Refining 3D Medical Segmentation with Verbal Instruction

Kangxian Xie[1 [0009−0007−2879−1448]], Jiancheng Yang[2,3⋆ [0000−0003−4455−7145]], Nandor Pinter[1,5 [0000−0002−7714−143X]], Chao Wu[1 [0009−0003−7508−2012]], Behzad Bozorgtabar[4 [0000−0002−5759−4896]], and Mingchen Gao[1 [0000−0002−5488−8514]]

[1] University at Buffalo, Buffalo NY 14260, USA
[2] ELLIS Institute Finland, Espoo 02150, Finland
[3] Aalto University, Espoo 02150, Finland
[4] Aarhus University, Aarhus 8000, Denmark
[5] Dent Neurologic Institute, Amherst, NY

Abstract. Accurate 3D anatomical segmentation is essential for clinical diagnosis and surgical planning. However, automated models frequently generate suboptimal shape predictions due to factors such as limited and imbalanced training data, inadequate labeling quality, and distribution shifts between training and deployment settings. A natural solution is to iteratively refine the predicted shape based on the radiologists' verbal instructions. However, this is hindered by the scarcity of paired data that explicitly links erroneous shapes to corresponding corrective instructions. As an initial step toward addressing this limitation, we introduce **CoWTalk**, a benchmark comprising 3D arterial anatomies with controllable synthesized anatomical errors and their corresponding repairing instructions. Building on this benchmark, we further propose an iterative refinement model that represents 3D shapes as vector sets and interacts with textual instructions to progressively update the target shape. Experimental results demonstrate that our method achieves significant improvements over corrupted inputs and competitive baselines, highlighting the feasibility of language-driven clinician-in-the-loop refinement for 3D medical shapes modeling. Code and dataset will be released at https://github.com/HINTLab/CowTalk.

Keywords: Medical Shape · Vision-Language Model · Shape Editing.

1 Introduction

Accurate 3D reconstruction of anatomical shapes from medical images is essential to diagnosis, measurement, and intervention planning. Despite the strong progress of modern segmentation and reconstruction networks, radiologists still frequently encounter suboptimal 3D segmentation shapes in practice, due to domain variations (e.g., different scanners, protocols, and populations), limited and imbalanced training data, or inaccurate annotations.

⋆ Correspondence to: Jiancheng Yang (jiancheng.yang@aalto.fi).

However, radiologists typically have a clear expectation of what the underlying anatomy looks like while inspecting the 3D image volume. They naturally compare the expected shape with the predicted shape and can articulate the differences (e.g., the lumen is disconnected near the proximal end). Motivated by this observation, we address the segmentation problem from a new perspective by iteratively refining the suboptimal shape according to clinicians' instructions.

Research in text-guided shape editing is largely constrained by the availability of datasets like ShapeTalk [1] and ShapeWalk [16] that align linguistic context with geometric changes. While ShapeTalk utilizes manual descriptions of shape differences, and ShapeWalk [16] relies on CAD model interpolations for fine-grained control, the medical domain remains underexplored because there is no publicly available 3D dataset of paired medical shapes with natural language descriptions for geometric refinement. We bridge this gap by establishing the **CoWTalk** benchmark on the arterial segments mask of TopCoW [21] dataset. We simulate controlled segmentation errors by introducing parameterized morphological perturbations to the ground truth segments (Fig. 2), and subsequently associate them with LLM-generated corrective instructions (Fig. 3). For **CoWTalk**, we generate 15 erroneous variants for each subject in TopCoW, and construct an (Erroneous shape, Instruction, GT Shape) tuple per variant, resulting in a paired dataset comprising 3,750 samples.

Early research on text-shape interaction primarily focused on language-based shape generation [7, 19], texture editing [9] based on CLIP [14], and shape editing (ShapeTalk [1], ShapeWalk [16]), with the latter two being the most relevant to ours. While ShapeTalk aligns text and geometry via feature concatenation, ShapeWalk adopts a latent diffusion-based mechanism. Implicit-based approach [2] leverages boundary sensitivity to link neural parameters to surface update. Interactive paradigms such as PrEditor3D [4] explore synchronized 2D multi-view editing. In contrast, the proposed architecture adopts an intuitive encode-fusion-decode structure for shape refinement. First, the medical structure is encoded into vector sets [22], and the vector sets undergo multi-modal fusion with text features [3] before implicit decoding. This pipeline translates high-level verbal instructions into 3D shape edits by conditioning shape representations on the contextual information provided by radiologists.

