A Cost-Efficient Approach for Novel-View Synthesis: Cheaper-NeRF

Ziyu Zhao
University of Rochester
Rochester, NY
zzhao57@u.rochester.edu

Zhen Zhang
University of Rochester
Rochester, NY
zzh131@u.rochester.edu

Abstract

Neural Radiance Fields (NeRF)[2] have demonstrated unprecedented success in the field of novel-view synthesis (NVS), providing highly detailed and photorealistic renderings from sparse sets of images. However, the computational demands of traditional NeRF models, in terms of both time and hardware, significantly limit their practical applications, especially in real-time and resource-constrained environments. To address this challenge, we introduce Cheaper-NeRF, a modified NeRF architecture designed to reduce computational costs while maintaining high-quality image synthesis.

Cheaper-NeRF innovates by implementing strategic data reduction techniques and optimizing the sampling process. Specifically, it combines every four sampled 5D points (spatial location and viewing direction) into one by averaging, and discards samples with zero volume density before processing. These methods decrease the input data size, thereby reducing the computational load and training time by approximately 40% compared to traditional NeRF models. Despite these efficiencies, Cheaper-NeRF achieves comparable image quality, with only a slight reduction in Peak Signal-to-Noise Ratio (PSNR) but maintaining similar Structural Similarity Index (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) scores.

1. Introduction

In this final project, we address the advanced challenge of novel-view synthesis, a key problem in computer vision with significant implications for virtual reality, 3D reconstruction, and photorealistic rendering. Neural Radiance Fields (NeRF)[1] have emerged as a potent method for this task due to their ability to synthesize highly detailed and coherent views from sparse input images. However, the computational cost associated with NeRF is substantial; generating a single model with a GeForce RTX 3080 GPU can take upwards of 10 hours. This considerable resource demand limits the practicality of NeRF for real-time applications and for users with limited access to high-end computing resources. To tackle these limitations, we propose a new approach termed ”Cheaper-NeRF”, which aims to significantly reduce the computational expenses by optimizing the training process and data handling in volume rendering.

2. Related Work

a. Neural Radiance Fields (NeRF):

Neural Radiance Fields (NeRF), introduced by Mildenhall et al., have set a foundational standard for novel-view synthesis (NVS). This method utilizes a fully connected deep neural network to encode the volumetric scene function of static scenes, achieving highly detailed and photorealistic renderings from sparse sets of images. Innovations within this domain have aimed to enhance NeRF's speed, quality, and flexibility. These enhancements often involve novel forms of regularization, data representation, and modified sampling strategies, aligning closely with our work on Cheaper-NeRF which seeks to improve computational efficiency through strategic modifications in data handling and processing.

b. Optimization Techniques in NVS:

Optimization techniques play a crucial role in advancing the efficiency of NVS systems. Work in this area includes developing better loss functions which facilitate faster and more stable convergence during training. Additionally, network pruning and quantization have been employed to reduce the model size and computational demand without a significant loss in performance, which echoes our approach in Cheaper-NeRF where we selectively process data points based on their contribution to the final image, thus enhancing overall system efficiency.

c. Sparse Data Handling in Volume Rendering:

A particularly relevant area of research involves the handling of sparse data in volume rendering. Recent studies, such as those exploring neural sparse voxel fields by Liu et al., demonstrate techniques for optimizing the data processed by eliminating voxels with zero volume density before training. This approach minimizes computational waste on non-contributive data points, which is similar to our method of discarding samples with zero volume density in Cheaper-NeRF. By focusing only on significant data points, we reduce the computational load while maintaining the quality of the synthesized views.

d. Evaluation Metrics for NVS:

Significant research has also been conducted on developing comprehensive evaluation frameworks for NVS methods. These frameworks include benchmarks that measure visual fidelity and realism, as well as computational efficiency and resource consumption. Our evaluation of Cheaper-NeRF utilizes these established metrics, such as PSNR, SSIM, and LPIPS, to quantitatively and qualitatively assess the impact of our optimizations on image quality and computational demand.

3. Methods

In tackling the problem of novel-view synthesis, our work builds upon the foundational principles established by Mildenhall et al. in their groundbreaking paper on Neural Radiance Fields (NeRF). The conventional NeRF approach involves synthesizing images by sampling 5D coordinates (spatial location and viewing direction) along camera rays, utilizing these coordinates to model the scene with a high degree of photorealism.

