SYNTHESIZING 3D CONSISTENT HUMAN HEADS VIA 3DGS GANS

MLaftermath Workshop

Florian Barthel

1

Motivation

2

3DGS GAN

3

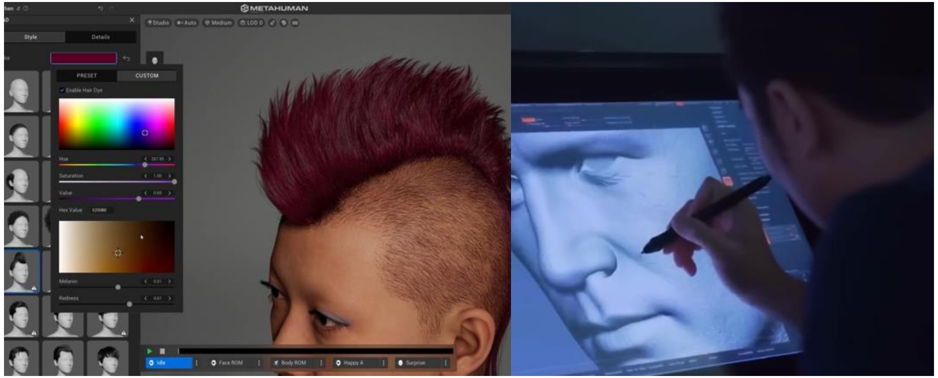
FFHQ Clean

4

Future Work

Motivation

3D Video Calls (Apple Persona)



