SPLATTER IMAGE: ULTRA-FAST SINGLE-VIEW 3D RECONSTRUCTION IMPROVEMENT BY TRAINING ON RGB-D DATASET

Dina Kogan[1] and Daniel Bobritski[2]

[1] *University of Haifa – <dinakogan90@gmail.com>*
[2] *University of Haifa – <danielbob32@gmail.com>*

ABSTRACT

This study investigates the performance of the Splatter Image, an ultra-fast approach for monocular 3D object reconstruction that operates at 38 FPS. The Splatter Image is based on Gaussian Splatting, which has been successful in real-time rendering, fast training, and scaling in multi-view reconstruction. For this work, we evaluate the application of Gaussian Splatting in a monocular reconstruction setting.
Our approach extends previous work by incorporating additional depth information into the model during training, aiming to enhance 3D reconstruction quality. Specifically, we modify the UNet architecture by adding an additional depth channel to improve the reconstruction performance. We evaluate these modifications on several standard datasets and compare the results with baselines using metrics such as PSNR, SSIM, and LPIPS.
Our experiments show that incorporating depth data leads to better performance, especially when using models like Depth Anything for depth estimation. We demonstrate that these improvements are consistent across multiple datasets, confirming that adding depth information during training can significantly enhance the quality of single-view 3D reconstructions.

1. INTRODUCTION

Reconstructing 3D objects from a single view is a challenging problem in computer vision, traditionally requiring extensive training on large datasets to achieve generalization across various object classes. Recent advancements, such as Gaussian Splatting, have demonstrated significant promise in multi-view reconstruction by efficiently handling real-time rendering and fast training. However, applying these methods to single-view reconstructions remains difficult, particularly when dealing with diverse and complex scenes.

In this work, we evaluate the Splatter Image framework, a technique based on Gaussian Splatting, but adapted for monocular 3D object reconstruction. The core of our investigation is to enhance the model's performance by integrating depth information during training. By incorporating additional depth channels, we aim to improve the model's ability to reconstruct high-quality 3D objects from single images, thus reducing the dependency on extensive multi-view data. Our experiments demonstrate how adding depth information can significantly improve the quality of 3D reconstructions in terms of standard evaluation metrics, such as PSNR, SSIM, and LPIPS.

2. RELATED WORK

3D object reconstruction from single-view images is a long-standing problem in computer vision. Traditional approaches often rely on shape priors, voxel grids, or point clouds to estimate the 3D structure from 2D images, but these methods tend to be computationally intensive and require large datasets for effective generalization. Popular methods such as ShapeNet or Pix2Vox leverage deep learning models but struggle with real-time performance and high computational costs.

Recent work has introduced Gaussian Splatting, an innovative approach for multi-view reconstruction that uses 3D Gaussian distributions to represent object geometry and appearance. This technique has shown promise in real-time rendering and scalability but has largely been applied to multi-view settings. Kerbl et al. (2024), for example, demonstrate the effectiveness of Gaussian Splatting in multi-view scenarios by leveraging real-time training capabilities and rendering at high frame rates.

While multi-view reconstruction methods are effective, their application to single-view reconstruction remains limited. Our work builds upon these prior methods by incorporating depth information directly during training, allowing for more robust single-view reconstructions. We also draw from Depth Anything, a model for depth estimation, to provide more accurate depth maps, which we use as additional input during training. By doing so, we aim to push the boundaries of monocular 3D reconstruction performance.

3. METHODOLOGY AND EXPERIMENTS

3.1. Datasets

In our experiments, we evaluated the performance of the Splatter Image approach on three different datasets:

- **SRN Cars**: This dataset was tested using three different subsets:
  - 100% of the data
  - 50% of the data
  - 20% of the data
- **CO3D Cars with Background**: This dataset includes real-world cars with diverse and complex backgrounds.

3.2. Experimental Setup

For each dataset, we ran a baseline model using RGB inputs and compared the results for three key metrics—PSNR, LPIPS, and SSIM—across different configurations:

- **RGB+D using Splatter Image Depth Output**: In this configuration, the depth channel was derived from the depth image produced by the Splatter Image model itself.
- **RGB+D using Depth Anything Model Output**: Here, the depth channel was generated using the Depth Anything model, which provides a more robust depth estimation based on external networks.

4. RESULTS AND ANALYSIS

In this section, we present the results of our experiments across all datasets and configurations. Our aim is to demonstrate the improvements in reconstruction quality by integrating depth information into the Splatter Image framework.

Table 1 summarizes the performance of the models for the experimental configurations across the datasets (SRN Cars, and CO3D Cars). We evaluated the models based on Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM), and Learned Perceptual Image Patch Similarity (LPIPS).

5. CONCLUSION

In this paper, we conducted a comprehensive evaluation of the Splatter Image framework, an ultra-fast approach for single-view 3D object reconstruction. This method, based on Gaussian Splatting, was tested in combination with additional depth information to improve reconstruction quality. Specifically, we explored the integration of depth maps into a modified UNet architecture to assess its impact on the accuracy of 3D reconstructions from single-view images.

Our experiments, which were constrained by limited computational resources, primarily focused on baseline models. Despite these limitations, our results showed that incorporating an additional depth channel led to improvements in PSNR, SSIM, and LPIPS across multiple datasets, with the Depth Anything model outperforming the Splatter Image's generated depth outputs. We believe that with more computational resources, our approach could sustain the performance improvement curve, further enhancing the accuracy and quality of the reconstructions.

Moreover, the visual and quantitative comparisons validated the effectiveness of depth data, particularly in improving object detail and reducing background artifacts. We hypothesize that scaling up the experiments, with larger datasets and more extensive iterations, would further substantiate the observed improvements.

