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

Hyunsoo Kim[a] and Jinah Park[a]

[a]School of Computing, KAIST, Daejeon, South Korea

## ABSTRACT

We present a robust vertebral segmentation framework that can bootstrap the segmentation task with little to no dataset. Our deformable model-based framework jointly optimizes the appearance and shape of the spine model using the novel differentiable appearance modeling method. Because our framework learns the appearance of the spine only from the given image, it does not rely on the dataset or handcrafted image features and adapts robustly to the appearance of the image. With our proposed differentiable signed distance operator and spectral mesh optimization, the shape of the spine model can be refined via a gradient-based optimizer. Our framework was tested on the VerSe'20 training dataset, and it achieved an average Dice score of up to 90% for selected vertebral labels. Our results suggest that utilizing the explicit knowledge from the template model can significantly reduce the need for a large training dataset.

**Keywords:** Deformable model, differentiable appearance modeling, differentiable signed distance operator, spectral mesh optimization, vertebral segmentation

## 1. INTRODUCTION

Vertebral segmentation, a task of retrieving the structure of vertebrae from computed tomography (CT) spinal images, has received constant attention due to the crucial role of vertebrae in supporting the body and protecting the spinal cord. Accurate automated segmentation facilitates the timely detection, prevention, and treatment of various spine-related diseases. Early approaches to automated vertebral segmentation, as explored in previous studies,[1–4] mainly based on classical image processing methods. With the emergence of deep learning and the release of public CT image datasets, deep learning-based vertebral segmentation methods[5,6] have shown better performance compared to the classical ones.

Although recent advances in vertebral segmentation are mainly driven by convolutional neural networks supported by extensive datasets, *deformable model-based methods*[7,8] and their applications to the vertebral segmentation task[9,10] have demonstrated distinct advantages that are not easily replicable by neural networks. Leveraging explicit knowledge about the target subject from a deformable template model, such model-based methods can achieve several benefits: (1) conducting segmentation with significantly smaller datasets, (2) transparently elucidating the reasoning behind results, and (3) offering correspondence information between segmentation outcomes. However, deformable model-based methods often rely on handcrafted energy for optimization, resulting in potential robustness issues.

Our goal is to improve robustness while retaining the inherent advantages of the deformable model-based method. Our main inspiration comes from the recent advancement of differentiable rendering[11] as an approach to inverse rendering problems in computer vision. Inspired by this optimization-based inverse problem solving approach, in this paper, we present a novel vertebral segmentation framework from CT spine images using differentiable appearance modeling. Our main contributions are threefold: (1) an appearance optimization framework for vertebral segmentation, which aligns the deformable spine template to the given image by learning the appearance of the spine without a training process; (2) a differentiable signed distance operator, which has

Further author information: (Send correspondence to Jinah Park)
Hyunsoo Kim: E-mail: khskhs@kaist.ac.kr, Telephone: +82 (0)42 350 7755
Jinah Park: E-mail: jinahpark@kaist.ac.kr, Telephone: +82 (0)42 350 3555
The source code is available at https://github.com/SSTDV-Project/DiffAM.

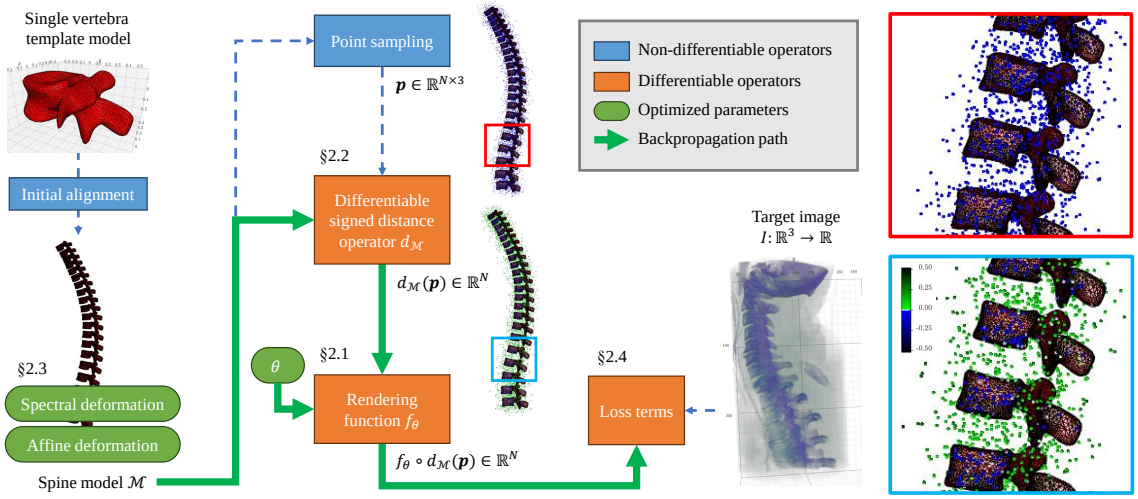



Figure 1: Overview of our vertebral segmentation framework.

