Paper #: **12926-93**

SPIE.

Vertebral Segmentation without Training using Differentiable Appearance Modeling of a Deformable Spine Template

Hyunsoo Kim, Jinah Park
Computer Graphics and Visualization Lab., School of Computing, KAIST

KAIST

TL; DR

Given a spine template as the only prior knowledge, we perform vertebral segmentation by optimizing the shape and appearance through simple gradient descent, enabled by our novel **differentiable template-to-image pipeline**

Introduction

Deformable model-based medical image analysis

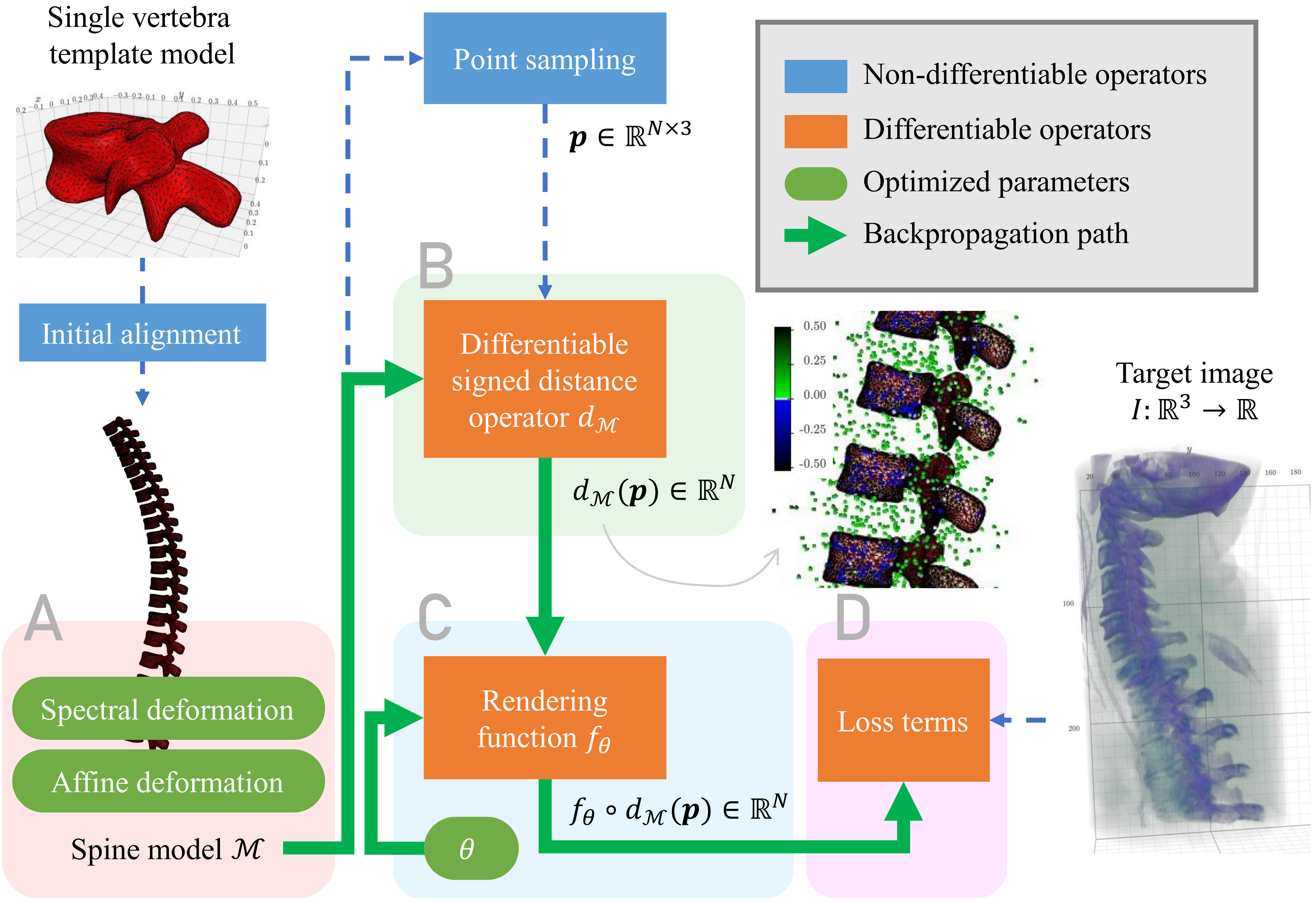
- Provides correspondence information
- More transparent on reasoning
- Can work with significantly smaller dataset

