LPM: 3D Lungs Reconstruction using Lungs Parametric Model

***Abstract*—Deep learning techniques have been proven fruitful for many researchers in the computer vision field for problems like image segmentation and object recognition. With the success of this, 3D reconstruction has been of great interest, and how Deep Learning can be used to implement the same. There have been plenty of successfully developed models and methods that take Computer Tomography (CT) scans as an input, then pre-processed to obtain the specific organ (lungs, liver, heart, etc.) and offer a stacked 3D representation of the organ as output. This paper reviews some of the recent papers that offer a different approach on how a 3D object can be constructed from an X-ray image alone, along with the challenges that come with such. This paper will provide insights into the development of a model capable of reconstructing 3D lung structures from point cloud data, utilizing real-world datasets. Additionally, it will explore the application of latent code for generating a broader dataset of lung models. This approach has the potential to aid in identifying the size and shape of lungs directly from X-ray images.**

***Index Terms*—Deep Learning, 3D Reconstruction, Single Image, Parametric Model, X-Ray**

I. Introduction

3D reconstruction from 2D images is a pivotal technique in computer vision, enabling the transformation of flat images into detailed three-dimensional models. This process is fundamental in various applications, including medical imaging, computer graphics, and autonomous navigation.

Inferring 3D structures from 2D images is an ill-posed inverse problem, often complicated by projection ambiguities and depth compression. Deep learning approaches, such as Pixel2Mesh [1], StructureNet [2], and PointNet [3], have been developed to generate 3D representations, though they frequently require significant post-processing. X-rays produce grayscale projections that lack surface color and texture, making depth perception difficult. Furthermore, X-ray images are 2D projections that compress depth information along the beam's path [4], necessitating advanced methodologies for accurate reconstruction.

This study addresses these gaps through advanced techniques and demonstrates the development of a neural network model capable of reconstructing 3D models from point cloud data derived from real lung dataset meshes.

II. Methodology

A. Problem Statement

The objective of this research is to be able to model the Signed Distance Function (SDF) for 3D lung surface reconstruction, with incorporation of point-cloud data from CT scans. A neural network is trained for prediction of the signed distance of any given point from the lung surface. In addition, latent code is to be implemented for encoding variability across multiple lung datasets.

B. Deep Neural Networks (DNNs)

DNNs are multi-layered models that learn hierarchical data representations, enabling them to capture complex patterns for tasks like classification and regression [5].

C. Latent Code

A latent vector $\mathbf{z} \in \mathbb{R}^d$ encodes individual shape geometry. Concatenated with spatial coordinates, it conditions the Signed Distance Field (SDF) predictions to reflect shape-specific features [6].

D. Auto-Decoder Framework

Each lung shape is associated with a unique latent code. An SDF network, acting as an auto-decoder, maps this code and spatial coordinates to SDF values, reconstructing the 3D lung surface. Training optimizes both network parameters and latent codes to model anatomical variations.

E. Signed Distance Function (SDF)

The SDF assigns each spatial point a scalar value indicating its distance to the object's surface: negative inside, positive outside, and zero on the surface. This continuous, differentiable representation is ideal for shape modeling and reconstruction [7].

F. Dataset Preparation

Raw Data: CT scans from the 'COVID-19 CT Lung and Infection Segmentation' dataset are normalized for consistency [8].

Mesh Conversion: NIfTI-formatted scans are converted to smooth meshes using adaptive Gaussian filtering for noise reduction [9] and Laplacian smoothing for refinement [10].

Point Sampling: Surface and non-surface points are sampled from the meshes. Surface points are assigned normals and zero SDF values; non-surface points have computed signed distances. Samples are stored in NPZ files.

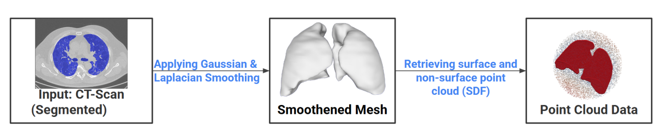

Fig. 1. Dataset generation framework. CT scan image from the COVID-19 CT Scan dataset [11].

G. Latent Code Assignment

During data loading, each lung is assigned a trainable latent vector to capture its unique geometry.

H. Marching Cubes Algorithm

Marching Cubes extracts polygonal meshes from volumetric data by evaluating scalar values at cube vertices, determining surface intersections, and constructing corresponding triangles to approximate the object's surface [12].

III. Model Architecture

A. Network Design

The architecture employs a Deep Neural Network (DNN) to predict Signed Distance Function (SDF) values, integrating a latent code to condition outputs for various 3D lung shapes. Key components include:

1) Input: The network accepts 3D point coordinates concatenated with a latent vector $\mathbf{z} \in \mathbb{R}^{256}$, representing shape-specific characteristics.

