Neural Scene Representation, Rendering, and Generation
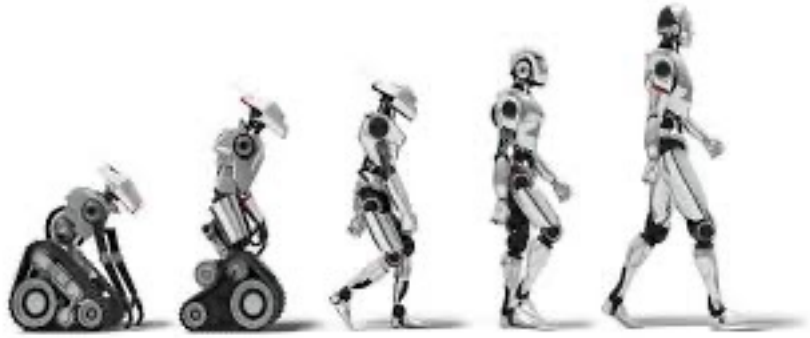
Gordon Wetzstein

Stanford EE & CS

www.computationalimaging.org

Neural Radiance Fields for View Interpolation
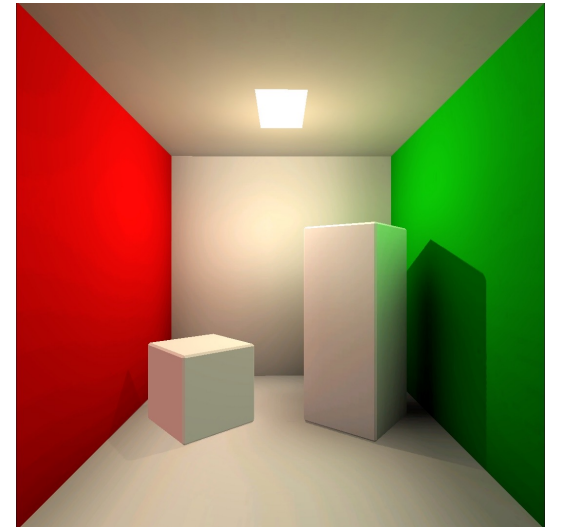
Robotics

Autonomous Driving

VR/AR

Neural Scene Representation, Rendering, and Generation

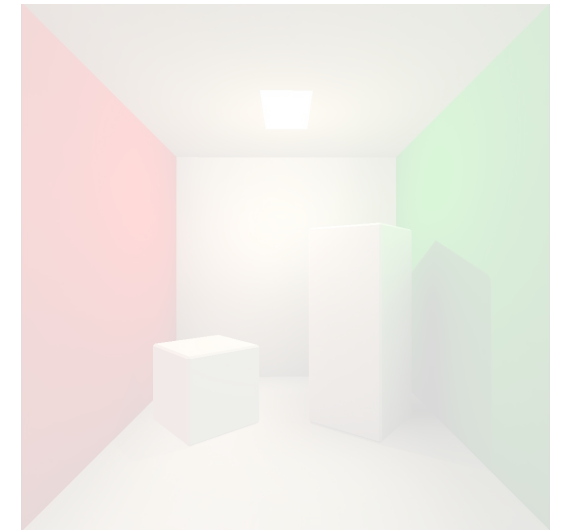Computer Vision

Computer Graphics

Robotics

Autonomous Driving

VR/AR

Neural Scene Representation, Rendering, and Generation

Computer Vision

Computer Graphics

Efficient Geometry-aware 3D Generative Adversarial Networks

Potential Training Datasets

STUDIO-CAPTURED MULTIVIEW

Triplegangers
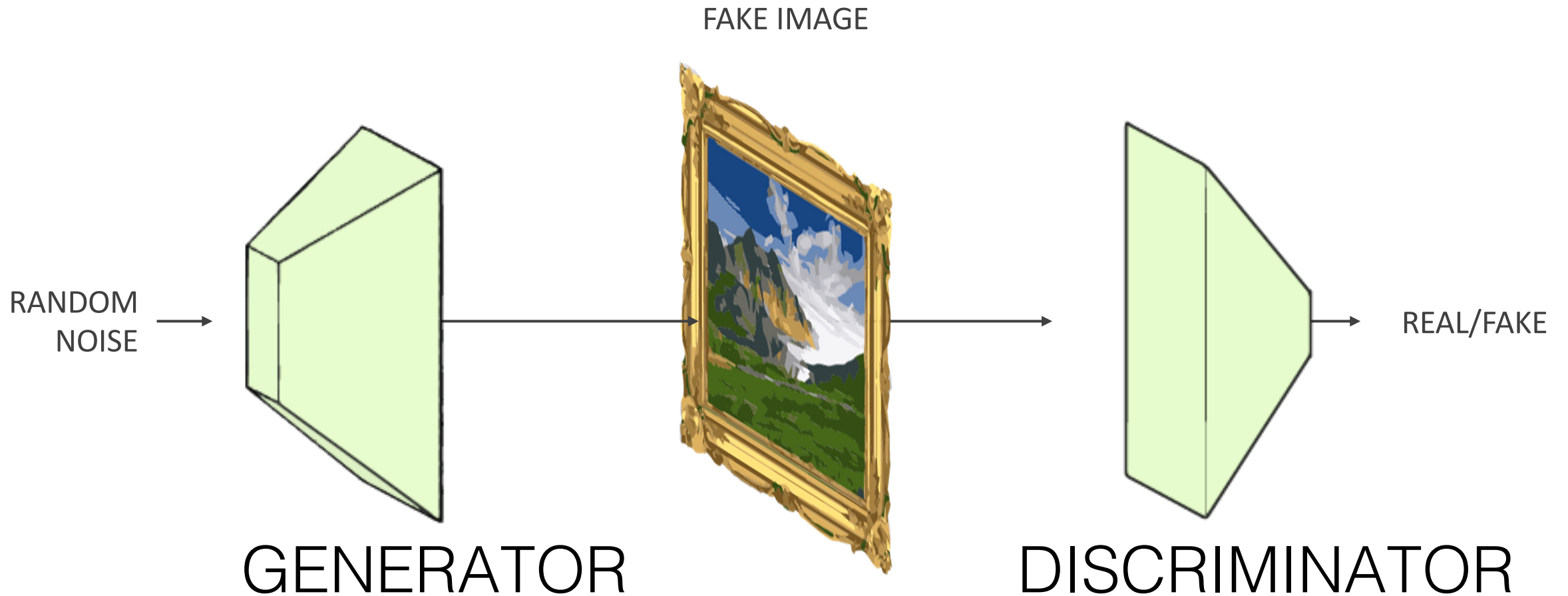515 Identities
3D ground truth