Our contributions are threefold. First, we introduce **CoWTalk**, a new benchmark for instruction-following medical shape refinement based on TopCoW [21]. Second, we propose an end-to-end 3D vision-language iterative refinement framework (Fig. 1) that updates a multi-class 3D segmentation shape according to clinician instructions, enabling human-in-the-loop correction of complex vascular anatomy. Finally, we conduct a comprehensive evaluation to validate the advantages and effectiveness of the proposed method.

2 Problem Formulation

There are two stages to achieving instruction-guided medical shape refinement. The first stage synthesizes errors into ground truth (GT) anatomical shapes and

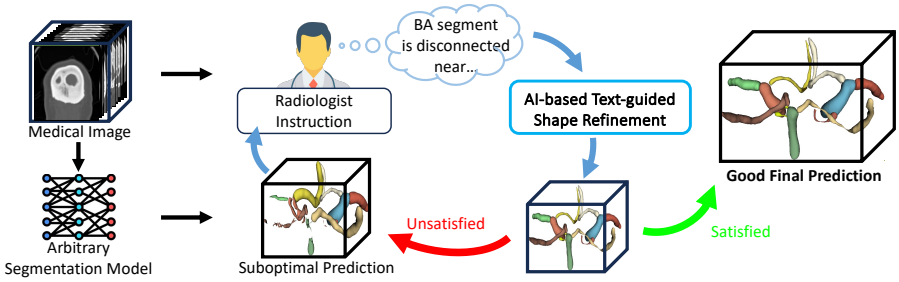

Fig. 1. The overall workflow. A medical image is processed by an arbitrary segmentation model for an initial/suboptimal shape prediction. Based on the shape, the radiologist provides instructions to iteratively correct the shape until satisfied.

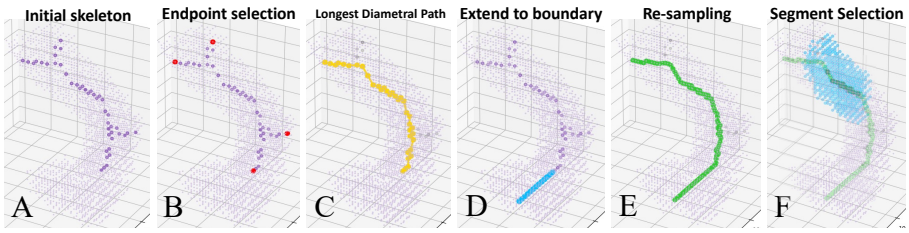

Fig. 2. Centerline refinement and Segment Selection. A to E shows the steps for refining the centerline representation, while F illustrates the partial selection of a segment (blue).

pairs each corrupted shape with a natural-language corrective instruction. The second stage establishes the learning objective, where the model learns to recover the GT shape from a corrupted version given natural language instructions.

2.1 CoWTalk Dataset Synthesis

The **CoWTalk** benchmark is developed based on the artery label masks of the Circle of Willis (CoW) from the TopCoW dataset [21], which provides the 14-class masks for the artery structure. For each subject, we iterate through its 13 arterial tubular segments, and synthesize error into the segment. The errors are randomized from a designated list of global errors: global thickening/thinning, missing segments, and local errors: local thickening/thinning, shortening, disconnection, fragmentation. For global errors, 3D morphological operations (dilation, erosion and etc.) [18] are applied to the entire segment, while a local error affects only a contiguous portion of the tube. With each error type, error-related parameters (Fig.3, left) are also randomly generated as a numerical description.

