

🔥 FLAME-*in*-NeRF 🔥: NEURAL CONTROL OF RADIANCE FIELDS FOR FREE VIEW FACE ANIMATION

-SUPPLEMENTARY-

Anonymous authors

Paper under double-blind review

1 REANIMATION RESULTS (VIDEO)

In the accompanying videos in supplementary material, we show results of reanimation on four different subjects. The videos with name `Reanimation_Subject_{i}.mp4` contains the video of the i 'th subject reanimated using a self-captured video and a video from the internet. Stills from the reanimated videos are shown in Fig 1 and Fig 2. As can be seen from both Fig 1 and Fig 2, the rendered frames faithfully capture the target expression of the driving frame while retaining the individual characteristics of each subject being reanimated.

The videos `Reanimation_View_Subject_2.avi` and `Reanimation_View_Subject_4.avi` contain videos of subject 2 and subject 4, respectively, reanimated from different views using the same driving frames. From the videos, one can see that expression of the reanimated frames are consistent across views and maintain high fidelity to the driving frames.

Finally, the video `Spatial_Ray_Prior_ConstantView.avi` reanimates the subject from a constant view with and without using the spatial ray prior (Section 3.3) in the paper. As can be seen from the video, not using the spatial ray prior makes the appearance of the background dependent on the expression parameters of the driving frame and significantly hurts the quality of reanimation.

2 QUALITATIVE COMPARISON WITH FIRST ORDER MOTION MODEL

In Fig 5 and Fig 6 we provide qualitative comparisons with First Order Motion Model (Siarohin et al., 2019). We use [their publicly available code](#) to generate the results. As can be seen, FOMM generates significant artefacts on the face when performing expression reanimation, especially for more extreme expressions as can be seen in the second column of Fig 5 and Fig 6. Artefacts also appear on the background of all columns of Fig 5 and columns 2,3 and 4 of Fig 5. Further, since FOMM is an image based method, it cannot perform novel-view-synthesis like FLAME-*in*-NeRF can.

3 EXPERIMENTAL CONFIGURATION

All models were trained on 7 Titan RTX GPUs for 1.5 days. Both the Coarse and Fine NeRF models Mildenhall et al. (2020) used 64 points along the ray. The positional is encoded using 10 frequencies while the view is encoded using 4. The Adam optimizer was used for all experiments with a starting learning rate of $1e-3$ which was decayed to $5e-4$ over 150k epochs. Coarse-to-fine regularization was applied for 50k epochs (i.e $N = 50k$, see Eq. 4 of the paper). The network architecture for the canonical NeRF that gives as output the RGB color and density is shown in Fig 7 and the architecture of the deformation network is shown in Fig 8.

Subject	Method	Epochs Trained	App Code dim	Def Code dim	FRR Coeff
Subject 1	FLAME- <i>in</i> -NeRF	150000	8	128	1e-1
	Nerfies	200000	8	128	1e-1
Subject 2	FLAME- <i>in</i> -NeRF	150000	8	128	10
	Nerfies	150000	8	128	10
Subject 3	FLAME- <i>in</i> -NeRF	150000	8	128	10.0
	Nerfies	150000	8	128	10.0
Subject 4	FLAME- <i>in</i> -NeRF	150000	8	128	1.0
	Nerfies	150000	8	128	1.0

Table 1: Trainig configuration for all the experiments.

REFERENCES

- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. 2020. [1](#), [8](#)
- 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. [9](#)
- Aliaksandr Siarohin, Stéphane Lathuilière, Sergey Tulyakov, Elisa Ricci, and Nicu Sebe. First order motion model for image animation. 2019. [1](#), [7](#), [8](#)



Figure 1: **Reanimation using FLAME-in-NeRF**. Results of reanimating 4 subjects using a self-captured video. The reanimated frames retain high fidelity to the target expression of the driving frame while, while simultaneously, respecting the individual characteristics of each subject. For example, in column 2, the rounding of the mouth is faithfully rendered across all the subjects but is individualistic. Subject 3 (in column 2) has her teeth showing as her mouth is rounded, while the others do not. Similarly, the half-open mouth of the last column is also faithfully rendered across all subjects while retaining individual characteristics.



