time, N_b is usually fixed to a nominal value N_{train}. At inference, however, we allow the input cardinality to vary. Let $X_N^b \subseteq S^b$ denote an N-point sampling of the same underlying anatomy. A resolution-robust correspondence predictor should produce ordered correspondence sets

$$C_N^b = \mathcal{F}_\theta(X_N^b)$$

that remain accurate and anatomically consistent as N and the sampling density change, without retraining the model.

2.2 Neural-Operator Background

Neural operators learn mappings between function spaces rather than mappings between fixed-dimensional vectors [12]. Let $\mathcal{G}^\dagger : a \mapsto u$ denote an unknown operator that maps an input function a defined on a domain D to an output function u. A generic neural-operator layer updates a feature field $v_\ell : D \to \mathbb{R}^{d_\ell}$ by

$$v_{\ell+1}(x) = \sigma\left(W_\ell v_\ell(x) + \int_D \kappa_\ell(x,y) v_\ell(y)\, d\mu(y)\right), \tag{4}$$

where W_ℓ is a pointwise linear map, κ_ℓ is a learned kernel, μ is a measure on D, and σ is a nonlinear activation.

The Fourier Neural Operator (FNO) parameterizes the integral term on a regular grid in the spectral domain [15]. For a grid feature field Z^ℓ, an FNO layer can be written as

$$Z^{\ell+1} = \sigma\left(W_\ell Z^\ell + \mathcal{F}^{-1}\left(\Theta_\ell \cdot \mathcal{F}(Z^\ell)\right)\right), \tag{5}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and inverse Fourier transform, and Θ_ℓ is a learned complex-valued tensor applied only to a truncated set of low-frequency Fourier modes. This spectral parameterization provides global mixing on the latent grid.

The Geometry-Informed Neural Operator (GINO) combines graph neural operator (GNO) integral transforms with FNO layers so that irregular geometric data can be lifted to a regular latent grid, processed spectrally, and decoded back to irregular query locations [16]. NO2SSM follows this GINO-style point–grid–point design, but uses it as a geometry encoder for SSM rather than as a generic field regressor. In the current implementation, the GNO transfer layers use K_G-nearest-neighbor neighborhoods for efficient batched point-cloud processing:

$$\mathcal{N}_{K_G}(q; Y) = \mathrm{kNN}_{K_G}(q, Y), \tag{6}$$

where q is a query coordinate and $Y = \{y_j\}_{j=1}^{n_Y}$ is the source coordinate set. The resulting graph integral layer is

$$(\mathcal{K}_\phi v)(q) = \frac{1}{K_G} \sum_{y_j \in \mathcal{N}_{K_G}(q;Y)} \kappa_\phi\left(q,\ y_j,\ q - y_j,\ v(y_j)\right), \tag{7}$$

where κ_ϕ is an MLP kernel whose inputs are the query coordinate, source coordinate, relative coordinate, and source feature. This gives a GINO-style kernel-integral transfer between irregular points and the regular latent grid. Since the current implementation uses fixed K_G-nearest neighbors rather than fixed-radius neighborhoods, we treat resolution robustness as an empirical property to be evaluated directly through resolution sweeps.

2.3 NO2SSM Architecture

NO2SSM replaces the fixed-discretization point-cloud encoder used in Point2SSM with a GINO-style neural-operator encoder while retaining the Point2SSM attention correspondence head and training losses. Given a normalized input point cloud

$$X = \{x_i\}_{i=1}^{N} \in \mathbb{R}^{N\times 3},$$

the model consists of six stages: coordinate lifting, GNO encoding from the input points to a latent grid, FNO processing on the latent grid, GNO decoding from the latent grid back to the input points, attention-based probability-map prediction, and convex-combination correspondence generation.

Coordinate lifting. Each input coordinate is embedded by a shared MLP:

$$h_i = \rho_\theta(x_i) \in \mathbb{R}^{C_h}, \qquad i = 1, \ldots, N. \tag{8}$$

In the implementation, ρ_θ is a two-layer MLP with GELU activation and hidden width C_h.

