

Rendering 3D CT Scans through 3D Gaussian Splatting Initialized with Points Sampled by Cube-based Neural Radiance Fields

Sanghyuk Roy Choi, Chanhoe Gu, Sun Jae Baek
Department of Intelligent Semiconductor Engineering
Chung-Ang University
Seoul, Korea
{choiroy, kum0100, baechu}@cau.ac.kr

Minhyeok Lee*
School of Electrical and Electronics Engineering
Chung-Ang University
Seoul, Korea
mlee@cau.ac.kr

Abstract—In this paper, we present Cube-Cloud 3D Gaussian Splatting (Cube-Cloud 3DGS), a novel framework designed to render medical image that inherently structured with three fixed axes. The methods such as COLMAP, Neural Radiance Field (NeRF), and 3D Gaussian Splatting are unsuitable for medical image reconstruction due to the lack of diverse viewpoints. To address this challenge, we propose Cube-Cloud 3DGS that leverages Cube-based Neural Radiance Field (CuNeRF) for cube-based sampling to generate point clouds from medical data. CuNeRF in Cube-Cloud 3DGS generates point cloud and renders images through various viewpoints which can be used as camera poses. We integrate the point cloud with 3D Gaussian Splatting that is initializing 3D gaussians. By utilizing the viewpoints extracted from CuNeRF, the parameters of 3D gaussians are refined. Cube-Cloud 3DGS renders images through its 3D gaussians while traditional models fail to render based on medical images. We evaluated Cube-Cloud 3DGS on the Kidney and Kidney Tumor Segmentation (KiTS23) dataset, demonstrating that our model reconstructs 3D medical volumes effectively. Therefore, our model resolves the limitation focusing on the internal features for medical images. Our model achieves higher performance of 2.707 in PSNR and 0.0504 in SSIM over existing 3D Gaussian Splatting.

Keywords—*NeRF, Cube-based NeRF, COLMAP, 3D Gaussian Splatting,*

I. INTRODUCTION

In recent advances in novel view synthesis, 3D Gaussian Splatting has gained attention as a novel approach for efficient and high-quality image rendering in real-time applications [1]. Traditional methods like Neural Radiance Fields (NeRF) rely on volumetric ray marching and Multi-Layer Perceptron (MLP) which suffer from significant computational overhead during both training and inference [2]. 3D Gaussian Splatting alleviates these challenges by introducing an efficient scene representation and rendering algorithm that reduces computational costs showing superior performance. 3D Gaussian Splatting utilizes a sparse point cloud derived from Structure-from-Motion (SfM) tool COLMAP, which serves as the initial set of 3D gaussians [3]. Each Gaussian is determined by a position, anisotropic covariance matrix, and opacity. This

representation efficiently captures the geometry and radiance information of the scenes.

In the context of novel view synthesis and 3D reconstruction, COLMAP is widely used for generating sparse point clouds and corresponding camera poses. COLMAP provides an efficient pipeline for reconstructing 3D scenes from 3D images, which is crucial stage for 3D Gaussian Splatting. COLMAP detects key features in the input images and matches the extracted features across multiple views. COLMAP computes camera poses and sparse 3D points through refining the spatial and optical information of the images. By using the point clouds calculated from COLMAP, 3D Gaussian Splatting creates a precise and optimized scene representation.

Applying COLMAP to medical image datasets, such as Computed Tomography (CT) scan and Magnetic Resonance Imaging (MRI), presents significant challenges [4, 5]. In 3D Gaussian Splatting, the point cloud generated by COLMAP is crucial for initializing the 3D gaussians. COLMAP relies on multiple 2D images of a scene taken from diverse viewpoints with the corresponding camera poses, to generate the point clouds. In contrast, medical image datasets inherently represent 3D volumes captured from fixed axis, such as x, y, and z axes, without the diversity of viewpoints required for COLMAP to generate point clouds. Therefore, attempting to apply COLMAP to such datasets is limited to producing point clouds.

To address the limitations of COLMAP, we employ Cube-based Neural Radiance Field (Cube-NeRF) [5]. Cube-NeRF overcomes the challenge posed by conventional NeRF models. Cube-NeRF introduces cube-based point sampling that allows the model to learn efficiently from the 3D volumetric data without diverse viewpoints, allowing the model suitable for medical image datasets.