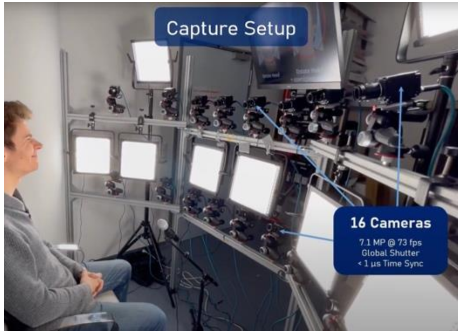
Movie / Gaming Industry

Education

Psychology Experiments

Method Selection

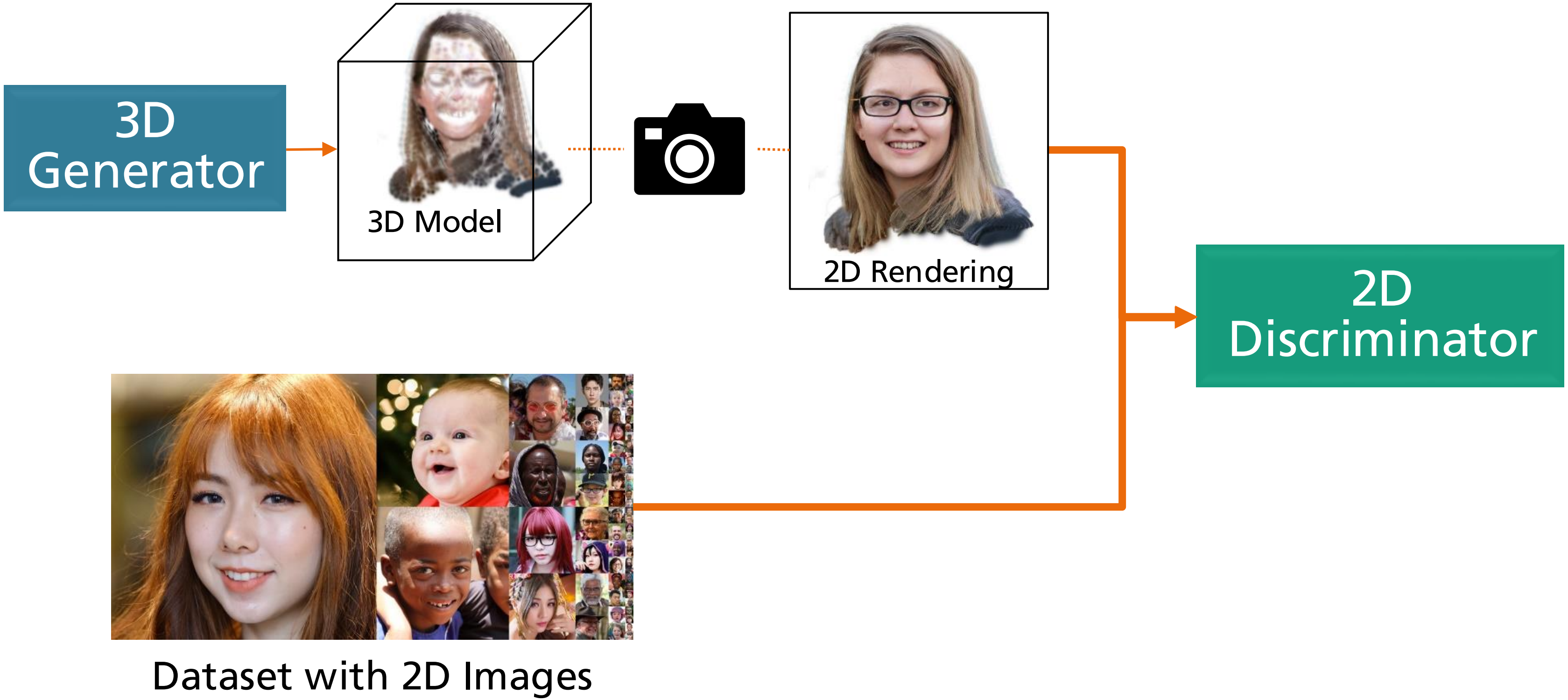
3D Reconstruction



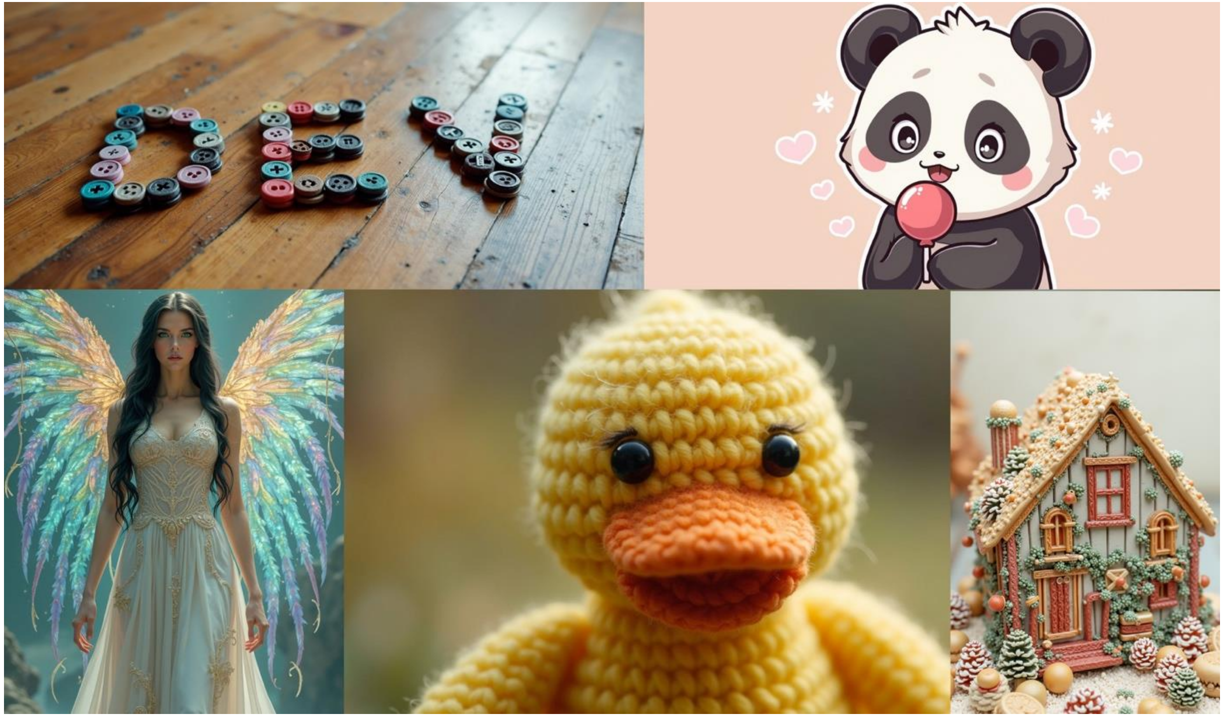
Generative Model

latent → Generator →

3D GANs

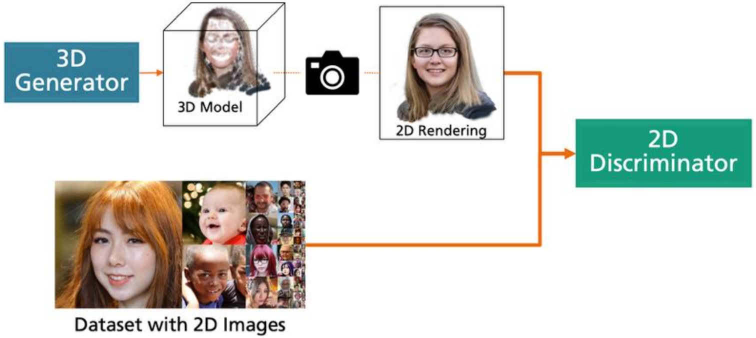

3D Generator
3D Model
2D Rendering
2D Discriminator
Dataset with 2D Images

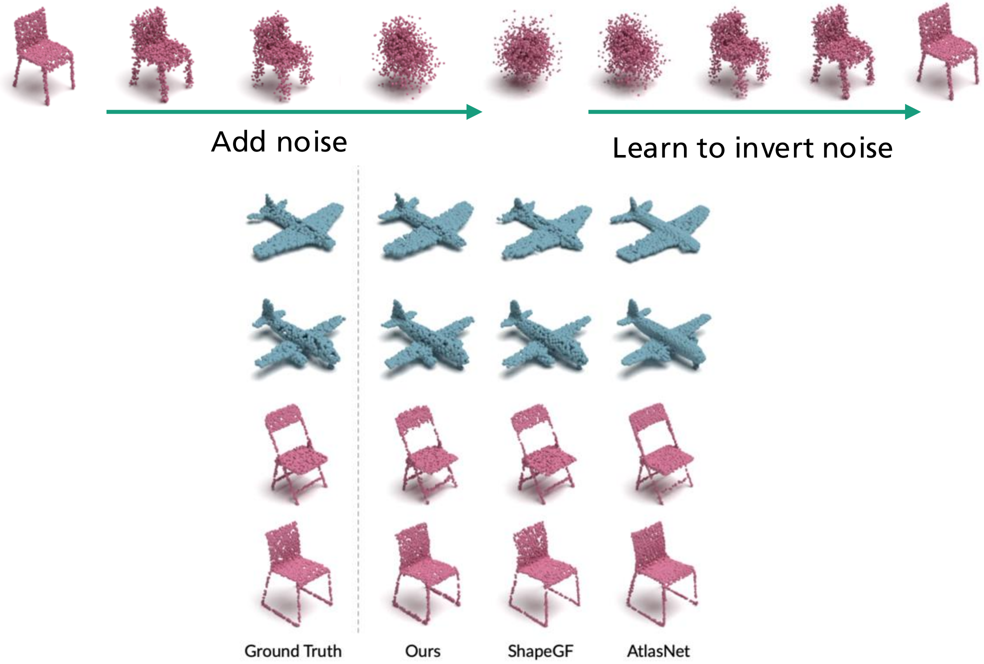
Why not Diffusion Models?



FLUX.1 [dev]

3D GANs vs 3D Diffusion Models





Learns 3D from **2D**

Learns 3D from **3D**

3DGS

- Scene build from a set of 3D Gaussian primitives
- Each 3D Gaussian primitive is described by:
  - Position
  - Color
  - Opacity
  - Scale
  - Rotation

Florian Barthel

3DGS Example

Florian Barthel

Fraunhofer HHI

1

Motivation

2

3DGS GAN

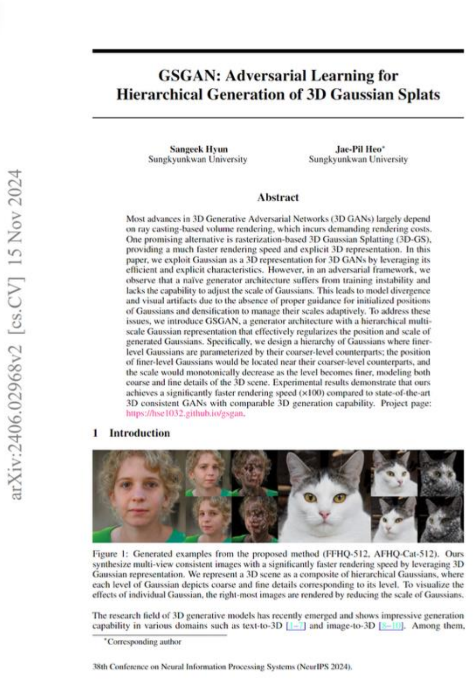
3

FFHQ Clean

4

Future Work

Existing 3DGS GANs



arXiv:2406.02968v2 [cs.CV] 15 Nov 2024

GSGAN: Adversarial Learning for Hierarchical Generation of 3D Gaussian Splats

Sangeek Hyun
Sungkyunkwan University

Jae-Pil Heo*
Sungkyunkwan University

Abstract

Most advances in 3D Generative Adversarial Networks (3D GANs) largely depend on ray casting-based volume rendering, which incurs demanding rendering costs. One promising alternative is rasterization-based 3D Gaussian Splatting (3D-GS), providing a much faster rendering speed and explicit 3D representation. In this paper, we exploit Gaussian as a 3D representation for 3D GANs by leveraging its efficient and explicit characteristics. However, in an adversarial framework, we observe that a naïve generator architecture suffers from training instability and lacks the capability to adjust the scale of Gaussians. This leads to model divergence and visual artifacts due to the absence of proper guidance for initialized positions of Gaussians and densification to manage their scales adaptively. To address these issues, we introduce GSGAN, a generator architecture with a hierarchical multi-scale Gaussian representation that effectively regularizes the position and scale of generated Gaussians. Specifically, we design a hierarchy of Gaussians where finer-level Gaussians are parameterized by their coarser-level counterparts; the position of finer-level Gaussians would be located near their coarser-level counterparts, and the scale would monotonically decrease as the level becomes finer, modeling both coarse and fine details of the 3D scene. Experimental results demonstrate that ours achieves a significantly faster rendering speed (×100) compared to state-of-the-art 3D consistent GANs with comparable 3D generation capability. Project page: https://hse1032.github.io/gsgan.

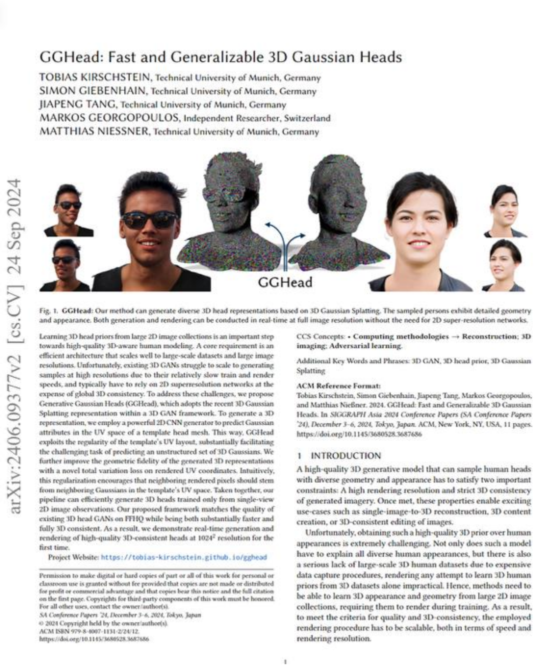
1 Introduction

Figure 1: Generated examples from the proposed method (FFHQ-512, AFHQ-Cat-512). Ours synthesize multi-view consistent images with a significantly faster rendering speed by leveraging 3D Gaussian representation. We represent a 3D scene as a composite of hierarchical Gaussians, where each level of Gaussian depicts coarse and fine details corresponding to its level. To visualize the effects of individual Gaussian, the right-most images are rendered by reducing the scale of Gaussians.

The research field of 3D generative models has recently emerged and shows impressive generation capability in various domains such as text-to-3D [1–7] and image-to-3D [8–10]. Among them,

*Corresponding author

38th Conference on Neural Information Processing Systems (NeurIPS 2024).



arXiv:2406.09377v2 [cs.CV] 24 Sep 2024

GGHead: Fast and Generalizable 3D Gaussian Heads

TOBIAS KIRSCHSTEIN, Technical University of Munich, Germany
SIMON GIEBENHAIN, Technical University of Munich, Germany
JIAPENG TANG, Technical University of Munich, Germany
MARKOS GEORGOPOULOS, Independent Researcher, Switzerland
MATTHIAS NIESSNER, Technical University of Munich, Germany

Fig. 1. **GGHead**: Our method can generate diverse 3D head representations based on 3D Gaussian Splatting. The sampled persons exhibit detailed geometry and appearance. Both generation and rendering can be conducted in real-time at full image resolution without the need for 2D super-resolution networks.

Learning 3D head priors from large 2D image collections is an important step towards high-quality 3D-aware human modeling. A core requirement is an efficient architecture that scales well to large-scale datasets and large image resolutions. Unfortunately, existing 3D GANs struggle to scale to generating samples at high resolutions due to their relatively slow train and render speeds, and typically have to rely on 2D superresolution networks at the expense of global 3D consistency. To address these challenges, we propose Generative Gaussian Heads (GGHead), which adopts the recent 3D Gaussian Splatting representation within a 3D GAN framework. To generate a 3D representation, we employ a powerful 2D CNN generator to predict Gaussian attributes in the UV space of a template head mesh. This way, GGHead exploits the regularity of the template's UV layout, substantially facilitating the challenging task of predicting an unstructured set of 3D Gaussians. We further improve the geometric fidelity of the generated 3D representations with a novel total variation loss on rendered UV coordinates. Intuitively, this regularization encourages that neighboring rendered pixels should stem from neighboring Gaussians in the template's UV space. Taken together, our pipeline can efficiently generate 3D heads trained only from single-view 2D image observations. Our proposed framework matches the quality of existing 3D head GANs on FFHQ while being both substantially faster and fully 3D consistent. As a result, we demonstrate real-time generation and rendering of high-quality 3D-consistent heads at 1024^2 resolution for the first time.

