SAMSUNG SAIT

Stanford University

SCI STANFORD COMPUTATIONAL IMAGING LAB

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

SinGRAF: Learning a 3D Generative Radiance Field for a Single Scene

WED-AM-027

Minjung Son*[1,2] Jeong Joon Park*[2] Leonidas Guibas[2] Gordon Wetzstein[2]

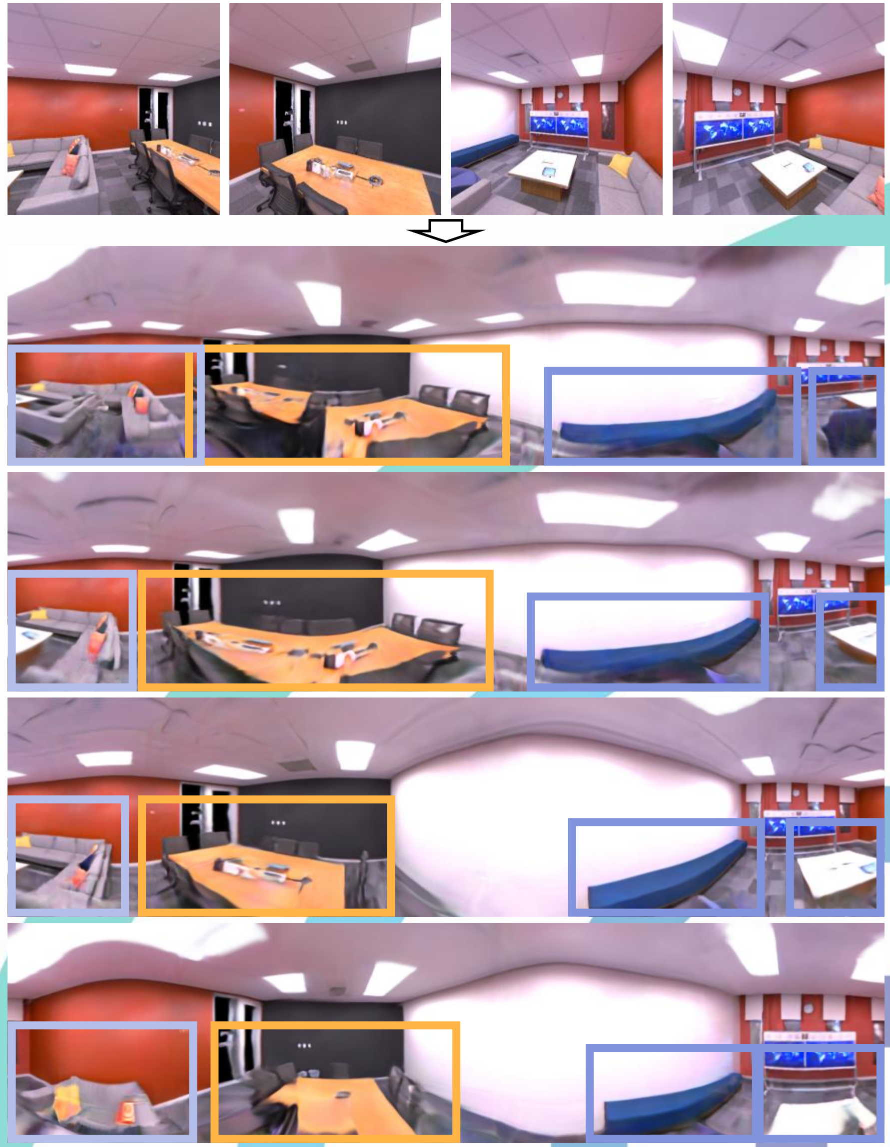
[1]Samsung Advanced Institute of Technology (SAIT)

[2]Stanford University

June 21 Wed, 2023

Content Creation
from a Few Unposed Images?