For future work, we recommend expanding the dataset and conducting comparisons with equal amounts of computational resources across all configurations. Additionally, investigating more advanced neural architectures that can dynamically adapt to varying depth inputs and extending the framework to handle multi-view inputs and real-time dynamic object reconstruction will be critical for realizing the full potential of this approach.

Dataset	Configuration	PSNR	SSIM	LPIPS
SRN Cars (100%)	Baseline (RGB only)	**19.5569**	0.8334	0.2559
	RGB+D using Splatter Image Depth Output	18.9316	0.8244	0.2639
	RGB+D using Depth Anything Model Output	19.4645	**0.8361**	**0.2530**
SRN Cars (50%)	Baseline (RGB only)	**19.5290**	0.8326	0.2539
	RGB+D using Splatter Image Depth Output	18.9742	0.8225	0.2651
	RGB+D using Depth Anything Model Output	19.4829	**0.8374**	**0.2494**
SRN Cars (20%)	Baseline (RGB only)	19.3081	0.8298	**0.2554**
	RGB+D using Splatter Image Depth Output	18.7255	0.8193	0.2663
	RGB+D using Depth Anything Model Output	**19.3170**	**0.8329**	0.2567
CO3D Cars	Baseline (RGB only)	**14.0015**	**0.3806**	**0.6762**
	RGB+D using CO3D Depth Output	13.9242	0.3730	0.6883

Table 1. Performance metrics across datasets and experimental configurations.

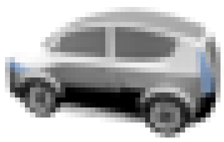

Figure 1. SRN Cars - Ground Truth (GT)

Figure 2. SRN Cars - Splatter Image Depth Output

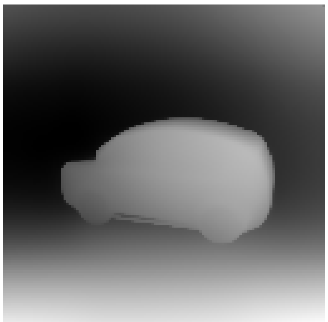

Figure 3. SRN Cars - Depth Anything Depth Output

Figure 4. CO3D Cars - Depth Input from Dataset

Figure 5. Comparison of depth images: Ground Truth, Splatter Image Depth Output, Depth Anything Depth Output, and CO3D Depth Input.

Figure 6. SRN Cars (100%) - Baseline (RGB)

Figure 7. SRN Cars (100%) - RGB+D Splatter Image Depth Output

Figure 8. SRN Cars (100%) - RGB+D Depth Anything Model Output

Figure 9. SRN Cars (100%) dataset: Baseline, RGB+D Splatter Image Output, and RGB+D Depth Anything Output.

Figure 10. SRN Cars (50%) - Baseline (RGB)

Figure 11. SRN Cars (50%) - RGB+D Splatter Image Depth Output

Figure 12. SRN Cars (50%) - RGB+D Depth Anything Model Output

Figure 13. SRN Cars (50%) dataset: Baseline, RGB+D Splatter Image Output, and RGB+D Depth Anything Output.

Figure 14. SRN Cars (20%) - Baseline (RGB)

Figure 15. SRN Cars (20%) - RGB+D Splatter Image Depth Output

Figure 16. SRN Cars (20%) - RGB+D Depth Anything Model Output

Figure 17. SRN Cars (20%) dataset: Baseline, RGB+D Splatter Image Output, and RGB+D Depth Anything Output.

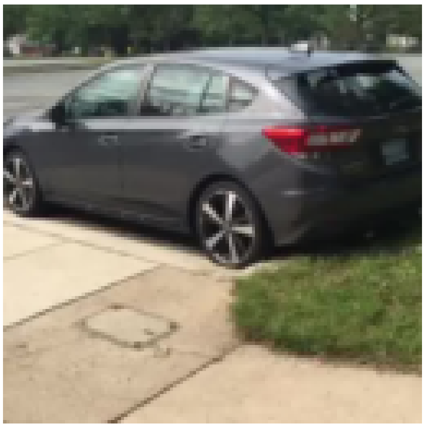

Figure 18. CO3D Cars - GT

Figure 19. CO3D Cars - Baseline (RGB)

Figure 20. CO3D Cars - RGB+D

Figure 21. CO3D Cars dataset: CO3D Cars - GT, CO3D Cars - Baseline (RGB), and CO3D Cars - RGB+D.

References

Szymanowicz, S., et al. (2024). Splatter Image: Ultra-Fast Single-View 3D Reconstruction. *arXiv preprint*, arXiv:2312.13150.

Szymanowicz, S., et al. (2024). Splatter Image GitHub Repository.

Kerbl, B., et al. (2024). Gaussian Splatting for Real-Time Rendering. *Graph Deco Inria GitHub Repository*.

Li, H., et al. (2024). Depth Anything v2: Advanced Depth Estimation Models. *GitHub Repository*.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. *arXiv preprint*, arXiv:1505.04597.

[1]Splatter Image arXiv preprint: <https://arxiv.org/pdf/2312.13150>
[2]Splatter Image GitHub Repository: <https://github.com/szymanowiczs/splatter-image>
[3]Gaussian Splatting GitHub Repository: <https://github.com/graphdeco-inria/gaussian-splatting>
[4]Depth Anything v2 GitHub Repository: <https://github.com/DepthAnything/Depth-Anything-V2>
[5]U-Net arXiv preprint: <https://arxiv.org/pdf/1505.04597>