A MedShapeNet Foundation Model
Learning-Based Multimodal Medical Point Cloud Completion

Ahmed Yassin[1], Gijs Luijten[1,2,3], Abdelrahman Elsakka[1], André Ferreira[1,4],
Behrus Puldai[5,6], Victor Alves[4], and Jan Egger[1,2,3,7]

[1] Institute for AI in Medicine (IKIM), University Hospital Essen, Essen, Germany
[2] Center for Virtual and Extended Reality in Medicine (ZvRM), University Medicine Essen, Essen, Germany
[3] Institute of Computer Graphics and Vision (ICG), Graz University of Technology, Graz, Austria
[4] Center Algoritmi / LASI, University of Minho, Braga, Portugal
[5] Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany
[6] Department of Oral and Maxillofacial Surgery, University Hospital RWTH Aachen, Aachen, Germany
[7] Cancer Research Center Cologne Essen (CCCE), University Medicine Essen (AöR), Essen, Germany

Abstract. Point clouds provide an efficient and natural representation of 3D anatomy, capturing fine details of complex structures with minimal computational overhead. However, most learning-based approaches in medical imaging focus on 2D and 3D gridded data, disregarding the potential of point clouds. In this paper, we present a multimodal foundation model for medical point cloud completion, trained on the MedShapeNet dataset using over 200,000 shapes across 240+ anatomical classes. Our transformer-based autoencoder, combined with a BERT text encoder, is designed to take multiple variations of incomplete point clouds and map them to their corresponding complete shapes. This allows the model to consolidate different incomplete representations of an object into a single, fully reconstructed version. The model leverages Density-Aware Chamfer Distance loss to handle varying point densities effectively. The results demonstrate strong performance across a wide range of anatomical categories, establishing this foundation model as a valuable tool for future research and downstream tasks. The source code for this contribution is available at:
`https://github.com/jrHoss/MedShapeNet-Foundation-Model`

Keywords: Point Cloud Completion · Transformer · Autoencoder · Multimodal Learning · Density-Aware Chamfer Distance

1 Introduction

Over the past several decades, technological advancements have significantly shaped healthcare, with automated and computer-assisted medical image analysis becoming indispensable tools in clinical practice. These innovations have not

only enhanced diagnostic accuracy, but also streamlined workflows in preoperative planning, intraoperative guidance, and postoperative monitoring. Recently, deep learning (DL) techniques have revolutionized medical image processing and analysis. From disease detection to personalized treatment planning, these methods have consistently set new benchmarks in healthcare, and their impact continues to grow as the technology evolves [8, 11, 25, 27, 30].

Medical imaging data, traditionally in the form of 2D and 3D images such as X-rays, ultrasound, MRI, and CT, can require significant computational resources and pose challenges in terms of memory usage and spatial resolution [39]. Research has primarily focused on processing this dense data [3], but point clouds—collections of 3D data points representing anatomical structures—offer a compelling alternative due to their efficiency [24].

Point clouds provide an efficient representation of 3D anatomy, capturing fine details with minimal computational overhead [39]. Compared to traditional grid-based approaches, they enable faster processing and better scalability [39]. However, challenges such as order invariance persist [43], especially when dealing with medical data, where imaging artifacts and segmentation errors introduce gaps that complicate reconstruction. Existing point cloud completion methods, developed for general 3D objects, often fail to address the complexity of medical datasets [37, 43] .

This work presents a multimodal foundation model aimed at solving the challenge of medical point cloud completion. The model is trained using the MedShapeNet dataset [22] and employs a transformer-based autoencoder (AE) to reconstruct a wide variety of anatomical structures. By incorporating multiple data modalities, the model ensures accurate completion of incomplete point clouds [16].

Our results demonstrate that the model generalizes effectively across diverse anatomical classes, as confirmed by both quantitative and qualitative evaluations. This model can be fine-tuned for future research and downstream tasks, with potential applications in medical education, extended reality (XR) for training, and custom bone implant design.

2 Related Work

Point cloud completion techniques can be divided into two main categories: Traditional and data-driven approaches. Traditional methods rely on established mathematical and geometric principles, using techniques like interpolation, extrapolation, or geometric fitting. While effective in many cases, these approaches often face limitations when dealing with complex shapes [33]. In contrast, data-driven methods utilize machine learning to learn from large datasets, capturing intricate patterns and structures. These methods typically outperform traditional approaches in handling complex geometries, offering more accurate and realistic completions [15, 43].

2.1 Traditional Approaches

Various techniques exist for optimizing the parameters of a shape model in traditional approaches. These include methods based on interpolation [5, 17, 19, 45], matching patches [4, 9, 32], alignment [9, 23, 28, 31], and geometric principles [6, 13, 18, 21]. Each of these strategies addresses different aspects of shape modeling, with interpolation handling missing data, matching-patches methods comparing local regions, alignment methods focusing on spatial transformations, and geometric-based approaches emphasizing the underlying structure of the shape.

2.2 Learning-based Approaches

Learning-based approaches utilize data-driven parameterized models, often DL networks, to map incomplete inputs to full shapes, offering rapid inference and strong generalization. These models can encode the geometric structure of point clouds in high-dimensional spaces and do not assume any specific object shape unless trained to do so. Learning-based methods are categorized into fully supervised, semi-supervised, and unsupervised techniques [38]. Fully supervised approaches dominate the field, requiring extensive training data and computational resources [15, 33, 36].

