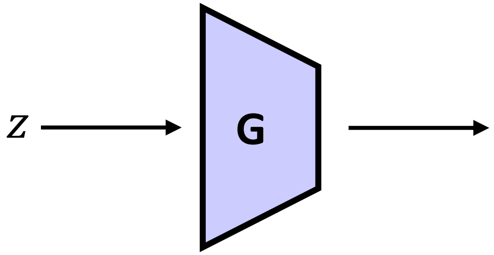
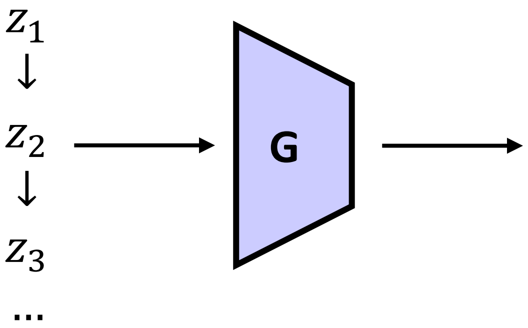
Near-perfect GAN Inversion

GANs



Images from StyleGAN3

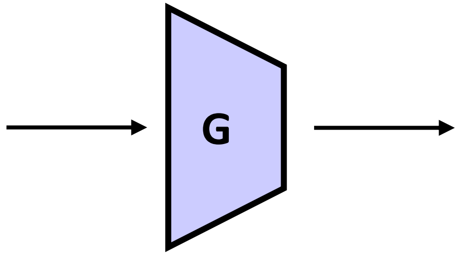
GANs: Smooth Image Manifold



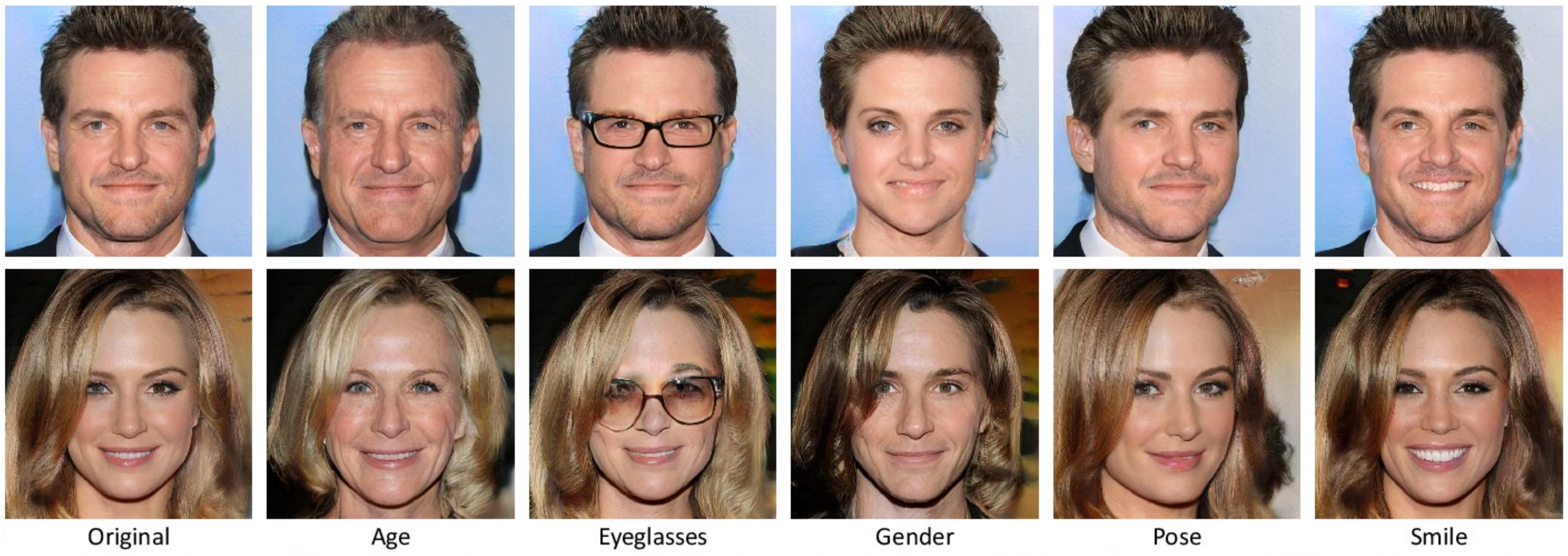
Images from StyleGAN3

GANs: Image editing

z_1
$\downarrow$
z_2
$\downarrow$
z_3
...



z_1 $\longleftarrow$ z_i $\longrightarrow$



Possible to edit

any

face image?

Images from [InterFaceGAN]

Pipeline for Editing Unseen Image via GANs

Image Generation	Image Inversion	Image Manipulation
Achieve a capability to **generate** diverse set of photo-realistic images	Locate the given unseen image in the **range of the Generator**	**Traverse** the Generator space to obtain **meaningful** image manipulations

Tasks are not necessarily sequential.

Can be tackled separately.