For local error synthesis, selecting a local tube span from error parameters for error synthesis requires special attention. We use each class segments' skeletal

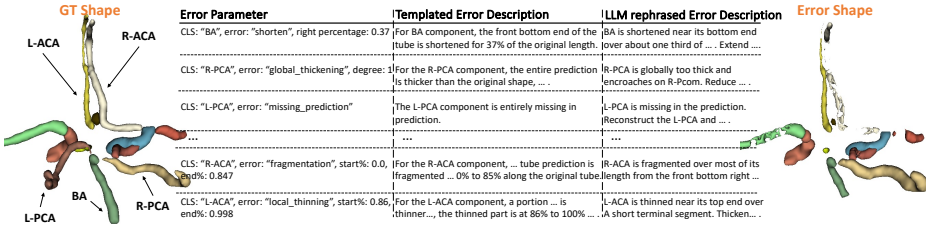

Fig. 3. Examples of generated instructions in **CoWTalk**.

centerlines from the TopCoW centerline dataset [10] and improve upon (Fig.2 A-E) them to be the abstract spatial and trajectory representation of the shape for local region identification. Given the percentage range parameters (Fig.3, left), we treat the centerline points as connected graph nodes with edge weights 1, and select the points located within the desired range from an anchor endpoint. For example, a 86%-to-99% tube local selection corresponds to selecting the 86 to 99 percentile points in terms of their distance away from an anchor endpoint. For each selected point, we generate a cylinder shape pointing at the local trajectory of the centerline. The selected mask (Fig.2 F) is identified as the intersection between these cylinders and the original ground truth mask.

As shown in Fig. 3, given error parameters, we instantiate a template sentence that describes the error, and fix a canonical posterior viewing direction so that directional phrases (e.g., left/right, proximal/distal) are unambiguous. Subsequently, a dedicated prompt is created for ChatGPT 5.2 [11] to paraphrase the templated error description into radiologist-style narratives, followed by corrective instructions with two variants: a concise (general) and a comprehensive (detailed) instruction.

2.2 Interactive Shape Refinement Objective

The refinement objective is to learn a mapping f_θ that iteratively recovers the ground-truth anatomy S_{gt} from an erroneous input shape by resolving the errors specified in a sequence of natural-language instructions $\{t^k\}_{k=0}^{K-1}$. We formulate this as an iterative process:

$$S^{k+1} = f_\theta(S^k, t^k), \quad k = 0, \ldots, K-1, \tag{1}$$

where S^0 is the initial erroneous shape and S^K is the final refined output, expected to match the ground truth on the targeted arterial structures.

3 Methods

We propose a language-guided end-to-end shape refinement network (Fig.4) that takes a suboptimal segmentation shape and the radiologist's instruction as input and outputs a refined segmentation shape according to Equation. 1. The

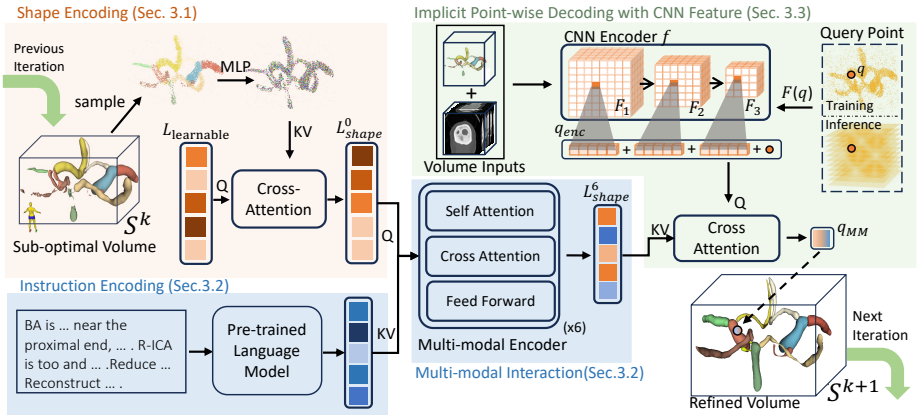

Fig. 4. Overview of the proposed instruction-guided medical shape refinement pipeline.

proposed architecture contains four major components: shape encoding, text encoding, multi-modal interaction, and implicit point-wise decoding with CNN features.

