Point2SSM: Learning Morphological Variations of Anatomies from Point Clouds

Jadie Adams
Scientific Computing and Imaging Institute
and Kahlert School of Computing
University of Utah
Salt Lake City, UT 84112
jadie.adams@utah.edu

Shireen Elhabian
Scientific Computing and Imaging Institute
and Kahlert School of Computing
University of Utah
Salt Lake City, UT 84112
shireen@sci.utah.edu

Abstract

We introduce Point2SSM, a novel unsupervised learning approach that can accurately construct correspondence-based statistical shape models (SSMs) of anatomy directly from point clouds. SSMs are crucial in clinical research for analyzing the population-level morphological variation in bones and organs. However, traditional methods for creating SSMs have limitations that hinder their widespread adoption, such as the need for noise-free surface meshes or binary volumes, reliance on assumptions or predefined templates, and simultaneous optimization of the entire cohort leading to lengthy inference times given new data. Point2SSM overcomes these barriers by providing a data-driven solution that infers SSMs directly from raw point clouds, reducing inference burdens and increasing applicability as point clouds are more easily acquired. Deep learning on 3D point clouds has seen recent success in unsupervised representation learning, point-to-point matching, and shape correspondence; however, their application to constructing SSMs of anatomies is largely unexplored. In this work, we benchmark state-of-the-art point cloud deep networks on the task of SSM and demonstrate that they are not robust to the challenges of anatomical SSM, such as noisy, sparse, or incomplete input and significantly limited training data. Point2SSM addresses these challenges via an attention-based module that provides correspondence mappings from learned point features. We demonstrate that the proposed method significantly outperforms existing networks in terms of both accurate surface sampling and correspondence, better capturing population-level statistics.

1 Introduction

Statistical Shape Modeling (SSM) enables quantifying and characterizing the morphological variations in a group of shapes. SSM captures the inherent characteristics of a shape class or the underlying parameters that remain when global geometric information (i.e., location and orientation) are factored out [40]. It is a powerful tool in medical research, enabling population-level analysis of anatomies such as bones or organs, revealing correlations between shape variations and clinical outcomes. Despite being successfully applied in a wide range of tasks (including pathology detection [7, 15, 65], disease biomarker identification [16, 19, 45, 46], surgical/treatment planning [11, 17, 73, 70], and implant designs [71]), practical limitations have prevented its widespread adoption. The conventional SSM approach entails analyzing a group of *complete* surface representations obtained from 3D medical images (e.g., CT or MRI) in the form of binary volumes or meshes. Correspondence-based SSM establishes sets of geometrically and semantically consistent landmarks or correspondence points on the shape surfaces, providing an interpretable explicit model. Various optimization schemes have automated the correspondence point generation process [18, 27, 51, 57] to provide dense sets of correspondence points for each shape, including particle-based shape modeling (PSM) [18, 20]. However, such optimization techniques have *three significant limitations*. Firstly, they require *complete*

Preprint. Under review.

shape representations in the form of high-resolution meshes or binary volumes that are *free from noise and artifacts*, which are difficult to acquire, prohibiting many use cases. Secondly, the optimization process is time-consuming and must be performed on the entire cohort simultaneously, which greatly *hinders inference* when adding a new shape to the SSM. Lastly, these methods utilize metrics such as Gaussian entropy [18] or parametric representations [51, 57] to define optimization objectives, which may *bias or restrict* the types of variation captured by the SSM (i.e., via enforcing linearity or inheriting the topology of the pre-defined template). Deep learning methods have been proposed to predict SSM from meshes [8, 43, 38]addressing some of these limitations. However, such approaches rely on mesh connectivity, enforcing the same input restrictions as optimization-based methods.

Point cloud deep learning has recently shown success in tasks such as unsupervised representation learning, shape generation, point cloud up-sampling and completion, and point-to-point matching [31, 67, 6]. Point clouds can be readily obtained from full-shape segmentations such as meshes when available. But they can also be obtained from more lightweight shape acquisition methods (e.g., thresholding clinical images, anatomical surface scanning, and combining 2D contour representations [59, 61]). Thus generating SSM directly from point clouds would significantly expand the potential clinical use cases of shape analysis. Recently, Adams and Elhabian [4] demonstrated that existing point cloud encoder-decoder-based completion networks perform reasonably well at SSM generation out-of-the-box. Although such architectures were not designed for correspondence tasks, the bottleneck captures a population-specific shape prior, and the continuous-mapping decoder results in ordered output, providing correspondence as a by-product. However, such methods have not been benchmarked against point correspondence approaches or tested for robustness to noise, partiality, or sparse input [4]. A myriad of challenges accompanies applying point cloud deep learning approaches to anatomical SSM - most notably, the issue of data scarcity. Deep network training requires a large cohort of representative anatomies defined from volumetric medical images, which is difficult to acquire, especially if modeling an uncommon disease or pathology. In addition, point cloud shape representations obtained from medical images may suffer from various issues, such as noise from the acquisition process, missing regions outside the scanner field of view, or sparse point clouds due to low image resolution [54].

In this paper, we introduce Point2SSM, an unsupervised deep learning framework for learning correspondence-based SSM of anatomy directly from point clouds. Point2SSM overcomes the limitations of existing optimization-based SSM methods and point cloud networks by providing a data-driven solution that operates on unordered point clouds and infers SSMs that capture population-level statistics with good surface sampling. Point2SSM removes biases imposed by optimization assumptions and/or templates and significantly relaxes the optimization input requirement. Moreover, a trained Point2SSM provides a fast and efficient way to predict SSM from a new unseen point cloud without re-optimization. We benchmark existing state-of-the-art (SOTA) point cloud networks against the proposed method for the SSM application. This benchmark reveals that although these methods achieve some success on this new task, they have limited ability to address the challenges that arise in clinical scenarios. We show that Point2SSM is more robust to a limited training budget as well as to sparse, noisy, and incomplete input than existing point methods and achieves similar statistical compactness to an optimization-based SSM method. Hence, our contributions can be summarized as follows.

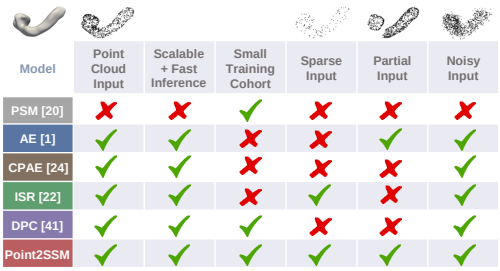

Figure 1: Comparison of partitcle-based modeling[20] (PSM), point autoencoders [1] (AE), canonical point autoencoder [24] (CPAE), Chen et al. [22] (ISR), deep point correspondence [41] (DPC), and Point2SSM (ours).

(1) We introduce Point2SSM, a novel point cloud approach for anatomical SSM that removes the limitations of optimization-based SSM generation techniques without hindering statistical accuracy.

(2) We provide the first benchmark of SOTA point cloud networks on the anatomical SSM task.

(3) We demonstrate Point2SSM outperforms existing methods in terms of surface sampling and correspondence accuracy and is more robust to limited data and sparse, noisy, or incomplete input.

2 Related Work

2.1 Optimization-based SSM

Correspondence-based SSM optimization techniques constrain points to the shape surfaces and establish inter-subject correspondence via metrics such as entropy [18, 20, 50] or minimum description length [27], or via parametric representations [51, 57]. These approaches require *complete, faultless* shape inputs (surface meshes or binary segmentations) and operate on the entire cohort simultaneously, preventing the addition of a new shape without rerunning the optimization process. Convolutional deep learning approaches for predicting SSMs directly from unsegmented images have been proposed to reduce these burdens [10, 5, 2, 3, 12, 62, 60]. However, such approaches are supervised and thus require a traditional optimization scheme for generating a training data cohort, limiting their accuracy potential based on the training set.