Typical GAN setup

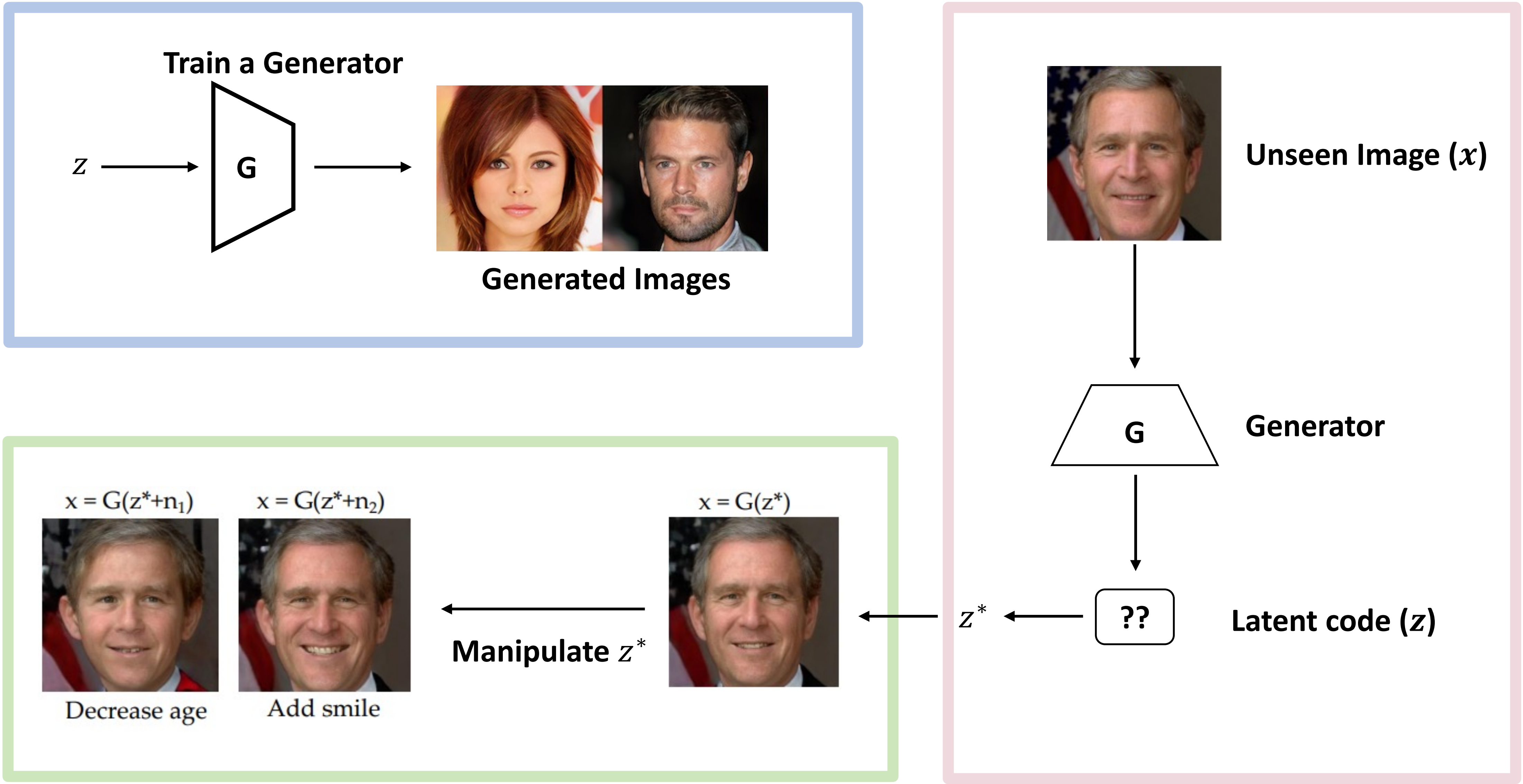


Image from [GAN Inversion: A Survey, Xia et al. arXiv '21]

Image Inversion: Common Techniques [1]

- Invert a given image back into the **latent space**
- Several methods:
  - Optimization based
  - Learning based
- Can be done in any **intermediate** latent space

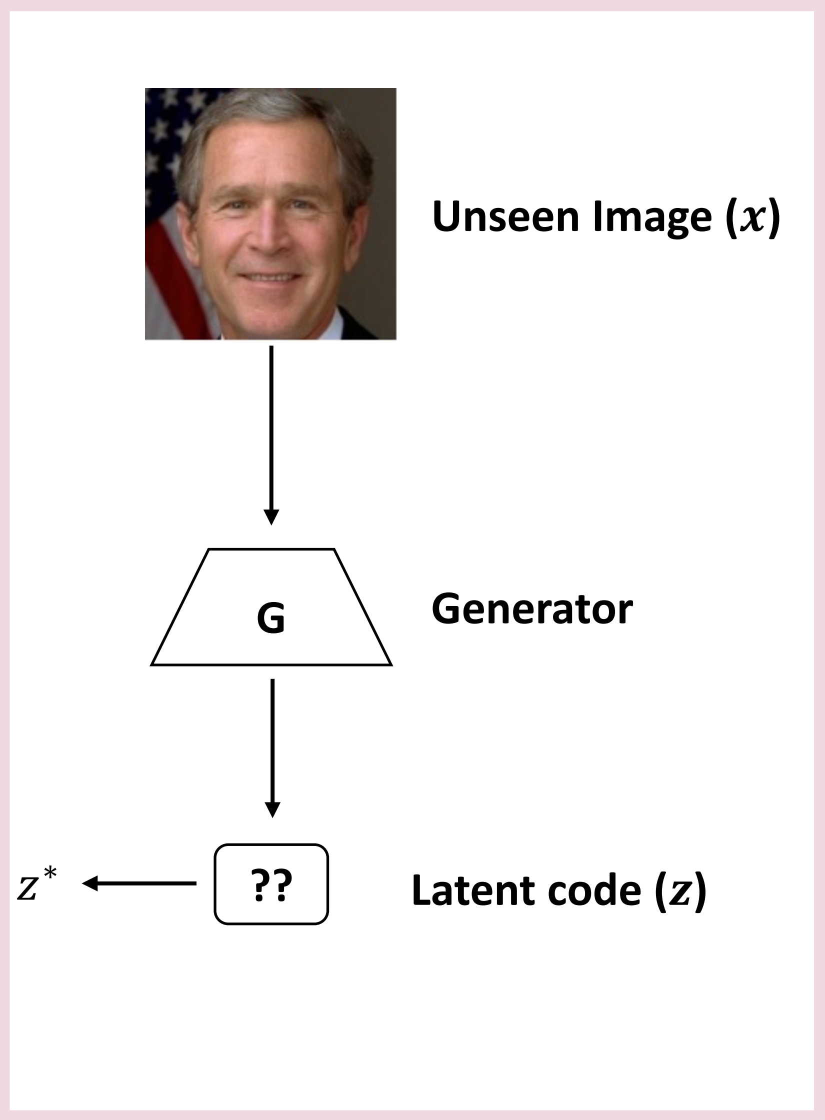


Image Inversion: Common Techniques

- **Optimize** a loss function over z:
  - $z^* = \underset{z}{\operatorname{argmin}}\, l(x, G(z;\theta))$

Iterative methods mainly using gradient descent

Highly non-covex, computationally **expensive**

- **Learning-based** method:

Train **encoder-decoder** model

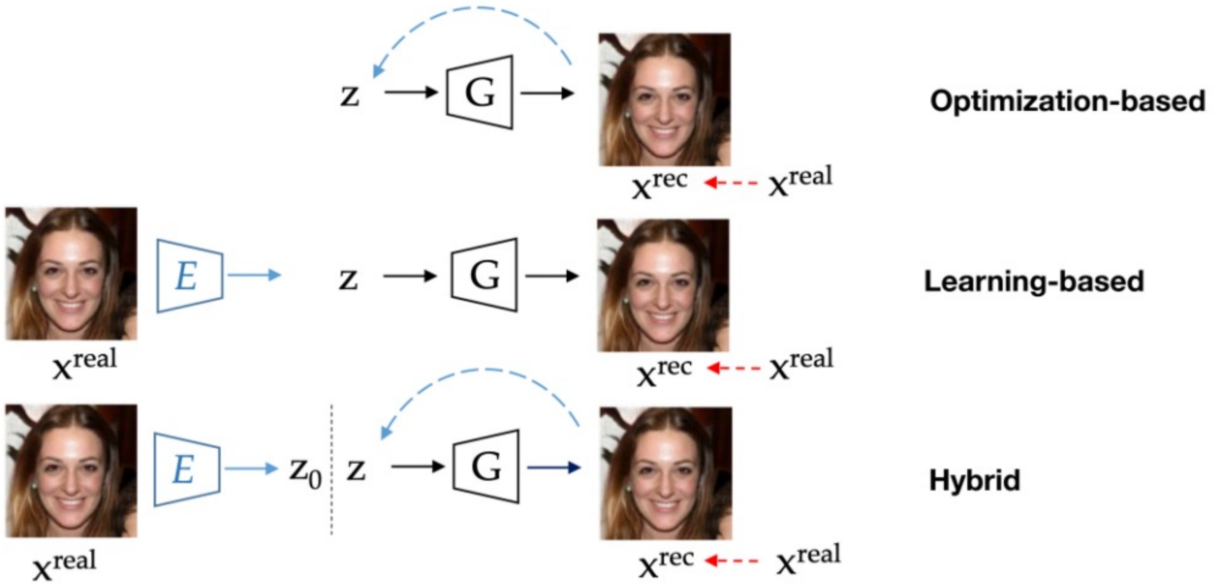


Image from GAN Inversion: A Survey, Xia et al. arXiv '21

Image Inversion: Prior Work

