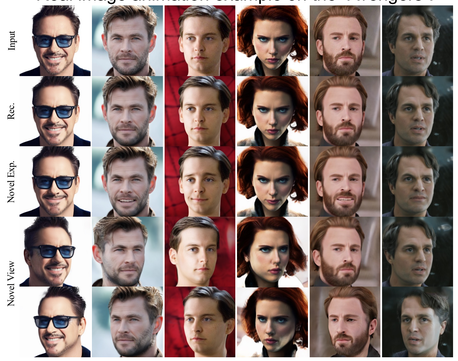
NeRFInvertor: High Fidelity NeRF-GAN Inversion for Single-shot Real Image Animation

Yu Yin[1,3], Kamran Ghasedi[2], HsiangTao Wu[2], Jiaolong Yang[2], Xin Tong[2], Yun Fu[1]
[1] Northeastern University, [2] Microsoft , [3] Case Western Reserve University

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

Real image animation example on the "Avengers".



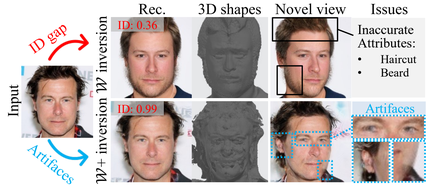
- *Input & Output*:
  Single reference image → Realistic **novel views** and **expressions** animation
- *Goal*:
  High-fidelity, 3D consistent, and identity-preserving talking head synthesis of real person/subjects

Contributions

- Universal and easy-to-apply inversion method for NeRF-GAN models
- High-fidelity and identity-preserving animation of real subjects given only a single image
- Novel geometric constraint by leveraging density of in-domain samples

Backgrounds

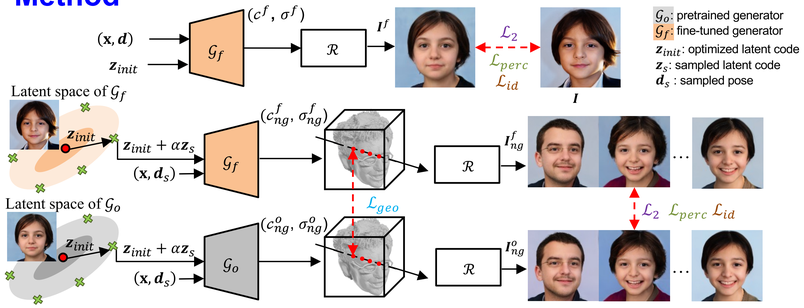
- Challenges of traditional W [1] and W+ [1] inversion
  Trade-off between ID-preserving and removing artifacts.



Specifically, W space inversion results in clean geometry but identity gap between real and generated images, and W+ space inversion causes preserving of identity attributes but inaccurate geometry and visual artifacts.

- Application:
  - Metaverse, talking head synthesis, AR/VR, video editing …
  - 3D data augmentation for AI model training

Method



- **Framework of NeRFInvertor**: Given the optimized latent code z_{init}, we fine-tune the generator with *image space loss* functions to reduce the identity gap. We also apply an *explicit and implicit geometrical constraint* to maintain the model's ability to produce high-quality and 3D-consistent images.
- *Image Space Supervision*

We first apply image space supervision to push the generated image to match the input image in the original view d.

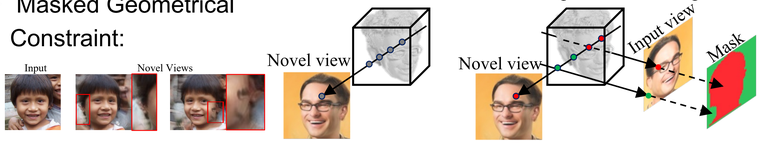
$$\mathcal{L}_{img} = \lambda_1 \mathcal{L}_{pix}(\boldsymbol{I}^f, \boldsymbol{I}) + \lambda_2 \mathcal{L}_{perc}(\boldsymbol{I}^f, \boldsymbol{I}) + \lambda_3 \mathcal{L}_{id}(\boldsymbol{I}^f, \boldsymbol{I})$$

- *Explicit and Implicit Geometrical Regularization*

To augment the NeRF-GAN manifold without worrying about visual artifacts in novel views, we then leverage the surrounding samples of the optimized latent code to regularize the realism and fidelity of the novel view and expression synthesis. The neighborhood latent codes can be obtained by:

$$\mathbf{z}_{ng} = \mathbf{z}_{init} + \alpha \frac{\mathbf{z}_{smp} - \mathbf{z}_{init}}{||\mathbf{z}_{smp} - \mathbf{z}_{init}||_2}, \quad \mathbf{z}_{smp} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$$

- Masked Geometrical Constraint:

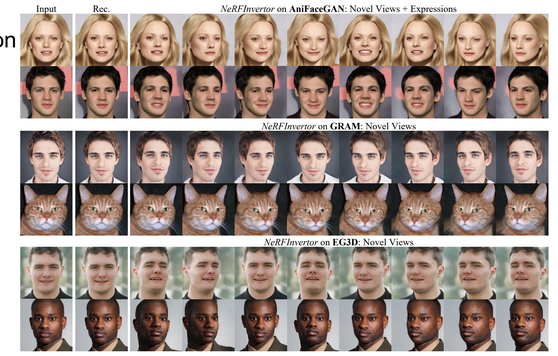


(a) Traditional Rendering (b) Masked Geometrical Constraint

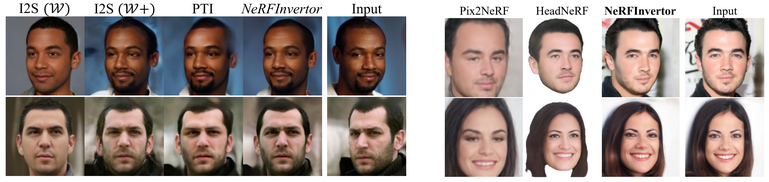
Experiments

- Validating our method on multiple NeRF-GANs
  - AniFaceGAN [2]
  - GRAM [3]
  - EG3D [4]

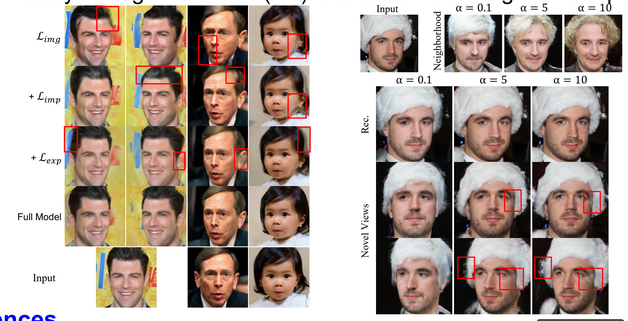
  And multiple datasets
  - FFHQ
  - CelebAHQ
  - Cats
  - Collection of notable individuals



- Comparison with prior inversion methods (left) and single-shot NeRF methods (right)



- Ablation study on regularizations (left) and distance of neighbor samples (right)



References

[1] Abdal et al., Image2stylegan++: How to edit the embedded images? In CVPR, 2020.
[2] Wu et al., Animatable 3d-aware face image generation for video avatars. In NeurIPS, 2022.
[3] Deng et al., Gram: Generative radiance manifolds for 3d-aware image generation. In CVPR, 2022.
[4] Chan et al., Efficient geometry-aware 3d generative adversarial networks. In CVPR, 2022.

Contact Project Page