Project Website: https://tobias-kirschstein.github.io/gghead

SA Conference Papers '24, December 3–6, 2024, Tokyo, Japan

ACM ISBN 979-8-4007-1131-2/24/12.
https://doi.org/10.1145/3680528.3687686

CCS Concepts: • **Computing methodologies → Reconstruction; 3D imaging; Adversarial learning.**

Additional Key Words and Phrases: 3D GAN, 3D head prior, 3D Gaussian Splatting

ACM Reference Format:
Tobias Kirschstein, Simon Giebenhain, Jiapeng Tang, Markos Georgopoulos, and Matthias Nießner. 2024. GGHead: Fast and Generalizable 3D Gaussian Heads. In *SIGGRAPH Asia 2024 Conference Papers (SA Conference Papers '24), December 3–6, 2024, Tokyo, Japan.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3680528.3687686

1 INTRODUCTION

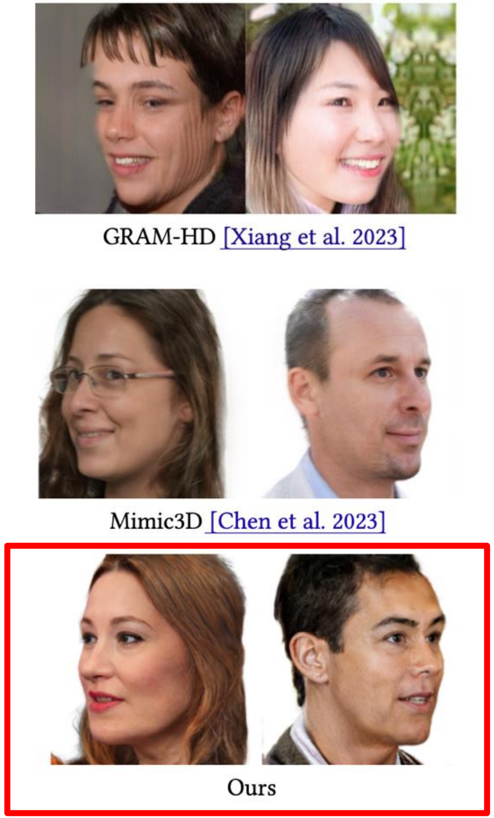
A high-quality 3D generative model that can sample human heads with diverse geometry and appearance has to satisfy two important constraints: A high rendering resolution and strict 3D consistency of generated imagery. Once met, these properties enable exciting use-cases such as single-image-to-3D reconstruction, 3D content creation, or 3D-consistent editing of images.

Unfortunately, obtaining such a high-quality 3D prior over human appearances is extremely challenging. Not only does such a model have to explain all diverse human appearances, but there is also a serious lack of large-scale 3D human datasets due to expensive data capture procedures, rendering any attempt to learn 3D human priors from 3D datasets alone impractical. Hence, methods need to be able to learn 3D appearance and geometry from large 2D image collections, requiring them to render during training. As a result, to meet the criteria for quality and 3D-consistency, the employed rendering procedure has to be scalable, both in terms of speed and rendering resolution.