2) Hidden Layers: Two fully connected layers with ReLU activations [13] learn abstract representations, modeling complex relationships between spatial points and SDF values.

3) Output Layer: A final fully connected layer outputs a single SDF value per input point, with Tanh activation ensuring values lie within $[-1, 1]$.

4) Latent Code Integration: During training, the latent code is treated as a learned parameter, enabling the network to generate distinct surface shapes conditioned on this code.

B. Objective

The model aims to predict SDF values for 3D points using both the latent code and point coordinates. The loss function comprises:

1) SDF Surface Loss: Ensures predicted SDF for surface points remains close to zero:

$$\mathcal{L}_{\text{surface}} = \frac{1}{N} \sum_{i=1}^{N} \left| S\hat{D}F_i \right| \tag{1}$$

2) SDF Non-Surface Loss: Encourages accurate SDF prediction for non-surface points using L_1 norm:

$$\mathcal{L}_{\text{non-surface}} = \frac{1}{N} \sum_{i=1}^{N} \left| S\hat{D}F_i - SDF_i \right| \tag{2}$$

3) Eikonal Loss: Promotes smoothness and valid distance fields by enforcing the eikonal equation:

$$\mathcal{L}_{\text{eikonal}} = \frac{1}{N} \sum_{i=1}^{N} \left| \|\nabla S\hat{D}F_i\| - 1 \right| \tag{3}$$

C. Surface Extraction

Post-training, the Marching Cubes algorithm is employed to extract 3D surfaces from the predicted SDF field:

1) Grid Construction: A 3D grid is established over the region of interest.

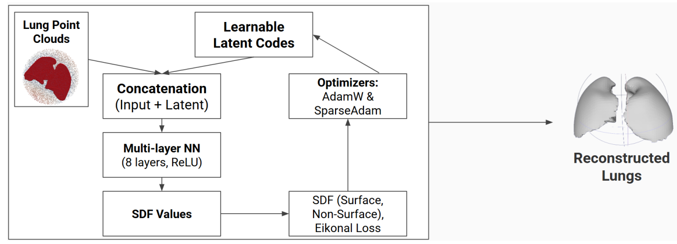

Fig. 2. Model Architecture: The network processes point cloud data with a latent dimension of 256 and 8 neural network layers to predict SDF values for reconstruction.

2) SDF Evaluation: The trained model evaluates SDF at each grid point.

3) Marching Cubes: Processes each cube in the grid, identifying surface intersections and generating a triangular mesh representing the 3D surface of the lungs.

D. Testing the Model via SDF Ground Truth

For evaluation, unseen datasets with X-ray images as primary input are used to assess the model's ability to reconstruct accurate 3D lung meshes:

1) Freezing Model Parameters: All network parameters are frozen to preserve learned features, while the latent code remains trainable to adapt to new lung shapes.

2) Partial Point Cloud Sampling: Reduces the number of sampled points to test the model's generalization with limited input data.

3) Evaluation Procedure: Performance is assessed both quantitatively and qualitatively:

a) Quantitative Evaluation: Segmentation accuracy is measured using the F-Score:

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

where,

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

and the Chamfer Distance:

$$D_{\text{Chamfer}}(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \|a - b\|_2^2 + \frac{1}{|B|} \sum_{b \in B} \min_{a \in A} \|a - b\|_2^2 \tag{7}$$

with $|A|$ and $|B|$ denoting the number of points in sets A and B, respectively.

b) Qualitative Evaluation: Visual inspections of reconstructed meshes are conducted to assess anatomical accuracy and structural detail compared to CT-derived references.

IV. Results

A. Dataset Preparation

Segmented lung CT scans were processed into meshes using an Adaptive Gaussian Filter [9], followed by a Laplacian Filter

Fig. 3. Testing Phase: Unseen data is input into the pre-trained model to evaluate reconstruction performance.

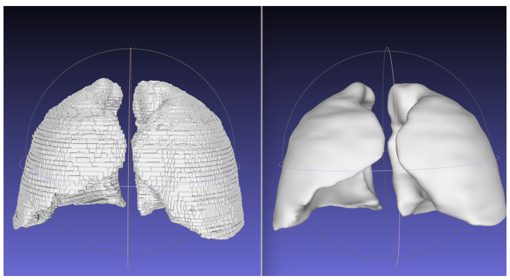

Fig. 4. Left: Direct NifTi to mesh conversion. Right: Application of Adaptive Gaussian and Laplacian Filters for refined lung meshes.