- State-of-the-art learning-based method: **Restyle-encoder [2]**

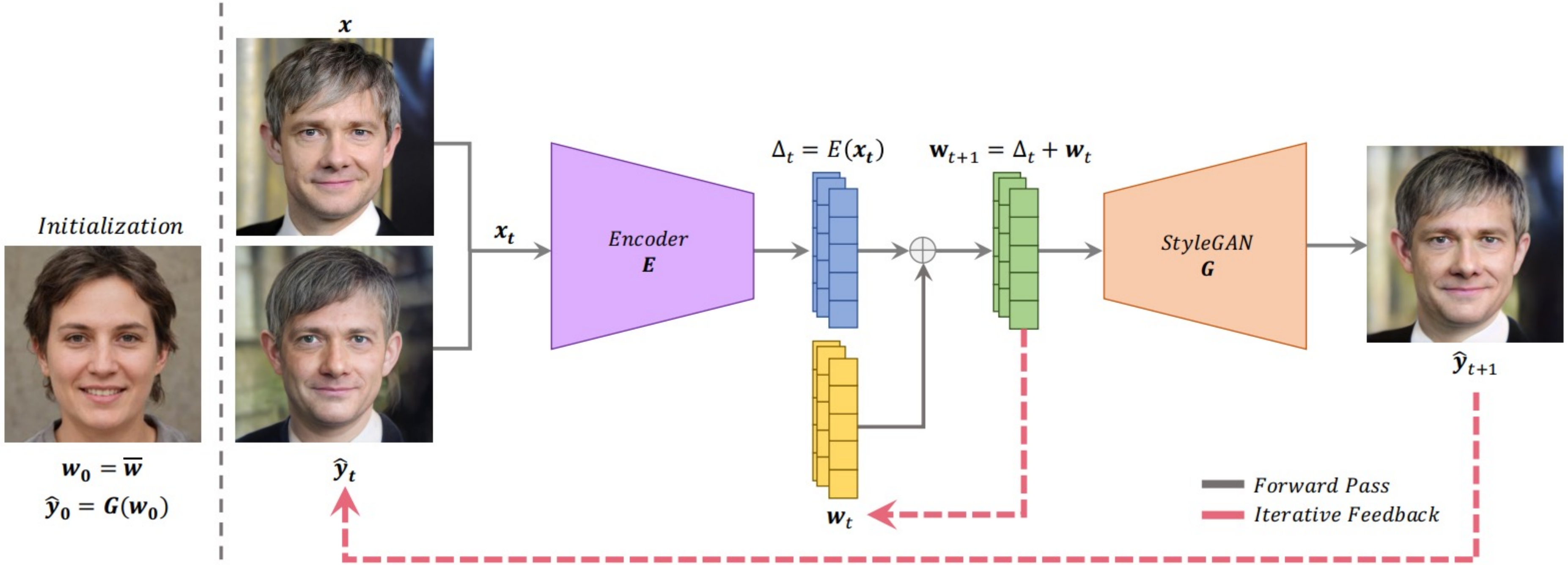
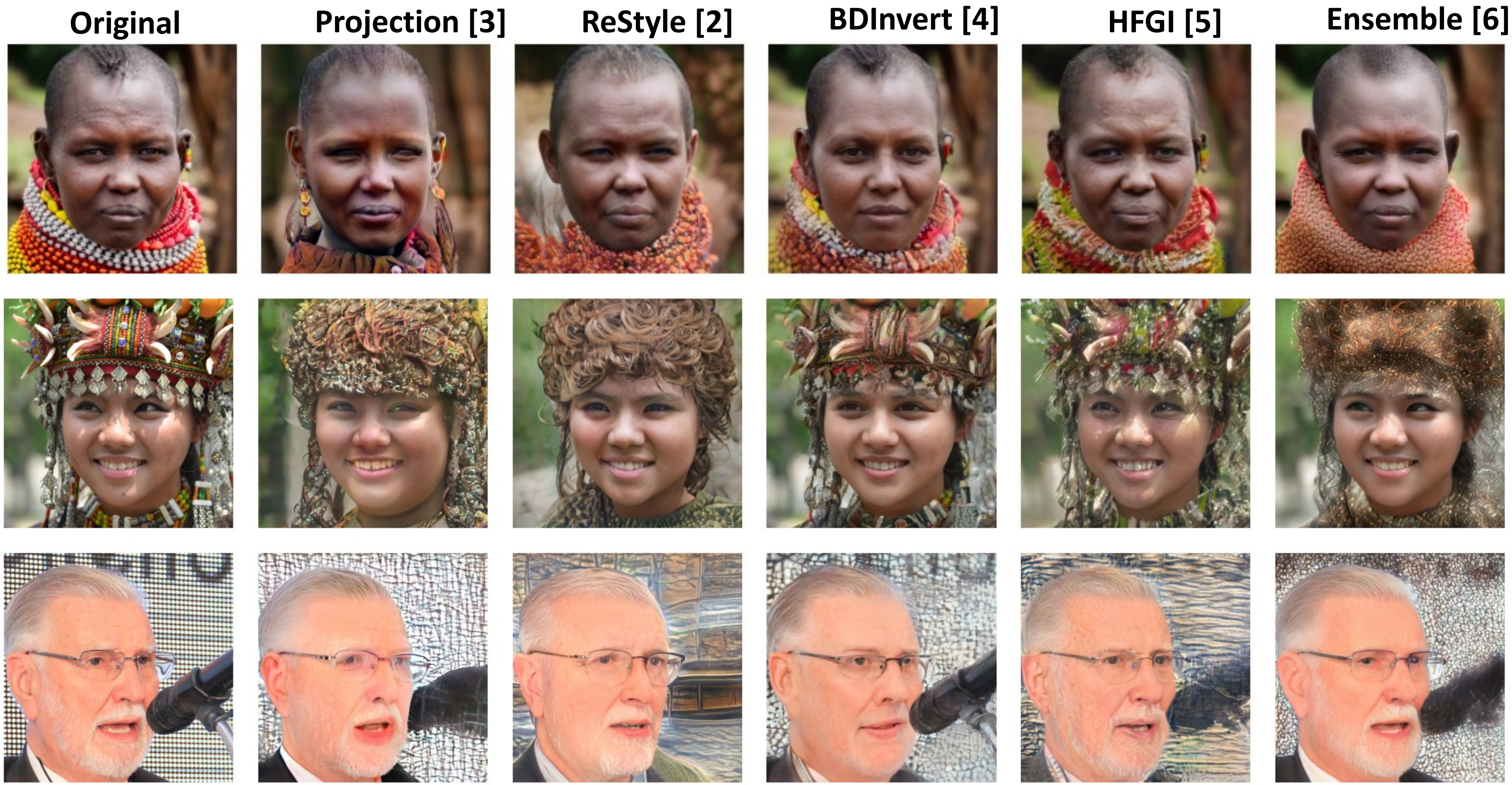