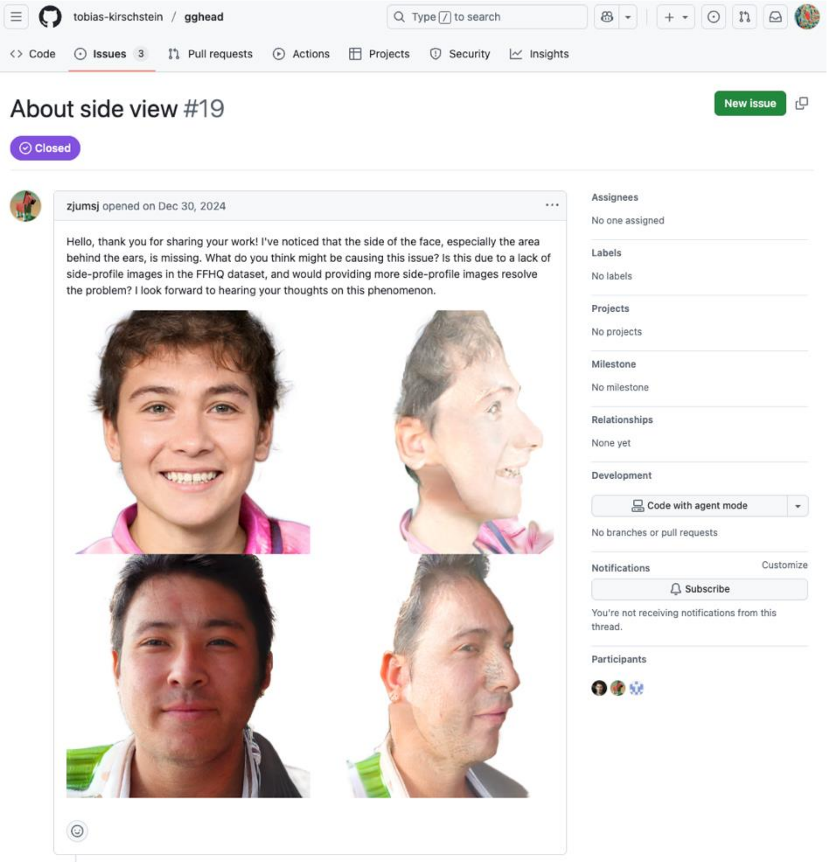
1

Florian Barthel

Fraunhofer HHI

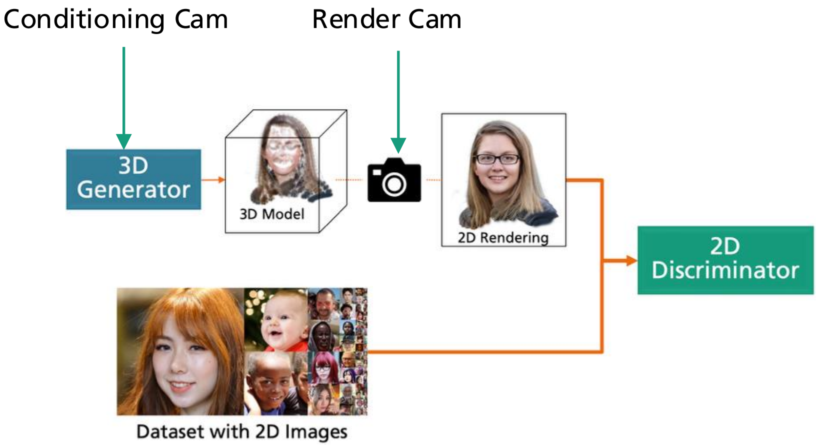


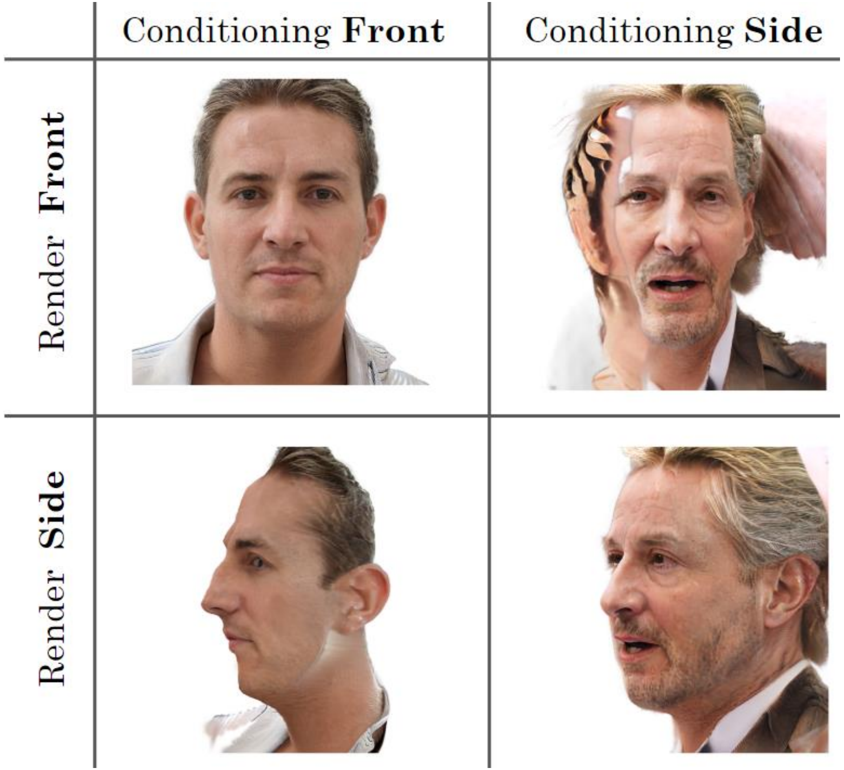
Paper Results



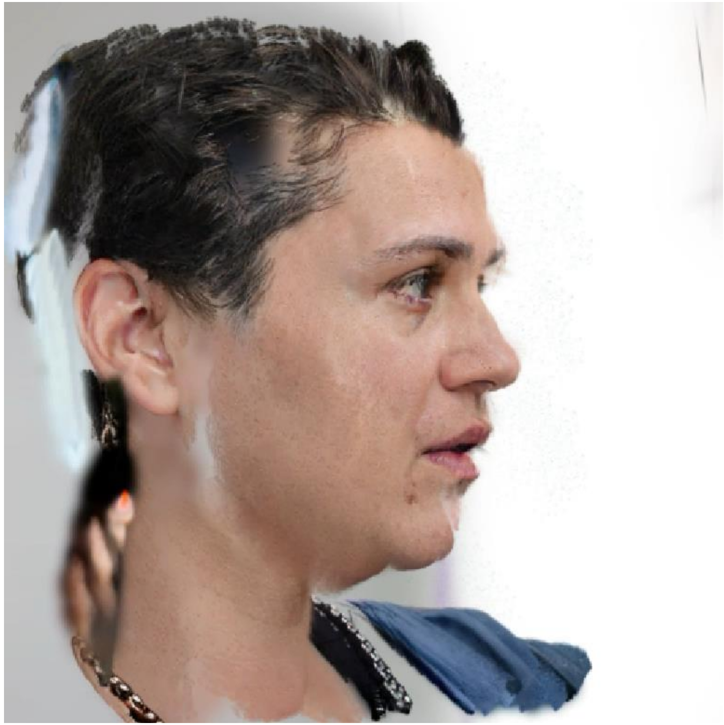
Reality

Camera Conditioning

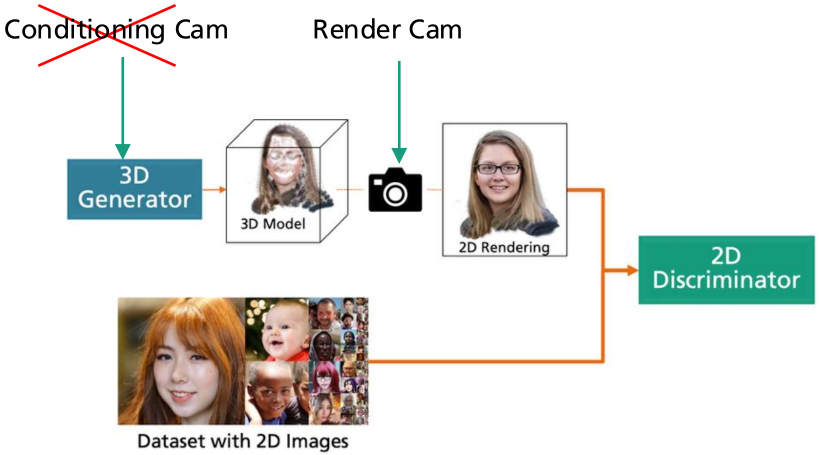




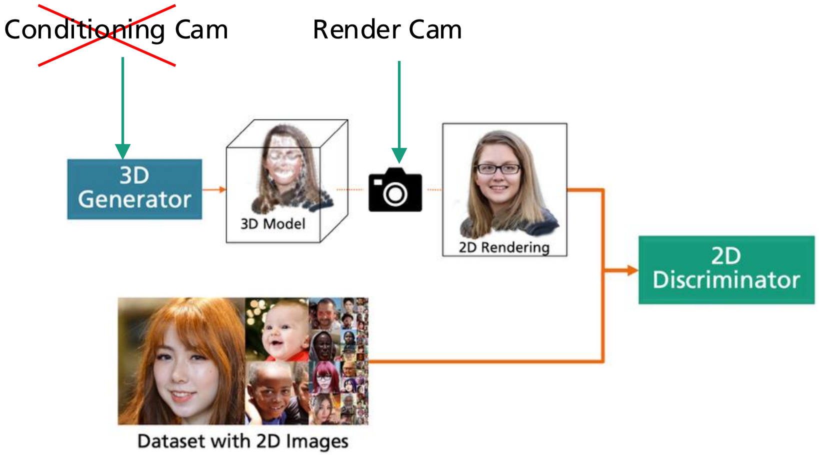
GSGAN

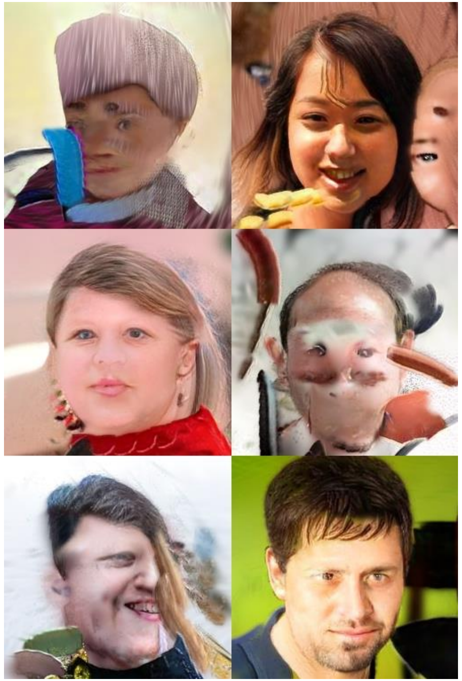


Remove View-Conditioning

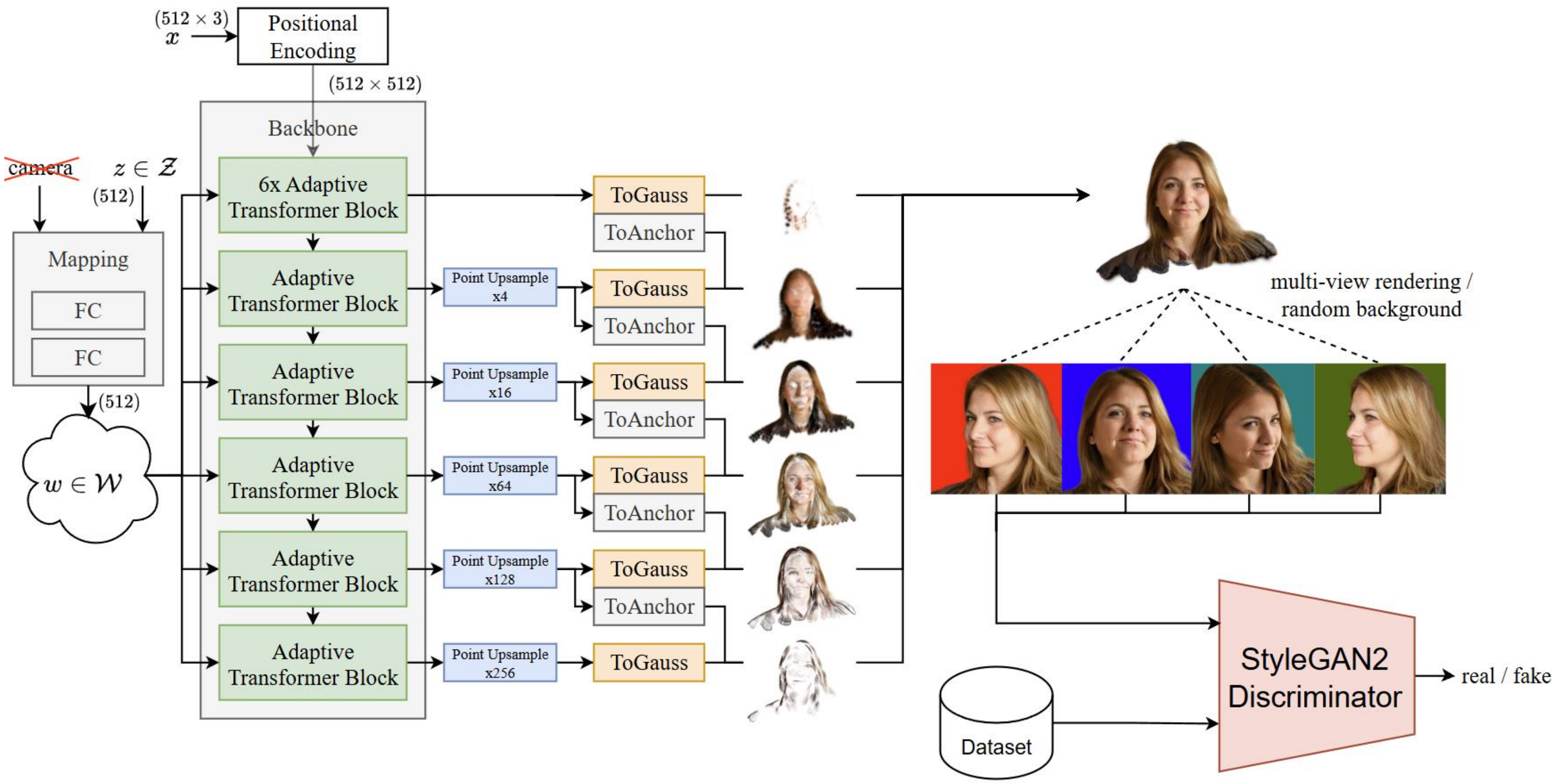


Remove View-Conditioning





CGS-GAN



Florian Barthel

Fraunhofer HHI

Comparison to GSGAN

Florian Barthel

Fraunhofer HHI

1

Motivation

2

3DGS GAN

3

FFHQ Clean

4

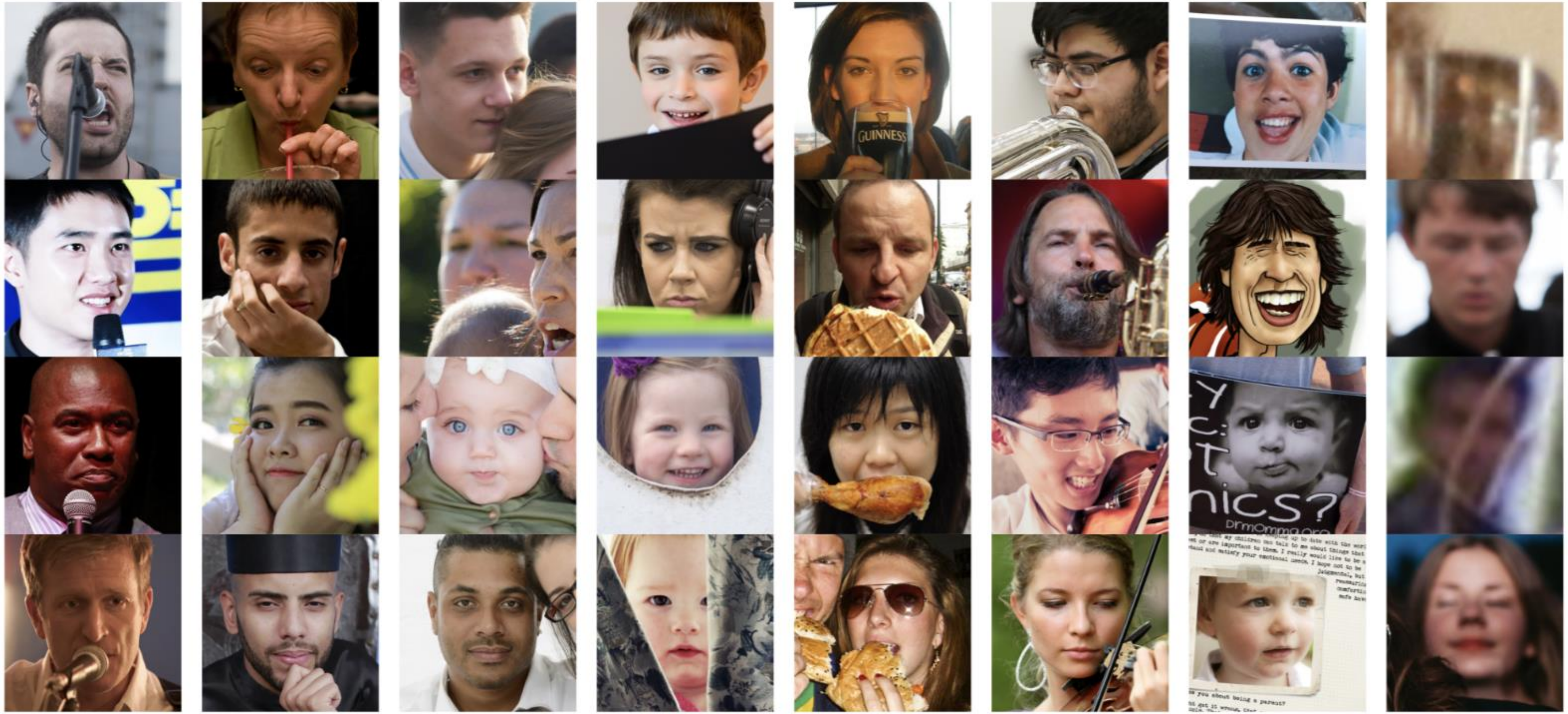
Future Work

Florian Barthel

Bad Renderings

Florian Barthel

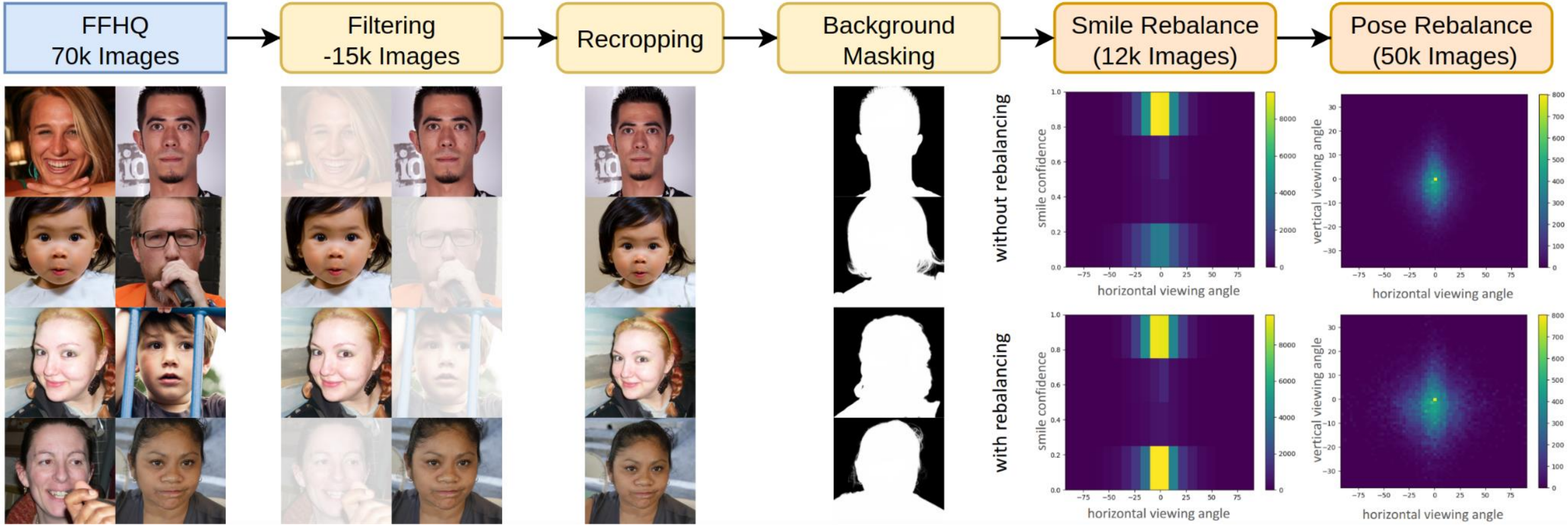
Bad Training Images



