HFNeRF: Learning Human Biomechanic Features with Neural Radiance Fields

Arnab Dey*, Di Yang†, Antitza Dantcheva† and Jean Martinet*
*I3S-CNRS/Université Côte d'Azur †INRIA Center at Université Côte d'Azur

***Abstract*—In recent advancements in novel view synthesis, generalizable Neural Radiance Fields (NeRF) based methods applied to human subjects have shown remarkable results in generating novel views from few images. However, this generalization ability cannot capture the underlying structural features of the skeleton shared across all instances. Building upon this, we introduce HFNeRF: a novel human feature NeRF aimed at generating human biomechanic features using a pre-trained image encoder. While previous human NeRF methods have shown promising results in the generation of photorealistic virtual avatars, such methods lack underlying human structure or biomechanic features such as skeleton or joint information that are crucial for downstream applications including Augmented Reality (AR)/Virtual Reality (VR). HFNeRF leverages 2D pre-trained foundation models toward learning human features in 3D using neural rendering, and then volume rendering towards generating 2D feature maps. We evaluate HFNeRF in the skeleton estimation task by predicting heatmaps as features. The proposed method is fully differentiable, allowing to successfully learn color, geometry, and human skeleton in a simultaneous manner. This paper presents preliminary results of HFNeRF, illustrating its potential in generating realistic virtual avatars with biomechanic features using NeRF.**

***Index Terms*—Computer Vision, Augmented Reality, Virtual Reality, Neural Radiance Fields**

I. Introduction

The development of custom virtual avatars capable of achieving photorealism is essential for realistic AR/VR environments. Moreover, it is a significant challenge to create a photorealistic virtual human avatar from a sparse set of images captured by a smartphone or a single camera. Previously, creating personalized virtual avatars with an underlying structure, such as a skeleton, required the use of costly camera setups that were only within the reach of a limited group of people. Furthermore, the labor-intensive process of body marker capture, extraction, and fitting of parametric models, such as SMPL [5], is not scalable for widespread use.

The recent progress in Neural Radiance Fields has demonstrated significant potential in creating highly realistic virtual avatars using few images [4], [6]. However, previous NeRF-based methods do not provide any underlying structure, which is crucial for AR/VR applications and animation. We introduce a novel approach named HFNeRF: Learning Human Biomechanic Features with Neural Radiance Fields, a unified framework to learn human biomechanic features such as the human skeleton with NeRF. Inspired by previous NeRF-based methods [7], [8] that utilize 2D encoders to generalize NeRF by conditioning input images or learning scene features. Our method uses a 2D pre-trained encoder to learn human features using NeRF architecture. Specifically, HFNeRF predicts heatmap features of human joints, aiding in skeleton detection. Our method adopts two different types of encoder to generate features from images. HFNeRF estimate separate heatmaps corresponding to each joint along with color and volume density. Our NeRF model takes as input the image feature of the 3D query point x, along with its frequency encoding and view direction. The final heatmaps are generated using volume rendering inspired by the pixel color generation process of NeRF.

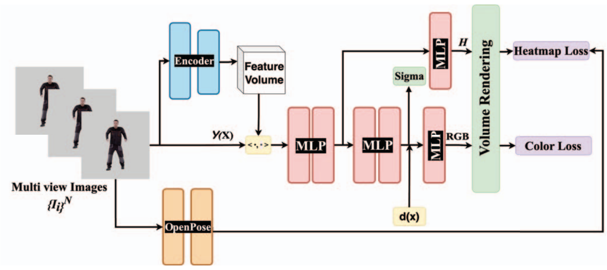

Fig. 1. Proposed pipeline of HFNeRF. During training, we generate ground truth heatmaps using OpenPose and compare these with the heatmaps predicted by our network to compute the heatmap loss.

In this paper, we present the initial results of HFNeRF obtained with the RenderPeople dataset by distilling a state-of-the-art pose estimation algorithm based on heatmaps. To the best of our knowledge, our method is the first to estimate human biomechanic features with NeRF. Our contributions are as follows:

- We present a new method for estimating human biomechanic features using NeRF.
- We show that our model successfully learns to predict skeleton information from 2D images.

II. Method

This section introduces HFNeRF, a unified framework utilizing the NeRF architecture for learning human features. It begins with a brief overview of NeRF, followed by the methodology for feature extraction using a 2D encoder, and concludes with a detailed description of skeleton estimation.

Neural Radiance Fields: The NeRF algorithm utilizes a multilayer perceptron (MLP) to map the 3D coordinate $\mathrm{x} = (x, y, z)$ and view direction $\mathbf{d} = (\theta, \phi)$ to the corresponding color c and volume density σ. This mapping can be expressed

Contact email: adey@i3s.unice.fr

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00131

as $F(\mathrm{x}, \mathbf{d}) \rightarrow (c, \sigma)$. Subsequently, volume rendering is employed to generate the pixel color by performing alpha composition of the volume density σ and color c using samples taken along the ray.

