This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Joint Segmentation and Graph-Based Skeletal Representation Learning with Geometric Priors

Yinhao Wu, Haiqing Li, Hehuan Ma, and Junzhou Huang(✉)

Department of Computer Science and Engineering
The University of Texas at Arlington, Arlington, TX, USA
jzhuang@uta.edu

Abstract. Skeletal representations (s-reps) provide a powerful tool for anatomical shape analysis by explicitly capturing object widths and boundary curvatures. Existing s-rep construction methods follow an evolutionary framework for gradual deformation from a reference ellipsoid and remain computationally demanding and slow. In this work, we propose a unified end-to-end deep learning framework for joint volumetric segmentation and skeletal modeling directly from 3D medical images. Built upon a general segmentation backbone, the framework adopts a hybrid variational encoder–decoder architecture that integrates skeletal priors with explicit geometric constraints to promote anatomically consistent skeletal representations. Experiments on hippocampus analysis demonstrate that the proposed framework produces s-reps with accuracy comparable to state-of-the-art evolutionary fitting approaches, while substantially reducing inference time from minutes to seconds. Moreover, the learned s-reps preserve discriminative geometric information that enhances downstream Alzheimer's disease classification and improves segmentation robustness, particularly in limited-data regimes.

Keywords: Skeletal Representations · Shape Analysis · Geometric Learning · Image Segmentation · Graph Neural Network.

1 Introduction

Skeletonization [24,16] plays a vital role in characterizing the interior geometry of anatomical objects by extracting their medial axes. For bent slab-like structures such as the hippocampus, skeletal representations are particularly effective, as they explicitly capture local width variations and boundary curvature along the medial axis, which constitute key geometric properties for neurodegenerative disease analysis [21]. A classical formulation of object skeletons is Blum's medial axis transform (MAT) [2], which represents an object by a medial skeleton and associated spoke vectors whose endpoints lie orthogonally on the boundary (Fig. 1(a)). However, MAT is highly sensitive to boundary noise, where small surface perturbations can lead to unstable and spurious skeletal branches [22]. To address these limitations, skeletal representations (s-reps) [13,19] were introduced as a quasi-medial framework that relaxes strict medial constraints

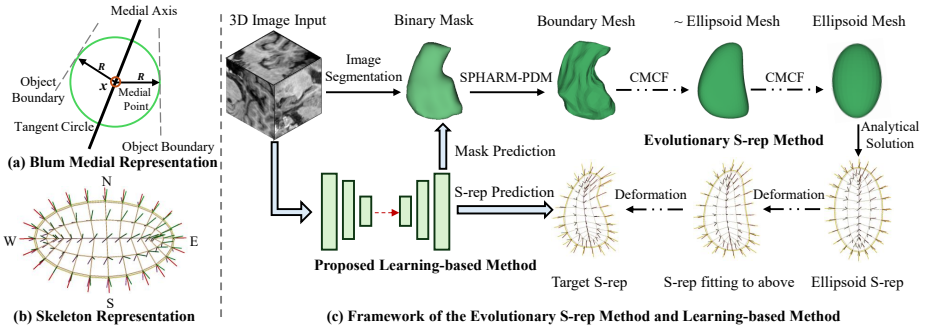

Fig. 1. (a) Blum medial representation, where a medial point x is associated with two orthogonal boundary spokes of equal length R. (b) Skeletal representation (s-rep) discretized as a grid of skeletal points with paired boundary spokes and fold-curve spokes (red). (c) Comparison of evolutionary and learning-based s-rep fitting, where the proposed framework predicts s-reps from 3D images in a single forward pass.

while preserving essential geometric properties (Fig. 1(b)). By employing a fixed-topology skeletal template, s-reps enable population-consistent correspondence and improved robustness to local boundary variations. As a continuous alternative, the Continuous Medial Representation (CM-Rep) [26] models the medial axis as a continuous parametric manifold via inverse skeletonization.