SYNTHETIC DATA

Face Synthetics
100K Identities
Multi-view

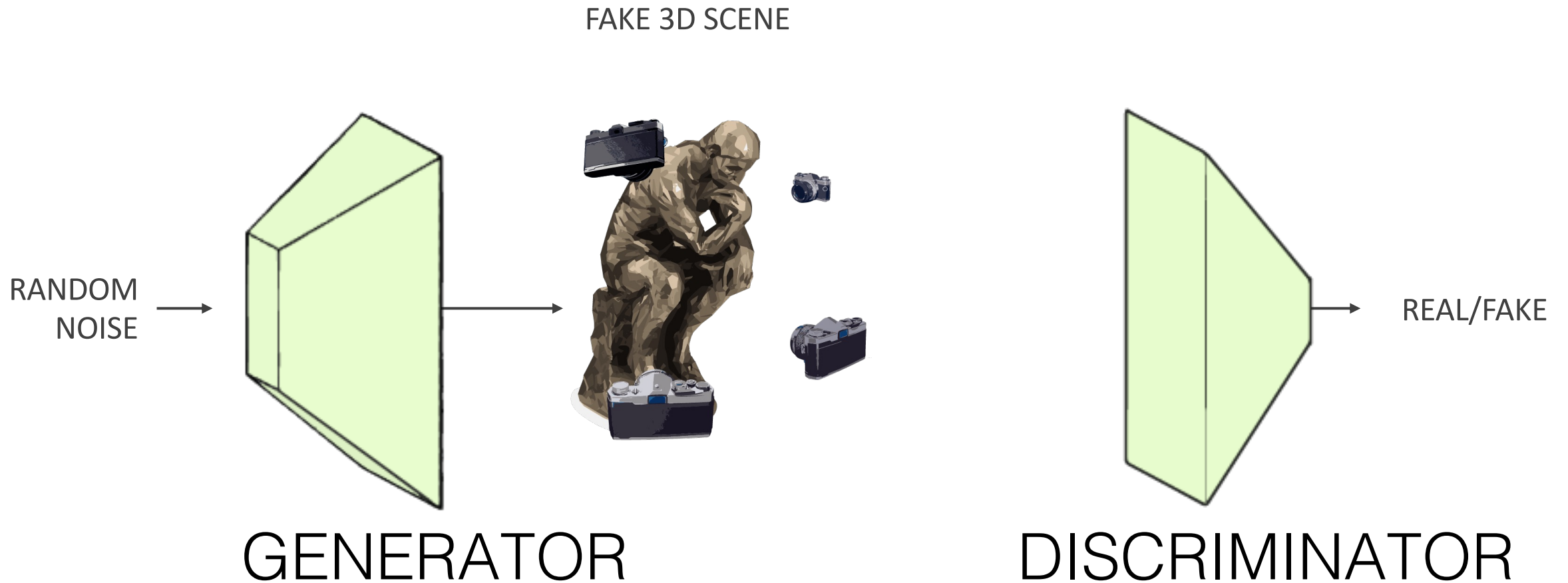**UNSTRUCTURED SINGLE-VIEW**

FFHQ
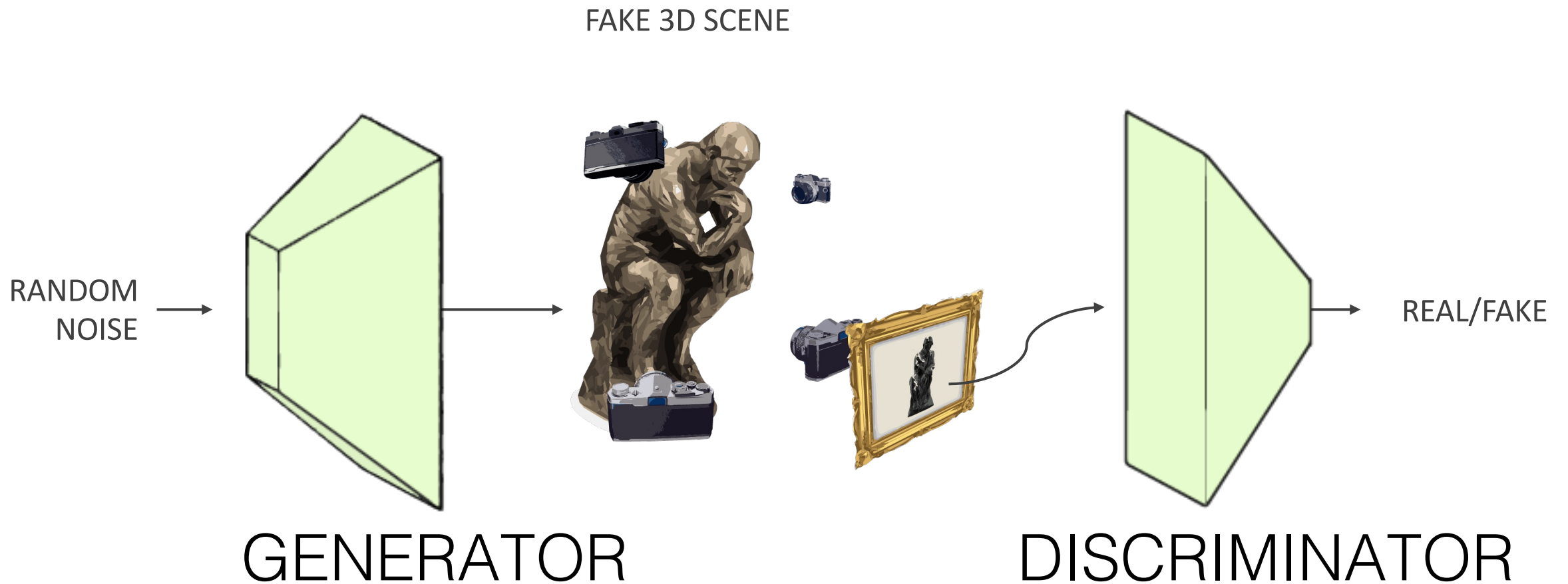70K Identities
Scraped from the internet