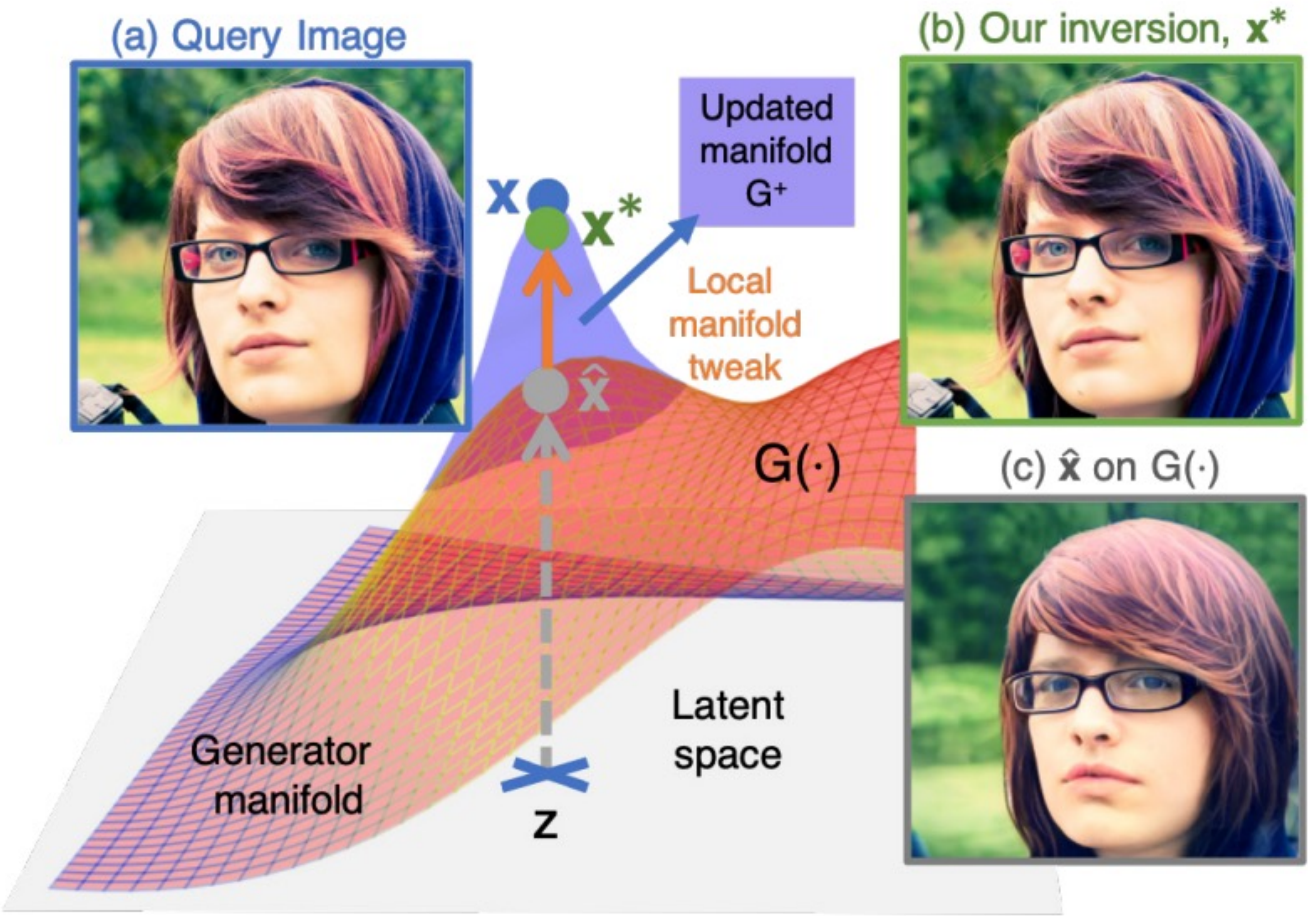
Image from [Restyle Encoder, Alaluf et al., ICCV'21]

Image Inversion: Most techniques fail on Unseen Images



Idea: Fine tuning the Generator

- Initial estimate is obtained using a **learning-based method**
- Generator is **fine-tuned** so that the given image lies on the Generator manifold
- Key is to update the Generator manifold **without affecting the disentanglement** characteristics



How to preserve photo-realism in fine-tuning?