Existing state-of-the-art s-rep construction follows an evolutionary fitting paradigm that models anatomical shapes as diffeomorphic deformations [4] of a reference ellipsoid. As shown in Fig. 1(c), the pipeline applies a forward flow based on Kazhdan's conformalized mean curvature flow (CMCF) [11] to obtain a near-ellipsoidal shape, followed by a multi-stage backward flow that alternates deformation and geometric correction to preserve population-level correspondence. Despite its effectiveness for statistical shape analysis, this evolutionary process is computationally demanding, requiring approximately 5.5 minutes per case in the public SlicerSALT implementation.[1]

Recent deep learning–based methods have shown promising performance for 3D skeletal modeling and shape analysis, while substantially improving computational efficiency and scalability. For instance, Menten *et al.* [14] proposed a topology-preserving 3D skeletonization framework compatible with gradient-based optimization, while Gaggion *et al.* introduced HybridGNet and Hybrid-VNet [6,7], demonstrating the feasibility of directly learning anatomical landmarks and surface meshes from images via hybrid graph neural networks. More recently, s-rep prediction has been formulated as a graph regression problem using GNNs [5]. However, existing learning-based approaches still depend on intermediate representations such as binary masks or surface meshes, which require time-consuming annotations and limit clinical scalability. Moreover, the absence of an end-to-end framework that directly generates s-reps from images makes it

[1] https://github.com/Kitware/SlicerSALT

difficult to integrate skeletal learning into established deep learning workflows in a unified manner.

In this paper, we propose a unified end-to-end framework for joint volumetric segmentation and skeletal modeling from 3D medical images. Built upon a general segmentation backbone (i.e., U-Net), the framework adopts a hybrid variational encoder–decoder architecture that learns latent representations from foreground probability maps and generates object s-reps via a topology-aware spectral graph convolutional decoder. To incorporate anatomical priors, a differentiable skeletonization module extracts medial skeletal cues from the probability map, while learnable node-wise positional embeddings enable position-aware cross-attention to produce node-specific features for graph decoding. To further enforce geometric consistency, we introduce a spoke direction loss and an orthogonality loss that regularize spoke orientations and their alignment with object boundaries. Experimental results show that the proposed framework achieves s-rep accuracy comparable to state-of-the-art evolutionary methods while reducing inference time from minutes to seconds. Moreover, the learned s-reps preserve discriminative geometric information that benefits downstream Alzheimer's disease classification and improves segmentation robustness, particularly in low-data regimes. The main contributions are as follows:

- We propose an end-to-end framework that produces s-reps directly from 3D images, achieving comparable accuracy with substantially faster inference.
- We formulate s-rep prediction as a graph-based learning task that integrates skeletal priors with explicit geometric constraints, yielding anatomically consistent skeletal representations.
- Experiments show that the learned s-reps preserve the geometric sensitivity required for downstream disease classification and improve segmentation performance in low-data regimes.

2 Methodology

2.1 Discrete S-rep Representation and Problem Definition

A discrete skeletal representation (s-rep) is defined by a set of n_a skeletal points $\mathbf{h}_i \in \mathbb{R}^3$ arranged in a fixed grid topology to ensure population-consistent correspondence. Each skeletal point is associated with one or more spokes $S(\mathbf{h}_i) = (\mathbf{U}_i, r_i)$, where $\mathbf{U}_i \in \mathbb{S}^2$ denotes a unit direction and $r_i \in \mathbb{R}_+$ the spoke length, mapping $\mathbf{h}_i$ to a boundary point $\mathbf{b}_i = \mathbf{h}_i + r_i\mathbf{U}_i$. Interior skeletal points have two opposing spokes, while fold skeletal points include an additional spoke connecting the skeletal sheet fold to the boundary crest. As a result, an s-rep can be represented as a fixed-topology graph with $n_s = 2n_a^{\text{int}} + 3n_a^{\text{fold}}$ spokes. In our implementation, each s-rep consists of $n_a = 61$ skeletal points (37 interior and 24 fold points), yielding 146 spokes and 207 skeletal and boundary points.