Training a 2D GAN

FAKE IMAGE

RANDOM
NOISE

REAL/FAKE

GENERATOR

DISCRIMINATOR

Training a 3D-aware GAN

FAKE 3D SCENE

RANDOM NOISE

REAL/FAKE

GENERATOR

DISCRIMINATOR

Training a 3D-aware GAN

FAKE 3D SCENE

RANDOM
NOISE

REAL/FAKE

GENERATOR

DISCRIMINATOR

Training a 3D-aware GAN
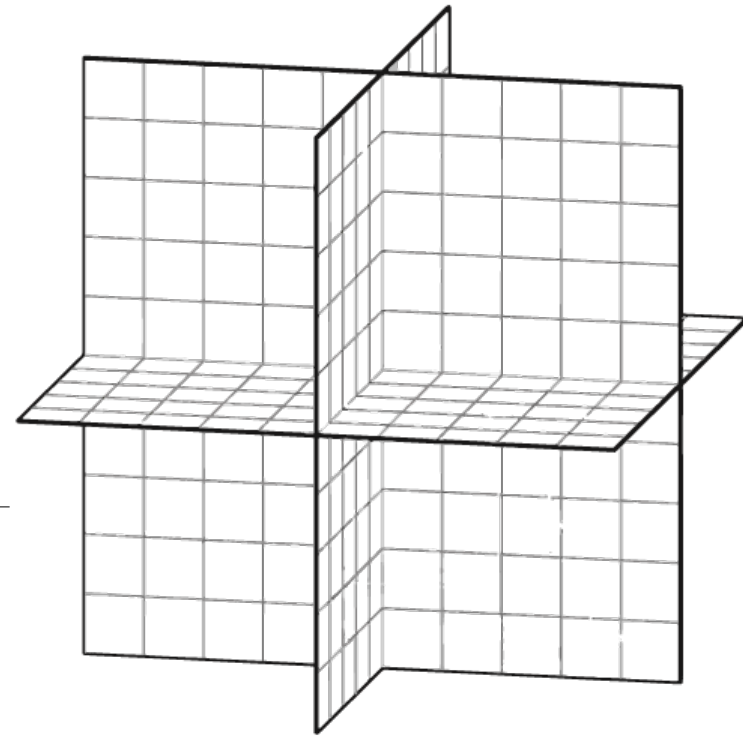

FAKE 3D SCENE

RANDOM
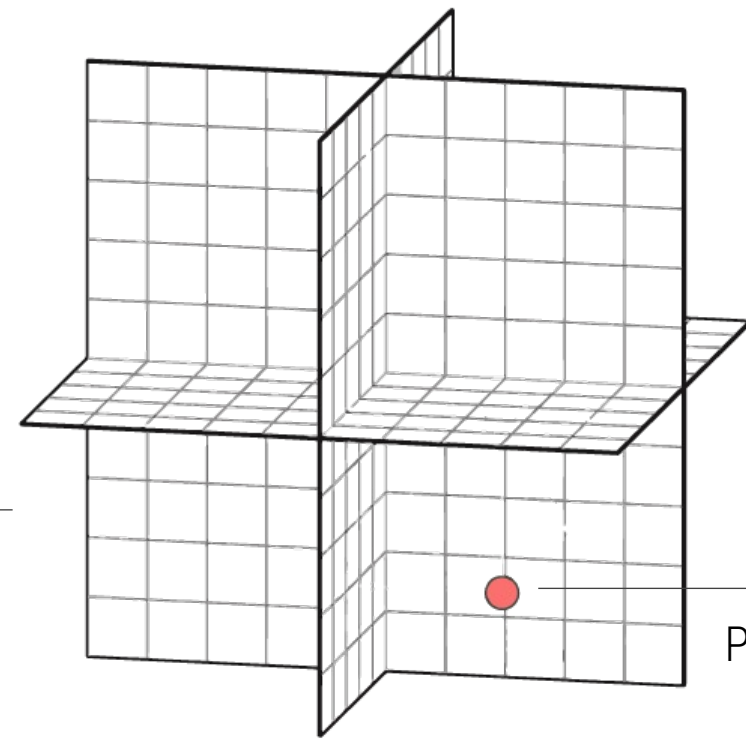NOISE

GENERATOR

REAL/FAKE

DISCRIMINATOR

Tri-Plane Representation

Explicit features
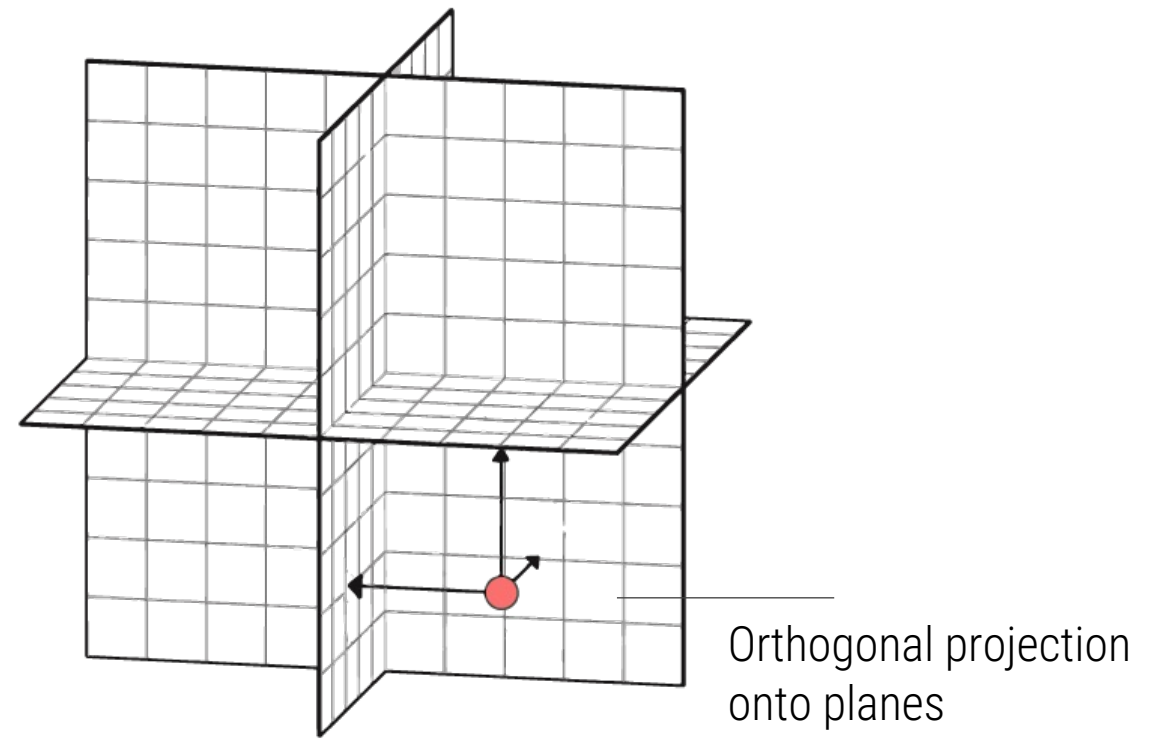aligned along three
orthogonal planes