an analytic gradient formulation, sublinear time complexity, and GPU accelerated implementation; and (3) a spectral mesh optimization, which enables gradient-based optimizers to refine the deformable mesh in a stable, robust, and coarse-to-fine manner.

## 2. METHOD

*Without* using any training process or predefined image feature, our framework can extract spine shape and correspondence by aligning a 3D deformable spine template to a volumetric CT image. To achieve this, our framework exploits the following properties of spine imagery: (1) the repetitive structure of vertebrae and (2) the high contrast of the bone-tissue boundary. Given a coarsely aligned spine template, our framework refines the alignment by jointly optimizing the appearance (*i.e.*, voxel intensity) and the shape of the spine model using gradient descent, minimizing the difference between the predicted image of the spine model and the given target image. Figure 1 illustrates our framework.

### 2.1 Appearance modeling

We assume that the voxel intensity is predominantly determined by the signed distance from the bone-tissue boundary. With this assumption, our framework models the appearance of the spine model with a rendering function $f_\theta : \mathbb{R} \to \mathbb{R}$ from the signed distance to a voxel intensity. Defined by a multilayer perceptron, the rendering function $f_\theta$ can be optimized using gradient descent. Additionally, Fourier feature encoding[12] is used to let the neural function learn sharp boundaries efficiently.

### 2.2 Differentiable signed distance operator

We propose a differentiable signed distance operator, which enables optimizing the shape of the deformable spine model by backpropagating the image loss. Computing a signed distance $d$ at a point $p$ from a closed mesh $\mathcal{M}$ is performed in two parts: computing an unsigned distance and determining the sign by checking whether $p$ is inside $\mathcal{M}$ or not. The unsigned distance from $\mathcal{M}$ to $p$ can be easily calculated by finding the minimum among the distances from $p$ to each triangle consisting of $\mathcal{M}$. To determine the sign, the winding numbers[13] are used for robustness. Using octrees and the Barnes-Hut approximation,[14] the computation of both unsigned distances and winding numbers can be done in sub-linear time and easily accelerated by GPUs.

Given the definition of the signed distance $d$, the gradient $(\partial d/\partial p, \partial d/\partial v_i)$ can be derived. A simple geometric proof gives $\partial d/\partial p = \texttt{sign}(d)\texttt{normalize}(p - p_*)$, and $\partial d/\partial v_i = -w_i \partial d/\partial p$, where $p_*$ is the closest point from $p$ on $\mathcal{M}$, $v_i$ is a vertex of the face containing $p_*$ and $w_i$ is a barycentric coordinate of $p_*$ with respect to $v_i$.

### 2.3 Spectral mesh optimization

For a stable and accurate optimization of the vertebral mesh consisting of $K$ vertices, we propose spectral mesh optimization, where instead of the vertex positions $V \in \mathbb{R}^{K\times 3}$, the spectral coefficients $U \in \mathbb{R}^{K\times 3}$ of the Laplace-Beltrami eigenbases are used to parameterize the shape. For background on the use of Laplace-Beltrami eigenbases, refer to the course by Lévy and Zhang.[15] In essence, the eigenbases $\Phi$ of the Laplace-Beltrami operator $\Delta$ on a given mesh form a set of orthonormal bases for a set of functions on the mesh, and the eigenvalues $\Lambda$ encode the *frequency* of the corresponding eigenbasis, similar to Fourier bases for a flat surface. Thus, the vertex positions $V$, which is a function on the mesh surface, can be written as a linear combination of eigenbases. Those coefficients $U = \Phi^T V$ reparameterize the shape into low- to high-frequency signal magnitudes, providing coarse-to-fine control over the shape.