Figure 2: **Reanimation using FLAME-in-NeRF.** Results of reanimating 4 subjects using a video from the internet. Despite being an in-the-wild video with a wide variety of expressions, the reanimated frames retain high fidelity to the target expression of the driving frame while, while simultaneously, respecting the individual characteristics of each subject. For example, in column 1, the half open mouth is faithfully rendered across all the subjects but is individualistic. Subjects 1 and 3 (in column 1) have their teeth showing prominently while Subjects 2 and 4 do not.

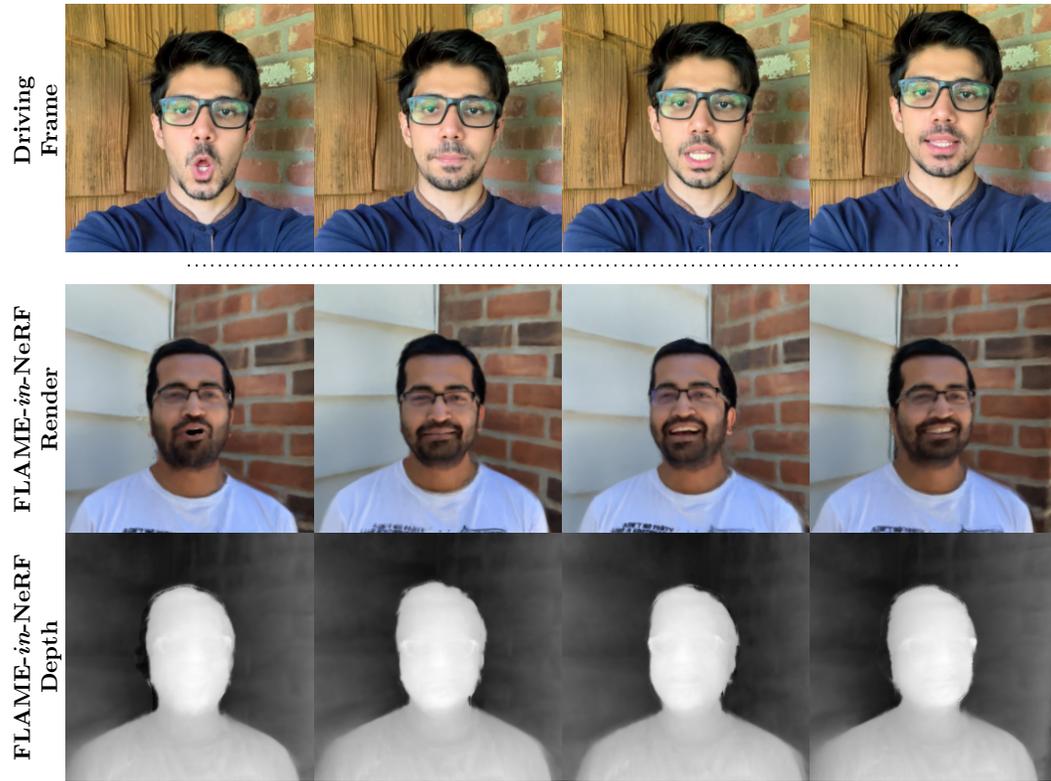


Figure 3: Reanimation using FLAME-in-NeRF.



Figure 4: Reanimation using FLAME-*in*-NeRF.