3.1 Shape Encoding via Latent Query Cross-Attention

Given an input volume containing the corrupted segmentation shape S_{error}, we sample a point cloud $P = \{(\mathbf{x}_i, \mathbf{p}_i)\}_{i=1}^{N}$, where $\mathbf{x}_i \in \mathbb{R}^3$ denotes the coordinate and $\mathbf{p}_i$ is the label at $\mathbf{x}_i$ in the corrupted volume. Inspired by 3DS2VS [22], we first project the input points into point-wise embedding features $\phi(P)$, and embed the point features $\phi(P)$ into a learnable latent vector set $L_{learnable} \in \mathbb{R}^{M\times d}$ with a cross-attention [17] module to summarize the shape:

$$L_{shape} = \text{CrossAttn}(L_{learnable}, \phi(P)), \tag{2}$$

where $L_{shape} \in \mathbb{R}^{M\times d}$ ($M = 512$, $d = 512$) is referred to as shape latents.

3.2 Instruction Encoding and Multi-modal Interaction

A radiologist provides a plain-text instruction t describing what is wrong and how to correct it. We tokenize t and encode it with a pre-trained BERT encoder [3] to obtain per-token features in $\mathbb{R}^{768}$, then project token-wise to $d = 512$ to form a text feature vector set $T \in \mathbb{R}^{L\times 512}$. We refer to T as the *text feature*.

To inject text feature into the *Shape Latents* L_{shape}, we employ a multi-modal encoder consisting of $K = 6$ attention blocks. Each block contains: a self-attention [17] layer over the current L_{shape}, a cross-attention layer where L_{shape} attends to text features, and a feed-forward network. Let L^k_{shape} denote the shape latent after the k-th block, we compute

$$L^{k+1}_{shape} = \text{FFN}\big(\text{CrossAttn}(\text{SelfAttn}(L^k_{shape}), T)\big), \quad k = 0, \ldots, K-1. \tag{3}$$

The output L^K_{shape} is a set of *Shape Latents* conditioned on the user instruction.

3.3 Implicit Point-wise Decoding with CNN Feature

Arterial geometries are locally smooth and near-linear, following well-defined trajectories in 3D space. Implicit functions are well-known to naturally preserve continuity [2, 12, 20, 8, 15], but they still require spatially coherent context. To generate such a continuous feature field, we feed the error volume and the image (optional) through a 3-layer CNN f to obtain feature pyramids $F = \{F_1, F_2, F_3\}$. For each query location $\mathbf{q}$, we interpolate F at $\mathbf{q}$ for multi-level features, and form the point encoding: $\mathbf{q}_{enc} = [F_1(\mathbf{q}_m), F_2(\mathbf{q}_m), F_3(\mathbf{q}), \mathbf{q}, S_{error}(\mathbf{q})]$. Then, an cross-attention-based implicit-function decodes $\mathbf{q}_{enc}$ based on the *Shape Latents* L^K_{shape} for an updated point encoding with multi-modal context:

$$\mathbf{q}_{MM} = \text{CrossAttn}(\mathbf{q}_{enc}, L^K_{shape}) \tag{4}$$

We then apply an MLP to predict class logits $\hat{\mathbf{y}}_m = \text{MLP}(\mathbf{q}_{MM})$. Finally, the predicted labels are assembled into the refined segmentation $S_{refined}$.

3.4 Training and Inference

During training, query points are sampled from the near-surface region, foreground, and random background to ensure sufficient coverage of key locations. We mix the long and short versions of instruction during training. The model is trained end-to-end with CE loss for 200 epochs. During inference, all 3D locations are fed as query points to enable full-volume reconstruction.

As a training data augmentation, we randomly remove each error instance with probability 0.2 (i.e., replace it with the correct class shape) as model input. For each input error, the model learns to correct the error with probability 0.5, guided by the corresponding instruction, leaving the remaining error intact.

4 Experiments

We perform experiments against baselines (Sec.4.1), analyze robustness across error types and input modalities (Sec. 4.2), and evaluate the network on real errors with instructions by a radiologist (Sec. 4.3). In primary experiments (Tab.1) and error type analysis (Fig.6) on **CoWTalk**, we drop each error with probability 20% to simulate variable input segmentation quality.