Tri-Plane Representation

Explicit features
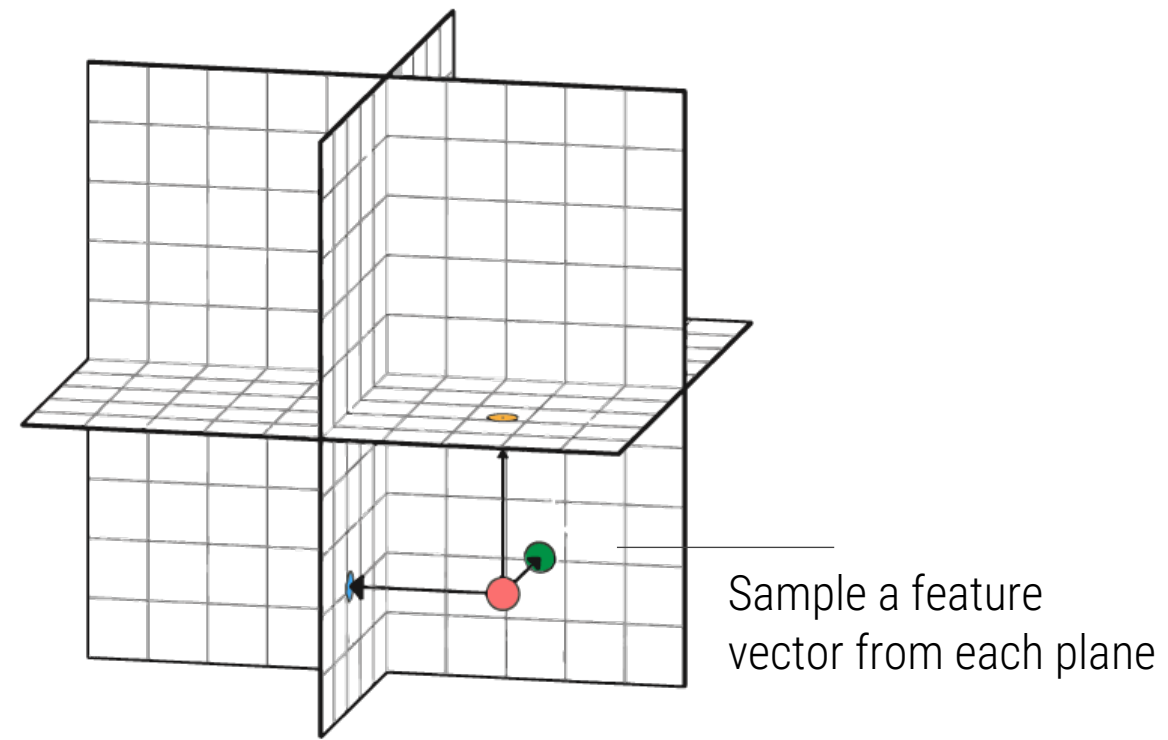aligned along three
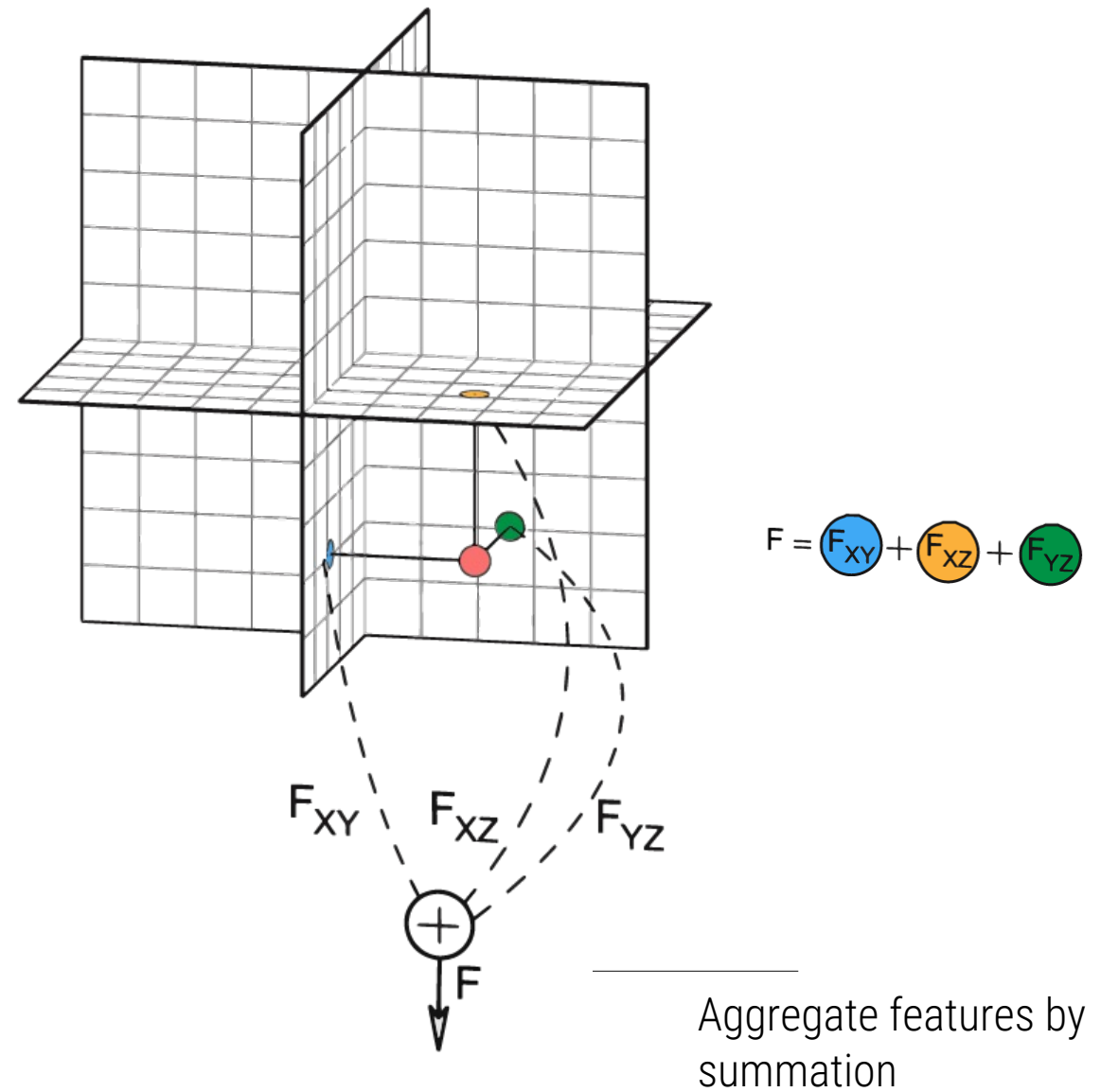orthogonal planes

Position in space (**XYZ**)