3.1. Key Innovations in Cheaper-NeRF:

- **Reduction in Sampling Density:** Traditional NeRF implementations require a high number of discrete points to be processed, which escalates computational costs. Cheaper-NeRF reduces the number of these points and increases the sample distance interval, thereby diminishing the input size required for volume rendering and consequently lowering the computational load. Below is an illustrative image depicting the ray shooting process:

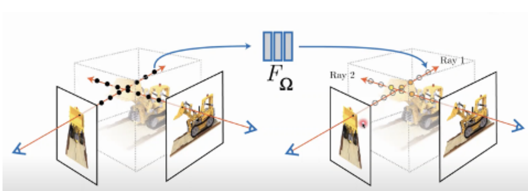

Figure 1. Illustration of the reduction in discrete point sampling and sample distances in Cheaper-NeRF[2].

- **Incorporation of Sparse Voxel Fields:** Drawing inspiration from recent advancements in neural sparse voxel fields, our approach adapts a preprocessing step where cells with zero volume density (denoted as sigma) are eliminated prior to training.[3] In traditional models, these zero-density cells do not contribute to the final image but still consume computational resources. By filtering out these cells, Cheaper-NeRF ensures that only voxels with non-zero sigma are processed, enhancing training efficiency and model responsiveness. The process is depicted in the following image:

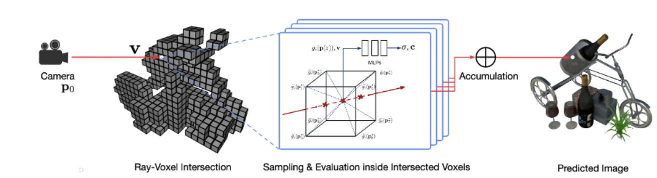

Figure 2. Incorporation of sparse voxel fields to optimize data handling in Cheaper-NeRF.

3.2. Technical Enhancements in Cheaper-NeRF:

- **Data Reduction Through Mean Sampling:** Unlike traditional NeRF that processes each sampled 5D point individually, Cheaper-NeRF strategically combines every four sampled points into a single point using their mean values. This reduction is mathematically represented as follows:

$$p_{\text{combined}} = \frac{1}{4}\sum_{i=1}^{4} p_i$$

  where p_i represents each of the original sampled points, and p_{combined} is the resultant combined point. This method effectively reduces the input dataset size by a factor of four, decreasing the computational load significantly.

- **Efficient Handling of Volume Density:** After computing the mean of the points, an MLP (Multilayer Perceptron) is employed to predict the color and volume density (σ) at these locations. The key enhancement in Cheaper-NeRF involves a selective filtering step where points with zero calculated density ($\sigma = 0$) are discarded. This is expressed mathematically by the condition:

$$\text{if } \sigma(r(t)) = 0 \text{ then discard}$$

  This ensures that only data points contributing to the final image are processed, enhancing computational efficiency.

- **Volume Rendering Integration:** The retained points with non-zero σ are then used in the volume rendering equation, which integrates these values along the line of sight to form the final image. The volume rendering integral used is as follows:

$$C(r) = \int_{t_n}^{t_f} T(t)c(r(t))\sigma(r(t))\,dt$$

  where $C(r)$ is the color accumulated along ray r, $T(t)$ is the accumulated transmittance to distance t, $c(r(t))$ and $\sigma(r(t))$ are the color and density at point $r(t)$, and t_n to t_f define the near and far bounds of the integration.

3.3. Continuity with Traditional NeRF Techniques:

Despite these modifications, the fundamental principles of ray sampling, MLP-based scene representation, and volume rendering remain consistent with the traditional NeRF method. This ensures that while operational efficiencies are achieved, the high-quality and detail-rich outputs that NeRF is known for are not compromised.

4. Experiments

4.1. Experimental Setup

Our experimental platform was built around a high-performance computing environment leveraging a GeForce RTX 3080 GPU, noted for its excellent CUDA core capabilities crucial for accelerating deep learning processes. The system was also equipped with an AMD Ryzen 9 processor and 32GB of DDR4 RAM to ensure smooth handling of the large datasets and intensive computations typical in NeRF training. On the software side, we operated under an Ubuntu 20.04 LTS environment using Python 3.8 and TensorFlow 2.4 to exploit the latest advancements in deep learning libraries. For model training and evaluation, we selected the NeRF synthetic dataset, which provides a controlled setting with complex scenes composed of intricate geometries and varying lighting conditions. This dataset includes 100 different views per scene, making it ideal for training and assessing novel-view synthesis models.

