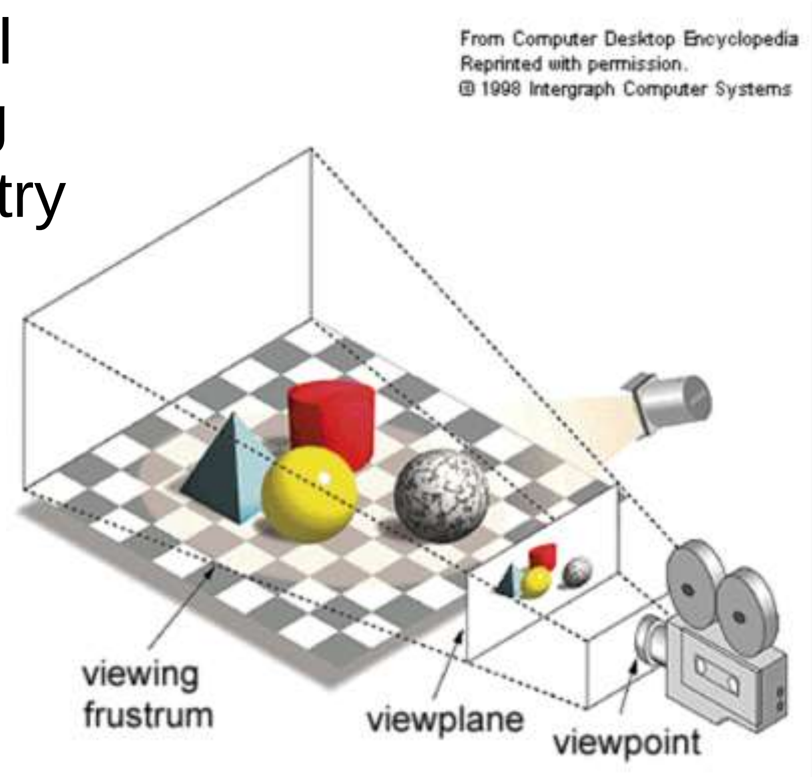
Lecture 20: Advanced Topics: Neural Rendering and Modeling

Lan Xu
SIST, ShanghaiTech
Fall, 2023

What's Rendering

3D scene
- Material
- Lighting
- Geometry
- …



Camera Def.
- Intrinsics
- Focal length
- Principal point
- …

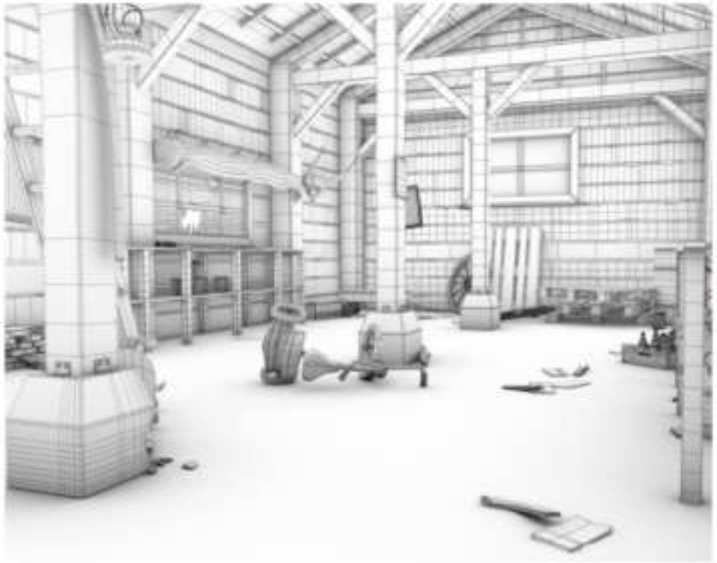
View point
- Extrinsic
- 6DoF(rot + trans)
- …

Photo-realistic Image Synthesis

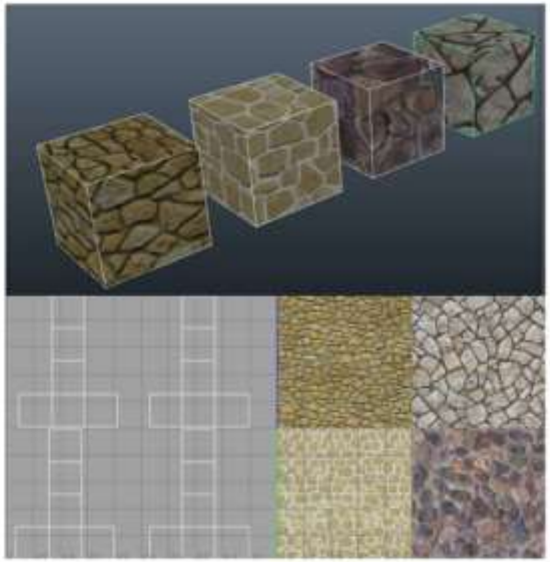
- The Rendering Equation [Kajiya 86]

$$L_o(\mathbf{x}, \omega_o, \lambda, t) = L_e(\mathbf{x}, \omega_o, \lambda, t) + \int_\Omega f_r(\mathbf{x}, \omega_i, \omega_o, \lambda, t)\, L_i(\mathbf{x}, \omega_i, \lambda, t)\, (\omega_i \cdot \mathbf{n})\, \mathrm{d}\,\omega_i$$

Need 3D Content for Rendering



Geometry



Textures

Material & Lighting

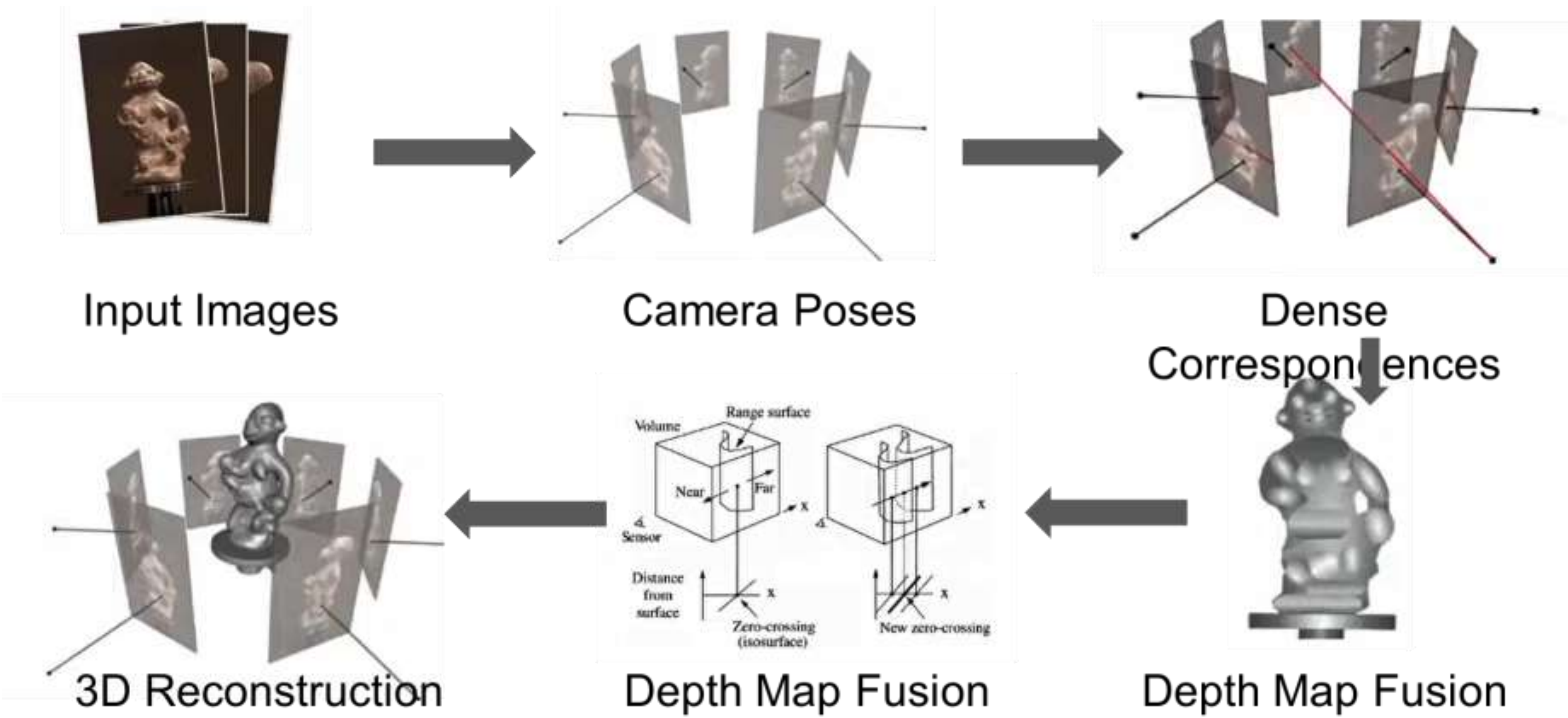
Computer Vision for Reconstruction

- Building Rome in a day

Sameer Agarwala , Yasutaka Furukawaa , Noah Snavely, Ian Simonb , Brian Curless, Steven M. Seitz and Richard Szeliski, *Communications of the ACM, 2011*

Traditional Reconstruction

- Recall traditional reconstruction pipeline
- Take **geometry** as example



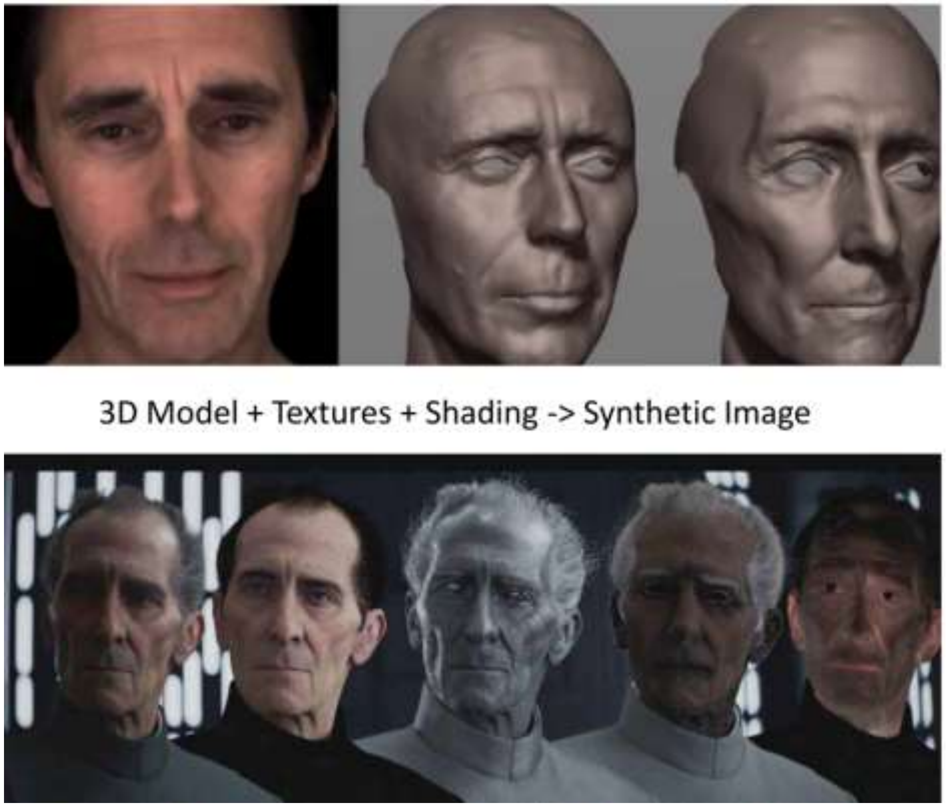
3D Digitization

- Various tastes for CV and CG

Computer Graphics

Computer Vision

Traditional Graphics v.s Deep Learning



Generative Adversarial Networks

Discriminator loss

$$J^{(D)} = -\frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}} \log D(\boldsymbol{x}) - \frac{1}{2}\mathbb{E}_{\boldsymbol{z}} \log\left(1 - D\left(G(\boldsymbol{z})\right)\right)$$