Tri-Plane Representation

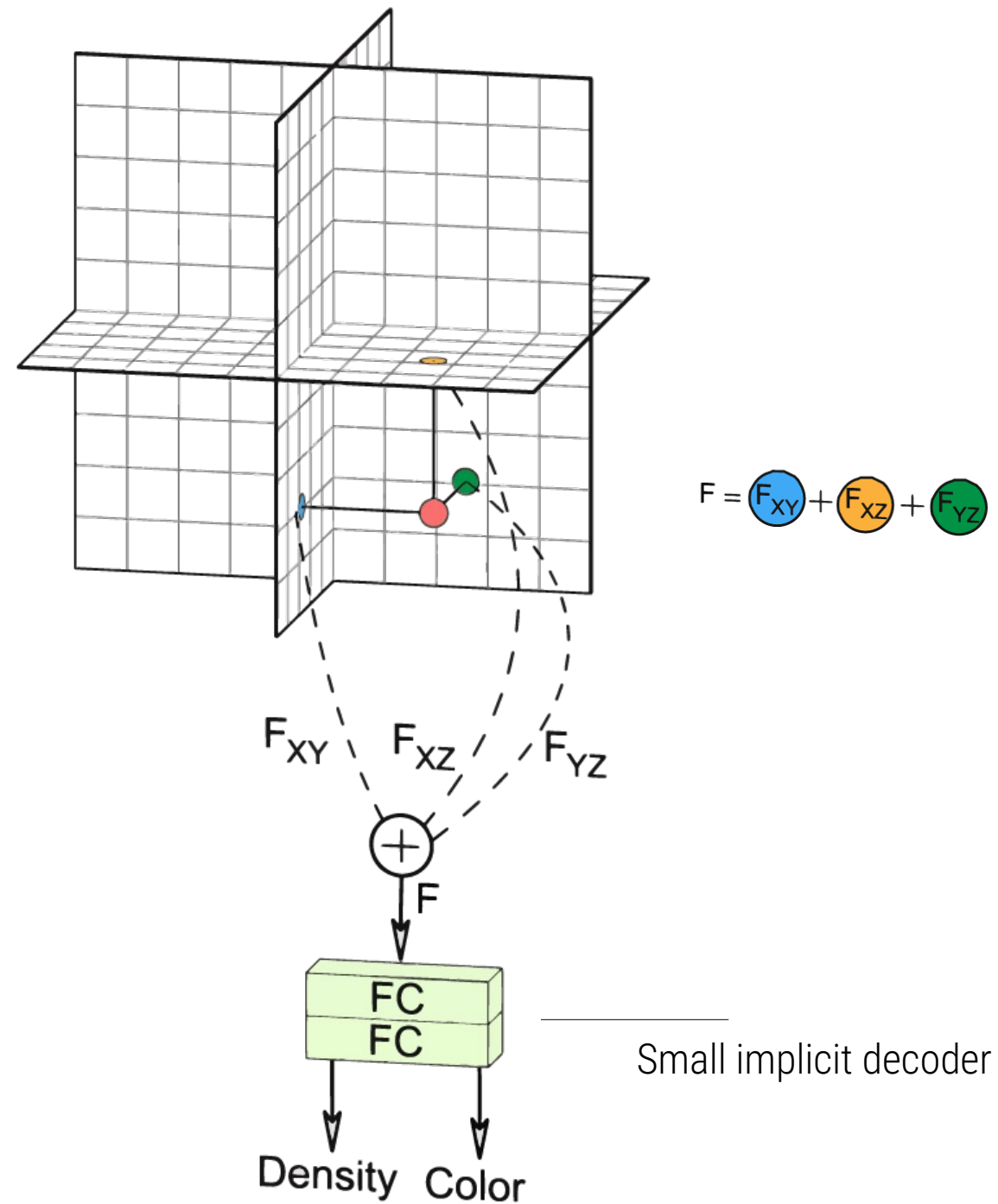Orthogonal projection
onto planes

Tri-Plane Representation

Sample a feature
vector from each plane

Tri-Plane Representation

$$F = F_{XY} + F_{XZ} + F_{YZ}$$

F_{XY} F_{XZ} F_{YZ}

$+$

F

Aggregate features by summation

Chan et al. CVPR 2022

Tri-Plane Representation

$$F = F_{XY} + F_{XZ} + F_{YZ}$$

Small implicit decoder
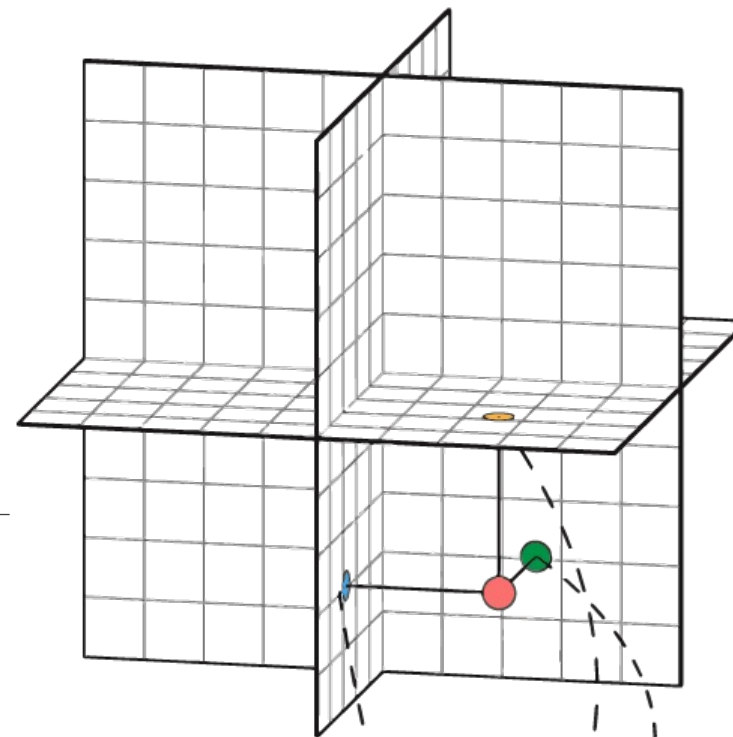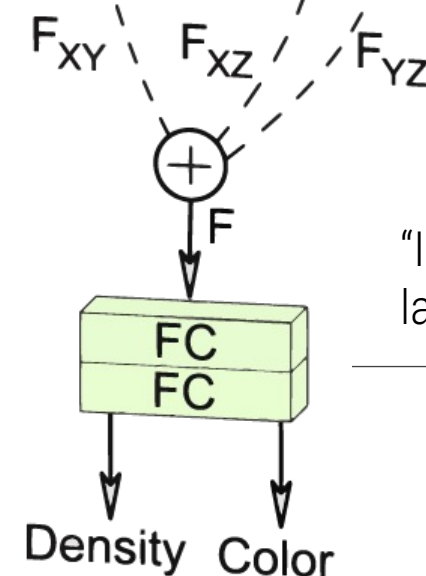
Tri-Plane Representation

"Explicit Features" aligned along three orthogonal planes

$F = F_{XY} + F_{XZ} + F_{YZ}$

+ Memory efficient
  (scales quadratically vs. cubically)

+ Easy to generalize

+ Compatible with off-the-shelf
  2D backbones, e.g. StyleGAN

F_{XY} F_{XZ} F_{YZ}

$\oplus$

F

"Implicit Decoder": two layers of 64 hidden units

FC
FC

Density Color

Mip-NeRF (fully implicit)

Tri-Plane (ours)

7.8X faster!