In addition to the well-established spectral mesh deformation technique, we introduce the following adaptations for a gradient-based optimization framework. During optimization, our method utilizes eigenvalues to control the coarse-to-fine manner of the optimizer. Given the eigenvalues $\Lambda$, our method uses adaptive step sizes for each frequency band as follows:

$$U \leftarrow U - \texttt{lr} \times (\frac{\Lambda}{\lambda_0})^{-\alpha} \times \frac{\partial \mathcal{L}}{\partial U}, \tag{1}$$

where $\lambda_0$ is the smallest nontrivial eigenvalue and $\alpha$ controls how much details are suppressed. In our experiments, the optimization starts with $\alpha = 1$ and linearly decreases to $\alpha = 0.4$ at the end of the optimization. Our method also exploits the repetitiveness of the spine model by approximating individual eigenbases with a single set of eigenbases from the average shape of vertebrae.

### 2.4 Losses

The image similarity loss is defined as the mean $L_1$ distance between the predicted appearance of sampled points and the voxel intensity from the target image at the same sampled points, or

$$\mathcal{L}_I = \frac{1}{N}\sum_i^N ||f_\theta(d_\mathcal{M}(p_i)) - I(p_i)||_1, \tag{2}$$

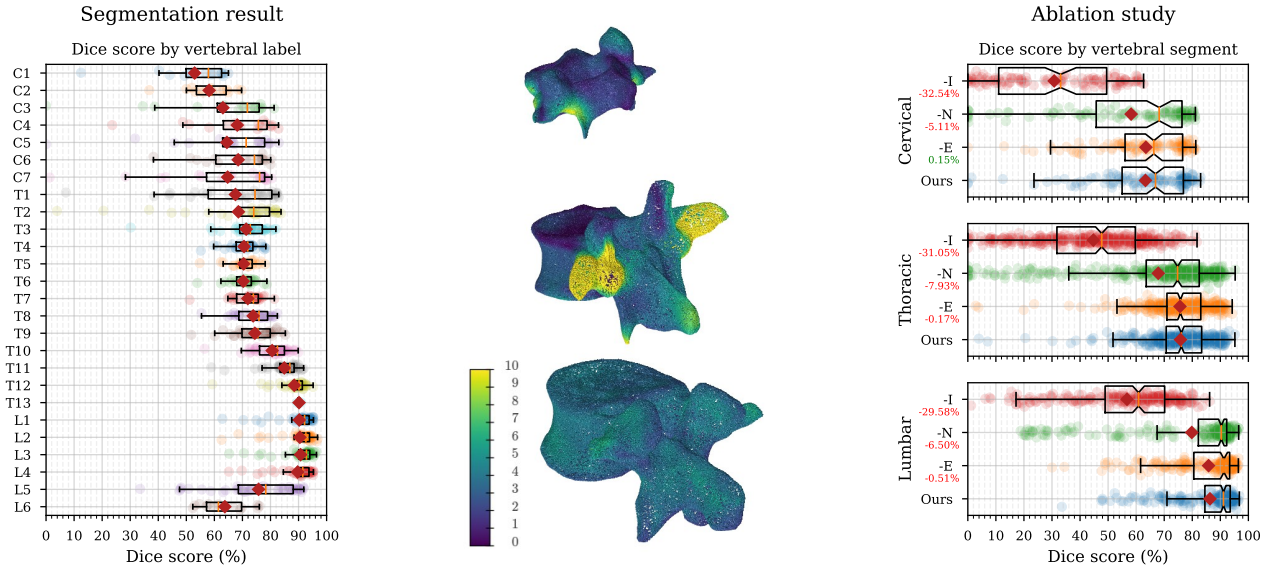
where $N$ denotes the number of sampled points, and $I(p_i)$ is trilinearly interpolated voxel intensity at point $p_i$ from the target image $I$. In our experiments, $N = 10{,}000$ is used.

In addition to image similarity loss, the following four regularizers are used to stabilize the optimization process and the result. The edge length regularizer $\mathcal{L}_E = \frac{1}{|E|}\sum_i^{|E|} ||e_i||_2^2$ smooths the mesh and evenly distributes the vertices. The normal regularizer $\mathcal{L}_N = \frac{1}{M}\sum_i^M \exp(-|n(q_i)\cdot\nabla I(q_i)|)$, calculated with additional $M$ points $q_i$ sampled on the mesh surface, aligns the mesh surface with the local maxima of the image gradient. The overlap regularizer $\mathcal{L}_O = \sum_i^N \sum_{\mathcal{M}_a,\mathcal{M}_b} ReLU(-d_{\mathcal{M}_a}(p_i)) \times ReLU(-d_{\mathcal{M}_b}(p_i))$ penalizes overlaps between any pair of meshes $\mathcal{M}_a, \mathcal{M}_b$ (*i.e.*, sampling points with negative signed distances from multiple meshes). The variance regularizer $\mathcal{L}_V = \sum_m ||\Lambda^{0.5}(U_m - U_{m+1})||_2^2$ assimilates the shapes of the neighboring vertebra.