GNO encoder: point cloud to latent grid. Let

$$\mathit{\Gamma}_R = \{g_q\}_{q=1}^{R^3} \subset [-1,1]^3 \tag{9}$$

denote a fixed regular latent grid of resolution $R\times R\times R$. The first graph integral operator transfers lifted point features from the irregular input point cloud to the latent grid:

$$z_q^0 = \frac{1}{K_G} \sum_{x_i \in \mathcal{N}_{K_G}(g_q;X)} \kappa_\theta^{\mathrm{in}}\left(g_q,\ x_i,\ g_q - x_i,\ h_i\right), \qquad q = 1, \ldots, R^3. \tag{10}$$

Here $\mathcal{N}_{K_G}(g_q;X)$ denotes the K_G nearest input points to latent-grid coordinate g_q, and $\kappa_\theta^{\mathrm{in}}$ is an MLP kernel.

FNO processing on the latent grid. The grid features $\{z_q^0\}_{q=1}^{R^3}$ are reshaped into

$$Z^0 \in \mathbb{R}^{C_h\times R\times R\times R}$$

and processed by L_F three-dimensional FNO layers:

$$Z^{\ell+1} = \mathrm{GELU}\left(W_\ell Z^\ell + \mathcal{F}^{-1}\left(\mathit{\Theta}_\ell \cdot \mathcal{F}(Z^\ell)\right)\right), \qquad \ell = 0, \ldots, L_F - 1. \tag{11}$$

Only the lowest K_F Fourier modes are retained along each spatial dimension, with K_F capped by the Nyquist limit of the latent grid.

GNO decoder: latent grid to input point cloud. The second graph integral operator decodes the globally mixed latent-grid field back to the original input coordinates:

$$\tilde{h}_i = \frac{1}{K_G} \sum_{g_q \in \mathcal{N}_{K_G}(x_i;\mathit{\Gamma}_R)} \kappa_\theta^{\mathrm{out}}\left(x_i,\ g_q,\ x_i - g_q,\ z_q^{L_F}\right), \qquad i = 1, \ldots, N. \tag{12}$$

This yields a per-point operator feature $\tilde{h}_i \in \mathbb{R}^{C_h}$. To preserve high-frequency local coordinate information that may be attenuated by the latent-grid bottleneck, NO2SSM concatenates the decoded operator feature with the original lifted coordinate feature:

$$\bar{h}_i = [\tilde{h}_i, h_i] \in \mathbb{R}^{2C_h}. \tag{13}$$

A final projection MLP maps this concatenated feature to the latent dimension expected by the correspondence head:

$$f_i = \pi_\theta(\bar{h}_i) \in \mathbb{R}^L. \tag{14}$$

The encoder output is therefore

$$F = [f_1, \ldots, f_N] \in \mathbb{R}^{L \times N}.$$

Attention correspondence head. The correspondence head is the Point2SSM attention module. It maps the per-point feature tensor F to M landmark-specific logits over the input samples. We write the three attention blocks as

$$\Phi_1 : \mathbb{R}^{L \times N} \to \mathbb{R}^{M \times N}, \qquad \Phi_2, \Phi_3 : \mathbb{R}^{M \times N} \to \mathbb{R}^{M \times N}.$$

Each block is a cross-transformer block applied with identical source and target features, using multi-head attention over the point dimension followed by a feed-forward layer. The final probability map is

$$A = \mathrm{softmax}_N \left(\Phi_3 \left(\Phi_2 \left(\Phi_1(F)\right)\right)\right) \in [0,1]^{M \times N}, \tag{15}$$

where the softmax is applied across the N input points. Thus, each row A_m is a probability distribution over the observed point cloud for landmark identity m. The final correspondence prediction is computed by Eq. (3).

2.4 Training Objective

NO2SSM is trained with the same reconstruction and correspondence-regularizing terms used by Point2SSM [1]. The reconstruction term is a Chamfer distance between the predicted correspondence set C^b and the full target point cloud S^b. For $p \in \{1, 2\}$, we define

$$\mathrm{CD}_p(C, S) = \frac{1}{|C|} \sum_{c \in C} \min_{s \in S} \|c - s\|_p^p + \frac{1}{|S|} \sum_{s \in S} \min_{c \in C} \|s - c\|_p^p. \tag{16}$$

The implementation computes both CD_1 and CD_2, and uses the version specified by the training configuration.