3D GAN Pipeline
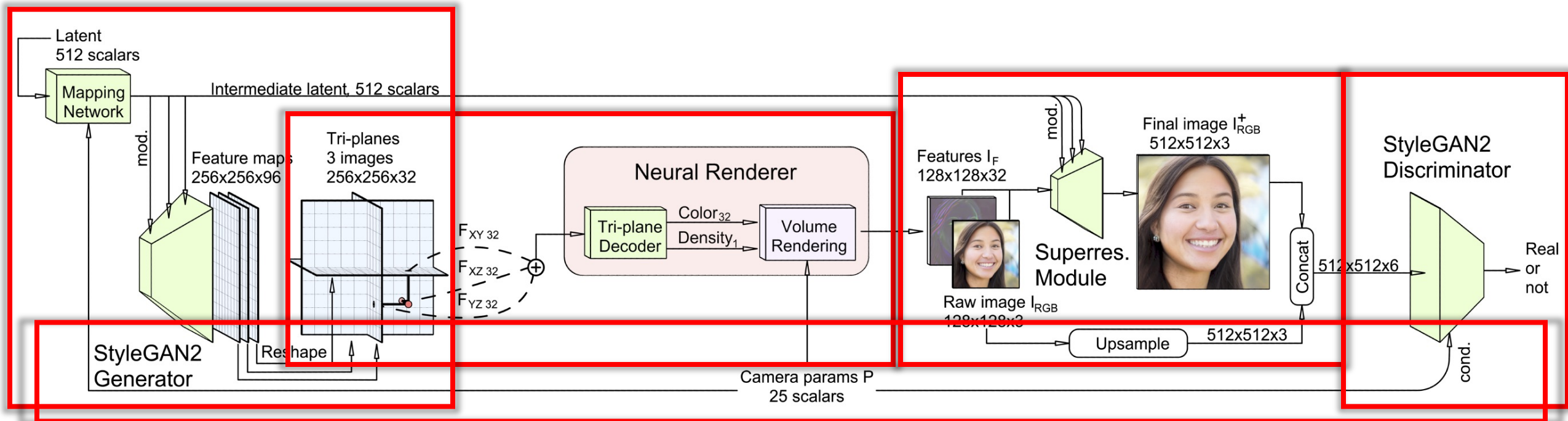
Standard Discriminator

Dataset

Generator

Real Image
(512 x 512 x 3)

Discriminator

Synthesized Final Image I⁺_RGB
(512 x 512 x 3)

Real or Fake

Dual Discriminator

Dataset

Generator

Discriminator

Real Image
(512 x 512 x 3)

Blurred Real Image
(512 x 512 x 3)

Synthesized Final Image I*$_{RGB}$
(512 x 512 x 3)

Synthesized
Raw Rendering I$_{RGB}$
(128 x 128 x 3)

Upsample

Upsampled Raw Rendering
(512 x 512 x 3)

Real or Fake

Chan et al. CVPR 2022
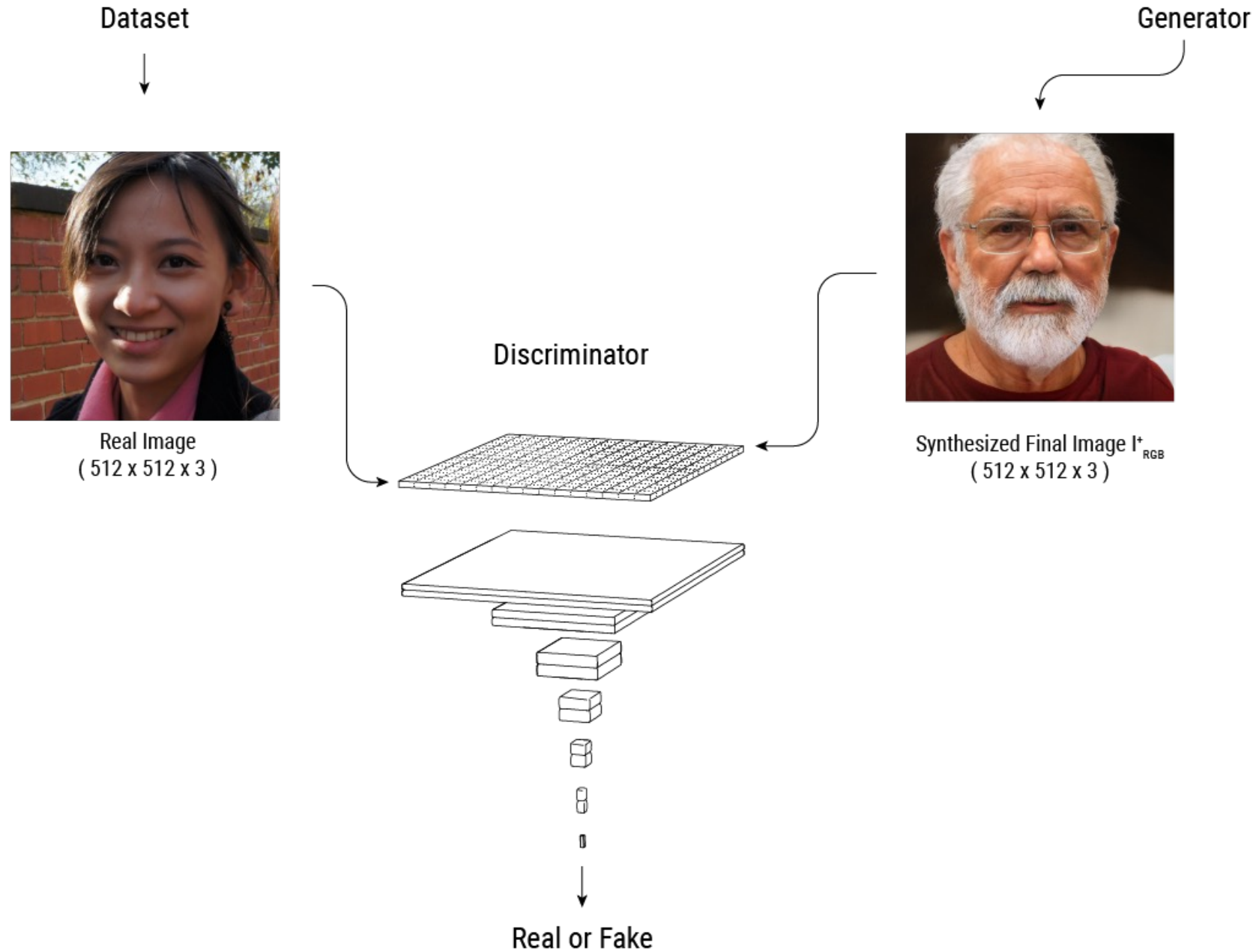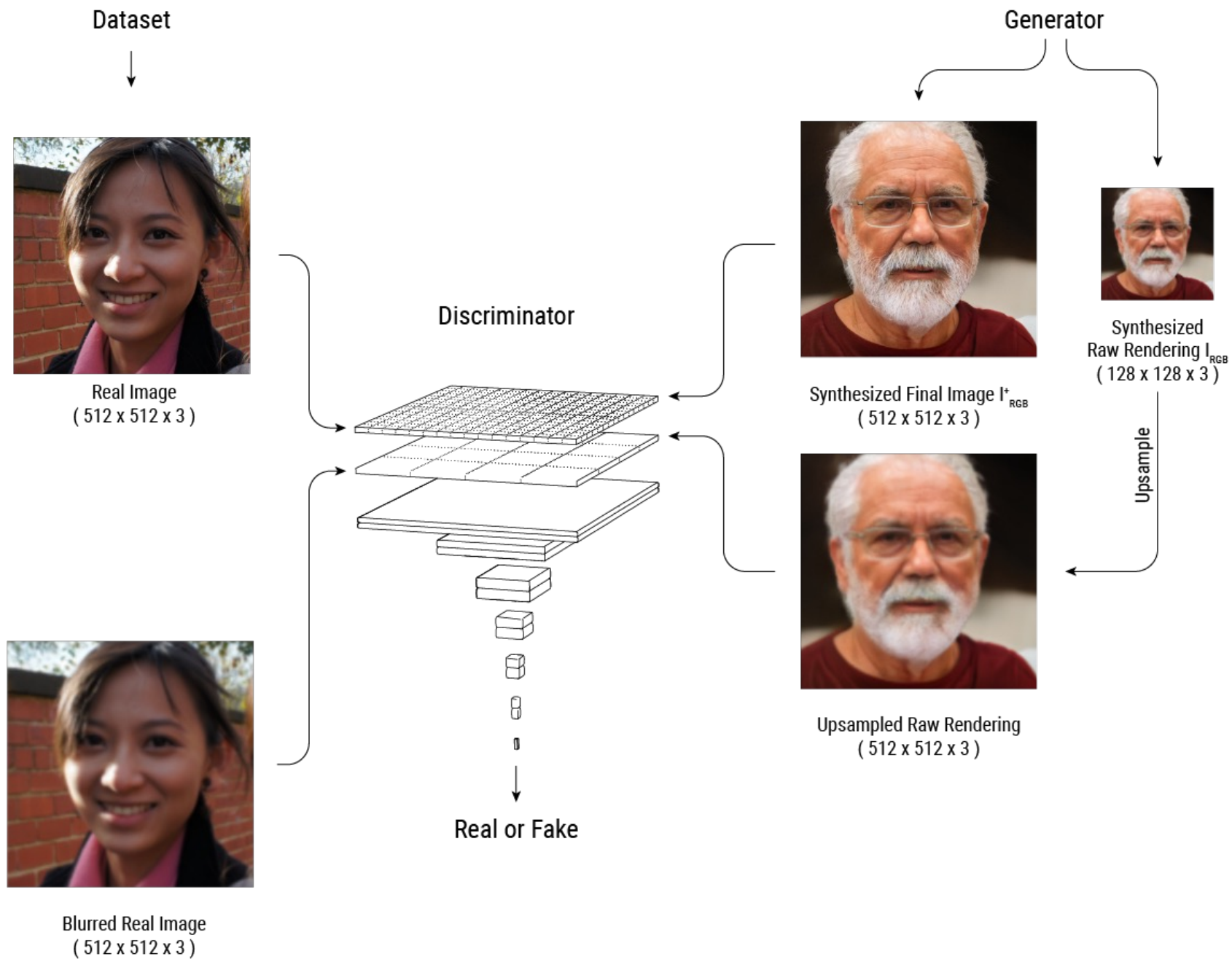
Raw Neural Rendering (learned underlying 3D representation)
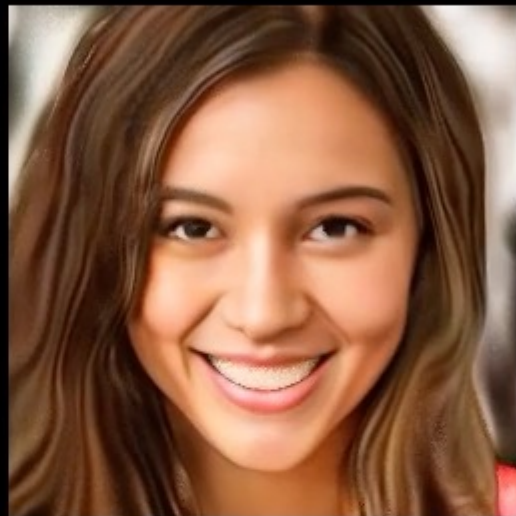
Chan et al. CVPR 2022