2.2 Deep Learning on Point Clouds

Deep learning directly on point clouds is a relatively new research field. Traditional 3D vision techniques for processing point clouds utilized structures like Octrees [36] or voxel hashing [49]. PointNet [52] was the first deep network designed to process raw point clouds. It employs Multi-Layer Perceptrons (MLPs) and symmetric aggregation functions to learn permutation invariant features. PointNet++ [53] extended this architecture to have a hierarchical structure for learning multi-scale geometric information. Dynamic graph convolution (DGCNN) [66] utilized nearest neighbors to construct point cloud graphs and apply edge convolution. Transformer-based methods, including PointTransformer [74] and Point-Bert [68], have also been developed, which apply self-attention to 3D point cloud processing to learn underlying structure. To date, there is no ubiquitous 3D backbone for point cloud networks, but the aforementioned networks are common [67].

The bulk of the proposed point cloud deep networks are trained in a supervised manner via large-scale, densely-labeled datasets. This annotation burden has inspired the exploration of unsupervised learning of robust feature representations for transfer learning [67]. Common unsupervised pretext tasks include self-reconstruction, point cloud up-sampling, and completion of partial point clouds. Achlioptas et al. [1] proposed the first point cloud autoencoder (AE) and demonstrated the generative power of the learned latent space and many generative architectures have since been developed [67]. Point completion networks have widely adopted an encoder and coarse-to-fine decoder architecture, allowing the network to learn the general shape first and then refine it [69, 31]. Point cloud encoder-decoder-based networks such as these have been shown to perform reasonably well at SSM generation Adams and Elhabian [4]. Though these methods were not designed for SSM, the continuous mapping from the learned latent space to output space results in decoded ordered point clouds, providing correspondence as a by product. However, such methods require a sufficiently large and representative training dataset and have yet to be compared to point correspondence approaches [4].

2.3 Learning 3D Point Cloud Dense Correspondence

Defining dense correspondence between 3D shapes is a fundamental task in computer vision and geometric processing with various applications, including robot grasping, non-rigid human body alignment, articulated motion transfer, face swapping, object manipulation, and attribute mapping [14, 55]. While point cloud learning for anatomical SSM is largely unexplored, point networks have been developed for establishing shape correspondence. This task has been approached in a supervised manner through point cloud registration [25, 21, 34, 37], via unified embeddings of multiple shape representations [48], and via part labels [13]. Recently, unsupervised methods have been developed which formulate shape correspondence from either a *pairwise* or class-level, *global* standpoint.

The pairwise approaches seek to find a point-to-point mapping from a source shape to a target shape. Many mesh-based methods have been established for this task using functional maps [51, 29, 33, 30], which require connectivity. The idea of functional maps has been extended to point clouds, using spectral matching to define correspondence [44]. Other approaches extract correspondence by learned deformations to a predefined template [28, 26]. Recent methods utilize deep networks to learn a matching permutation between a source and target point cloud [72, 41]. Zeng et al. [72] utilize an encoder-decoder architecture to regress the shape coordinates for permuted reconstruction, while

Lang et al. [41] opt to drop the decoder and use the original point cloud in reconstruction, achieving better performance by leveraging similarity in the learned feature space.

Prior global correspondence approaches have focused on discovering class-specific keypoints (a.k.a. ordered stable interest or structure points) from point clouds [58, 32, 39]. These are typically very sparse semantically consistent points. More relevant works have sought to define dense keypoints or structure points - a task highly related to correspondence-based SSM. Chen et al. [22] developed a network that reconstructs point clouds in a consistent way using an input subset and learned features, providing a correspondence model with meaningful principal component analysis embedding. Liu and Liu [42] leveraged part features learned by a branched AE[23], to establish intra-class correspondence. However, this approach requires additional knowledge of the shape surface to compute the occupancy for training the implicit function. Cheng et al. [24] developed a self-supervised canonical point autoencoder that utilizes mapping to a canonical primitive (sphere) to establish order across classes of shapes. These methods establish correspondence but are prone to overfitting given a limited training budget and are not robust to noise, missingness, and sparsity in the input point cloud.

3 Methods

3.1 Proposed Approach: Point2SSM

Let $\mathcal{S}$ denote a point cloud of P unordered points representing an anatomical shape: $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_P\}$ with $\mathbf{s}_i \in \mathbb{R}^3$. Given a subset of N points in $\mathcal{S}$, the goal of Point2SSM is to predict a set of M correspondence points, denoted $\mathcal{C}$. Point2SSM learns correspondence in a self-supervised manner by estimating a sparse set of points $\mathcal{C}$ that best reconstructs the full point clouds $\mathcal{S}$. It is comprised of a DGCNN [66] encoder and a transformer-like attention module, as shown in Figure 2. The encoder learns an L-dimensional feature vector for each point, incorporating local neighborhood information to learn global shape properties via edge convolution. The attention module predicts a correspondence map from the feature representation via self-attention. The output correspondence points are then computed via matrix multiplication between the correspondence map and the input point cloud. Because the output correspondence points are a convex combination of the input points, they are located within the convex hull of the input points. This effect combined with the attention module increase surface sampling accuracy.

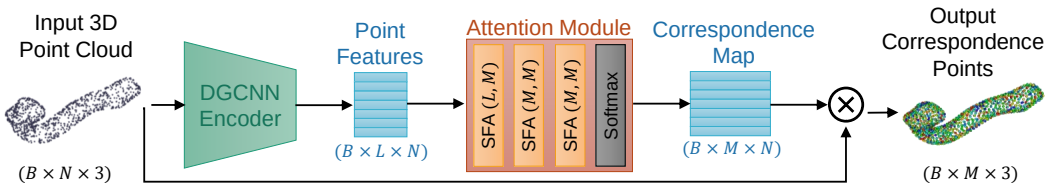

Figure 2: Point2SSM Architecture

The attention module must capture the structural characteristic of the shape from the features to learn correspondence. Hence, we leverage the Self-Feature Augment (SFA) blocks introduced in Wang et al. [64]. The SFA block is a multi-head, self-attention mechanism drawn on the core idea of transformers [63], but without the encoder-decoder structure. The SFA block integrates the information from different point features and establishes the spatial relationship among points by introducing self-attention. In this manner, SFA captures global information and reveals detailed shape geometry. The attention module generates semantically consistent probability maps for all shapes in a cohort, leading to the consistent ordering of the output points and promoting correspondence. A visualization of these attention maps is provided in the supplementary material.

The Point2SSM loss function is comprised of two terms - one which encourages $\mathcal{C}$ to reconstruct $\mathcal{S}$ and one which regularizes $\mathcal{C}$, encouraging better correspondence. Chamfer distance (CD) is used to measure the difference between the point clouds in a permutation-invariant way:

$$\mathrm{CD}(\mathcal{C}, \mathcal{S}) = \frac{1}{|\mathcal{C}|} \sum_{\mathbf{c} \in \mathcal{C}} \min_{\mathbf{s} \in \mathcal{S}} ||\mathbf{c} - \mathbf{s}||_2^2 + \frac{1}{|\mathcal{S}|} \sum_{\mathbf{s} \in \mathcal{S}} \min_{c \in \mathcal{C}} ||\mathbf{s} - \mathbf{c}||_2^2 \tag{1}$$