Chamfer distance is permutation invariant and therefore does not by itself enforce consistent correspondence ordering. We therefore include the mapping error regularizer used in DPC and Point2SSM [14, 1]. For a source prediction C^a and target prediction C^b, let $\mathcal{N}_K(c_m^a; C^a)$ denote the K nearest neighbors of

c_m^a in C^a, including c_m^a itself. The self-neighbor is removed when computing the loss. The mapping error from C^a to C^b is

$$\mathrm{ME}(C^a, C^b) = \sum_{m=1}^{M} \sum_{j \in \mathcal{N}_K(c_m^a; C^a) \setminus \{m\}} \exp\left(-\frac{\|c_m^a - c_j^a\|_2^2}{\tau} \right) \|c_m^b - c_j^b\|_2^2. \quad (17)$$

This loss encourages local neighborhoods in one subject to remain local at the same correspondence indices in another subject, promoting consistent anatomical parameterization across the cohort. In the implementation, $\tau = 8.0$ and $K = 10$.

For a minibatch of size B, let

$$\mathcal{P} = \{(a, b) : a \neq b,\ y^a = y^b\} \quad (18)$$

be the set of ordered sample pairs with the same anatomy label. In single-anatomy training, all distinct pairs are included. The total loss is

$$\mathcal{L} = \frac{1}{B} \sum_{b=1}^{B} \mathrm{CD}_p(C^b, S^b) + \alpha \frac{1}{|\mathcal{P}|} \sum_{(a,b) \in \mathcal{P}} \mathrm{ME}(C^a, C^b), \quad (19)$$

where α is the mapping-error weight. If no valid same-label pair is present in a minibatch, the mapping-error term is set to zero.

2.5 Evaluation Metrics

An accurate SSM must satisfy two complementary requirements. First, the predicted point set should reconstruct the anatomical surface. Second, the ordered points should define consistent correspondences that yield compact and meaningful population statistics. We therefore report both point-surface accuracy metrics and SSM correspondence metrics.

Point-surface accuracy. We report Chamfer distance, computed as in Eq. (16). When triangular surface meshes are available, we also compute point-to-face distance:

$$\mathrm{P2F}(C, \mathcal{M}) = \frac{1}{M} \sum_{m=1}^{M} \min_{f \in \mathcal{F}(\mathcal{M})} d(c_m, f), \quad (20)$$

where $\mathcal{F}(\mathcal{M})$ is the set of triangular faces in mesh $\mathcal{M}$, and $d(c_m, f)$ is the Euclidean distance from point c_m to face f.

Distributional reconstruction. When comparing point distributions, we also report the Cauchy–Schwarz divergence. Let p_C and p_S be kernel-density estimates of the predicted correspondence set C and target point cloud S, respectively. The Cauchy–Schwarz divergence is

$$\mathrm{CS}(C, S) = -\log \int p_C(x) p_S(x)\, dx + \frac{1}{2} \log \int p_C(x)^2\, dx + \frac{1}{2} \log \int p_S(x)^2\, dx. \quad (21)$$

Lower values indicate greater similarity between the predicted and target point distributions.

SSM correspondence quality. To evaluate whether the predicted ordered points form a useful statistical shape model, each correspondence set is vectorized as

$$\mathbf{c} = [c_1^\top, \dots, c_M^\top]^\top \in \mathbb{R}^{3M},$$

and PCA is performed on the training correspondence cohort. Let $\lambda_1 \geq \lambda_2 \geq \dots$ be the PCA eigenvalues and u_r the corresponding eigenvectors.

Compactness is the number of PCA modes required to explain 95% of the training shape variation:

$$\text{Comp.} = \min \left\{ q : \frac{\sum_{r=1}^{q} \lambda_r}{\sum_r \lambda_r} \geq 0.95 \right\}. \tag{22}$$

A smaller value indicates that the model captures population variation with fewer parameters.