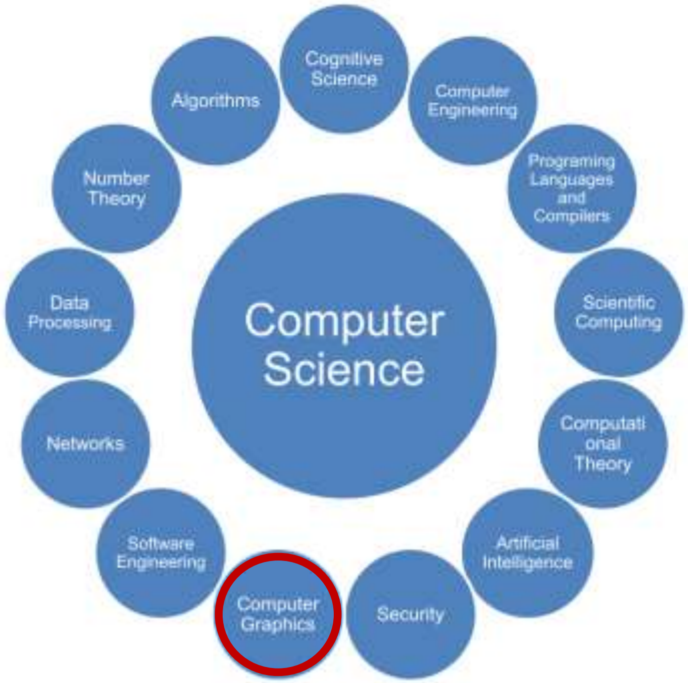
Generator loss

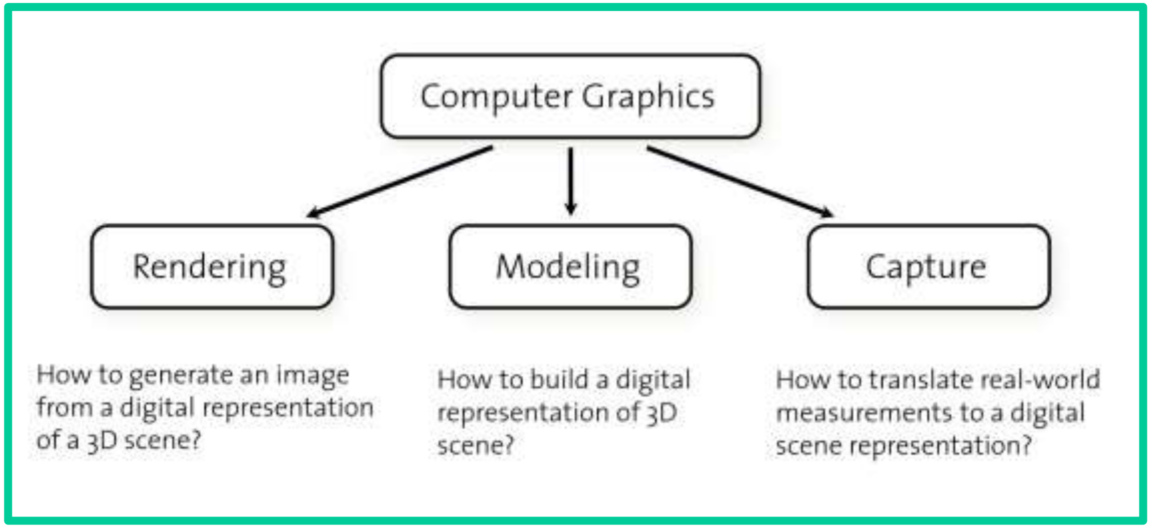
$$J^{(G)} = -J^{(D)}$$

7

Computer Graphics?

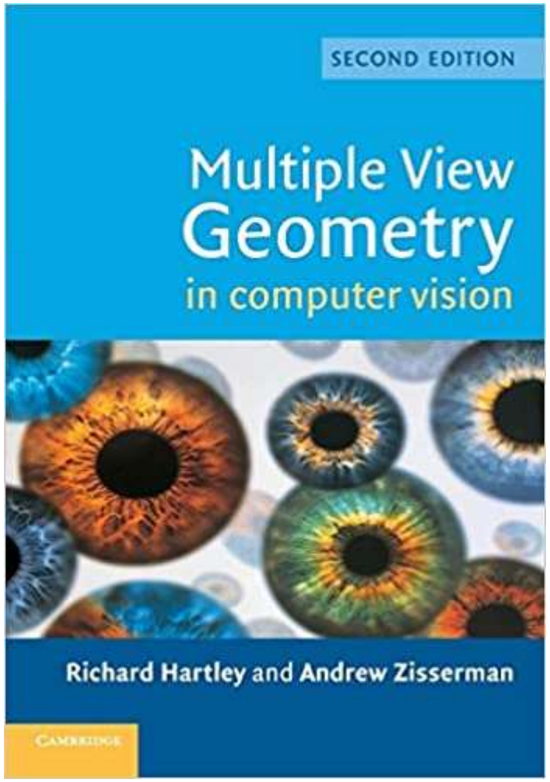
- Both inverse process and forward process
- From **real world** to **virtual representation**, then to **vivid rendering**





Recall traditional pipeline

- Systematic knowledge with representative methods



Cambridge University Press, March 2004.

Building Rome in a day, Sameer Agarwala , Yasutaka Furukawaa , Noah Snavely, Ian Simonb , Brian Curless, Steven M. Seitz and Richard Szeliski, *Communications of the ACM, 2011*

Recall traditional pipeline

- Well-developed capture devices across the world
- Two riding horses: muti-view/photometric Stereos