4.1 Baseline Comparisons

Table 1 reports the results for erroneous shape refinement. We compare our method against CNN-based refinement, 3D reconstruction + text refinement, and specialized text-guided shape editing methods in general-purpose domain.

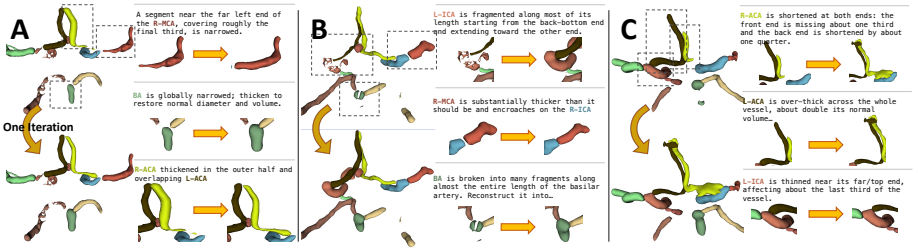

Fig. 5. Qualitative refinement results with shape-only input from **CoWTalk**, and only 3 corrections are highlighted for each sample. Some correction descriptions are omitted.

Table 1. CoWTalk refinement performances using F1 score, Dice, Normalized Surface Dice (NSD) and Chamfer's Distance (CD) as metrics. Results are presented both without and with images. Iter.: Iterative; Text: if guided by text. *: methods with specialized enhancements.

Method	Iter.	Text	F1	Dice	NSD	CD(↓)
Input Error	–	–	86.0	81.9	86.7	0.795
nnUNetv2 [6]	–	–	84.6 / **94.2**	79.5 / 90.0	86.4 / 93.2	1.84 / **0.29**
SwinUnetrv2 [5]	–	–	86.6 / 89.5	79.7 / 84.7	84.5 / 90.0	0.83 / 0.77
3DS2VS [22]	–	✓	39.6 / –	36.3 / –	34.2 / –	5.88 / –
ShapeTalk [1]*	✓	✓	72.4 / 79.3	71.0 / 78.7	77.6 / 83.6	3.13 / 3.28
ShapeWalk [16]*	✓	✓	71.6 / 77.0	68.5 / 74.4	75.4 / 81.4	5.58 / 3.45
Ours	✓	✓	**87.7** / 91.8	**87.5** / **91.8**	**90.7** / **94.3**	1.70 / 1.33

For CNN-based refinement, we tested nnUNetv2 [6] and SwinUNetrv2 [5] on two input settings: shape-only input, and 2-channel input with concatenated medical image. We observe that with the image, nnUNetv2 achieves the highest F1, and improves upon the erroneous shape, likely due to its reliance on image.

Next, we focus on methods based on the latent vector set representation [22]. For methods using vector sets for representation, we unify the vector dimensionality to $(512, 512)$ for fair comparison. We also split the refinement into 4 non-overlapping instruction sets for iterative operation, since the model performs increasingly well as more iterations divide the instruction into shorter text input (Iterations: 1,2,3,4 $\rightarrow$ $91.0, 91.3, 91.6, 91.8$ in Dice %). First, we evaluate a straightforward baseline that injects textual context via cross-attention into the 3DS2VS [22] reconstruction pipeline to perform edits; however, it fails to reconstruct the original shape. This suggests that naïvely conditioning generic reconstruction features on text can be unstable, and that the MLP-only implicit decoding feature lacks continuous feature context. We also compare against two general-purpose text-guided shape editing methods, ShapeTalk (S.T.) [1] and ShapeWalk (S.W.) [16]. To enable fair comparisons and continuous feature context, we replace the original PointNet [13] auto-encoder in S.T. with a 3DS2VS-style vector-set auto-encoder [22], and further augment both S.T. and S.W. with CNN-derived local features (Sec.3.3) for reconstruction pre-training. Nev-