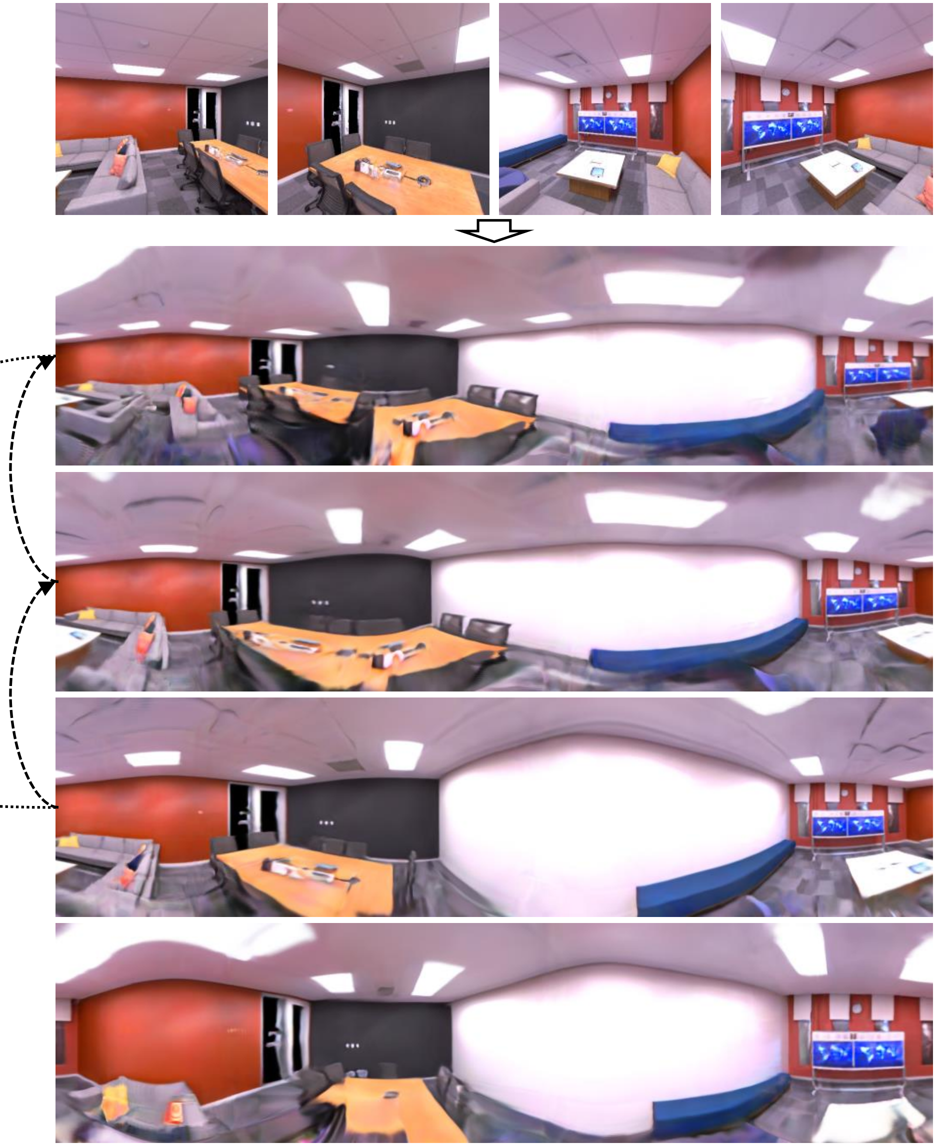
Yes, SinGRAF



01 | SinGRAF

First 3D-Aware GAN for Individual Scenes

Learning 3D generative radiance field from a few unposed images
Creating realistic variations of a single 3D scene
Realistic and diverse results with 3D view consistency

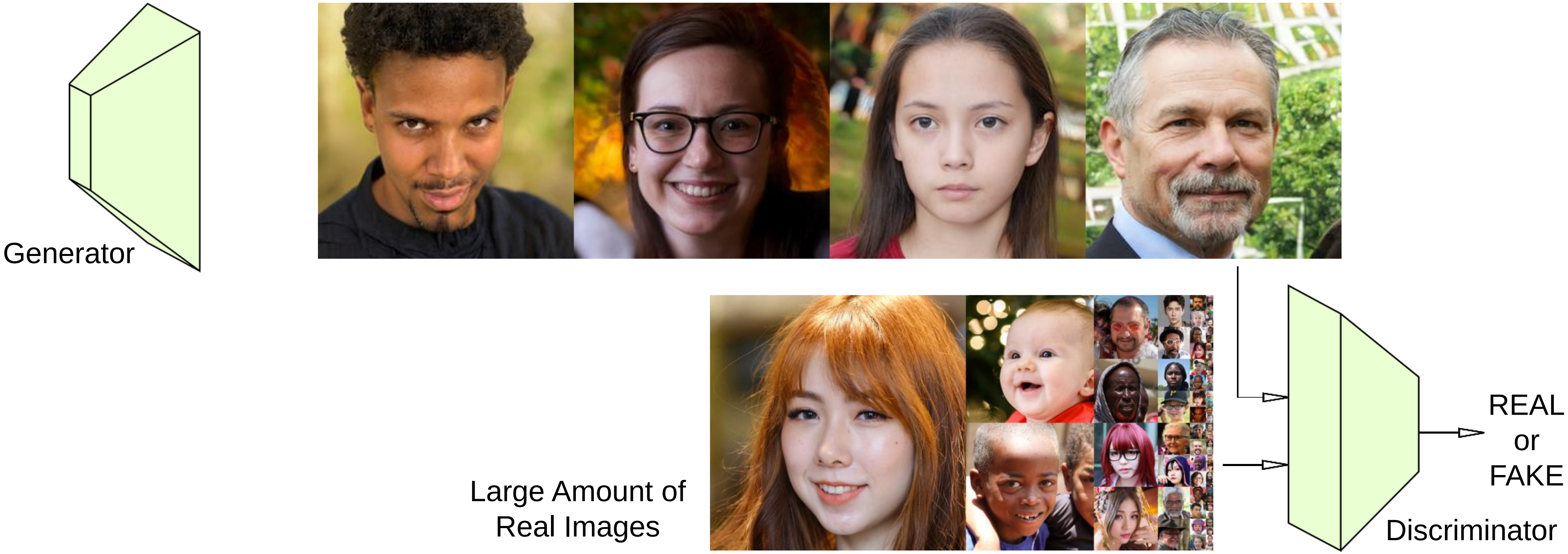


SinGRAF Result #1 → #2
with Latent Interpolation (Fixed Camera)

02 | Recent 3D-Aware GANs

Learning 3D Generative Radiance Field from a Set of Single-View Images

Projecting 3D generative radiance fields into 2D images using volume rendering
Supervised adversarially on 2D without any 3D supervision
High-quality images with 3D view consistency