Our final optimization loss is defined as $\mathcal{L} = \mathcal{L}_\mathcal{I} + \lambda_E\mathcal{L}_E + \lambda_N\mathcal{L}_N + \lambda_O\mathcal{L}_O + \lambda_V\mathcal{L}_V$, where $\lambda_{\{E,N,O,V\}}$ are the regularization weights for the corresponding regularizer. In our experiments, $\lambda_E = 0.02, \lambda_N = 0.1, \lambda_O = 1, \lambda_V = 0.02, M = 10{,}000$ is used.

### 2.5 Optimization process

**Point sampling** On every iteration, points are sampled near the boundary of the spine model as follows. First, the $N$ points are uniformly sampled on the model surface. Then, the normal distribution noise $\epsilon \sim N(0, \sigma^2)$ is added to the sampled points. In our experiments, $N = 10{,}000$, $\sigma = 9.6mm$ is used.




(a) (b) (c)

Figure 2: (a) Box plot of Dice score by vertebral label in VerSe'20 training dataset. Semi-transparent dots denote instances, and red diamonds denote the average scores of each vertebral label. (b) Selected segmentation results and their correspondence error ($mm$) from cervical, thoracic, and lumbar segments, from top to bottom. (c) Box plot of the Dice score by vertebral segments, performed by our framework and ablated variants. -I denotes our framework without image loss, -N for without normal loss, and -E for without edge length loss. For each variant, changes in the average Dice score are marked below the label, red and green for a decrease and an increase in performance, respectively.

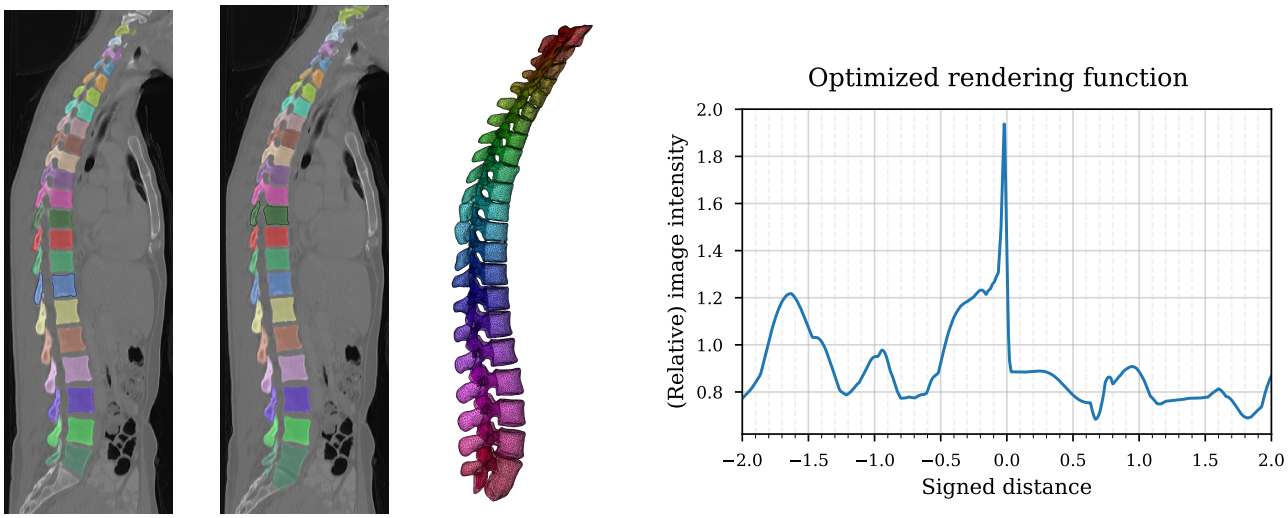
**Optimizer** For the rendering function $f_\theta$, the Adam optimizer[16] with $(\texttt{lr}, \beta_1, \beta_2) = (0.01, 0.9, 0.9)$ is used. For affine and spectral deformations, a vector-wise normalization variant[17,18] of the Adam optimizer with the same hyperparameters is instead used to reduce the artifacts aligned to the grid.[17–19] For any positional values, *e.g.* the voxel size or the vertex position, we used $64mm$ as a unit of length to match the scale of values similar to the weights of the neural network and the affine transformation matrices. No weight decay is used for both optimizers. In our experiments, a fixed number of optimizer iterations of 3,000 is used.