A pairwise mapping error (ME), originally proposed in [41], is adapted to provide an output regularization loss term. The ME between two output point clouds $\mathcal{C}'$ and $\mathcal{C}''$ is defined as:

$$\text{ME}(\mathcal{C}', \mathcal{C}'') = \frac{1}{M * K} \sum_{i=1}^{M} \sum_{j \in \mathcal{N}_{\mathcal{C}'}(\mathbf{c}'_i)} v'_{ij} \left\| \mathbf{c}''_i - \mathbf{c}''_j \right\|_2^2 \tag{2}$$

where $\mathcal{N}_{\mathcal{C}'}(\mathbf{c}'_i)$ are the K indices of the Euclidean neighbors of point $\mathbf{c}'^1_i$ in $\mathcal{C}'$. Here $v'_{il} = e^{-\left\|\mathbf{c}'_i - \mathbf{c}'_l\right\|_2^2}$ weights the loss elements according to the proximity of the neighbor points. ME loss encourages the point neighborhoods in $\mathcal{C}'$ to be similar to $\mathcal{C}''$, promoting correspondence. We compute ME between all pairs of output point sets $\mathcal{C}$ in a minibatch of size B. The Point2SSM loss is thus defined as:

$$\mathcal{L} = \frac{1}{B} \sum_{i=1}^{B} \text{CD}(\mathcal{C}^i, \mathcal{S}^i) + \alpha \left(\frac{1}{(B-1)^2} \sum_{i=1}^{B} \sum_{j=1, j \neq i}^{B} \text{ME}(\mathcal{C}^i, \mathcal{C}^j) + \text{ME}(\mathcal{C}^j, \mathcal{C}^i) \right) \tag{3}$$

where α is a hyperparameter that controls the effect of the regularization. Ablation experiments that demonstrate the impact of the encoder architecture, attention module architecture, and ME loss are provided in the supplementary material.

3.2 Comparison Models

We benchmark the following SOTA methods on the anatomical SSM tasks and compare their performance to that of Point2SSM.

PSM [20] denotes Particle-based Shape Modeling, the SOTA optimization-based technique for generating correspondence points from complete shape surface representations [35]. We use the mesh-based implementation of PSM available in the open-source toolkit, ShapeWorks [20]. We include this method to provide context regarding the expected modes of variation compactness of SSM for a given anatomy.

PN-AE [1] is the autoencoder formulated by Achlioptas et al. [1] with PointNet [52] encoder. The combination of bottleneck and MLP decoder results in the consistent output ordering $\mathcal{C}$ [4].

DG-AE [66] is a variant of PN-AE [1] where the PointNet [52] encoder is replaced with a DGCNN [66] encoder. It is included to assist in comparing the AE framework with Point2SSM and DPC [41].

CPAE [24] is the Canonical Point Autoencoder that maps points to a sphere template in the bottleneck and then reconstructs the points in an ordered fashion.

ISR [22] denotes the method for learning Intrinsic Structural Representation (ISR) points proposed by Chen et al. [22]. This network utilizes a PointNet++ [53] encoder and a Point Integration Module that maps the features to a probability map, which is multiplied by a point subset to provide output.

DPC [41] is the Deep Point Correspondence proposed by Lang et al. [41]. This is a pairwise-correspondence method that takes the source and target point clouds as input and outputs the source reordered to match the target. The architecture comprises a DGCNN [66] encoder and cross and self-construction modules that utilize latent similarity. To adapt DPC [41] to be a global correspondence method, the same target point cloud is used for every shape in inference.

3.3 Evaluation Metrics

An ideal SSM accurately samples the shape surface via uniformly distributed points that are constrained to lie on the surface. Simultaneously, it captures anatomically relevant mappings between shapes by establishing consistent, invariant points (i.e., correspondences) across diverse populations with varying forms. Consequently, the evaluation of an SSM should encompass both aspects: the accuracy of surface sampling and the efficacy of the extracted shape statistics.

Three metrics are used to define the *point surface sampling accuracy*: $\text{CD}(\mathcal{C}, \mathcal{S})$, Earth Movers Distance (EMD), and point-to-face distance (P2F). EMD requires both point clouds to have the same number of points; thus, a subset of points in $\mathcal{S}$ is selected using farthest point sampling [53]. In this way, EMD captures whether the predicted correspondence points cover the entire shape. P2F distance is calculated as the distance of each point in $\mathcal{C}$ to the closest face of the ground truth mesh, indicating how well points are constrained to the surface.

In an ideal SSM, point locations and neighborhoods are preserved across all shapes in the cohort, leading to a meaningful model with compact modes of shape variation. In SSM analysis, correspondence points are averaged to generate a representative mean shape and principal component analysis (PCA) is performed to compute the significant modes of variation. These modes can be used in medical hypothesis testing and visualized by deforming the mean shape along each basis of the linear subspace [19]. Three statistical metrics are used to evaluate *SSM correspondence accuracy*: compactness, generalization, and specificity [47]. A compact SSM represents the training data distribution using the minimum number of parameters; thus, compactness is quantified as the number of PCA modes required to capture 95% of the variation in the correspondence points. Moreover, a good SSM should generalize well from training examples to unseen examples. The generalization metric quantifies how well the SSM generalizes from training examples to unseen examples via the reconstruction error ($L2$) between held-out correspondence points and those reconstructed via the training SSM. The specificity metric measures the degree to which the SSM generates valid instances of the shape class presented in the training set. It is computed as the average distance between correspondences sampled from the training SSM and the closest existing training correspondences. The equations of these metrics are available in [47] and the supplementary material.

Both categories of metrics must be used to evaluate the accuracy of SSM, as one does not imply the other. For example, a network that yields identical output given any input will perform very well on statistical metrics but poorly on sampling metrics. Conversely, a network that perfectly reconstructs the input in an unordered fashion will succeed at point sampling metrics but fail at statistical metrics.

4 Experiments and Analysis

We utilize three organ mesh datasets of various sample sizes to benchmark the performance of Point2SSM and the comparison methods: spleen [56], pancreas [56], and left atrium of the heart. A visualization of the organ cohorts is provided in the supplementary materials, demonstrating the large morphological variation. We elect to acquire point clouds from meshes to enable benchmarking against PSM [20], which requires mesh connectivity. The size of the spleen dataset is representative of a typical SSM scenario with 40 total shapes, the pancreas dataset is medium-sized with 272 total shapes, and the left atrium dataset provides an example of an unusually large and higly-variable medical dataset with 1096 total shapes. Note that the non-anatomical shape datasets previously used to benchmark comparison methods are much larger (i.e., tens of thousands).

As preprocessing, we align the meshes to factor out global geometric information via iterative closest points [9] (utilizing the ShapeWorks [20] toolkit). The aligned, unordered mesh vertices serve as ground truth complete point clouds, $\mathcal{S}$. In all experiments, we set $N = 1024$, $L = 128$, $M = 1024$, and batch size $B = 8$, unless otherwise specified. During model training, input point clouds are generated by randomly selecting N points from $\mathcal{S}$ each iteration and uniformly scaling them to be between -1 and 1. The datasets are randomly split into a training, validation, and test set using an 80%, 10%, 10% split. Adam optimization with a constant learning rate of 0.0001 is used for all models. Model training is run until convergence via validation assessment. Specifically, a model is considered to have converged if the validation CD has not improved in 100 epochs. Models resulting from the epoch with the best validation CD are used in the evaluation. A single 4x TITAN V GPU was used to train all models. For Point2SSM loss (Eq. 3), α is set to 0.1 based on tuning using the validation set. For comparison models, the originally proposed loss is used with the reported tuned hyperparameter values. All code, parameters, and trained models as well as the spleen data are provided in the supplementary material, as well as a comparison of model memory footprint.

4.1 Results

Figure 3 provides an overview of the results on all datasets. Point2SSM significantly outperforms the existing point methods with regard to the surface sampling metrics: CD, EMD, and P2F distance. This is further illustrated in Figure 4, which provides a point-level visualization of the test example with median P2F distance output by each model. Figure 3 additionally demonstrates that Point2SSM performs comparably with respect to statistical metrics to PSM [20] and provides the best compactness on the left atrium dataset. Out of the point-based methods, Point2SSM achieves the best compactness on all datasets - with the exception of the spleen DG-AE [66], which greatly suffers in terms of point sampling metrics. The AE methods aggregate features into a global ($L \times 1$) feature in the

bottleneck. This restriction enforces a shape prior resulting in compactness, but it greatly limits model expressivity hindering accurate surface sampling. The CPAE output reconstructs the point cloud to some degree but does not provide correspondence. The resulting CPAE SSM is not compact or interpretable, likely because learning a canonical mapping is too complex given a small training set. ISR [22] and DPC [41] performs reasonably well on the surface sampling metrics but do not provide as compact, generalizable, or specific of a model as Point2SSM. Point2SSM combines the strengths of the comparison methods, providing the best overall accuracy.