Reconstruction	Photo-realism
Most loss functions **ignore high-frequencies** thus, use: **Laplacian Pyramids** $\boldsymbol{L_{recon} = Lap.Pyd.(x, G(z))}$	Degree of realism **is governed by Discriminator** Thus, Use: **Discriminator Loss** $\boldsymbol{L_{adv\_local} = log\left(D(x)\right) + log\left(1 - D(G(z))\right)}$

How to prevent overfitting?

- Key is to update the Generator manifold **without affecting the disentanglement** characteristics

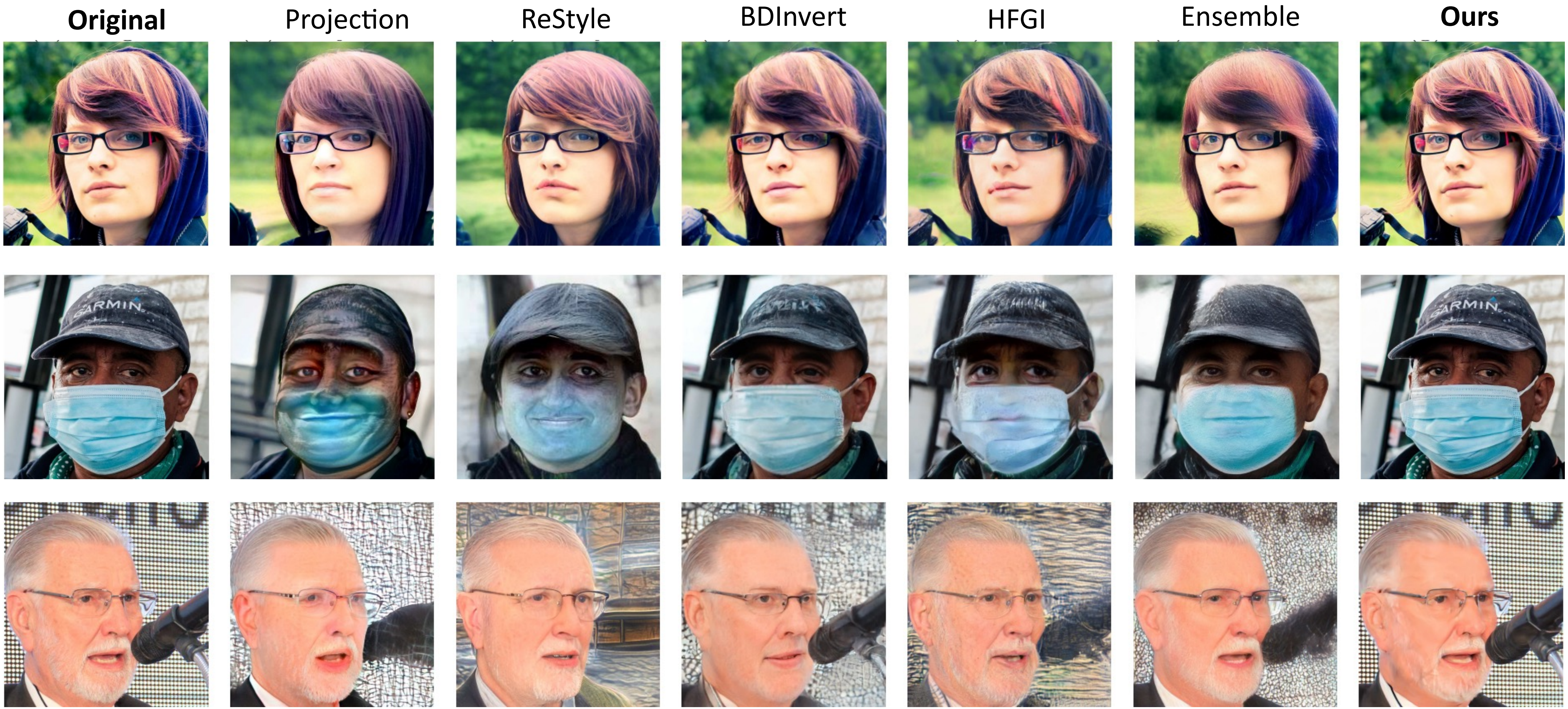
Global Cohesion Loss

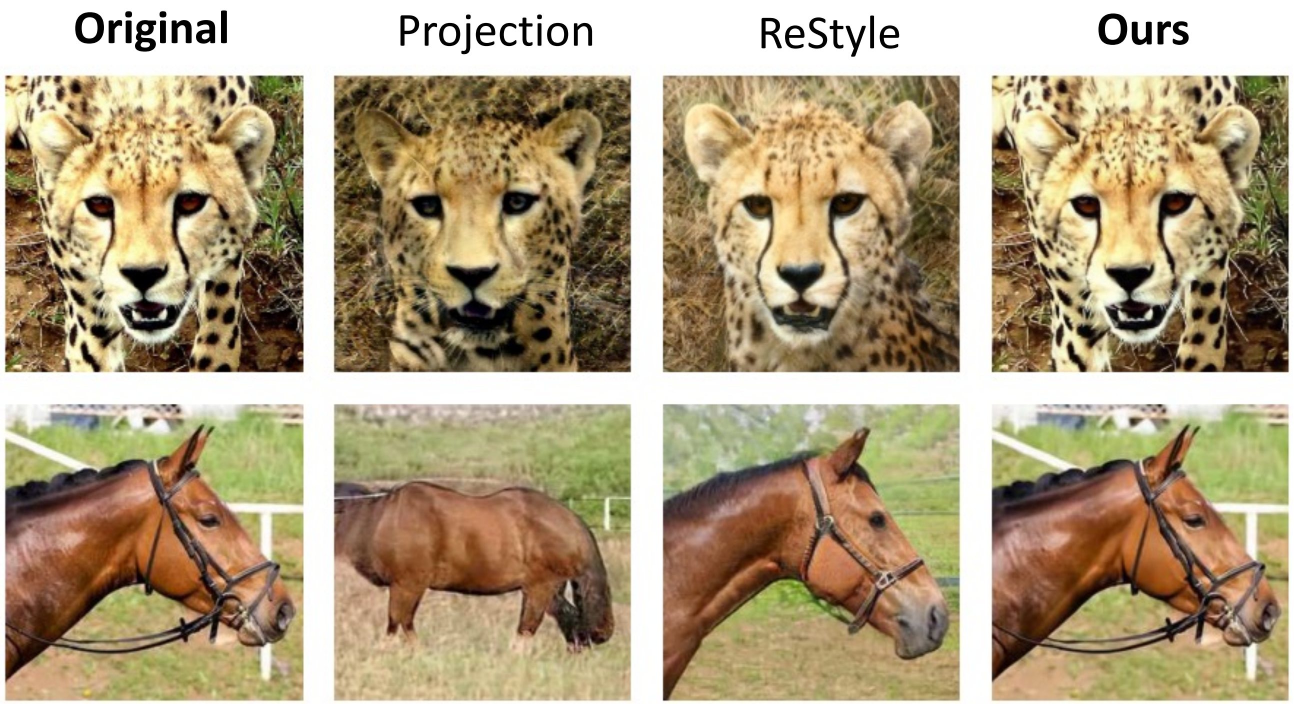
Use Discriminator Loss on other parts of the GAN manifold

$$\boldsymbol{L_{global}} = \mathbb{E}_x\,[\boldsymbol{log(D(x))}] + \mathbb{E}_z[\boldsymbol{log\,(1 - D(G(z)))}]$$

$$\boldsymbol{L_{total}} = \mathbb{I}_p[\boldsymbol{L_{local}} + \boldsymbol{L_{adv_{local}}}] + \boldsymbol{L_{global}}$$