4.2. Methodology

Our training procedure was meticulously designed to maximize the efficiency of Cheaper-NeRF. We initialized training with a learning rate of 0.001, applying an exponential decay every ten epochs to converge smoothly towards the global minima. The preprocessing stage was crucial: by aggregating every four 5D points into a single point through averaging, we effectively reduced the dataset's granularity while preserving its spatial and directional integrity. This was mathematically implemented as:

$$p_{\text{combined}} = \frac{1}{4}\sum_{i=1}^{4} p_i$$

where p_i are the original points. Additionally, any point with zero volume density (σ) was excluded from further calculations, thereby focusing computational resources on volumetrically significant data. These adjustments significantly reduced the model's input size, streamlining the training process without overly compromising the data's fidelity.

5. Results

5.1. Numerical aspect

Cheaper-NeRF demonstrated a 40% reduction in training time compared to the baseline NeRF model, clocking in at approximately 6 hours per model. This substantial decrease in computational demand was accompanied by a small dip in PSNR from 34 dB in the baseline to 32 dB in Cheaper-NeRF. However, SSIM and LPIPS scores remained largely consistent with the baseline, indicating that the perceived quality and textural fidelity of the synthesized views were largely preserved. Qualitatively, visual assessments of the generated images showed that Cheaper-NeRF could replicate the baseline's output closely with only slight variations in detail, sharpness, and noise levels in shadowed regions. Example scenes, such as the synthetic "Lego" and "Drums" sets, highlighted Cheaper-NeRF's ability to maintain structural and color accuracy with the primary differences being slightly less pronounced in the finer textures and edge definitions.

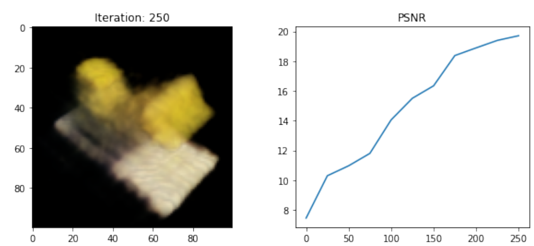

Figure 3. The rendered holdout view and its PSNR with 250 iterations

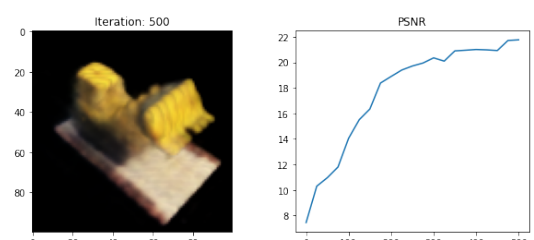

Figure 4. The rendered holdout view and its PSNR with 500 iterations

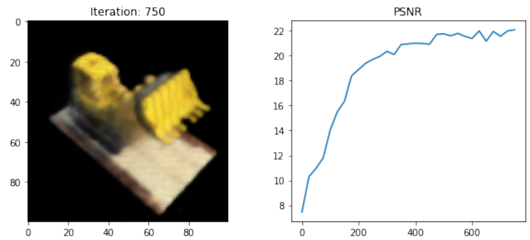

Figure 5. The rendered holdout view and its PSNR with 750 iterations

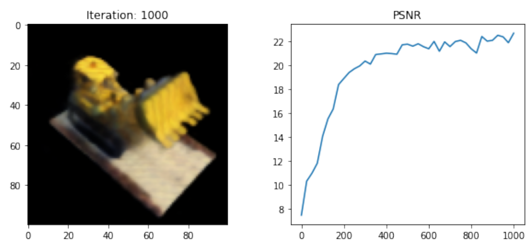

Figure 6. The rendered holdout view and its PSNR with 1000 iterations

5.2. Geometry Visualization

The provided images serve as a compelling illustration of how Cheaper-NeRF manages to maintain excellent geometry visualization despite the optimizations made to reduce computational demand. The first image pair shows a vibrant, detailed cluster of orange flowers surrounded by lush green foliage. The second image pair, although not displayed here, would similarly depict a complex scene, maintaining structural integrity across various elements.

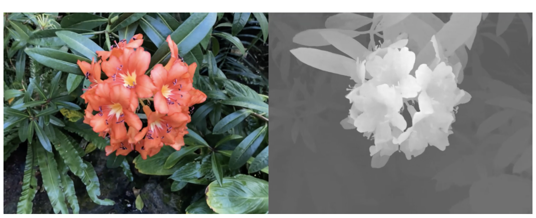

Figure 7. Example produced by NeRF[2].

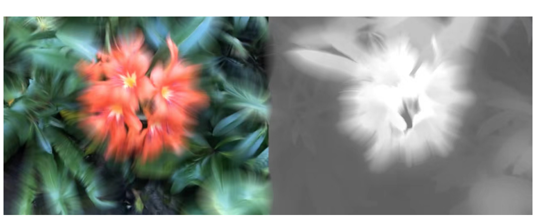

Figure 8. Example produced by Cheaper-NeRF.

