

Moving forward with StyleGAN to Real Data and New Domains

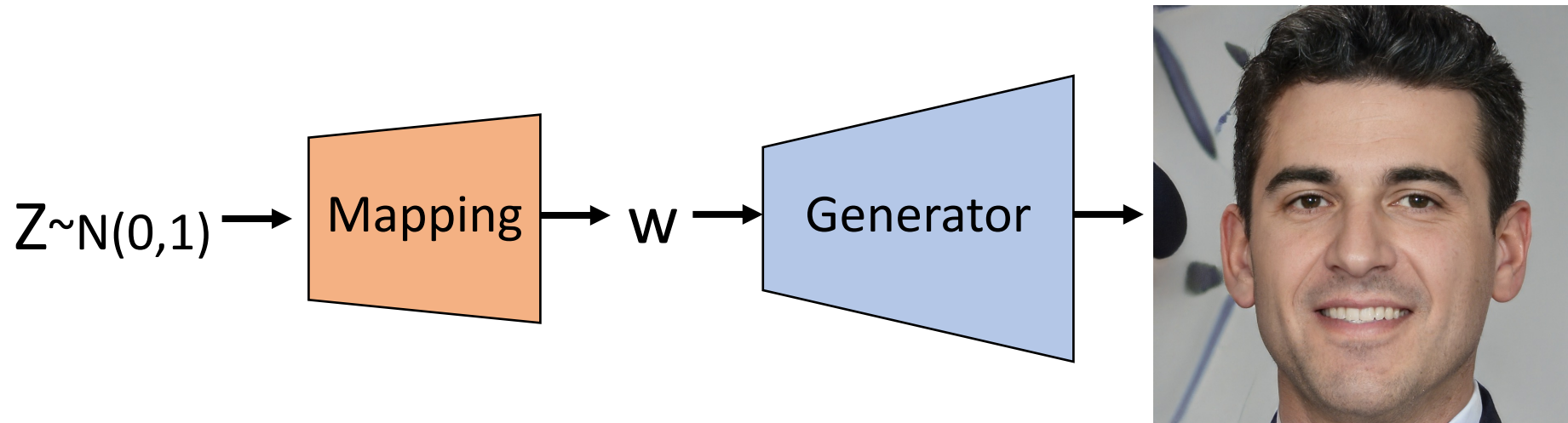
Ron Mokady

- Background
- Pivotal tuning Inversion
- Stitch it in time
- Self-Distilled StyleGAN

StyleGAN



StyleGAN





חדשות 20 - ערוץ החדשות שלי

August 6 at 12:51 PM · 🌐



***** לא ברור לנו עד כמה הפוסט אמין אבל התחברנו מאוד

למה שכתוב. *****

#שתפו #רץברשת

דנה אוברלנדר על מחנה השמאל: ברשת

אני מתבאסת מהמחנה שלי. כל חיי הפגנתי למען השלום, כל חיי

הצבעתי לשמאל - ממפלגת העבודה-מרצ... [More](#)

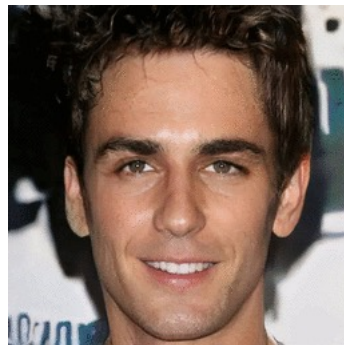
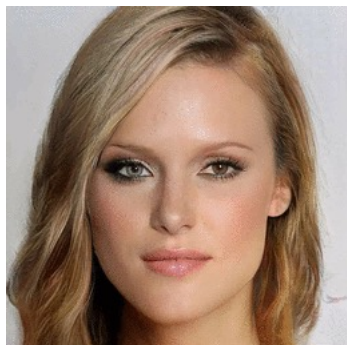


156

50 Comments 50 Shares

Latent-based Editing

$$w + \Delta w$$



Interpreting the Latent Space of GANs for Semantic Face Editing. Shen et al.

Editing real image

GAN Inversion

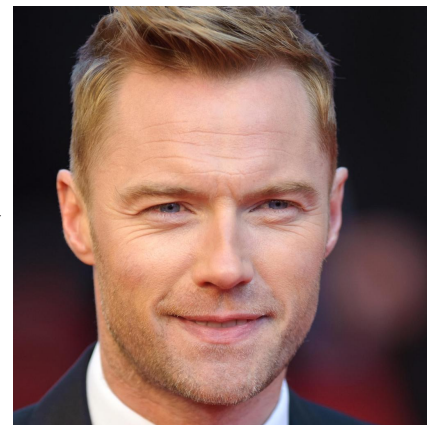
Input



Output:

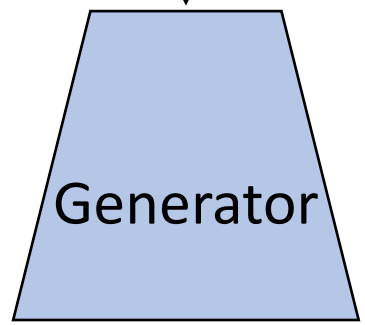
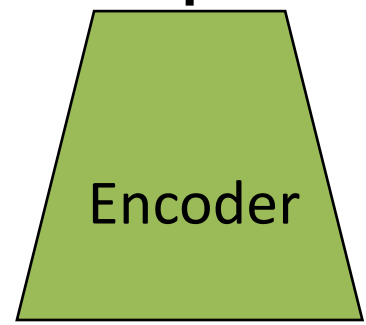
W^*

Generator



Encoder

Output: W^*



Input

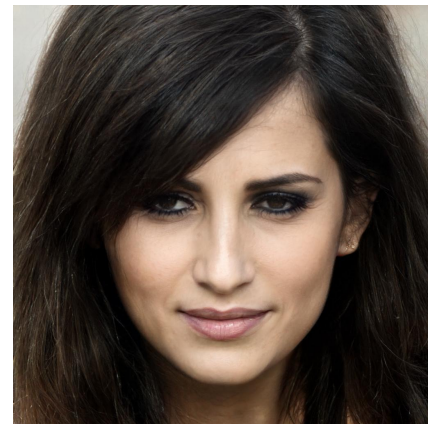
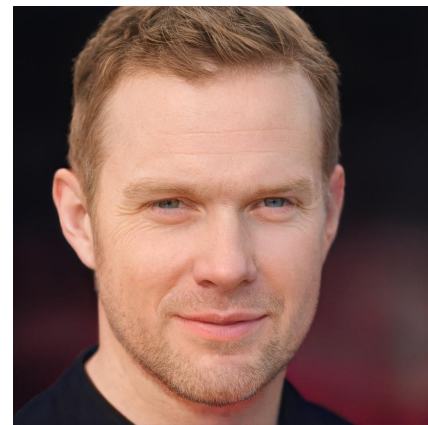


Is W space is expressive enough?

Original



Inversion



Why?

$W+$

Image2StyleGAN: How to Embed
Images Into the StyleGAN Latent
Space?

Abdal et al.

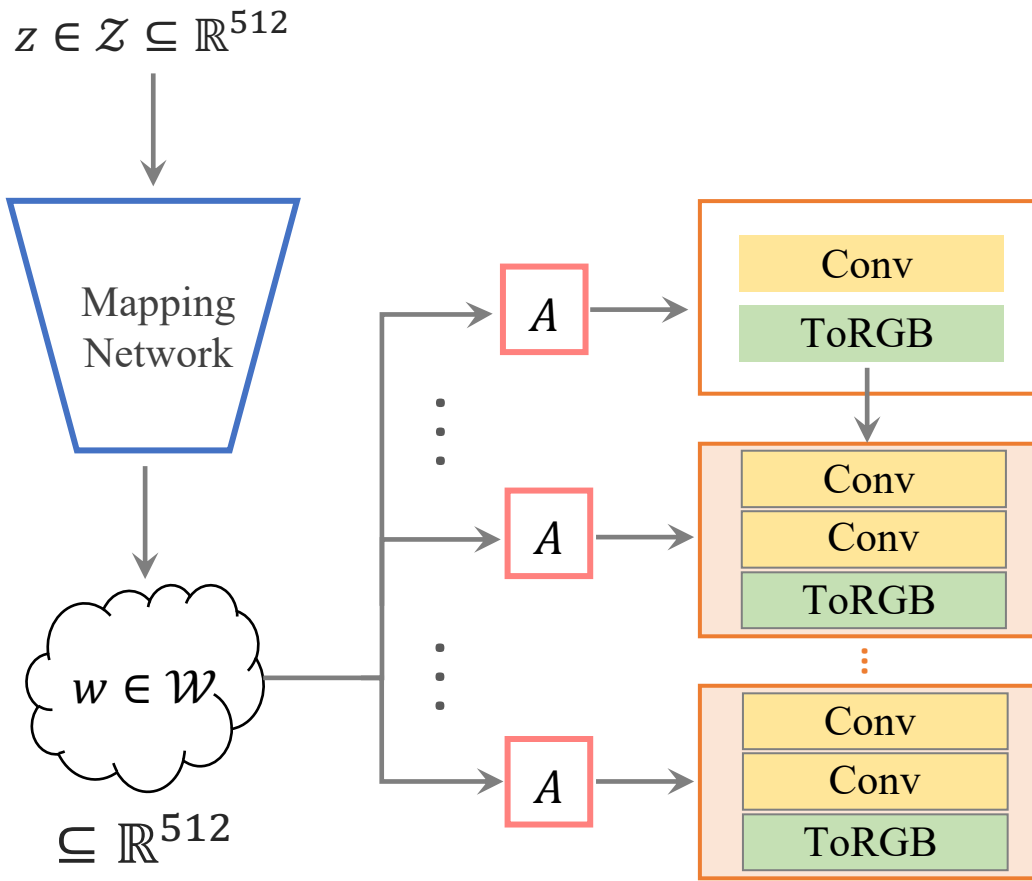
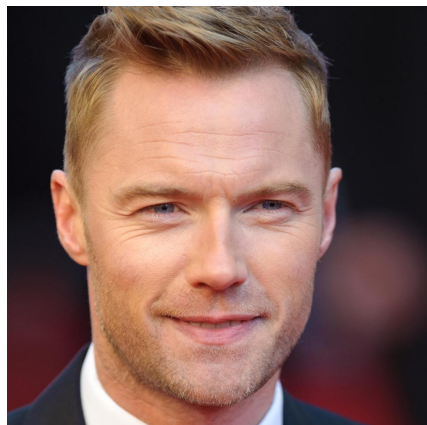


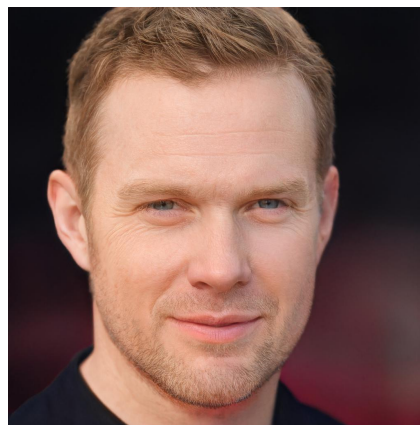
Figure credit - Yuval Alaluf

W+ Optimization

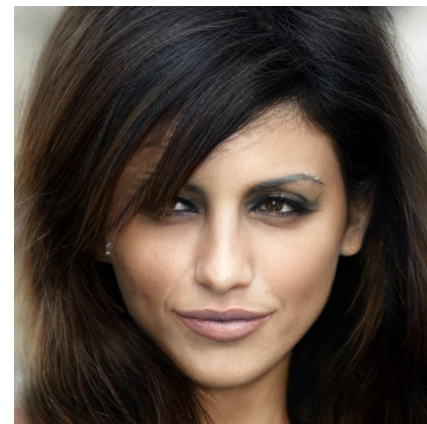
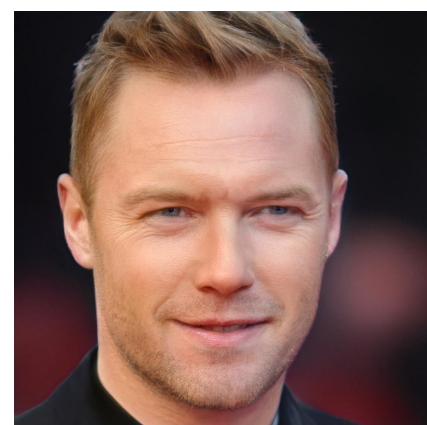
Original