Generalization measures how well the training SSM reconstructs unseen test correspondences. For a test correspondence vector $\mathbf{c}$, let

$$\hat{\mathbf{c}} = \bar{\mathbf{c}} + \sum_{r=1}^{q} \left[(\mathbf{c} - \bar{\mathbf{c}})^\top u_r \right] u_r \tag{23}$$

be its reconstruction using the first q modes. The generalization error is

$$\text{Gen.} = \frac{1}{|\mathcal{D}_{\text{test}}|} \sum_{\mathbf{c} \in \mathcal{D}_{\text{test}}} \|\mathbf{c} - \hat{\mathbf{c}}\|_2^2. \tag{24}$$

Specificity measures whether samples drawn from the learned SSM resemble valid training anatomies. We sample

$$\mathbf{c}' = \bar{\mathbf{c}} + \sum_{r=1}^{q} \xi_r \sqrt{\lambda_r} u_r, \qquad \xi_r \sim \mathcal{N}(0, 1), \tag{25}$$

and compute

$$\text{Spec.} = \frac{1}{|\mathcal{D}_{\text{sample}}|} \sum_{\mathbf{c}' \in \mathcal{D}_{\text{sample}}} \min_{\mathbf{c} \in \mathcal{D}_{\text{train}}} \|\mathbf{c}' - \mathbf{c}\|_2^2. \tag{26}$$

Lower specificity indicates that generated samples lie closer to the empirical training distribution. Finally, we report the average pairwise mapping error in Eq. (17) as an explicit correspondence-neighborhood consistency metric.

2.6 Implementation Details

NO2SSM was implemented in PyTorch. Unless otherwise specified, each training input contained $N_{\text{train}} = 1024$ points and the network predicted $M = 1024$ ordered correspondence points. The latent feature dimension passed to the attention head was $L = 128$. The neural-operator encoder used hidden width $C_h = 64$, a fixed latent grid of resolution $R = 16$ per axis, four FNO layers ($L_F = 4$), and $K_F = 8$ retained Fourier modes along each spatial dimension. Since $R = 16$, this mode count is capped at the latent-grid Nyquist limit. The GNO encoder and decoder each used $K_G = 8$ nearest neighbors and a kernel-MLP hidden width of 64. The skip connection from the lifted coordinate features to the decoded operator features was enabled.

The model was trained with the squared Chamfer variant CD_2, specified in the configuration as `cd_l2`, together with the mapping-error regularizer. The mapping-error weight was $\alpha = 0.1$, and the mapping-error neighborhood size was $K = 10$. Training used Adam with learning rate 5×10^{-5}, $\beta_1 = 0.9$, $\beta_2 = 0.999$, and zero weight decay. The batch size was 4. Models were trained for at most 3000 epochs with early stopping enabled; early stopping began after epoch 500 and used a patience of 100 epochs.

3 Results

3.1 Dataset

We evaluated NO2SSM on three anatomical datasets from the Medical Segmentation Decathlon [3]: pancreas, liver, and left atrium. These anatomies were selected to cover different organ scales, surface complexity, and population variability. For each subject, the segmentation was converted to a surface representation and sampled to obtain a full target point cloud S^b. During training, an input point cloud $X^b \subseteq S^b$ was sampled from the target and the model predicted an ordered correspondence set C^b. Unless otherwise stated, all models were trained with $N_{\text{train}} = 1024$ input points and predicted $M = 1024$ correspondence points.

The same preprocessing, train/validation/test split, reconstruction loss, mapping-error regularizer, correspondence head, optimizer, and training schedule were used for NO2SSM and the Point2SSM-family baselines.

3.2 Results

Point-surface and correspondence metrics We first compared the native-resolution performance of NO2SSM against Point2SSM and Point2SSM++ at $N_{\text{test}} = N_{\text{train}} = 1024$. For each dataset, we report box plots of the surface reconstruction and correspondence metrics over the test subjects. Specifically, we report the squared Chamfer distance CD_2, Cauchy–Schwarz divergence (CS), point-to-face distance (P2F), and mapping error. Lower values indicate better performance for all four metrics.

