Journal of Digital Media & Culture Technology Vol.6, No.1, June (2026), pp.25-32
ISSN 2983-0664(p) 2983-0850(e) *http://dx.doi.org/10.29056/jdmct.2026.06.03*

Assessment of Foundation Models' Applicability to Visualization

Junseo Choi[1†] , Juhan Lim[2†] , Suhyeon Kim[3], Younhyun Jung[4*]

Abstract

Direct Volume Rendering (DVR) and Cinematic Rendering (CR) are representative visualization techniques that transform Computed Tomography (CT) and Magnetic Resonance Imaging (MRI)-based 3D volume data into intuitive images using color, lighting, and shading information. These visualization images allow efficient understanding of complex anatomical structures, but the results vary greatly depending on user-defined parameters such as transfer functions and lighting. Existing Region of Interest (ROI)-based automatic enhancement methods primarily rely on scribble inputs, leading to limited accuracy and consistency. Recently, foundation models capable of performing segmentation using simple prompts such as points or boxes have opened new possibilities. SAM has demonstrated strong segmentation performance on natural images, while MedSAM has shown high accuracy on medical images, suggesting the potential to apply intuitive ROI specification to visualization images as well. However, visualization images form a hybrid domain combining characteristics of natural images and medical images, making it unclear which model is more suitable. This study compares and evaluates the segmentation performance of SAM and MedSAM on DVR and CR-based visualization images to identify the model best suited for the hybrid domain and analyze its applicability.

Keyword : Cinematic Rendering, Direct Volume Rendering, Foundation Model, Vision Transformer, Visualization

1. Introduction

Medical image visualization is a rendering technique that transforms 3D volume data obtained from Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) into a form that can be understood intuitively [1]. Direct Volume Rendering (DVR) and Cinematic Rendering (CR) are representative

1 School of Computing, Gachon University, Gyeonggi-do, Republic of Korea [Graduate Student]
e-mail: wnstl1234@gachon.ac.kr
2 School of Computing, Gachon University, Gyeonggi-do, Republic of Korea [Graduate Student]
e-mail: dlawhks@gachon.ac.kr
3 School of Computing, Gachon University, Gyeonggi-do, Republic of Korea [Graduate Student]
e-mail: kih629@gachon.ac.kr
4 School of Computing, Gachon University, Gyeonggi-do, Republic of Korea [Professor]
e-mail: younhyun.jung@gachon.ac.kr (Corresponding author)

† These authors equally contributed to this manuscript

* This research was supported by the Regional Innovation System & Education (RISE) program through the Gyeonggi RISE Center, funded by the Ministry of Education (MOE) and Gyeonggi-do, Republic of Korea. (2025-RISE-09-A01), and by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (RS-2025-00554526).

Received(March 29, 2026), Review Result(1st: April 15, 2026), Accepted(June 13, 2026), Published(June 30, 2026)

volume visualization approaches. DVR is a technique for visualizing internal structures by using a transfer function to map voxel intensities in volume data to optical properties such as opacity and color [2], while CR applies physically based lighting to provide high-quality visualization images that are more realistic and convey a greater sense of depth [3-6]. Because such visualization images integrate high-dimensional spatial information that is difficult to obtain from a single cross-sectional slice into a single image, they enable users to perceive complex anatomical structures at a glance and interpret volumetric data through familiar visual cues such as color, lighting, and shading. These methods are essential for the efficient and accurate interpretation of volumetric medical data by clinicians and researchers [7].

Since it is impossible to simultaneously emphasize all information within volumetric data, users must specify regions of interest (ROIs) for visualization. User intent is expressed through factors such as transfer functions, lighting, and viewpoints, all of which significantly affect the visualization result. In anatomically complex regions such as the thorax, different transfer function settings may either cause the ribs and sternum to occlude the lungs and heart or increase their transparency to emphasize cardiac structures. Thus, identical volumetric data may yield different results depending on the user's ROI.

Manually adjusting visualization parameters to emphasize ROIs is unintuitive and time-consuming [8]. To enable more effective ROI highlighting, many studies have explored the automatic adjustment of visualization parameters to enhance the visibility of user-specified structures. However, because ROIs are typically specified directly within volumetric data in 3D space, mismatches frequently arise between the user's intended region and the actually selected area. To address this limitation, several approaches first visualize the volumetric data and then allow users to specify ROIs directly on the visualization images. Although highly intuitive, these methods primarily rely on manual scribble inputs, which limits their accuracy and consistency [9][10].

