EFFICIENT GEOMETRY-AWARE 3D GENERATIVE ADVERSARIAL NETWORKS

Eric R. Chan, Connor Z. Lin, Matthew A. Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas Guibas, Jonathan Tremblay, Sameh Khamis, Tero Karras, and *Gordon Wetzstein*

Stanford University and NVIDIA, 2021

GOAL

Unsupervised generative model for high-quality multi-view consistent imagery and 3D shapes, given single-view 2D photographs

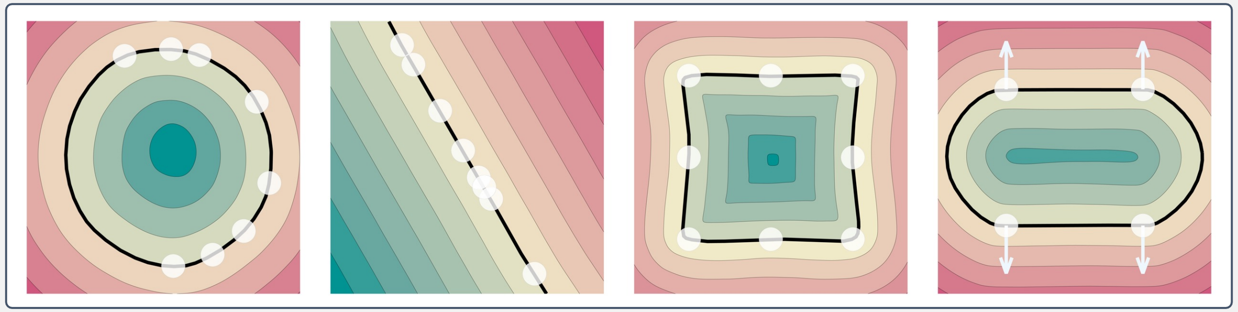
EG3D : expressive hybrid explicit-implicit network architecture

RELATED WORK

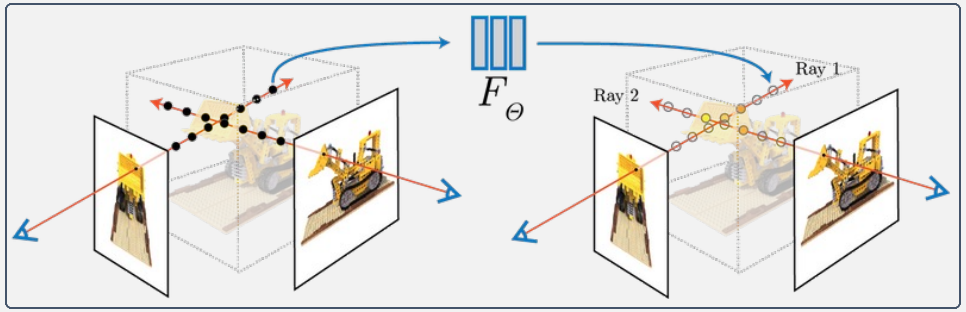
Neural scene representation and rendering

Learning 3D scene representations from 2D multi-view images via neural rendering : SDFs, Implicit Fields or Neural Network Level Sets

Neural rendering : *NeRF*, *FastNeRF*, *DONeRF*, *PlenOctrees* or *SDFDiff*



IGR : Implicit Geometric Regularization for Learning Shapes



NeRF : Representing Scenes as Neural Radiance Fields for View Synthesis

RELATED WORK

Explicit vs implicit 3D representations

METHOD	REPRESENTATION	PROS	CONS
Explicit	Voxel grid	Fast to evaluate	Heavy memory overhead (not scalable)
Implicit	Fully connected layer (MLP)	Memory efficient continuous function	Slow forward pass

Generative 3D-aware image synthesis

2D photorealistic image synthesis : *StyleGAN2*

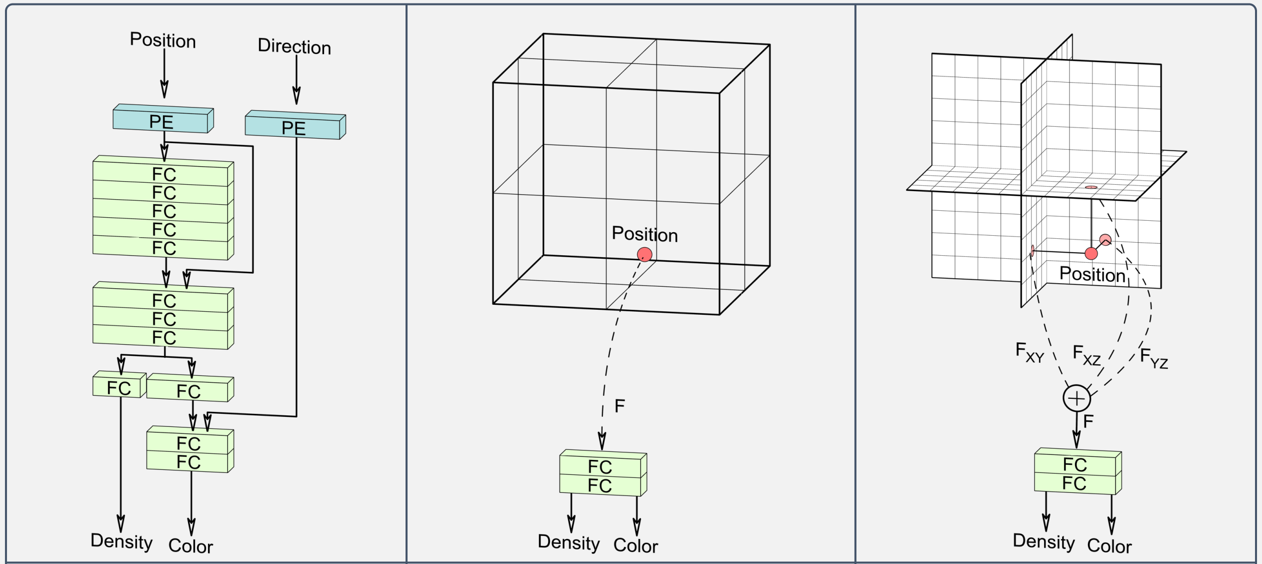
3D GANs : voxel-based extensions of classical 2D CNN-based generators, e.g., *PlatonicGAN*, *HoloGAN* or *BlockGAN*