W



W+



Analyzing and improving the image quality of stylegan. *Karras et al.*

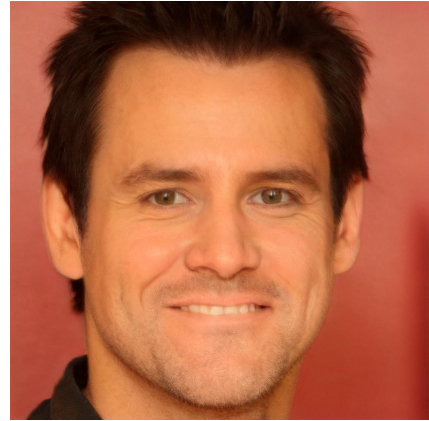
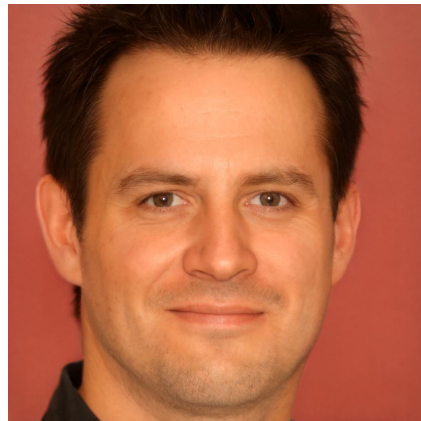
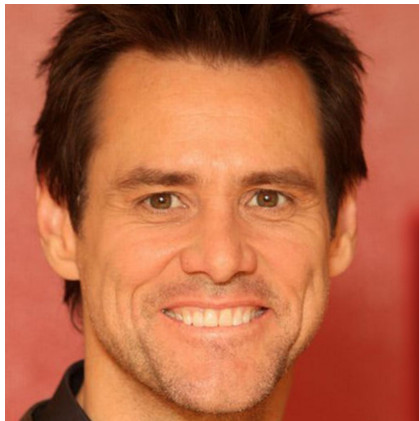
W+ editability

Original

W

W+

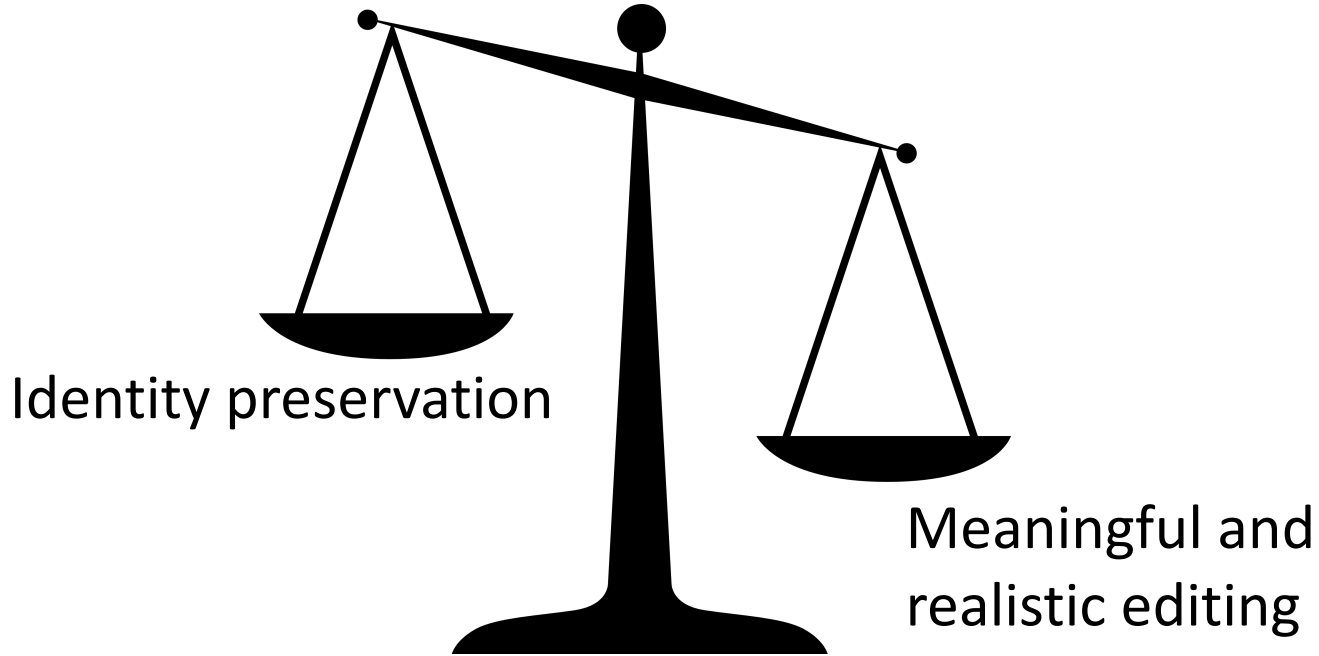
*Smile
Removal*



Frontalization



Distortion-Editability Tradeoff





Tradeoff sweet spot - e4e

Original

Inversion

No smile

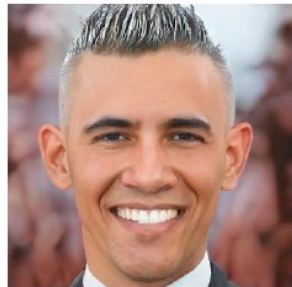
Gender

Glasses

Pose



Unsupervised Editing



“Emma Stone”

“Mohawk hairstyle”

“Without makeup”

“Cute cat”

“Lion”

“Gothic church”

StyleCLIP: Text-Driven Manipulation of StyleGAN Imagery. Patashnik et al.

Inverting real images?



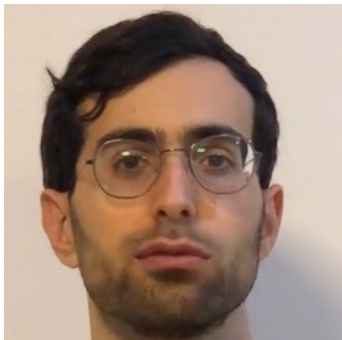
Original



e4e



StyleCLIP editing



Pivotal Tuning for Latent-based Editing of Real Images

Daniel Roich, **Ron Mokady**, Amit H. Bermano, Daniel Cohen-Or

Original



e4e



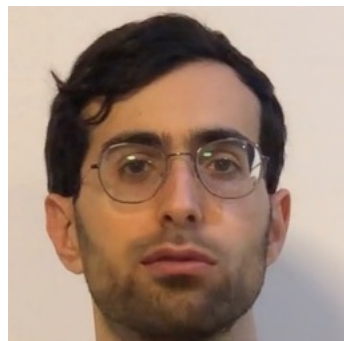
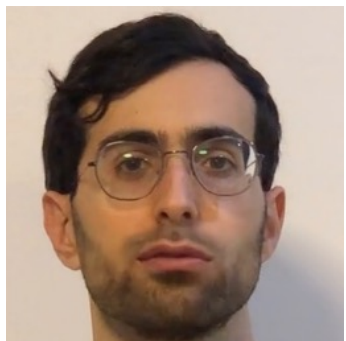
e4e+ editing



PTI



PTI + Editing



Original



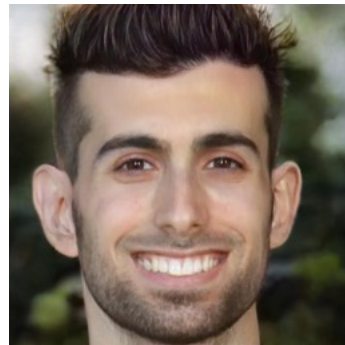
-smile



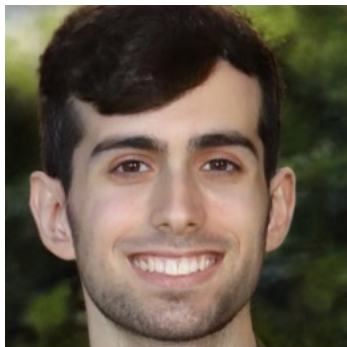
+smile



Mohawk



Young



Pose



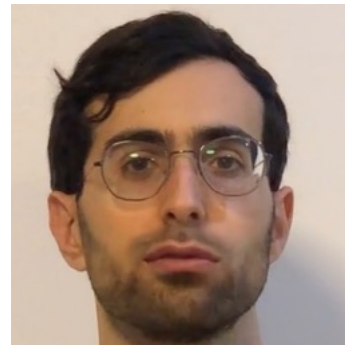
Old



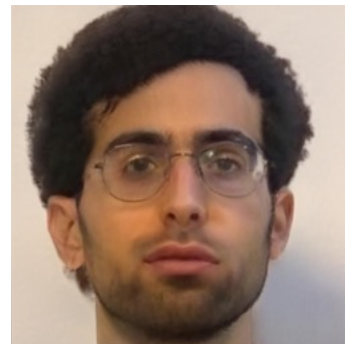
Afro

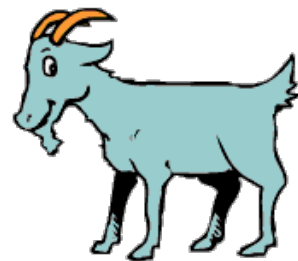
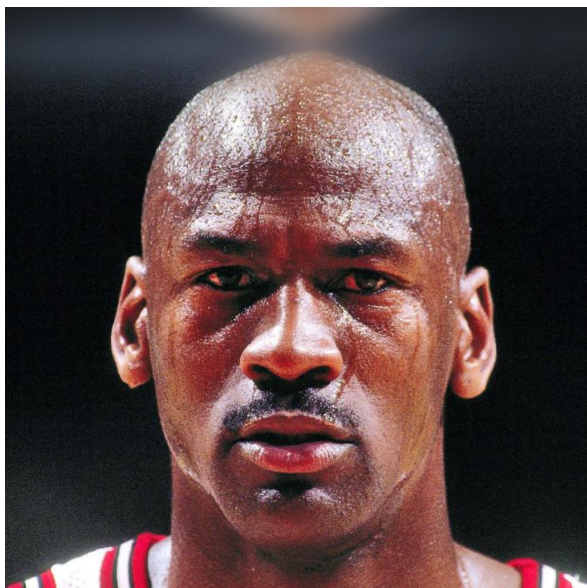


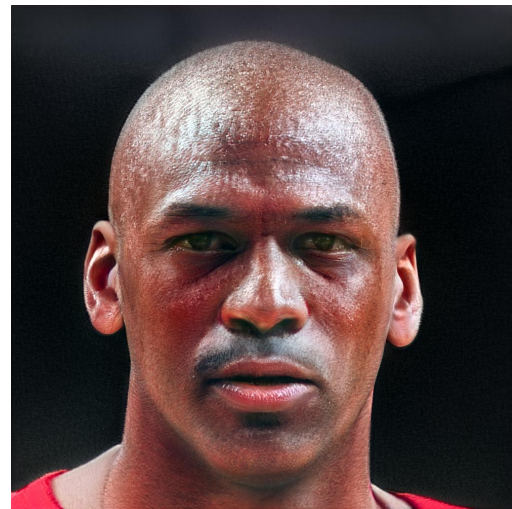
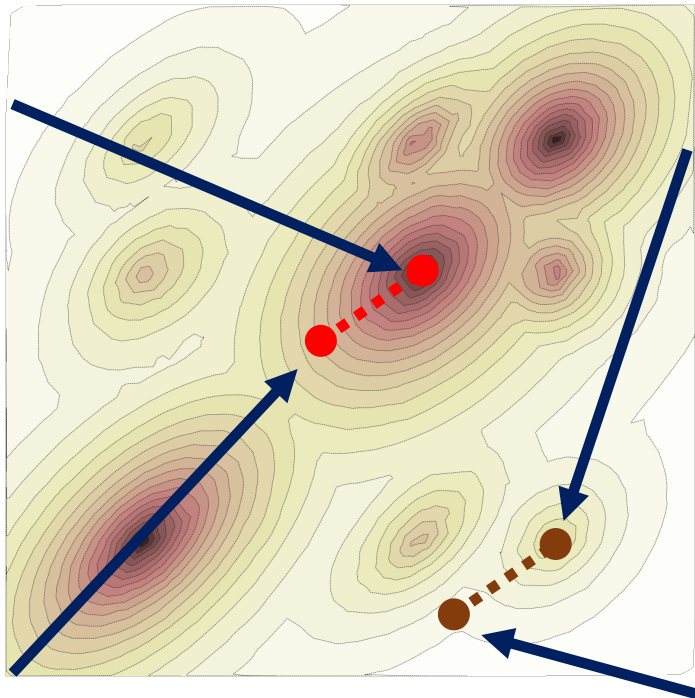
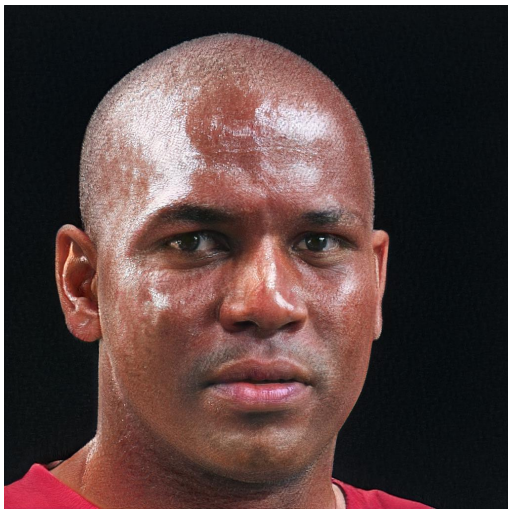
Original