Several DL models have been applied to point cloud analysis in the medical field, though this area is still emerging compared to other applications like autonomous driving or 3D object recognition. Some of the notable models used for medical point clouds include [15]:

– **"Multi-class point cloud completion networks for 3D cardiac anatomy reconstruction from cine magnetic resonance images"**: Introduces a fully automated pipeline for reconstructing 3D cardiac anatomy from cine MRI scans. The method addresses the challenges of sparse and misaligned data by leveraging a point cloud completion network (PCCN) to create complete point clouds of cardiac structures [42].

– **3D Medical Point Transformer (3DMedPT)**: Presents an attention-
  based DL model designed specifically for medical point clouds, which rep-
  resent 3D biological structures. The model focuses on capturing both local
  context and global feature interactions to analyze complex medical data. It
  incorporates position embeddings to handle limited training data by learn-
  ing accurate local geometries, and it uses Multi-Graph Reasoning (MGR) to
  propagate global knowledge across channel graphs, enriching feature repre-
  sentations [2].
– **Neural Shape Completion for Personalized Maxillofacial Surgery**:
  Investigates the use of neural networks for reconstructing and completing 3D
  shapes in the context of maxillofacial surgery. The goal is to create personal-
  ized, accurate models for use in surgical planning by reconstructing missing
  or incomplete parts of anatomical structures, such as facial bones. This work
  focuses on developing shape completion techniques that can aid clinicians in
  planning complex surgeries by generating complete and realistic 3D models
  from partial or damaged scans [26].

3   Methods

3.1   Problem Formulation

Shape completion from sparse and partial point clouds involves training the AE
to map multiple incomplete instances to their corresponding complete ground
truth (GT) shapes. Let each GT shape be denoted by y_n, where n indexes
the specific objects in the dataset. For each shape y_n, there exist several in-
complete instances, x_n^m, where m indicates different partial instances. The goal
of the mapping function F is to transform the set of incomplete observations,
$x_n^1, x_n^2, \ldots, x_n^M$, into the corresponding complete shape, y_n. This process is de-
picted in Figure 1, which visually represents the transformation, and is mathe-
matically formalized in Equation 1.

$$F : x_n^m \to y_n, \quad n = 1, 2, \ldots, N \quad m = 1, 2, \ldots, M \tag{1}$$

Fig. 1. The first point cloud represents the complete structure, while the second, third,
and fourth point clouds are three different incomplete instances, each missing different
regions.

3.2   Dataset

In this work, we utilize the MedShapeNet dataset [22], a comprehensive collection of 3D medical shapes representing various anatomical structures such as bones, organs, and other body parts. The dataset contains a total of 114,075 instances spanning over 240 unique anatomical classes. This extensive dataset is specifically designed for medical shape analysis tasks such as 3D shape completion, segmentation, and classification, making it an ideal resource for training models that generalize across different anatomical structures. Through augmentation techniques, which are explained in detail in subsection 3.3, the dataset is doubled from 104,807 to 209,614 instances. The augmented dataset is then split into two subsets: 90% for training (188,652), and 10% for validation (20,962), ensuring a robust evaluation.
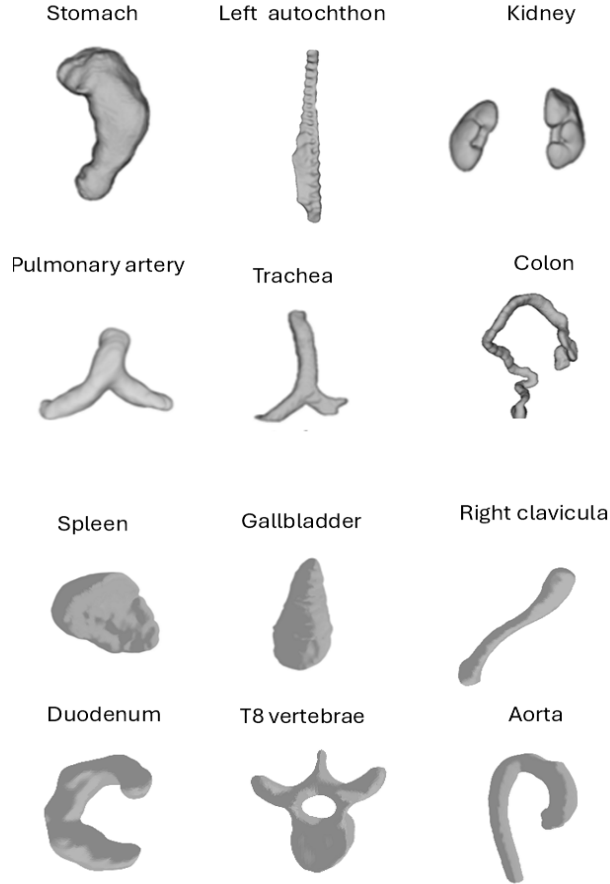
Fig. 2. Examples of anatomical shapes from the MedShapeNet dataset [22].

3.3   Pre-processing

To create the point cloud dataset, each mesh from the MedShapeNet dataset was sampled to generate corresponding point clouds. Specifically, 6,144 points were uniformly sampled from each mesh to represent the GT point clouds. This consistent sampling process ensures that all shapes are accurately represented in a format well-suited for point cloud-based tasks, maintaining both accuracy and uniformity across the dataset for effective processing and analysis.

Each point cloud, denoted as $P = \{p_i\}_{i=1}^{K}$, consists of K points, where each point $p_i = (x_i, y_i, z_i)$ is normalized so that its coordinates lie within the range $\{-1, 1\}$. The normalization is performed by subtracting the centroid to align the point cloud with the origin $(0, 0, 0)$. Additionally, the point cloud is scaled by a factor L, defined as:

$$L = \max_{p_i \in P} \left(\sqrt{x_i^2 + y_i^2 + z_i^2} \right) \tag{2}$$

This normalization and scaling process ensures consistency across the dataset.