Multi-camera Dome

Multi-View Stereo

Light stage

Photometric Stereo

Recall traditional pipeline

- Various Applications
- Yet time-consuming → artist in-the-loop

architecture

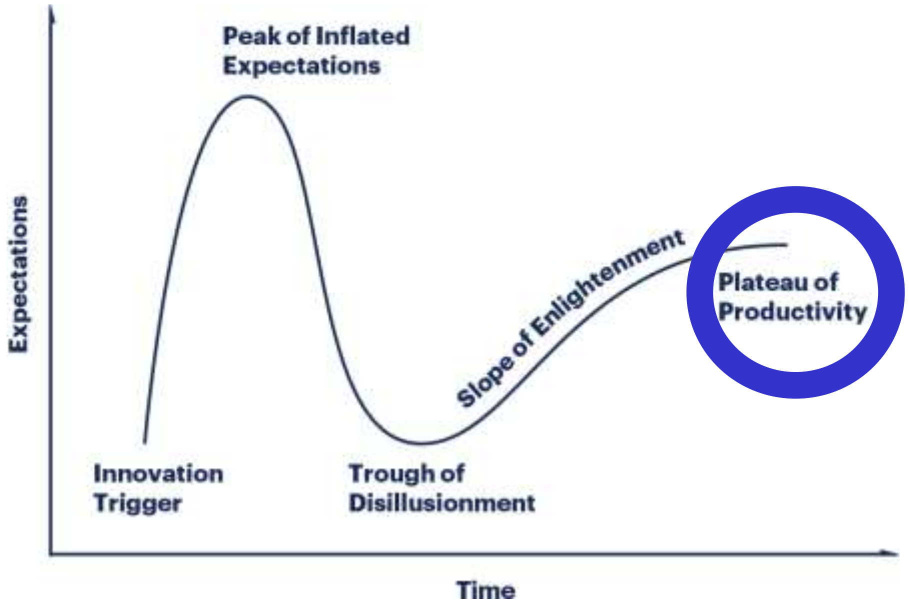
digital twin

Movie

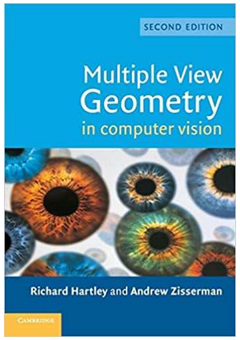
E-commerce

Recall traditional pipeline

- Traditional Pipeline: mature in the past decades



2021: Epic Games buys Capturing Reality

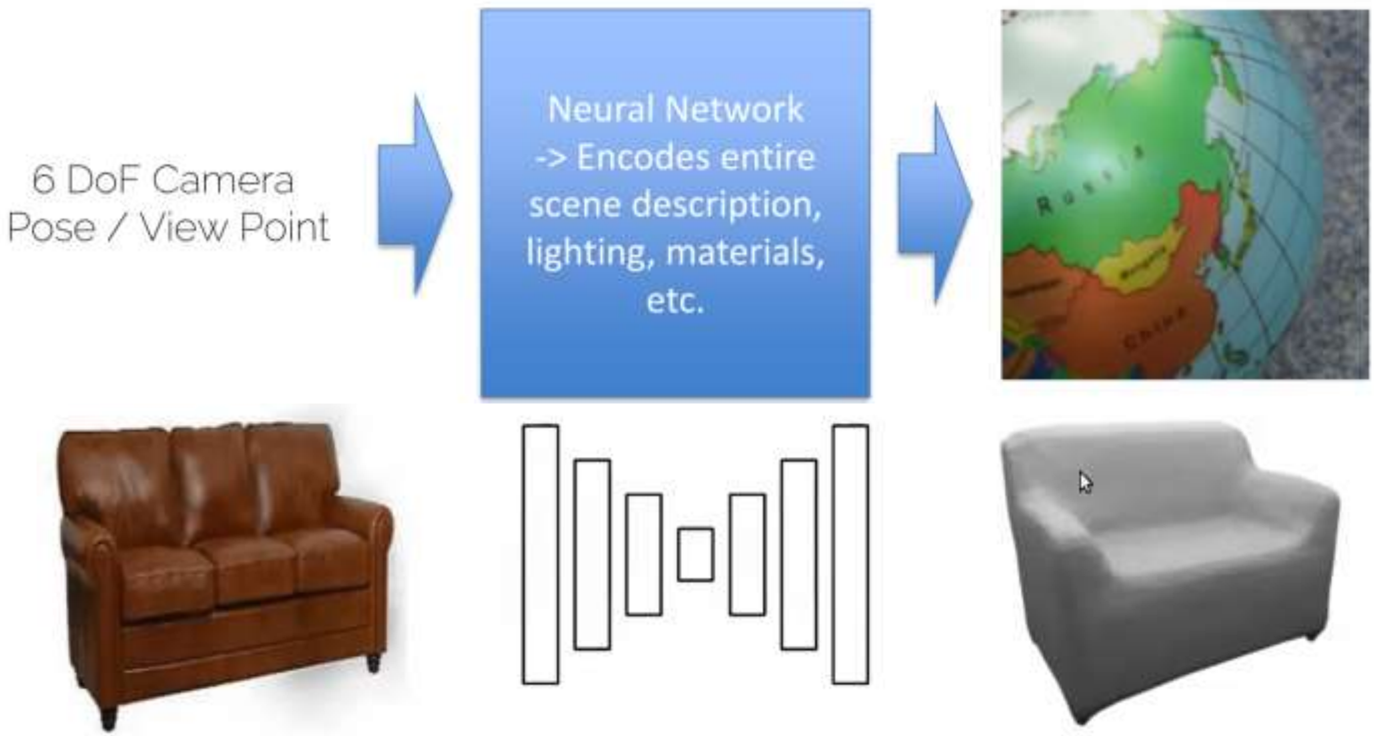


Symbol

2004

Idea of Neural Rendering

- Neural reconstruction from 2D images directly
- Novel view point synthesis



Input Images Neural Networks Reconstruction/Rendering

Idea of Neural Rendering

- Definition: Deep neural network for **image or video generation** that enable **explicit or implicit control** of **scene properties**

1)
Generative networks that synthesis raw pixel output

2)
Output contrillable by interpretable paras or by video/audio input.

3)
Illumination, camera para., pose, geometry, appearance, or semantic structure controlable

Idea of Neural Rendering

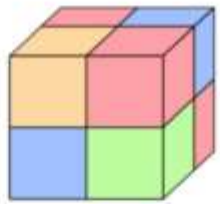
- Neural Rendering Taxonomy

- Required Data (image, video, mesh, etc.)
- Controllable Parameters (camera, pose, lighting, etc.)
- Multi-modal Synthesis
- Temporal Coherence
- **Computer Graphics Module**
- Generality
- …

Neural Representation History

Scene Representation

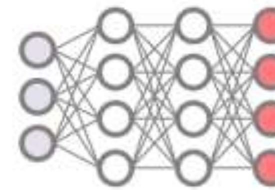
Multi-Plane Images



Voxelgrids

Image-based

Point Clouds



Implicit Function

Renderer

(Alpha) compositing

Volumetric Ray-based

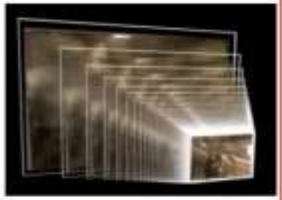
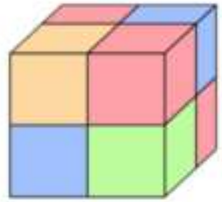
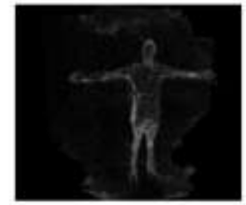
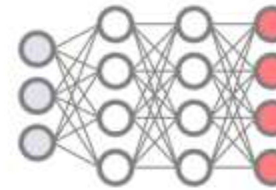
Rasterization

Splatting

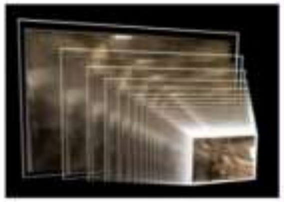
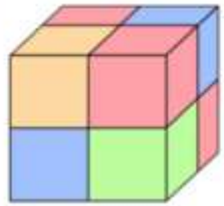
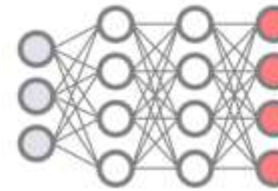
Sphere-Traced Volumetric

Scene Representation Differentiable Renderer

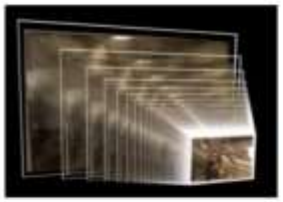
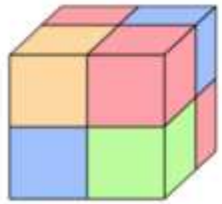
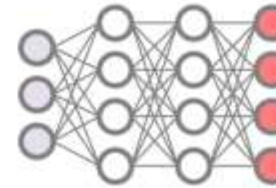
Neural Representation History

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes				
Cons	Only 2.5D Size				

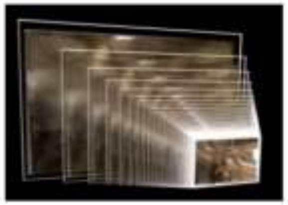
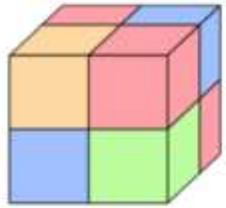
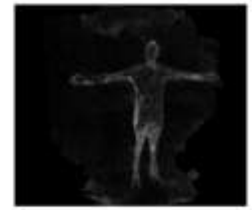
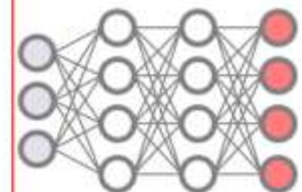
Neural Representation History

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality			
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$			

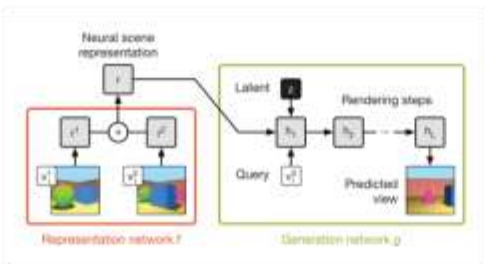
Neural Representation History

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality		
Cons	Only 2.5D Size	No reconstruction priors Memory O(n^3)	Requires good SFM No compact representation		

Neural Representation History

Scene Representation	 Multi-Plane Images	 Voxelgrids	 Image-based	 Point Clouds	 Implicit Function
Renderer	(Alpha) compositing	Volumetric Ray-based	Rasterization	Splatting	Sphere-Traced Volumetric
Pros	Fast rendering High quality Generalizes	"True 3D" High quality	High quality	High quality	
Cons	Only 2.5D Size	No reconstruction priors Memory $O(n^3)$	Requires good SFM No compact representation	Requires good SFM	

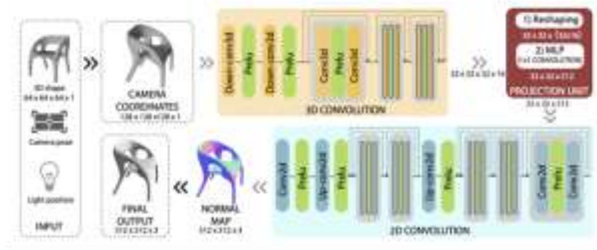
Neural Representation History



Generative Query Networks
[Eslami et al. 2018]

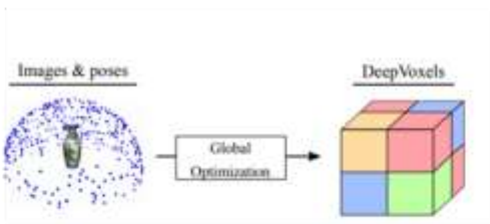
[Flynn et al., 2016; Zhou et al., 2018b; Mildenhall et al. 2019]

Multiplane Images (MPIs)

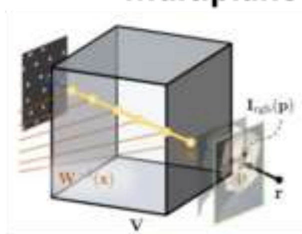


RenderNet [Nguyen-Phuoc et al. 2018]

Voxel Grids + CNN decoder



DeepVoxels
[Sitzmann et al. 2019]



Neural Volumes
[Lombardi et al. 2019]

Voxel Grids + Ray Marching

SRN
[Sitzmann et al. 2019b]

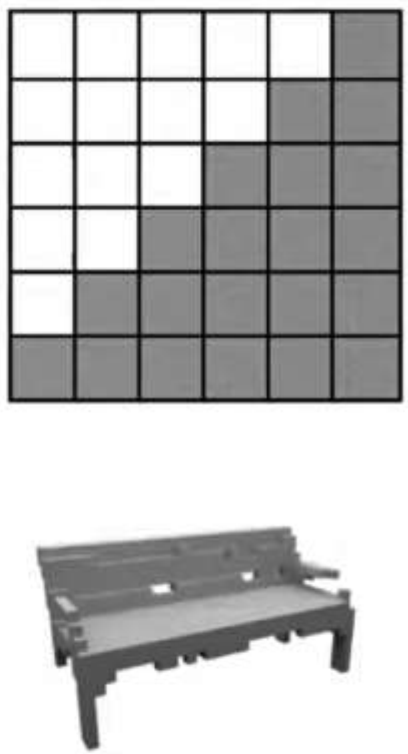
NeRF
[Mildenhall et al. 2020]

IDR
[Yariv et al. 2020]

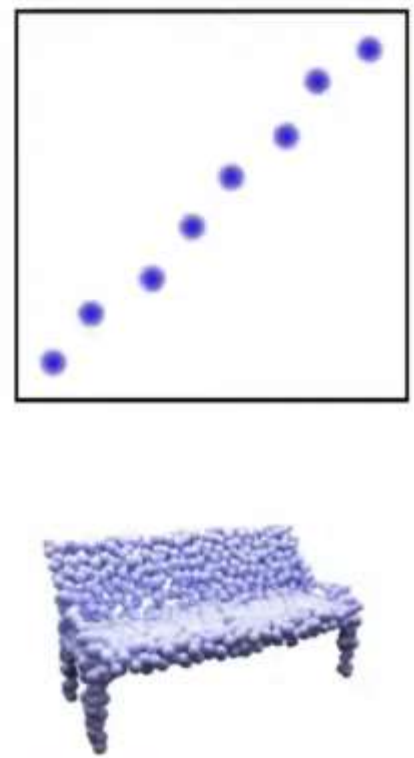
Implicit Fields

Neural Representation History

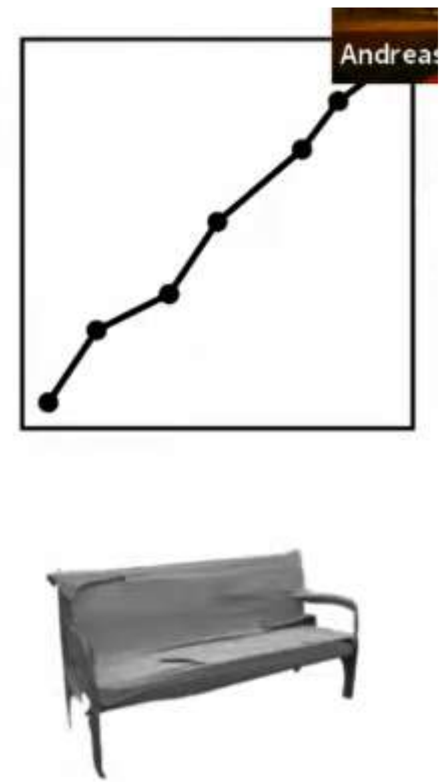
- The trend in recent years: from explicit to implicit
- Various explicit representations