Given a 3D input image $\mathbf{I} \in \mathbb{R}^{H\times W\times D}$, our goal is to jointly predict volumetric segmentation and its corresponding s-rep via a mapping $\mathcal{F} : \mathbf{I} \to \{\mathbf{P}, \mathcal{G}\}$. Here, $\mathbf{P} \in [0,1]^{H\times W\times D}$ denotes the voxel-wise foreground probability map,

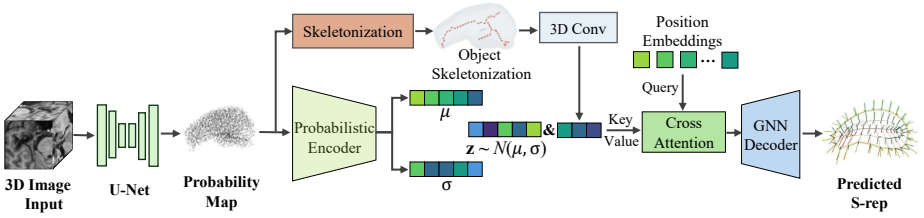

Fig. 2. Overview of the proposed end-to-end framework for joint volumetric segmentation and s-rep learning. A 3D U-Net first predicts a foreground probability map, which is encoded by a variational encoder to capture global anatomical context. In parallel, a differentiable skeletonization module extracts medial structural cues that are embedded and fused with the latent representation. Learnable s-rep node position embeddings query the fused features via cross-attention, and a graph convolutional decoder produces the final s-rep.

and $\mathcal{G} = \{\mathcal{V}, \mathbf{A}, \mathbf{X}\}$ is the graph representation of the s-rep. The vertex set $\mathcal{V}$ includes skeletal nodes $\mathbf{h}_i$ and their associated boundary nodes $\mathbf{b}_{ij}$, while $\mathbf{A} \in \{0,1\}^{|\mathcal{V}|\times|\mathcal{V}|}$ encodes the fixed s-rep connectivity. Following [5], adjacency is defined by connecting each quadrilateral of neighboring skeletal points to its corresponding boundary quadrilateral, forming local volumetric elements that induce consistent neighborhood relations. Finally, $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}|\times 3}$ stores the 3D coordinates of all graph nodes.

2.2 Anatomy-Aware Probabilistic S-rep Learning

Probabilistic Encoding with Skeletal Priors. An overview of the proposed end-to-end framework is illustrated in Fig. 2. We first apply a 3D U-Net [3] to obtain a foreground probability map $\mathbf{P} \in [0,1]^{H\times W\times D}$, which serves as a probabilistic representation of the target anatomy. A variational encoder [12] then maps $\mathbf{P}$ to a latent distribution parameterized by $(\boldsymbol{\mu}, \boldsymbol{\sigma})$ with $\boldsymbol{\mu}, \boldsymbol{\sigma} \in \mathbb{R}^d$, from which a latent code $\mathbf{z} \in \mathbb{R}^d$ is sampled via the reparameterization trick $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where d denotes the latent feature dimension. In parallel, a differentiable skeletonization operator $\mathcal{T}(\cdot)$ is applied to $\mathbf{P}$ to extract a skeletal map $\mathcal{K} \in [0,1]^{H\times W\times D}$, following [5]. The skeletal map is encoded by a lightweight 3D convolutional encoder $\phi_K(\cdot)$ into M skeletal tokens, denoted $\phi_K(\mathcal{K}) \in \mathbb{R}^{M\times d}$. The global latent code $\mathbf{z}$ is broadcast across all M tokens and concatenated with $\phi_K(\mathcal{K})$ along the feature dimension, followed by a linear projection, yielding a unified representation $\tilde{\mathbf{z}} \in \mathbb{R}^{M\times d}$ that jointly captures global anatomical context and skeletal priors for subsequent s-rep decoding.