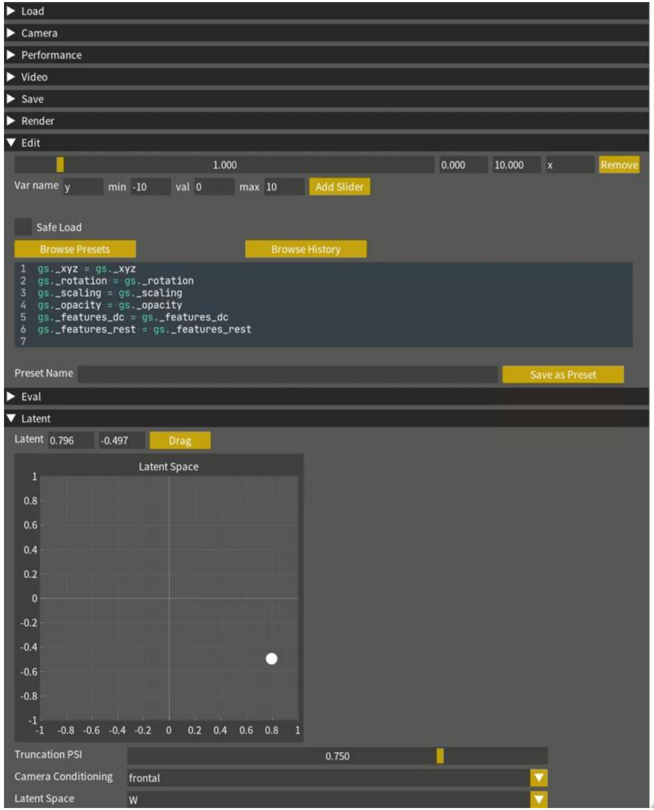
Microphone | Hand | >1 Face | Behind Obj. | Food | Instrument | Depiction | Blurry

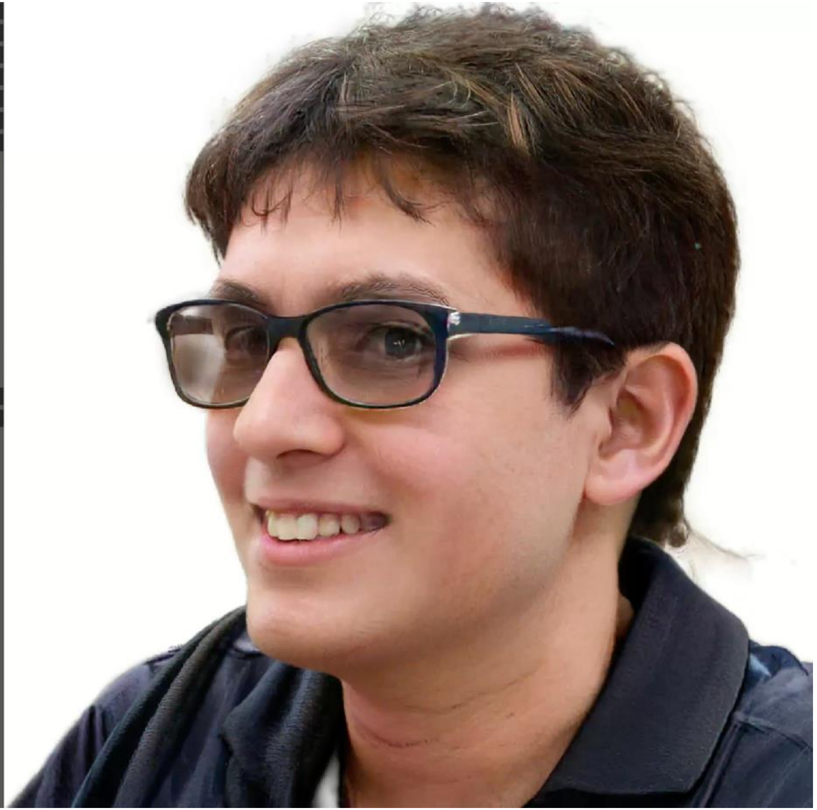
Florian Barthel

Preprocessing Overview



Results





Florian Barthel

Comparison

GGHead GSGAN Ours Ours FFHQC

▼ Load

Load pkl

▶ Camera

▶ Performance

▶ Save

▶ Render

▶ Edit

▶ Eval

▶ Latent

▼ Inversion

1. Load Images

Open Images from Files

▶ Webcam

2. Image Preprocessing

Preprocess

3. Run Latent Inversion

ID Similarity	0.10000	0.10000
MSE	0.01000	0.01000
LPIPS	1.00000	1.00000
Learning Rate	0.00050	0.00500
Batch Size	1 - +	1

Start Inversion

4. Run Generator Tuning

ID Similarity	0.10000	0.10000
MSE	0.01000	0.01000
LPIPS	1.00000	1.00000
Learning Rate	0.00050	0.00050
Batch Size	1 - +	1