3.4   Defect injection

Since the dataset contains only full meshes, partial point clouds were synthetically generated from the GT point clouds using a two-step process:

1. **Farthest Point Sampling (FPS)** [10]: This algorithm was applied to the GT point cloud, which consisted of 6,144 points, to sample two points that are the farthest apart from each other. These points were chosen to ensure spatially distinct regions in the point cloud.
2. **K-Nearest Neighbor (KNN)** [12]: Using the indices of the farthest points obtained from FPS, the KNN algorithm was applied to remove each of these points along with the 2,047 nearest neighboring points. This guarantees input point clouds with different missing regions, leaving a partial point cloud with 4,096 points.

This approach ensured that each instance in the GT dataset had two corresponding partial point clouds as input, each with 4096 points as shown in Figure 3. Consequently, the total number of instances in the point cloud dataset is doubled from 104,807 to 209,614 instances.
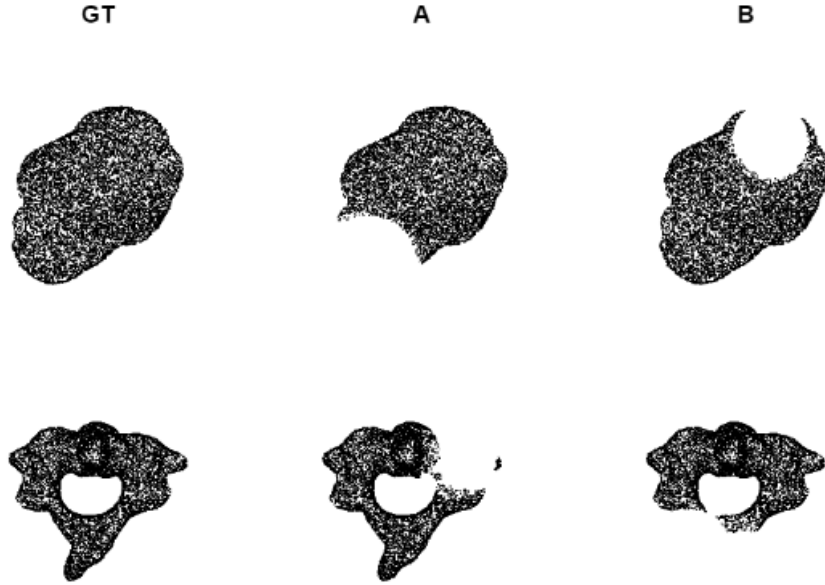
Fig. 3. Examples of GT and inputs from the dataset. The first column shows the GT, the second and third columns display two incomplete examples. Each row corresponds to a different organ.

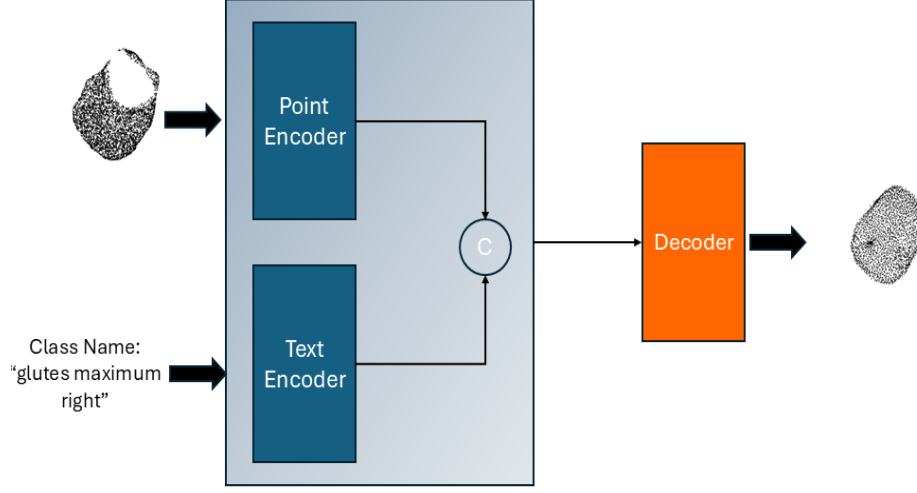### 3.5   Transformer-Based Autoencoder

As shown in the previous section, the aim of the model is to learn the function F such that:

$$\hat{y}_n^m = F(x_n^m) \approx y_n, \quad n = 1, 2, \ldots, N \quad m = 1, 2, \ldots, M \tag{3}$$

where x_n^m are the incomplete instances, y_n represent the complete point clouds, and $\hat{y}_n^m$ represent the model predictions.

Multi-Modality: Integrating text and point cloud modalities in deep learning significantly enhances point cloud completion by enriching semantic understanding and improving reconstruction accuracy. Text data provides contextual descriptions that help resolve ambiguities in geometric data, enabling more precise reconstruction of missing parts [1]. For instance, in medical imaging, textual descriptions can clarify differences between a healthy liver and one with tumors, aiding in more accurate and context-aware 3D reconstructions [7, 29, 35].

In our model, the global feature vectors from both the text and point cloud encoders are concatenated to form a unified representation. This combination allows the model to leverage information from both modalities for more accurate and contextually rich point cloud reconstructions, as shown in Figure 4."