Position-Aware Skeletal Context Aggregation. To enable flexible and node-specific s-rep learning, each graph node $v_i \in \mathcal{V}$ is associated with a learnable positional embedding $\mathbf{e}_i \in \mathbb{R}^d$ that encodes its canonical anatomical location within the fixed s-rep topology. These embeddings act as queries in a cross-attention module [25], while the fused token matrix $\tilde{\mathbf{z}} \in \mathbb{R}^{M\times d}$ serves as both keys and values. Through this mechanism, each node selectively aggregates spatially relevant

skeletal context as

$$\mathbf{f}_i = \mathrm{softmax}\left(\frac{\mathbf{e}_i \tilde{\mathbf{z}}^\top}{\sqrt{d}}\right) \tilde{\mathbf{z}} \in \mathbb{R}^d, \tag{1}$$

where $\mathbf{f}_i$ denotes the skeletal-aware feature for node v_i.

Graph Convolutional Decoder. Given the fixed s-rep topology, we employ a graph convolutional network to decode node-wise s-rep representations. Each node $v_i \in \mathcal{V}$ is initialized with a feature vector $\mathbf{f}_i$ that fuses global latent features and skeletal context. A stack of five graph convolutional layers with Layer Normalization propagates structural information over the fixed adjacency graph $\mathbf{A}$ and directly regresses the 3D coordinates $\mathbf{X} \in \mathbb{R}^{|\mathcal{V}|\times 3}$ of all skeletal and boundary nodes, yielding the final s-rep representation.

2.3 Loss Functions

The proposed framework is trained end-to-end by jointly optimizing volumetric segmentation and s-rep generation. The overall objective is defined as

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\text{seg}} + \lambda_2 \mathcal{L}_{\text{KL}} + \lambda_3 \mathcal{L}_{\text{mse}} + \lambda_4 \mathcal{L}_{\text{spoke}} + \lambda_5 \mathcal{L}_{\text{orth}}. \tag{2}$$

Here, the segmentation loss $\mathcal{L}_{\text{seg}}$ supervises accurate volumetric segmentation. The Kullback–Leibler (KL) divergence loss $\mathcal{L}_{\text{KL}}$ imposes a unit Gaussian prior $\mathcal{N}(0, 1)$ on the latent posteriors, regularizing the variational space and stabilizing global shape learning. The mean squared error loss $\mathcal{L}_{\text{mse}}$ enforces coordinate-level fidelity of the generated s-reps. In addition, we introduce a spoke direction loss $\mathcal{L}_{\text{spoke}}$ and an orthogonality loss $\mathcal{L}_{\text{orth}}$ to encourage consistent spoke orientations and orthogonality between skeletal spokes and object boundaries, respectively. They are formulated as

$$\mathcal{L}_{\text{spoke}} = \frac{1}{n_s}\sum_{i=1}^{n_s}\big(1 - \langle \hat{\mathbf{s}}_i, \mathbf{s}_i\rangle\big), \quad \mathcal{L}_{\text{orth}} = \frac{1}{n_s}\sum_{i=1}^{n_s}\big(1 - \langle \hat{\mathbf{s}}_i, \mathbf{n}_i\rangle\big), \tag{3}$$

where $\hat{\mathbf{s}}_i$ and $\mathbf{s}_i$ denote the predicted and reference spoke direction vectors, respectively, both normalized to unit length, and $\mathbf{n}_i$ denotes the outward surface normal at the corresponding boundary point.

3 Experiments

3.1 Datasets and Implementation Details

We evaluate the proposed framework on two hippocampus benchmarks. The first is the Medical Segmentation Decathlon (MSD) dataset [1] with 520 volumes, and the second is the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [15] with 194 volumes, which is additionally used for downstream classification to distinguish Alzheimer's disease (AD) from normal controls (NC). We perform patient-level five-fold cross-validation on both datasets, ensuring that left and right hippocampi from the same subject are assigned to the same fold

Table 1. Ablation study on geometric components for s-rep prediction. SP denotes skeletal priors. $\mathcal{L}_{\text{spoke}}$ and $\mathcal{L}_{\text{orth}}$ denote spoke direction and orthogonality constraints, respectively. Results are reported on MSD and ADNI datasets.