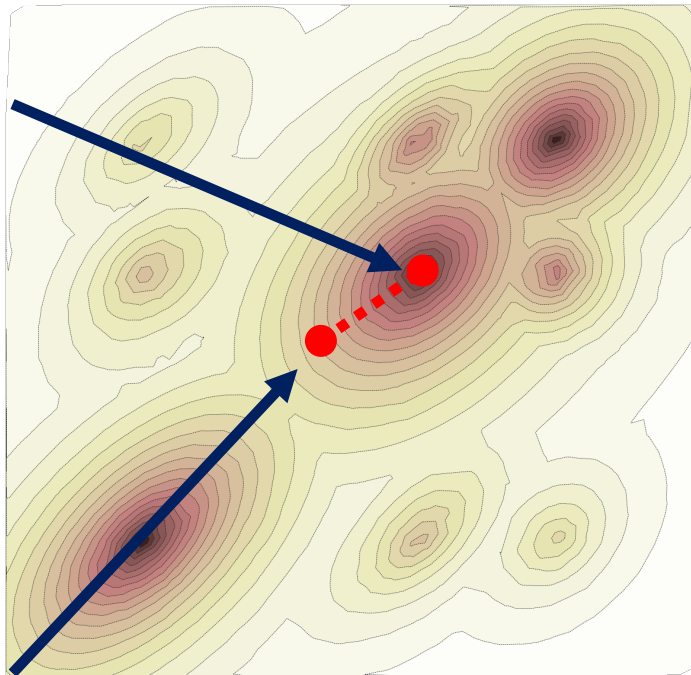
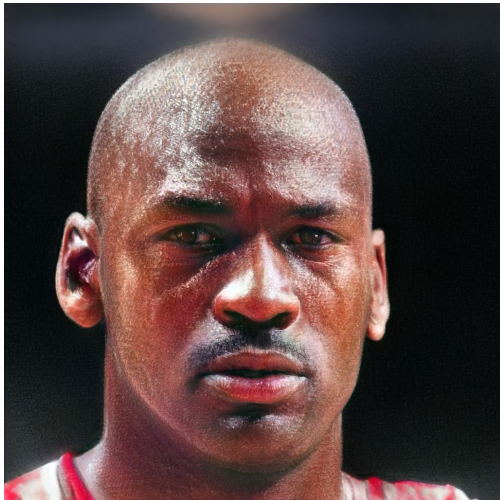
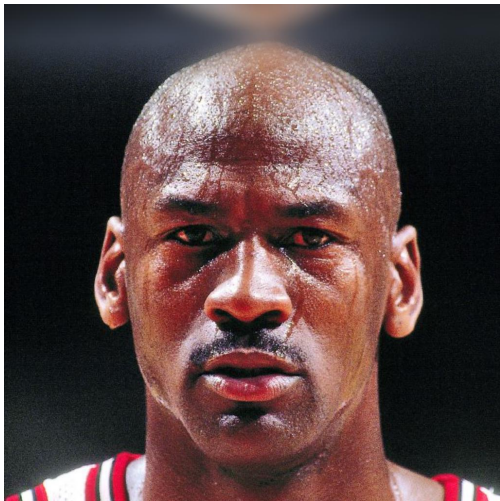


Afro



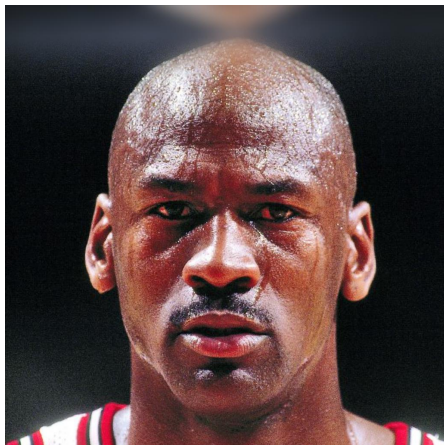




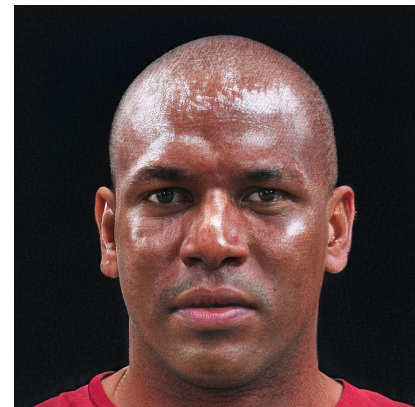
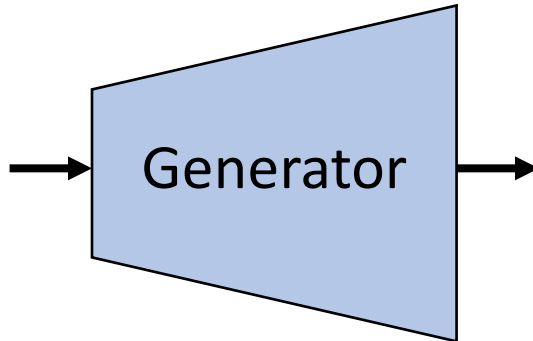


GAN Inversion

Input

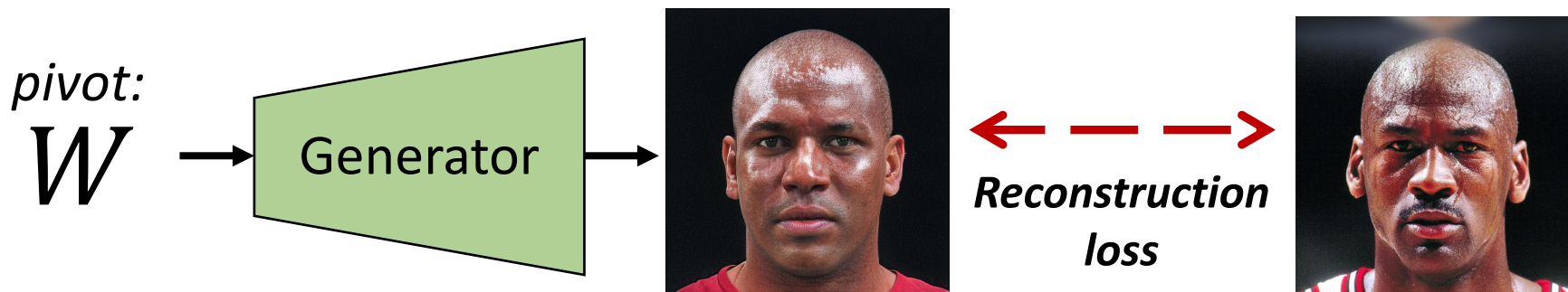


pivot:
 W

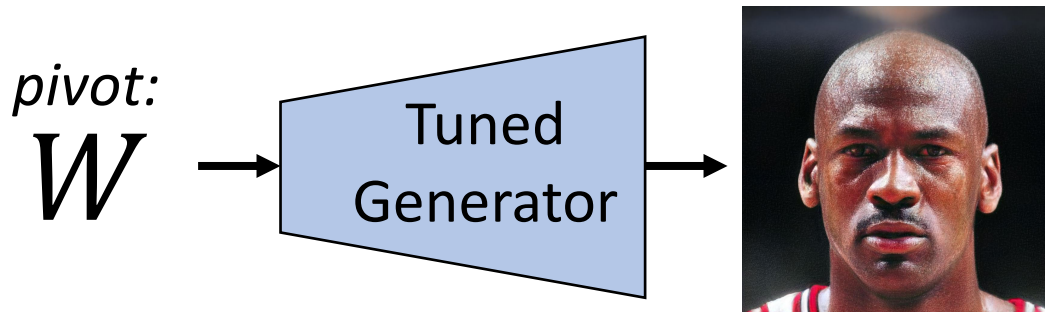


$$\hat{w} = \arg \min_w \mathcal{L}(x, G(w; \theta))$$

Pivotal Tuning

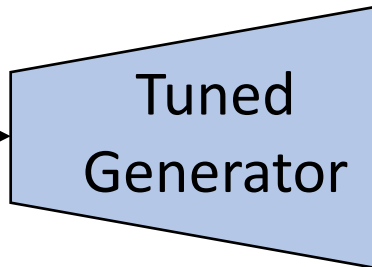


$$\hat{\theta} = \arg \min_{\theta} \mathcal{L}(x, G(\hat{w}_{init}; \theta)) \quad \hat{y} = G(\hat{w}_{init}; \hat{\theta})$$

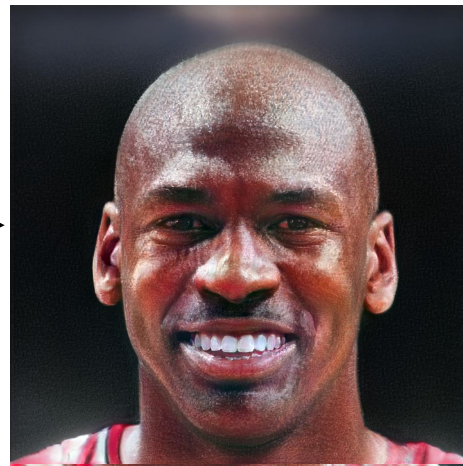


Editing

$w + \Delta w$



Smile



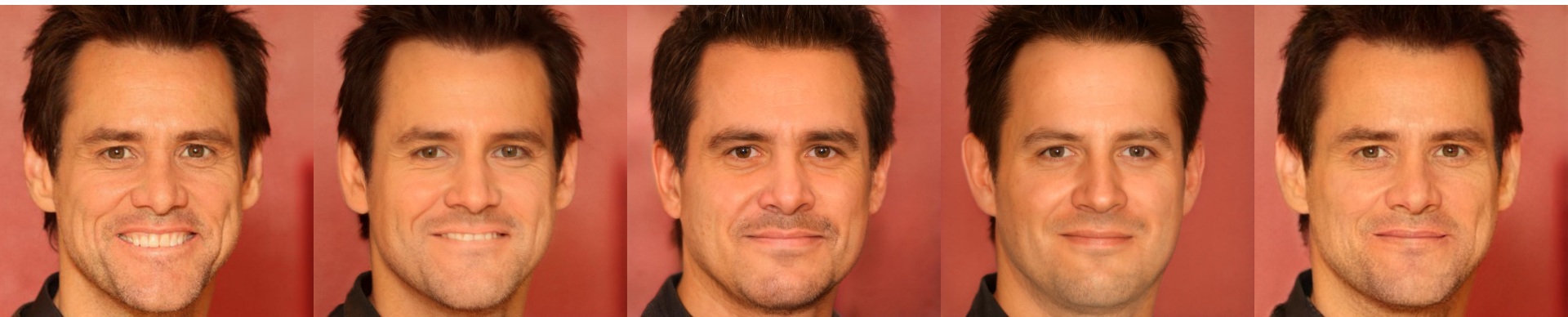
Original

W+ Optimization

e4e

W Optimization

PTI



Out of Distribution

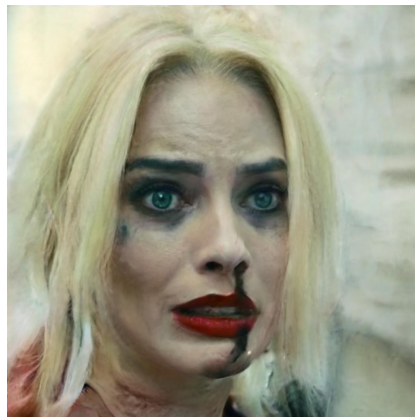
Original



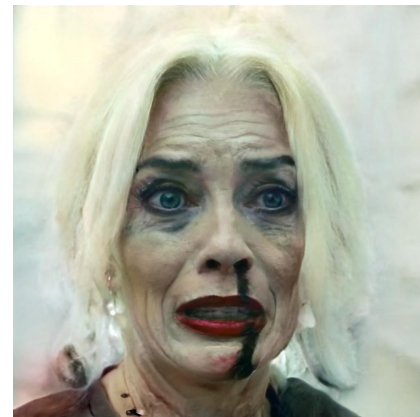
Smile



Pose



Age



Original

W+ Optimization

e4e

W Optimization

PTI



Original

StyleCLIP

StyleCLIP + PTI

Hair

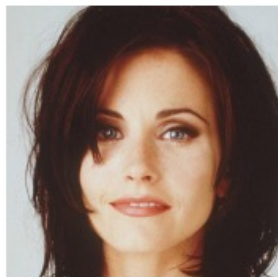


Hair+Age



Multiple Image Inversion

Locality Regularization – keep the tuning local.
Refer to the paper for more details





EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks. Chan et al.