Comparisons with State-of-the-art 3D GANs

GIRAFFE

pi-GAN

Lifting SG.

Ours256

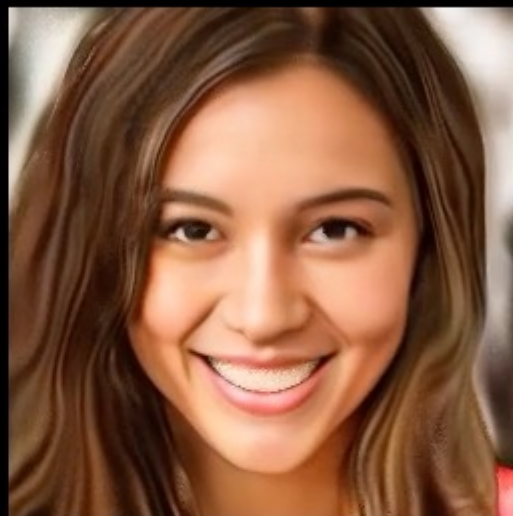# Comparisons with State-of-the-art 3D GANs

Framerate (↑)   FID (↓)

1 fps   29.9

35 fps   4.7

pi-GAN

Ours256

Latent Code Interpolation

Real-time Demonstration

GAN Inversion

3D GAN inversion

Source Image

Controllable Portrait Image Animation via 3D GAN Inversion

Attribute-Edited Images

Source Image

Original   Age   Beard

Glasses   Gender   Hair

Lin et al. arxiv 2022

Controllable Portrait Image Animation via 3D GAN Inversion

Attribute-Edited Images

Source Image   Original   Age   Beard

Target Video   Glasses   Gender   Hair

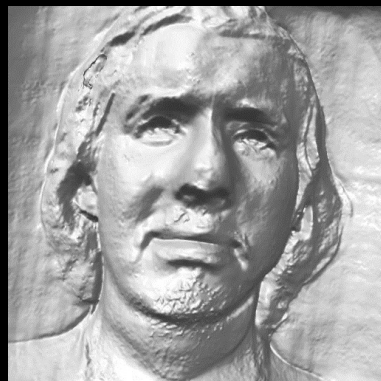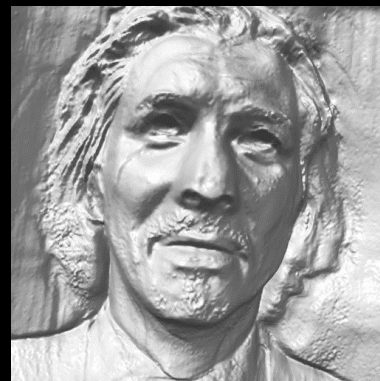Controllable Portrait Image Animation via 3D GAN Inversion

Attribute-Edited Images

Source Image

Target Video

Original

Age

Beard

Glasses

Gender

Hair

Controllable Portrait Image Animation via 3D GAN Inversion

Lin et al. arxiv 2022

Predicted 3D Geometry

Source Image

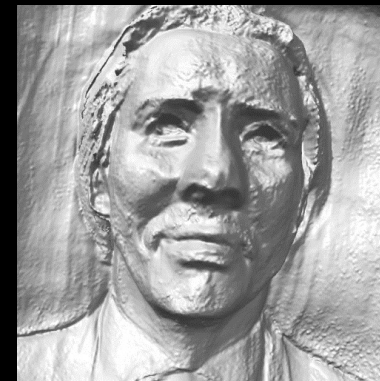**Target Video**

Original Age Beard

Glasses Gender Hair

Controllable Portrait Image Animation via 3D GAN Inversion

3D GAN for Human Bodies?

3D GAN for Human Bodies?

More challenging than faces:

- Much higher diversity in articulation (body poses)

- Challenging to get "clean" datasets with 3D joint / pose annotation, good segmentation masks, camera poses, not too many self-occlusions, etc.