Components			MSD Dataset				ADNI Dataset			
SP	$\mathcal{L}_{\text{spoke}}$	$\mathcal{L}_{\text{orth}}$	RMSE ↓	Medialness ↓	Angle ↓	Orthog. ↓	RMSE ↓	Medialness ↓	Angle ↓	Orthog. ↓
✗	✗	✗	0.073(0.01)	1.282(0.31)	1.053(0.11)	0.831(0.06)	0.067(0.01)	1.294(0.02)	1.059(0.07)	0.851(0.03)
✓	✗	✗	0.032(0.01)	1.124(0.01)	0.582(0.04)	0.669(0.01)	0.062(0.01)	1.195(0.06)	1.006(0.11)	0.827(0.02)
✓	✓	✗	0.055(0.02)	**1.028**(0.03)	0.399(0.04)	0.720(0.04)	0.089(0.02)	1.088(0.10)	**0.627**(0.61)	0.822(0.08)
✓	✓	✓	0.101(0.03)	1.032(0.07)	**0.381**(0.03)	**0.535**(0.02)	0.113(0.01)	**1.043**(0.14)	0.660(0.05)	**0.651**(0.01)
Reference S-reps			–	0.969(0.06)	–	0.506(0.06)	–	1.028(0.01)	–	0.567(0.01)

to avoid data leakage. Reference s-reps for skeletal supervision are generated on the training set using the evolutionary s-rep fitting framework [18], based on hippocampal surface meshes parameterized via SPHARM-PDM [23]. All input images are preprocessed by normalizing intensities to the range $[0, 1]$ and resampling to an isotropic voxel spacing of $[1.0, 1.0, 1.0]$ mm. Volumes are then zero-padded to a fixed size of $64 \times 64 \times 64$. Loss weights $\{\lambda_i\}_{i=1}^{5}$ are determined by grid search on the validation set to balance segmentation accuracy and s-rep geometric validity. Specifically, $\lambda_1 = 1.0$, $\lambda_2 = 0.001$, $\lambda_3 = 0.01$, $\lambda_4 = \lambda_5 = 0.001$. Training is conducted for 70 epochs using Adam with a learning rate of 1×10^{-4}, weight decay 1×10^{-5}, and a batch size of 4.

3.2 Evaluation Metrics

We evaluate the proposed framework on both volumetric segmentation and s-rep prediction tasks. Segmentation performance is measured using the Dice similarity coefficient and the 95th percentile Hausdorff distance (HD95). For s-rep prediction, root mean squared error (RMSE) is used to assess coordinate accuracy. Following [5], we adopt three skeleton-specific geometric metrics. *Medialness* is defined as the ratio between paired top and bottom spoke lengths, with values closer to 1 indicating better medial alignment. *Angle* measures the angular deviation between corresponding spokes. *Orthogonality* is defined as the angle between spoke directions and boundary surface normals, where smaller values indicate better boundary alignment.

3.3 Ablation Study

To investigate the contribution of each geometric component, we conduct an ablation study by progressively introducing skeletal priors (SP), the spoke direction constraint ($\mathcal{L}_{\text{spoke}}$), and the orthogonality constraint ($\mathcal{L}_{\text{orth}}$). As shown in Table 1, incorporating skeletal priors substantially improves s-rep prediction accuracy over the baseline, particularly in terms of medialness (e.g., from 1.282 to 1.124 on MSD and from 1.294 to 1.195 on ADNI), indicating that geometry-informed initialization provides a strong inductive bias for stable s-rep learning. Adding $\mathcal{L}_{\text{spoke}}$ further enhances spoke orientation consistency, yielding marked

Table 2. Downstream Alzheimer's disease classification performance using different s-rep configurations on the ADNI dataset.