**Fig. 4.** Illustration of the multimodal foundation model.

Encoder: The transformer-based AE's encoder is adapted from the Point Cloud Transformer (PCT) [14] to meet the specific needs of our task, as shown in Figure 5. It encodes point clouds into a higher-dimensional feature space, capturing the semantic relationships between points, which improves point cloud processing across various tasks. Initially, the PCT encoder embeds the input coordinates into a new feature space. The embedded features then pass through four consecutive attention modules, refining the representation of each point to make it more detailed and discriminative. Finally, a linear layer generates the output features. Positional embeddings are omitted, as point cloud coordinates inherently provide positional information.
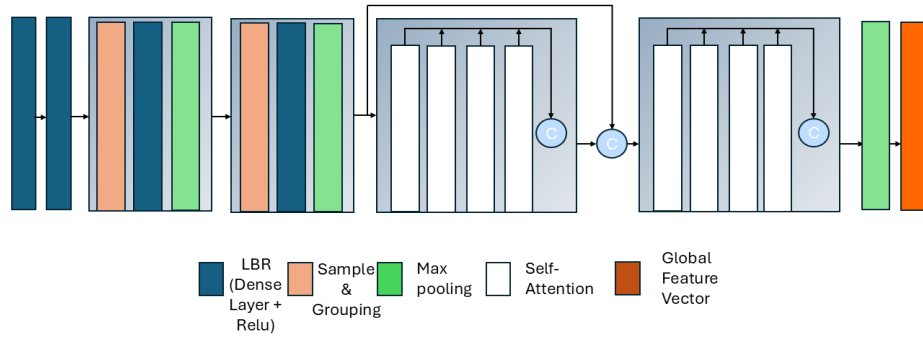
Fig. 5. Detailed illustration of the encoder.

Decoder: The decoder takes the concatenated global feature vectors from the text and point cloud encoders and initializes the point set by tiling this combined representation across 1,024 points. This forms the starting point for the reconstruction. An identity matrix is used to initialize the attention mechanism. The input then passes through four stages of cross-attention layers, where queries are derived from the decoder's current features and keys/values from the encoder's output. This process refines the point-wise features by allowing the decoder to focus on relevant parts of the global feature vector. The refined features are concatenated and up-sampled to 3,072 points to progressively increase the number of points for a complete point cloud shape. This procedure is repeated in a second block of cross-attention layers, followed by another up-sampling stage. Finally, the features are processed through several linear + ReLU layers, and the output is generated by a dense layer, producing the 3D coordinates (x, y, z) for the reconstructed point cloud, resulting in a final output of 6,144 points.
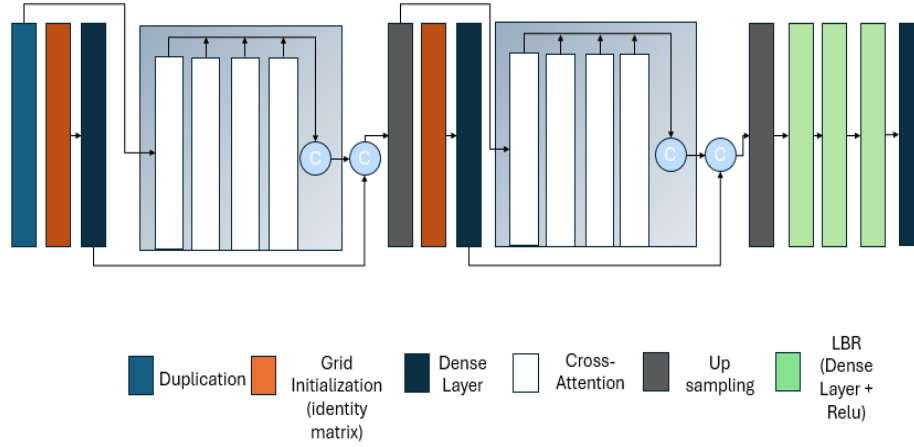
Fig. 6. Detailed illustration of the decoder.

3.6   Evaluation Metric

To evaluate the dissimilarity between the predicted point clouds and the corresponding GT point clouds, we utilized the Density-Aware Chamfer Distance (DCD) loss. This metric, an extension of the traditional Chamfer Distance [40, 41], accounts for the varying point densities in the dataset, enabling a more accurate comparison between dense and sparse regions of the point clouds [40]. The DCD between two point sets S_1 and S_2 is defined as:

$$L_{DCD}(S_1, S_2) = \frac{1}{2} \left(\frac{1}{S_1} \sum_{x \in S_1} e^{-\alpha\|x_i - \hat{y}\|_2^2} + \frac{1}{S_2} \sum_{x \in S_2} e^{-\alpha\|y - \hat{x}\|_2^2} \right) \tag{4}$$

Where $\hat{y} = \min_{y \in S_2} \|x - y\|^2$ and $\hat{x} = \min_{x \in S_1} \|y - x\|^2$, with α representing a temperature scalar. Without loss of generality, focusing on the first term, each y contributes $\left| -\frac{1}{n_y} \sum_{x \in S_1} e^{-\|x - y\|^2} \right| \in [0, 1]$ to the overall distance metric before averaging [40].

The loss is aggregated over M versions of incomplete instances, where $\hat{y}_n^m$ represents the model predictions:

$$L_F = \sum_{m=1}^{N} \sum_{n=1}^{M} L_{DCD}(y_n, \hat{y}_n^m) \tag{5}$$

3.7   Implementation Details

The model, previously described in subsection 3.5, contains approximately 187.5 million trainable parameters and is implemented using TensorFlow [34]. Training was conducted on 12 NVIDIA RTX A6000 GPUs, each with 40 GB of VRAM, and 8 CPUs with 1000 GB of RAM. Training was halted after 322 epochs as there were no noticeable improvements in performance for 12 consecutive epochs. The model was optimized using the ADAM optimizer [20], with an initial learning rate of 1e-7. To improve training efficiency, the ReduceLROnPlateau callback was used to dynamically adjust the learning rate based on validation performance, helping the model converge more effectively.