Feature Extraction: We propose a novel architecture to estimate human features using the NeRF framework. NeRF models estimate the color c and density σ using an MLP, where the input is the positional encoding $\gamma(\mathrm{x})$ of the query point $\mathrm{x} = (x, y, z)$. The human feature $f(\mathrm{x})$, generated by the encoder corresponding to the input point x, is concatenated with the positional encoding $\gamma(\mathrm{x})$ before being fed into the MLP. In this work, we experimented with 2 different kinds of encoders, namely: ResNet [3] and DINO [2].

Learning Human Biomechanics features Previous methods [8] used encoded features to generalize NeRF, producing color and density as output. In this work, we extend the NeRF architecture to estimate human features by generating heatmaps of skeleton joints, as shown in Figure 1. We used an MLP with a skip connection, where the view direction is incorporated into the final layer before producing the color output. To generate heatmaps, we extract NeRF features from an intermediate layer and process them through a smaller secondary MLP. We employ volume rendering to produce the final pixel color and heatmap values. This method is fully differentiable and optimized using a combined loss function: $L = l_c + \lambda_h l_h$, where λ_h is the weighting factor and l_h represents mean squared error between predicted and ground truth heatmap.

Skeleton prediction: We estimate the human skeleton by predicting joint locations from heatmaps. For each heatmap channel, which corresponds to a specific joint, a binary mask is generated through Gaussian filtering and thresholding. The joint locations are then identified as the pixels with the peak heatmap values within these mask regions.

III. Experiments and Discussions

In this section, we present our preliminary results, focusing on skeleton detection and novel view synthesis.

Dataset. We trained our model on the RenderPeople [4] dataset, consists of multi-view image sequences of animated characters performing various actions. We use 34 cameras for training and 2 cameras for testing.

Experimental setup. All experiments were conducted using a PyTorch implementation on an RTX 3090 GPU. For our experiments, the value of λ_h was set to 0.5. We used the Adam optimizer for 100,000 iterations. We learn the heatmap features by distilling OpenPose [1].

Results. This section details the initial results obtained using our HFNeRF method. Quantitative results from the RenderPeople dataset are summarized in Table I. The predicted and ground truth heatmaps are compared with the Mean Squared Error (MSE). The results indicate that ResNet features improve visual quality, while Vision Transformer-based DINO features lead to better heatmap predictions. Figure 2 visually demonstrates these findings. In the future, we intend to expand our experimentation to encompass various datasets and perform additional comparisons with other methods.

Dataset	PSNR↑	SSIM↑	LPIPS↓	MSE↓
RenderPeople+ResNet	46.421	0.9996	0.0024	0.0003
RenderPeople+DINO	35.928	0.9914	0.0345	0.0001

TABLE I
Quantitative results on RenderPeople dataset.

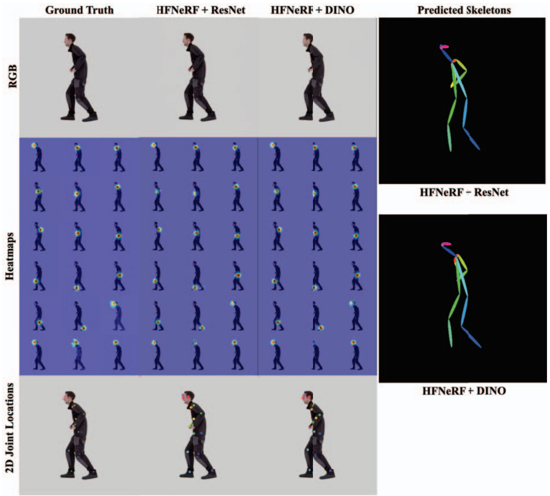

Fig. 2. Qualitative comparison on RenderPeople dataset.

IV. Conclusion

This paper presents a novel framework called HFNeRF, which uses NeRF to learn human biomechanic features. Our initial findings demonstrate the effectiveness of HFNeRF in predicting human features, a significant improvement over previous NeRF methods for humans. Although our focus was on human skeleton detection, we believe that this architecture can be extended to other generalizable human features, such as body part detection.

V. Acknowledgements

This project has received funding from the H2020 COFUND program BoostUrCareer under MSCA no.847581.

References

[1] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE TPAMI*, pages 172–186, 2021.

[2] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin. Emerging properties in self-supervised vision transformers. In *ICCV*, 2021.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778, 2016.

[4] S. Hu, F. Hong, L. Pan, H. Mei, L. Yang, and Z. Liu. Sherf: Generalizable human nerf from a single image. *arXiv:2303.12791*, 2023.

[5] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 851–866. 2023.

[6] S.-Y. Su, F. Yu, M. Zollhöfer, and H. Rhodin. A-nerf: Articulated neural radiance fields for learning human shape, appearance, and pose. *NeurIPS*, 34:12278–12291, 2021.

[7] J. Ye, N. Wang, and X. Wang. Featurenerf: Learning generalizable nerfs by distilling foundation models. *arXiv:2303.12786*, 2023.

[8] A. Yu, V. Ye, M. Tancik, and A. Kanazawa. pixelnerf: Neural radiance fields from one or few images. In *CVPR*, pages 4578–4587, 2021.