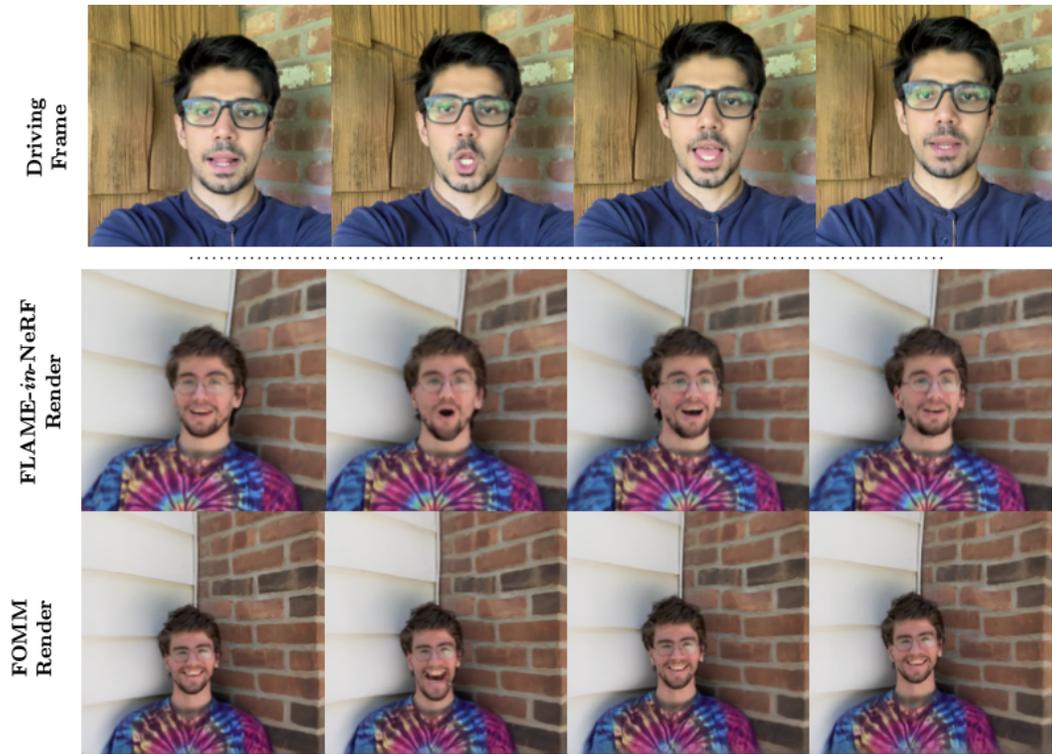


Figure 5: **Comparison with FOMM (Siarohin et al., 2019)**: Here we show a qualitative comparison with FOMM (Siarohin et al., 2019). The first row is the driving frame, the second is the render of FLAME-in-NeRF in different views and the third is the render of FOMM (Siarohin et al., 2019). As can be seen in column 2, there are significant artefacts on the face of the results from FOMM. FOMM also generates artefacts on the background, as can be seen in all columns.