4   Results

4.1   Quantitative Results

A quantitative analysis was conducted to assess the model's performance objectively. The CD [41] evaluates the model's capacity to complete the general shape correctly, whereas the DCD [40] assesses the model's capability to address uneven point distributions. Furthermore, two F-scores were computed to ascertain the balance between precision and recall of the model's predictions. Lower CD

and DCD values, coupled with a higher F-score, indicate better performance and a closer alignment between the predicted and GT point clouds.

The results of the quantitative analysis, summarized in Table 1, are based on the entire validation dataset, which contains 20,962 point clouds. The CD, with a mean value of 0.00170, demonstrates minimal average distance between the predicted and GT point clouds, indicating high accuracy in overall shape reconstruction. The DCD yielded a value of 1.41269, reflecting the model's ability to handle varying point densities.

In terms of F-scores, the model achieved a high mean score of 0.937 with a threshold of 0.05, indicating strong performance in capturing both coarse and fine details in the normalized point cloud space, where a threshold of 0.05 represents 5% of the total possible distance. Additionally, with a more stringent threshold of 0.03, the model achieved a lower F-score of 0.7408, showing that while the model performs well in general, there is a slight decrease in accuracy when evaluated on finer, more granular details.

Table 1. Quantitative results for point cloud completion.

Metric	Mean Value
CD	0.00170
DCD	1.41269
F-score (threshold=0.05)	0.9370
F-score (threshold=0.03)	0.7408

4.2   Qualitative Results

Visual inspection complements quantitative metrics, offering insights into the model's ability to reconstruct point clouds. This section compares the model's outputs to GT, evaluating its performance in capturing fine details, handling complex structures, and generating realistic shapes from incomplete inputs. These examples reveal the model's strengths and limitations across various anatomical shapes.

Figure 7 presents a multi-angle comparison of the partial input, predicted point cloud, and the GT for a heart myocardium instance. By showcasing the point cloud from different perspectives, this visualization highlights the model's ability to generate a complete and accurate 3D reconstruction of the heart myocardium based on incomplete input data. The comparison demonstrates how well the predicted point cloud aligns with the GT, providing valuable insight into the model's performance across different regions of the myocardium.

Input   Prediction   GT

Fig. 7. Multi-angle comparison of partial input, predicted point cloud, and GT for a heart myocardium instance.

Figure 8 demonstrates the model's capability to reconstruct the overall structure of the right iliac vein with a reasonable degree of accuracy. While the model effectively captures the main anatomical features, it struggles with the lower section of the vein, resulting in an imprecise reconstruction of this region.
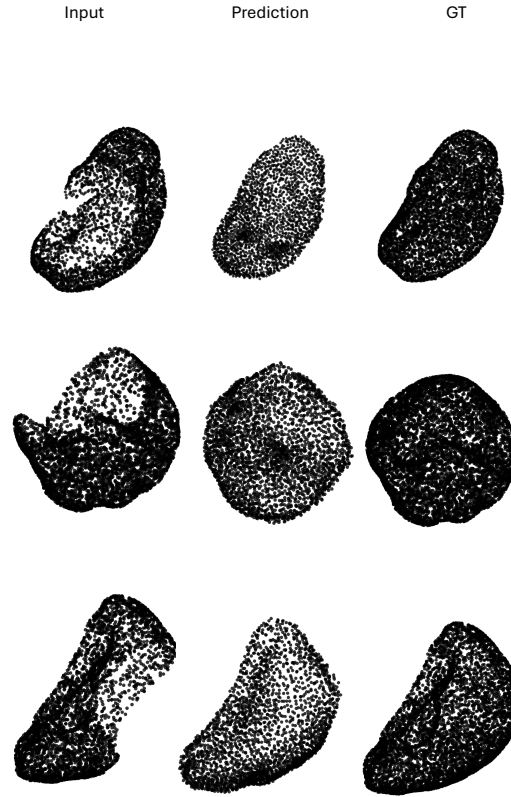

Fig. 8. Multi-angle comparison of partial input, predicted point cloud, and GT for a right iliac vena instance.

The model exhibits a commendable ability in reconstructing the right scapula's overall shape, as depicted in Figure 9. It effectively captures the major structures, including the prominent curves and edges that define the general form. While some of the finer details appear less distinct, these minor variances do not significantly impact the accuracy or quality of the reconstructed shape, reflecting the model's robustness in handling complex anatomical features.

Fig. 9. Multi-angle comparison of partial input, predicted point cloud, and GT for a right scapula example.

The model performs exceptionally well in completing the spleen, producing results that are nearly perfect in terms of shape and structure. As shown in Figure 10, the overall geometry of the organ is accurately captured, with the predicted point cloud closely matching the GT. The model demonstrates a high level of precision in handling the spleen's complexities, making it one of the best reconstructions in the dataset.

Fig. 10. Multi-angle comparison of partial input, predicted point cloud, and GT of a spleen.

