Project Report

SHERF: Generalizable Human NeRF from a Single Image

Sukhdev Kolli
Nandini Chalicham
Sindhu Malisetty
Tanvi Jonnada

Background:
Title: "SHERF: Generalizable Human NERF from a Single Image"

Abstract:
The paper "SHERF: NERF from a Single Image" presents a novel approach to addressing the challenging problem of Neural Radiance Fields (NERF) reconstruction from a single image. The authors introduce the SHERF (Single-view Human Eye-like Radiance Fields) framework, which leverages human eye-inspired priors and attention mechanisms to enhance the reconstruction quality of NERF models. This review paper provides an in-depth analysis of the key contributions, methodologies, strengths, and potential limitations of the SHERF framework.

Overview:
SHERF - Generalizable Human NERF from a Single Image" is a research project that focuses on Neural Radiance Fields (NERF) and aims to create a more generalizable and robust system for synthesizing 3D reconstructions of human subjects from a single 2D image. NERF is a technology used in computer graphics and computer vision to generate 3D models of objects or scenes from 2D images or videos.

This research extends the capabilities of NERF to handle human subjects by addressing some of its limitations. One of the primary challenges is that NERF models tend to overfit and require a large amount of training data. "SHERF" stands for "Single-Human Extrapolatable Radiance Fields," and it focuses on the following key aspects:

1. **Generalization:** The system is designed to work effectively with various poses and clothing, improving its generalizability.

2. **Single-Image Reconstruction:** The goal is to generate 3D reconstructions of human subjects from a single 2D image, making it more practical for real-world applications.

3. **Realism:** The research aims to create more realistic and accurate 3D reconstructions of human subjects.

1. Introduction:
The project begins by highlighting the significance of single-image NERF reconstruction in computer vision and graphics applications, emphasizing the limitations of existing methods. It sets the stage for the introduction of SHERF as a solution that draws inspiration from the human visual system to improve the fidelity of scene reconstructions.

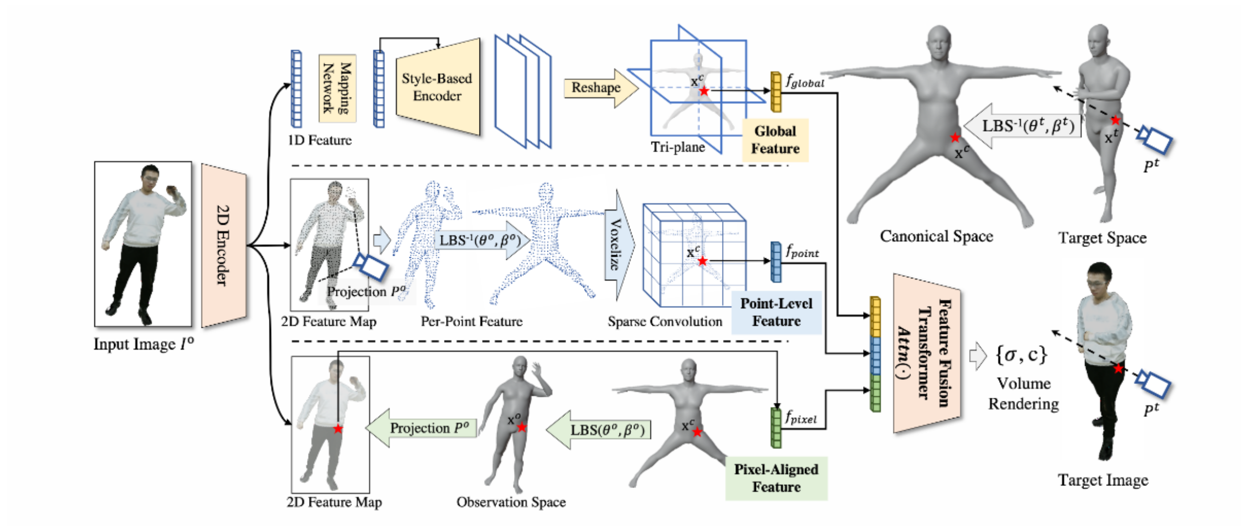
2. Key Contributions:
SHERF introduces several key contributions to the field of NERF reconstruction:

a. Human Eye-inspired Priors: The authors incorporate insights from human vision to guide the reconstruction process. By integrating attention mechanisms and perceptual priors, SHERF aims to generate more visually plausible reconstructions.

b. Adaptive Sampling: The paper introduces an adaptive sampling strategy that optimizes the efficiency of the reconstruction process. This is achieved by selectively sampling views based on the relevance to the scene content, reducing computational overhead.

c. Learning Scene Semantics: SHERF focuses on learning semantic information about the scene to improve the accuracy of the reconstructed radiance fields. This involves training the model to recognize and incorporate semantic cues from the input image.