Original

Inversion

Edit



StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets. Sauer *et al.*

Previous Works:

Image-adaptive GAN based reconstruction. Hussein *et al.*

Semantic photo manipulation with a generative image prior. Bau *et al.*

Exploiting deep generative prior for versatile image restoration and manipulation. Pan *et al.*

Our method:

- Simpler approach
- Maintain latent-based editing
- New inversion Standard for StyleGAN

Questions?



How long is this gonna take?

HyperStyle: StyleGAN Inversion with HyperNetworks for Real Image Editing

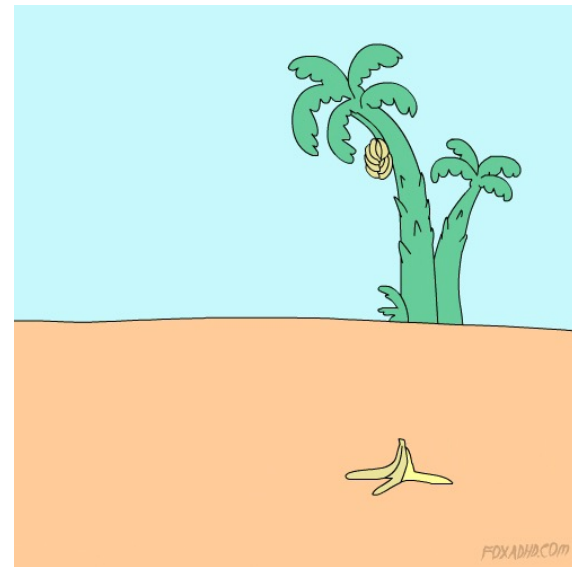
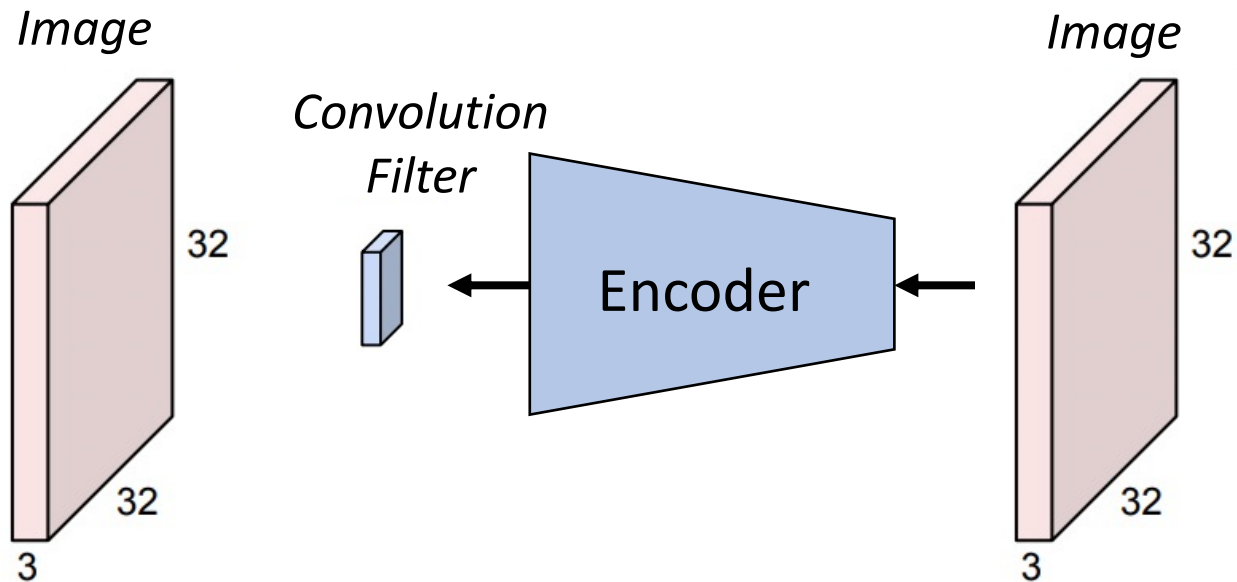
Yuval Alaluf, Omer Tov, **Ron Mokady**, Rinon Gal, Amit H. Bermano

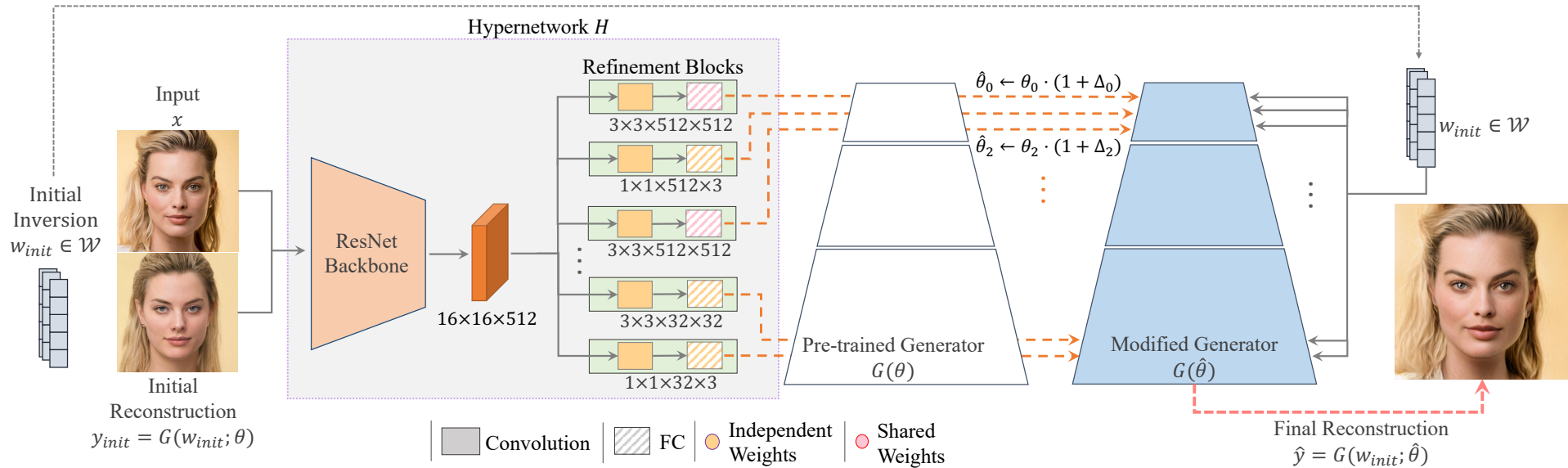
Concurrent work:

HyperInverter: Improving StyleGAN Inversion via Hypernetwork, Dinh *et al.*



Hyper Networks





The challenge: naïve approach requires 3B parameters.



Refer to the paper for more details

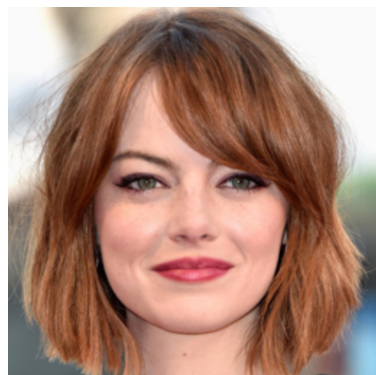
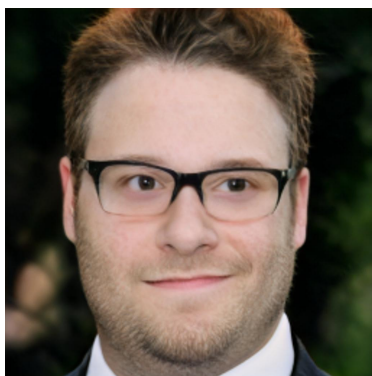
Original

Inversion

Pose

Smile

age



Original

HyperStyle

PTI

Smile



Frontalization

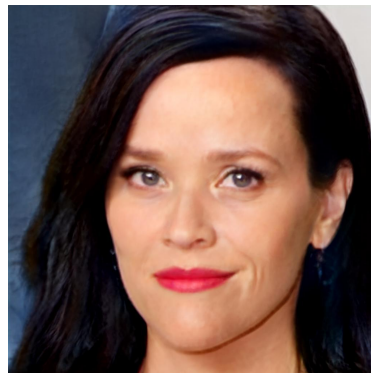
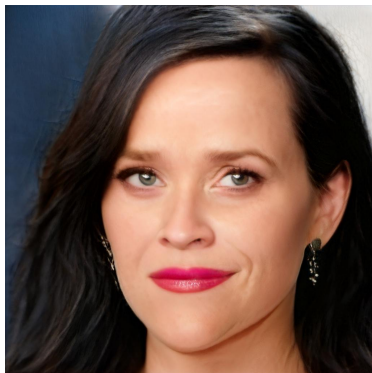


Original

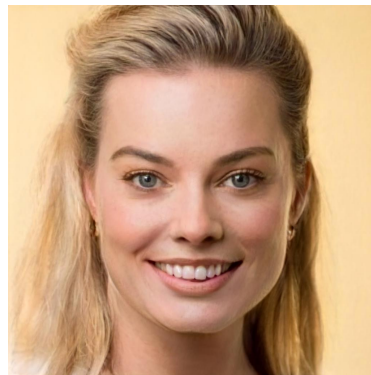
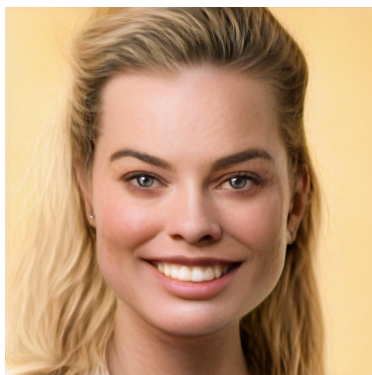
HyperStyle

PTI

Hair



Smile



Concurrent works: F space



Barbershop: GAN-based Image Compositing using Segmentation Masks. Zhu *et al.*

High-Fidelity GAN Inversion for Image Attribute Editing. Wang *et al.*

Questions?

What's next?