Method	AUC	AUPR	Accuracy
Baseline	0.536 ± 0.049	0.466 ± 0.077	0.522 ± 0.026
Ours (SP)	0.557 ± 0.095	0.465 ± 0.087	0.556 ± 0.045
Ours (SP + $\mathcal{L}_{\text{spoke}}$)	0.593 ± 0.026	0.487 ± 0.048	0.569 ± 0.050
Ours (SP + $\mathcal{L}_{\text{spoke}}$ + $\mathcal{L}_{\text{orth}}$)	0.612 ± 0.074	0.519 ± 0.097	0.583 ± 0.067
Liu et al. [13]	0.606 ± 0.080	0.519 ± 0.095	0.584 ± 0.076
Evolutionary S-reps [18]	0.619 ± 0.066	0.525 ± 0.083	0.567 ± 0.067

reductions in Angle (e.g., from 0.582 to 0.399 on MSD and from 1.006 to 0.627 on ADNI), while $\mathcal{L}_{\text{orth}}$ consistently achieves the lowest orthogonality error on both datasets (e.g., from 0.669 to 0.535 on MSD and from 0.827 to 0.651 on ADNI). Notably, these results reveal a trade-off between coordinate accuracy (RMSE) and geometric validity, suggesting that enforcing anatomically meaningful constraints favors s-reps that better conform to the quasi-medial definition.

3.4 Downstream Evaluations

To assess whether the learned s-reps preserve the statistical properties required for population-level shape analysis, we evaluate them on a downstream disease classification task using the ADNI dataset, distinguishing 39 early-stage Alzheimer's disease (AD) subjects from 58 normal controls (NC) based on hippocampal shape. This task is particularly challenging because early neurodegenerative shape differences are subtle, making it a sensitive test of statistical shape representations. Following standard practice in s-rep–based shape analysis [17], spoke directions are Euclideanized using Principal Nested Spheres (PNS) [10], and spoke radii are log-transformed. For all s-rep–based classifiers, we use the same skeletal features following [18]: each hippocampal s-rep consists of 61 interior skeletal points and 24 fold skeletal points, yielding 8,076 Euclideanized features per hippocampus, which are PCA-reduced to 177 dimensions. The resulting features are used to train a Distance Weighted Discrimination (DWD) [20] classifier. The evolutionary s-rep baseline and the proposed method share the same statistical analysis pipeline, differing only at test time: evolutionary s-reps are replaced by network-predicted s-reps for the proposed approach. As reported in Table 2, classification using predicted s-reps achieves performance comparable to evolutionary and Liu's fitted s-reps in terms of AUC and AUPR, while yielding slightly higher balanced accuracy. These results demonstrate that the learned s-reps preserve geometric validity and retain the statistical sensitivity necessary for downstream disease classification and population-level shape analysis.

3.5 S-rep Improves Data-Efficient Segmentation

An important observation from our experiments is that joint s-rep learning improves volumetric segmentation performance, particularly in data-scarce regimes.

Table 3. Inference-time computational efficiency comparison between evolutionary s-rep fitting and the proposed learning-based method.

Method	Inference Time (s/case)	Speed-up	#Params (M)	Hardware
Evolutionary S-rep	310.0	1×	–	CPU
Ours	2.1	148×	15.89	CPU
Ours	0.27	1148×	15.89	GPU

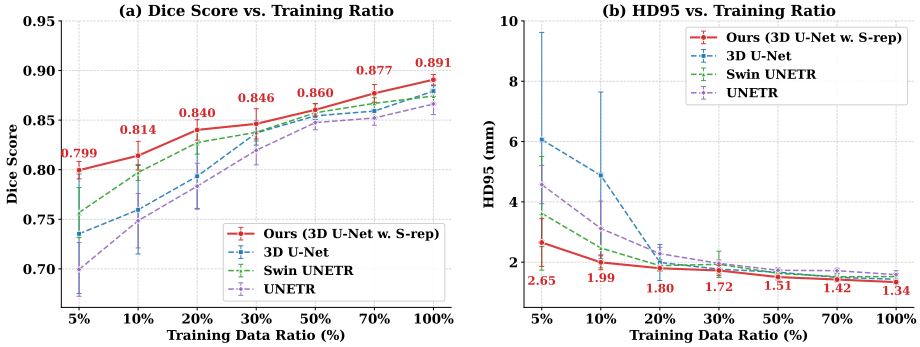

Fig. 3. Segmentation performance on the MSD dataset across different training data ratios. Joint s-rep learning improves Dice scores and reduces HD95 compared to 3D U-Net, Swin UNETR, and UNETR, especially when limited training data are available.

