NeRFlect: Reflecting NeRFs as Dataset Generators

Korhan Kemal Kaya*

kemal.kaya@ug.bilkent.edu.tr

1. Introduction

Neural Radiance Fields (NeRF) utilize an advanced method for 3D scene representation, facilitating the generation of novel views from limited 2D images [5]. This approach is capable of producing high-resolution images and effectively dealing with complex geometries.

Despite NeRF's advancements, a significant challenge remains in the area of image datasets. The scarcity of diverse and comprehensive image datasets limits the development and assessment of various applications and algorithms, such as virtual reality applications and body-pose estimation algorithms.

NeRF models offer a robust and efficient solution to address the existing gaps in image datasets. By utilizing NeRF's inherent strengths in capturing and reconstructing fine scene details, it is possible to create new, high-quality image datasets. Moreover, using NeRF as a generative model enables the production of images that are not only visually precise but also diverse, supporting the enhancement of more reliable and stable algorithms. This research introduces NeRFlect, a tool designed to aid in generating diverse datasets that may further drive advancements in related applications by modifying and utilizing various NeRF models. Fig. 1 illustrates the concept of the work accomplished.

For more details, please refer to our GitHub repository.

2. Related Work

2.1. Neural Radiance Fields

Trained on a set of input images of a scene, NeRF operates a fully connected deep network (non-convolutional in nature) to map a 3D coordinate (**x, y, z**) and a viewing direction (θ, ϕ) to corresponding RGB color and opacity values. It employs classic volume rendering techniques to composite single pixels and generate novel views of the scene. As the rendering process is inherently differentiable, NeRF can optimize scene representations based solely on the camera poses from the input images. This method outperforms previous techniques in novel view synthesis and captures complex geometries and textures effectively [5]."

*Bilkent University, Ankara, Turkey

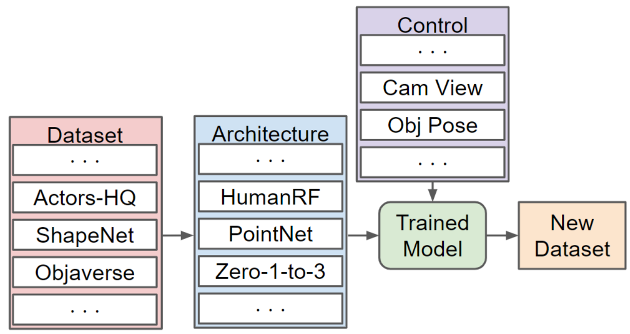

Figure 1. A figure illustrating the mission of the work.

2.2. Generative AI Models

Numerous deep generative models, including Generative Adversarial Networks (GAN) and Variational Autoencoders (VAE), have been developed for image generation [3, 4]. These models excel in producing high-quality images and offer significant advantages in semantic image manipulation, allowing control over the generation process through conditioning on additional information [3,4]. However, despite their effectiveness in image inversion, generation, and editing, NeRF models demonstrate superior accuracy and efficiency in generating images of specific scenes or objects [1].

Traditional deep generative techniques primarily generate 2D images from a single viewpoint, whereas NeRF can construct detailed 3D scenes. This capability enables NeRF to provide a richer representation of a scene's appearance. NeRF is tailored to 'overfit' specific input images of a scene, allowing it to capture even the most intricate details and complex structures. Additionally, NeRF requires a smaller dataset for training, leading to faster and more adaptable processes. These qualities distinguish NeRF as a particularly powerful tool in the realm of generative technology.

2.3. HumanRF

Building on NeRF's principles, HumanRF introduces a specialized approach for representing human actions with high fidelity. This 4D dynamic neural scene representation is trained using multi-view video inputs to capture the full-body motions and appearances of humans. It utilizes ActorsHQ, a dataset comprising high-resolution videos of 8 actors performing various movements [2]. HumanRF can generate photorealistic views from novel points and poses of human actors, making it valuable in industries like computer games and film.

One of HumanRF's major advantages is its ability to render images from any specified camera position, allowing users to generate images from any location and angle, which facilitates the creation of a comprehensive dataset. However, it does not provide a visual camera viewer or a guide for orienting the camera to match the original setup used in creating the ActorsHQ dataset. This limitation requires users to experiment with various camera settings to achieve the desired images. Additionally, its automated rendering processes are limited by reading from a calibration file and interpolating between input key cameras.

2.4. nerfstudio

Nerfstudio is a comprehensive and extensible framework designed to accelerate both the research and application of NeRFs [6]. Its standout feature is a modular design that supports real-time visualization tools and streamlines the import of real-world data. Additionally, Nerfstudio offers diverse export options, including video, point cloud, and mesh formats, providing a robust platform for developing custom NeRF methods and processing real-world data.

A notable feature of NerfStudio is its detailed documentation and active community channels, which significantly simplify NeRF research and enhance user-friendliness by providing remarkable support. However, NerfStudio has limitations, such as its inability to render from a single camera, restricting automated rendering processes. Like HumanRF, it also lacks diverse automated rendering options, such as circular path rendering from an initial camera position. Positively, NerfStudio's viewer is a powerful tool that allows users to precisely set camera locations and angles, improving control over rendering. It also enables users to define any trajectory using the viewer, offering greater flexibility in rendering and analysis. It's important to acknowledge that the platform is continuously evolving, and these observations reflect the current state of NerfStudio as of the date this report is authored.