Stitch it in Time: GAN-Based Facial Editing of Real Videos

Rotem Tzaban, **Ron Mokady**, Rinon Gal, Amit H. Bermano, Daniel Cohen-Or



Problem: Temporal Coherence

Optimization per frame



PTI per frame

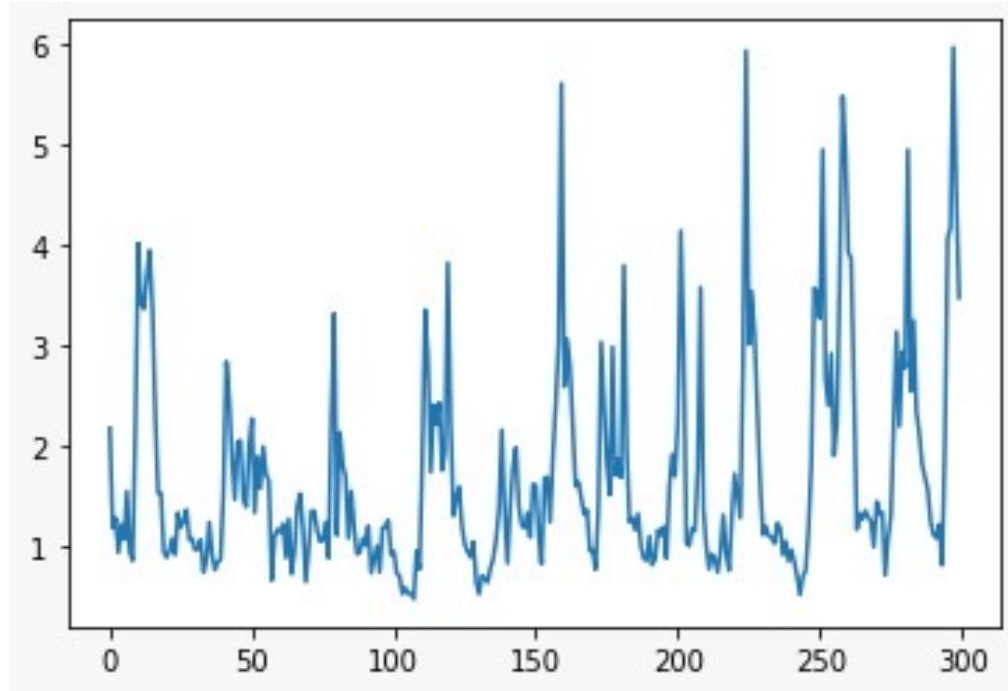


Popular approach: Optical Flow



Another idea: latent-based smoothing

*Distance to
next frame*



Frame

Observation: Input is already temporal coherent



Solution: using only consistent tools



Per-frame optimization



Encoder-based inversion

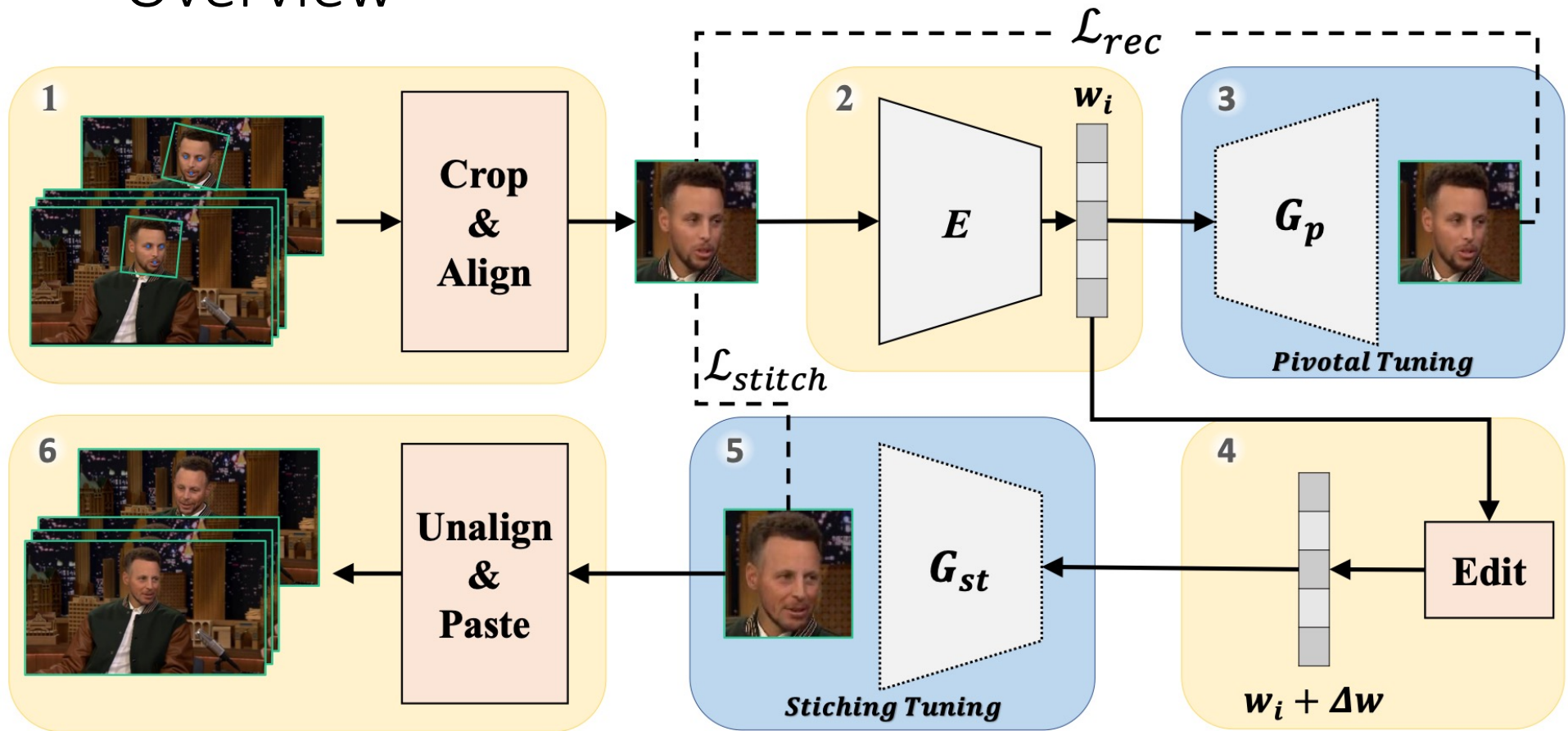


StyleGAN's latent-based editing

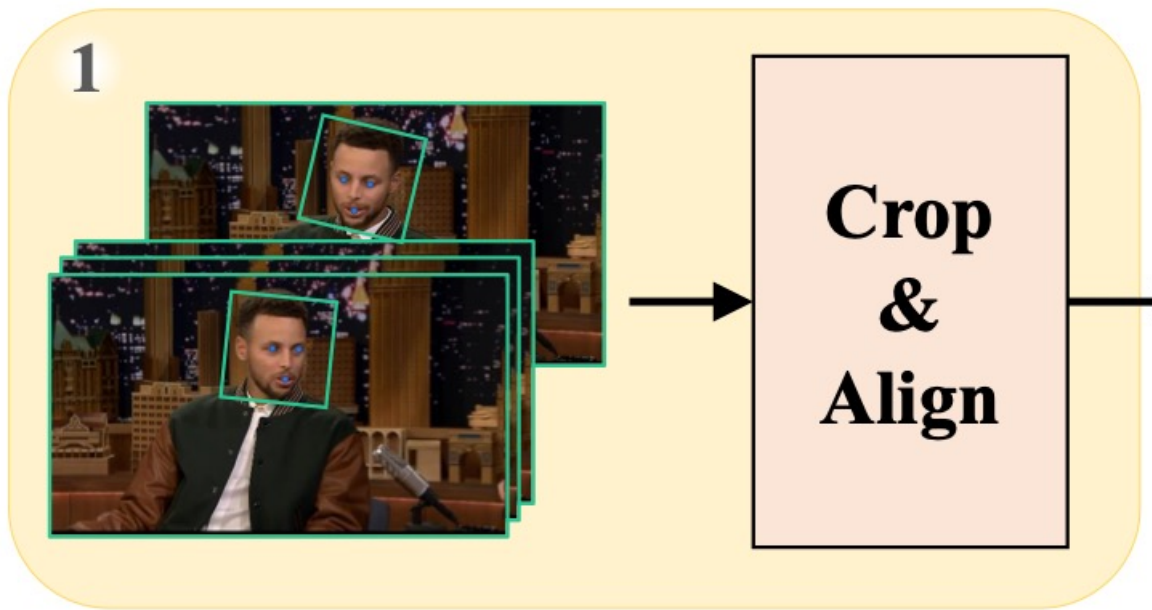


Fine-tuning StyleGAN

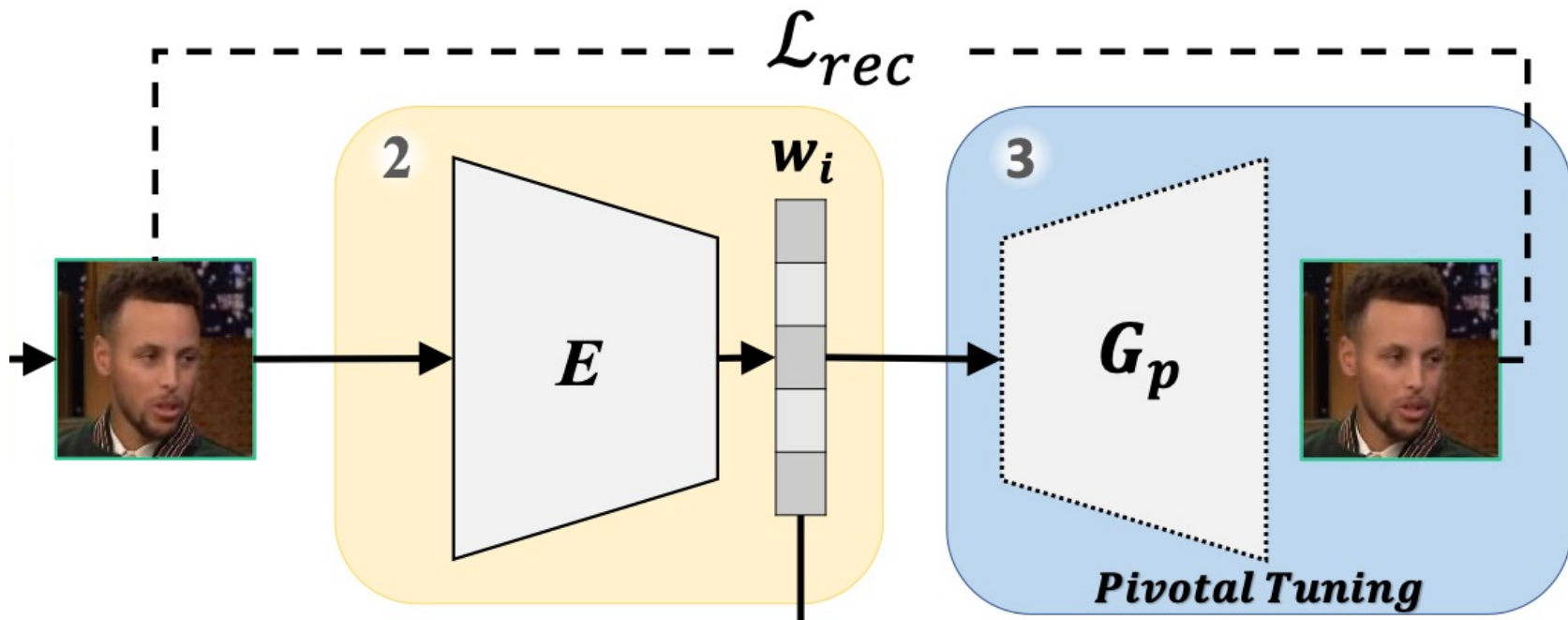
Overview



1. Cropping and alignment



2. Inversion using encoder (e4e)



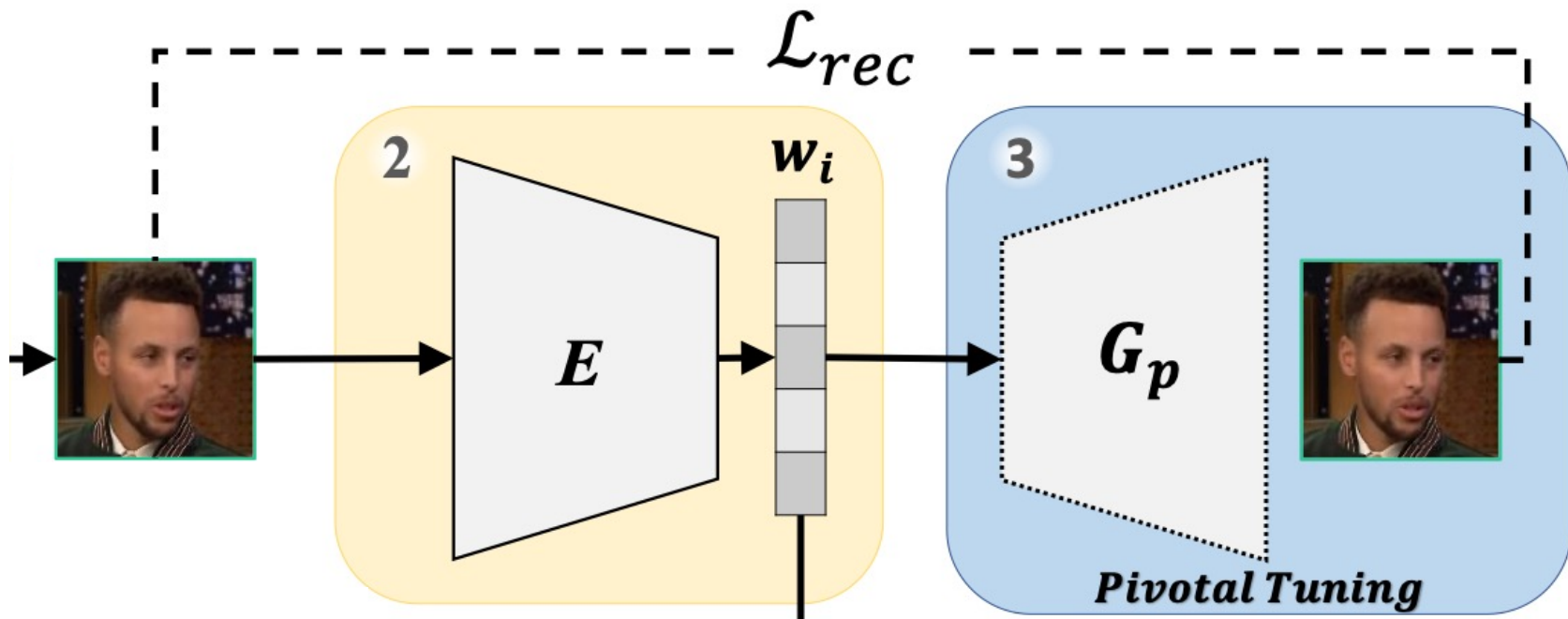
Original



e4e