Start Tuning

1

Motivation

2

3DGS GAN

3

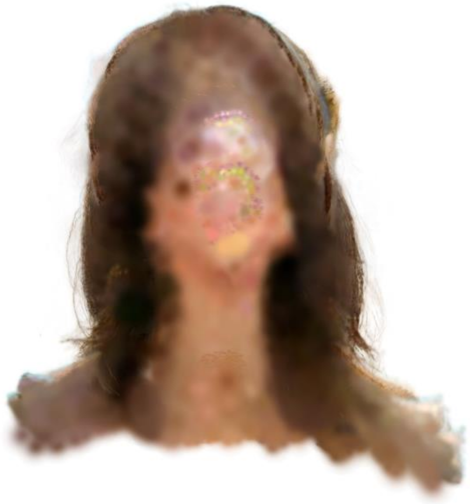
FFHQ Clean

4

Future Work

Florian Barthel

360° Data



Flux.1

Train on Large Datasets

FFHQ Dataset

55k Images

2k Resolution

Getty Images Dataset

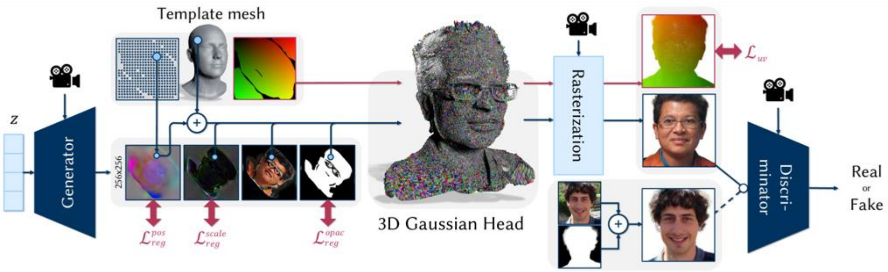
1.6M Images

4k Resolution

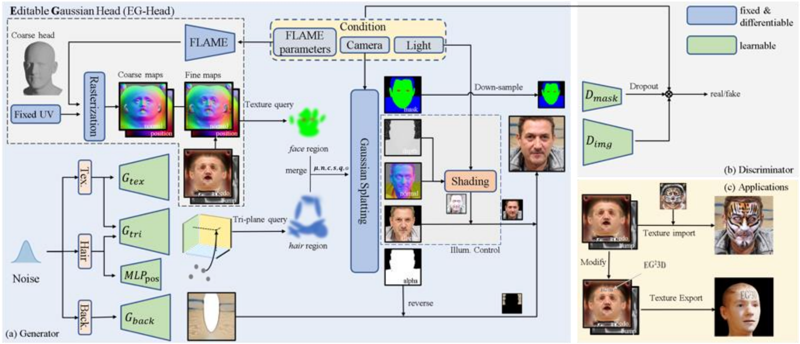
Florian Barthel

Train with Different Data Domains

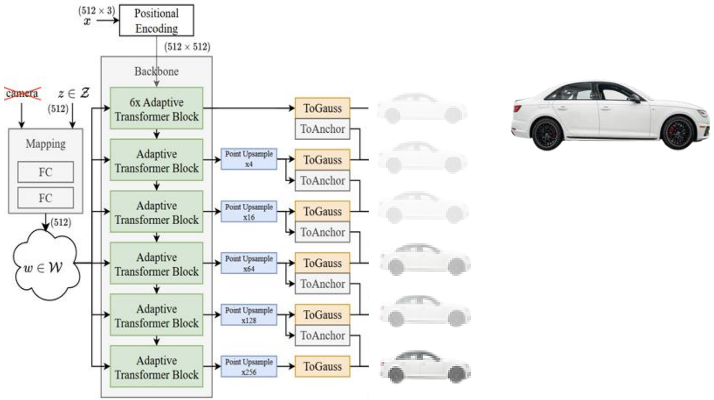
GGHead



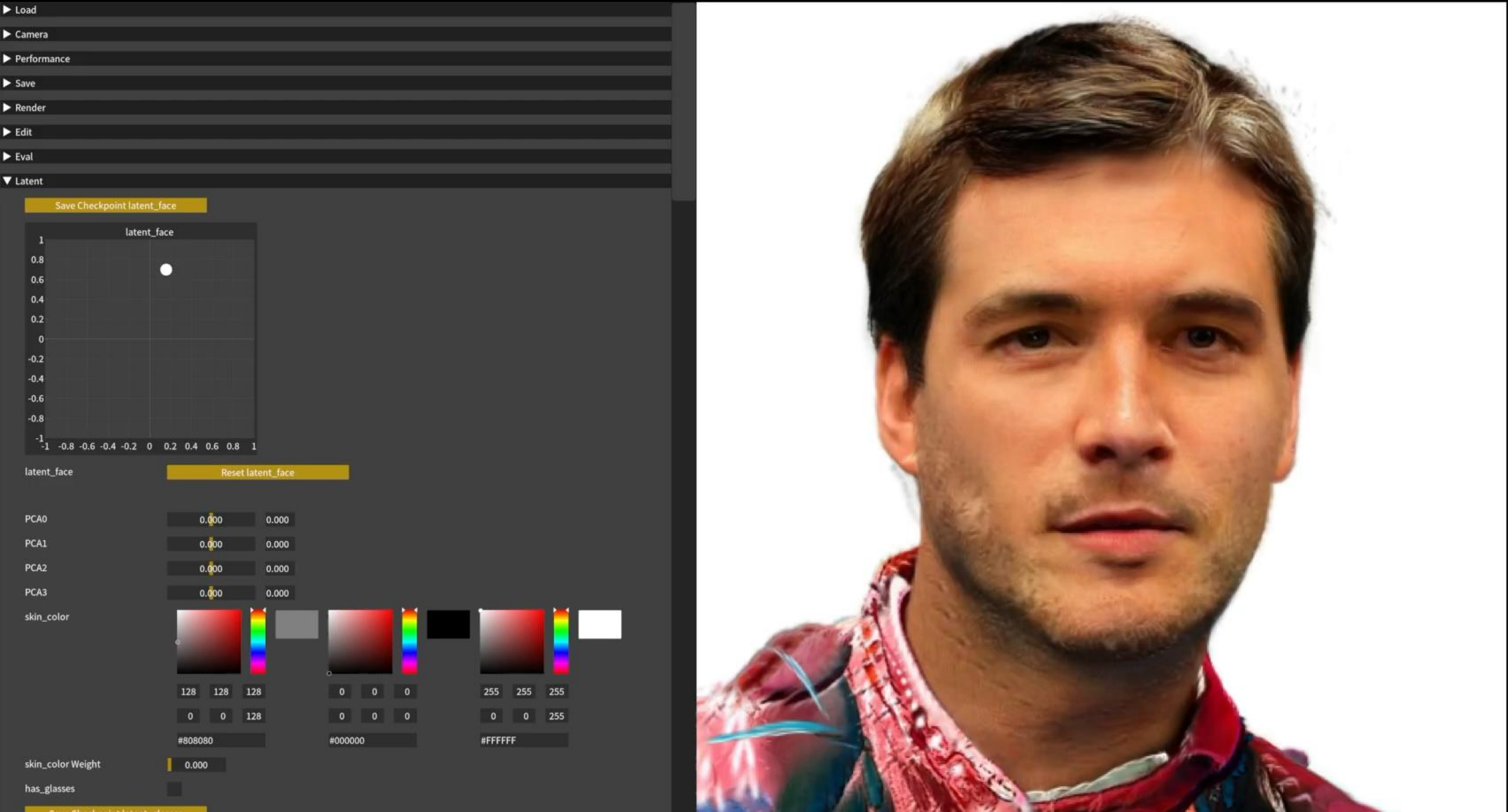
EGG3D



Ours

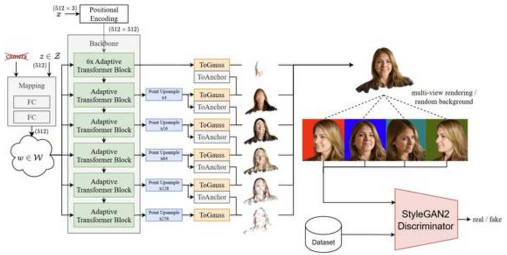


Towards building 3D training environments for robots…


Load
Camera
Performance
Save
Render
Edit
Eval
Latent
Save Checkpoint latent_face
latent_face
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1
-1 -0.8 -0.6 -0.4 -0.2 0 0.2 0.4 0.6 0.8 1
latent_face
Reset latent_face
PCA0 0.000 0.000
PCA1 0.000 0.000
PCA2 0.000 0.000
PCA3 0.000 0.000
skin_color
128 128 128
0 0 128
#808080
0 0 0
0 0 0
#000000
255 255 255
0 0 255
#FFFFFF
skin_color Weight 0.000
has_glasses

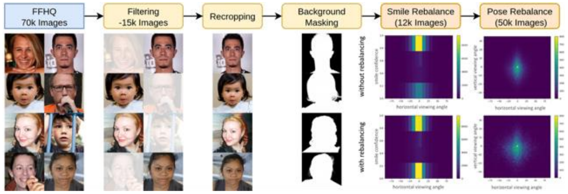
Conclusion

GAN Architecture



- 3D consistent rendering
- Stable training
- Scales to high resolutions

FFHQ Clean



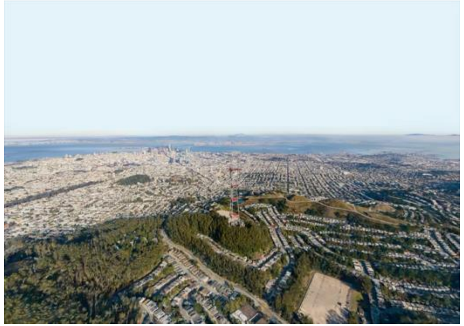
- Removes occluders
- Removes view-dependent biases
- High resolution