3. NeRFlect

NeRFlect is a pipeline designed to enhance the capabilities of NeRF models as generative tools. It enables users to generate new images of scenes in specific ways. Users can choose to orbit circularly around an object from a fixed camera or uniformly sample new camera positions on a spherical surface, which allows for the formation of images from various perspectives of the object in 3D space.

3.1. Orbit Sampling

Orbit Sampling in NeRFlect involves positioning the camera at various points along a circular trajectory around the object, thus capturing images from multiple angles. This method is beneficial in a dataset creation pipeline as it systematically covers a broad range of perspectives around the object starting from a specified location, enriching the dataset with diverse viewpoints. Sample images generated using Orbit Sampling can be found in Fig. 2."

The process of orbit sampling can be mathematically represented as follows:

$$\theta_i = i \cdot \Delta\theta \tag{1}$$

$$(x_i, y_i, z_i) = (x_t, y_t, z_t) + \text{radius} \cdot [\cos(\theta_i), 0, \sin(\theta_i)] \tag{2}$$

where:

- $\Delta\theta = \frac{2\pi}{\text{num\_samples}}$ is the fixed angle denoting the rotation angle per sample, where num_samples defines the number of samples desired.
- θ_i is the angle of rotation around the object for the i^{th} camera, calculated by multiplying i by $\Delta\theta$, with i varying from 0 to num_samples-1.
- (x_i, y_i, z_i) is the 3D position of the camera along the circular orbit for the i^{th} sampled camera.
- (x_t, y_t, z_t) is the 3D position of the target point. It denotes where the camera is oriented towards.
- radius is the distance from the target position to the i^{th} camera's position, which remains constant along the orbit.

The rotation matrix for orienting the camera can be derived using the following expression:

$$R = \begin{bmatrix} \text{right}_x & \text{right}_y & \text{right}_z \\ \text{actual\_up}_x & \text{actual\_up}_y & \text{actual\_up}_z \\ \text{forward}_x & \text{forward}_y & \text{forward}_z \end{bmatrix} \tag{3}$$

where:

$$\text{forward} = \frac{(x_t, y_t, z_t) - (x_i, y_i, z_i)}{\|(x_t, y_t, z_t) - (x_i, y_i, z_i)\|},$$
$$\text{right} = \frac{\text{up} \times \text{forward}}{\|\text{up} \times \text{forward}\|},$$
$$\text{actual\_up} = \text{forward} \times \text{right}.$$

Here, up is a predefined global up direction and can change between models that implement different axes conventions. Also, it is essential to note that all the equations in this report are designed assuming a Right-Down-Forward (RDF) axes convention. The cross-product operation $\times$ denotes the vector cross-product.

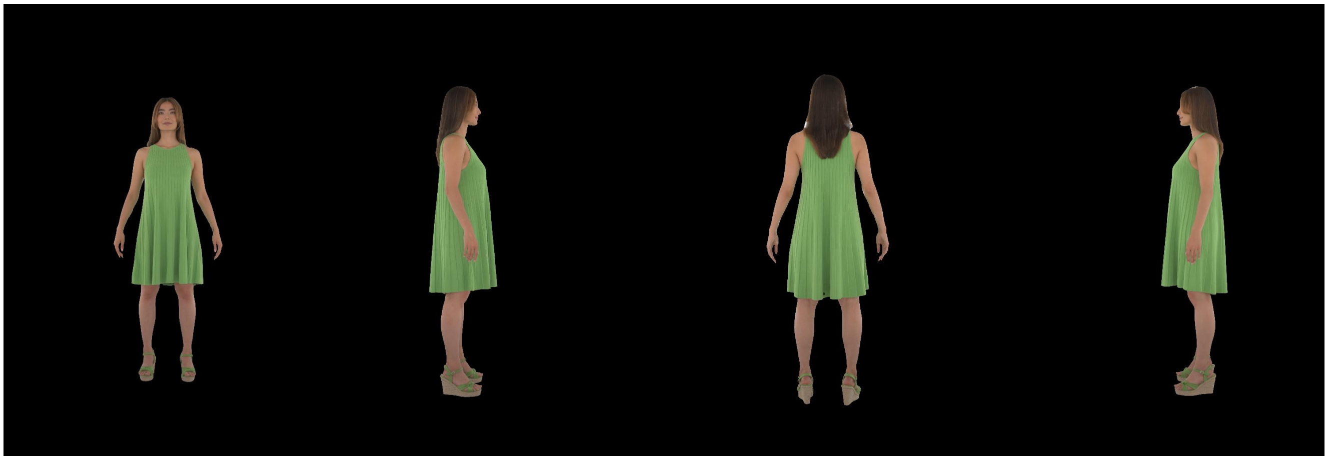

Figure 2. Example images of a human actor generated by orbit sampling method.