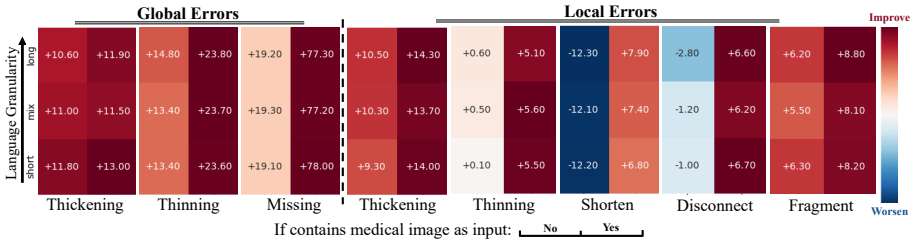

Fig. 6. Refinement (F1 %) with the proposed method under different visual and language input, showing the difference between refined results and synthesized error (±).

ertheless, under pretrained-latent-editing protocol, neither method sufficiently reaches comparable refinement performance. Conversely, without medical images as guidance, our text-guided model achieves 6.8% Dice refinement improvement over the error samples, 10% advantage over CNN-refinement, and top performances across the first 3 metrics. With image input, our method still outperforms segmentation-focused methods [6, 5] in Dice and NSD.

4.2 Multi-modal Input Ablation on Error Categories

We ablate visual and language inputs to quantify the impact of image context and instruction granularity, and display results as heatmaps by error type in Fig. 6. Without image as ground truth, errors that remove anatomy (Disconnection, Missing, Shorten) are harder to correct as they require fabrication (Fig. 5 C. top); in contrast, shape-only refinement works well at errors where topology is largely preserved, such as fragmentation and local thickening (Fig. 5 A. Bottom; B. Top). Additionally, across most error types, our performance is consistent and robust under various instruction granularities.

4.3 Radiologist Instruction on Real Error

To assess real-world applicability, we evaluate our method on cases pre-segmented by nnUNetv2 [6] with instructions provided by a researcher without a medical background, and by a practicing neuroradiologist with 13 years of experience. The qualitative example in Fig. 7 shows that both instructions correct the error with professional instruction yielding more topologically robust results.

5 Conclusion

This work aims to enable clinician-in-the-loop refinement of medical segmentations through iterative, instruction-following updates. To this end, we introduced **CoWTalk**, a benchmark built on CoW tubular structures, where centerlines provide a principled basis for controllable error synthesis and radiologist-like corrective instructions. An iterative refinement framework is proposed with

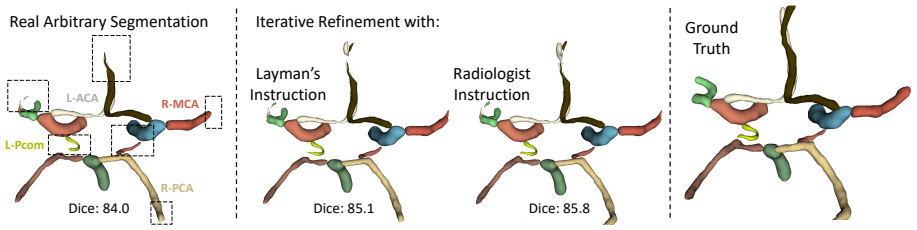

Fig. 7. Example of refining a real, suboptimal segmentation [6] prediction using instructions given by a layperson and a radiologist.

multi-modal interaction to apply corrective edits. Experiments demonstrate that this pipeline effectively follows instructions and improves segmentation quality. We present this work as a focused proof-of-concept for language-guided refinement of 3D medical segmentation shapes, with the synthesized error distribution serving as a well-defined testbed for this novel task.

Acknowledgments. This study was supported by ELLIS Institute Finland and the School of Electrical Engineering, Aalto University. We acknowledge CSC-IT Center for Science, Finland, for providing access to the supercomputers Mahti and Puhti, as well as LUMI, owned by the European High Performance Computing Joint Undertaking (EuroHPC JU) and hosted by CSC Finland in collaboration with the LUMI consortium. We also acknowledge the computational resources provided by the Aalto Science-IT project through the Triton cluster. The work was further supported by the US NSF CAREER award IIS-2239537. This work was also supported in part by Dent Family Foundation.

Disclosure of Interests The authors have no competing interests to declare.