[10] to enhance edge definition and reduce pixelation. This yielded smoother lung meshes, as illustrated in Fig. 4.

Using the Mesh to SDF library [14], point clouds were sampled from these meshes, separating surface and non-surface points into distinct `.npz` files. The resulting dataset, depicted in Fig. 5, served as input for model training.

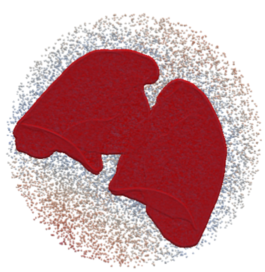

Fig. 5. Extracted point cloud data: red indicates surface points; a gradient from blue to warm red represents non-surface points based on distance.

B. Model Training and Evaluation

The SDF Latent Model was trained on 75% of the dataset, inspired by Neural Parametric Head Models (NPHM) [15]. Initial configurations included a 4-layer network with 512 neurons per layer and a latent dimension of 64. However, this setup resulted in suboptimal accuracy and merging issues in reconstructed lung regions (Fig. 6).

Subsequently, the network was expanded to 8 layers with 1024 neurons each and a latent dimension of 256. This enhancement significantly improved the model's capacity to learn complex structures, yielding more accurate lung mesh reconstructions (Fig. 7).

An attempt to incorporate normal loss alongside SDF and Eikonal losses did not yield satisfactory results, as shown in Fig. 8. Consequently, the final model excluded normal loss.

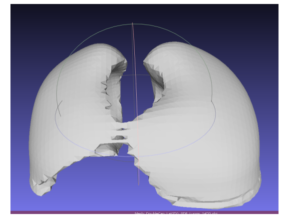

Fig. 6. Initial reconstruction: Latent dimension 64, 4 layers, 512 neurons per layer.

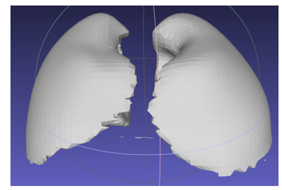

Fig. 7. Enhanced reconstruction: Latent dimension 256, 8 layers, 1024 neurons per layer.

Training loss metrics are presented in Fig. 12, showcasing overall loss, SDF loss (surface and non-surface), and Eikonal and latent regularization losses.

C. Model Testing on Unseen Data

The remaining 25% of the dataset, comprising only surface points, was used to evaluate the model's performance on unseen data. To maintain the integrity of the trained network, all model parameters were frozen, allowing only the latent code to adapt during testing.

From an initial set of 250,000 surface points, a subset of 5,000 points was sampled to create partial point cloud inputs. The pre-trained model successfully reconstructed lung meshes without merging issues, although there is potential for further improvement in reconstruction quality.

Quantitative evaluation metrics included the F-score and Chamfer Distance [16], assessing the accuracy of lung region segmentation.

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{6}$$

$$D_{\text{Chamfer}}(A, B) = \frac{1}{|A|} \sum_{a \in A} \min_{b \in B} \|a - b\|_2^2 + \frac{1}{|B|} \sum_{b \in B} \min_{a \in A} \|a - b\|_2^2 \tag{7}$$

An example reconstruction from unseen data is presented in Fig. 13, demonstrating the model's capability to generalize to new inputs.

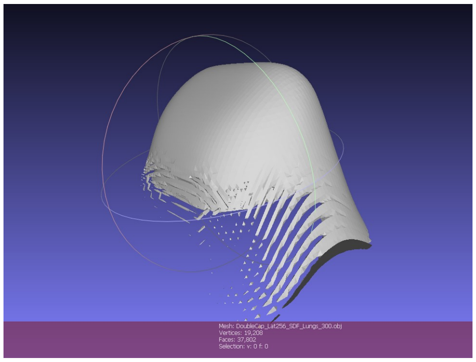

Fig. 8. Unsuccessful reconstruction with the inclusion of normal loss.

Fig. 9. Overall training loss.

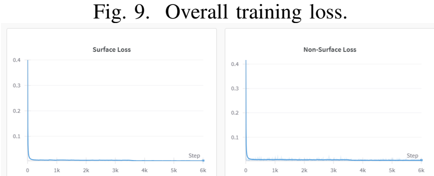

Fig. 10. SDF loss (surface and non-surface).

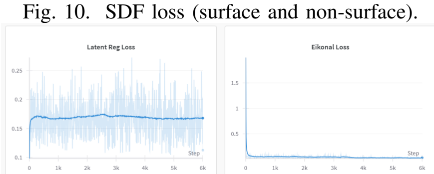

Fig. 11. Eikonal and latent regularization losses.

Fig. 12. Training loss metrics.