Recently, foundation models have emerged as a powerful approach for segmentation tasks in computer vision [11][12]. Foundation models learn general-purpose visual representations through large-scale data-driven pretraining and can flexibly interpret and segment ROIs using simple prompts such as points, boxes, and text. In particular, the Segment Anything Model (SAM) [13] demonstrates robust segmentation performance across diverse image domains without additional training, introducing a prompt-based segmentation paradigm that directly reflects user-specified locations or extents. Meanwhile, MedSAM [14], which extends SAM to the medical domain, accurately delineates major anatomical structures in CT and MRI cross-sectional images, suggesting the potential applicability of foundation models to medical visualization.

Leveraging the prompt-based segmentation capability of foundation models may overcome the limitations of conventional scribble-based ROI specification, enabling more intuitive and consistent user interaction. However, visualization images form a hybrid domain that combines natural-image characteristics, such as color, lighting, shading, and texture, with medical-image characteristics derived from complex anatomical structures. Consequently, it remains unclear whether visualization images are more closely associated with natural-image or medical-image distributions. Therefore, insufficient evidence exists regarding whether the natural-image-based SAM or the medical-image-based MedSAM is more suitable for segmentation of visualization images.

Accordingly, this study investigates whether prompt-based foundation models can be effectively applied to visualization images generated using DVR and CR techniques, and systematically compares the structure selection capabilities of SAM and MedSAM in this hybrid domain. Furthermore, we evaluate the robustness and consistency of both models for segmenting visualization images with substantial visual diversity and complex overlapping anatomical structures.

2. Related Work

2.1 Segment Anything Model

The Segment Anything Model (SAM) is a segmentation foundation model pretrained on the large-scale SA-1B dataset for interactive object segmentation through prompts such as points and boxes. SAM extracts global image features using a Vision Transformer (ViT)-based encoder and generates segmentation results by directly incorporating the user's intent through prompt features. This prompt-based design enables fast and intuitive segmentation of user-specified ROI, allowing the model to respond flexibly on the basis of user prompts even when data distributions differ greatly across domains.

2.2 Segment Anything Model for Medical Images

MedSAM initializes its encoder with SAM ViT-Base weights and fine-tunes the image encoder and mask decoder on medical images while keeping the prompt encoder frozen. This training strategy enables the model to learn medical image-specific representations while preserving prompt-based user interaction. Furthermore, MedSAM was trained on a large-scale dataset spanning diverse medical imaging domains, including CT, MRI, X-ray, ultrasound, endoscopy, dermoscopy, fundus, and pathology images, and demonstrated strong performance in cross-sectional anatomical segmentation. These results suggest

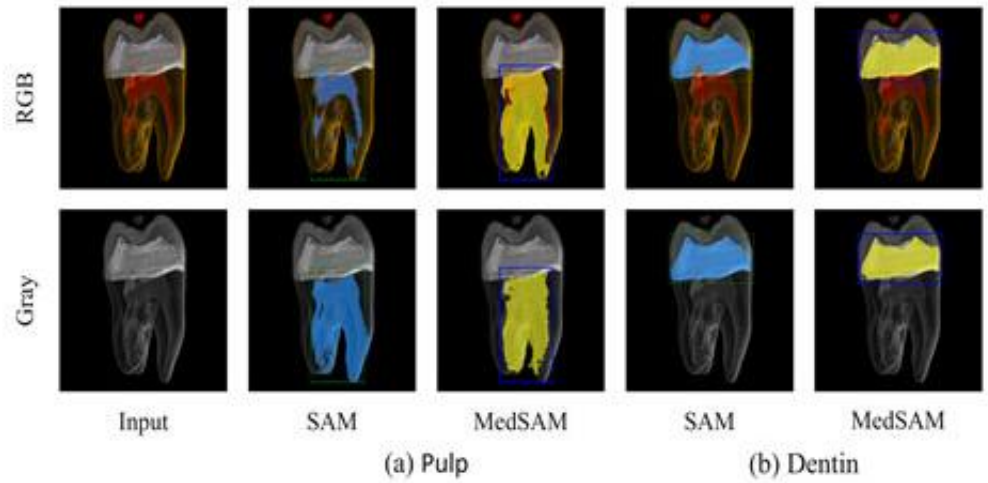
that the generalization capability of SAM can be effectively extended to the medical imaging domain.