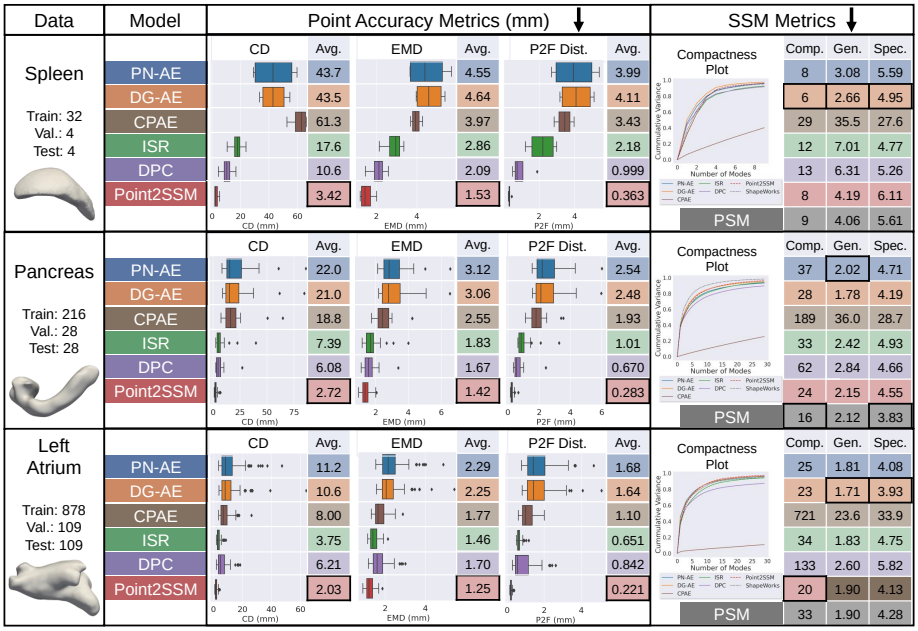

Figure 3: Accuracy metrics are reported for all models on all three datasets; best values are outlined. Boxplots show the distribution across the test set with the average value reported to the right. Compactness plots show cumulative population variation captured by PCA modes, where a larger area under the curve indicates a more compact model.

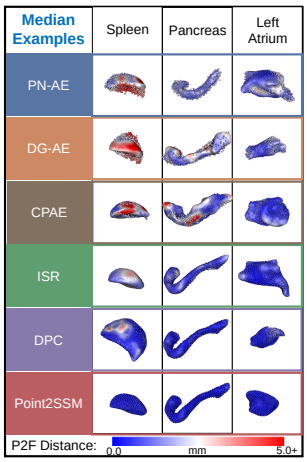

Figure 4: The test examples with median P2F distance output from each model are shown over ground truth meshes. P2F distance is displayed via a color map.

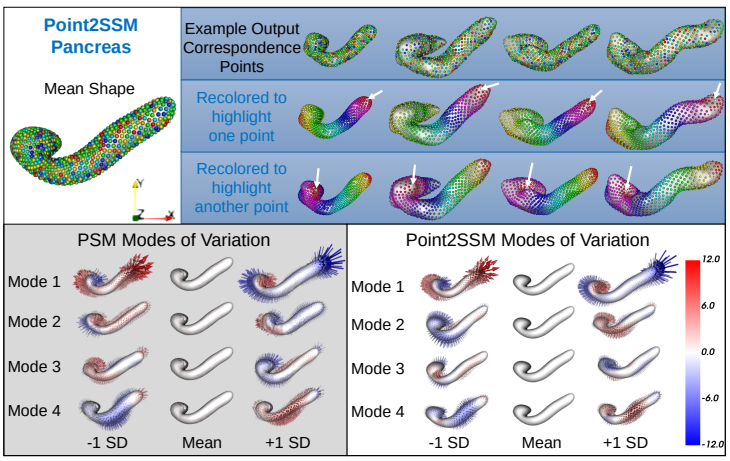

Figure 5: The pancreas SSM from Point2SSM is displayed. Point color denotes correspondence. Recoloring according to the distance to a selected point is provided for further illustration. The first four modes of variation are shown for the PSM [20] and Point2SSM model at ± 1 standard deviation from the mean. The heatmap and vector arrows display the distance to the mean.

Figure 5 displays the SSM resulting from Point2SSM on the pancreas dataset. The mean shape is plausible and interpretable and the individual point locations are geometrically consistent across shapes. The primary modes of variation in the pancreas are semantically similar to those resulting from the PSM [20] method, suggesting they accurately capture population statistics and could be similarly used in downstream tasks. Figure 6 displays the first two modes of variation captured by Point2SSM and all comparison models on the spleen dataset. The CPAE [24] results are excluded from this visualization because the resulting SSM is uninterpretable. The meshes in Figure 6 are constructed using the output correspondence points. Mesh artifacts or implausible morphologies are an indication of output miscorrespondence. Point2SSM is the only point-based method that provides similar, if not more smooth and interpretable, modes to the PSM [20] method. Not only does Point2SSM allow for much faster inference than PSM [20], but it is also more scalable. While PSM [20] accuracy is not largely affected by the shape cohort size, the optimization process is much slower given a large cohort. Fitting the ShapeWorks PSM [20] model to the large left atrium cohort required running optimization in an incremental fashion over the course of four days. In contrast, the Point2SSM left atrium model required under eight hours to train on a GPU.

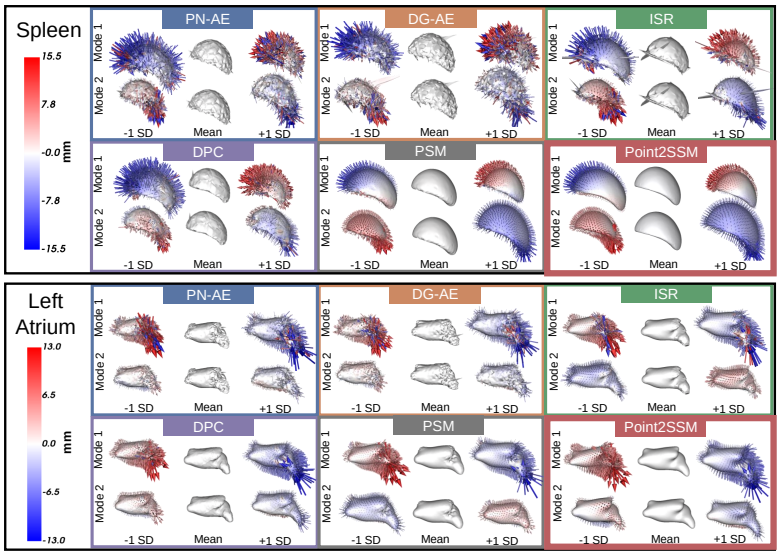

Figure 6: Primary and secondary modes of variation captured by spleen and left atrium SSMs output by each model. Point2SSM provides the smoothest and most plausible mean and modes of variation.

4.2 Robustness Evaluation

We perform experiments to demonstrate model robustness using the medium-sized pancreas dataset. These results are plotted in Figure 7. In all robustness experiments, the training and testing point clouds are corrupted in the same manner.

Robustness against input noise. To analyze the effect of noise on performance, we apply random Gaussian noise to input point clouds at various levels: 0.25, 0.5, 1, and 2 mm standard deviation. Point2SSM and the comparison models are not significantly impacted by input noise and Point2SSM achieves the best accuracy at all noise levels.