References

1. Achlioptas, P., Huang, I., Sung, M., Tulyakov, S., Guibas, L.: ShapeTalk: A language dataset and framework for 3d shape edits and deformations. In: Conference on Computer Vision and Pattern Recognition (CVPR) (2023)
2. Berzins, A., Ibing, M., Kobbelt, L.: Neural implicit shape editing using boundary sensitivity. In: The Eleventh International Conference on Learning Representations (2023), https://openreview.net/forum?id=CMPIBjmhpo
3. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). pp. 4171–4186 (2019)
4. Erkoç, Z., Gümeli, C., Wang, C., Nießner, M., Dai, A., Wonka, P., Lee, H.Y., Zhuang, P.: Preditor3d: Fast and precise 3d shape editing. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 640–649 (June 2025)

5. He, Y., Nath, V., Yang, D., Tang, Y., Myronenko, A., Xu, D.: Swinunetr-v2: Stronger swin transformers with stagewise convolutions for 3d medical image segmentation. In: Greenspan, H., Madabhushi, A., Mousavi, P., Salcudean, S., Duncan, J., Syeda-Mahmood, T., Taylor, R. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2023. pp. 416–426. Springer Nature Switzerland, Cham (2023)
6. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jaeger, P.F.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 488–498. Springer (2024)
7. Jain, A., Mildenhall, B., Barron, J.T., Abbeel, P., Poole, B.: Zero-shot text-guided object generation with dream fields. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 867–876 (2022)
8. Mescheder, L., Oechsle, M., Niemeyer, M., Nowozin, S., Geiger, A.: Occupancy networks: Learning 3d reconstruction in function space. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4460–4470 (2019)
9. Michel, O., Bar-On, R., Liu, R., Benaim, S., Hanocka, R.: Text2mesh: Text-driven neural stylization for meshes. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 13492–13502 (2022)
10. Musio, F., Juchler, N., Yang, K., Shit, S., Prabhakar, C., Menze, B., Hirsch, S.: Circle of willis centerline graphs: A dataset and baseline algorithm. arXiv preprint arXiv:2510.13720 (2025)
11. OpenAI: Openai api. https://platform.openai.com (2026), accessed February 2026
12. Park, J.J., Florence, P., Straub, J., Newcombe, R., Lovegrove, S.: Deepsdf: Learning continuous signed distance functions for shape representation. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)
13. 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)
14. Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al.: Learning transferable visual models from natural language supervision. In: International conference on machine learning. pp. 8748–8763. PmLR (2021)
15. Sitzmann, V., Martel, J., Bergman, A., Lindell, D., Wetzstein, G.: Implicit neural representations with periodic activation functions. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) Advances in Neural Information Processing Systems. vol. 33, pp. 7462–7473. Curran Associates, Inc. (2020)
16. Slim, H., Elhoseiny, M.: Shapewalk: Compositional shape editing through language-guided chains. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 22574–22583 (June 2024)
17. 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)
18. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R.,

Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

19. Wang, C., Chai, M., He, M., Chen, D., Liao, J.: Clip-nerf: Text-and-image driven manipulation of neural radiance fields. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 3835–3844 (2022)
20. Yang, J., Shi, R., Wickramasinghe, U., Zhu, Q., Ni, B., Fua, P.: Neural annotation refinement: Development of a new 3d dataset for adrenal gland analysis. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. pp. 503–513. Springer Nature Switzerland, Cham (2022)
21. Yang, K., Musio, F., Ma, Y., Juchler, N., Paetzold, J., Al-Maskari, R., Scheffler, K., Gross, O., Rusu, R.T., Konukoglu, E., Grange, B., Merck, H., Sznitman, R., Wiest, R., Li, H., Chung, C.A., Schneider, M.A., Petersen, J., Menze, B.: Benchmarking the cow with the topcow challenge: Topology-aware anatomical segmentation of the circle of willis for cta and mra (2023)
22. Zhang, B., Tang, J., Niessner, M., Wonka, P.: 3dshape2vecset: A 3d shape representation for neural fields and generative diffusion models. ACM Transactions On Graphics (TOG) **42**(4), 1–16 (2023)