Study on reducing the impact of irrelevant Objects in 3D Reconstruction Using NeRF

Qihui Fan
Student, MS Data Science
fan.qih@northeastern.edu

Zuyu Guo
Student, MS Computer Science
guo.zuy@northeastern.edu

Yiqian Li
Student, MS Computer Science
li.yiqian@northeastern.edu

Abstract

3D Reconstruction of objects from 2D images has always been an important topic in the field of Computer Vision. The technology has been widely applied across various industries, ranging from medical research and map applications to the increasingly popular fields of virtual reality **(VR)** and augmented reality **(AR)**. **Neural Radiance Fields (NeRF)** as the state-of-art 3D reconstruction model requires a large amount of high-quality training images to produce an acceptable reconstruction. However, irrelevant objects in the training images significantly impact the quality of synthesized views of NeRF. Motivated by various approaches to best segment such elements present in the training images, this project therefore aims to enhance the performance of NeRF by such approaches. These approaches, such as the Mask R-CNN, are comprehensively tested and discussed in this project. The robustness of NeRF in real-world scenarios, 3D reconstruction in particular, is the priority of this project.

Keywords: 3D reconstruction, NeRF, Mask R-CNN, Image Segmentation, Scene Optimization

1 Introduction

3D reconstruction is a process of creating a three-dimensional model from two-dimensional images or other data sources. The concept has its roots in the fields of computer vision, photogrammetry, and graphics, with significant advancements occurring as computing power and imaging technologies improved. Early techniques primarily relied on stereoscopic images and manual measurements, but modern methods leverage algorithms like Structure from Motion (SfM), Simultaneous Localization and Mapping (SLAM), and deep learning models to achieve more accurate and automated reconstructions. These techniques are widely used in applications ranging from medical imaging and archaeology to virtual reality and robotics, enabling precise digital representations of real-world objects and environments.

2 Related Works

NeRF: Neural Radiance Field (NeRF) is an advanced 3D reconstruction technique based on Neural Network, which can generate high-quality 3D models from the related group 2D images. Unlike

Preprint. Under review.

traditional reconstruction methods that aim to recover the full geometry of a 3D scene, NeRF instead generates a 'radiance field', a volumetric model that provides the color and density for each point within the 3D space. (Mildenhall, Srinivasan, Tancik, Barron, Ramamoorthi, & Ng, 2021).

The core of this project is to reduce the impact of irrelevant objects in the NeRF reconstruction. We have reviewed the following previous works that offer valuable insights:

- **U-Net**: The U-Net architecture consists of a symmetric design with two main parts: a contracting path (encoder) and an expansive path (decoder). The contracting path is similar to a traditional convolutional network and is responsible for feature extraction; it repeatedly applies convolutions followed by max-pooling operations, progressively reducing the spatial dimensions while increasing feature depth. The expansive path performs upsampling of the feature maps, using transposed convolutions to restore the image resolution. It combines these upsampled features with high-resolution features from the contracting path, allowing the network to refine and localize the segmentation precisely. This design excels at preserving spatial information and ensures accurate pixel-level classification, making it highly effective for segmentation tasks. (Ronneberger, Fischer, & Brox, 2015)
- **Mask RCNN**: Mask R-CNN is developed based on its predecessor, Faster R-CNN. Faster R-CNN focuses on object detection, which involves locating and identifying objects in an image. However, it cannot distinguish different instances of the same class or produce pixel-level contours of objects. Mask R-CNN improves upon this by introducing pixel-level segmentation capabilities, making it excel in instance segmentation tasks. Mask R-CNN can not only detect objects in an image but also generate pixel-level segmentation masks for each detected object, which is essential for instance segmentation. (He, Gkioxari, Dollár, & Girshick, 2017).
- **GAN**(Generative Adversarial Network): As generative neural networks, GAN are introduced to generate new, synthetic data that resembles real data. A GAN consists of two competing networks: a generator and a discriminator. The generator attempts to produce realistic data, while the discriminator tries to differentiate between real and generated data. This adversarial process leads the generator to produce increasingly realistic data over time. In this project, GANs can be used to refine the 3D reconstruction process by synthesizing realistic textures or removing unwanted objects from the training data, allowing for improved scene reconstruction with cleaner inputs. (Goodfellow et al., 2014).

3 Project Goal

This project would attempt to best eliminate the impact of irrelevant objects from the main subject matter. One draft approach is to process the training images leaving only interested objects. **U-Net** and **Mask RCNN** could be utilized to filter and remove such unrelated objects for cleaner training data.

In addition to the potential usage of U-Net and Mask RCNN, **Generative Adversarial Network(GAN)** may be able to optimize such mask-cleaning on the training data, by comparing and converging the mask results leading to optimized input data for further NeRF.

4 Dataset Description

The training dataset consists of a large amount of photos/images about a particular outdoor architecture, such as the Rome Colosseum, from various viewing points. Data will be obtained on the Internet. The raw dataset may contain noise such as walking people and different shapes of vegetation.

5 Plan of activities

1. **Initial Research and Model Setup:**
    - Conduct a comprehensive review of the NeRF (Neural Radiance Fields) model to understand its capabilities and limitations.
    - Collect a dataset of high-quality images of the Colosseum from various internet sources.

2. **Preliminary 3D Reconstruction:**
    - Apply the NeRF model to the collected images to generate an initial 3D reconstruction of the Colosseum.
    - Evaluate the initial model output, focusing on identifying areas of slow processing or inaccuracies, particularly those attributed to low-quality inputs or the presence of noise.
3. **Noise Segmentation and Data Cleansing:**
    - Implement advanced segmentation algorithms such as U-Net and Mask R-CNN to isolate and remove noise elements (e.g., pedestrians, vegetation) from the dataset.
    - Utilize Generative Adversarial Networks (GANs) to synthesize realistic image fill-ins for the areas removed during the segmentation process.
4. **Refined 3D Reconstruction:**
    - Re-apply the NeRF model using the cleansed and enhanced dataset to assess improvements in the speed and accuracy of the 3D reconstruction.
    - Document the performance improvements and any remaining challenges.
5. **Performance Evaluation and Reporting:**
    - Compare the outputs before and after the implementation of noise management strategies.
    - Prepare a detailed report documenting the methodologies used, results obtained, and insights gained from the project.

Expected Outcomes:
The project aims to demonstrate the effectiveness of combining noise segmentation and data enhancement techniques with NeRF to improve the quality and efficiency of 3D reconstructions from noisy datasets.

References

[1]Rabby, A. K. M., & Zhang, C. (2023). BeyondPixels: A comprehensive review of the evolution of neural radiance fields. *arXiv preprint arXiv:2306.03000.*

[2] Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, **65(1)**, 99-106.

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing, 2015.

[4] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969).

[5]Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems, 27.*