Figure 1 summarizes these metrics across the pancreas, liver, and left-atrium datasets. Each row corresponds to one dataset and each column corresponds to one metric. Each panel compares Point2SSM, Point2SSM++, and NO2SSM. Across the native-resolution setting, NO2SSM remains competitive with the Point2SSM-family baselines while preserving the same ordered PDM output

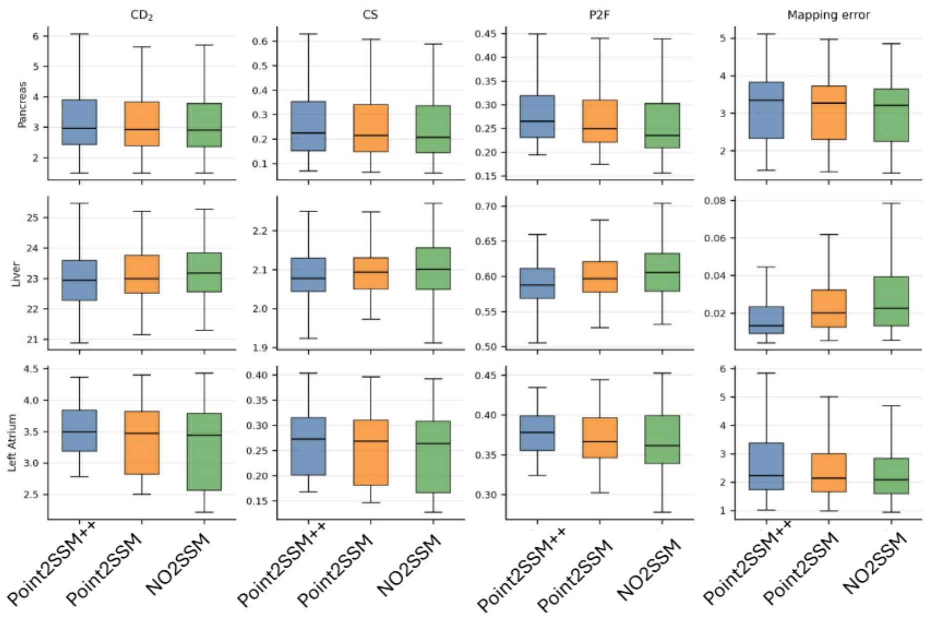

Fig. 1. Native-resolution point-surface and correspondence metrics. Each box summarizes the distribution over test subjects at $N_{\text{test}} = 1024$. Lower values are better for all metrics.

structure. This result is important because it shows that replacing the point-cloud encoder with the neural-operator encoder does not compromise the standard SSM reconstruction setting. The main advantage of NO2SSM is therefore evaluated through resolution generalization rather than claimed solely from native-resolution accuracy.

Table 1 shows that NO2SSM produces SSMs with quality comparable to the Point2SSM-family baselines. Point2SSM++ achieves the best compactness and generalization on all three datasets, which is expected given its additional robustness and consistency terms. NO2SSM remains close to these values, with similar compactness and generalization on the pancreas and liver datasets, and only a small degradation on the left atrium. Notably, NO2SSM achieves the best specificity on the pancreas dataset and remains competitive on the liver and left atrium. These results indicate that replacing the fixed point-cloud encoder with the neural-operator encoder does not substantially degrade the statistical quality of the ordered correspondence model, while the resolution-sweep results demonstrate its stronger robustness to changes in input point count.

Resolution generalization The central claim of NO2SSM is that the learned correspondence predictor should be less tied to the input point count used during training. We therefore trained each model at $N_{\text{train}} = 1024$ and evaluated the

Table 1. SSM quality metrics computed from the ordered correspondence sets. Compactness is the number of PCA modes required to explain 95% of training shape variation. Lower values are better for generalization, specificity, and compactness.

Dataset	Method	Compactness ↓	Generalization ↓	Specificity ↓
Pancreas	Point2SSM	16	2.12	3.83
	Point2SSM++	15	1.99	3.8
	NO2SSM	16	2.17	3.75
Liver	Point2SSM	46	6.12	7.59
	Point2SSM++	42	5.98	7.28
	NO2SSM	45	6.01	7.44
Left atrium	Point2SSM	20	1.9	4.3
	Point2SSM++	18	1.79	4.25
	NO2SSM	21	1.93	4.39

same checkpoint without retraining or fine-tuning at

$$N_{\text{test}} \in \{64, 128, 256, 512, 1024, 2048, 4096\}.$$

For each test resolution, the input point cloud was resampled to the specified cardinality while the target point cloud used for computing reconstruction error was kept fixed. The cases $N_{\text{test}} < 1024$ test sparse-input correspondence prediction, while $N_{\text{test}} > 1024$ tests whether the model remains stable under denser input sampling.