3. PTI



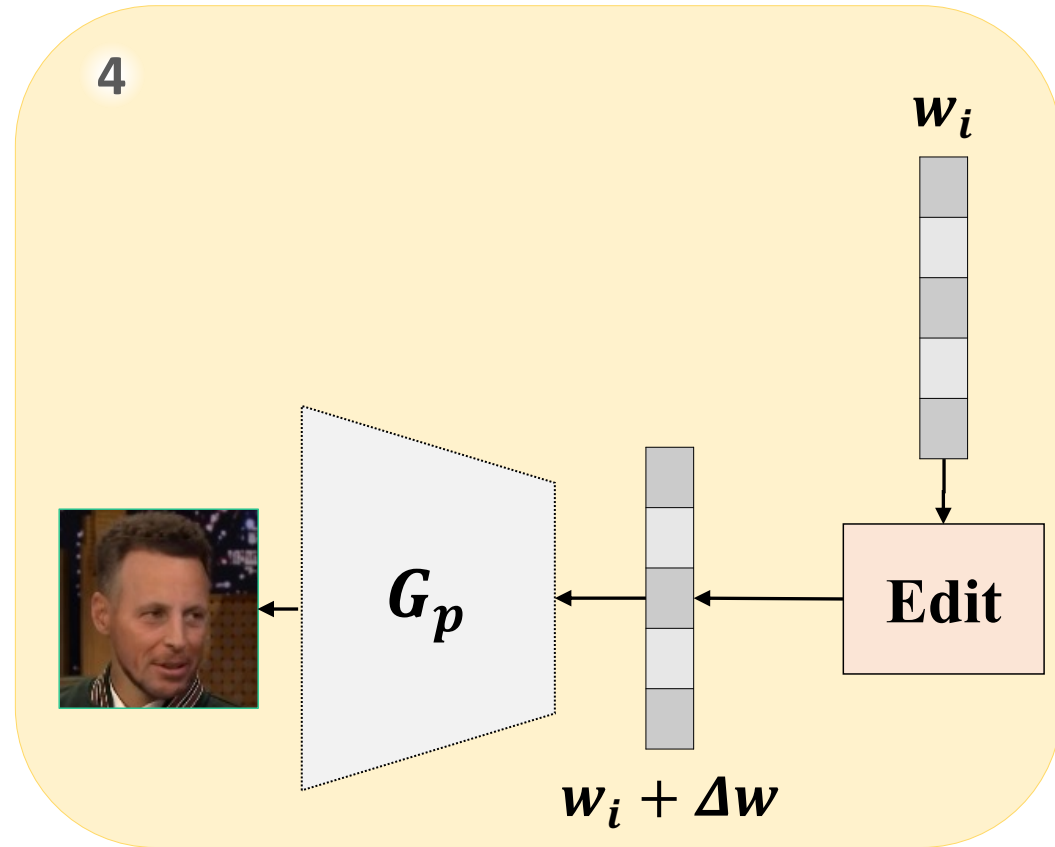
Original



PTI



4. Editing



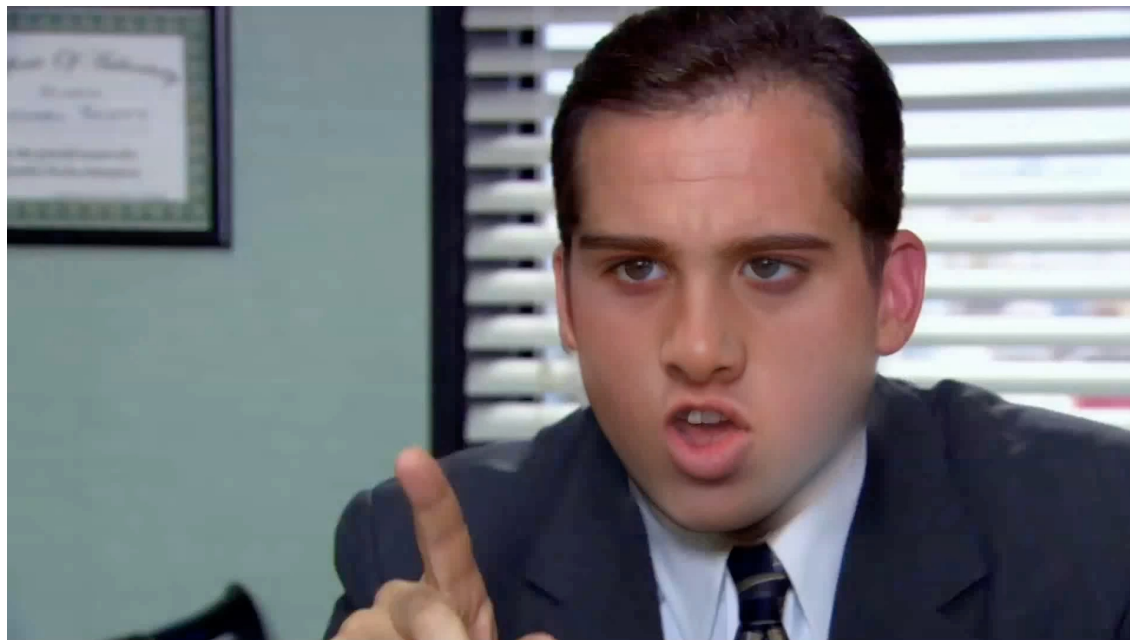
Problem: Blending

Original



Edit

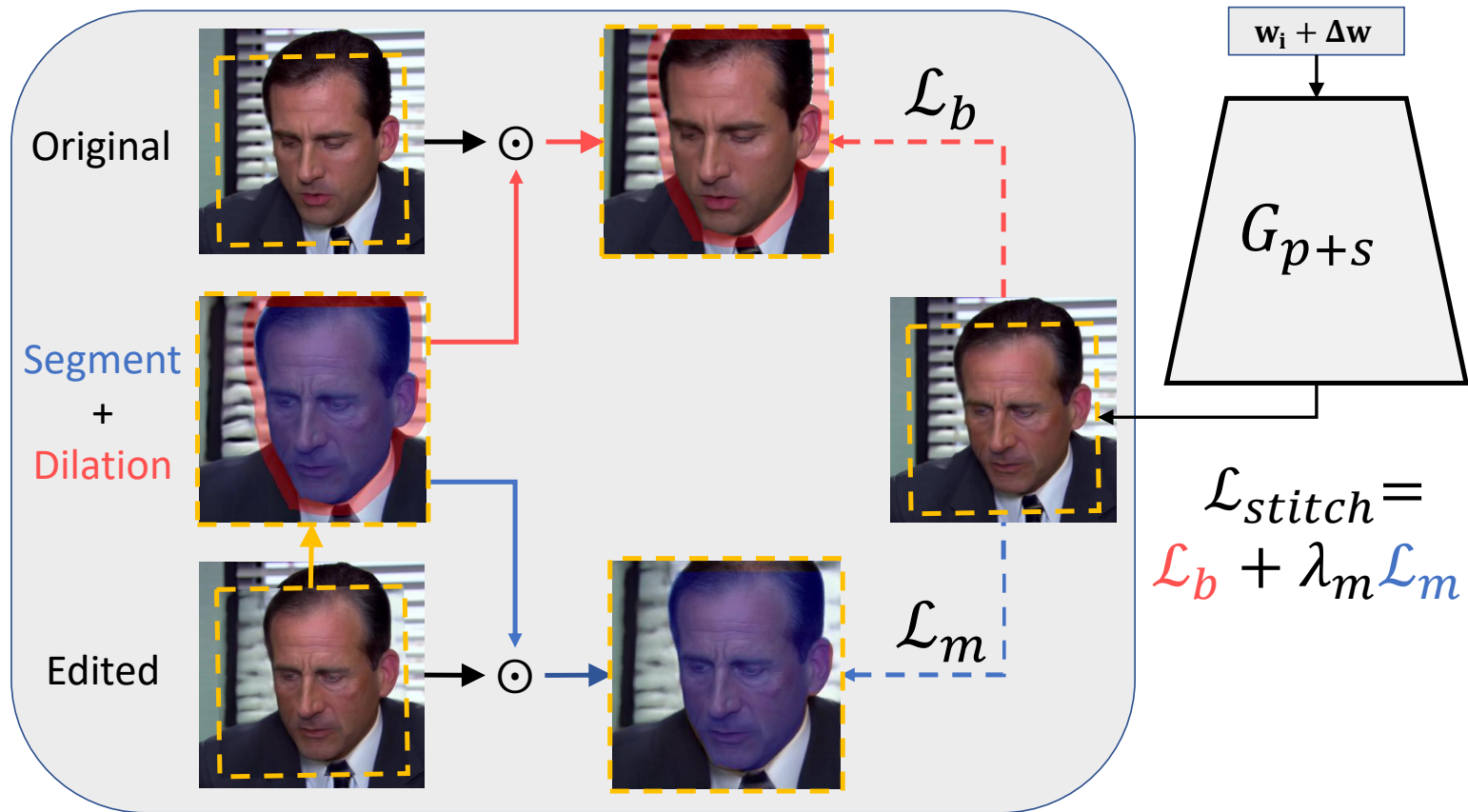




A latent transformer for disentangled face editing in images and videos. *Yao et al.*



Stitching Tuning



w/o Stitching Tuning

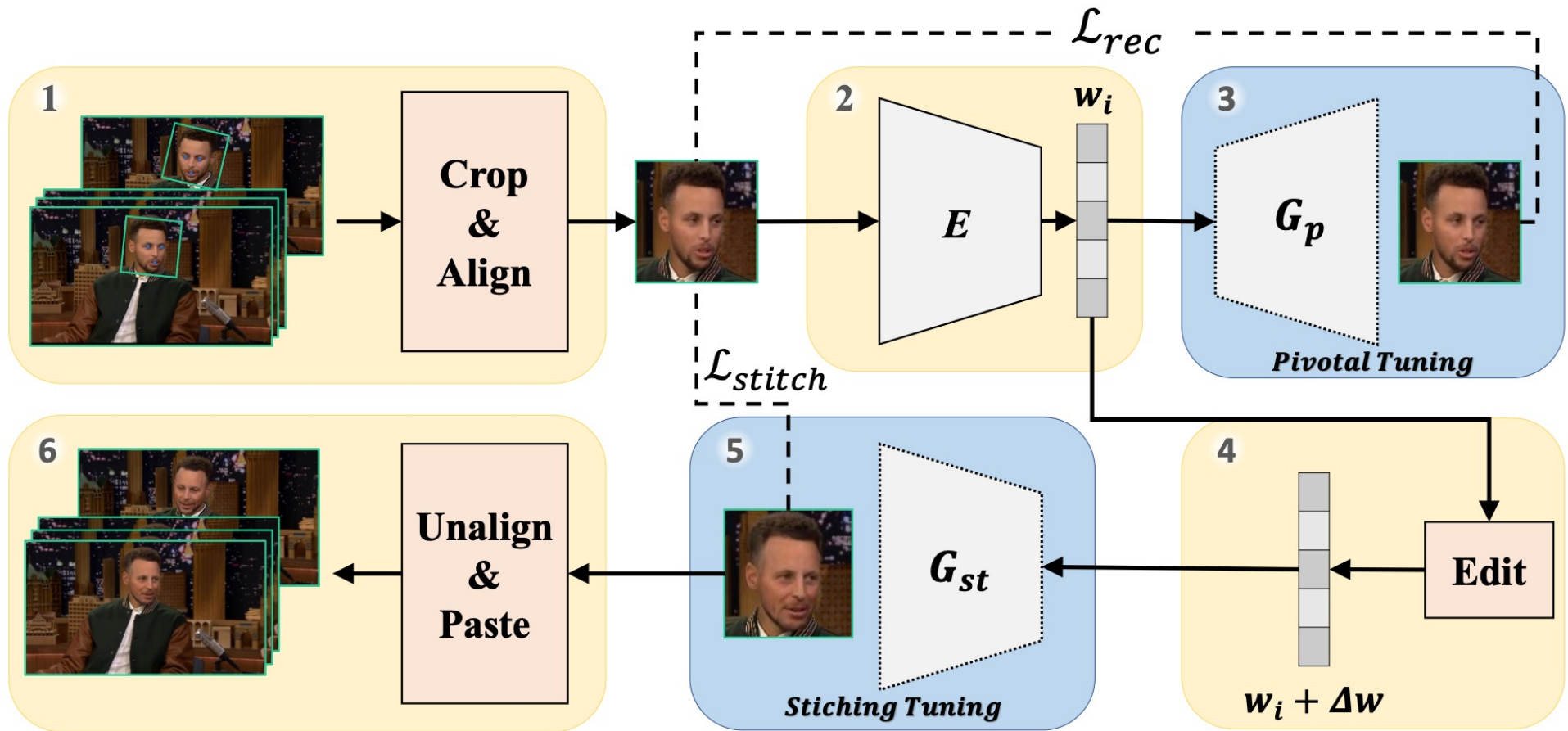


w/ Stitching Tuning



Recall: FFHQ





Original

Young



Original



Smile



Original



Angry



Young



Lipstick



Original



Smile

Gender



Old

Original



Old



Original



Smile

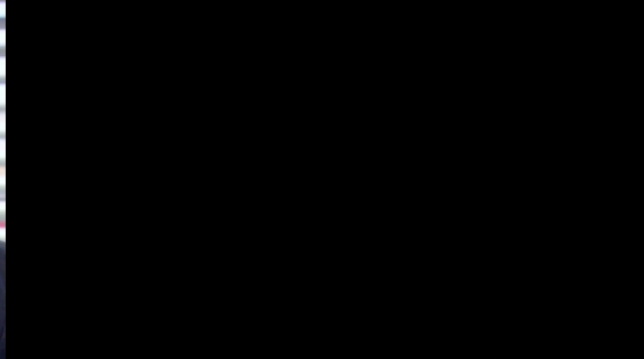
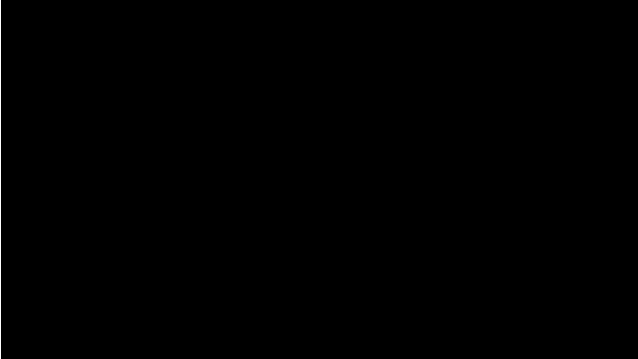


Original



Gender





Future Work

Replacing PTI with Encoder

Better editing technique instead of Sticking-Tuning

Better alignment scheme for StyleGAN

Questions?