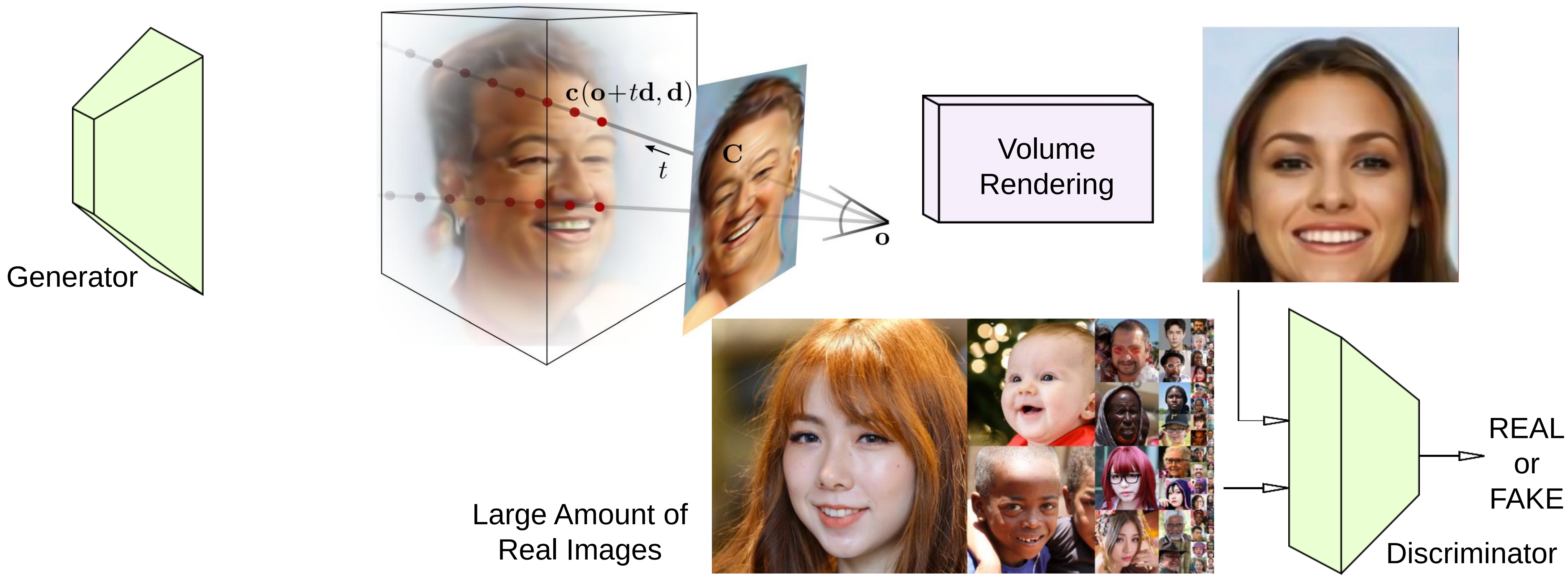


[StyleGAN2] Analyzing and Improving the Image Quality of StyleGAN, CVPR2020

02 | Recent 3D-Aware GANs

Learning 3D Generative Radiance Field from a Set of Single-View Images

Projecting 3D generative radiance fields into 2D images using volume rendering
Supervised adversarially on 2D without any 3D supervision
High-quality images with 3D view consistency

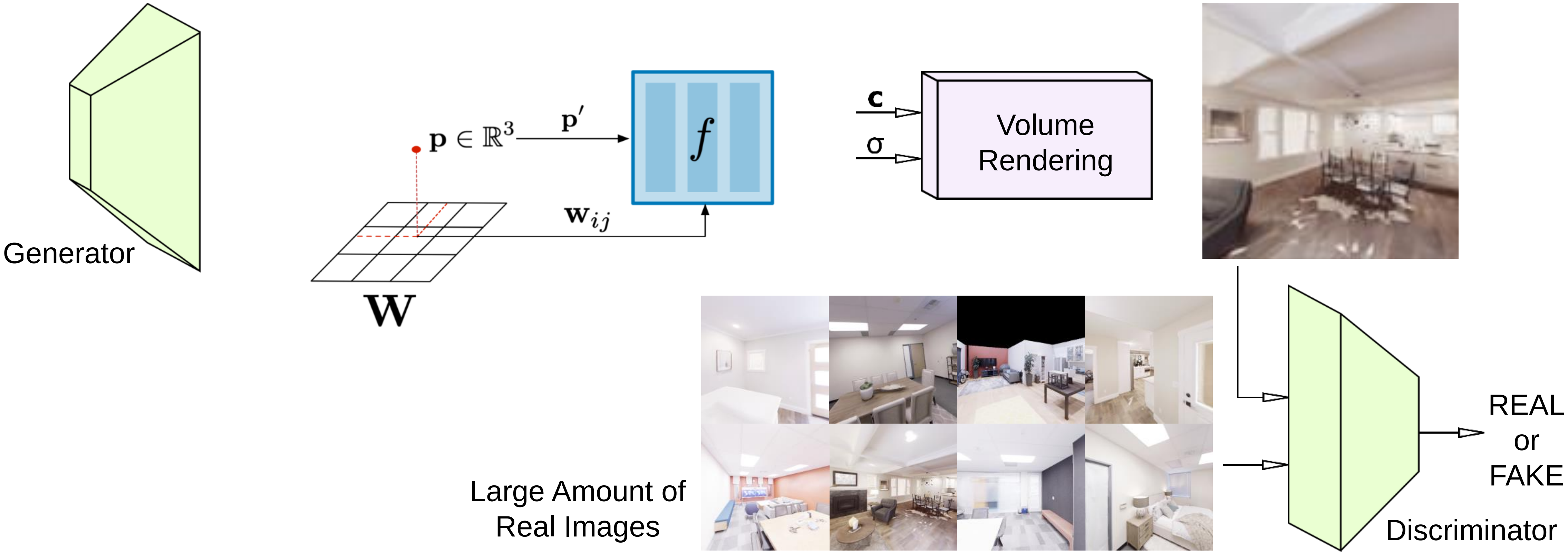


[piGAN] pi-GAN: Periodic Implicit Generative Adversarial Networks for 3D-Aware Image Synthesis, CVPR2021

02 | Recent 3D-Aware GANs

Learning 3D Generative Radiance Field from a Set of Single-View Images

Projecting 3D generative radiance fields into 2D images using volume rendering
Supervised adversarially on 2D without any 3D supervision
High-quality images with 3D view consistency