We propose Cube Point Cloud 3D Gaussian Splatting (Cube-Cloud 3DGS) By integrating the point cloud sampled from Cube-NeRF, 3D gaussians are initialized effectively. This approach allows 3D Gaussian Splatting to perform view synthesis tasks resolving the limitation of COLMAP. Furthermore, we utilize the images rendered from Cube-NeRF as an input for 3D Gaussian Splatting. These images correspond to different viewpoints of the medical dataset and an significant advantage of Cube-NeRF is its ability to generate an unlimited number of images from various viewpoints. This unlimited ability implies that an unlimited number of input images for 3D

* Corresponding Author

This work was supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) (P0020967, Advanced Training Program for Smart Sensor Engineers) and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2024-00337250).

Gaussian Splatting can be generated. Therefore, our Cube-Cloud 3DGS effectively overcomes the limitation. We perform comprehensive evaluation on 2023 Kidney and Kidney Tumor Segmentation (KiTS23) dataset [6, 7]. The results demonstrate that Cube-Cloud 3DGS effectively captures the 3D representation of CT images, proposing detailed reconstruction of the volumetric medical dataset.

II. RELATED WORKS

A. Neural Radiance Field

NeRF have emerged as a powerful model for rendering the scenes from sparse input images [2]. NeRF reconstructs a scene as a continuous implicit representation calculated by a neural network composed of MLP layers. By sampling rays cast through pixels in the input images, NeRF utilizes volumetric rendering to predict the color of each pixel. NeRF effectively reconstructs the 3D scene by optimizing across multiple views. The advantage of NeRF lies in its ability to learn complex scene representation and render novel views by querying the network with positional encodings of 3D coordinates and view directions. NeRF exploits camera poses and input images from viewpoints for training. Therefore, NeRF is able to produce high-quality rendering images from a limited set of input images.

B. Cube-Based Neural Radiance Field

CuNeRF is designed to address a fundamental limitation that is inherent in medical image dataset, specifically that the data is organized along the three orthogonal x, y, and z axes [5]. Traditional methods are inferior to perform with this limitation due to the lack of diverse viewpoints and images. However, CuNeRF introduces a novel cube-based sampling approach that allows it to overcome the limitation. By sampling in 3D volumetric cube space, CuNeRF ensures extensive spatial coverage of the 3D representations and alleviates the issues caused by the fixed views typical in medical dataset. This approach enables CuNeRF to synthesize high-quality medical image from arbitrary viewpoints rendering increased resolution of the images with improved quality.

C. COLMAP

COLMAP is SfM tool designed for generating 3D reconstructions from 2D images [3]. COLMAP efficiently extracts features by matching the features across multiple views to identify key points that are common across the image dataset. Using the matched features, COLMAP computes camera poses through a nonlinear optimization technique that refines the camera parameters and 3D point coordinates. As a result, COLMAP produces a sparse point cloud representing the 3D structure of the scene. In the context of 3D scene reconstruction, COLMAP performs a crucial role by providing the initial sparse point cloud.

D. 3D Gaussian Splatting

The 3D Gaussian Splatting utilizes a sparse point cloud generated from SfM and converts the scene data into a 3D gaussians. The 3D gaussians represents complex scene geometry without dense voxel grids or ray-marching used in NeRF [1]. Each 3D gaussians are defined by their position, covariance matrix, opacity, and color. 3D

Gaussian Splatting learns both spatial and radiance information of the scene. During the optimization, the parameters of 3D gaussians are adjusted and additional 3D gaussians are copied in that the scene requires more detailed representation while the 3D gaussians of which opacity is lower than the pre-defined threshold are eliminated. A key component of 3D Gaussian Splatting is the tile-based rasterization algorithm which significantly accelerates both training and rendering speed. The 3D gaussians are projected into 2D plane and the projected 2D gaussians are rasterized using alpha blending. This process enables real-time rendering of complex scenes at high resolution with maintaining high visual quality. By adopting 3D gaussians and rasterization algorithm, 3D gaussian Splatting achieves competitive training time and real-time novel view synthesis.

