GCV @ CVPR2023

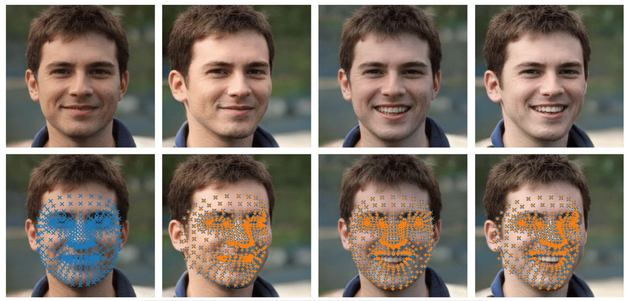
Controllable GAN Synthesis Using Non-Rigid Structure-from-Motion

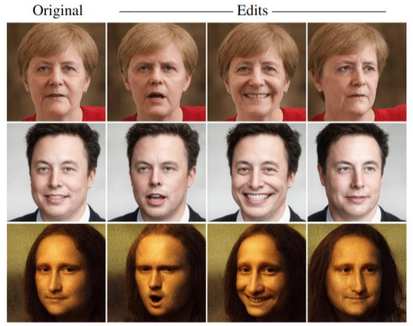
René Haas, Stella Graßhof, and Sami S. Brandt

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

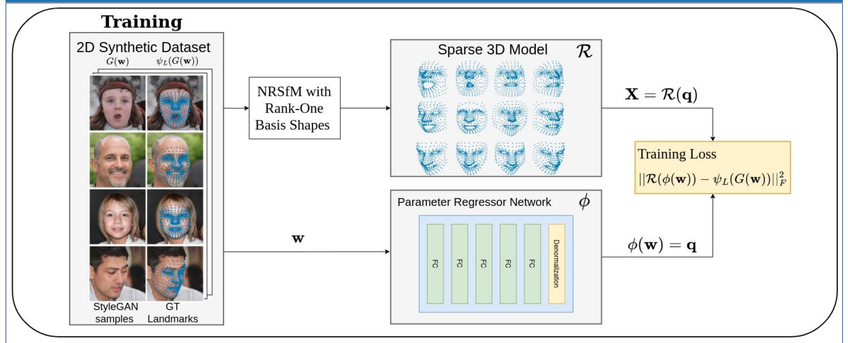
Contributions

- We use a lightweight method for Non-Rigid Structure-from-Motion (NRSfM) to edit latent codes in a pretrained StyleGAN model.
- This allows for explicit 3D control over images synthesized by the 2D StyleGAN generator.
- Coupled with techniques for GAN inversion our method can edit real images.
- Our method does not require any adaptations to the StyleGAN architecture and can be applied to existing pretrained models.





Method. NRSfM with rank-one basis shapes



- We annotate StyleGAN samples using a pretrained landmark extractor ψ_L.
- From this data we fit a NRSfM model

$$\mathcal{R}(\mathbf{q}) = \mathbf{K}[\mathbf{I}_2|\mathbf{0}]\mathbf{R}(\boldsymbol{\theta})\left[\mathbf{B}_0 + \sum_{k=1}^{K} \alpha_k \mathbf{B}_k\right] + \mathbf{t} \otimes \mathbf{1}_L^{\mathrm{T}},$$

where $\mathbf{q} = (\mathbf{k}, \boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{t})$ is an attribute vector describing camera parameters $\mathbf{k}$, rotation $\boldsymbol{\theta}$, non-rigid shape coefficients $\boldsymbol{\alpha}$, and translation $\mathbf{t}$.

- We train an attribute regressor using the loss

$$\mathcal{L}(\mathbf{w}) = \|\mathcal{R}(\phi(\mathbf{w})) - \psi_L(G(\mathbf{w}))\|_F^2 ,$$

to predict the attribute vectors directly from the latent codes as $\hat{\mathbf{q}} = \phi(\mathbf{w})$.

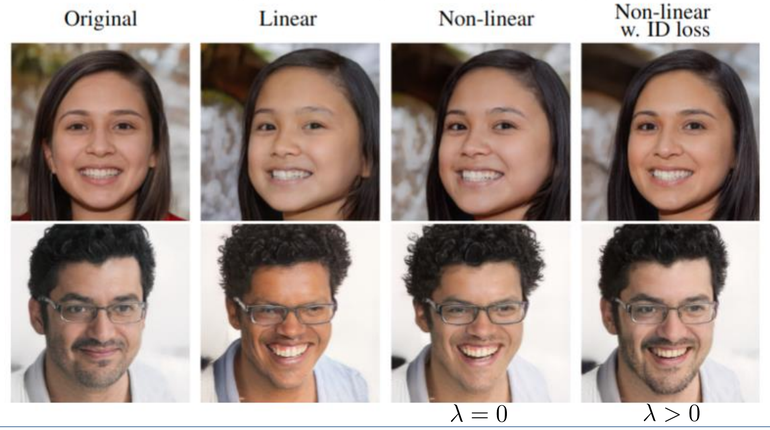
Semantic Editing

- Based on the trained regressor network ϕ, we propose two editing techniques:
- **A linear method,** from inverting the first-order Taylor expansion of the regressor

$$\phi(\mathbf{w}) = \phi(\mathbf{w}_0) + \mathbf{J}|_{\mathbf{w}=\mathbf{w}_0}(\mathbf{w} - \mathbf{w}_0) \quad \rightarrow \quad \mathbf{w}_{\mathrm{edit}} = \mathbf{w}_0 + \mathbf{J}^{\dagger}(\mathbf{q}_{\mathrm{edit}} - \mathbf{q}_0)$$

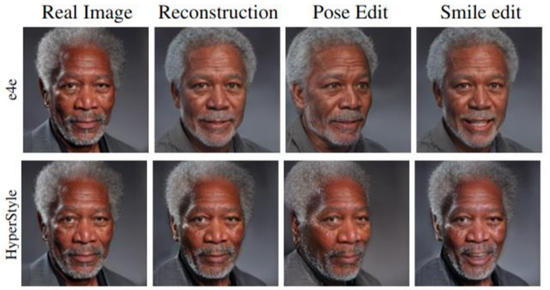
- **A Gradient-based method** which allows for easy identity regularization by using ArcFace [1] to increase the degree of identity preservation during edits.

$$\mathbf{w}_{\mathrm{edit}} = \min_{\mathbf{w}} \|\phi(\mathbf{w}) - \mathbf{q}_{\mathrm{edit}}\|^2 + \lambda \mathcal{D}_{\mathrm{ID}}(G(\mathbf{w}), G(\mathbf{w}_0))$$



Editing real images

- Coupled with GAN inversion techniques like e4e [2] Hyperstyle [3] our method can be used to edit real images.



Attribute transfer

- Our method allows for attribute transfer where pose and facial expression is transferred from a target image onto a source image.

René Haas: renha@itu.dk
Stella Graßhof: stgr@itu.dk
Sami S. Brandt: sambr@itu.dk

IT UNIVERSITY OF COPENHAGEN

References:
[1] Arcface: Additive angular margin loss for deep face recognition. Deng et. al. CVPR 2019.
[2] Designing an Encoder for StyleGAN Image Manipulation. Tov et. al. SIGGRAPH 2021.
[3] Hyperstyle: Stylegan inversion with hypernetworks for real image editing. Alaluf & Tov et.al., CVPR 2022