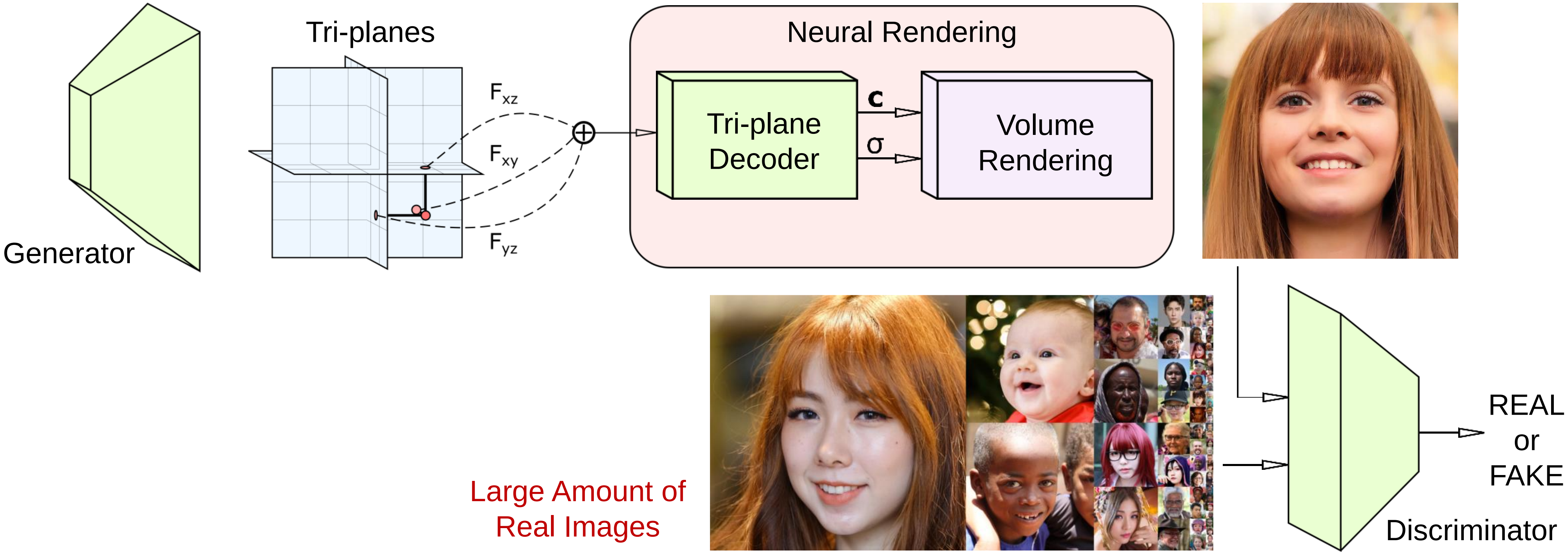


[GSN] Unconstrained Scene Generation with Locally Conditioned Radiance Fields, ICCV2021

02 | Recent 3D-Aware GANs

Learning 3D Generative Radiance Field from a Set of Single-View Images

Projecting 3D generative radiance fields into 2D images using volume rendering
Supervised adversarially on 2D without any 3D supervision
High-quality images with 3D view consistency

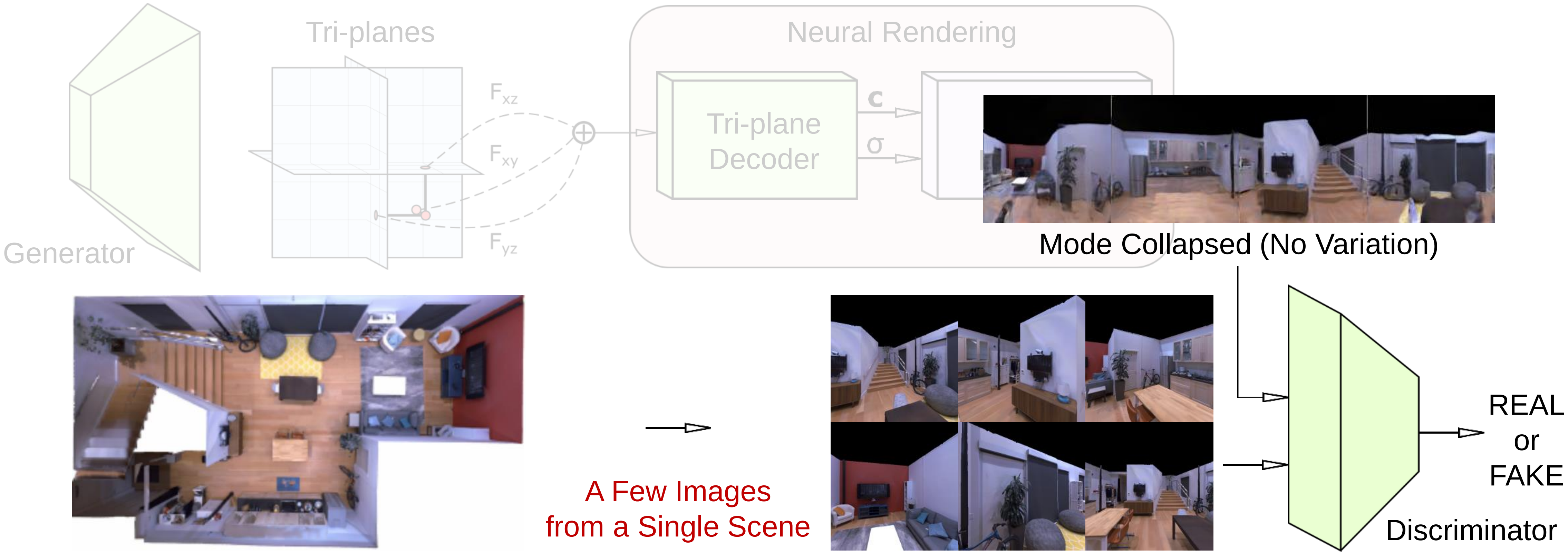


[EG3D] Efficient Geometry-aware 3D Generative Adversarial Networks, CVPR2022

02 | Recent 3D-Aware GANs

Learning 3D Generative Radiance Field from a Set of Single-View Images

Projecting 3D generative radiance fields into 2D images using volume rendering
Supervised adversarially on 2D without any 3D supervision
High-quality images with 3D view consistency