Differentiable rendering [1] in computer vision

- Finds optimal scene parameters based on the rendering process
- Gradient-based optimization: simple, powerful, and extensible

Our approach
Differentiable Appearance Modeling

- Jointly optimizes the shape and appearance of the spine model to match the given CT image
- *No training dataset required:* Learning appearance only from the given image
- Why spine? High contrast surface + repetitive structures are easily exploitable to achieve the segmentation without large dataset



↑ Overview of our vertebral segmentation framework

Method

A Spectral mesh optimization

$$\mathcal{M}\begin{pmatrix} A_i \in \mathbb{R}^{3\times 3} \\ U_i \in \mathbb{R}^{V\times 3} \end{pmatrix}$$

Spine model $\mathcal{M}$ is described by affine deformation A_i and spectral coefficients U_i
→ **Shape parameters** with natural coarse-to-fine hierarchy

$$U_i \leftarrow U_i - lr \times \left(\frac{\Lambda}{\lambda_0}\right)^{-\alpha} \times \frac{\partial \mathcal{L}}{\partial U_i}$$

Gradient is rescaled proportional to the eigenvalues
Large α → coarse optimization
Small α → fine control

B Differentiable signed distance operator

$$d_{\mathcal{M}}: \mathbb{R}^3 \to \mathbb{R}$$

The operator takes point positions and gives the signed distance from the surface mesh $\mathcal{M}$
→ **Bridging shape & appearance**

$$\frac{\partial d_{\mathcal{M}}}{\partial v_k}$$

Our backward step is allowing losses to propagate to the shape, accelerated by GPU + spatial structure

C Appearance modeling

$$f_\theta: \mathbb{R} \to \mathbb{R}$$

Rendering function takes the signed distance and gives the appearance (voxel intensity)
→ **Appearance model** that fits high-contrast surfaces

Learnable function f_θ is modeled using a shallow neural network (Fourier feature transform + 3-layer MLP)

D Losses & optimization

$\mathcal{L}_I$ - Image similarity loss
Guide appearance to match the given image (L1 loss)

$\mathcal{L}_E$ - Edge length regularizer
Regulate mesh complexity

$\mathcal{L}_N$ - Normal regularizer
Align surface normal to image gradient

$\mathcal{L}_O$ - Overlap regularizer
Penalize overlaps between meshes

$\mathcal{L}_V$ - Variance regularizer
Regulate shape variance between neighboring vertebra

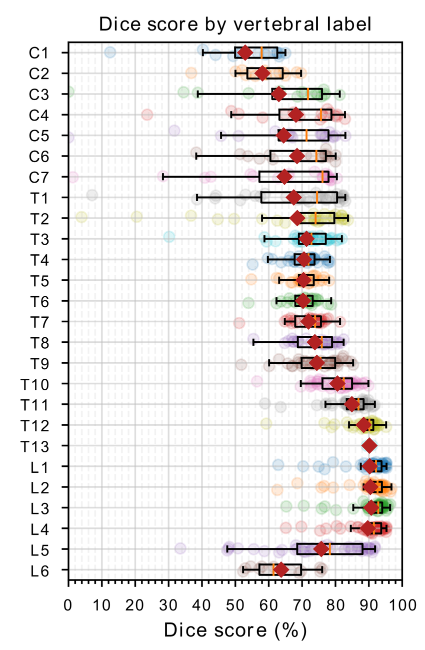
Point sampling
Increase sampling rate around the mesh surface

AdamVector optimizer
Reduce artifacts with vector-wise normalization

Coarse-to-fine optimization
Start optimization with low-passed image, gradually introduce high-freq. features