Current / Future Work

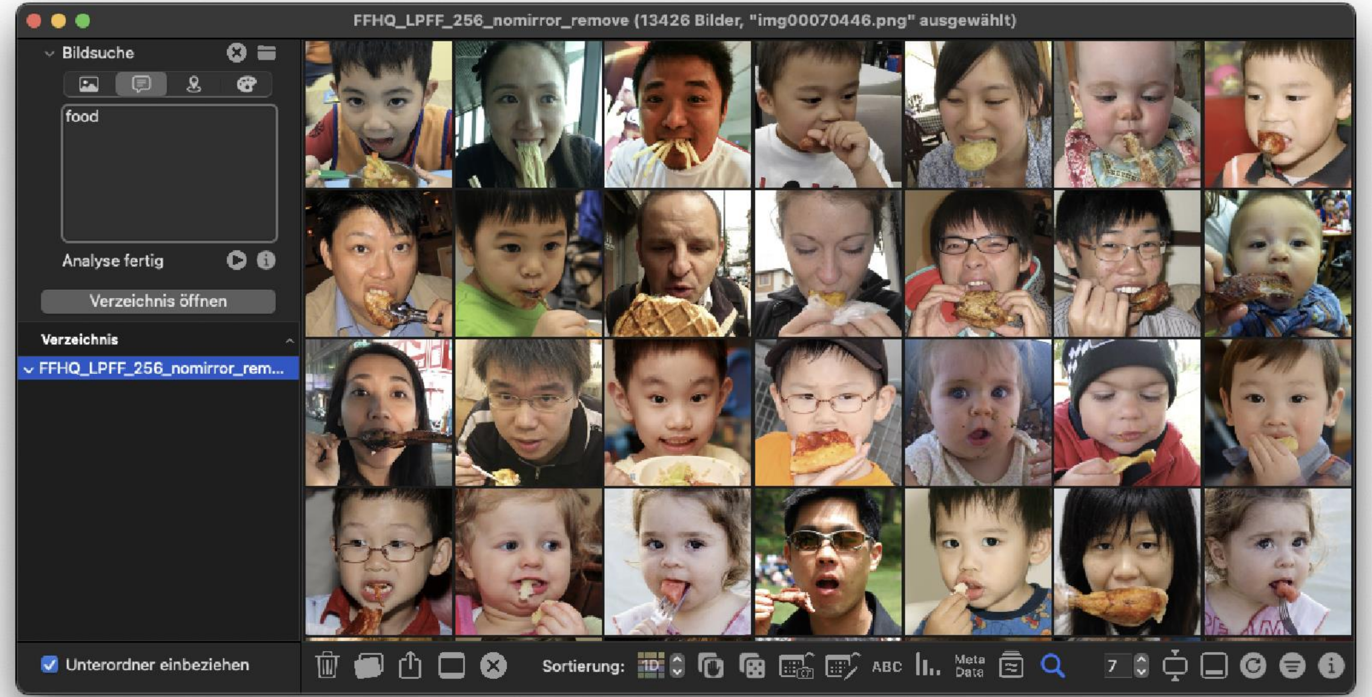
- Disentangled editing
- 360° rendering
- More training data
- Test new Data Domains