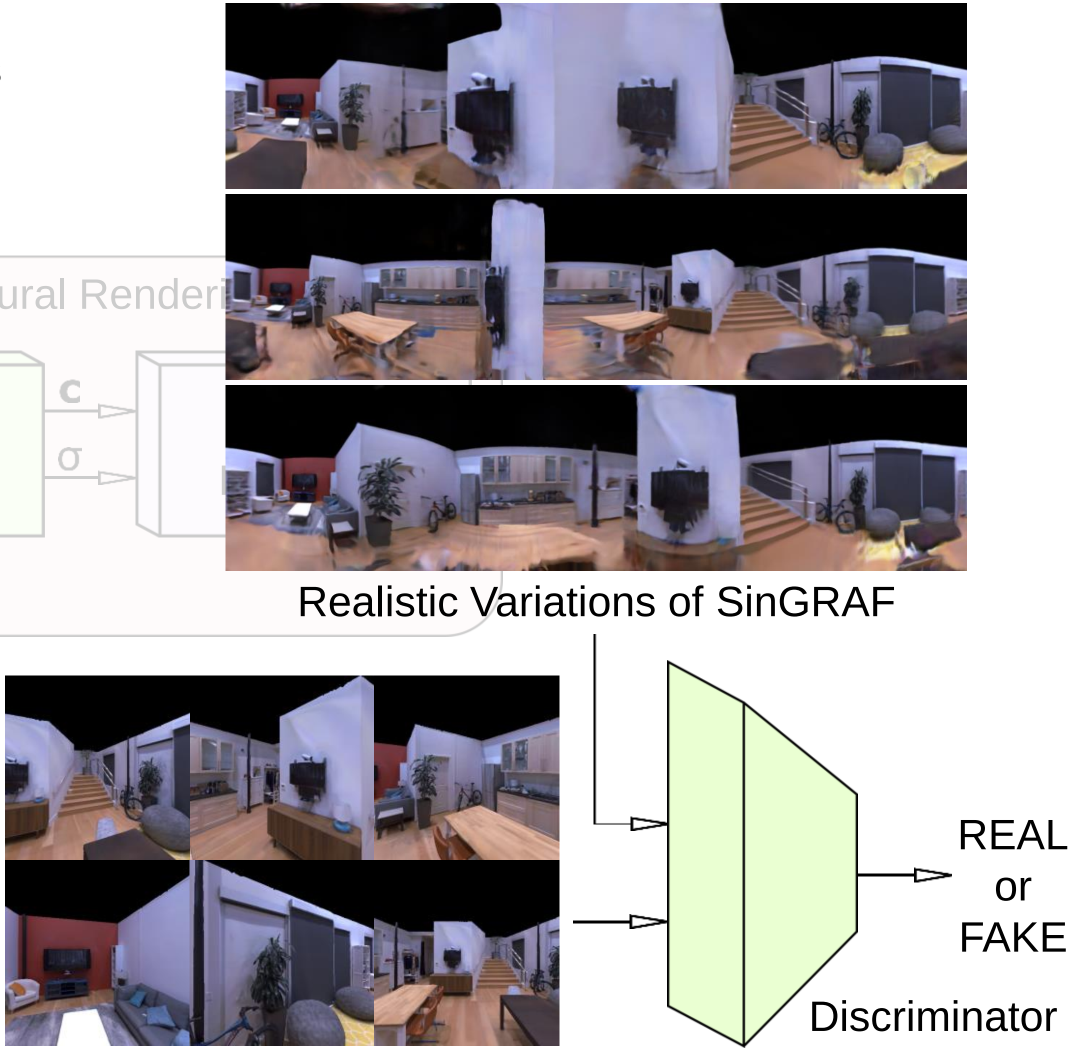
03 | Single Scene 3D GAN

SinGRAF

Learning 3D generative radiance field from a few unposed images
Creating realistic variations of a single 3D scene
Novel continuous-scale patch-based adversarial training

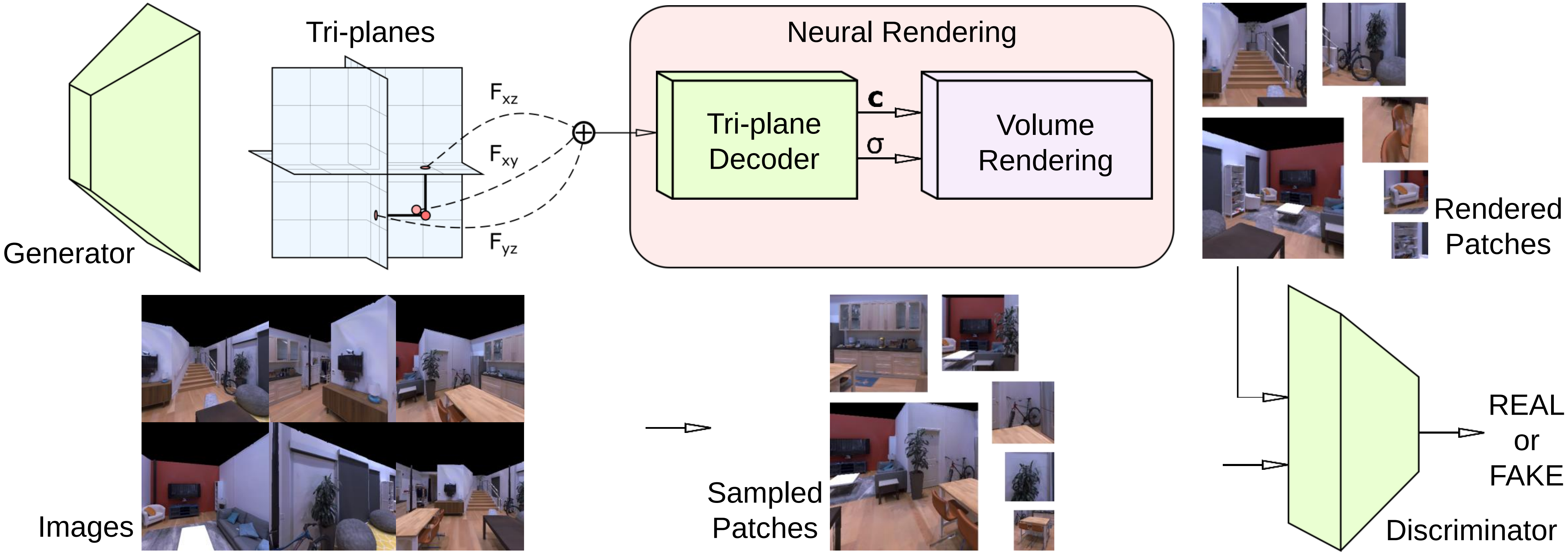
GSN SinGRAF

Latent Interpolation



Progressive-Scale Patch Discrimination

Volume rendering of patches with fixed resolution but different scales s
Random scale $s \sim U(s_{min}(t), s_{max}(t))$ with gradually decreasing along training epoch t
Discriminating w/ scale conditioning for enhanced quality