TRI-PLANE HYBRID 3D REPRESENTATION

Align explicit features along three axis-aligned orthogonal feature planes

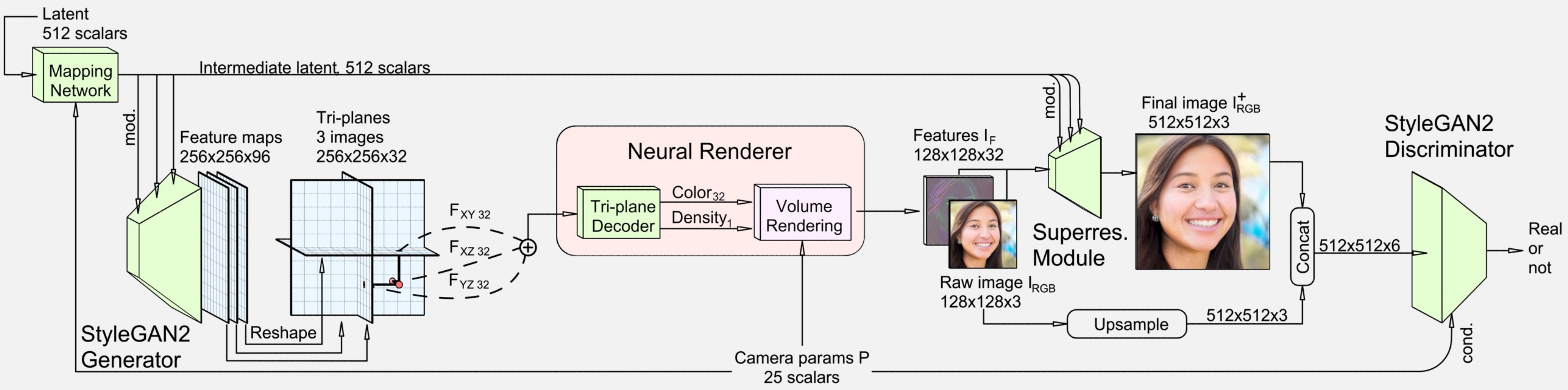
Each feature plane : N x N x C

Query 3D position, retrieve feature vectors, sum vectors and pass-through lightweight decoder



NeRF (Implicit)	Voxels (Explicit or Hybrid)	*EG3D* (Hybrid)

3D GAN FRAMEWORK



3D GAN FRAMEWORK

CNN generator backbone and rendering

Edit *StyleGAN2* to output 96 channel images. Split and reshape into three 32-channel planes

Decoder = single hidden layer MLP with 64 units + softplus activation

Super-resolution

Interactive framerate : volume rendering @ spatial resolution 128^2 + up-sampling