Voxels



Points



Meshes

Neural Implicit Representation

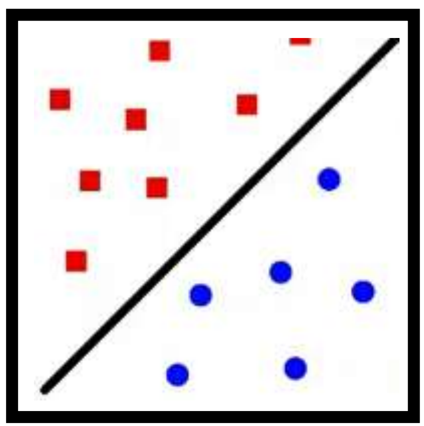
- The trend in recent years: from explicit to implicit
- Do not represent the 3D attribute explicitly
- Instead, consider a continues manifold space, with continues implicit attributes at the coordinates of the space

$$f_\theta : \mathbb{R}^3 \times \mathcal{X} \to [0, 1]$$

3D Location

Condition (eg, Image)

Occupancy Probability



Neural Implicit Representation

- NeRF: Neural Radiance Field
- 1) Color + Density;   2)- Positional Encoding + Volume Rendering

Representing Scenes as Neural Radiance Fields for View Synthesis, Mildenhall et al., *ECCV 2020 Oral - Best Paper Honorable Mention*

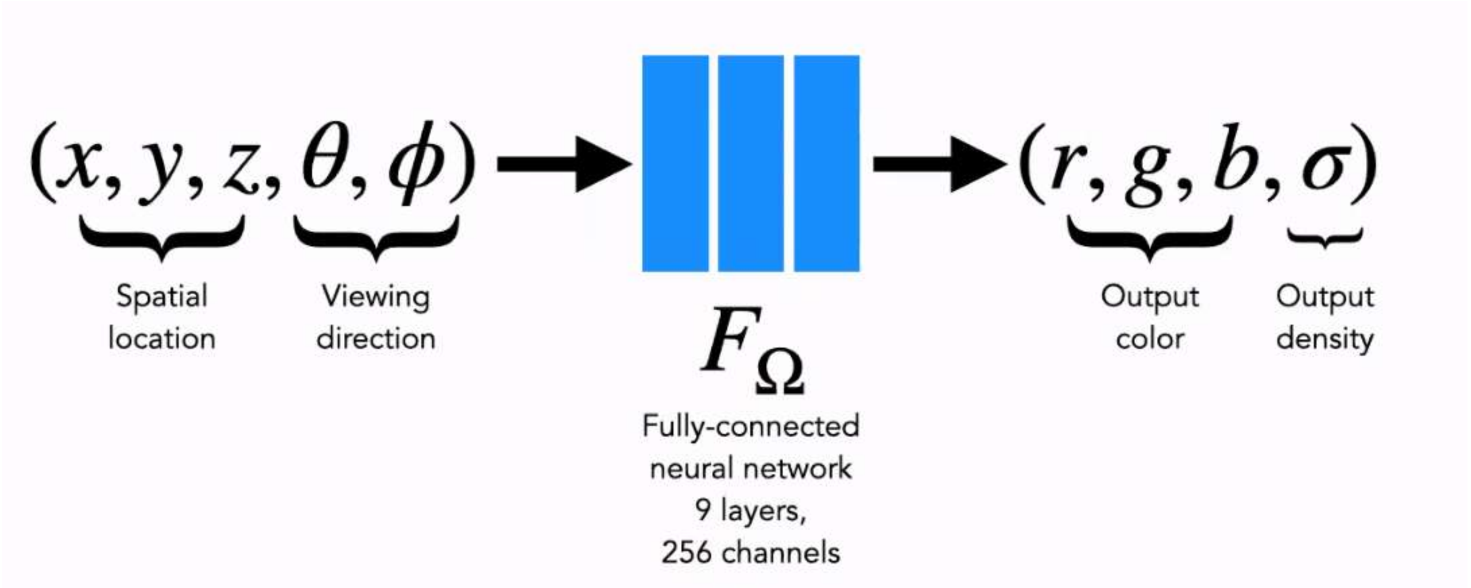
Neural Implicit Representation

- NeRF: Neural Radiance Field
- 1) Color + Density;   2)- Positional Encoding + Volume Rendering

Representing Scenes as Neural Radiance Fields for View Synthesis, Mildenhall et al., *ECCV 2020 Oral - Best Paper Honorable Mention*

Neural Radiance Field

- A scene is a continues 5D function

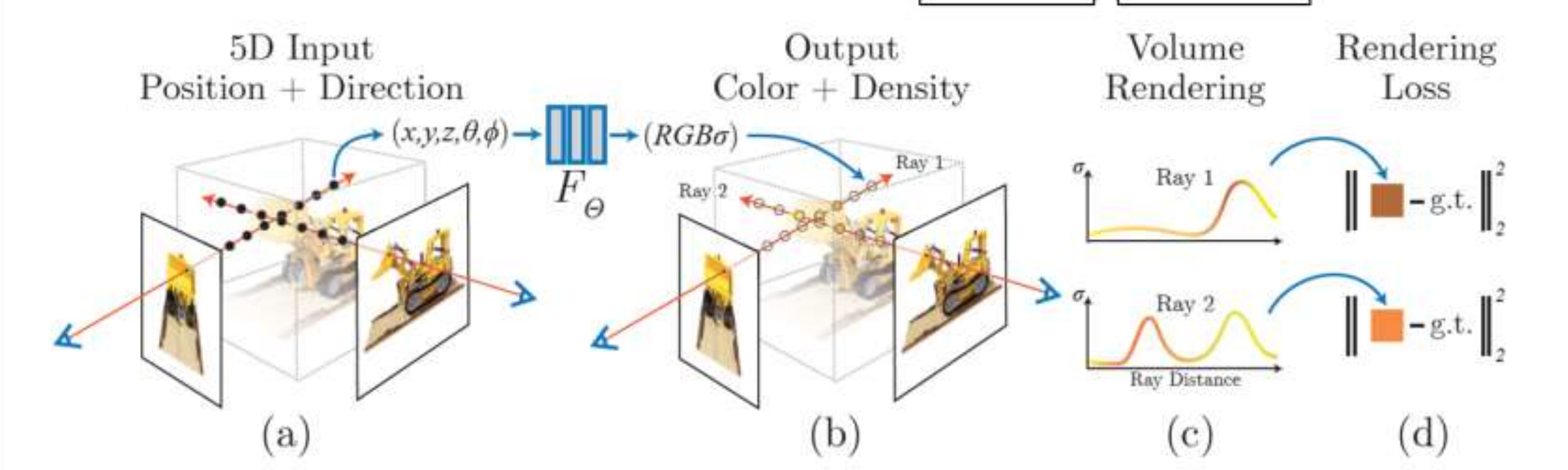


Neural Radiance Field

- Generate views with traditional **volume rendering**

$$\underbrace{C(\mathbf{r})}_{\text{Ray color}} = \int_{t_n}^{t_f} \underbrace{T(t)}_{\text{Transmittance}} \overbrace{\underbrace{\sigma(\mathbf{r}(t))}_{\text{Density}} \underbrace{\mathbf{c}(\mathbf{r}(t),\mathbf{d})}_{\text{Color}}}^{\text{Per volume}} dt\,, \text{ where } T(t) = \exp\left(-\int_{t_n}^{t}\sigma(\mathbf{r}(s))ds\right)$$

$$\mathbf{r}(t) = \underbrace{\mathbf{o}}_{\text{Position}} + t\underbrace{\mathbf{d}}_{\text{Direction}}$$



Neural Radiance Field

- Generate views with traditional **volume rendering**

Rendering model for ray r(t) = o + td:

$$C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$

colors

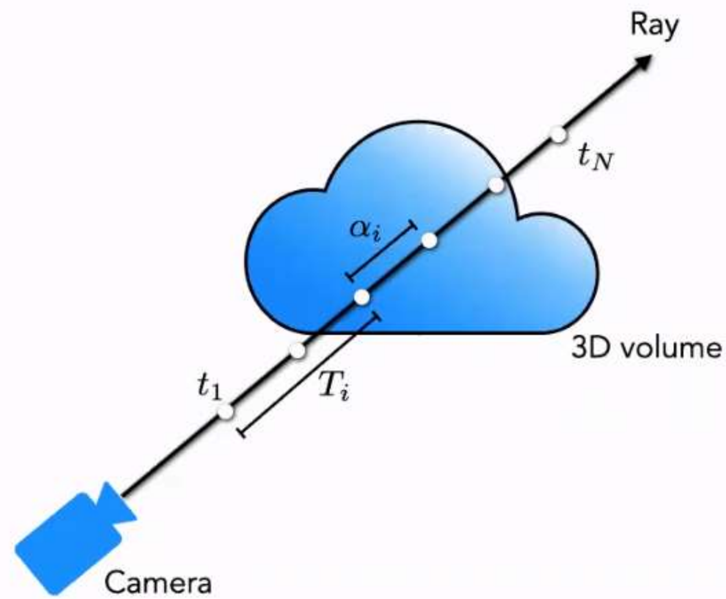
weights

How much light is blocked earlier along ray:

$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

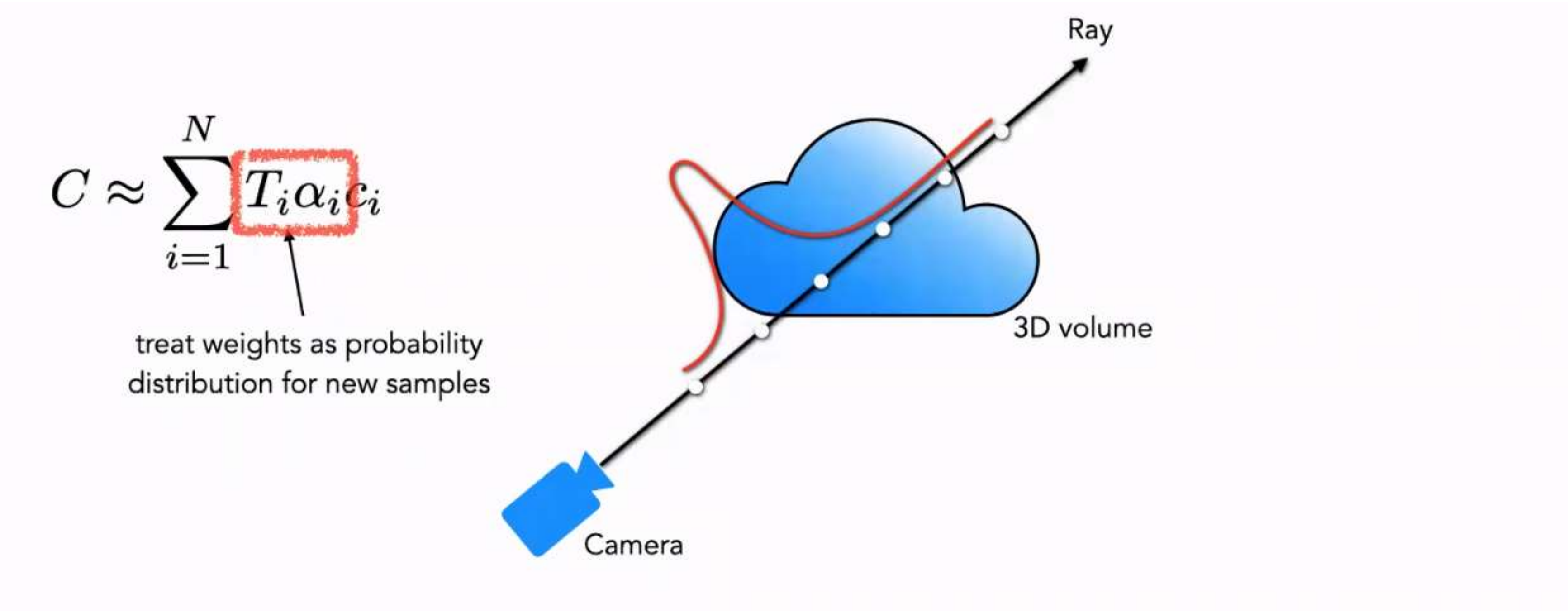
How much light is contributed by ray segment *i*:

$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$



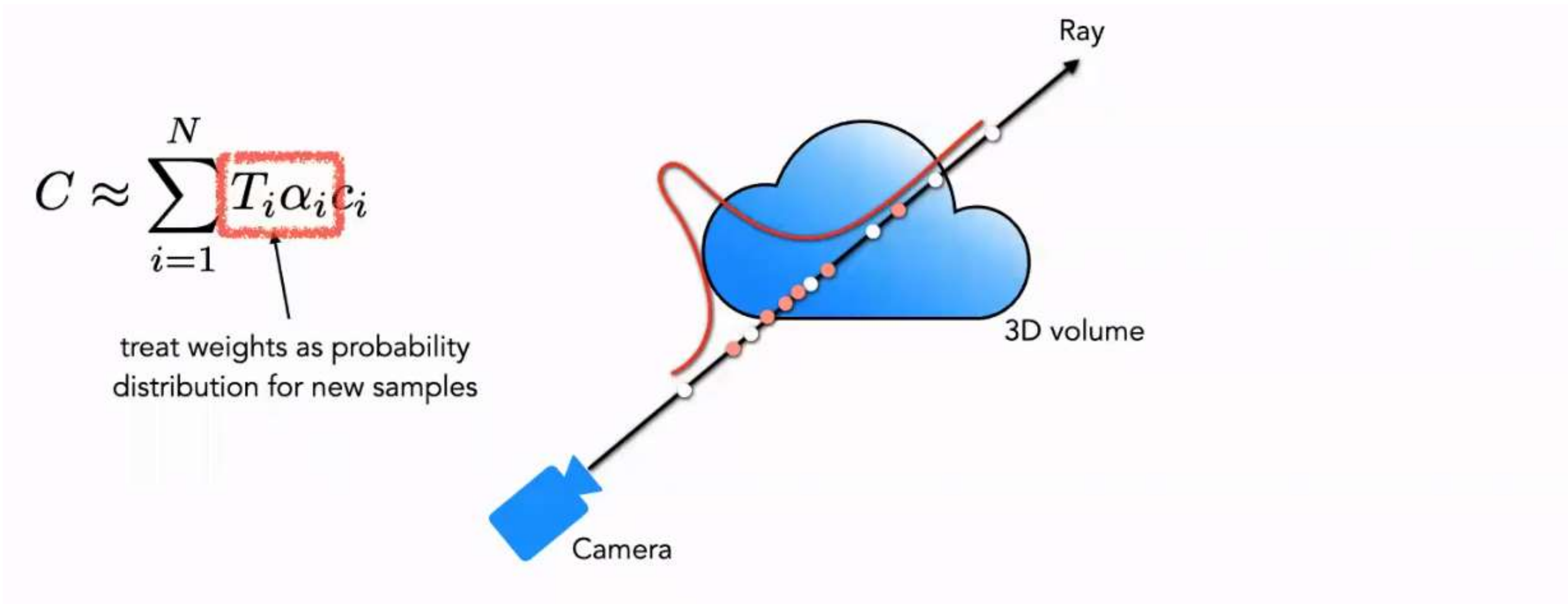
Neural Radiance Field

- Two pass rendering: **coarse**



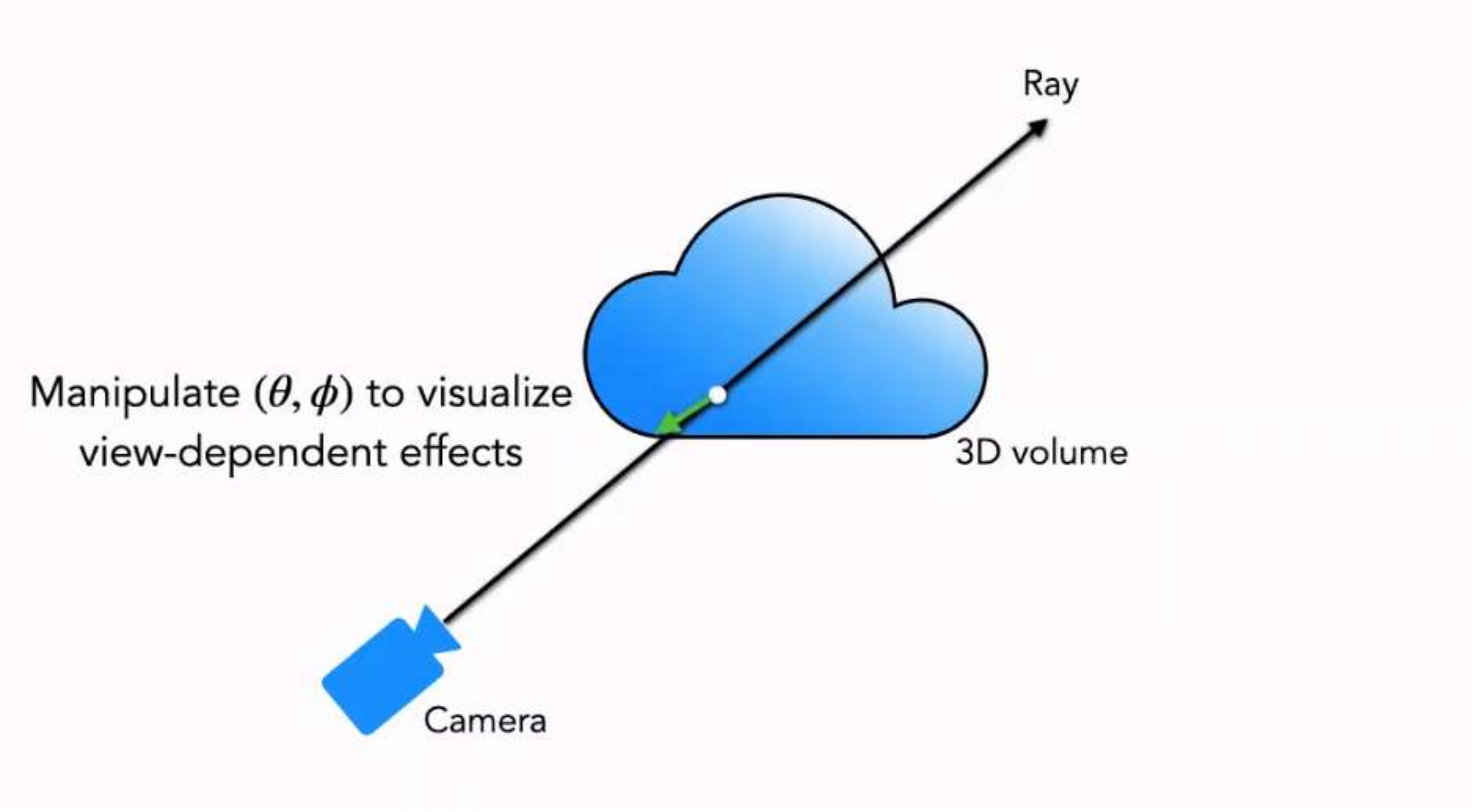
Neural Radiance Field

- Two pass rendering: **fine**



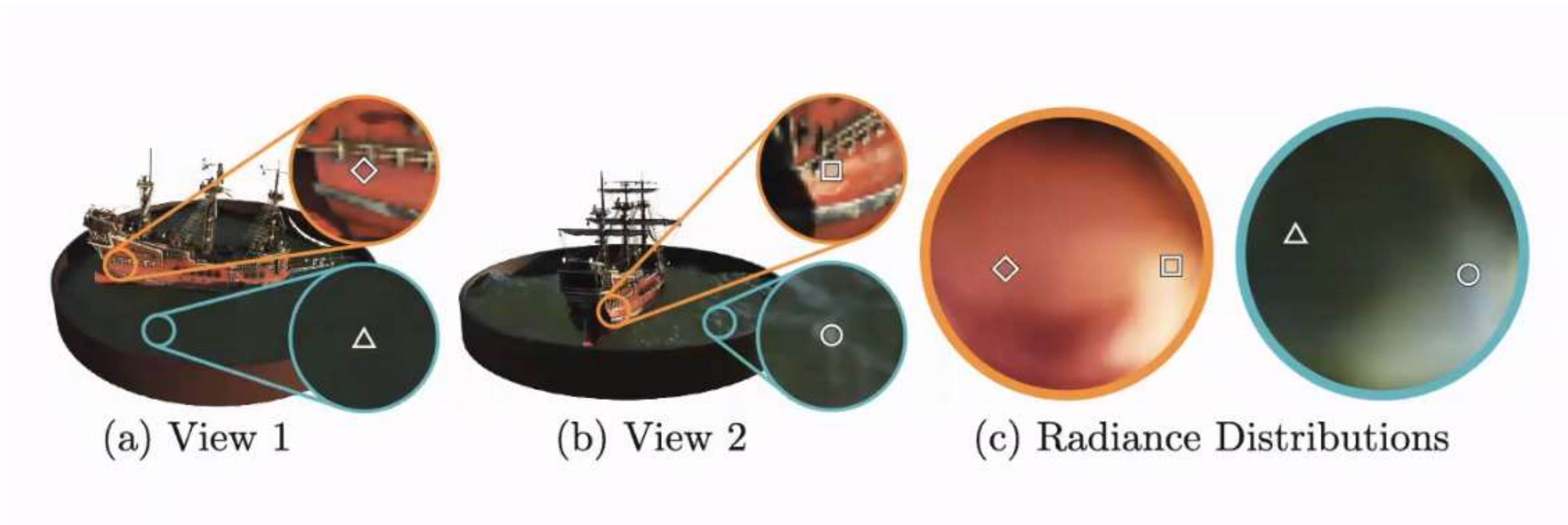
Neural Radiance Field

- Viewing directions as input



Neural Radiance Field

- Viewing directions as input



Neural Radiance Field

- Volume rendering is trivially differentiable

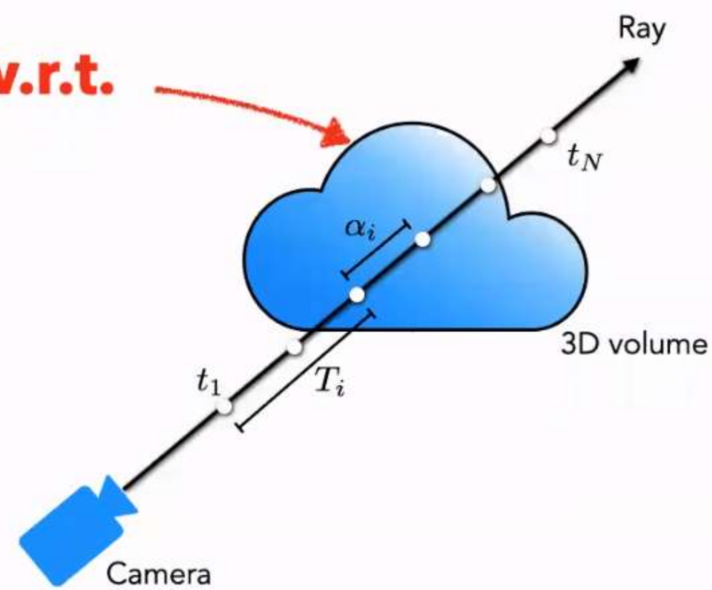
Rendering model for ray r(t) = o + td:

$$\boxed{C \approx \sum_{i=1}^{N} T_i \alpha_i c_i}$$

differentiable w.r.t.

weights

colors



How much light is blocked earlier along ray:

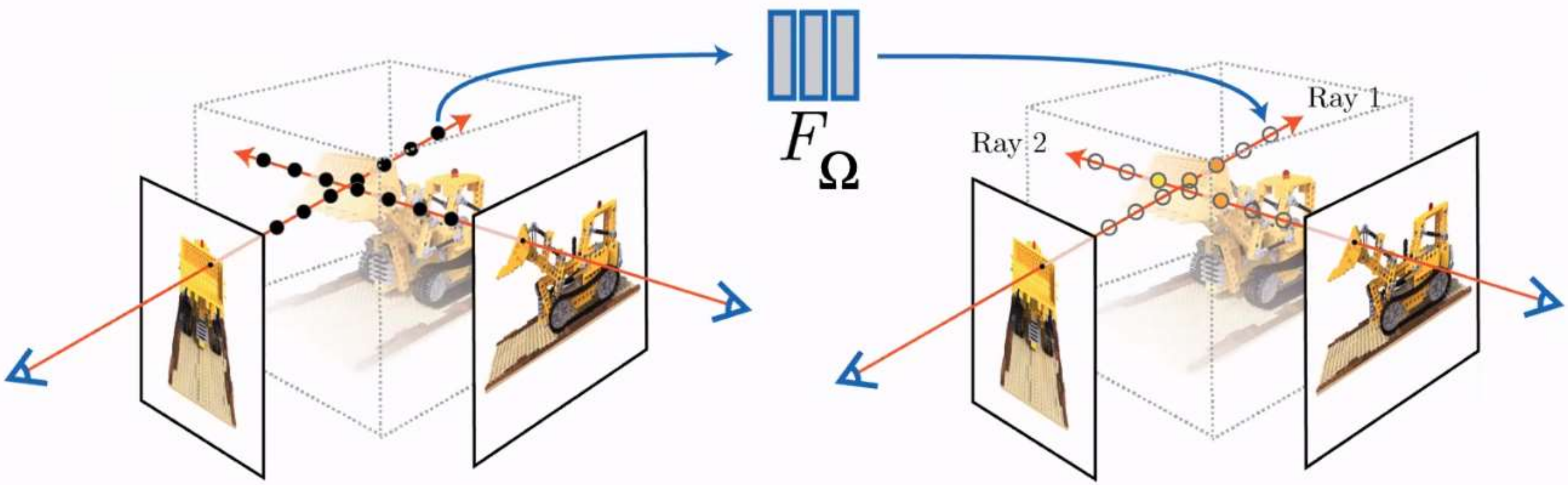
$$T_i = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

How much light is contributed by ray segment *i*:

$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$

Neural Radiance Field

- Optimize with gradient descent on rendering loss



$$\min_{\Omega} \sum_{i} \|\text{render}^{(i)}(F_{\Omega}) - I_{\text{gt}}^{(i)}\|^2$$

Neural Radiance Field

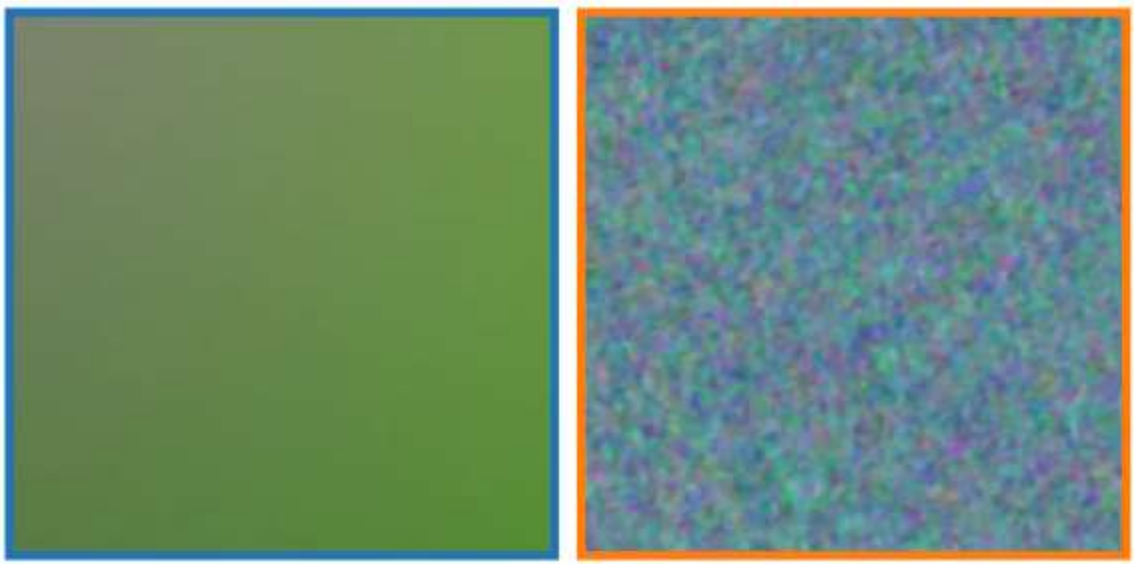
- Positional encoding

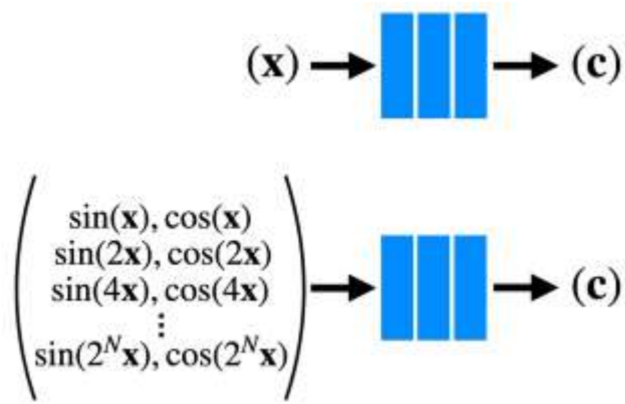
NeRF (Naive)

NeRF (with positional encoding)