04 | Key Idea

WED-AM-027

Progressive-Scale Patch Discrimination

Volume rendering of patches with fixed resolution but different scales s
Random scale $s \sim U(s_{min}(t), s_{max}(t))$ with gradually decreasing along training epoch t
Discriminating w/ scale conditioning for enhanced quality

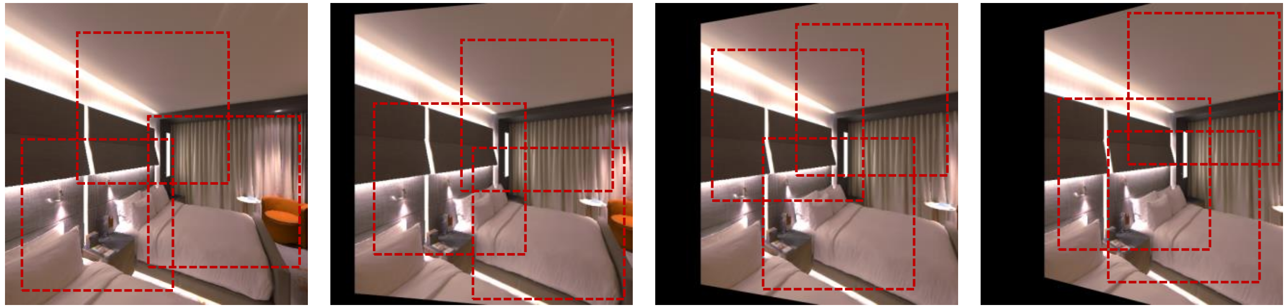
04 | Key Idea

Progressive-Scale Patch Discrimination

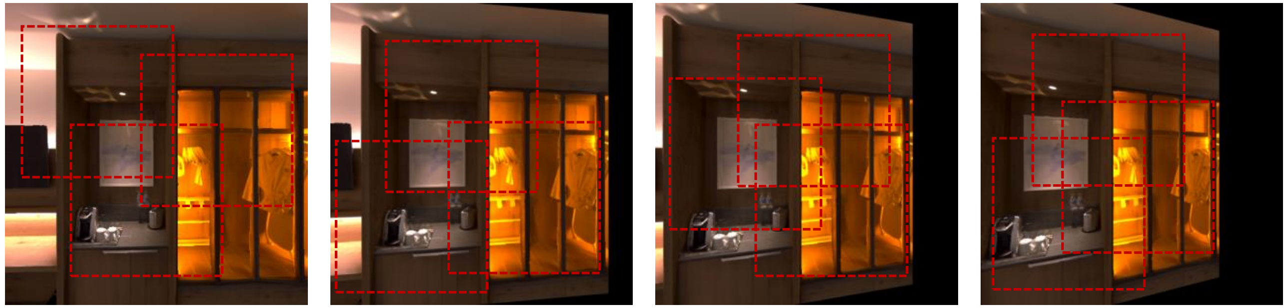
Volume rendering of patches with fixed resolution but different scales s
Random scale $s \sim U(s_{min}(t), s_{max}(t))$ with gradually decreasing along training epoch t
Discriminating w/ scale conditioning for enhanced quality

Perspective Augmentation

Imitating camera rotation with patch cropping



Original | Angle 5° | Angle 10° | Angle 15°



Original | Angle -5° | Angle -10° | Angle -15°

04 | Key Idea

Progressive-Scale Patch Discrimination

Volume rendering of patches with fixed resolution but different scales s
Random scale $s \sim U(s_{min}(t), s_{max}(t))$ with gradually decreasing along training epoch t
Discriminating w/ scale conditioning for enhanced quality

Perspective Augmentation

Imitating camera rotation with patch cropping

Camera Distribution Optimization

Using adversarial loss in the early training stage

	128×128	
	KID↓	Div.↑
full & half-scale patches	.183	.001
progressive patches	.046	$\boxed{.308}$
+ camera opt.	$\boxed{.037}$	.295
+ perspective aug.	.037	$\boxed{.335}$

05 | Results

Scene Generation Results

Scenes from Replica and Matterport3D
Various scenes with structural diversity & view consistency

Input Images from “hotel_0” Scene

GSN

SinGRAF

Latent Interpolation

05 | Results

Scene Generation Results

Scenes from Replica and Matterport3D
Various scenes with structural diversity & view consistency

Input Images from “apartment_0” Scene

GSN

SinGRAF

Latent Interpolation

Scene Generation Results

Scenes from Replica and Matterport3D
Various scenes with structural diversity & view consistency

Input Images from “castle” Scene

GSN **SinGRAF**

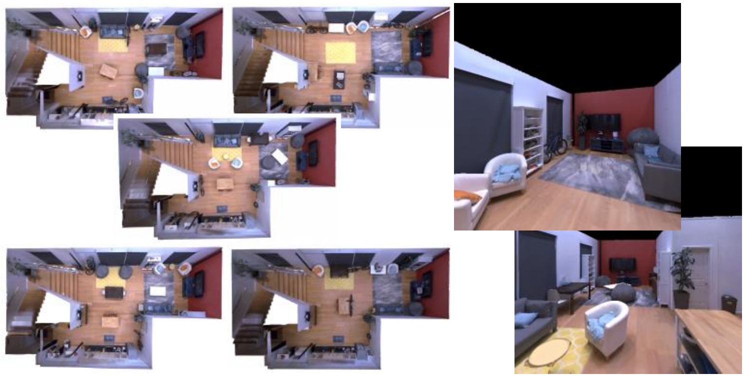
Latent Interpolation

05 | Results

Modeling Scene Dynamics

5 different configurations from Replica dataset
Robust for scene dynamics without any additional setting