**Coarse-to-fine optimization** In addition to spectral mesh optimization, we encourage the optimization process to align shapes in a coarse-to-fine manner by blurring the target image in the early stages and slowly decreasing the blur radius as the iteration progresses. In our experiments, the initial blur radius $\sigma_0$ is set to $2.56mm$, and the blur radius is linearly decreased until optimization is 80% done, which is when the blur radius becomes 0 and the original target image is used instead.

## 3. EXPERIMENTS AND RESULTS

### 3.1 Experiments Setup

We used VerSe'20[20] training dataset, which is preprocessed to have an isometric voxel size of $1mm^3$. For initial spine alignment, we used centroids in the dataset. We designed the spine template model by stacking a single vertebral mesh shown at top left of Figure 1, which is constructed from the average shape of xVertSeg[21] training dataset. We deliberately chose the lumbar vertebra as the shared vertebral template model, to show the performance differences by shape similarities between the lumbar-based template models and the cervical, thoracic, and lumbar vertebrae in the image.




(a) (b) (c) (d)

Figure 3: (a) A sample of CT spine image (sagittal slice with GT segmentation masks overlayed) from VerSe'20 dataset. (b) Segmentation output from our framework. (c) Optimized vertebral mesh. (d) Optimized rendering function $f_\theta$.

Table 1: Average Dice score by vertebral segments, performed by our framework and ablated variants.

| | Ours | -I | -N | -E |
|---|---|---|---|---|
| Cervical | 63.3186 | 30.8736 | 58.2077 | **63.4729** |
| Thoracic | **75.8013** | 44.7484 | 67.8749 | 75.6270 |
| Lumbar | **86.3177** | 56.7357 | 79.8192 | 85.8110 |
| Total | **77.3202** | 46.5200 | 70.2868 | 77.0879 |

### 3.2 Results

**Vertebral segmentation** We measured the Dice score of the aligned spine model and the ground truth labels. The result is shown in Figure 2a and Figure 2b. Our framework took on average one hour per image. The lumbar vertebrae, especially from L1 to L4, show up to 90% Dice score, higher than the other two segments, as they have a shape similar to the template model. Within a vertebra, our framework identified the correspondence more accurately for vertebral bodies than vertebral arches due to the more complex structures in vertebral arches. We expect that the result would be more accurate if the spine model were constructed from multiple label-specific meshes.

Figure 3 depicts the output mesh and the optimized rendering function for a selected CT spine image from the VerSe'20 dataset. Note that our framework can correctly learn the sharp appearance of bone-tissue boundary on the zero-isosurface due to Fourier feature encoding and the normal regularizer. Although it shows great results on the lumbar segments, our framework produces a noticeable error near the first or last vertebra (the L5 vertebra shown in green in Figure 3b) or the cervical segments which have highly divergent shapes from the template.

**Ablation study** We ran the same segmentation task with variants of our framework to see the effects of each loss term. The result, described in Figure 2c and Table 1, shows that the image similarity loss (Equation (2)) is critical for shape alignment, as it can provide volumetric visibility of the target image compared to the normal regularizer, which only provides rather sparse surface visibility. The normal regularizer also has a considerable impact on performance because it resolves the ambiguous zero-isosurface of the signed distance for the rendering function $f_\theta$. The edge length regularizer, compared to the previous two regularizers, shows little impact on the

Dice score, which measures the similarity of two volumetric regions. However, it affects the overall quality of the shape and the surface correspondence between the aligned shape and the ground truth.

## 4. CONCLUSION

In this paper, we introduced a differentiable appearance modeling of a deformable spine template, a *dataset-free* framework to vertebral segmentation. Exploiting the repetitive structure and high-contrast boundary of the spine, our framework can align a spine template model onto a given CT image by jointly optimizing the appearance and shape of the model. Our experimental results show that explicit knowledge from a well-tailored template model can significantly reduce the need for a large training dataset in segmentation tasks. We believe that our framework can bootstrap the segmentation task with little to no dataset.

In future work, we will explore various differentiable operators to allow the rendering function to utilize richer information beyond the signed distance. Additionally, because our framework can generate the appearance of a given model, we expect that it can also perform image synthesis when combined with a more sophisticated rendering function.

## ACKNOWLEDGMENTS

This work was supported by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korean government (MSIT) (No.00223446, Development of object-oriented synthetic data generation and evaluation methods).