The examples presented in this paper were randomly selected to offer a representative overview of the model's performance across various structures.
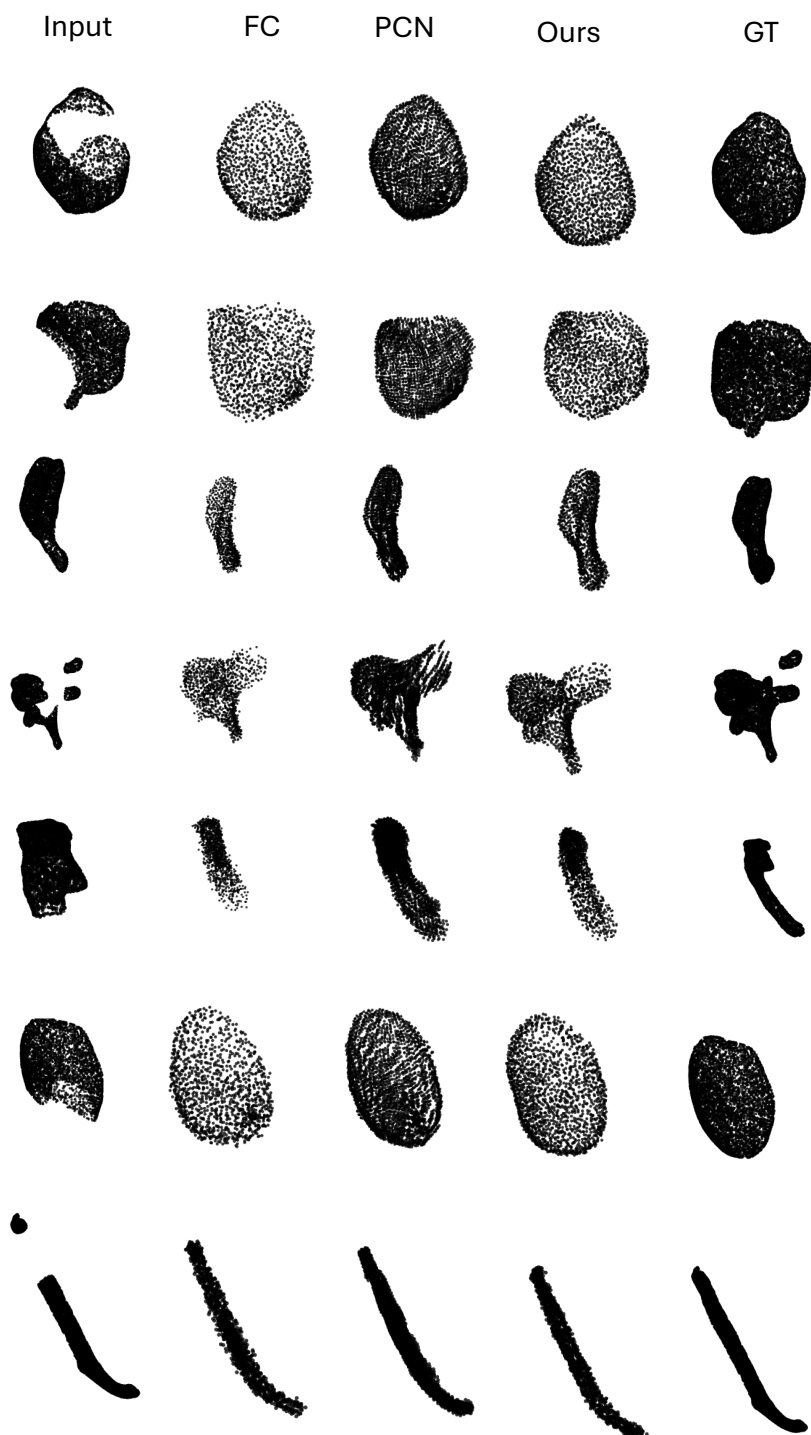
4.3   Comparison

In order to assess the performance of our model against other architectures, we trained three models—our model, the Fully Connected (FC) network [44], and the Point Completion Network (PCN) [44]—on a subset of the MedShapeNet dataset. We used a total of 4,800 shapes from this dataset to ensure consistency across all experiments. Each model was trained under identical conditions to provide a fair comparison of their capabilities in point cloud completion.

The evaluation was conducted using several metrics, including CD, DCD, and F-score with thresholds of 0.05 and 0.03, to compare the models. These metrics were chosen to capture both the global shape accuracy and fine details of the reconstructed point clouds. The results, summarized in Table 2 and Figure 11, highlight the differences in performance among the models.

Table 2. Quantitative comparison of point cloud completion models (*DCD requires the number of Prediction and GT point clouds to match. The PCN produces 6,084 points while the GTs contain 6,144 points).

Metric	FC	PCN	Our Model
CD	0.003194	0.002758	**0.002327**
DCD	1.70910	N/A*	**1.60467**
F-score (threshold = 0.05)	0.83295	0.88799	**0.89202**
F-score (threshold = 0.03)	0.5201	**0.6630**	0.6234

Fig. 11. Shape comparison between the FC, PCN networks and our model on a subset of the MedShapeNet dataset.

5   Discussion

The results of our experiments underscore the strong capabilities of the proposed multimodal foundation model for point cloud completion. Quantitatively, the model achieved a mean CD value of 0.00170, reflecting high accuracy in reconstructing overall shapes across the dataset. Additionally, the model's DCD, with a mean value of 1.41269, demonstrates its effectiveness in handling varying point densities—a crucial factor in medical datasets with complex geometries. However, these findings also reveal areas for improvement, as the model occasionally struggles to capture fine details in regions with uneven point distributions.

The F-score, calculated at two thresholds, offers deeper insight into the model's performance. At a threshold of 0.05, the F-score reached 0.937, showcasing strong precision and recall for the majority of anatomical structures. Even under the more stringent threshold of 0.03, the F-score remained robust at 0.7408, highlighting the model's resilience in tackling more intricate and detailed structures. This demonstrates that while the model excels at capturing general anatomical shapes, there remains room for refinement in handling finer details.

Qualitative evaluations further support these findings, illustrating the model's ability to accurately reconstruct missing portions across diverse anatomical categories. In many cases, the model demonstrated high fidelity in restoring complex structures. However, in more intricate regions, characterized by complex geometries, it occasionally encountered challenges in capturing delicate details, suggesting avenues for future improvements.