E. Kidney and Kidney Tumor Segmentation Dataset

The KiTS23 is a medical dataset that is composed of 300 CT scans with the expert annotations provided for kidney and kidney tumor segmentation [6, 7]. The KiTS23 includes a diverse range of patients with different organ and tumor information. The KiTS23 is widely used in models that perform tumor segmentation tasks or medical image rendering tasks [8]. The KiTS23 dataset is available in <https://github.com/neheller/kits23>.

III. METHOD

Medical images including CT and MRI scans present unique challenges in applying traditional models such as NeRF or COLMAP. These methods rely on external surface information of an object for image reconstruction. In contrast, medical image primarily focuses on the internal structure rather than external surfaces. Also, medical image is inherently organized along limited three orthogonal axes making it difficult for NeRF and COLMAP to capture the volumetric information of the data. To address this limitation, we propose Cube-Cloud 3DGS that adapts CuNeRF that is suitable for medical image dataset.

In our proposed model in Figure 1, we utilize CuNeRF to create initial point cloud and camera pose for 3D Gaussian Splatting. The input medical volume has dimensions of $(512 \times 512 \times 512)$. Before training, we normalize medical volume into a range of $[-1, 1]$ which the center of medical volume is located to $(0, 0, 0)$. Once normalized, the volume is divided into cubes of size $(16 \times 16 \times 16)$. For each cube, the center is denoted as $\hat{x}_c = (x_t, y_t, z_t)$, and the normalization is performed using the following equation:

$$\hat{x}_c = \left(\frac{2(x_t - H)}{H + 2P}, \frac{2(y_t - W)}{W + 2P}, \frac{2(z_t - S)}{S + 2P} \right), \quad (1)$$

where H is height, W is width, S is the number of slices in medical image, and P is hyperparameter of the padding size. This normalization ensures precise point sampling each cube within the aligned medical volume. In each cube, we extract a fixed number of N sampling points. These points are uniformly sampled from the cube space $B(\hat{x}_c, \frac{l}{2})$, where l is the length of the normalized medical image.

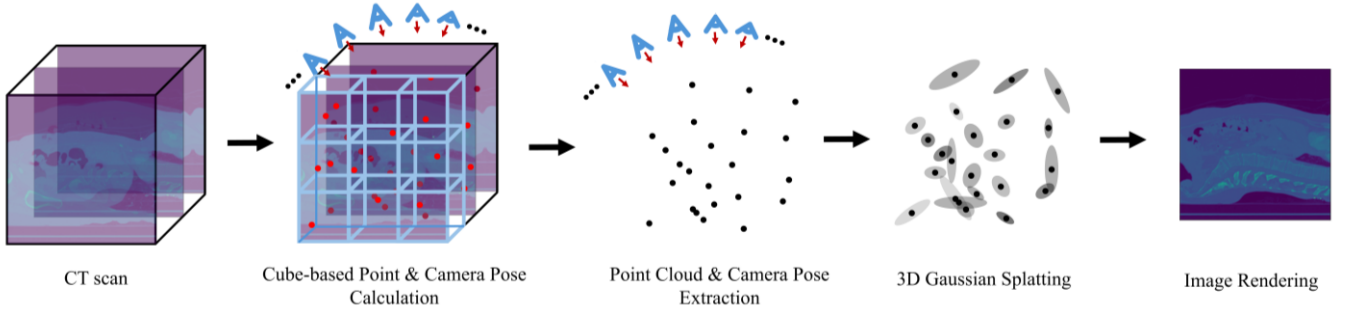


Fig 1. The structure of our proposed model. The red dots indicate the sampled points from corresponding view points. The black dots indicated extracted points and blue eye shaped figure indicates camera pose. The black dots are used for initial position of 3D gaussians. 3D CT images are rendered by projecting 3D gaussians and rasterization.

Within the each cube space, the sampled points $\hat{x}_i$ are selected under the uniform distribution U by:

$$\hat{x}_i \sim U \left[B(\hat{x}_c, \frac{l}{2}) \right]. \quad (2)$$

These sampled points are assembled into a point cloud effectively performing as the output of COLMAP point cloud. The corresponding view point parameters and the rendered images from CuNeRF are utilized as inputs for the 3D Gaussian Splatting component of Cube-Cloud 3DGS.