THANK YOU

MLaftermath Workshop



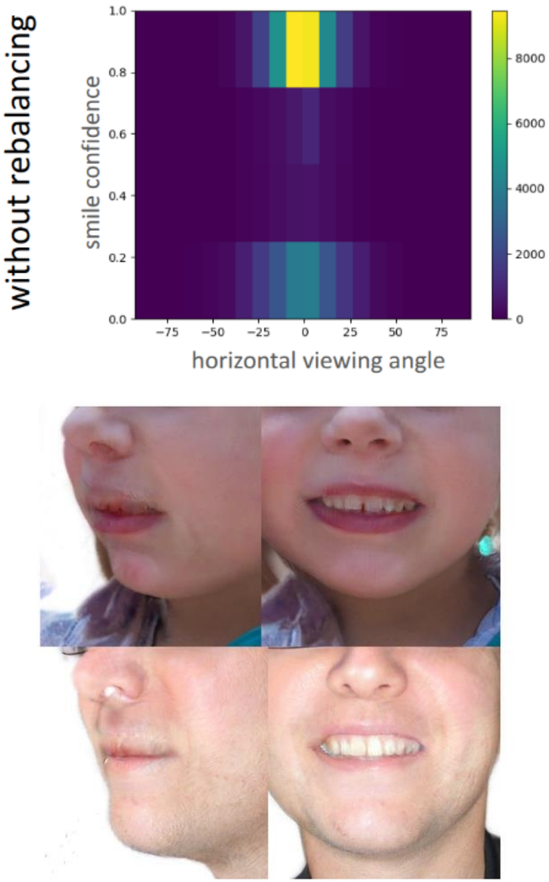
Filtering with CLIP



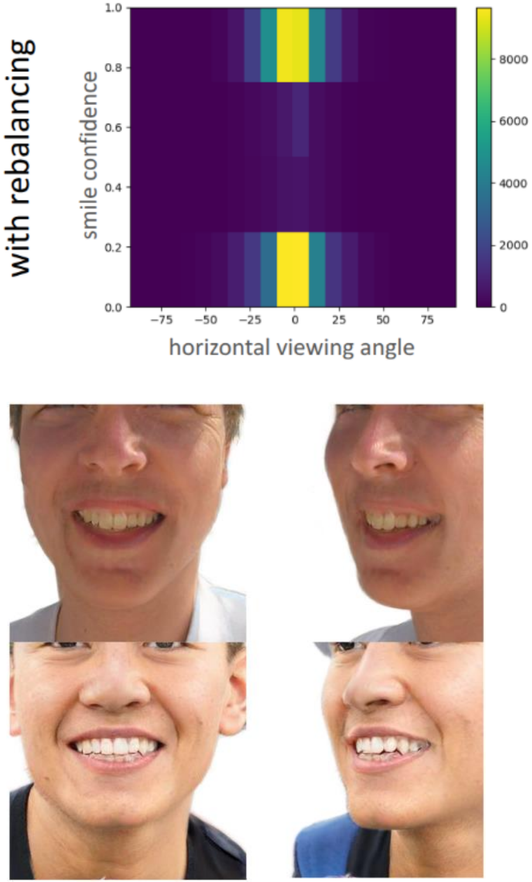
Florian Barthel

Fraunhofer HHI

Smiling Bias

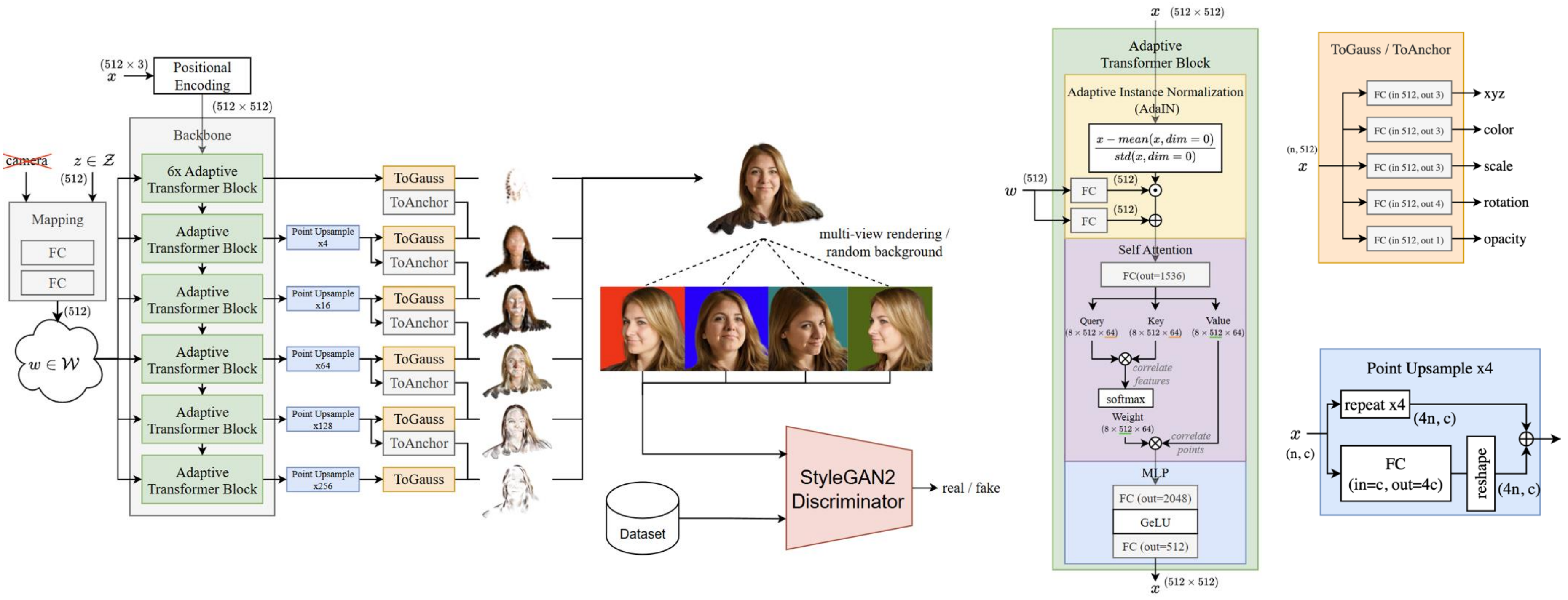


w/o smile rebalancing

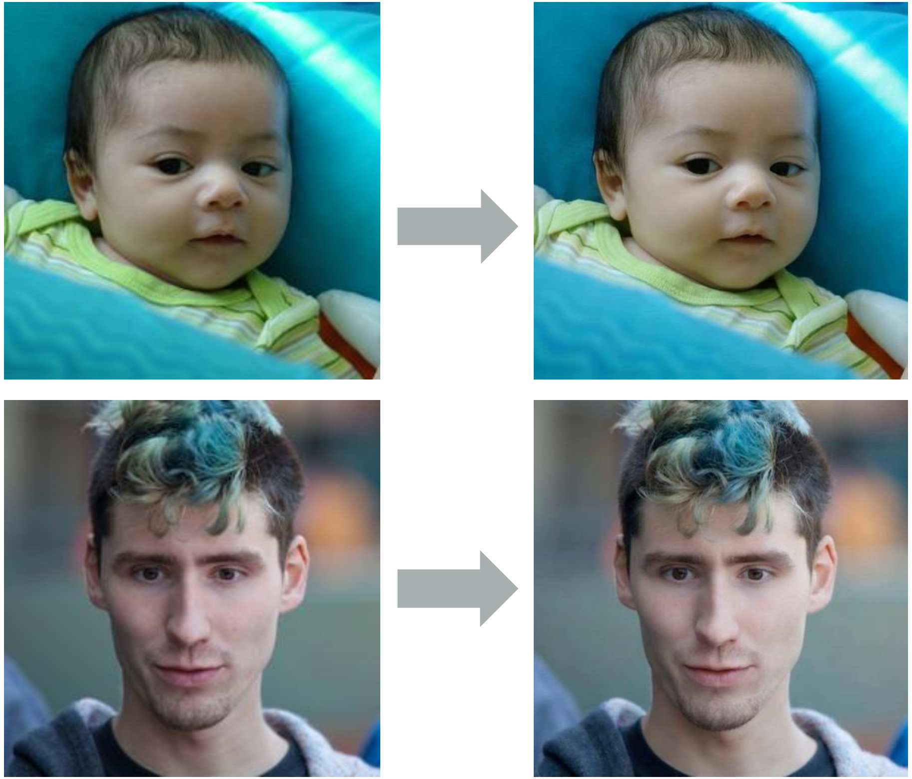


w/ smile rebalancing

CGS-GAN

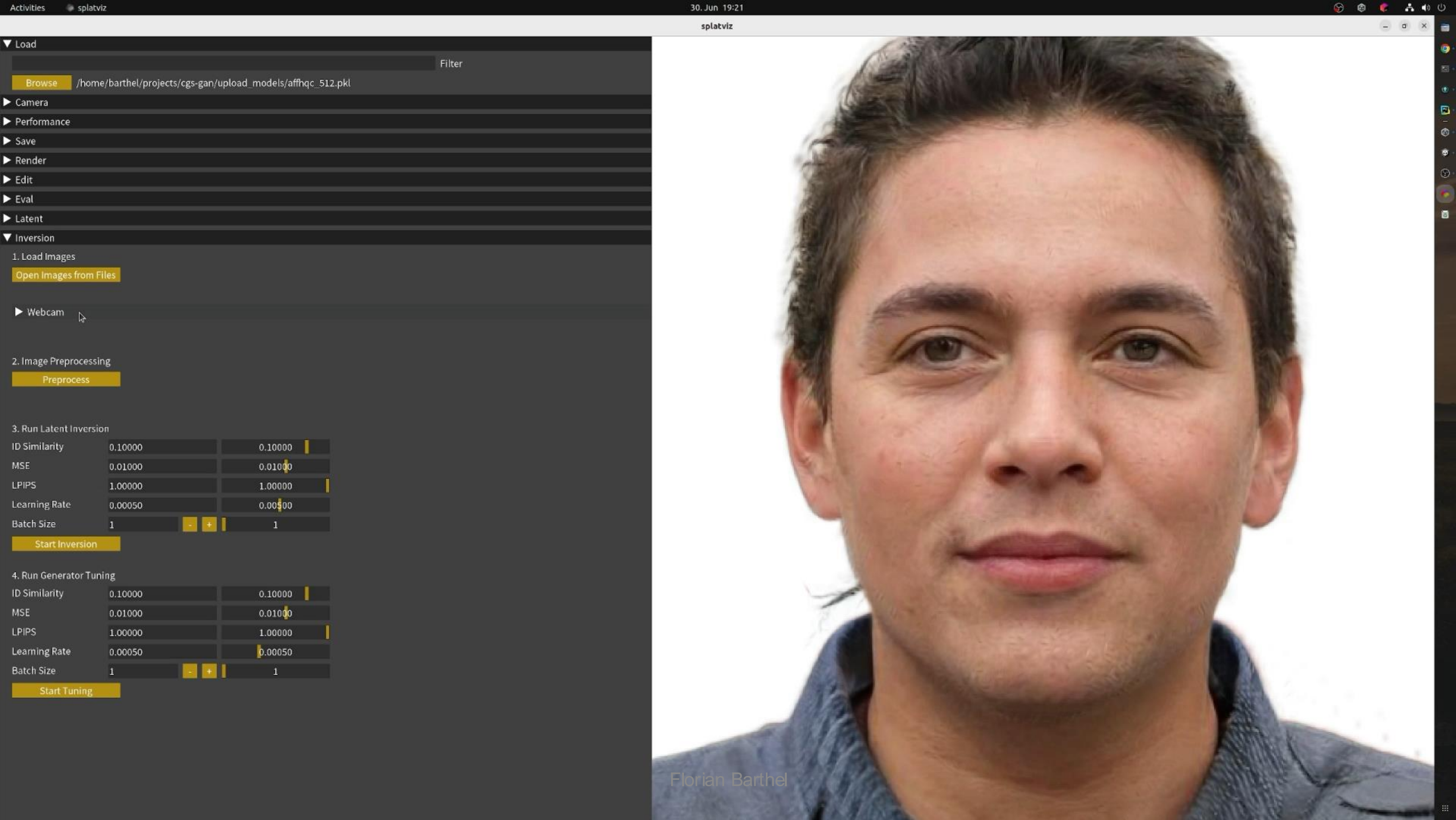

(512 × 3)
x
Positional Encoding
(512 × 512)
Backbone
camera
z ∈ Z
(512)
Mapping
FC
FC
(512)
w ∈ W
6x Adaptive Transformer Block
Adaptive Transformer Block
Adaptive Transformer Block
Adaptive Transformer Block
Adaptive Transformer Block
Adaptive Transformer Block
Point Upsample x4
Point Upsample x16
Point Upsample x64
Point Upsample x128
Point Upsample x256
ToGauss
ToAnchor
multi-view rendering / random background
Dataset
StyleGAN2 Discriminator
real / fake
x (512 × 512)
Adaptive Transformer Block
Adaptive Instance Normalization (AdaIN)
x − mean(x, dim = 0) / std(x, dim = 0)
w
(512)
FC
FC
(512)
(512)
Self Attention
FC(out=1536)
Query (8 × 512 × 64)
Key (8 × 512 × 64)
Value (8 × 512 × 64)
correlate features
softmax
Weight (8 × 512 × 64)
correlate points
MLP
FC (out=2048)
GeLU
FC (out=512)
(512 × 512)
x
ToGauss / ToAnchor
(n, 512)
x
FC (in 512, out 3)
FC (in 512, out 3)
FC (in 512, out 3)
FC (in 512, out 4)
FC (in 512, out 1)
xyz
color
scale
rotation
opacity
Point Upsample x4
x
(n, c)
repeat x4
(4n, c)
FC (in=c, out=4c)
reshape
(4n, c)

Diffusion Dataset



"Create a realistic photo of a person **${prop1}** **${prop1}** *in front of a white background viewed from the* **${cam_pose}** *"*

prop1: with a red hat
prop2: wearing sunglasses
cam_pose: front left


Activities
splatviz
30. Jun 19:21
splatviz
Load
Filter
Browse
/home/barthel/projects/cgs-gan/upload_models/affhqc_512.pkl
Camera
Performance
Save
Render
Edit
Eval
Latent
Inversion
1. Load Images
Open Images from Files
Webcam
2. Image Preprocessing
Preprocess
3. Run Latent Inversion
ID Similarity 0.10000 0.10000
MSE 0.01000 0.01000
LPIPS 1.00000 1.00000
Learning Rate 0.00050 0.00500
Batch Size 1 - + 1
Start Inversion
4. Run Generator Tuning
ID Similarity 0.10000 0.10000
MSE 0.01000 0.01000
LPIPS 1.00000 1.00000
Learning Rate 0.00050 0.00050
Batch Size 1 - + 1
Start Tuning
Florian Barthel

Results

FID	FFHQ 512
GSGAN	5.02
GGHead	**4.34**
Ours	4.94

$\mathbf{FID_{3D}}$	FFHQ 512
GSGAN	10.50
GGHead	7.90
Ours	**4.94**

Visual Results

GGHead

GSGAN

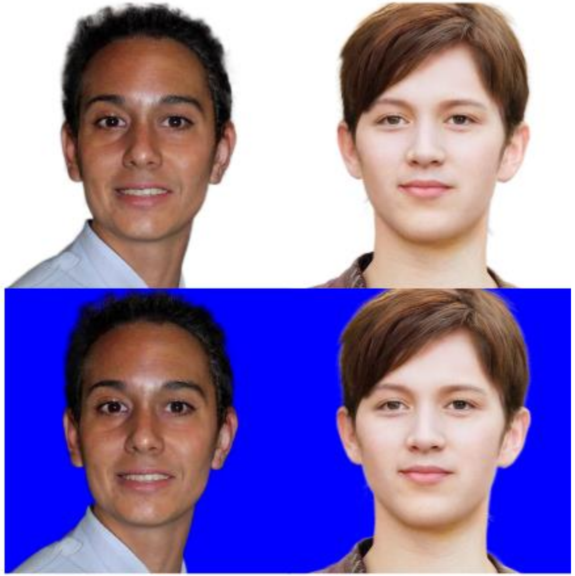
Comparison

FID$_{3D}$	FFHQ	FFHQ Clean		
	512	512	1024	2048
GSGAN	10.50	7.68	/	/
GGHead	7.90	7.78	14.27	/
Ours	**4.94**	**4.53**	**5.25**	7.8

Florian Barthel

Background Color Augmentation

trained with **white** backgrounds



trained with **random** backgrounds