We evaluate this effect on the MSD dataset under varying training data ratios, comparing against 3D U-Net [3], Swin UNETR [8], and UNETR [9]. As shown in Fig. 3, incorporating s-rep supervision consistently improves Dice scores and reduces HD95 across all ratios, with the largest gains observed in low-data settings (e.g., 5% and 10%). These improvements stem from two complementary factors: (i) the skeletal representation provides an explicit structural prior that enforces topological consistency and suppresses anatomically implausible predictions under limited supervision; and (ii) geometric constraints such as $\mathcal{L}_{\text{spoke}}$ and $\mathcal{L}_{\text{orth}}$ introduce shape-aware supervision beyond voxel-wise objectives, strengthening boundary regularization.

3.6 Computational Efficiency Analysis

We evaluate inference-time computational efficiency by comparing the proposed method with the evolutionary s-rep fitting approach. All experiments are conducted on a server equipped with an Intel Xeon Gold 5420 CPU and an NVIDIA H100 PCIe GPU. As reported in Table 3, the evolutionary method requires approximately 310 seconds per case, due to its reliance on iterative, per-instance geometric processing without learned parameters. In contrast, the learning–based framework performs s-rep prediction via a single forward pass, reducing inference time to 2.1 seconds per case on CPU and 0.27 seconds per case on GPU.

This corresponds to a 148× speedup on CPU and a 1148× speedup on GPU, demonstrating the substantial efficiency advantage of the proposed approach while maintaining comparable geometric and statistical performance.

4 Conclusions

In this work, we present an end-to-end framework for direct inference of skeletal representations (s-reps) from 3D volumes, unifying volumetric segmentation with skeletal geometry modeling and alleviating the reliance on evolutionary fitting and intermediate boundary meshes. Experimental results show that our pipeline reduces inference time from minutes to seconds while preserving sensitivity to subtle morphological changes for clinical classification. Moreover, joint skeletal learning introduces strong geometric priors that improve segmentation accuracy, particularly in data-scarce regimes.

Acknowledgments. This work was partially supported by US National Science Foundation IIS-2412195, CCF-2400785, the Cancer Prevention and Research Institute of Texas (CPRIT) award (RP230363), the National Institutes of Health (NIH) R01 award (1R01AI190103-01) and Microsoft Accelerate Foundation Models Research (2024).

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

References

1. 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)
2. Blum, H.: A transformation for extracting new descriptions of shape. Models for the perception of speech and visual form pp. 362–380 (1967)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International conference on medical image computing and computer-assisted intervention. pp. 424–432. Springer (2016)
4. Durrleman, S., Prastawa, M., Charon, N., Korenberg, J.R., Joshi, S., Gerig, G., Trouvé, A.: Morphometry of anatomical shape complexes with dense deformations and sparse parameters. NeuroImage **101**, 35–49 (2014)
5. Gaggion, N., Ferrante, E., Paniagua, B., Vicory, J.: Fitting skeletal models via graph-based learning. In: 2024 IEEE International Symposium on Biomedical Imaging (ISBI). pp. 1–4. IEEE (2024)
6. Gaggion, N., Mansilla, L., Mosquera, C., Milone, D.H., Ferrante, E.: Improving anatomical plausibility in medical image segmentation via hybrid graph neural networks: Applications to chest x-ray analysis. IEEE Transactions on Medical Imaging **42**, 546–556 (2022)
7. Gaggion, N., Matheson, B.A., Xia, Y., Bonazzola, R., Ravikumar, N., Taylor, Z.A., Milone, D.H., Frangi, A.F., Ferrante, E.: Multi-view hybrid graph convolutional network for volume-to-mesh reconstruction in cardiovascular mri. Medical Image Analysis **104**, 103630 (2025). https://doi.org/10.1016/j.media.2025.103630