Once the 3D gaussians are initialized, the training stage focuses on optimizing the parameters of each 3D gaussians. Each 3D gaussians initialized in 3D space are projected on the 2D plane based on camera pose. To render multiple 2D gaussians in efficient manner, 3D Gaussian Splatting employs a differentiable tile rasterizer. This rasterization method divides the image projected on 2D plane into multiple tiles and calculates the color and opacity of 2D gaussians in each tile independently. The adaptive density control performs by evaluating the importance of each 3D gaussians during the rendering process. The important 3D gaussians that significantly contribute to the rendered image are retained and refined. Conversely, the 3D gaussians with low contribution are removed. The adaptive density control ensures that the balance between the quality of rendered image and computational efficiency.

We compute the loss by comparing the rendered images with the ground truth. The loss function is a combination of L_1 loss and Differential Structural Similarity Index (D-SSIM) and the equation is shown below:

$$L_t = (1 - \alpha)L_1 + \alpha L_{D-SSIM}, \quad (3)$$

L_{D-SSIM} indicates D-SSIM loss term. L_1 loss ensures that the pixel-wise difference between the rendered image and ground truth. D-SSIM is employed to measure perceptual quality. The value of α is set to 0.2 as inherited in the previous 3D Gaussian Splatting. Through minimizing the L_t loss, the spatial distribution of 3D gaussians and the perceptual quality of the rendered images are successfully optimized.

IV. RESULTS

We utilized KiTS23 dataset for training and evaluating our proposed model Cube-Cloud 3DGS. The CT scan of KiTS23 has a resolution of 512×512 pixels, and the number of slices per CT scan varies depending on each CT scan. In our experiments, we resized all volumes to a

uniform dimension of $(512 \times 512 \times 512)$. This preprocessing ensures that the data is compatible with our model while maintaining the essential features of the medical images. We applied min-max scaling to our dataset adjusting the pixel values into a range of 0 to 1.

We extracted 60 images and viewpoints through cube-based point sampling. We focused on the sagittal axis and performed 360-degree rotation capturing 60 images and viewpoints. While generating more viewpoints was feasible, we limited the number of images to 60 for computational efficiency.

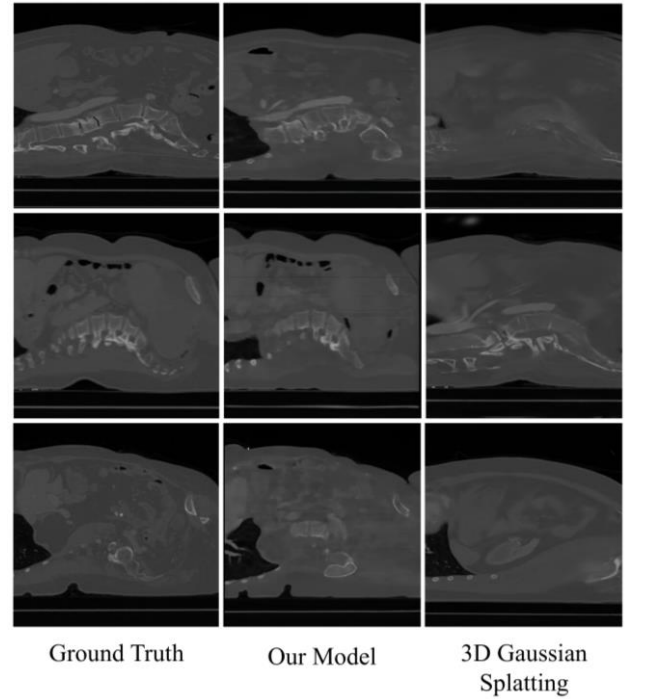


Fig 2. The rendering image from KiTS23 dataset of our model and 3D Gaussian Splatting. The first column is ground truth image for our model and the second column is the rendering image of our model. The third row is the ground truth image for the 3D Gaussian Splatting and the fourth row is the rendering image of the 3D Gaussian Splatting.