Out-Of-Domain

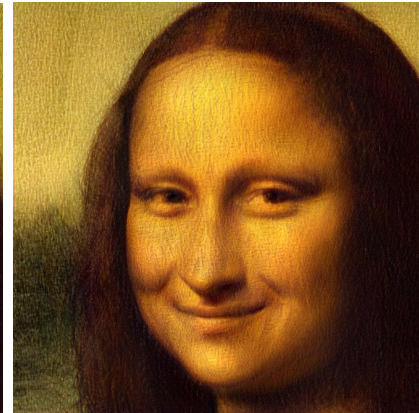
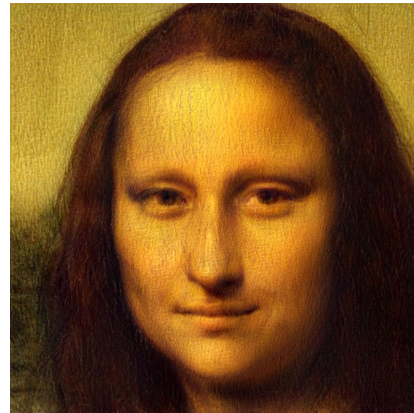
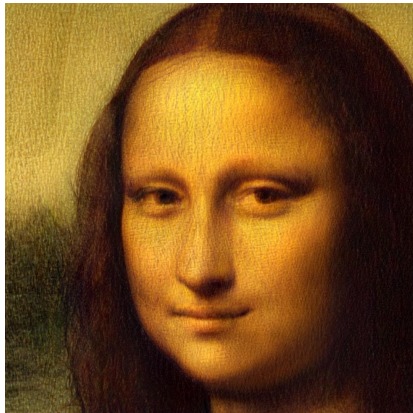
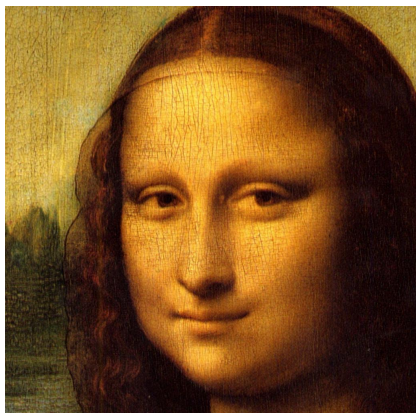
Original

Inversion

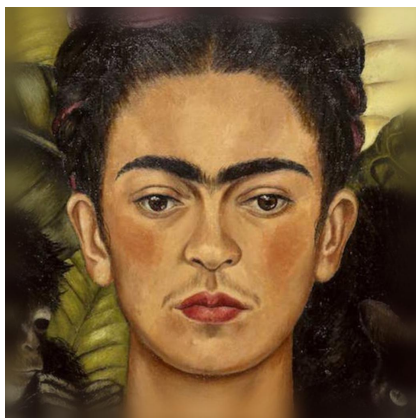
Pose

Smile

PTI



Hyper Style



Original

Inversion

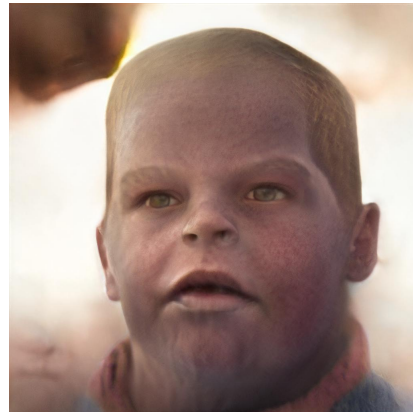
Age

Smile

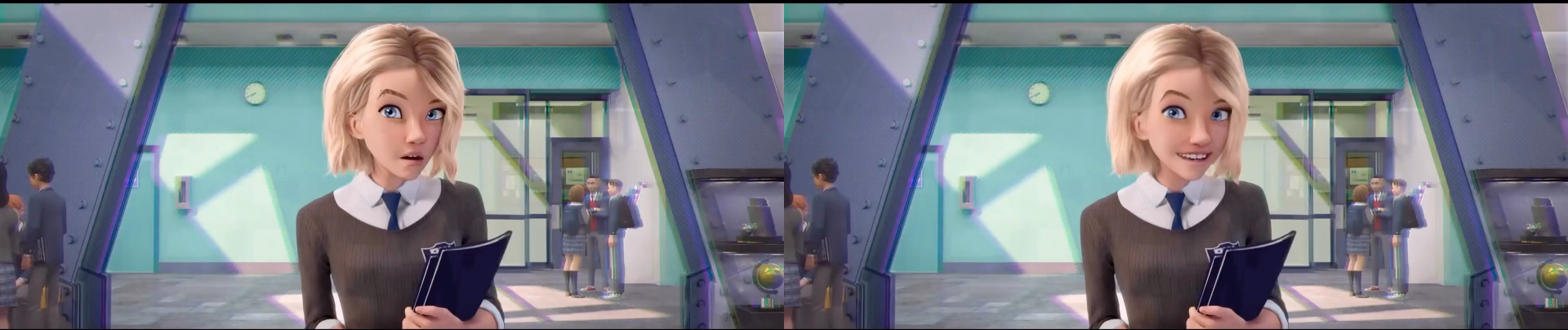
PTI



*Hyper
Style*



STIT



Different Domains

HyperStyle

Original

Inversion

Pose

no roof

cube

color



PTI

Original

Inversion

Pose

no roof

cube

color

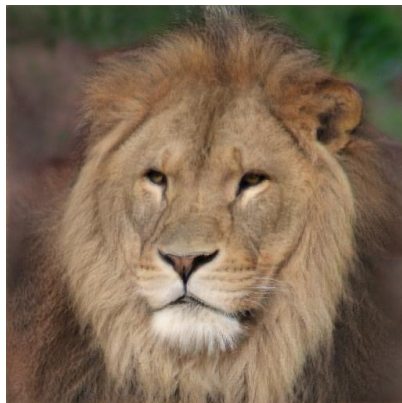


HyperStyle

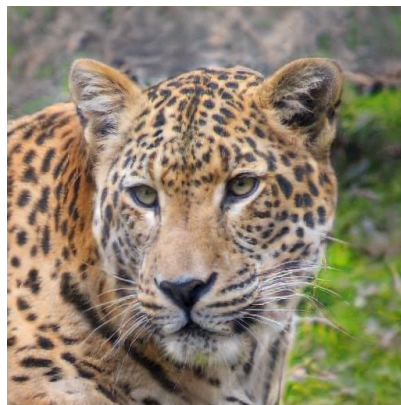
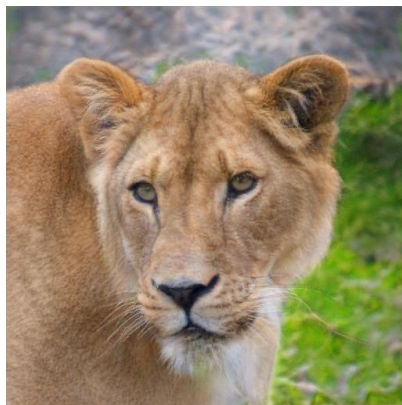
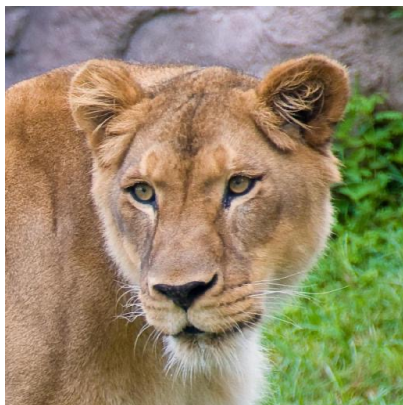
Original



Inversion



Edit



More challenging datasets?

Self-Distilled StyleGAN: Towards Generation from Internet Photos

Ron Mokady, Michal Yarom, Omer Tov, Oran Lang, Daniel Cohen-Or,
Tali Dekel, Michal Irani, Inbar Mosseri

Internet Photo Collection





StyleGAN Trained on Uncurated Internet Photos



Filtering

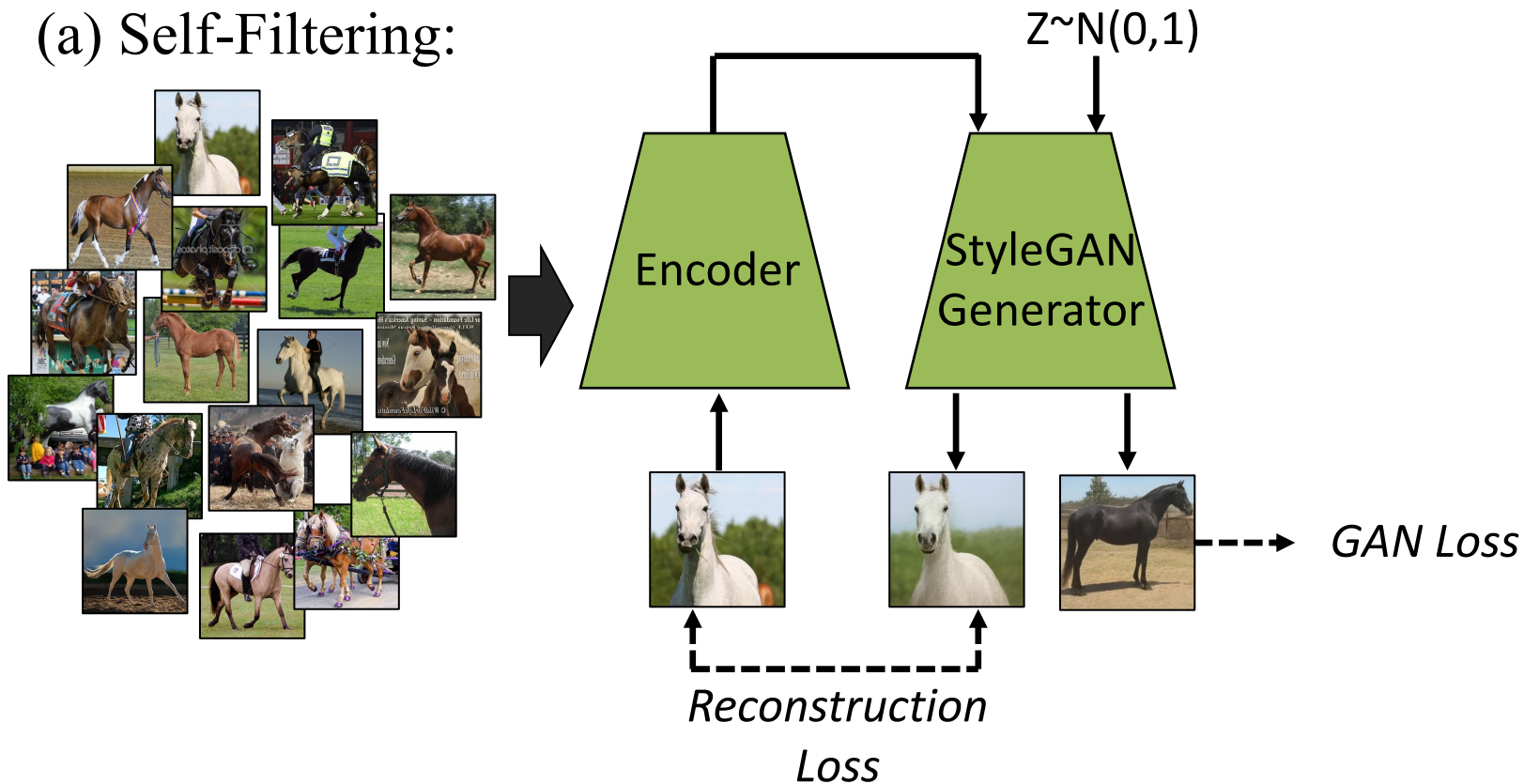
- Classifier
- Detection network
- Manually