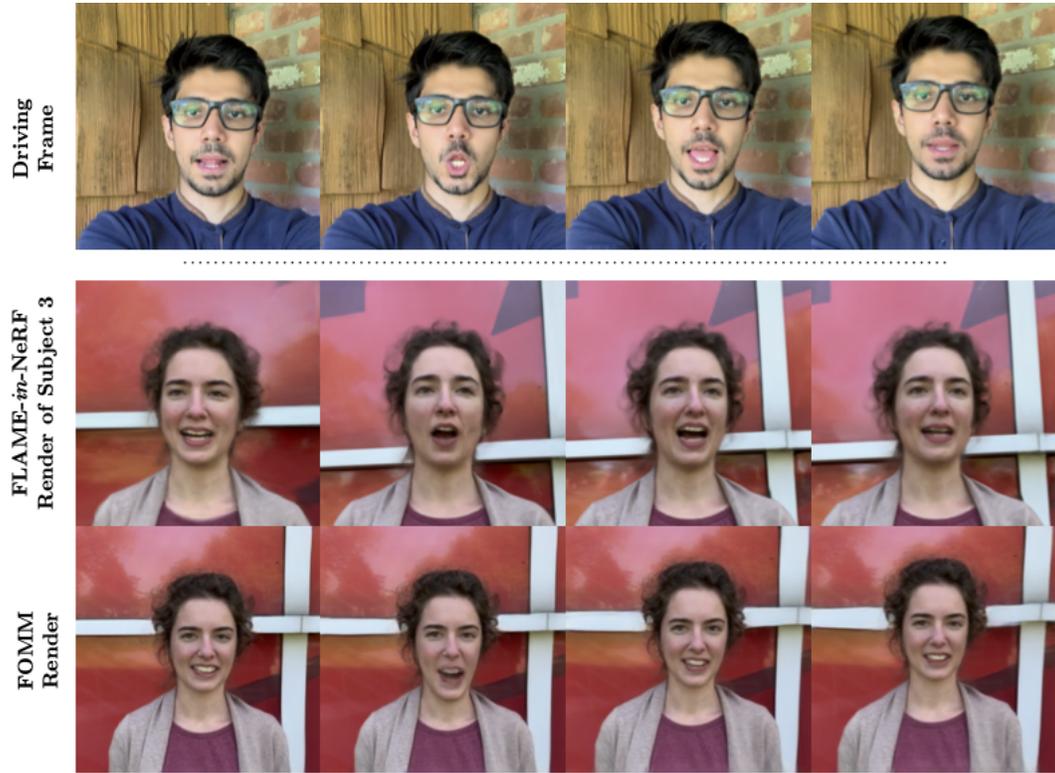


Figure 6: **Comparison with FOMM (Siarohin et al., 2019)**: Here we show a qualitative comparison with FOMM (Siarohin et al., 2019). The first row is the driving frame, the second is the render of FLAME-*in*-NeRF in different views and the third is the render of FOMM (Siarohin et al., 2019). As can be seen in columns 2 and 3, there are significant artefacts on the face of the results from FOMM. FOMM also generates artefacts on the background, as can be seen in columns 2,3 and 4.

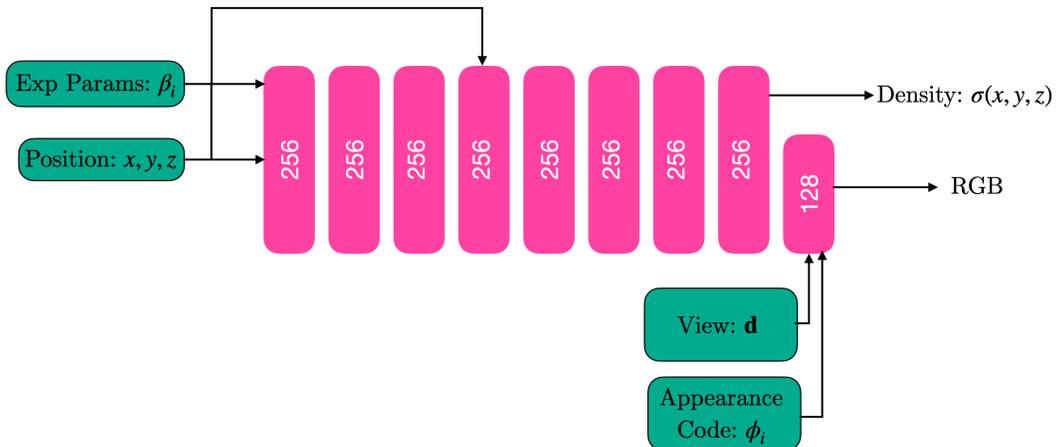


Figure 7: **Canonical NeRF architecture used in FLAME-*in*-NeRF**. FLAME-*in*-NeRF uses the canonical NeRF architecture Mildenhall et al. (2020) with a hidden layer size of 256. Both the position and view direction are encoded using positional encoding.

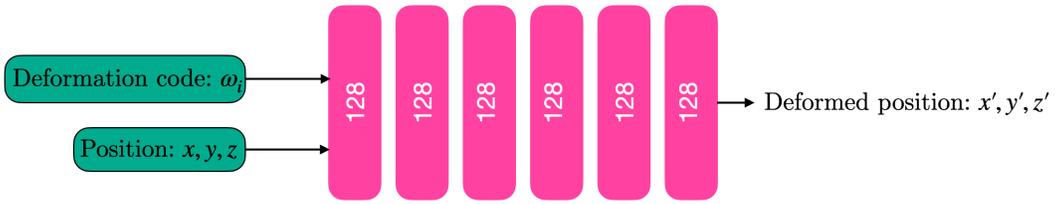


Figure 8: **Deformation network architecture used in FLAME-*in*-NeRF.** FLAME-*in*-NeRF uses the deformation network architecture from [Park et al. \(2021\)](#) with a hidden layer size of 128. The position is encoded using positional encoding with coarse-to-fine regularization [Park et al. \(2021\)](#) (See Section 3.1 in the paper for details).