3. Experimental Evaluation

3.1 Experiment Setup

This study uses Tooth, Head, and Abdominal CT datasets. The Tooth images were generated using DVR, whereas the Head and Abdominal images were generated using CR. For comparative analysis, SAM and MedSAM were initialized with pretrained ViT-Base weights provided by open-source. Furthermore, RGB images containing rich color information and grayscale images emphasizing anatomical structures were used to analyze the features prioritized by each model based on their segmentation results.

3.2 Image Segmentation Results and Analysis

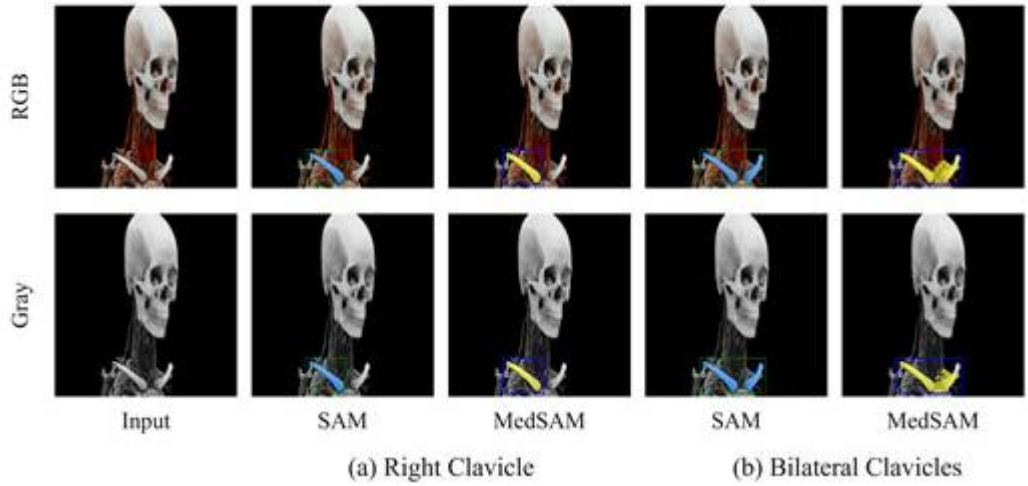


[Fig. 1] Segmentation results of Tooth data with direct volume rendering

[Fig. 1] compares the segmentation performance of SAM and MedSAM for pulp and dentin structures in Tooth DVR images. SAM produced more stable segmentation results under both RGB and grayscale conditions. The model preserved the overall contours of the pulp and dentin regions, demonstrating strong structural generalization from natural-image pretraining. However, under grayscale conditions, the lack of color cues caused over-segmentation around the cementum and crown regions. This result suggests degraded generalization for complex overlapping anatomical structures in DVR images.

MedSAM over-segmented the pulp and cementum regions and produced blurred dentin boundaries. This result suggests reduced segmentation stability caused by the distributional gap between the optical characteristics of DVR images and grayscale medical images.

Overall, DVR images represent complex anatomical structures through the transfer function, requiring segmentation methods that can handle such composite structures.



[Fig. 2] Segmentation results of Head data with cinematic rendering

[Fig. 2] compares the segmentation performance of SAM and MedSAM for clavicle structures in Head CR images. For the right clavicle, both models produced stable segmentation results. For the bilateral clavicles, SAM leveraged color information to segment both structures but also included parts of the similarly colored scapula, indicating limited use of anatomical structure cues. MedSAM, despite operating under conditions suited to its grayscale-based features, produced errors that included surrounding tissue. This result suggests that the contrast discrepancy between medical training images and CR visualizations reduces the model's ability to exploit anatomical structure information.

Therefore, SAM provides higher structural reliability for CR images with strong optical characteristics, whereas MedSAM is vulnerable to the distributional gap between CR and medical images.

[Fig. 3] compares the segmentation performance of SAM and MedSAM on kidney CR images. Both models captured the shape of the kidney, but tended to over-segment by including the internal vascular structures and soft tissue of the kidney. SAM partially separated surface vessels using color and intensity cues, whereas MedSAM was comparatively unable to separate the vascular structures within the kidney. These results suggest that the optical characteristics and structural complexity of CR images

reduce the generalization capability of features learned from medical images.

[Fig. 3] Segmentation results of Abdominal data with cinematic rendering

In the experiments on the three medical visualization datasets, SAM consistently leveraged the optical characteristics of DVR and CR images and produced stable segmentation results. However, it often fails to consistently capture the complex anatomical structures synthesized by transfer functions. Although MedSAM learned optical and anatomical features from medical images, its generalization capability degraded in transformed domains generated by volume rendering.