Our model outperforms both the Fully Connected (FC) and Point Completion Network (PCN) architectures in most metrics, achieving lower CD and DCD values, indicating greater accuracy in point cloud reconstruction. Although PCN achieved a slightly higher F-score at the strict threshold of 0.03, our model excelled at the 0.05 threshold, demonstrating a stronger balance between capturing global shapes and preserving fine details. These quantitative results are further confirmed by qualitative comparisons, which consistently showed our model's superior ability to reconstruct intricate structures.

These results underscore both the strengths and limitations of the model. Its ability to complete a wide range of medical structures with high accuracy highlights its potential for use in medical imaging tasks. Nonetheless, addressing the challenge of reconstructing finer details offers an opportunity for enhancement.

Beyond its technical performance, the model holds significant promise for practical applications. It can serve as a valuable tool in educational settings, allowing medical students and professionals to better visualize and understand complex anatomical structures. Additionally, its potential for use in XR applications opens doors for immersive, real-time interactions with 3D anatomical

data, enhancing medical training. Finally, the model paves the way for innovations in bone implant design, enabling the creation of custom implants tailored to patient-specific anatomy through the reconstruction of complete 3D shapes from incomplete scans.

Future work could focus on architectural improvements, such as the incorporation of residual connections or advanced loss functions that specifically target the preservation of intricate details in complex anatomical regions.

6 Conclusion

This paper introduces a multimodal foundation model for medical point cloud completion, trained on the comprehensive MedShapeNet dataset, which encompasses over 200,000 point clouds spanning over more than 240 anatomical classes. By integrating a transformer-based AE with a BERT text encoder, the model effectively leverages both geometric and textual data, achieving high accuracy in reconstructing diverse anatomical structures. Quantitative evaluations affirm the model's robust performance, while qualitative assessments further highlight its ability to generalize across a wide range of anatomical categories.

While the model demonstrates strong overall performance, challenges remain in reconstructing finer, more intricate anatomical details, particularly in regions with complex geometries and uneven point densities. Future work may focus on refining the model's architecture, such as incorporating residual connections to enhance the preservation of intricate details, and developing more advanced loss functions specifically tailored to complex shapes and varying point densities.

This model establishes a robust foundation for future research in medical point cloud completion, with significant potential applications in areas such as medical diagnostics, surgical planning, and custom implant design. By enabling the reconstruction of complete 3D shapes from incomplete scans, it also opens up new opportunities for advancements in personalized medicine and medical training. Additionally, the model holds promise for enhancing educational tools in XR, making it a valuable resource for both practical medical applications and immersive learning environments. Thus, its potential impact spans from the design of custom bone implants to broader uses in the field of medical education and XR.

References

1. Baltrušaitis, T., Ahuja, C., Morency, L.P.: Multimodal machine learning: A survey and taxonomy. IEEE transactions on pattern analysis and machine intelligence **41**(2), 423–443 (2018)
2. Beetz, M., Banerjee, A., Ossenberg-Engels, J., Grau, V.: Multi-class point cloud completion networks for 3d cardiac anatomy reconstruction from cine magnetic resonance images. Medical Image Analysis **90**, 102975 (2023)
3. Bello, S.A., Yu, S., Wang, C., Adam, J.M., Li, J.: Deep learning on 3d point clouds. Remote Sensing **12**(11), 1729 (2020)
4. Cai, Z., Wang, C., Wen, C., Li, J.: 3d-patchmatch: An optimization algorithm for point cloud completion. In: 2015 2nd IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services (ICSDM). pp. 157–161. IEEE (2015)
5. Caruso, C., Quarta, F.: Interpolation methods comparison. Computers & Mathematics with Applications **35**(12), 109–126 (1998)
6. Chauve, A.L., Labatut, P., Pons, J.P.: Robust piecewise-planar 3d reconstruction and completion from large-scale unstructured point data. In: 2010 IEEE computer society conference on computer vision and pattern recognition. pp. 1261–1268. IEEE (2010)
7. Chen, X., Ma, H., Wan, J., Li, B., Xia, T.: Multi-view 3d object detection network for autonomous driving. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 1907–1915 (2017)
8. De Fauw, J., Ledsam, J.R., Romera-Paredes, B., Nikolov, S., Tomasev, N., Blackwell, S., Askham, H., Glorot, X., O'Donoghue, B., Visentin, D., et al.: Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine **24**(9), 1342–1350 (2018)
9. Doria, D., Radke, R.J.: Filling large holes in lidar data by inpainting depth gradients. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 65–72. IEEE (2012)
10. Eldar, Y., Lindenbaum, M., Porat, M., Zeevi, Y.Y.: The farthest point strategy for progressive image sampling. IEEE transactions on image processing **6**(9), 1305–1315 (1997)
11. Fauser, J., Stenin, I., Bauer, M., Hsu, W.H., Kristin, J., Klenzner, T., Schipper, J., Mukhopadhyay, A.: Toward an automatic preoperative pipeline for image-guided temporal bone surgery. International journal of computer assisted radiology and surgery **14**, 967–976 (2019)
12. Fix, E.: Discriminatory analysis: nonparametric discrimination, consistency properties, vol. 1. USAF school of Aviation Medicine (1985)
13. Friedman, S., Stamos, I.: Online facade reconstruction from dominant frequencies in structured point clouds. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. pp. 1–8. IEEE (2012)
14. Guo, M.H., Cai, J.X., Liu, Z.N., Mu, T.J., Martin, R.R., Hu, S.M.: Pct: Point cloud transformer. Computational Visual Media **7**, 187–199 (2021)
15. Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., Bennamoun, M.: Deep learning for 3d point clouds: A survey. IEEE transactions on pattern analysis and machine intelligence **43**(12), 4338–4364 (2020)