We trained our model with KiTS23 dataset and data extracted from CuNeRF such as point cloud, camera pose parameters, and images from the various viewpoints. For 3D Gaussian Splatting, we initialized the 3D gaussians randomly. We used the 2D images in the direction of a sagittal plane. For the camera pose parameter, we utilized the identical camera pose in direction of z axis. We

trained 3D gaussians with 30,000 iterations for both models.

In Figure 2, we compared the quality of images rendered by our model and 3D Gaussian Splatting. Our model captures various features in the image. However, 3D Gaussian Splatting has challenges in capturing features due to the restricted viewpoints. Additionally, while our model effectively handles major structures, our model fails to accurately capture small details, resulting in blurry renderings for fine-grained features.

In Table 1, we evaluated our model and 3D Gaussian Splatting using the metrics that calculate the quality of rendered images such as Peak Signal-to-noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). Our model demonstrates superior performance across all metrics compared to 3D Gaussian Splatting. In terms of PSNR, our model achieves a higher performance by 2.707, and for SSIM, our model outperforms by 0.0504. Additionally, our model shows a lower Mean Square Error (MSE), outperforming by 0.0181.

TABLE I. EVALUATION ON KITS23 DATASET

Model	Evaluation Metrics		
	PSNR	SSIM	MSE
Our Model	28.104	0.7856	0.0501
3D Gaussian Splatting	25.397	0.7352	0.0682

V. CONCLUSION

We propose Cube-Cloud 3DGS that utilizes cube-based point sampling to generate point clouds from CuNeRF. The generated point cloud serves as initialization for 3D gaussians which are optimized through a differentiable tile rasterizer and adaptive density control mechanism. Through rendering 2D images by CuNeRF, we are able to create input images for training 3D Gaussian Splatting. These enable rendering high quality volumetric medical images that is structured with three axes.

Through evaluation using the KiTS23 dataset, we show that Cube-Cloud 3DGS successfully captures 3D representation of medical images. Our Cube-Cloud 3DGS successfully renders images by focusing on analyzing the internal structure through CuNeRF which is unique approach overcoming the limitation of traditional models that are typically designed to capture and reconstruct external surface.

REFERENCES

- [1] Kerbl, B., Kopanas, G., Leimkuehler, T., and Drettakis, G.: '3D Gaussian Splatting for Real-Time Radiance Field Rendering', *ACM Trans. Graph.*, 2023, 42, (4), pp. Article 139
- [2] Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., and Ng, R.: 'Nerf: Representing scenes as neural radiance fields for view synthesis', *Communications of the ACM*, 2021, 65, (1), pp. 99-106
- [3] Schonberger, J.L., and Frahm, J.-M.: 'Structure-from-motion revisited', in Editor (Ed.)^(Eds.): 'Book Structure-from-motion revisited' (2016, edn.), pp. 4104-4113
- [4] Guo, J., Wang, J., Kang, D., Dong, W., Wang, W., and Liu, Y.-h.: 'Free-SurGS: SfM-Free 3D Gaussian Splatting for Surgical Scene Reconstruction', in Editor (Ed.)^(Eds.): 'Book Free-SurGS: SfM-Free 3D Gaussian Splatting for Surgical Scene Reconstruction' (Springer, 2024, edn.), pp. 350-360
- [5] Chen, Z., Yang, L., Lai, J.-H., and Xie, X.: 'CuNeRF: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution', in Editor (Ed.)^(Eds.): 'Book CuNeRF: Cube-based neural radiance field for zero-shot medical image arbitrary-scale super resolution' (2023, edn.), pp. 21185-21195
- [6] Alzu'bi, D., Abdullah, M., Hmeidi, I., AlAzab, R., Gharaibeh, M., El-Heis, M., Almotairi, K.H., Forestiero, A., Hussein, A.M., and Abualigah, L.: 'Kidney tumor detection and classification based on deep learning approaches: a new dataset in CT scans', *Journal of Healthcare Engineering*, 2022, 2022, (1), pp. 3861161
- [7] Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., and Han, M.: 'The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge', *Medical image analysis*, 2021, 67, pp. 101821
- [8] Choi, S.R., Lee, J., and Lee, M.: 'OrgUNETR: Utilizing Organ Information and Squeeze and Excitation Block for Improved Tumor Segmentation', *IEEE Access*, 2024