Figure 2 shows the resulting resolution sweep. At the native training resolution, the three methods have similar Chamfer distance. Away from the training resolution, the baselines degrade more strongly, especially under sparse input sampling. NO2SSM maintains substantially lower Chamfer distance at $N_{\text{test}} \leq 512$ and remains stable for dense inputs, supporting the claim that the neural-operator encoder improves resolution generalization relative to fixed point-cloud encoders.

Table 2 summarizes the same experiment. Although Point2SSM and Point2SSM++ have slightly lower native-resolution Chamfer distance, NO2SSM has the lowest mean error over the full resolution sweep and the lowest error under both sparse and dense input settings.

3.3 Ablation Studies

We performed ablation studies on the pancreas dataset to evaluate the two main hyperparameters of the NO2SSM encoder: the latent-grid resolution R and the GNO neighbor count K_G. In both ablations, the model was trained at $N_{\text{train}} = 1024$, and the figures report the native-resolution result at $N_{\text{test}} = 1024$. The correspondence head, output cardinality, training loss, optimizer, FNO depth, hidden width, and training split were held fixed.

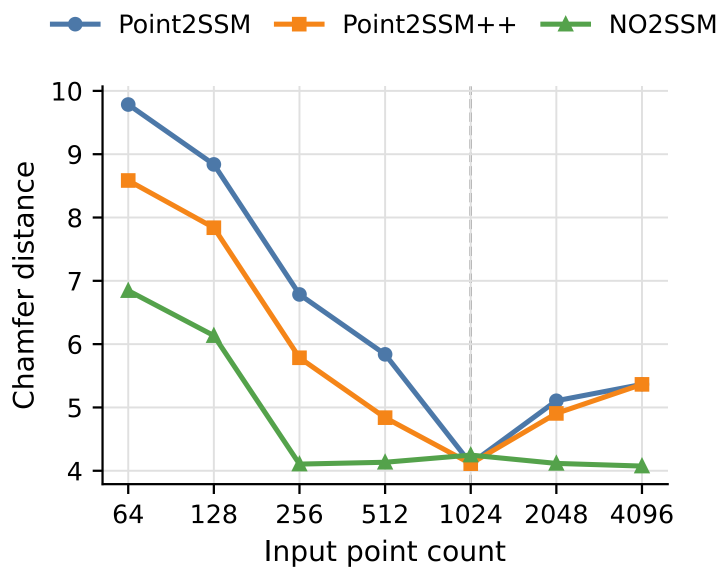

Fig. 2. Resolution-generalization experiment. Each model was trained at $N_{\mathrm{train}} = 1024$ and evaluated without retraining at the input point counts shown. The dashed vertical line indicates the training input resolution.

Latent-grid resolution $\boldsymbol{R}$ The latent grid determines the resolution of the regular feature field on which the FNO performs global spectral mixing. We varied $R \in \{8, 16, 24, 32\}$ while holding all other hyperparameters fixed. Figure 3(A) shows that increasing R improves native-resolution Chamfer distance, with the largest gain observed when increasing from $R = 8$ to $R = 16$. Larger grids continue to improve the metric, but with smaller incremental gains. We therefore use $R = 16$ as the default configuration because it provides a strong accuracy–capacity trade-off while avoiding the substantially larger latent grids required by $R = 24$ and $R = 32$.

GNO neighbor count $\boldsymbol{K_G}$ We next varied the number of nearest neighbors used by the GNO point–grid and grid–point transfer layers. The latent-grid resolution was fixed at $R = 16$, and the GNO neighbor count was varied over $K_G \in \{4, 8, 16, 32\}$. As shown in Figure 3(B), increasing K_G from 4 to 8 substantially improves native-resolution Chamfer distance, while further increasing the neighborhood size produces only marginal gains. This suggests that $K_G = 8$ is sufficient for stable point–grid feature transfer in the default setting, and that larger neighborhoods provide limited additional benefit relative to their increased aggregation cost.

Table 2. Resolution-generalization summary over $N_{\text{test}} \in \{64, 128, 256, 512, 1024, 2048, 4096\}$. Native CD is measured at $N_{\text{test}} = 1024$. Sparse CD averages $N_{\text{test}} \leq 256$, and dense CD averages $N_{\text{test}} \geq 2048$. Max. relative degradation is computed relative to the native-resolution CD for each method. Lower values are better.