Neural Radiance Field

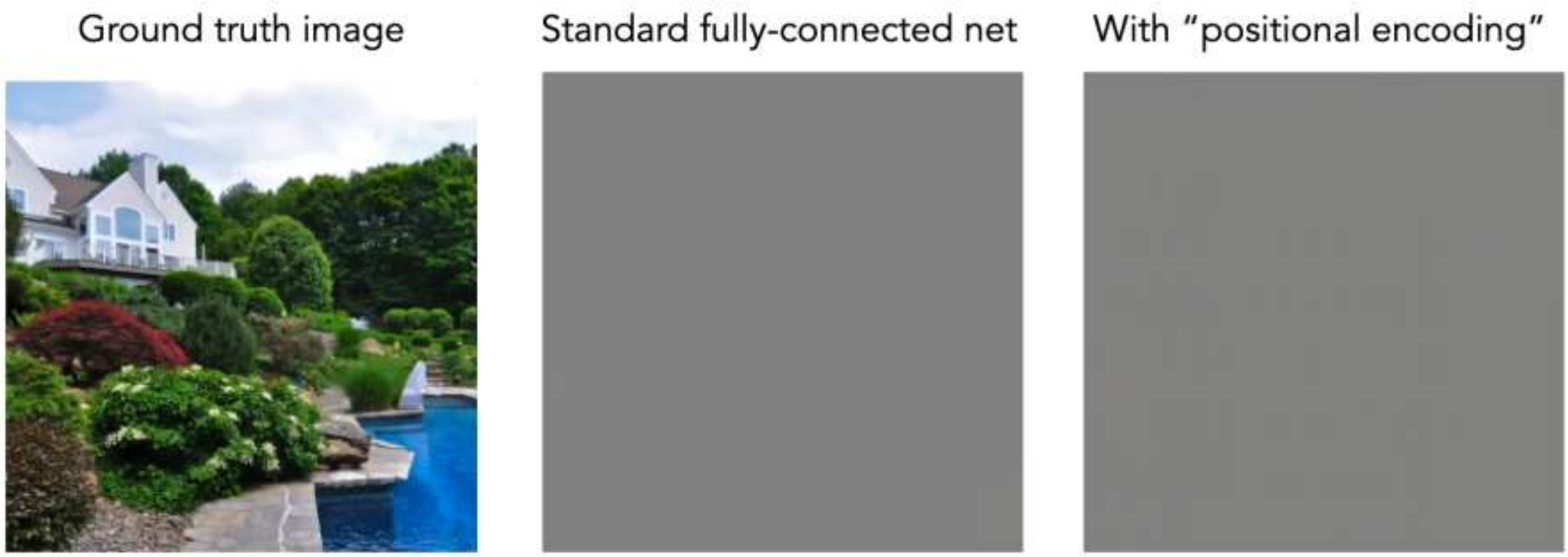
- Positional encoding: 2D toy example





Neural Radiance Field

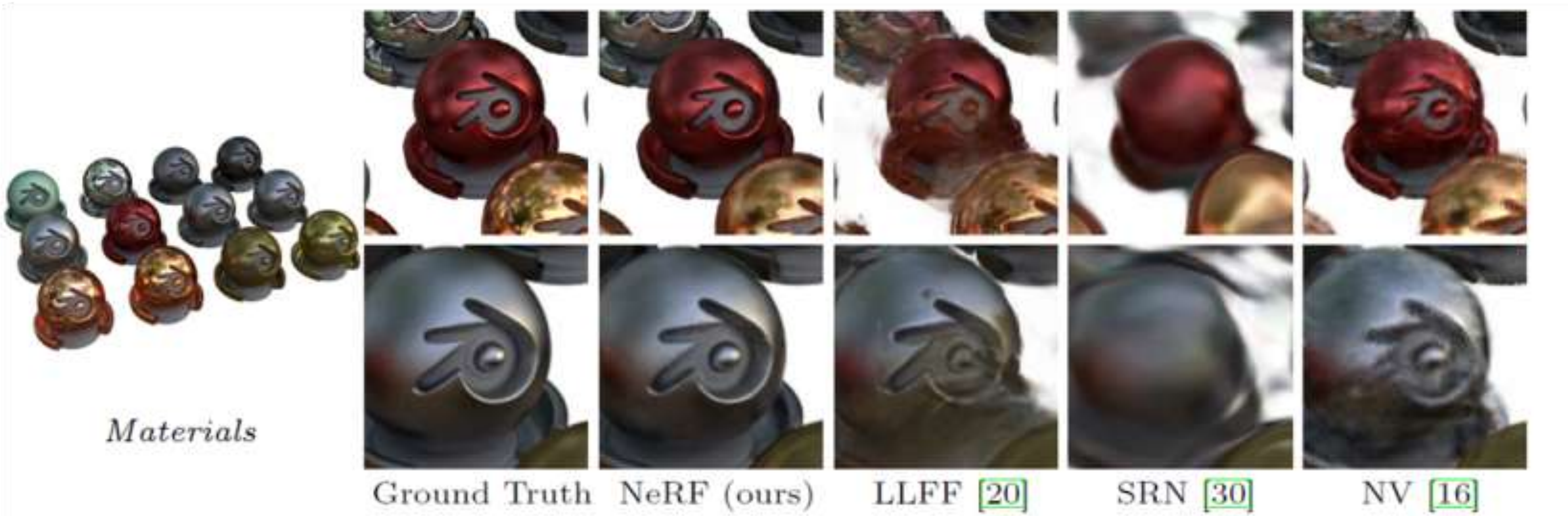
- Positional encoding: simple trick enable network to memorize



Neural Radiance Field

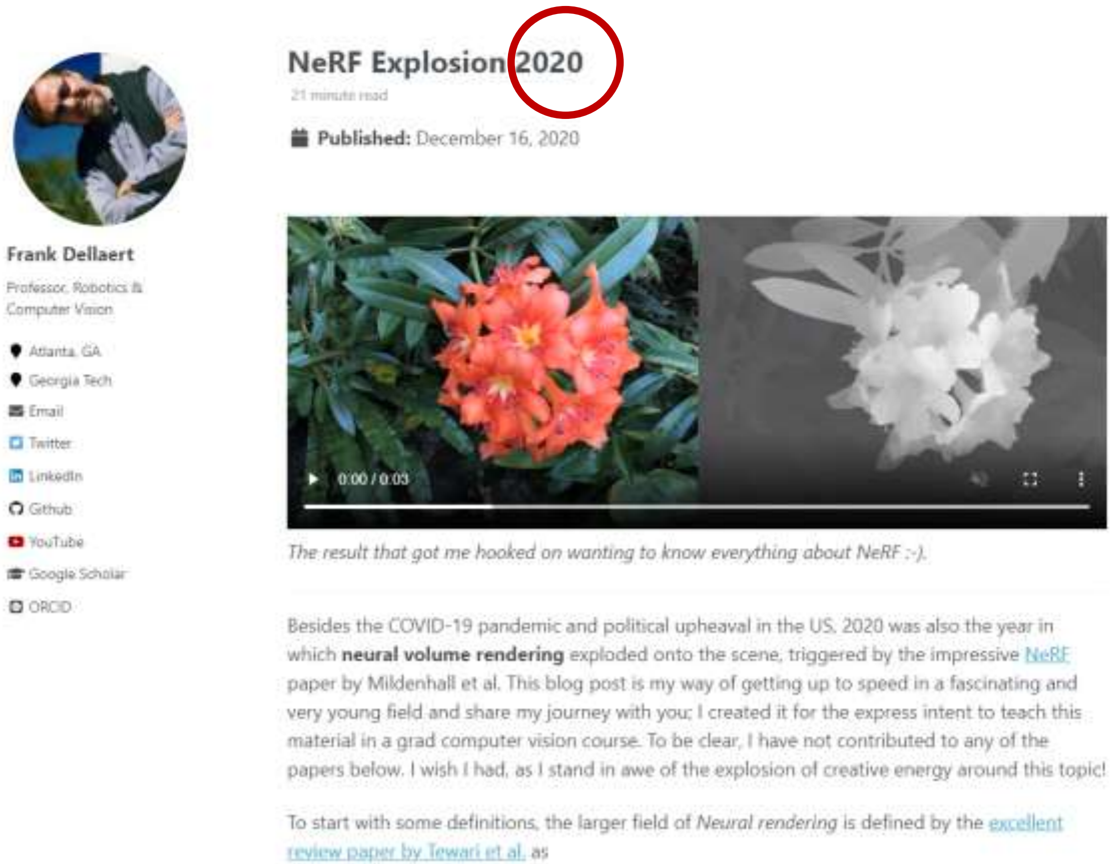
- Comparison Results

Method	Diffuse Synthetic 360° [29]			Realistic Synthetic 360°			Real Forward-Facing [20]		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SRN [30]	33.20	0.986	0.073	22.26	0.867	0.170	22.84	0.866	0.378
NV [16]	29.62	0.946	0.099	26.05	0.944	0.160	-	-	-
LLFF [20]	34.38	0.995	0.048	24.88	0.935	0.114	24.13	0.909	**0.212**
Ours	**40.15**	**0.998**	**0.023**	**31.01**	**0.977**	**0.081**	**26.50**	**0.935**	0.250



Neural Radiance Field

- A lot of know-how within the pass years
- Re-new a lot of topics in CV/CG ……



Frank Dellaert

Professor, Robotics & Computer Vision

Atlanta, GA

Georgia Tech

Email

Twitter

LinkedIn

Github

YouTube

Google Scholar

ORCID

NeRF Explosion 2020

21 minute read

Published: December 16, 2020

0:00 / 0:03

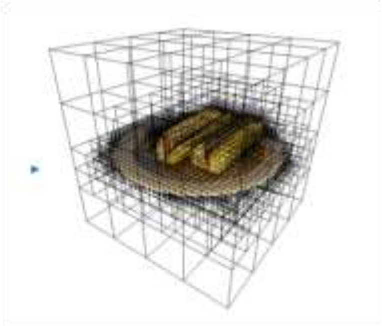
The result that got me hooked on wanting to know everything about NeRF :-).

Besides the COVID-19 pandemic and political upheaval in the US, 2020 was also the year in which **neural volume rendering** exploded onto the scene, triggered by the impressive NeRF paper by Mildenhall et al. This blog post is my way of getting up to speed in a fascinating and very young field and share my journey with you; I created it for the express intent to teach this material in a grad computer vision course. To be clear, I have not contributed to any of the papers below. I wish I had, as I stand in awe of the explosion of creative energy around this topic!

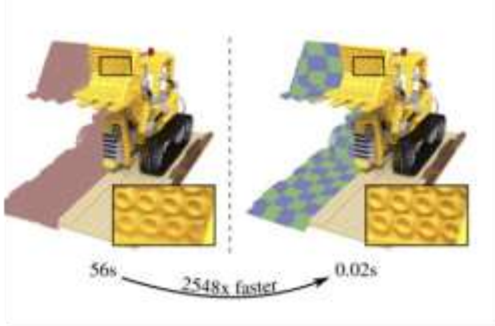
To start with some definitions, the larger field of *Neural rendering* is defined by the excellent review paper by Tewari et al. as

Powerful NeRF everywhere

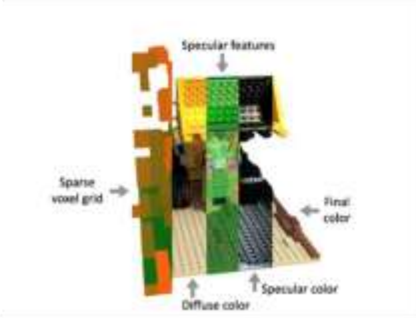
- Fast Rendering and Fast Training



Yu et. al, 2021



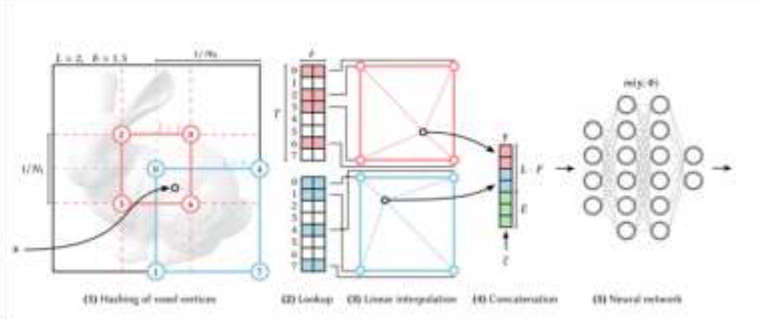
Reiser et. al, 2021



Hedman et. al, 2021

Garbin et. al, 2021

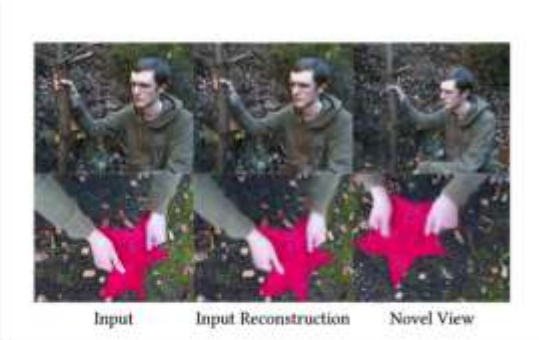
Wang et. al, 2022



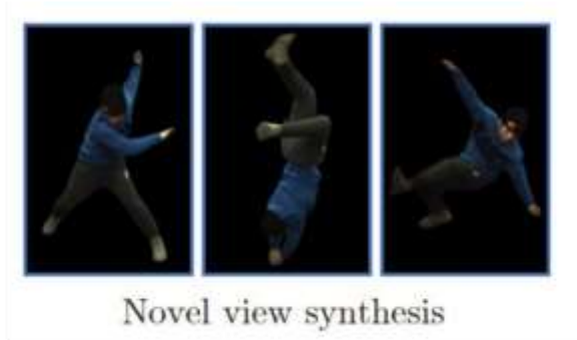
Müller, et. al, 2022

Powerful NeRF everywhere

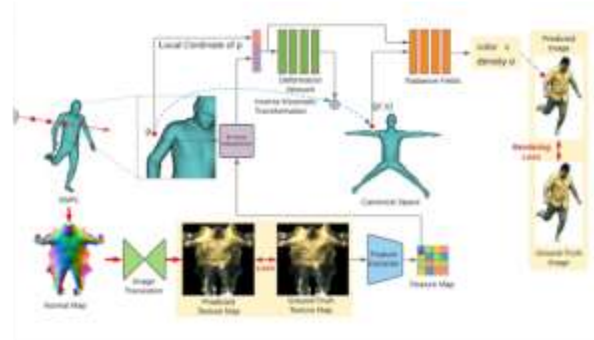
- Dynamic Modeling



Tretschk et al. 2019, Park et al. 2020, Pumarola et al. 2020, Li et al. 2020, Xian et al. 2020