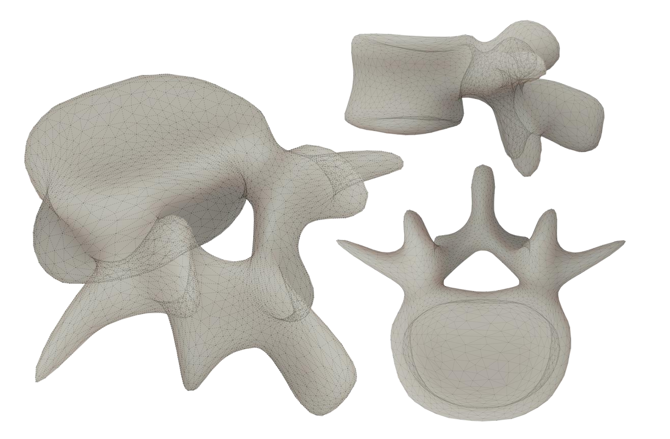
Experiments & Results

Our pipeline is tested on VerSe20' dataset [2]
Vertebra template model is made from the average shape of xVertSeg lumbars [3]

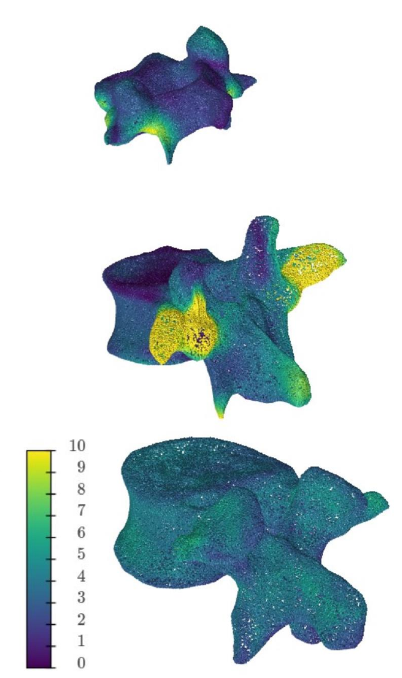


Segmentation

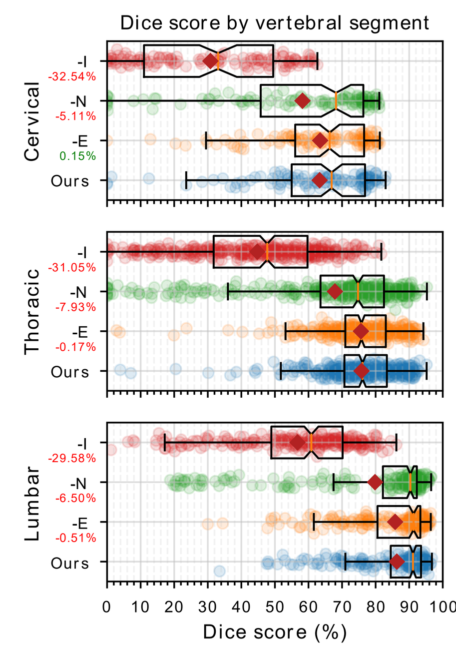
By vertebral label, lumbars have highest accuracy as their shapes are similar to lumbar-based template model.
Within a vertebra, vertebral bodies have higher correspondence accuracy than vertebral arches due to less complex shape & varying appearance.



↑ Vertebra template model used for the segmentation task



Correspondence error within a vertebra, selected from each segment (Top: cervical, Middle: thoracic, Bottom: lumbar) →



Ablation study

$\mathcal{L}_I$ is crucial for our optimization framework
$\mathcal{L}_N$ resolves ambiguous zero-isosurface for the rendering function
$\mathcal{L}_E$ improves the mesh quality, rather than segmentation accuracy

Conclusion

Differentiable Appearance Modeling allows **robust and dataset-free analysis** and **easily adaptable** to various neural network training/applications by being differentiable

Future works include developing holistic template models/sophisticated rendering functions and applying to image synthesis tasks

References

[1] Kato, H. et al., "Differentiable rendering: A survey," arXiv preprint arXiv:2006.12057 (2020).
[2] Sekuboyina, A. et al., "Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images," Medical image analysis 73, 102166 (2021).
[3] Korez, R. et al., "A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation," IEEE transactions on medical imaging 34(8), 1649–1662 (2015).

More info?

Contact me!
khskhs@kaist.ac.kr
@k_ _hyunsoo