## REFERENCES

[1] Mastmeyer, A., Engelke, K., Fuchs, C., and Kalender, W. A., "A hierarchical 3d segmentation method and the definition of vertebral body coordinate systems for qct of the lumbar spine," *Medical image analysis* **10**(4), 560–577 (2006).

[2] Aslan, M. S., Ali, A., Rara, H., Arnold, B., Farag, A. A., Fahmi, R., and Xiang, P., "A novel 3d segmentation of vertebral bones from volumetric ct images using graph cuts," in [*Advances in Visual Computing: 5th International Symposium, ISVC 2009, Las Vegas, NV, USA, November 30-December 2, 2009. Proceedings, Part II 5*], 519–528, Springer (2009).

[3] Lim, P. H., Bagci, U., and Bai, L., "Introducing willmore flow into level set segmentation of spinal vertebrae," *IEEE Transactions on Biomedical Engineering* **60**(1), 115–122 (2012).

[4] Huang, J., Jian, F., Wu, H., and Li, H., "An improved level set method for vertebra ct image segmentation," *Biomedical engineering online* **12**(1), 1–16 (2013).

[5] Chen, H., Shen, C., Qin, J., Ni, D., Shi, L., Cheng, J. C., and Heng, P.-A., "Automatic localization and identification of vertebrae in spine ct via a joint learning model with deep neural networks," in [*Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part I 18*], 515–522, Springer (2015).

[6] Cheng, P., Yang, Y., Yu, H., and He, Y., "Automatic vertebrae localization and segmentation in ct with a two-stage dense-u-net," *Scientific Reports* **11**(1), 22156 (2021).

[7] McInerney, T. and Terzopoulos, D., "Deformable models in medical image analysis," in [*Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis*], 171–180 (1996).

[8] Jayadevappa, D., Kumar, S. S., and Murty, D. S., "Medical image segmentation algorithms using deformable models: A review," *IETE Technical Review* **28**(3), 248–255 (2011).

[9] Rasoulian, A., Rohling, R., and Abolmaesumi, P., "Lumbar spine segmentation using a statistical multi-vertebrae anatomical shape+ pose model," *IEEE transactions on medical imaging* **32**(10), 1890–1900 (2013).

[10] Forsberg, D., "Atlas-based registration for accurate segmentation of thoracic and lumbar vertebrae in ct data," *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging* , 49–59 (2015).

[11] Kato, H., Beker, D., Morariu, M., Ando, T., Matsuoka, T., Kehl, W., and Gaidon, A., "Differentiable rendering: A survey," *arXiv preprint arXiv:2006.12057* (2020).

[12] Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., and Ng, R., "Fourier features let networks learn high frequency functions in low dimensional domains," *NeurIPS* (2020).

[13] Barill, G., Dickson, N., Schmidt, R., Levin, D. I., and Jacobson, A., "Fast winding numbers for soups and clouds," *ACM Transactions on Graphics* (2018).

[14] Barnes, J. and Hut, P., "A hierarchical o (n log n) force-calculation algorithm," *nature* **324**(6096), 446–449 (1986).

[15] Lévy, B. and Zhang, H. R., "Spectral mesh processing," in [*ACM SIGGRAPH 2010 Courses*], *SIGGRAPH '10*, Association for Computing Machinery, New York, NY, USA (2010).

[16] Kingma, D. and Ba, J., "Adam: A method for stochastic optimization," in [*International Conference on Learning Representations (ICLR)*], (2015).

[17] Palfinger, W., "Continuous remeshing for inverse rendering," *Computer Animation and Virtual Worlds* **33**(5), e2101 (2022).

[18] Ling, S., Sharp, N., and Jacobson, A., "Vectoradam for rotation equivariant geometry optimization," *arXiv preprint arXiv:2205.13599* (2022).

[19] Nicolet, B., Jacobson, A., and Jakob, W., "Large steps in inverse rendering of geometry," *ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia)* **40** (Dec. 2021).

[20] Sekuboyina, A., Husseini, M. E., Bayat, A., Löffler, M., Liebl, H., Li, H., Tetteh, G., Kukačka, J., Payer, C., Štern, D., et al., "Verse: a vertebrae labelling and segmentation benchmark for multi-detector ct images," *Medical image analysis* **73**, 102166 (2021).

[21] Korez, R., Ibragimov, B., Likar, B., Pernuš, F., and Vrtovec, T., "A framework for automated spine and vertebrae interpolation-based detection and model-based segmentation," *IEEE transactions on medical imaging* **34**(8), 1649–1662 (2015).