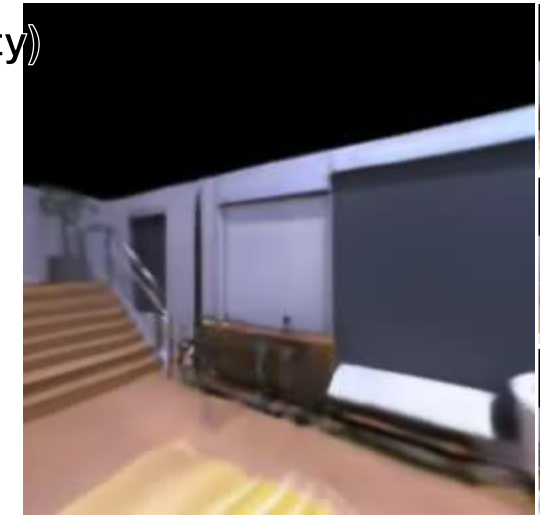
Input from “frl_apartment” Scenes

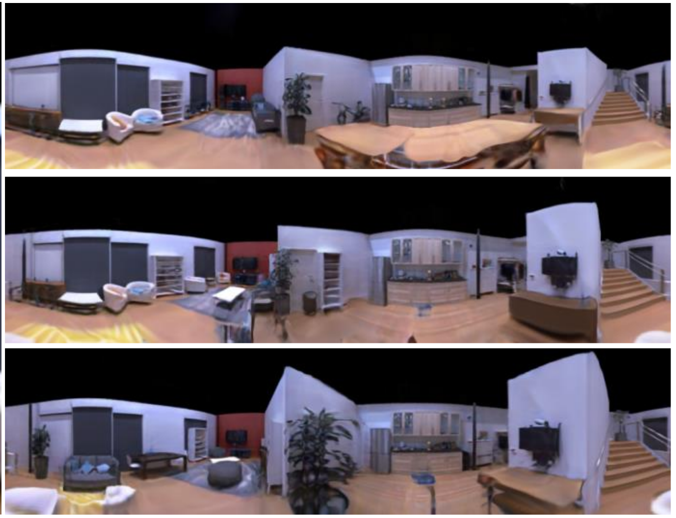


GSN

Mode collapsed (No Diversity)

SinGRAF





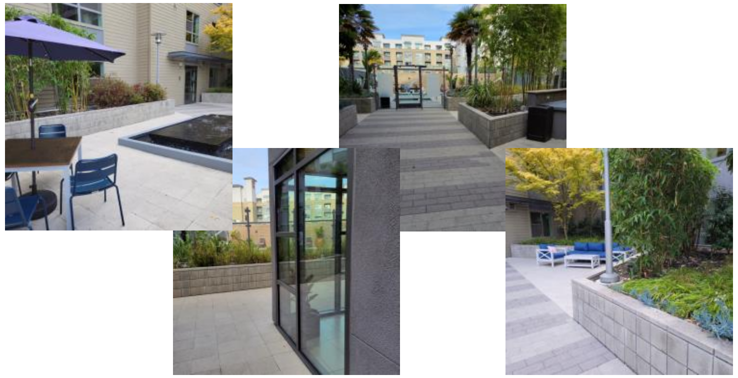
Latent Interpolation

05 | Results

Towards Casually-Captured Scenes

In-the-wild scene from consumer-level smartphone photographs
Potential for challenging outdoor scenes

Input from Captured Scene



GSN (Mode Collapsed)

SinGRAF

GSN

Mode collapsed (No Diversity)

SinGRAF

Latent Interpolation

Challenges

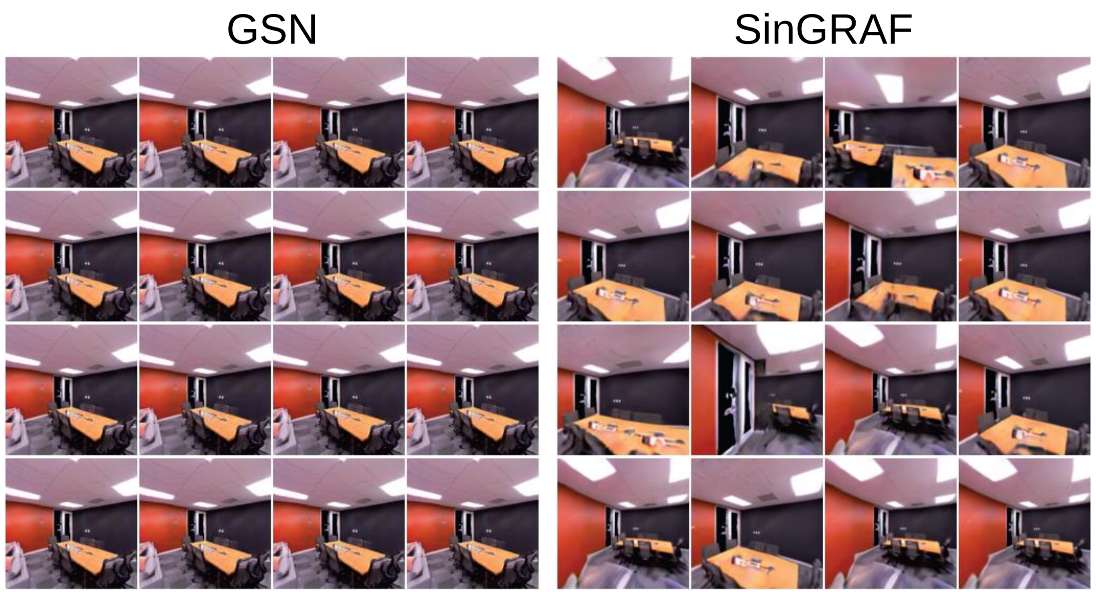
- Unknown camera intrinsic
- Camera lens distortion
- Auto exposure
- View-dependent reflection
- High-frequency textures