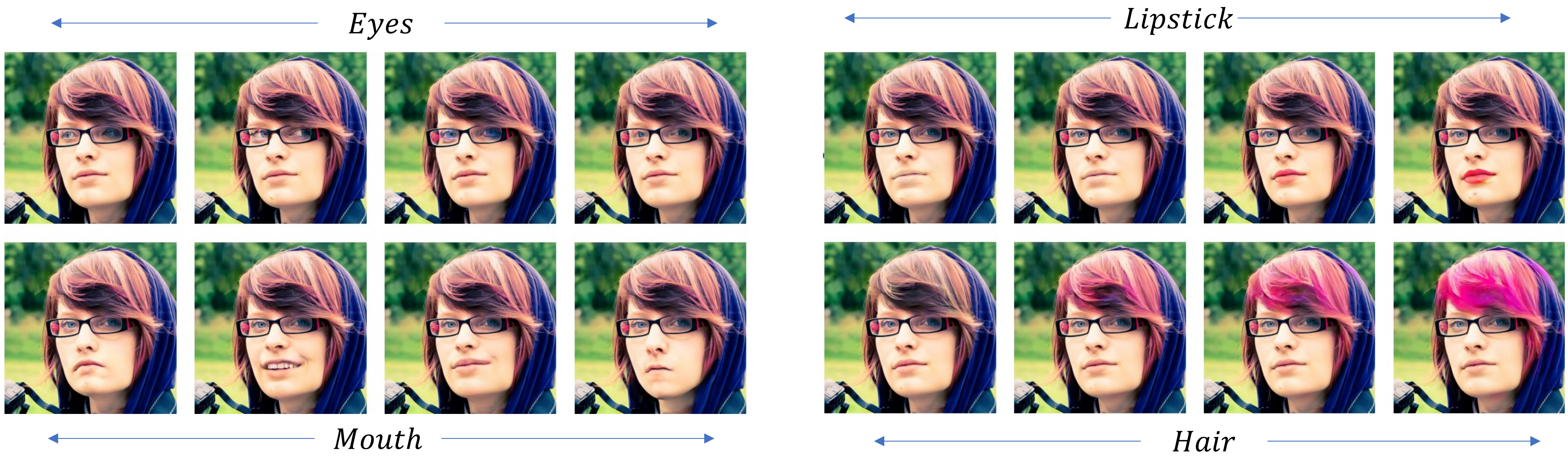
Results: FFHQ dataset



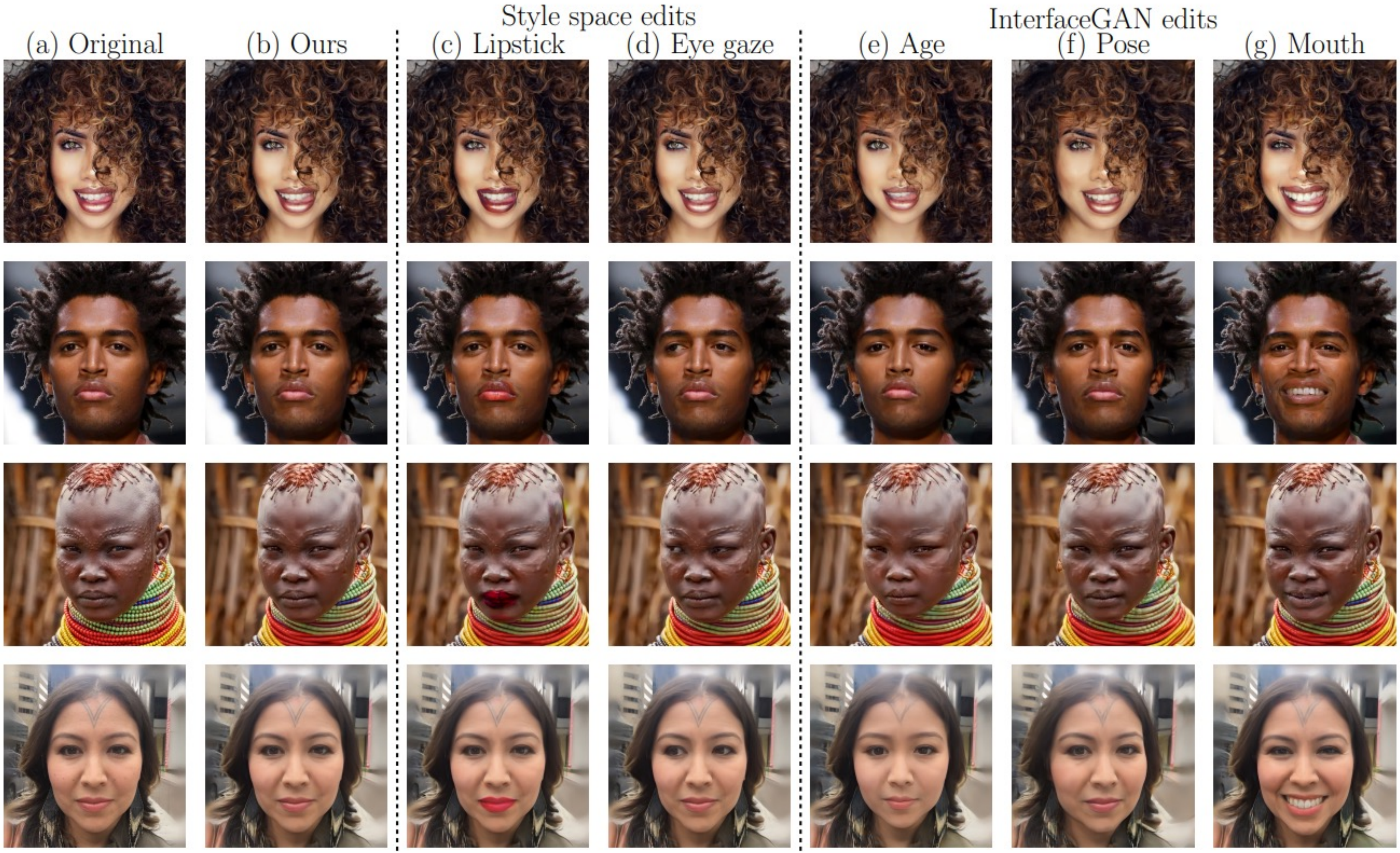

Original
Projection
ReStyle
Ours


Original
Projection
ReStyle
Ensemble
Ours

Most off-the-shelf Editing methods works!



Using **StyleSpace [7]**


Style space edits
InterfaceGAN edits
(a) Original
(b) Ours
(c) Lipstick
(d) Eye gaze
(e) Age
(f) Pose
(g) Mouth

Some quantitative results

Dataset	CelebA-HQ		Stanford Cars		AFHQ-Wild		LSUN-Horse	
Metric	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS	MSE	LPIPS
Projection O [22]	.074 ±.055	.429 ±.044	.318 ±.120	.486 ±.067	.126 ±.066	.491 ±.036	.240 ±.195	.454 ±.072
ReStyle E [4]	.050 ±.019	.475 ±.038	.082 ±.035	.352 ±.063	.085 ±.039	.509 ±.037	.159 ±.070	.525 ±.071
BDInvert E [16]	.016 ±.080	.373 ±.040	–	–	–	–	–	–
HFGI E [34]	.032 ±.054	.423 ±.045	–	–	–	–	–	–
Ensemble H [11]	.017 ±.011	.373 ±.038	.284 ±.025	.448 ±.053	–	–	–	–
Ours	**.004** ±.006	**.283** ±.050	**.006** ±.007	**.154** ±.046	**.014** ±.013	**.382** ±.087	**.005** ±.009	**.141** ±.043

Ablation on FFHQ

Loss	full	w/o $\mathcal{L}_{recon}$	w/o $\mathcal{L}_{adv\_local}$	w/o $\mathcal{L}_{global}$
MSE ↓	**.050**	**.155**	**.080**	.052
FID ↓	**5.21**	3.04	4.13	**164.7**