Dual discrimination

Concatenate bilinearly up-sampled I_{RGB} and I^{+}_{RGB} into a 6-channel image

Same for real images : concatenate image with blurred copy of itself

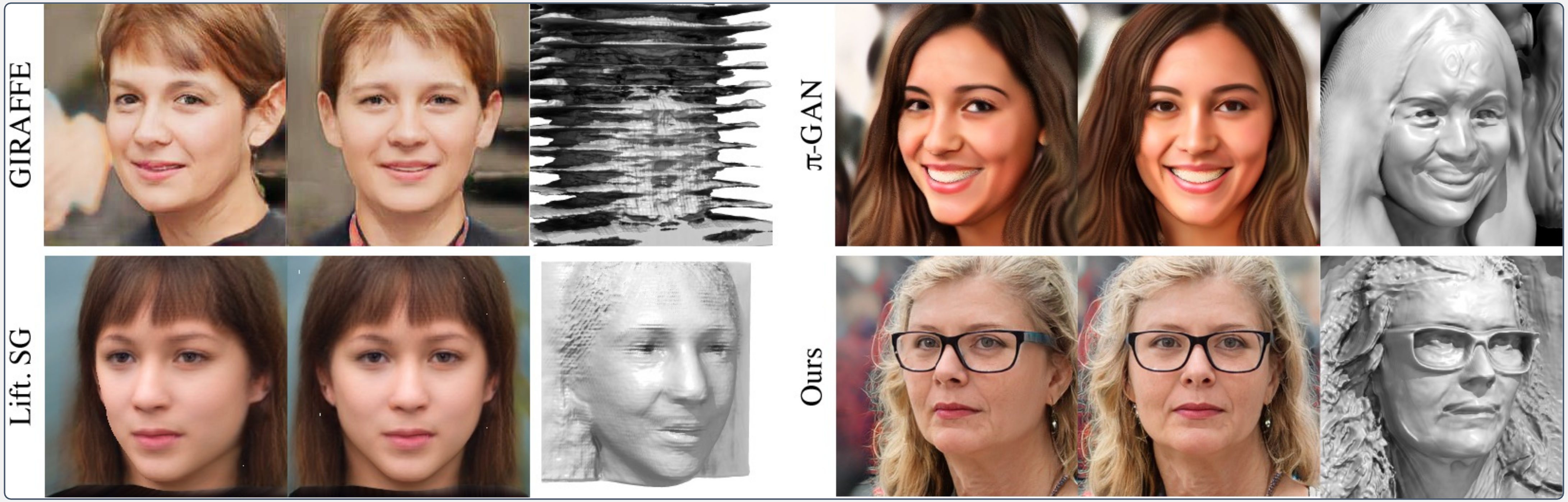
Encourages consistency between low resolution render and up-sampled image

Generator and discriminator pose conditioning

EXPERIMENTS AND RESULTS

Datasets : *FFHQ* and *AFHQv2 Cats*

Qualitative results



EXPERIMENTS AND RESULTS

EXPERIMENTS AND RESULTS

Qualitative effect of super-resolution network

EXPERIMENTS AND RESULTS

Qualitative effect of dual discrimination

EXPERIMENTS AND RESULTS

Quantitative results

	FFHQ FID↓	FFHQ ID↑	FFHQ Depth↓	FFHQ Pose↓	Cats FID↓
GIRAFFE 256^2	31.5	0.64	0.94	.089	16.1
π-GAN 128^2	29.9	0.67	0.44	.021	16.0
Lift. SG 256^2	29.8	0.58	0.40	.023	—
Ours 256^2	**4.8**	0.76	**0.31**	**.005**	**3.88**
Ours 512^2	**4.7**	**0.77**	0.39	**.005**	**2.77**†

Metrics evaluation

Res.	GIRAFFE	π-GAN	Lift. SG	Ours	Ours + TC
256^2	181	5	51	27	36
512^2	161	1	—	26	35

Runtime analysis (FPS on single RTX 3090)

	FID ↓	FACS Smile Std. ↓
Naive model	5.5	0.069
+ DD	6.5	0.054
+ DD, GPC (ours)	**4.7**	**0.031**

Ablation study

EXPERIMENTS AND RESULTS

Applications : latent interpolation

EXPERIMENTS AND RESULTS

Applications : image inversion and reconstruction

REFERENCES

Gropp, Amos, et al. “Implicit geometric regularization for learning shapes.” arXiv preprint arXiv:2002.10099 (2020).

Mildenhall, Ben, et al. “Nerf: Representing scenes as neural radiance fields for view synthesis.” European conference on computer vision. Springer, Cham, 2020.

Garbin, Stephan J., et al. “Fastnerf: High-fidelity neural rendering at 200fps.” *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021.

Neff, Thomas, et al. “DONeRF: Towards Real-Time Rendering of Compact Neural Radiance Fields using Depth Oracle Networks.” Computer Graphics Forum. Vol. 40. No. 4. 2021.

Yu, Alex, et al. “Plenoctrees for real-time rendering of neural radiance fields.” Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

Jiang, Yue, et al. “Sdfdiff: Differentiable rendering of signed distance fields for 3d shape optimization.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

Karras, Tero, et al. “Analyzing and improving the image quality of stylegan.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.

Henzler, Philipp, Niloy Mitra, and Tobias Ritschel. “Escaping plato's cave using adversarial training: 3d shape from unstructured 2d image collections.” Proceedings of the International Conference on Computer Vision 2019 (ICCV 2019). Vol. 2019. IEEE, 2019.

Nguyen-Phuoc, Thu, et al. “Hologan: Unsupervised learning of 3d representations from natural images.” Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.

Nguyen-Phuoc, Thu H., et al. “Blockgan: Learning 3d object-aware scene representations from unlabelled images.” Advances in Neural Information Processing Systems 33 (2020): 6767-6778.

Niemeyer, Michael, and Andreas Geiger. “Giraffe: Representing scenes as compositional generative neural feature fields.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.

Chan, Eric R., et al. “pi-gan: Periodic implicit generative adversarial networks for 3d-aware image synthesis.” Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.

Shi, Yichun, Divyansh Aggarwal, and Anil K. Jain. “Lifting 2D StyleGAN for 3D-Aware Face Generation.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.