3.2. Uniform Sampling

Uniform Sampling involves uniformly sampling camera positions on the surface of a sphere surrounding the object of interest. This technique is instrumental as it ensures a thorough and uniform coverage of all possible angles and perspectives of the object, which is vital for creating a rich and diverse dataset. A set of sample images demonstrating the use of uniform sampling can be found in Fig. 3. The process of sphere surface sampling can be mathematically represented as follows:

$$\phi_j = j \cdot \Delta\phi, \quad \theta_k = k \cdot \Delta\theta \tag{4}$$

where:

- $\Delta\phi = \frac{\pi}{\sqrt{\text{num\_samples}}}$ and $\Delta\theta = \frac{2\pi}{\sqrt{\text{num\_samples}}}$ are the fixed angles denoting the rotation angles per sample in polar and azimuthal directions, respectively.
- $i = j \cdot \sqrt{\text{num\_samples}} + k$, with k and j varying across 0 to $\sqrt{\text{num\_samples}}$.
- ϕ_j is the polar angle for the i^{th} camera, calculated by multiplying j by $\Delta\phi$.
- θ_k is the azimuthal angle for the i^{th} camera, calculated by multiplying k by $\Delta\theta$.

The Cartesian coordinates of the newly sampled cameras are then computed as follows:

$$x = x_t + r \cdot \sin(\phi_j) \cdot \cos(\theta_k) \tag{5}$$

$$y = y_t + r \cdot \sin(\phi_j) \cdot \sin(\theta_k) \tag{6}$$

$$z = z_t + r \cdot \cos(\phi_j) \tag{7}$$

where:

- (x_t, y_t, z_t) is the 3D position of the target, generally specified as object center.
- r is the radius of the sphere on which the cameras are being sampled.

The rotation matrix for orienting the camera towards the target point (object center) is computed using the same way described in Eq. (3).

3.3. Additional Improvements

While developing the NeRFlect pipeline, the sampling methods were integrated and enhanced based on the Nerf-Studio and HumanRF models. Additionally, the pipeline incorporates several other improvements: the capability to manually specify a camera in HumanRF, enhanced single-camera rendering in NerfStudio, the option to fix a single frame for rendering in both, and the ability to adjust the target point and radius for spheres or circles in both models. Images generated by specifying the radius and target point can be found in Fig. 4 and Fig. 5, respectively.

4. Conclusion

This research focuses on the capabilities of Neural Radiance Fields (NeRF) models, particularly in creating novel, high-quality image datasets, to address existing gaps in this domain. The study introduces a new tool, NeRFlect, designed to enhance the generative capacity of NeRF models. NeRFlect offers structured methods to produce images from various angles and locations, utilizing two main sampling techniques: Orbit Sampling and Uniform Sampling. These techniques ensure systematic coverage of a wide range of views for a scene or object.

By integrating these sampling methods with existing frameworks like NerfStudio and HumanRF, NeRFlect aims to broaden the scope of image production. This is achieved

Figure 3. Example images of a human actor generated by uniform sampling method.

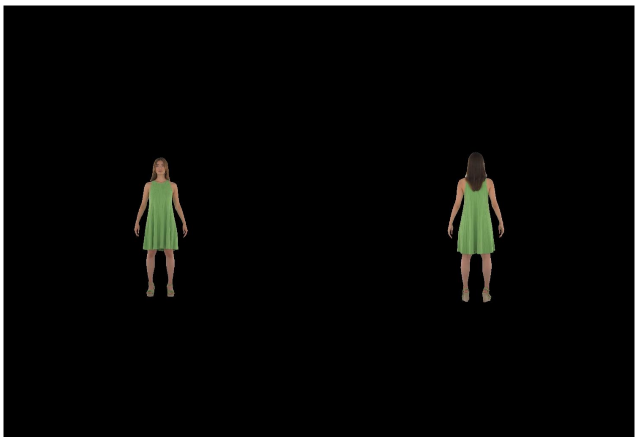

Figure 4. A pair of images illustrating the ability to specify the radius.

Figure 5. A pair of images illustrating the ability to specify the target point.

through a blend of innovative sampling techniques and enhancements to existing NeRF-based models, thereby addressing the data bottleneck in 3D scene representation and rendering.

In future of this research work, new sampling techniques and additional NeRF models can be investigated to boost the generative capability of the NeRFlect.

References

[1] Yifan Ge and et al. Neural-sim: Learning to generate training data with nerf. In *Lecture Notes in Computer Science*, pages 477–493, 2022. 1

[2] Mehmet Işık and et al. Humanrf: High-fidelity neural radiance fields for humans in motion. *ACM Transactions on Graphics*, 42(4):1–12, 2023. 2

[3] Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

[4] Diederik P Kingma and Max Welling. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019. 1

[5] Ben Mildenhall and et al. Nerf: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision – ECCV 2020*, pages 405–421, 2020. 1

[6] Matthew Tancik and et al. Nerfstudio: A modular framework for neural radiance field development. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Proceedings*, 2023. 2