Method	Native CD ↓	Mean sweep CD ↓	Sparse CD ↓	Dense CD ↓	Max. relative degradation ↓
Point2SSM	4.112	6.547	8.469	5.235	1.379
Point2SSM++	4.089	5.918	7.403	5.135	1.088
NO2SSM	4.248	4.808	5.695	4.095	0.612

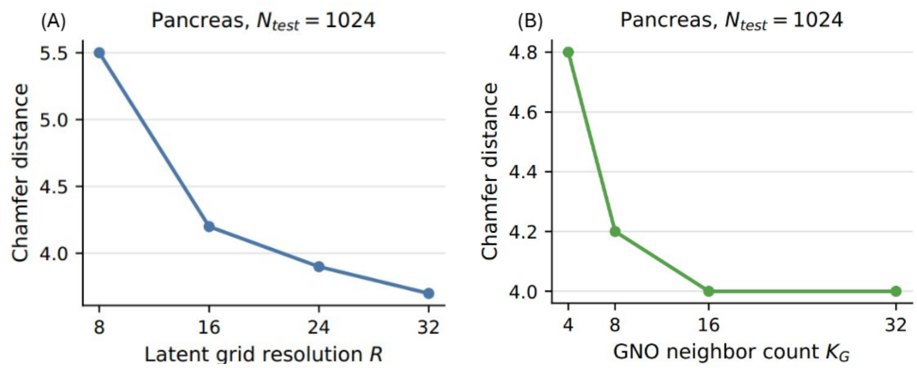

Fig. 3. (A)Latent-grid resolution ablation on the pancreas dataset at $N_{\text{test}} = 1024$. Increasing R improves native-resolution reconstruction, with the largest improvement occurring between $R = 8$ and $R = 16$. (B) GNO neighbor-count ablation on the pancreas dataset at $N_{\text{test}} = 1024$ with $R = 16$. Performance improves from $K_G = 4$ to $K_G = 8$ and then saturates for larger neighborhoods

4 Conclusion

We introduced NO2SSM, a neural-operator formulation of point-cloud statistical shape modeling that replaces the fixed point-cloud encoder of Point2SSM with a GINO-style GNO–FNO–GNO encoder while preserving the attention-based ordered PDM output. Across pancreas, liver, and left-atrium datasets, NO2SSM achieved native-resolution performance comparable to Point2SSM-family baselines and showed stronger robustness when the input point count changed at inference, highlighting resolution generalization as the main benefit of the operator encoder. Ablations on latent-grid resolution and GNO neighborhood size showed that $R = 16$ and $K_G = 8$ provide a practical accuracy–capacity trade-off. The results suggest that neural-operator encoders are a promising direction for SSM methods that must operate across heterogeneous anatomical point-cloud resolutions.