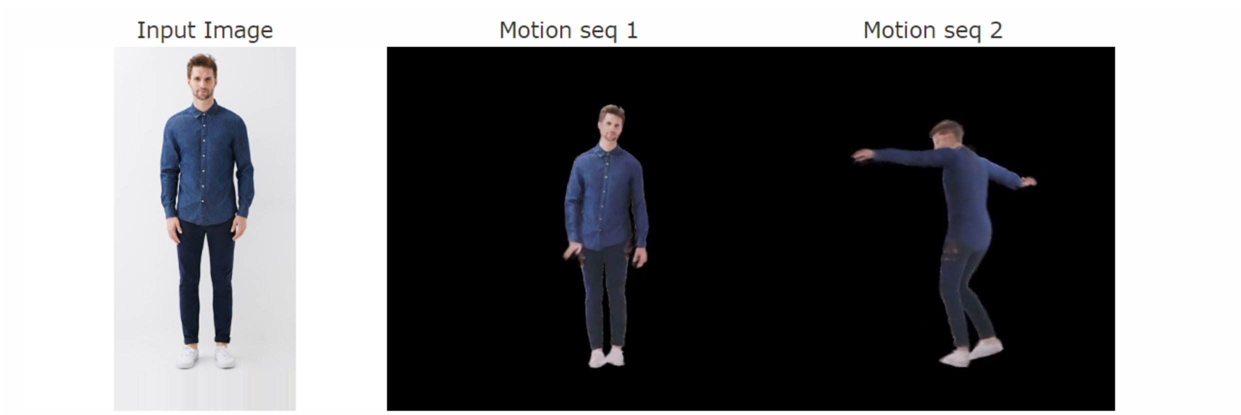
3. Methodology:
The paper provides a detailed overview of the SHERF framework, elucidating the mathematical formulations, training procedures, and implementation details. The incorporation of attention mechanisms, adaptive sampling, and semantic learning is thoroughly explained, allowing readers to grasp the technical intricacies of the proposed method.



To render the target image, we first cast rays and sample points in the target space. The sample points are transformed to the canonical space through inverse LBS. We then query the corresponding 3D-aware global, point-level, and pixel-aligned features. The deformed points, combined with the bank of features, are input into the feature fusion transformer and NeRF decoder to get the RGB and density, which are further used to produce the target image through volume rendering.

4. Experimental Evaluation:

We conducted comprehensive experiments to validate the efficacy of the SHERF framework. Results demonstrate improvements in reconstruction quality, especially in challenging scenarios with complex lighting and occlusions. The paper includes comparisons with state-of-the-art NERF methods, showcasing the advantages of SHERF in terms of visual realism and accuracy.



We researched and collected the dataset required for the project and trained the model accordingly. In addition to that we studied the SMPL models which are helpful for the project. The SMPL (Skinned Multi-Person Linear) model is a widely used 3D model for representing

human bodies and poses. SMPL models are essential tools in computer graphics, computer vision, and machine learning applications, particularly in areas like human pose estimation, 3D reconstruction, and virtual character animation. In addition to the base SMPL model, various extensions and pretrained models have been developed to improve the accuracy and versatility of human body representation and understanding.
Exploring NERF, Neural Radiance Fields (NeRF) is a powerful technique used in 3D scene reconstruction and rendering. In NeRF, a neural network learns to model a 3D scene from a set of 2D images taken from different viewpoints. This model can then be used to render novel views of the scene. If you're looking for a PyTorch implementation of NeRF, you can refer to various open-source repositories and libraries available.

5. Strengths and Limitations:
This section critically evaluates the strengths and potential limitations of the SHERF framework. While the human eye-inspired priors contribute to realistic reconstructions, questions may arise regarding the generalization of such priors across diverse scenes. Additionally, computational efficiency gains are reported, but the scalability of the proposed method to large-scale scenes needs further exploration.