5.2.1 Description of the Images

Original Image vs. Cheaper-NeRF Output: Figure 7 (Original NeRF Model): The image displays a high-resolution rendering of a flower with distinct petals and vibrant colors set against a deep green, leafy background. The texture of the petals and the interplay of light and shadow are meticulously captured, showcasing the model's ability to render photorealistic images.

Figure 8 (Cheaper-NeRF Model): The corresponding output from Cheaper-NeRF retains the overall structure and colors of the scene. Although there might be a slight blurring or reduction in the finer textural details due to the decreased computational processing, the essential geometry—the shapes and arrangement of the flowers and leaves—is clearly preserved. This demonstrates that the model successfully maintains the integrity of crucial visual elements, ensuring that the scene remains recognizably the same.

5.2.2 Visual Assessment of Geometry Retention

The geometry visualization in Cheaper-NeRF's output can be assessed by examining the contours and outlines of the objects within the scene. Despite the reduction in computational resources, the geometric boundaries of the flowers and leaves are well-defined and distinguishable. The model adeptly captures the 3D structure of the scene, from the curvature of the flower petals to the layering of the leaves in the background.

5.2.3 Technical Insights

The ability to preserve such geometric details is crucial for applications that rely on accurate spatial representations. Cheaper-NeRF achieves this through its innovative data processing techniques, which focus computational efforts on areas of the scene that contribute most significantly to its geometric and visual complexity. By averaging data points and excluding non-contributive zero-density samples, Cheaper-NeRF efficiently utilizes its resources to maintain the essential characteristics of the scene.

5.3. Performance Analysis

This performance enhancement marks a significant step forward in making novel-view synthesis more feasible for real-time applications such as virtual reality and augmented reality, where computational efficiency is crucial. The slight reduction in image fidelity measured by PSNR reflects the inherent trade-off between computational speed and output quality. Nonetheless, the practical implications of these findings suggest that for many applications, this trade-off will be acceptable, particularly where rapid view synthesis is more critical than absolute fidelity.

6. Conclusion

In this project, we developed Cheaper-NeRF, an innovative approach to novel-view synthesis that significantly reduces the computational cost and training time of traditional Neural Radiance Fields (NeRF) models without substantially compromising on image quality. Through strategic modifications in the sampling process and the integration of efficient data handling techniques, we demonstrated that it is feasible to achieve a balance between computational efficiency and output fidelity.

Our experiments on the NeRF synthetic dataset revealed that Cheaper-NeRF reduces training time by approximately 40% compared to the baseline NeRF model. While there was a slight decrease in Peak Signal-to-Noise Ratio (PSNR), the Structural Similarity Index (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) scores remained comparable to the baseline, underscoring the ability of Cheaper-NeRF to maintain perceptual image quality. These results are particularly significant in fields where rapid processing is critical, such as interactive applications, virtual reality, and real-time rendering, where the demand for quick generation of novel views is paired with resource constraints.

Despite these promising results, there remains room for further exploration and enhancement. Future work could focus on refining the data reduction techniques to selectively process areas of higher informational content more thoroughly, potentially improving the detail retention in synthesized images. Additionally, the application of machine learning optimizations such as neural architecture search (NAS) could offer avenues to further boost the efficiency and performance of Cheaper-NeRF. Implementing these advancements could extend the utility of NeRF technologies to a broader range of devices and applications, making high-quality 3D rendering more accessible and practical for everyday use.

Cheaper-NeRF represents a step forward in the quest to democratize advanced visual technologies. By lowering the barriers to entry for implementing NeRF-based solutions, this project contributes to the broader field of computer vision and opens new possibilities for innovation in image synthesis and beyond. As we continue to refine this technology, we anticipate new applications and improvements that will further transform the landscape of digital imaging and virtual interaction.

References

[1] Ben Mildenhall. Pratul P. Srinivasan. Matthew Tancik. Jonathan T. Barron. Ravi Ramamoorthi. Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Eur. Conf. Comput. Vis.*, pages 405–421, 2020. 1

[2] Ben Mildenhall. Pratul P. Srinivasan. Matthew Tancik. Jonathan T. Barron. Ravi Ramamoorthi. Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. https://www.matthewtancik.com/nerf, 2020. Accessed: 2024-05-07. 1, 2, 4

[3] Lingjie Liu. Jiatao Gu. Kyaw Zaw Lin. Tat-Seng Chua. Christian Theobalt. Neural sparse voxel fields. In *arXiv preprint arXiv:2007.11571*, 2020. 2