**NVIDIA 2024**

1. NeRF领域（文章说以下方法都不能泛化unseen identities during inference和time-consuming optimization，但是D-NeRF可以重演啊？耗时优化指的是什么？）

**1)**Many works [2; 3; 4; 19; 20; 21; 23; 31; 42; 43; 54; 55; 72; 73; ] attempt to apply NeRF to human portrait reconstruction and animation by extending it from static scenes to dynamic portrait videos. Although these methods demonstrate realistic reconstruction results, they inefficiently learn separate networks for different identities and require thousands of frames from a specific individual for training.

3, 31, 73

[2] ShahRukh Athar, Zexiang Xu, Kalyan Sunkavalli, Eli Shechtman, and Zhixin Shu. **Rignerf**: Fully controllable neural 3d portraits. In CVPR, **2022**. 2, 3

[4] Ziqian Bai, Feitong Tan, Zeng Huang, Kripasindhu Sarkar, Danhang Tang, Di Qiu, Abhimitra Meka, Ruofei Du, Mingsong Dou, Sergio Orts-Escolano, et al. Learning personalized high quality volumetric head avatars from monocular rgb videos. arXiv preprint arXiv:2304.01436, 2023. 2, 3

**[19]** Guy Gafni, Justus Thies, Michael Zollhöfer, and Matthias Nießner. **Dynamic** neural radiance fields for monocular 4d facial avatar reconstruction. In CVPR, 2021.

[20] Chen Gao, Yichang Shih, Wei-Sheng Lai, Chia-Kai Liang, and Jia-Bin Huang. Portrait neural radiance fields from a single image. arXiv preprint arXiv:2012.05903, 2020. 2, 3

[21] Xuan Gao, Chenglai Zhong, Jun Xiang, Yang Hong, Yudong Guo, and Juyong Zhang. Reconstructing personalized semantic facial nerf models from monocular video. SIGGRAPH Asia, 2022. 2, 3

[23] Yudong Guo, Keyu Chen, Sen Liang, Yong-Jin Liu, Hujun Bao, and Juyong Zhang. **Ad-nerf**: Audio driven neural radiance fields for talking head synthesis. In ICCV, **2021**. 2, 3

[42] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz,and Ricardo Martin-Brualla. **Nerfies**: Deformable neural radiance fields. ICCV, **2021**. 2, 3

[43] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M. Seitz. **Hypernerf**: A higher-dimensional representation for topologically varying neural radiance fields. ACMTrans. Graph., 2021. 2, 3

[54] Kartik Teotia, Mallikarjun B R, Xingang Pan, Hyeongwoo Kim, Pablo Garrido, Mohamed Elgharib, and Christian Theobalt. **Hq3davatar**: High quality controllable 3d head avatar. 2023. 2, 3

[55] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. **Non-rigid neural radiance fields**: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In ICCV, 2021. 2, 3

[72] Yufeng Zheng, Victoria Fernández Abrevaya, Marcel C. Bühler, Xu Chen, Michael J. Black, and Otmar Hilliges. **I M Avatar**: Implicit morphable head avatars from videos. In CVPR, 2022. 2, 3

**2)** Another line of works focus on generating a controllable 3D head avatar from **random noise** [33; 40; 50; 51; 53; 61; 62; 75].

[33] Yeonkyeong Lee, Taeho Choi, Hyunsung Go, Hyunjoon Lee, Sunghyun Cho, and Junho Kim. **Exp-gan**: 3d-aware facial image generation with expression control. In ACCV, **2022**. 3, 6

[40] Maryam Sadat Mirzaei, Kourosh Meshgi, Etienne Frigo, and Toyoaki Nishida. **Animgan:** A spatiotemporally-conditioned generative adversarial network for character animation. In ICIP, **2020**. 3

[50] Jingxiang Sun, Xuan Wang, Lizhen Wang, Xiaoyu Li, Yong Zhang, Hongwen Zhang, and Yebin Liu. **Next3d**: Generative neural texture rasterization for 3d-aware head avatars. arXiv preprint arXiv:2211.11208, **2022**. 3, 6

[51] Keqiang Sun, Shangzhe Wu, Zhaoyang Huang, Ning Zhang, Quan Wang, and HongSheng Li. Controllable 3d face synthesis with **conditional generative occupancy fields**. arXiv preprint arXiv:2206.08361, **2022**. 3, 6

[53] Junshu Tang, Bo Zhang, Binxin Yang, Ting Zhang, Dong Chen, Lizhuang Ma, and Fang Wen. Explicitly controllable 3d-aware portrait generation. arXiv preprint arXiv:2209.05434, **2022**. 3

[61] Eric Zhongcong Xu, Jianfeng Zhang, Junhao Liew, Wenqing Zhang, Song Bai, Jiashi Feng, and Mike Zheng Shou. **Pv3d:** A 3d generative model for portrait video generation. In ICLR, **2023**. 3

[62] Hongyi Xu, Guoxian Song, Zihang Jiang, Jianfeng Zhang, Yichun Shi, Jing Liu, Wanchun Ma, Jiashi Feng, and Linjie Luo. **Omniavatar**: Geometry-guided controllable 3d head synthesis. arXiv preprint arXiv:2303.15539, **2023.** 3

[75] Peiye Zhuang, Liqian Ma, Sanmi Koyejo, and Alexander Schwing. Controllable radiance fields for dynamic face synthesis. In 3DV, **2022**. 3

**3)** Intuitively, 3D face reconstruction and animation could be achieved by combining these generative methods with **GAN inversion** [18; 47; 60; 64]. However, the individual optimization process for each frame during GAN inversion is computationally infeasible for real-time performance in applications such as video conferencing.