Original

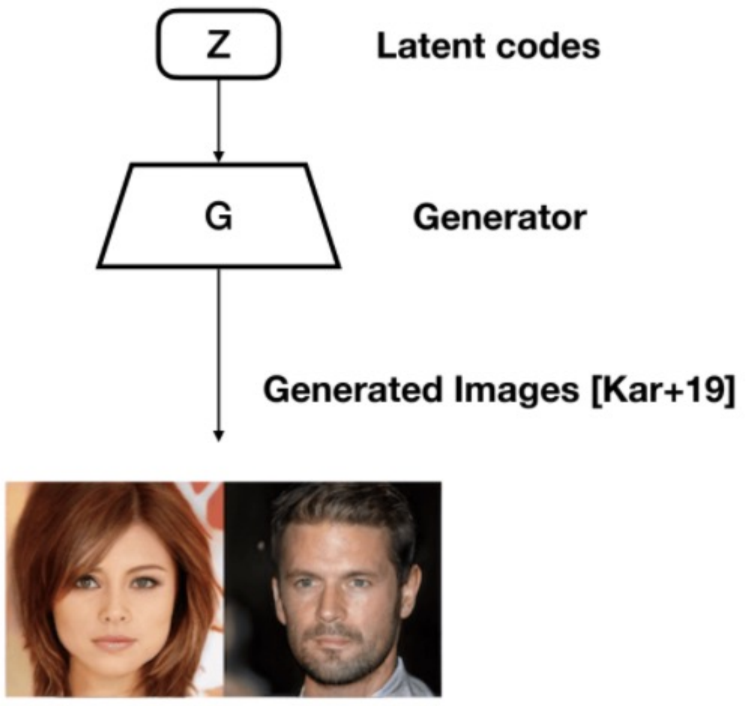
Eye-glasses

Angry

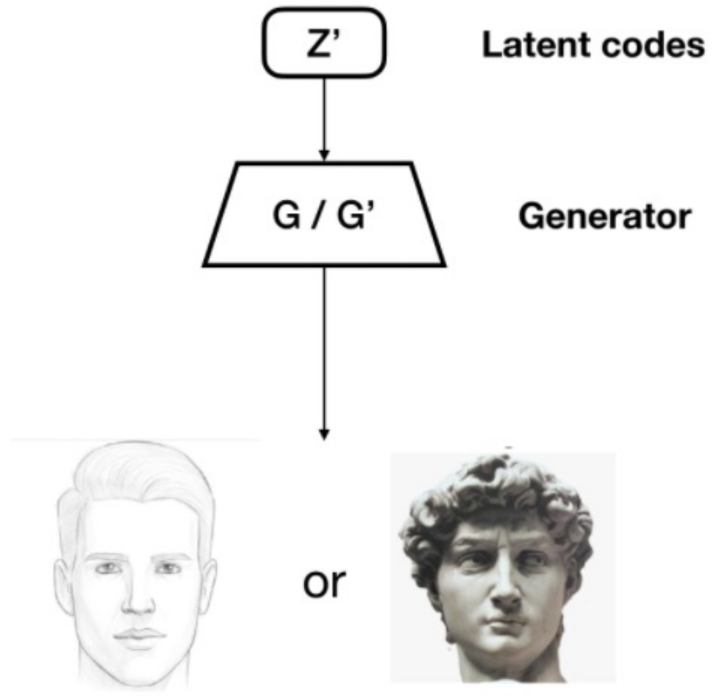
Smile

What can a pre-trained StyleGAN2 generate?

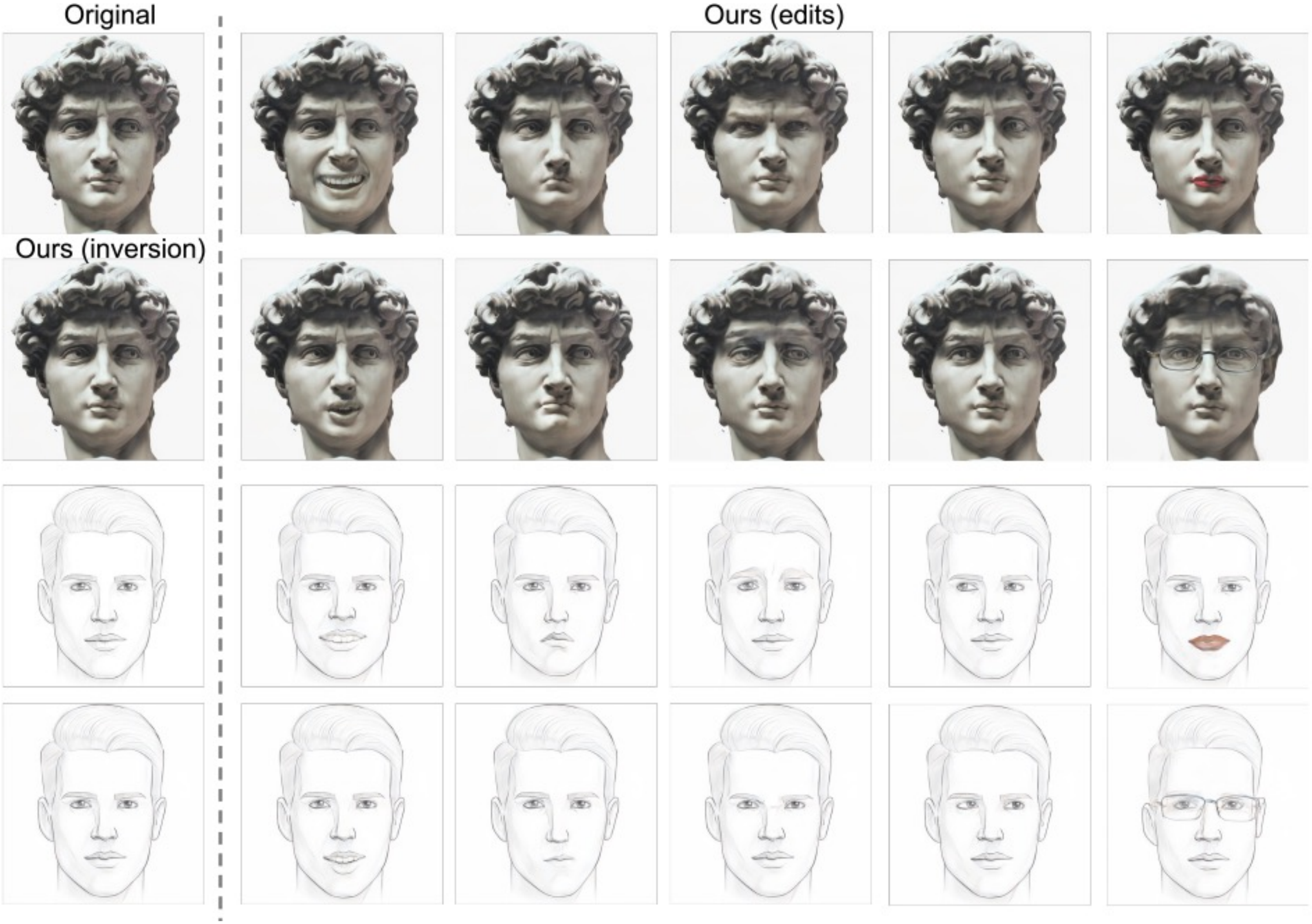
- Consider a StyleGAN2 model trained on FFHQ dataset.



- Can it generate out-of-domain images such as,
  - Sketches or Statues



Can be obtained by fine-tuning…



References

[VS+18] V. Shah and C. Hegde. “Solving Linear Inverse Problems using GAN Priors: An Algorithm with Provable Guarantees”. In: Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing (ICASSP). 2018.

[Kar+17] Tero Karras et al. “Progressive growing of gans for improved quality, stability, and variation”. In: arXiv preprint arXiv:1710.10196. 2017.

[Xia+21] W. Xia et al. GAN Inversion: A Survey, In: arXiv, 2021.

[Bor+17] A. Bora et al. “Compressed Sensing using Generative Models”. In: Proc. Int. Conf. Machine Learning. 2017.

[Goo+14] I. Goodfellow et al. “Generative adversarial nets”. In: Proc. Adv. in Neural Processing Systems (NIPS). 2014.

[EA+20] Erik Ha¨rko¨nen, Aaron Hertzmann, Jaakko Lehtinen, and Sylvain Paris. Ganspace: Discovering interpretable gan controls. arXiv.

[ZD+21] Z. Wu, D. Lischinski, E. Shechtman, Disentangled Controls for StyleGAN Image Generation, arXiv