Robustness against partial input. To analyze the impact of partial input with missing regions, we remove continuous regions at random locations of various sizes (5%, 10%, and 20% of the total points) from the input. The CD and EMD results capture how well the output fills in the missing regions to provide full coverage. The autoencoder architectures (PN-AE [1] and DG-AE [66]) are the least impacted by partial input, which is logical given this is the architecture used in point completion networks. Not only does Point2SSM perform similarly or better than all models regarding the distance metrics, but it also preserves compactness better than DPC [41] and ISR [22] with increasingly partial input.

Robustness against sparse input. To test the impact of sparse input, we train the models with input size N set to 128, 256, 512, and 1024, keeping the latent size L fixed at 128 and output size M fixed

at 1024. This experiment benchmarks the network's ability to both upsample and provide SSM. The CPAE [24] model and DPC [41] model require $N = M$ and are thus excluded from this experiment. Point2SSM achieves the best overall accuracy given sparse input; however, there is a decline in performance given very sparse input ($N = 128$).

Impact of training sample size. The final experiment benchmarks the effect of training size on model performance. We consistently define random subsets of the 216 training point clouds of size 100, 50, 25, 12, and 6. DPC [41] and Point2SSM perform the best on the distance metrics, demonstrating impressive robustness (generalizing to the test set with only 6 training examples). Compactness results are excluded, as compactness depends on the variation in the training cohort.

The robustness experiments demonstrate that Point2SSM combines the strengths of all existing models. It performs as well as PN-AE [1] and DG-AE [66] given large missing regions and as well as DPC [41] given a small training cohort. These comparisons are compiled in Figure 1.

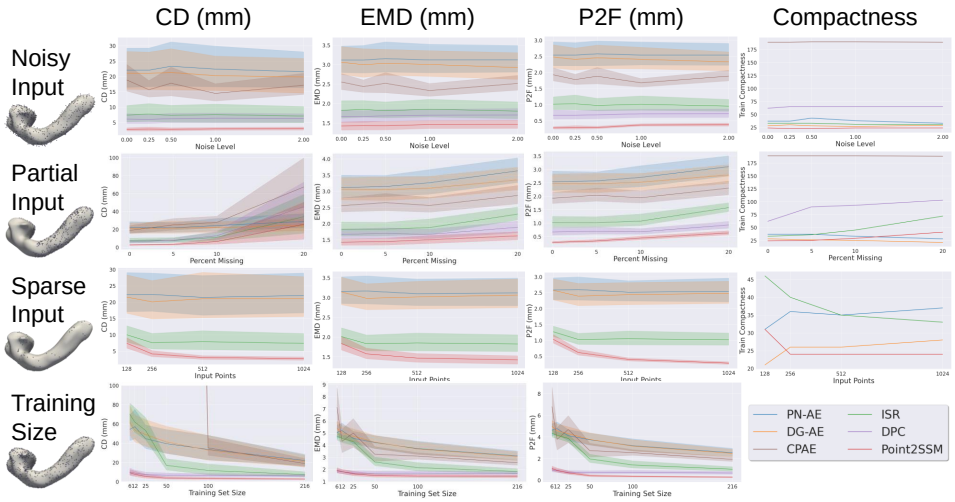

Figure 7: Results of robustness experiments on the pancreas test set. Example input point clouds are displayed over meshes. Distance metrics are shown with standard deviation error bands, and compactness is calculated at 95% variability.

4.3 Limitations and Future Directions

Deterministic deep learning frameworks like Point2SSM have some limitations. They produce overconfident estimates that could potentially misrepresent shapes, especially when dealing with noisy, partial, and sparse inputs. Incorporating uncertainty quantification would enhance the reliability of deploying Point2SSM in sensitive clinical decision-making scenarios. Additionally, Point2SSM requires roughly aligned input point clouds and is designed to produce SSM of a single anatomy (i.e., bone or organ). CPAE [24] and other methods [42] have demonstrated the ability to predict to inter-class or topology varying correspondence. Broadening the scope of Point2SSM to handle misalignment and multiple anatomies with a single network would increase applicability.

5 Conclusion

We introduced Point2SSM, the first deep learning method designed to produce 3D anatomical statistical shape models (SSM) directly from point clouds in an unsupervised manner. Point2SSM overcomes the prohibitive limitations of optimization-based SSM generation methods, offering an unbiased data-driven solution from point clouds with fast inference. It outperforms SOTA point cloud networks in surface sampling and correspondence accuracy. Moreover, Point2SSM is robust in challenging clinical modeling scenarios, handling limited data and noisy, incomplete, and sparse shape representations. Our proposed method increases the feasibility of SSM generation and expands its potential applications, potentially accelerating its adoption as a widely used tool in clinical research.

Acknowledgments and Disclosure of Funding

This work was supported by the National Institutes of Health under grant numbers NIBIB-U24EB029011, NIAMS-R01AR076120, and NHLBI-R01HL135568. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors would like to thank the University of Utah Division of Cardiovascular Medicine for providing left atrium MRI scans and segmentations from the Atrial Fibrillation projects.

Supplementary Material

A SSM Evalaluation Metrics

As is standard [47], we utilize three statistical metrics to evaluate correspondence accuracy: compactness, generalization, and specificity. A compact SSM represents the training data distribution using the minimum number of parameters. We quantify **compactness** as the number of PCA modes required to capture 95% of the total variation in the output training cohort correspondence points, where fewer modes indicate a more compact model. The compactness plots in Figure **??** show the cumulative explained variance as the number of modes increases to provide a full picture.

A good SSM should generalize well from training examples to unseen examples and be able to describe any valid instance of the shape class. Given an unseen test cohort of correspondence point sets, denoted $\mathcal{D}_{test}$, the **generalization** metric is defined as:

$$\text{Gen.} = \frac{1}{|\mathcal{D}_{test}|} \sum_{\mathcal{C} \in \mathcal{D}_{test}} ||\mathcal{C} - \hat{\mathcal{C}}||_2^2 \tag{4}$$

where $\hat{\mathcal{C}}$ is the point set reconstructed via the training cohort PCA eigenvalues and vectors that preserve 95% variability. A smaller average reconstruction error indicates that the SSM generalizes well to the unseen test set.

Finally, effective SSM is specific, generating only valid instances of the shape class presented in the training set. This metric is quantified by generating a set of new sample correspondence points, denoted $\mathcal{D}_{sample}$, from the SSM generated on the training cohort, denoted $\mathcal{D}_{train}$. The **specificity** metric is quantified as:

$$\text{Spec.} = \frac{1}{|\mathbf{D}_{\text{sample}}|} \sum_{\mathcal{C}' \in \mathbf{D}_{\text{sample}}} \min_{\mathcal{C} \in \mathcal{D}_{train}} ||\mathcal{C}' - \mathcal{C}||_2^2 \tag{5}$$

The average distance between correspondence points sampled from the training SSM and the closest existing training correspondence points provides the specificity metric. A small distance suggests the samples match the training distribution well, indicating the SSM is specific.

B Point2SSM Ablation Experiment

We perform an ablation experiment on the pancreas dataset to analyze the impact of each aspect of the Point2SSM model. To illustrate the impact of the DGCNN [66] encoder, we design a PointSSM variant with the PointNet [52] decoder used in PN-AE [1]. The Point2SSM attention module (denoted ATTN) is comprised of attention-based SFA[64] blocks. To analyze this impact, we design a variation of Point2SSM where the attention module is replaced with an MLP-based architecture. For the MLP-based architecture, we elect to use the Point Integration Module proposed in ISR [22], with three 1D convolution blocks. Finally, to test the impact of the ME loss in Equation 3, we test without it by setting $\alpha = 0$.