[18] Anna Frühstück, Nikolaos Sarafianos, Yuanlu Xu, Peter Wonka, and Tony Tung. **VIVE3D**: Viewpointindependent video editing using 3D-aware GANs. In CVPR, **2023**. 3

[47] Daniel Roich, Ron Mokady, Amit H Bermano, and Daniel Cohen-Or. Pivotal tuning for latent-based editing of real images. ACMTrans. Graph., **2021**. 3

[60] Jiaxin Xie, Hao Ouyang, Jingtan Piao, Chenyang Lei, and Qifeng Chen. High-fidelity **3d gan inversion** by pseudo-multi-view optimization. arXiv preprint arXiv:2211.15662, **2022**. 3

[64] Yiran Xu, Zhixin Shu, Cameron Smith, Jia-Bin Huang, and SeoungWug Oh. In-n-out: Face video inversion and editing with volumetric decomposition. arXiv preprint arXiv:2302.04871, **2023**. 3

**4)** Meanwhile, several works [6; 13; 56; 66;] focus on reconstructing 3D avatars from **arbitrary input images**, but they cannot animate or reenact these avatars.

[6] Ananta R Bhattarai, Matthias Nießner, and Artem Sevastopolsky. Triplanenet: An encoder for eg3d inversion. arXiv preprint arXiv:2303.13497, **2023**. 3

[13] Yu Deng, Baoyuan Wang, and Heung-Yeung Shum. Learning detailed radiance manifolds for high-fidelity and 3d-consistent portrait synthesis from monocular image. arXiv preprint arXiv:2211.13901, **2022**. 3, 4

[55] Edgar Tretschk, Ayush Tewari, Vladislav Golyanik, Michael Zollhöfer, Christoph Lassner, and Christian Theobalt. Non-rigid neural radiance fields: Reconstruction and novel view synthesis of a dynamic scene from monocular video. In ICCV, **2021.** 2, 3

[56] Alex Trevithick, Matthew Chan, Michael Stengel, Eric R Chan, Chao Liu, Zhiding Yu, Sameh Khamis, Manmohan Chandraker, Ravi Ramamoorthi, and Koki Nagano. Real-time radiance fields for single-image portrait view synthesis. arXiv preprint arXiv:2305.02310, **2023**. 3

**ROME** method [29] combines a learnable neural texture with explicit FLAME meshes [36] to reconstruct a 3D head avatar, encompassing areas beyond the face region. However, using meshes as the 3D shape representation prevents the model from producing high-fidelity geometry and appearance details.

Instead of using explicit meshes as 3D representation, the ***HeadNeRF [24] and MofaNeRF*** methods learn **implicit neural networks** that take 3DMM parameters (i.e. identity and expression coefficients or albedo and illumination parameters) as inputs to predict the density and color for each queried 3D point.

Additionally, the **OTAvatar [38]** method proposes to disentangle latent style codes from a pre-trained 3D-aware GAN [9] into separate motion and identity codes, enabling facial animation by exchanging the motion codes.

Nonetheless, all three models [24; 76; 38] require laborious test-time optimization, and struggle to reconstruct photo-realistic texture details of the given portrait image presumably because they encode the appearance using a compact latent vector.

In this paper, we propose the first 3D head neural avatar animation work that not only generalizes to unseen identities without test-time optimization, but also captures intricate details from the given portrait image, surpassing all previous works in quality.

1. 基于3DMM 显性重建mesh几何

numerous methods have been proposed to represent the shape and motion of human faces by 3D Morphable Models (3DMMs) [1; 36; 16; 7; 35].

Building upon 3DMMs, many works have been proposed to reconstruct and animate human faces by estimating the person-specific parameters given a single-view portrait image [14; 17; 11; 34].

In this work, we present a method that effectively exploits the strong prior in 3DMMs while addressing its geometry and texture fidelity limitation by employing neural radiance fields [39; 5; 41].*39是经典NeR，5/41也是神经辐射场的一种。F*

（3）2D GAN

The impressive performance of Generative Adversarial Networks (GANs) [22] spurred another line of head avatar animation methods [57; 65; 67; 48; 46; 70; 15].

本文改进点：

In this paper, we present a framework aiming at a more practical but challenging scenario – given an unseen single-view portrait image, we reconstruct an implicit 3D head avatar that not only captures photo-realistic details within and beyond the face region, but also is readily available for animation without requiring further optimization during inference.