3D Gaussian Splatting for PointNet Object Classification

D.J.A. van Dale

CSE3000 - Bsc. Computer Science & Engineering, TU Delft

June 20, 2024

1 Background

- **Structure from Motion (SfM),** a photogrammetric technique used to reconstruct 3D structures from a series of 2D images taken from different viewpoints. By analyzing the motion of features across the images, it estimates the 3D positions of points in the scene as well as the camera positions and orientations.
- **3D Gaussian Splatting,** a volume rendering technique which involves representing a volumetric scene using a set of continuous differentiable Gaussian functions, referred to as "Splatting" to approximate the volume's density and appearance [1], able to generate novel-view scenes.
- **Classification** is a machine learning process where a program learns to categorize objects into different predefined classes, using a large set of training data.

2 Research Question

- *Can 3D Gaussian Splatting be used to improve the accuracy of a PointNet classification model?*
    - *Can the PointNet architecture be applied on 3D Gaussian Splatted Point Clouds?*
    - *What is the difference in accuracy between PointNet models trained on 3D Gaussian Splatted data and non-3D Gaussian Splatted Data?*

3 Data Generation

- Princeton's **ModelNet10** dataset consists of hundreds of 3D CAD models, divided into 10 categories of the most common objects in the world. The dataset was specifically designed for research on computer vision & computer graphics [2].
- The dataset also includes a default training-test split which I use for my research.

Class Name	Train	Test	Total
Bathtub	106	50	156
Bed	515	100	615
Chair	889	100	989
Desk	200	86	286
Dresser	200	86	286
Monitor	465	100	565
Nightstand	200	86	286
Sofa	680	100	780
Table	392	100	492
Toilet	344	100	444

Table 1: Amount of models in Modelnet10 per category

- Python script rotates a camera around every object and captures a series of views, used as input for the 3D Gaussian Splatting optimizer, resulting in sparse point clouds.
- Each point is transformed into a 3D Gaussian with multiple parameters: Position, rotation, scale, opacity and spherical harmonics, which are subsequently optimized.

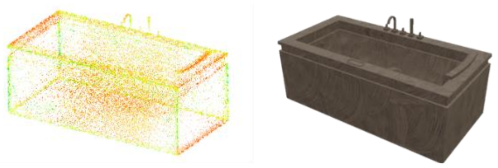

Figure 1: Visualization of 3D Gaussian Splatting point cloud of sample ModelNet10 bathtub object. (Left) The Gaussians at minimal scale. (Right) The Gaussians at maximal scale.

4 Methodology

- **PointNet/PointNet++** framework [3] to train the classification models
    - Uses point clouds as input directly
    - Number of features can be changed easily
    - PointNet++ addresses limitations by capturing local structures formed by neighbouring points
- Baseline is a model trained on points sampled from the surface of every mesh.
- 4 different configurations of features trained to test on
- Features are present as parameters in the Gaussians
    - Position (+3)
    - Rotation (+4) & Scale (+3)
    - Opacity (+1)
    - Spherical Harmonics (+45)
- Every cloud has been sampled using **Furthest Point Sampling (FPS)** with 8192 points

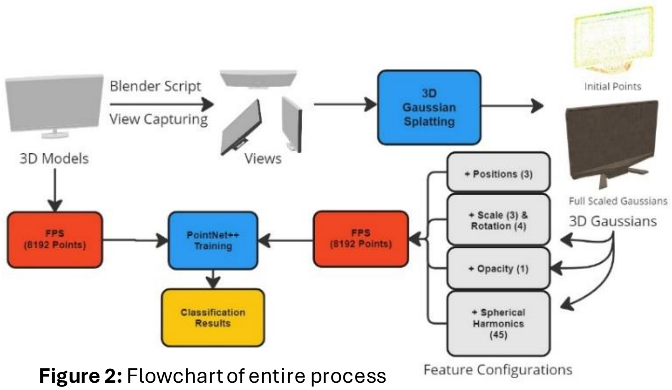

Figure 2: Flowchart of entire process

Figure 3: Furthest Point Sampling of 8192 points on sample object.

5 Results

- Configuration with the most optimal performance seems to be using positions, scales, rotations & opacity as features (11 total).

Network	Features	Accuracy
PointNet++	Positions	0.952

Table 2: Classification results of a PointNet++ model trained on points sampled directly from the ModelNet10 meshes. These results can be considered as the baseline for this research.

Network	Features	Accuracy
PointNet++	Positions	0.866
PointNet++	+ Scale & Rotation	0.887
PointNet++	+ Opacity	0.890
PointNet++	+ Sp. Harmonics	0.889

Table 3: Classification results of the PointNet++ trained on different configurations of 3D Gaussian Splatted ModelNet10 data

6 Conclusions

- Classification of 3D Gaussians achieves results on par with state-of-the-art models, but does not perform better than classifying on points directly sampled from the surface of the meshes.
- Spherical Harmonics, which allow for visibility-awareness, do not increase classification performance. The reason for this might be the fact that every model had the same light sources when capturing the views. Further research with real-world data could be done to determine the real impact of spherical harmonics as features.
- Adding Scale & Rotation as features gives the best accuracy improvement, suggesting that it defines the overall shape of the object better.

References

- **[1]:** Bernhard Kerbl et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering". In: ACM Transactions on Graphics 42.4 (July 2023).
- **[2]:** Zhirong Wu, Shuran Song, Aditya Khosla, Fisher Yu, Linguang Zhang, Xiaoou Tang, and Jianxiong Xiao. 3d shapenets: A deep representation for volumetric shapes, 2015.
- **[3]:** Charles R. Qi et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017. arXiv: 1612.00593 [cs.CV].