The results of this ablation are shown in Table 1, with the full proposed Point2SSM in the final row. The DGCNN-encoder and ATTN attention module both provide surface sampling and correspondence accuracy improvements. The addition of the ME loss ($\alpha = 0.1$) improves the correspondence accuracy without reducing the surface sampling accuracy.

Table 1: Point2SSM ablation experiment on the pancreas dataset. Average test set values are reported for point accuracy distance metrics in mm. SSM metrics are calculated at 95% variability.

Pancreas Point2SSM Ablation			Point Accuracy Metrics (mm) ↓			SSM Metrics ↓		
Encoder	Attention Module	α	CD	EMD	P2F	Comp.	Gen.	Spec.
PointNet	MLP	0	7.35	1.78	0.833	52	2.96	4.48
PointNet	ATTN	0	3.00	1.44	0.306	27	2.24	4.67
DGCNN	MLP	0	3.40	1.46	0.378	31	2.2	4.52
DGCNN	ATTN	0	2.87	1.43	0.283	26	2.32	4.80
DGCNN	ATTN	0.1	2.72	1.42	0.283	24	2.15	4.55

C Model Hyper-parameters

Model hyper-parameters are provided in the following tables and in the configuration files proved with the code. Table 2 displays the hyper-parameters that are consistent across all models. Note in the sparsity robustness experiments, the value of N varies. Tables 3 and 4 display the hyper-parameters specific to the CPAE [24] and DPC [41] models. These values match those originally tuned/reported. Finally, Table 5 displays the hyper-parameters specific to our Point2SSM model. The number of neighbors used in the ME loss is set to 10, as it is for DPC [41]. The value of α is determined via tuning based on the validation set performance.

Table 2: Hyper-parameters shared by all models.

Shared Hyper-parameters		
Parameter	Description	Value
N	Number of input points	1024
L	Number of per-point features output by encoder	128
M	Number of output points	1024
B	Batch size	8
LR	Learning rate	0.0001
ES	Early stopping patience (epochs)	100
β_1	Adam optimization first coefficient	0.9
β_2	Adam optimization second coefficients	0.999

Table 3: Hyper-parameters specific to the CPAE [24] model.

Additional CPAE [24] Hyper Parameters		
Parameter	Description	Value
λ_{MSE}	MSE loss weight	1000
λ_{CD}	CD loss weight	10
λ_{EMD}	EMD loss weight	1
λ_{cc}	Cross-construction loss weight	10
λ_{unfold}	Unfolding loss weight	10
e	Adaptive loss epoch	100

Table 4: Hyper-parameters specific to the DPC [41] model.

Additional DPC [41] Hyper-parameters		
Parameter	Description	Value
K	Neighborhood size for loss calculation	10
γ	Mapping loss neighbor sensitivity	8
λ_{cc}	Cross-construction loss weight	1
λ_{sc}	Self-construction loss weight	10
λ_m	Mapping loss weight	1

Table 5: Hyper-parameters specific to our Point2SSM model.

Additional Point2SSM Hyper-parameters		
Parameter	Description	Value
K	Neighborhood size for ME loss	10
α	ME loss weight	0.1

D Model Memory Comparison

Table 6 shows a comparison of the memory footprint of each model.

Table 6: Model memory footprint comparison. Size is reported in MB.

Model	Total Params	Forward/backward pass size	Params size	Total Size
PN-AE [1]	3,832,576	11.04	14.62	25.68
DG-AE [66]	4,702,336	609.54	17.94	627.5
CPAE [24]	156,652	19.58	0.60	56.18
ISR [22]	1,962,208	4.00	7.49	11.51
DPC [41]	962,176	609.50	3.67	613.21
Point2SSM	22,098,560	633.69	84.3	718.01

E Shape Dataset Visualization

Figure 8 displays example shapes from each organ dataset (spleen, pancreas, and left atrium) from multiple views. This illustrates the large amount of variation in these shape cohorts.

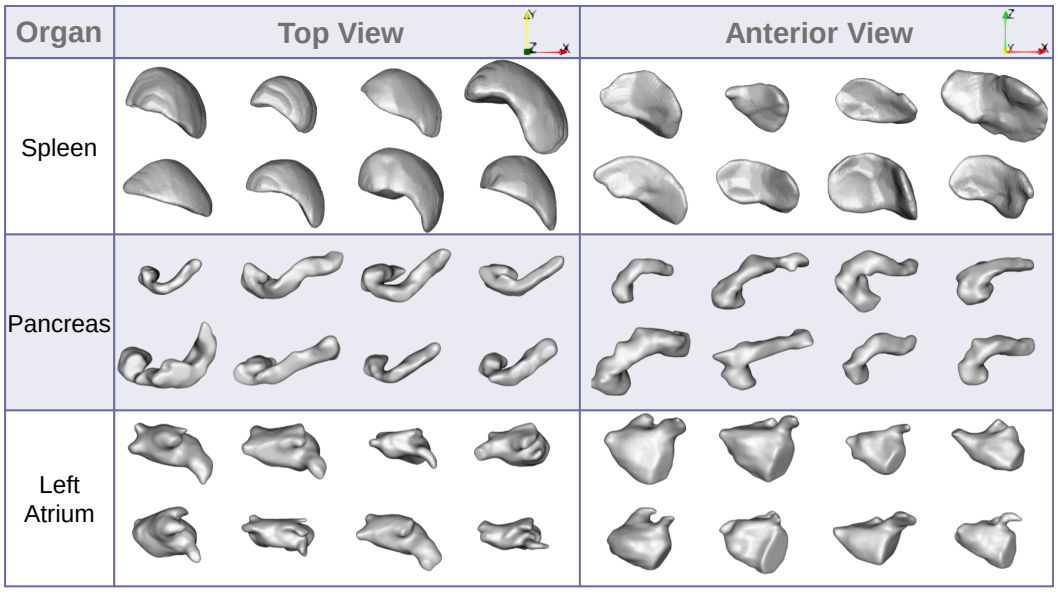

Figure 8: Example shapes are shown from each of the datasets from the top and anterior view.

F Attention Map Visualization

Figure 9 illustrates the correspondence map weights learned by the Point2SSM attention module on the pancreas dataset. Output correspondence points are a weighted combination of the input points, where the learned correspondence map defines the weights. Figure 9 highlights two output correspondence points across shapes. The attention maps show the weights on the input points (via color map) that generated the selected output point. This illustrates which input points were most

important to generating a given output point. Note the maps highlight similar anatomical regions across samples for a given output corresponding point.

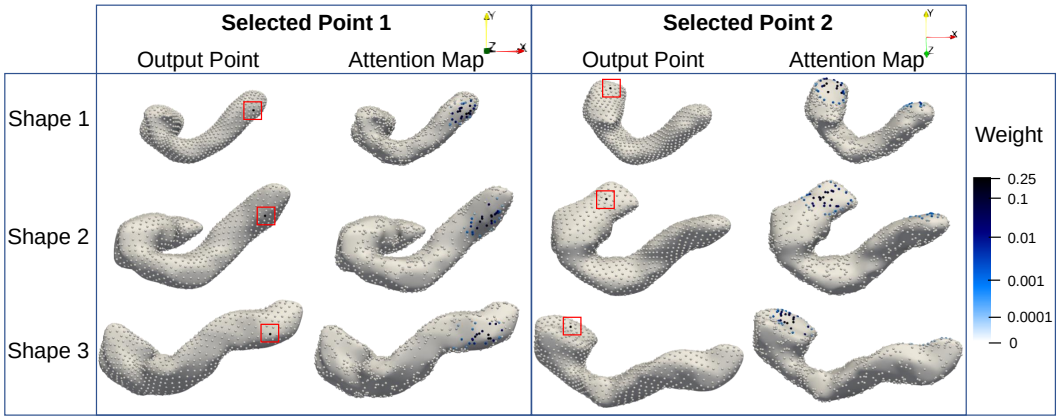

Figure 9: Two output points (highlighted in red boxes) across three pancreas shapes are shown with the corresponding weights (blue log scale color map) on the input point clouds.