16. Huang, Y., Du, C., Xue, Z., Chen, X., Zhao, H., Huang, L.: What makes multi-modal learning better than single (provably). Advances in Neural Information Processing Systems **34**, 10944–10956 (2021)
17. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Proceedings of the fourth Eurographics symposium on Geometry processing. vol. 7 (2006)
18. Kazhdan, M., Funkhouser, T., Rusinkiewicz, S.: Symmetry descriptors and 3d shape matching. In: Proceedings of the 2004 Eurographics/ACM SIGGRAPH symposium on Geometry processing. pp. 115–123 (2004)
19. Kimia, B.B., Frankel, I., Popescu, A.M.: Euler spiral for shape completion. International journal of computer vision **54**(1), 159–182 (2003)
20. Kingma, D.P.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Kroemer, O., Amor, H.B., Ewerton, M., Peters, J.: Point cloud completion using extrusions. In: 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). pp. 680–685. IEEE (2012)
22. Li, J., Zhou, Z., Yang, J., Pepe, A., Gsaxner, C., Luijten, G., Qu, C., Zhang, T., Chen, X., Li, W., et al.: Medshapenet–a large-scale dataset of 3d medical shapes for computer vision. arXiv preprint arXiv:2308.16139 (2023)
23. Li, Y., Dai, A., Guibas, L., Nießner, M.: Database-assisted object retrieval for real-time 3d reconstruction. In: Computer graphics forum. vol. 34, pp. 435–446. Wiley Online Library (2015)
24. Liu, Y., Cheng, M.M., Fan, D.P., Zhang, L., Bian, J.W., Tao, D.: Semantic edge detection with diverse deep supervision. International Journal of Computer Vision **130**(1), 179–198 (2022)
25. Lou, B., Doken, S., Zhuang, T., Wingerter, D., Gidwani, M., Mistry, N., Ladic, L., Kamen, A., Abazeed, M.E.: An image-based deep learning framework for individualising radiotherapy dose: a retrospective analysis of outcome prediction. The Lancet Digital Health **1**(3), e136–e147 (2019)
26. Mazzocchetti, S., Spezialetti, R., Bevini, M., Badiali, G., Lisanti, G., Salti, S., Di Stefano, L.: Neural shape completion for personalized maxillofacial surgery. Scientific Reports **14**(1), 19810 (2024)
27. Nguyen, M., He, T., An, L., Alexander, D.C., Feng, J., Yeo, B.T., Initiative, A.D.N., et al.: Predicting alzheimer's disease progression using deep recurrent neural networks. NeuroImage **222**, 117203 (2020)
28. Pauly, M., Mitra, N.J., Giesen, J., Gross, M.H., Guibas, L.J.: Example-based 3d scan completion. In: Symposium on geometry processing. pp. 23–32 (2005)
29. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
30. Sardar, P., Abbott, J.D., Kundu, A., Aronow, H.D., Granada, J.F., Giri, J.: Impact of artificial intelligence on interventional cardiology: from decision-making aid to advanced interventional procedure assistance. Cardiovascular interventions **12**(14), 1293–1303 (2019)
31. Schnabel, R., Degener, P., Klein, R.: Completion and reconstruction with primitive shapes. In: Computer graphics forum. vol. 28, pp. 503–512. Wiley Online Library (2009)
32. Sharf, A., Alexa, M., Cohen-Or, D.: Context-based surface completion. In: ACM SIGGRAPH 2004 Papers, pp. 878–887 (2004)
33. Stutz, D., Geiger, A.: Learning 3d shape completion from laser scan data with weak supervision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1955–1964 (2018)

34. TensorFlow, G.: Large-scale machine learning on heterogeneous systems. Google Research **10**, s15326985ep4001 (2015)
35. Vora, S., Lang, A.H., Helou, B., Beijbom, O.: Pointpainting: Sequential fusion for 3d object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4604–4612 (2020)
36. Wang, J., Cui, Y., Guo, D., Li, J., Liu, Q., Shen, C.: Pointattn: You only need attention for point cloud completion. In: Proceedings of the AAAI Conference on artificial intelligence. vol. 38, pp. 5472–5480 (2024)
37. Wang, Q., Tan, Y., Mei, Z.: Computational methods of acquisition and processing of 3d point cloud data for construction applications. Archives of computational methods in engineering **27**(2), 479–499 (2020)
38. Wang, X., Ang, M.H., Lee, G.H.: Voxel-based network for shape completion by leveraging edge generation. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 13189–13198 (2021)
39. Winter, J., Nowak, R.: Point cloud densification algorithm for multiple cameras and lidars data fusion (2024)
40. Wu, T., Pan, L., Zhang, J., Wang, T., Liu, Z., Lin, D.: Density-aware chamfer distance as a comprehensive metric for point cloud completion. arXiv preprint arXiv:2111.12702 (2021)
41. Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X., Xiao, J.: 3d shapenets: A deep representation for volumetric shapes. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1912–1920 (2015)
42. Yu, J., Zhang, C., Wang, H., Zhang, D., Song, Y., Xiang, T., Liu, D., Cai, W.: 3d medical point transformer: Introducing convolution to attention networks for medical point cloud analysis. arXiv preprint arXiv:2112.04863 (2021)
43. Yu, L., Li, X., Fu, C.W., Cohen-Or, D., Heng, P.A.: Pu-net: Point cloud upsampling network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2790–2799 (2018)
44. Yuan, W., Khot, T., Held, D., Mertz, C., Hebert, M.: Pcn: Point completion network. In: 2018 international conference on 3D vision (3DV). pp. 728–737. IEEE (2018)
45. Zhang, X., Feng, Y., Li, S., Zou, C., Wan, H., Zhao, X., Guo, Y., Gao, Y.: View-guided point cloud completion. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15890–15899 (2021)