- Usually lots of other items in the images: accessories, clothes, bikes, ...

GNARF: Generative Neural Articulated Radiance Fields

Neural radiance field generator

Z
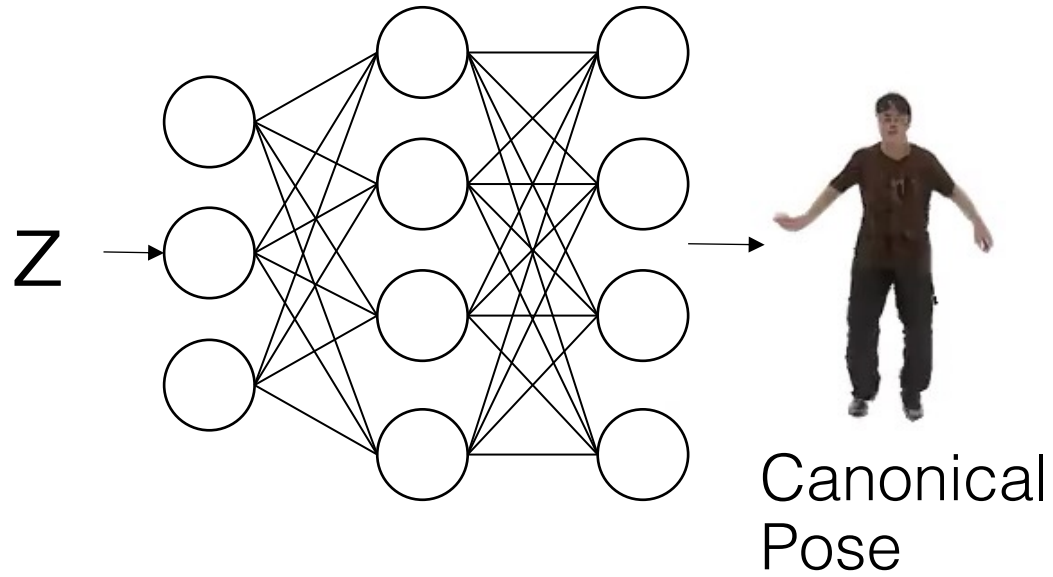
Canonical
Pose

- Generator controls identity in canonical space

GNARF: Generative Neural Articulated Radiance Fields

driving mesh

driven radiance fields

GNARF: Generative Neural Articulated Radiance Fields

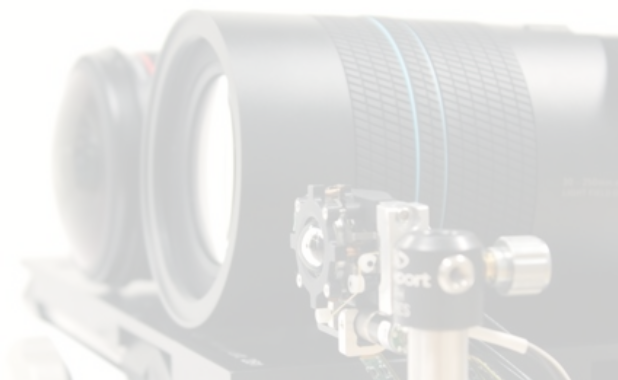Baseline: EG3D [Chan et al. 2022] + canonical pose estimation + deformation
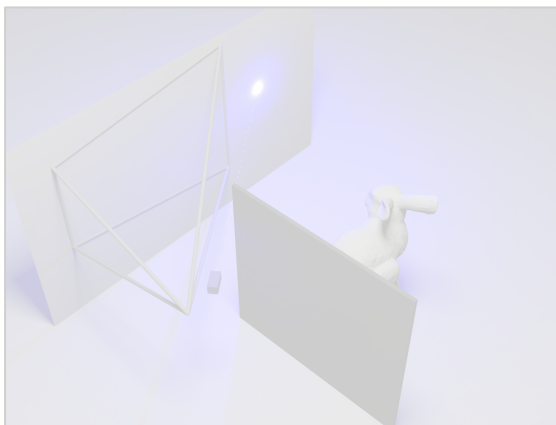
driving mesh

GNARF (ours)
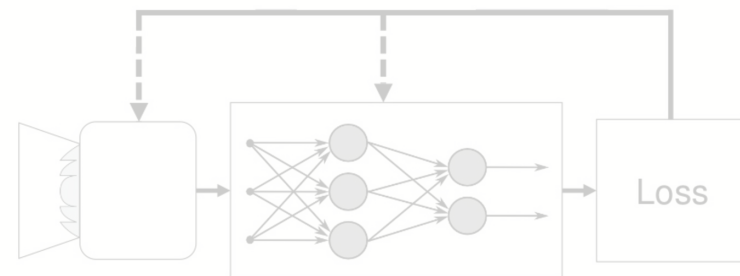
Stanford Computational Imaging Lab
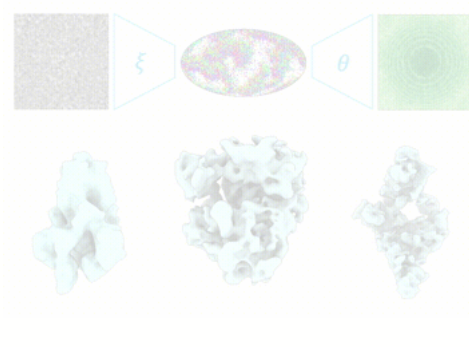

Computational Cameras

Single-photon Imaging

Deep Optics

VR/AR & Wearable Computing

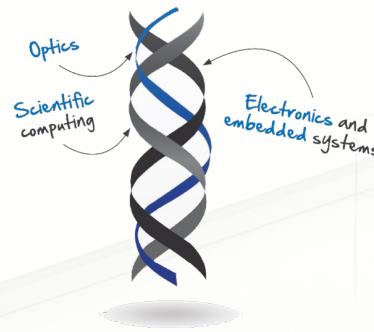Computational Microscopy

Neural Rendering

EG3D AUTHORS

Gordon Wetzstein
stanford.edu/~gordonwz

Computational Imaging Lab
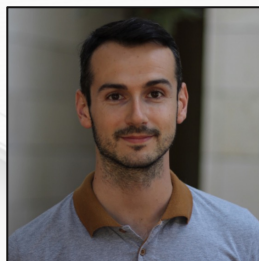Stanford University EE/CS

computationalimaging.org
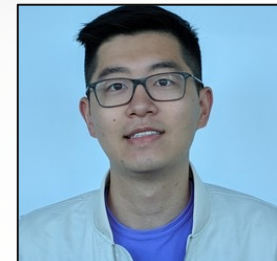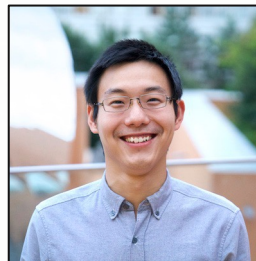

Vincent Sitzmann

Eric Chan

Julien Martel

David Lindell

Marco Monteiro

Petr Kellnhofer

Alex Bergman

Connor Lin

Jiajun Wu