Data Statement. This study relies exclusively on the publicly available Medical Segmentation Decathlon pancreas, liver and left-atrium datasets, which are fully deidentified. As no new data involving human participants were collected, no additional ethical approval or participant consent was required.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Adams, J., Elhabian, S.: Point2SSM: Learning morphological variations of anatomies from point clouds. In: International Conference on Learning Representations (ICLR) (2024), `https://openreview.net/forum?id=DqziS8DG4M`
2. Adams, J., Karanam, M.S.T., Elhabian, S.: Point2ssm++: Self-supervised learning of anatomical shape models from point clouds. Medical Image Analysis p. 104073 (2026)
3. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature communications **13**(1), 4128 (2022)
4. Atkins, P.R., Elhabian, S.Y., Agrawal, P., Harris, M.D., Whitaker, R.T., Weiss, J.A., Peters, C.L., Anderson, A.E.: Quantitative comparison of cortical bone thickness using correspondence-based shape modeling in patients with cam femoroacetabular impingement. Journal of Orthopaedic Research **35**(8), 1743–1753 (2017)
5. Atkins, P.R., Shin, Y., Agrawal, P., Elhabian, S.Y., Whitaker, R.T., Weiss, J.A., Aoki, S.K., Peters, C.L., Anderson, A.E.: Which two-dimensional radiographic measurements of cam femoroacetabular impingement best describe the three-dimensional shape of the proximal femur? Clinical Orthopaedics and Related Research® **477**(1), 242–253 (2019)
6. Bhalodia, R., Elhabian, S.Y., Kavan, L., Whitaker, R.T.: DeepSSM: A deep learning framework for statistical shape modeling from raw images. In: Shape in Medical Imaging (ShapeMI), MICCAI Workshop. pp. 244–257. Springer (2018)
7. Borotikar, B., Mutsvangwa, T.E., Elhabian, S.Y., Audenaert, E.A.: Statistical model-based computational biomechanics: applications in joints and internal organs. Frontiers in Bioengineering and Biotechnology **11**, 1232464 (2023)
8. Cates, J., Elhabian, S., Whitaker, R.: Shapeworks: Particle-based shape correspondence and visualization software. In: Statistical Shape and Deformation Analysis, pp. 257–298. Academic Press (2017)
9. Cates, J., Fletcher, P.T., Styner, M., Shenton, M., Whitaker, R.: Shape modeling and analysis with entropy-based particle systems. In: Information Processing in Medical Imaging: 20th International Conference, IPMI 2007, Kerkrade, The Netherlands, July 2-6, 2007. Proceedings 20. pp. 333–345. Springer (2007)
10. Davies, R.H., Twining, C.J., Cootes, T.F., Waterton, J.C., Taylor, C.J.: A minimum description length approach to statistical shape modeling. IEEE transactions on medical imaging **21**(5), 525–537 (2002)
11. Gerig, G., Styner, M., Shenton, M.E., Lieberman, J.A.: Shape versus size: Improved understanding of the morphology of brain structures. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2001: 4th International Conference Utrecht, The Netherlands, October 14–17, 2001 Proceedings 4. pp. 24–32. Springer (2001)
12. Kovachki, N., Li, Z., Liu, B., Azizzadenesheli, K., Bhattacharya, K., Stuart, A., Anandkumar, A.: Neural operator: Learning maps between function spaces with applications to PDEs. Journal of Machine Learning Research **24**(89), 1–97 (2023)

13. Kozic, N., Weber, S., Büchler, P., Lutz, C., Reimers, N., Ballester, M.Á.G., Reyes, M.: Optimisation of orthopaedic implant design using statistical shape space analysis based on level sets. Medical image analysis **14**(3), 265–275 (2010)
14. Lang, I., Ginzburg, D., Avidan, S., Raviv, D.: Dpc: Unsupervised deep point correspondence via cross and self construction. In: 2021 International Conference on 3D Vision (3DV). pp. 1442–1451. IEEE (2021)
15. Li, Z., Kovachki, N., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., Anandkumar, A.: Fourier neural operator for parametric partial differential equations. In: International Conference on Learning Representations (ICLR) (2021)
16. Li, Z., Kovachki, N., Choy, C., Li, B., Kossaifi, J., Otta, S., Nabian, M.A., Stadler, M., Hundt, C., Azizzadenesheli, K., et al.: Geometry-informed neural operator for large-scale 3d pdes. Advances in Neural Information Processing Systems **36**, 35836–35854 (2023)
17. Lu, L., Jin, P., Pang, G., Zhang, Z., Karniadakis, G.E.: Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence **3**(3), 218–229 (2021). https://doi.org/10.1038/s42256-021-00302-5
18. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: PointNet: Deep learning on point sets for 3D classification and segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 652–660 (2017)
19. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: PointNet++: Deep hierarchical feature learning on point sets in a metric space. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 30 (2017)
20. Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Levitt, J.J., Shenton, M.E., Gerig, G.: Framework for the statistical shape analysis of brain structures using spharm-pdm. The insight journal (1071), 242 (2006)
21. Wang, Y., Sun, Y., Liu, Z., Sarma, S.E., Bronstein, M.M., Solomon, J.M.: Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics (TOG) **38**(5), 1–12 (2019)
22. Zhu, C., Dev, H., Sharbatdaran, A., He, X., Shimonov, D., Chevalier, J.M., Blumenfeld, J.D., Wang, Y., Teichman, K., Shih, G., et al.: Clinical quality control of mri total kidney volume measurements in autosomal dominant polycystic kidney disease. Tomography **9**(4), 1341–1355 (2023)