References

[1] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. Guibas. Learning representations and generative models for 3d point clouds. In *International conference on machine learning*, pages 40–49. PMLR, 2018.

[2] J. Adams and S. Elhabian. From images to probabilistic anatomical shapes: A deep variational bottleneck approach. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*, pages 474–484. Springer, 2022.

[3] J. Adams and S. Elhabian. Fully bayesian vib-deepssm. *arXiv preprint arXiv:2305.05797*, 2023.

[4] J. Adams and S. Elhabian. Can point cloud networks learn statistical shape models of anatomies? *arXiv preprint arXiv:2305.05610*, 2023.

[5] J. Adams, R. Bhalodia, and S. Elhabian. Uncertain-deepssm: From images to probabilistic shape models. *Shape in Medical Imaging : International Workshop, ShapeMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings*, 12474:57–72, 2020.

[6] A. Akagic, S. Krivić, H. Dizdar, and J. Velagić. Computer vision with 3d point cloud data: Methods, datasets and challenges. In *2022 XXVIII International Conference on Information, Communication and Automation Technologies (ICAT)*, pages 1–8. IEEE, 2022.

[7] P. R. Atkins, S. Y. Elhabian, P. Agrawal, M. D. Harris, R. T. Whitaker, J. A. Weiss, C. L. Peters, and A. E. Anderson. 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.

[8] L. Bastian, A. Bauman, E. Hoppe, V. Bürgin, H. Y. Kim, M. Saleh, B. Busam, and N. Navab. S3m: Scalable statistical shape modeling through unsupervised correspondences. *arXiv preprint arXiv:2304.07515*, 2023.

[9] P. J. Besl and N. D. McKay. Method for registration of 3-d shapes. In *Sensor fusion IV: control paradigms and data structures*, volume 1611, pages 586–606. Spie, 1992.

[10] R. Bhalodia, S. Y. Elhabian, L. Kavan, and R. T. Whitaker. Deepssm: A deep learning framework for statistical shape modeling from raw images. In *Shape In Medical Imaging at MICCAI*, volume 11167 of *Lecture Notes in Computer Science*, pages 244–257. Springer, 2018.

[11] R. Bhalodia, L. A. Dvoracek, A. M. Ayyash, L. Kavan, R. Whitaker, and J. A. Goldstein. Quantifying the severity of metopic craniosynostosis: A pilot study application of machine learning in craniofacial surgery. *Journal of Craniofacial Surgery*, 2020.

[12] R. Bhalodia, S. Elhabian, J. Adams, W. Tao, L. Kavan, and R. Whitaker. Deepssm: A blueprint for image-to-shape deep learning models. *arXiv preprint arXiv:2110.07152*, 2021.

[13] B. L. Bhatnagar, C. Sminchisescu, C. Theobalt, and G. Pons-Moll. Combining implicit function learning and parametric models for 3d human reconstruction. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 311–329. Springer, 2020.

[14] S. Biasotti, A. Cerri, A. Bronstein, and M. Bronstein. Recent trends, applications, and perspectives in 3d shape similarity assessment. In *Computer graphics forum*, volume 35, pages 87–119. Wiley Online Library, 2016.

[15] J. E. Bischoff, Y. Dai, C. Goodlett, B. Davis, and M. Bandi. Incorporating population-level variability in orthopedic biomechanical analysis: a review. *Journal of biomechanical engineering*, 136(2):021004, 2014.

[16] J. L. Bruse, K. McLeod, G. Biglino, H. N. Ntsinjana, C. Capelli, T.-Y. Hsia, M. Sermesant, X. Pennec, A. M. Taylor, S. Schievano, et al. A statistical shape modelling framework to extract 3d shape biomarkers from medical imaging data: assessing arch morphology of repaired coarctation of the aorta. *BMC medical imaging*, 16:1–19, 2016.

[17] N. Carriere, P. Besson, K. Dujardin, A. Duhamel, L. Defebvre, C. Delmaire, and D. Devos. Apathy in parkinson's disease is associated with nucleus accumbens atrophy: a magnetic resonance imaging shape analysis. *Movement disorders*, 29(7):897–903, 2014.

[18] J. Cates, P. T. Fletcher, M. Styner, M. Shenton, and R. Whitaker. Shape modeling and analysis with entropy-based particle systems. In *IPMI*, pages 333–345. Springer, 2007.

[19] J. Cates, E. Bieging, A. Morris, G. Gardner, N. Akoum, E. Kholmovski, N. Marrouche, C. McGann, and R. S. MacLeod. Computational shape models characterize shape change of the left atrium in atrial fibrillation. *Clinical Medicine Insights: Cardiology*, 8:CMC–S15710, 2014.

[20] J. Cates, S. Elhabian, and R. Whitaker. Shapeworks: particle-based shape correspondence and visualization software. In *Statistical shape and deformation analysis*, pages 257–298. Elsevier, 2017.

[21] M. Chen, Q. Zou, C. Wang, and L. Liu. Edgenet: Deep metric learning for 3d shapes. *Computer Aided Geometric Design*, 72:19–33, 2019.

[22] N. Chen, L. Liu, Z. Cui, R. Chen, D. Ceylan, C. Tu, and W. Wang. Unsupervised learning of intrinsic structural representation points. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9121–9130, 2020.

[23] Z. Chen, K. Yin, M. Fisher, S. Chaudhuri, and H. Zhang. Bae-net: Branched autoencoder for shape co-segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8490–8499, 2019.

[24] A.-C. Cheng, X. Li, M. Sun, M.-H. Yang, and S. Liu. Learning 3d dense correspondence via canonical point autoencoder. *Advances in Neural Information Processing Systems*, 34:6608–6620, 2021.

[25] C. Choy, J. Park, and V. Koltun. Fully convolutional geometric features. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8958–8966, 2019.

[26] L. Cosmo, E. Rodola, J. Masci, A. Torsello, and M. M. Bronstein. Matching deformable objects in clutter. In *2016 Fourth international conference on 3D vision (3DV)*, pages 1–10. IEEE, 2016.

[27] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor. A minimum description length approach to statistical shape modeling. *IEEE transactions on medical imaging*, 21(5):525–537, 2002.

[28] T. Deprelle, T. Groueix, M. Fisher, V. Kim, B. Russell, and M. Aubry. Learning elementary structures for 3d shape generation and matching. *Advances in Neural Information Processing Systems*, 32, 2019.

[29] N. Donati, A. Sharma, and M. Ovsjanikov. Deep geometric functional maps: Robust feature learning for shape correspondence. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8592–8601, 2020.

[30] M. Eisenberger, A. Toker, L. Leal-Taixé, and D. Cremers. Deep shells: Unsupervised shape correspondence with optimal transport. *Advances in Neural information processing systems*, 33:10491–10502, 2020.

[31] B. Fei, W. Yang, W.-M. Chen, Z. Li, Y. Li, T. Ma, X. Hu, and L. Ma. Comprehensive review of deep learning-based 3d point cloud completion processing and analysis. *IEEE Transactions on Intelligent Transportation Systems*, 2022.

[32] C. Fernandez-Labrador, A. Chhatkuli, D. P. Paudel, J. J. Guerrero, C. Demonceaux, and L. V. Gool. Unsupervised learning of category-specific symmetric 3d keypoints from point sets. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 546–563. Springer, 2020.

[33] D. Ginzburg and D. Raviv. Cyclic functional mapping: Self-supervised correspondence between non-isometric deformable shapes. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pages 36–52. Springer, 2020.

[34] Z. Gojcic, C. Zhou, J. D. Wegner, and A. Wieser. The perfect match: 3d point cloud matching with smoothed densities. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5545–5554, 2019.