Our solution:

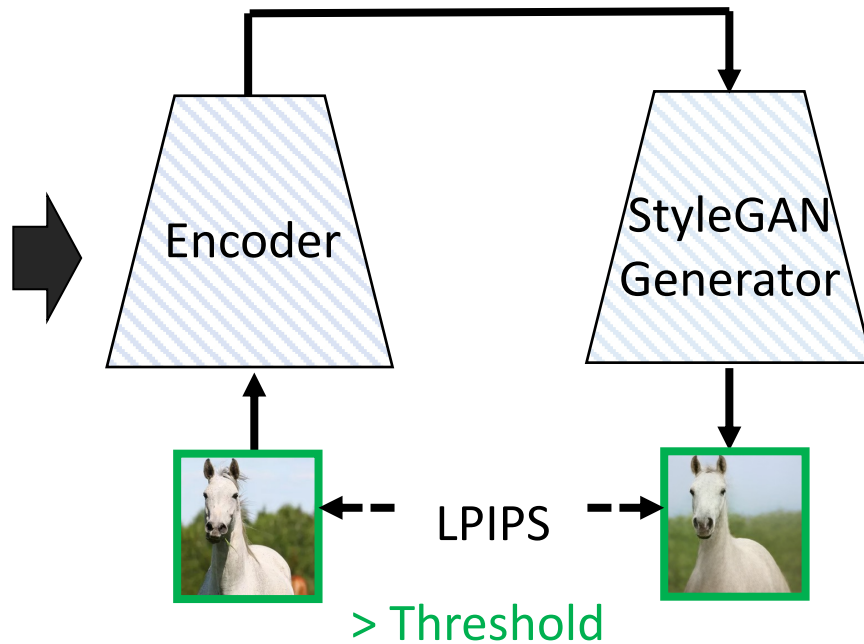
Let the generator itself filter the data

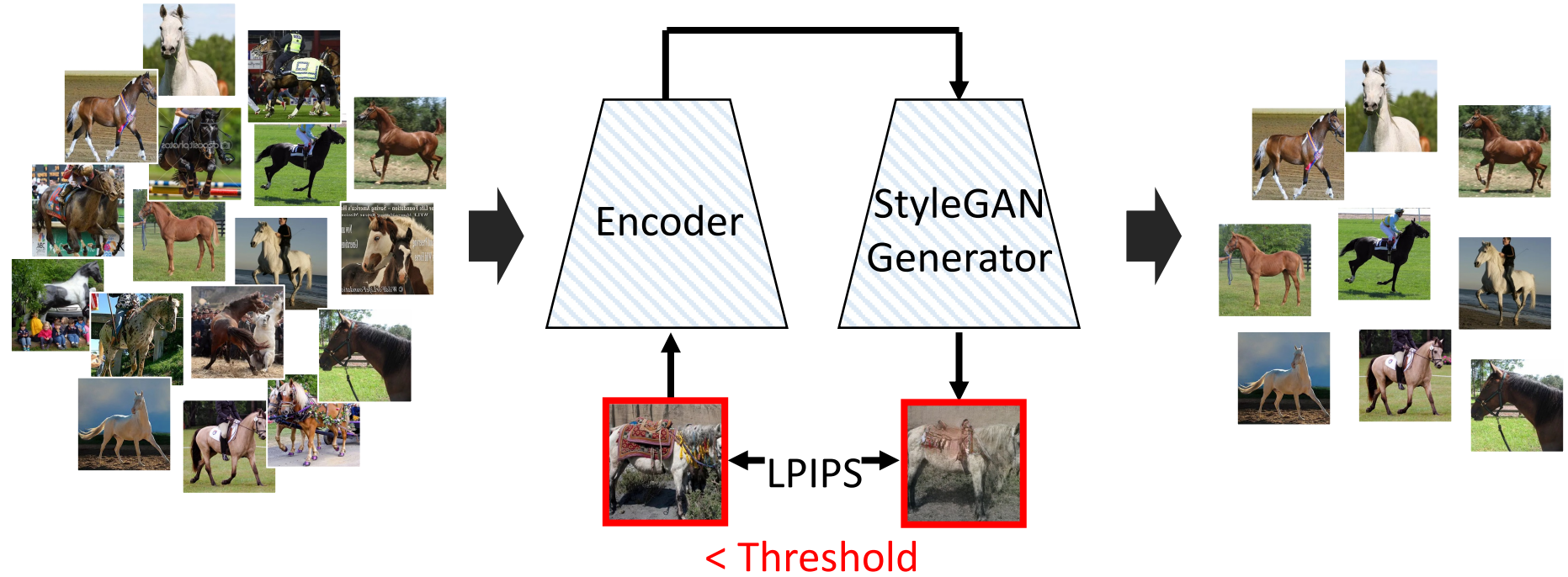
Self-Distilled StyleGAN

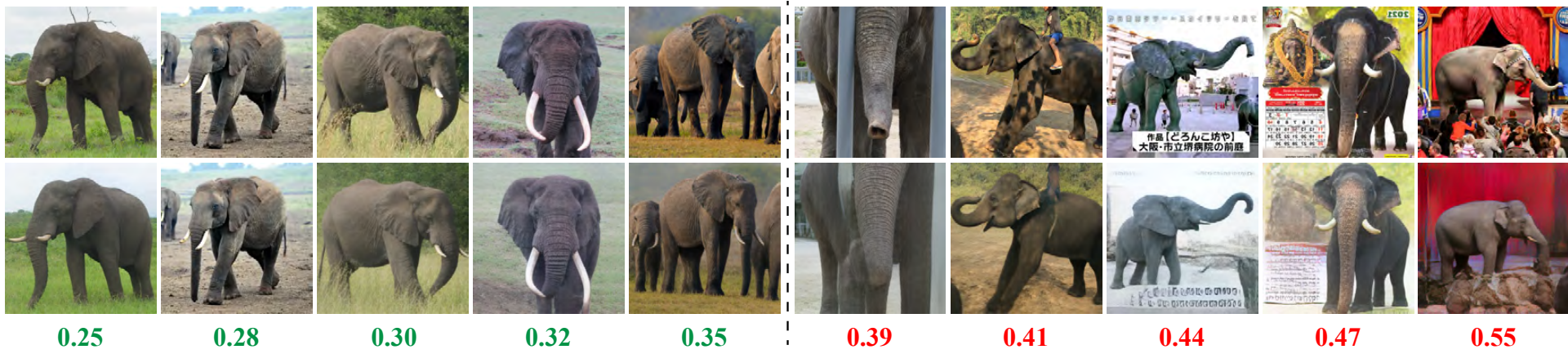
(a) Self-Filtering:



(a) Self-Filtering:







LPIPS

Questions?

Refer to the paper for more details

Truncation Trick: $w_t = \psi \cdot w + (1 - \psi) \cdot \bar{w}$



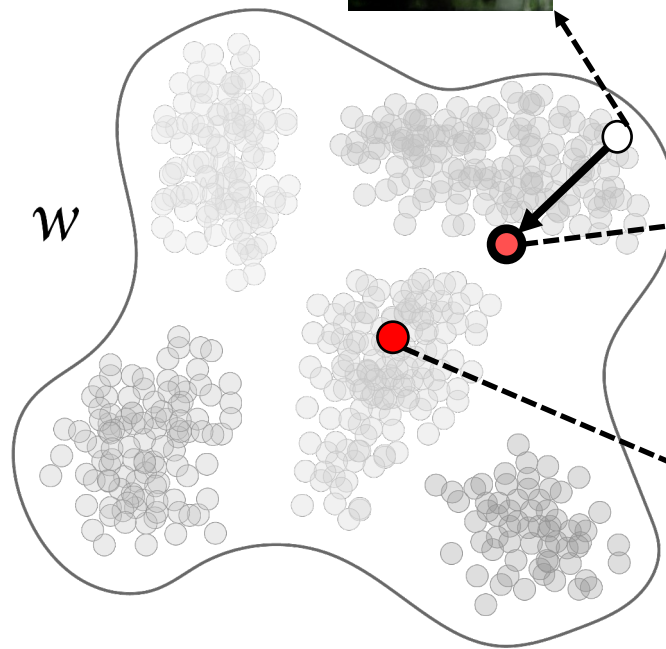
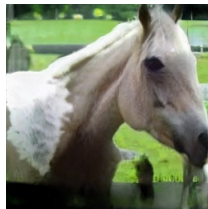
$\psi = 1$

$\psi = 0.7$

$\psi = 0.5$

$\psi = 0$

No
Truncation



Truncation to
Global Mean



Global Mean



(b) Multi-Modal Truncation:

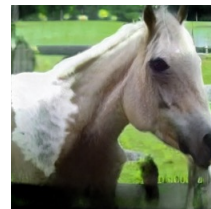
Cluster
Center



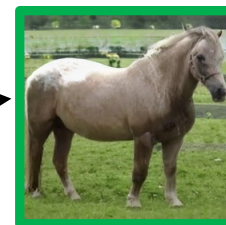
Truncation to
Cluster Center



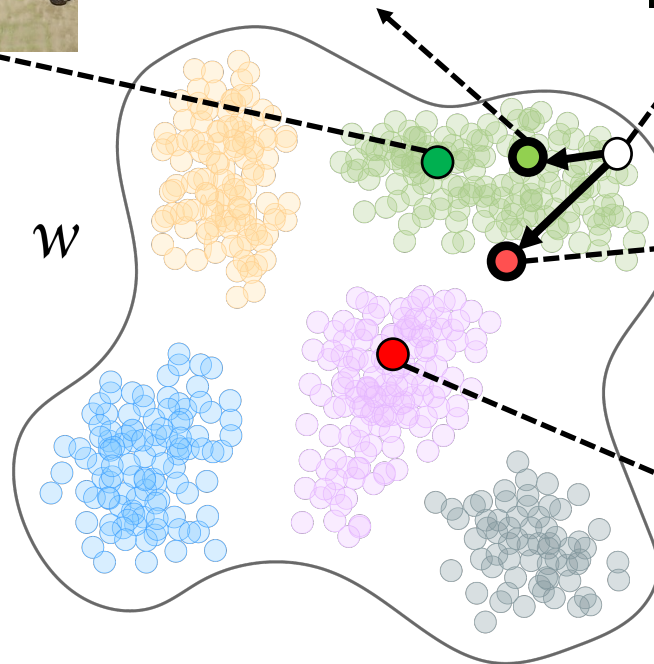
No
Truncation



Truncation to
Global Mean



Global Mean





(a) No Truncation



(b) Truncation to Global Mean



(c) Truncation to Cluster (Ours)

Multi-Modal Truncation – Cluster Centers

Dogs



Lions



Parrots



Plants



Results

Parrots



Elephants



Dogs





Lions



Bicycle



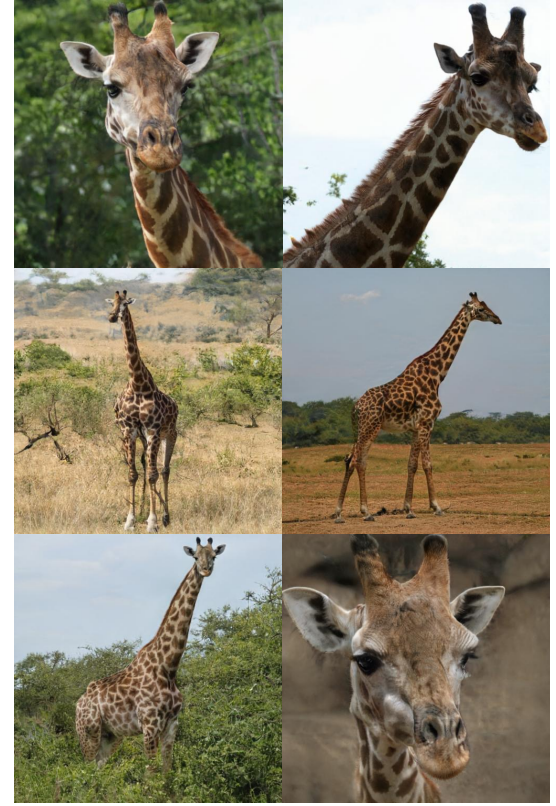
Potted Plants



Horses



Giraffes



Semantic Editing

Lions – Add/Remove Mane



Lions – Head turn



Lions – Roar



Parrots - Head pan



Horses

Head Pan



Running



Conclusions (Self-Distillation StyleGAN)

StyleGAN operate well on non-aligned data

Data quality is important

New Dataset! (Dogs, Elephants, Horses, Bicycles)

New Models! (Lions, Parrots, Dogs, Elephants, Horses, Bicycles).

Future Work (Self-Distillation StyleGAN)

StyleGAN for Human full-body

What is required for new editing directions?

Questions?