8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
9. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 574–584 (2022)
10. Jung, S., Dryden, I.L., Marron, J.S.: Analysis of principal nested spheres. Biometrika **99**(3), 551–568 (2012)
11. Kazhdan, M., Solomon, J., Ben-Chen, M.: Can mean-curvature flow be modified to be non-singular? In: Computer Graphics Forum. vol. 31, pp. 1745–1754. Wiley Online Library (2012)
12. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. CoRR **abs/1312.6114** (2013)
13. Liu, Z., Hong, J., Vicory, J., Damon, J.N., Pizer, S.M.: Fitting unbranching skeletal structures to objects. Medical Image Analysis **70**, 102020 (2021)
14. Menten, M.J., Paetzold, J.C., Zimmer, V.A., Shit, S., Ezhov, I., Holland, R., Probst, M., Schnabel, J.A., Rueckert, D.: A skeletonization algorithm for gradient-based optimization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21394–21403 (October 2023)
15. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack Jr, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., et al.: Alzheimer's disease neuroimaging initiative (adni) clinical characterization. Neurology **74**(3), 201–209 (2010)
16. Petrov, D., Goyal, P., Thamizharasan, V., Kim, V., Gadelha, M., Averkiou, M., Chaudhuri, S., Kalogerakis, E.: Gem3d: Generative medial abstractions for 3d shape synthesis. In: ACM SIGGRAPH 2024 Conference Papers. pp. 1–11 (2024)
17. Pizer, S.M., Hong, J., Vicory, J., Liu, Z., Marron, J., Choi, H.y., Damon, J., Jung, S., Paniagua, B., Schulz, J., et al.: Object shape representation via skeletal models (s-reps) and statistical analysis. In: Riemannian geometric statistics in medical image analysis, pp. 233–271. Elsevier (2020)
18. Pizer, S.M., Liu, Z., Zhao, J., Tapp-Hughes, N., Damon, J., Zhang, M., Marron, J., Taheri, M., Vicory, J.: Interior object geometry via fitted frames. Journal of Mathematical Imaging and Vision **67**(5), 51 (2025)
19. Pizer, S.M., Marron, J., Damon, J.N., Vicory, J., Krishna, A., Liu, Z., Taheri, M.: Skeletons, object shape, statistics. Frontiers in computer science **4**, 842637 (2022)
20. Qiao, X., Zhang, H.H., Liu, Y., Todd, M.J., Marron, J.S.: Weighted distance weighted discrimination and its asymptotic properties. Journal of the American Statistical Association **105**(489), 401–414 (2010)
21. Schulz, J., Pizer, S.M., Marron, J., Godtliebsen, F.: Non-linear hypothesis testing of geometric object properties of shapes applied to hippocampi. Journal of Mathematical Imaging and Vision **54**(1), 15–34 (2016)
22. Siddiqi, K., Pizer, S.: Medial representations: mathematics, algorithms and applications, vol. 37. Springer Science & Business Media (2008)
23. Styner, M., Oguz, I., Xu, S., Brechbühler, C., Pantazis, D., Gerig, G.: Statistical shape analysis of brain structures using spharm-pdm. In: Insight Journal, MICCAI 2006 Opensource Workshop. vol. 7 (2006)
24. Tagliasacchi, A., Alhashim, I., Olson, M., Zhang, H.: Mean curvature skeletons. In: Computer Graphics Forum. vol. 31, pp. 1735–1744. Wiley Online Library (2012)

25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
26. Yushkevich, P.A., Zhang, H., Gee, J.C.: Continuous medial representation for anatomical structures. IEEE transactions on medical imaging **25**(12), 1547–1564 (2006)