[35] A. Goparaju, K. Iyer, A. Bone, N. Hu, H. B. Henninger, A. E. Anderson, S. Durrleman, M. Jacxsens, A. Morris, I. Csecs, et al. Benchmarking off-the-shelf statistical shape modeling tools in clinical applications. *Medical Image Analysis*, 76:102271, 2022.

[36] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard. Octomap: An efficient probabilistic 3d mapping framework based on octrees. *Autonomous robots*, 34:189–206, 2013.

[37] H. Huang, E. Kalogerakis, S. Chaudhuri, D. Ceylan, V. G. Kim, and E. Yumer. Learning local shape descriptors from part correspondences with multiview convolutional networks. *ACM Transactions on Graphics (TOG)*, 37:1–14, 2017.

[38] K. Iyer and S. Elhabian. Mesh2ssm: From surface meshes to statistical shape models of anatomy. *arXiv preprint arXiv:2305.07805*, 2023.

[39] T. Jakab, R. Tucker, A. Makadia, J. Wu, N. Snavely, and A. Kanazawa. Keypointdeformer: Unsupervised 3d keypoint discovery for shape control. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12783–12792, 2021.

[40] D. G. Kendall. The diffusion of shape. *Advances in applied probability*, 9(3):428–430, 1977.

[41] I. Lang, D. Ginzburg, S. Avidan, and D. Raviv. Dpc: Unsupervised deep point correspondence via cross and self construction. In *2021 International Conference on 3D Vision (3DV)*, pages 1442–1451. IEEE, 2021.

[42] F. Liu and X. Liu. Learning implicit functions for topology-varying dense 3d shape correspondence. *Advances in Neural Information Processing Systems*, 33:4823–4834, 2020.

[43] D. Lüdke, T. Amiranashvili, F. Ambellan, I. Ezhov, B. H. Menze, and S. Zachow. Landmark-free statistical shape modeling via neural flow deformations. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2022: 25th International Conference, Singapore, September 18–22, 2022, Proceedings, Part II*, pages 453–463. Springer, 2022.

[44] R. Marin, M.-J. Rakotosaona, S. Melzi, and M. Ovsjanikov. Correspondence learning via linearly-invariant embedding. *Advances in Neural Information Processing Systems*, 33:1608–1620, 2020.

[45] C. S. Mendoza, N. Safdar, K. Okada, E. Myers, G. F. Rogers, and M. G. Linguraru. Personalized assessment of craniosynostosis via statistical shape modeling. *Medical Image Analysis*, 18(4):635–646, 2014.

[46] C. Merle, W. Waldstein, J. Gregory, S. Goodyear, R. Aspden, P. Aldinger, D. Murray, and H. Gill. How many different types of femora are there in primary hip osteoarthritis? an active shape modeling study. *Journal of Orthopaedic Research*, 32(3):413–422, 2014.

[47] B. C. Munsell, P. Dalal, and S. Wang. Evaluating shape correspondence for statistical shape analysis: A benchmark study. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(11):2023–2039, 2008.

[48] S. Muralikrishnan, V. G. Kim, M. Fisher, and S. Chaudhuri. Shape unicode: A unified shape representation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[49] M. Nießner, M. Zollhöfer, S. Izadi, and M. Stamminger. Real-time 3d reconstruction at scale using voxel hashing. *ACM Transactions on Graphics (ToG)*, 32(6):1–11, 2013.

[50] I. Oguz, J. Cates, M. Datar, B. Paniagua, T. Fletcher, C. Vachet, M. Styner, and R. Whitaker. Entropy-based particle correspondence for shape populations. *International journal of computer assisted radiology and surgery*, 11:1221–1232, 2016.

[51] M. Ovsjanikov, M. Ben-Chen, J. Solomon, A. Butscher, and L. Guibas. Functional maps: a flexible representation of maps between shapes. *ACM Transactions on Graphics (ToG)*, 31(4):1–11, 2012.

[52] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

[53] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. *Advances in neural information processing systems*, 30, 2017.

[54] M. I. Razzak, S. Naz, and A. Zaib. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps: Automation of Decision Making*, pages 323–350, 2018.

[55] Y. Sahillioğlu. Recent advances in shape correspondence. *The Visual Computer*, 36(8):1705–1721, 2020.

[56] A. L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063*, 2019.

[57] M. Styner, I. Oguz, S. Xu, C. Brechbühler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig. Framework for the statistical shape analysis of brain structures using spharm-pdm. *The insight journal*, page 242, 2006.

[58] S. Suwajanakorn, N. Snavely, J. J. Tompson, and M. Norouzi. Discovery of latent 3d keypoints via end-to-end geometric reasoning. *Advances in neural information processing systems*, 31, 2018.

[59] L. H. Timmins, H. Samady, and J. N. Oshinski. Effect of regional analysis methods on assessing the association between wall shear stress and coronary artery disease progression in the clinical setting. In *Biomechanics of Coronary Atherosclerotic Plaque*, pages 203–223. Elsevier, 2021.

[60] K. Tóthová, S. Parisot, M. Lee, E. Puyol-Antón, A. King, M. Pollefeys, and E. Konukoglu. Probabilistic 3d surface reconstruction from sparse mri information. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part I 23*, pages 813–823. Springer, 2020.

[61] P. Treleaven and J. Wells. 3d body scanning and healthcare applications. *Computer*, 40(7):28–34, 2007. doi: 10.1109/MC.2007.225.

[62] J. Ukey and S. Elhabian. Localization-aware deep learning framework for statistical shape modeling directly from images. In *Medical Imaging with Deep Learning*, 2023.

[63] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[64] J. Wang, Y. Cui, D. Guo, J. Li, Q. Liu, and C. Shen. Pointattn: You only need attention for point cloud completion. *arXiv preprint arXiv:2203.08485*, 2022.

[65] L. Wang, Y. Ren, Y. Gao, Z. Tang, K.-C. Chen, J. Li, S. G. Shen, J. Yan, P. K. Lee, B. Chow, et al. Estimating patient-specific and anatomically correct reference model for craniomaxillofacial deformity via sparse representation. *Medical physics*, 42(10):5809–5816, 2015.

[66] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic graph cnn for learning on point clouds. *Acm Transactions On Graphics (tog)*, 38(5):1–12, 2019.

[67] A. Xiao, J. Huang, D. Guan, X. Zhang, S. Lu, and L. Shao. Unsupervised point cloud representation learning with deep neural networks: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[68] X. Yu, L. Tang, Y. Rao, T. Huang, J. Zhou, and J. Lu. Point-bert: Pre-training 3d point cloud transformers with masked point modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19313–19322, 2022.

[69] W. Yuan, T. Khot, D. Held, C. Mertz, and M. Hebert. Pcn: Point completion network. In *2018 international conference on 3D vision (3DV)*, pages 728–737. IEEE, 2018.

[70] S. Zachow. Computational planning in facial surgery. *Facial Plastic Surgery*, 31(05):446–462, 2015.

[71] A. A. Zadpoor and H. Weinans. Patient-specific bone modeling and analysis: the role of integration and automation in clinical adoption. *Journal of biomechanics*, 48(5):750–760, 2015.

[72] Y. Zeng, Y. Qian, Z. Zhu, J. Hou, H. Yuan, and Y. He. Corrnet3d: Unsupervised end-to-end learning of dense correspondence for 3d point clouds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6052–6061, 2021.

[73] K. Zhang, W. K. Leow, and Y. Cheng. Performance analysis of active shape reconstruction of fractured, incomplete skulls. In *Computer Analysis of Images and Patterns: 16th International Conference, CAIP 2015, Valletta, Malta, September 2-4, 2015 Proceedings, Part I 16*, pages 312–324. Springer, 2015.

[74] H. Zhao, L. Jiang, J. Jia, P. H. Torr, and V. Koltun. Point transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 16259–16268, 2021.