4. Conclusion

We evaluated whether foundation models trained on natural and medical images can effectively segment complex anatomical structures in DVR and CR visualizations. These visualizations combine the structural complexity of volumetric medical data with optical characteristics similar to natural images, limiting the generalization capability of existing foundation models. In the experiments, SAM produced relatively stable segmentation in regions with clear optical cues but degraded in areas with ambiguous boundaries caused by anatomical complexity and transparency variations. In contrast, MedSAM showed unstable performance because DVR and CR images exhibit a distribution shift from its medical-image training domain.

These results suggest that foundation models trained on natural and medical images cannot fully generalize to the diverse characteristics of volume-rendered images, while highlighting the potential of foundation models specialized for medical visualization. We demonstrate the feasibility of foundation models for visualization-based medical image analysis and provide a foundation for future automated analysis in volume-rendering environments.

References

[1] Q. Zhang, R. Eagleson, and T. M. Peters, "Volume visualization: A technical overview with a focus on medical applications," *Journal of Digital Imaging*, vol. 24, no. 4, pp. 660-664, Aug. 2011, doi: 10.1007/s10278-010-9321-6.

[2] K. Engel, M. Hadwiger, J. M. Kniss, A. E. Lefohn, C. R. Salama, and D. Weiskopf, "Real-time volume graphics," in *ACM SIGGRAPH 2004 Course Notes*, Los Angeles, CA, USA, Aug. 8-12, 2004, pp. 29-es, doi: 10.1145/1103900.1103929.

[3] M. Eid et al., "Cinematic rendering in CT: A novel, lifelike 3D visualization technique," *American Journal of Roentgenology*, vol. 209, no. 2, pp. 370-379, Aug. 2017, doi: 10.2214/AJR.17.17850.

[4] D. Comaniciu, K. Engel, B. Georgescu, and T. Mansi, "Shaping the future through innovations: From medical imaging to precision medicine," *Medical Image Analysis*, vol. 33, no. 1, pp. 19-26, Oct. 2016, doi: 10.1016/j.media.2016.06.016.

[5] K. Engel, "Real-time Monte Carlo path tracing," *scribd.com*, https://www.scribd.com/document/714940911/s6535-klaus-engel-real-time-monte-carlo-path-tracing-medical-volume-data (accessed May 25, 2026).

[6] G. Paladini, K. Petkov, J. Paulus, and K. Engel, "Optimization techniques for cloud based interactive volume tric Monte Carlo path tracing," *researchgate.net*, https://www.researchgate.net/publication/299366162_Optimization_Techniques_for_Cloud_Based_Interactive_Volumetric_Monte_Carlo_Path_Tracing (accessed May 25, 2026).

[7] A. H. J. Koning, M. Rousian, C. M. V. Dikkeboom, L. Goedknegt, E. A. P. Steegers, and P. J. V. D. Spek, "V-Scope: Design and implementation of an immersive and desktop virtual reality volume visualization system," in *Medicine Meets Virtual Reality 17: NextMed: Design For/the Well Being*, vol. 142, J. D. Westwood et al., Eds., IOS Press, 2009, pp. 136-138.

[8] L. Cai, W. L. Tay, B. P. Nguyen, C. K. Chui, and S. H. Ong, "Automatic transfer function design for medical visualization using visibility distributions and projective color mapping," *Computerized Medical Imaging and Graphics*, vol. 37, no. 7-8, pp. 450-458, Oct.-Dec. 2013, doi: 10.1016/j.compmedimag.2013.08.008.

[9] X. Zhang et al., "HELPNet: Hierarchical perturbations consistency and entropy-guided ensemble for scribble supervised medical image segmentation," *Medical Image Analysis*, vol. 105, no. 1, Art. no. 103719, Oct. 2025, doi: 10.1016/j.media.2025.103719.

[10] Y. Jing and T. Stathaki, "Size aware cross-shape scribble supervision for medical image segmentation," unpublished.

[11] M. Awais et al., "Foundation models defining a new era in vision: A survey and outlook," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2245-2264, Apr. 2025, doi: 10.1109/TPAMI.2024.3506283.

[12] T. Zhou et al., "Image segmentation in foundation model era: A survey," unpublished.

[13] A. Kirillov et al., "Segment anything," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, Paris, France, Oct. 1-6, 2023, pp. 4015-4026.

[14] J. Ma, Y. He, F. Li, L. Han, C. You, and B. Wang, "Segment anything in medical images," *Nature Communications*, vol. 15, no. 1, Art. no. 654, Jan. 2024, doi: 10.1038/s41467-024-44824-z.