…

05 | Results

Quantitative Evaluation

[KID] Image quality with Kernel Inception Distance for sparsely sampled images
[Div.] Scene diversity with average pairwise LPIPS distance with sample images from fixed cameras
Outperforming for both realism and diversity



Visualization of Diversity Metric ("office_3")

	GSN (128^2)		SinGRAF (128^2)	
	KID↓	Div.↑	KID↓	Div.↑
office_3	.061	.001	**.044**	**.297**
hotel_0	.049	.012	**.037**	**.413**
apt.0	.069	.001	**.037**	**.401**
frl_apt.4	.052	.001	**.037**	**.335**
castle	**.050**	.001	.064	**.248**
office_0	.075	.001	**.053**	.001
dynamic	.089	.013	**.033**	**.298**

Quantitative Comparison

05 | Results

Quantitative Evaluation

[KID] Image quality with Kernel Inception Distance for sparsely sampled images
[Div.] Scene diversity with average pairwise LPIPS distance with sample images from fixed cameras
Outperforming for both realism and diversity

Failure Case

Detailed painting uniquely identifying patch locations
Possibility of unposed 3D reconstruction

Input Scene

SinGRAF with Mode Collapse

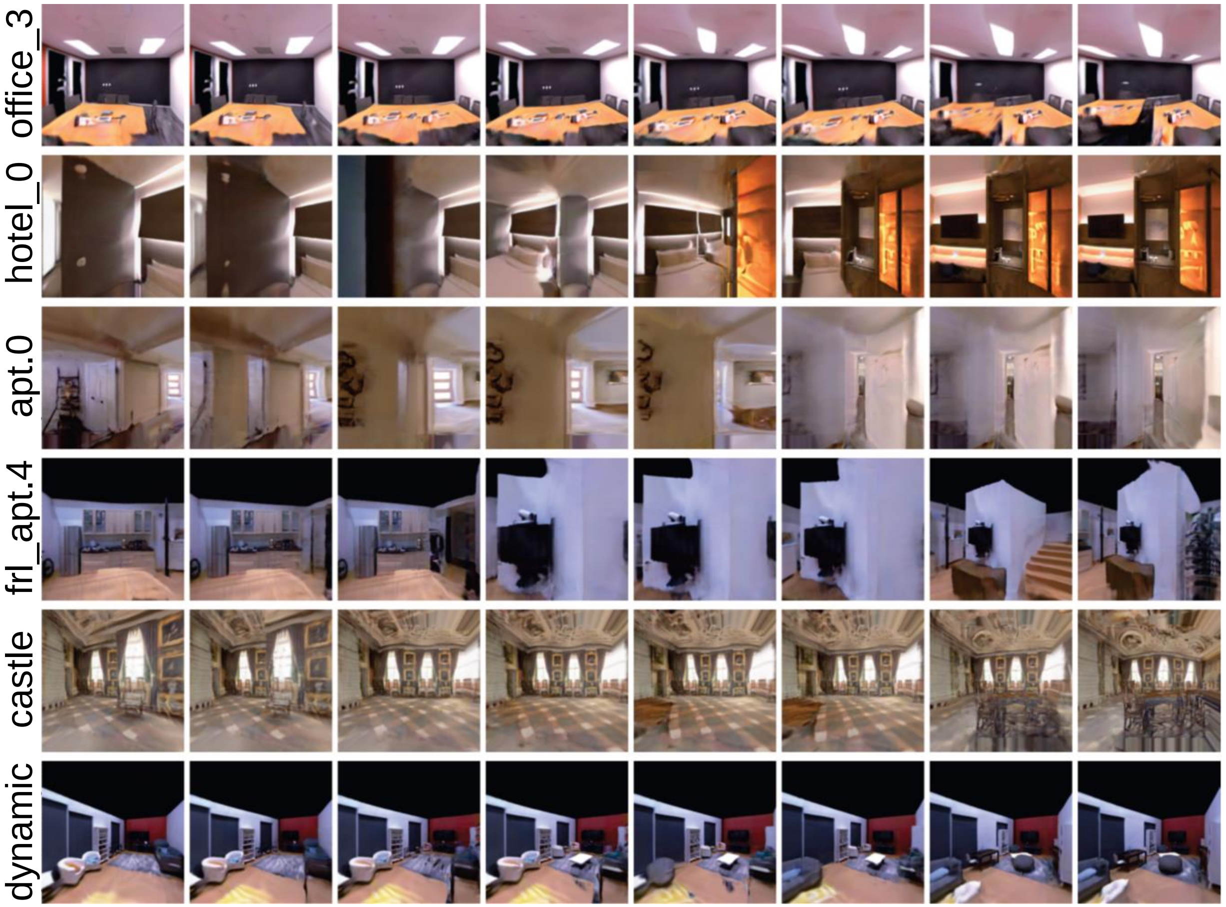
	GSN (128^2)		SinGRAF (128^2)	
	KID↓	Div.↑	KID↓	Div.↑
office_3	.061	.001	**.044**	**.297**
hotel_0	.049	.012	**.037**	**.413**
apt.0	.069	.001	**.037**	**.401**
frl_apt.4	.052	.001	**.037**	**.335**
castle	**.050**	.001	.064	**.248**
office_0	.075	.001	**.053**	$\boxed{.001}$
dynamic	.089	.013	**.033**	**.298**

Quantitative Comparison

06 | Conclusion

SinGRAF: Learning a 3D Generative Radiance Field for a Single Scene

First 3D-aware GAN from a few unposed images of a single 3D scene
Creating realistic variations w/ 3D view consistency
Novel continuous-scale patch-based training



Visualization of Latent Interpolation

Limitation

Limited predictability or controllability
Expensive per-scene training

Discussion & Future Work

Variational 3D reconstruction from unposed images
More in-the-wild and highly dynamic scenes
Advanced controllability

SinGRAF: Learning a 3D Generative Radiance Field for a Single Scene

WED-AM-027

Thank You!

Contact Information:

Minjung Son, PhD (SAIT)

Email: minjungs.son@samsung.com

Project Page: http://www.computationalimaging.org/publications/singraf