6. Future Directions:
The paper concludes by suggesting potential avenues for future research. Areas of improvement, such as scalability, robustness to diverse scenes, and integration with real-world applications, are identified. The authors emphasize the need for continued exploration and refinement of SHERF to address these challenges.

7. Conclusion:
In conclusion, "SHERF: NERF from a Single Image" presents a promising approach to single-image NERF reconstruction, leveraging human vision-inspired priors and attention mechanisms. The paper contributes to the ongoing efforts to enhance the realism and efficiency of neural radiance field models. While the proposed method demonstrates notable improvements, further research is required to address potential limitations and explore its applicability in diverse real-world scenarios.

Project Implementation:
Taking inspiration from the original authors and modifications made to the original code we have made the final GitHub repo of the final project.

Github link :

We have run the file in Google Colab and as well as the Visual Studio to run the project. However, the project requires high-level GPUs to run the project smoothly. We have downloaded the datasets provided in the original GitHub repo such as the HuMMan dataset, and THuman Dataset, and used the inference codes which is useful to learn and analyze large dataset population. The training codes for respective datasets are also given in the GitHub repository

Requirements
NVIDIA GPUs are required for this project.

Datasets
RenderPeople Dataset
Please download the rendered multi-view images of RenderPeople dataset created by original authors from (https://mycuhk-my.sharepoint.com/:f:/g/personal/1155098117_link_cuhk_edu_hk/ElL9IDDOaa5Hl785gvbqyEEB8ubdobyuMKqoDY3J85XStw?e=o2BUOt).

THuman Dataset
Please follow the instructions of MPS-NeRF (https://github.com/gaoxiangjun/MPS-NeRF) to download the THuman dataset (http://www.liuyebin.com/deephuman/deephuman.html) dataset. After that, please download the estimated SMPL Neutral parameters.

HuMMan Dataset
Please follow the instructions (https://caizhongang.github.io/projects/HuMMan/) to download the HuMMan dataset.

ZJU-Mocap dataset
Please follow the instructions of Neural Body to download the ZJU-Mocap dataset.

Download Models
The pre-trained models and SMPL model are needed for inference.

The pre-trained models are put in (https://mycuhk-my.sharepoint.com/:u:/g/personal/1155098117_link_cuhk_edu_hk/EU3RxpLuKmZImkdJbG8Y12EBZ9RxIfQiEx7ctt5obXUjzw?e=gXJbIQ) and Baidu Pan (pin:gu1q) for downloading.
Inference code with RenderPeople dataset
bash eval_renderpeople_512x512.sh
Inference code with THuman dataset
bash eval_THuman_512x512.sh
Inference code with HuMMan dataset

bash eval_HuMMan_640x360.sh
Inference code with ZJU-Mocap dataset
bash eval_zju_mocap_512x512.sh

Citation: The original project which is taken inspiration from original authors Shoukang Hu1Fangzhou Hong1Liang Pan1Haiyi Mei2Lei Yang2Ziwei Liu✉1, 1S-Lab, Nanyang Technological University 2Sensetime Research *Equal Contribution✉Corresponding Author publish in ICCV 2023.

This project is built on source codes shared by EG3D, MPS-NeRF, and Neural Body.

Challenges Faced :

- We have faced several challenges like not having sophisticated hardware to run the project but overcame them by using the computers provided and anaconda in the Python environments.
- We also had many errors while running the projects such as not having all the modules installed.
- We also stumbled upon various indentation errors that needed to be fixed and we fixed those errors as per our best knowledge. We tried our level best with the available resources to run the project.
- Optimizing the model and algorithms to meet real-time constraints was also one of the challenges we faced while implementing the project.
- Implementing SHERF (Single-Human Elevation Representations from a Single Image) or similar neural rendering techniques can be computationally intensive, and it often requires powerful GPUs for efficient training and inference. NVIDIA GPUs are the preferred choice for deep learning tasks due to their excellent support for CUDA. This would be a challenging aspect while implementing the code. In addition to that we faced Errors while training the model. Incompatibility between different libraries or versions also resulted in errors. Gradual memory consumption over time resulted in crashing.

Future Milestones :

We are working on the model and rectifying the errors. Training the datasets gives accurate results. In addition to that we will compare and evaluate the current model with the existing models so that we understand the model and know the proper differences between the xixting models. Also in the future we will complete the implementation of total project and running the code without any errors.