TABLE I
LOSS VALUES FOR INITIAL AND FINAL TRAINING PHASES WITH HIDDEN SIZE

Training Phase	Latent Dim	N-Layers	Hidden Size	Loss
Initial Training	64	4	512	0.2196
Final Training	256	8	1024	0.0225

TABLE II
F-SCORE AND CHAMFER DISTANCE METRICS FOR THE MODEL IN THE TESTING PHASE

Metric	Model Value	Acceptable Value
F-score	0.7047/0.8235	approx 0.80
Chamfer Distance	0.015/0.016	approx 0.02

V. CONCLUSION AND FUTURE WORK

This study developed a pipeline for anatomical shape generation, specifically lung segmentation from CT scans,

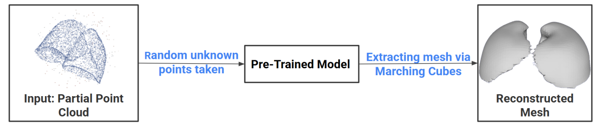

Fig. 13. Reconstruction from unseen lung data using only surface point cloud input.

Fig. 14. Here, only 5000 surface point cloud is sampled out of 250,000 points in order to test the robustness of my model to identify crucial parts

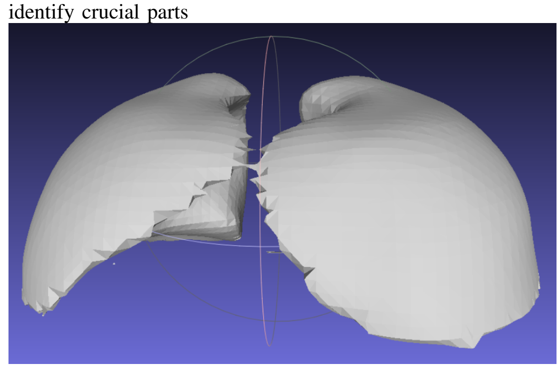

Fig. 15. The reconstructed lungs of the same unseen dataset, when applied into the pre-trained model and extracted via Marching Cubes

using mesh and point cloud data. The scalable Latent SDF model showed improved performance with increased latent dimension and network layers, achieving a final loss of 0.0225 with specific hyperparameters. Overfitting and computational resource demands increased with model complexity. Future work will focus on enhancing interpretability via semantic latent codes or conditional models, exploring alternative input functions, and multi-modal data. Generalizability will be assessed on other anatomical regions and larger datasets, adapting the architecture accordingly.

REFERENCES

[1] N. Wang, Y. Zhang, Z. Li, Y. Fu, W. Liu, and Y.-G. Jiang, "Pixel2mesh: Generating 3d mesh models from single rgb images," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 52–67.

[2] C. B. Choy, D. Xu, J. Gwak, K. Chen, and S. Savarese, "3d-r2n2: A unified approach for single and multi-view 3d object reconstruction," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VIII 14*. Springer, 2016, pp. 628–644.

[3] H. Fan, H. Su, and L. J. Guibas, "A point set generation network for 3d object reconstruction from a single image," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 605–613.

[4] G. T. Herman and G. T. Herman, "Physical problems associated with data collection in ct," *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, pp. 37–52, 2009.

[5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[6] R. Girdhar and D. Ramanan, "Attentional pooling for 3d shape recognition," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[7] Z. Liao and et al., "Deepsdf: Learning continuous signed distance functions for shape representation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[8] M. Jun, G. Cheng, W. Yixin, A. Xingle, G. Jiantao, Y. Ziqi, Z. Minqing, L. Xin, D. Xueyuan, C. Shucheng *et al.*, "Covid-19 ct lung and infection segmentation dataset," 2020.

[9] H. Wang and L. Xie, "Adaptive gaussian filtering and its applications," *Proceedings of the International Conference on Image Processing*, 2009.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd ed.)*. Pearson, 2008.

[11] D. Negroni, D. Zagaria, A. Paladini, Z. Falaschi, A. Arcoraci, M. Barini, and A. Carriero, "Covid-19 ct scan lung segmentation: how we do it," *Journal of Digital Imaging*, vol. 35, no. 3, pp. 424–431, 2022.

[12] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *Seminal graphics: pioneering efforts that shaped the field*, 1998, pp. 347–353.

[13] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

[14] T. Meshing, "Mesh to sdf: Converting meshes to signed distance functions," 2018. [Online]. Available: https://github.com/xyz/mesh-to-sdf

[15] A. Ranjan, M. Sela, J. Romero, X. Zhu, and M. J. Black, "A unified framework for human face reconstruction and animation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017.

[16] X. Zhang, L. Xie, X. Wei, and Z. Li, "Deep chamfer distance for learning 3d shape matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.