Peng et al. 2020, 2021



Liu et al. 2021

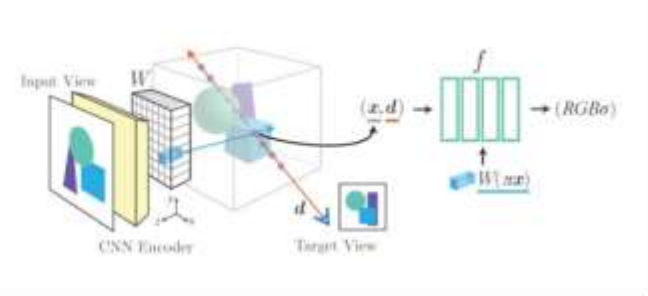
Zheng et al. 2022

HumanNeRF [Zhao et al. 2022]

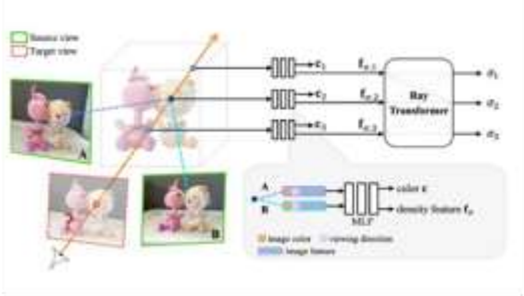
Artemis [Luo et al. 2022]

Powerful NeRF everywhere

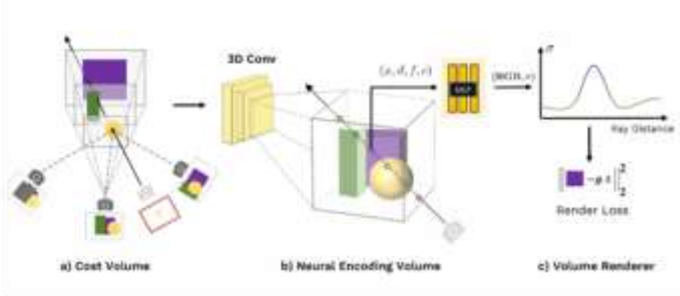
- Generalization



PixelNeRF [Yu et al. 2021]



IBRNet [Wang et al. 2021]



MVSNeRF [Chen et al. 2021]

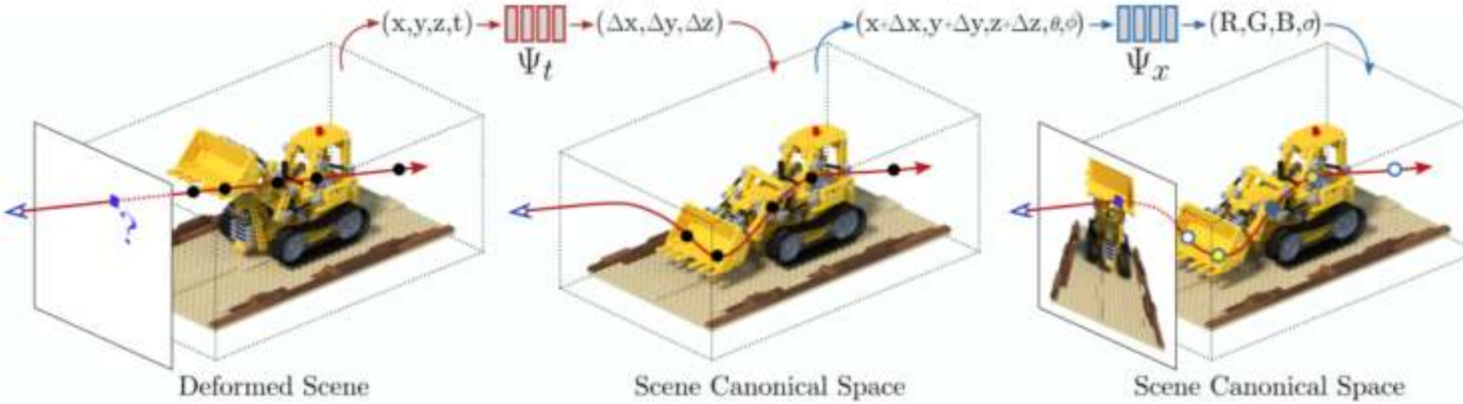
- Pose estimation
- Relighting
- Editing and Composition

…….

Some examples of NeRF-related topics

NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters



D-NeRF: Neural Radiance Fields for Dynamic Scene
Albert Pumarola, Enric Corona, Gerard Pons-Moll, Francesc Moreno-Noguer

NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters

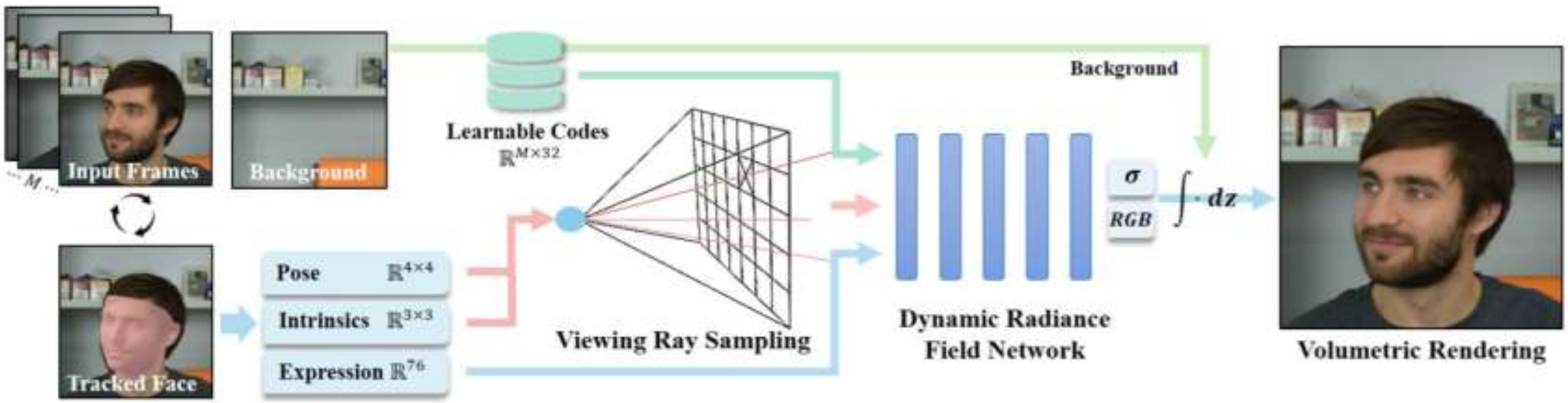
Non-Rigid Neural Radiance Fields:
Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video

Edgar Tretschk Ayush Tewari Vladislav Golyanik
Michael Zollhöfer Christoph Lassner Christian Theobalt

max planck institut informatik
Graphics Vision & Video
SIC Saarland Informatics Campus
facebook Reality Labs

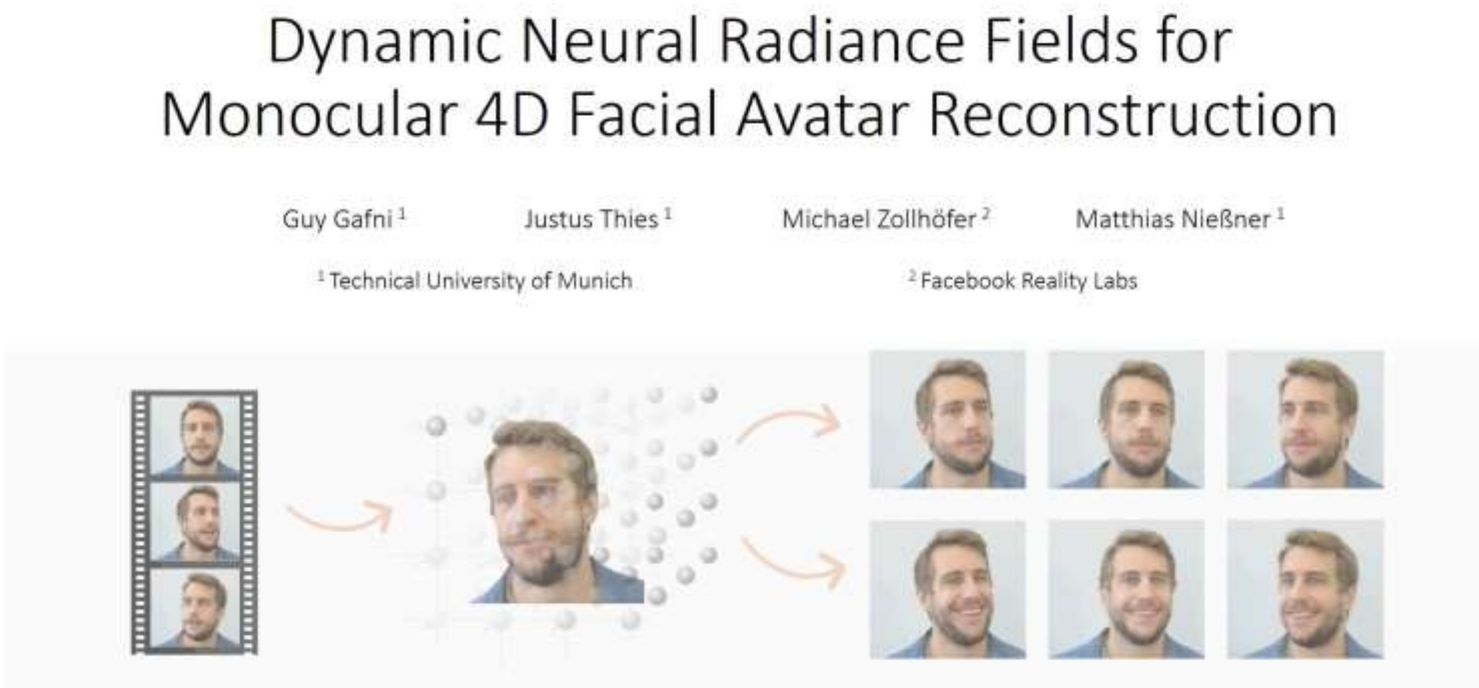
NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters



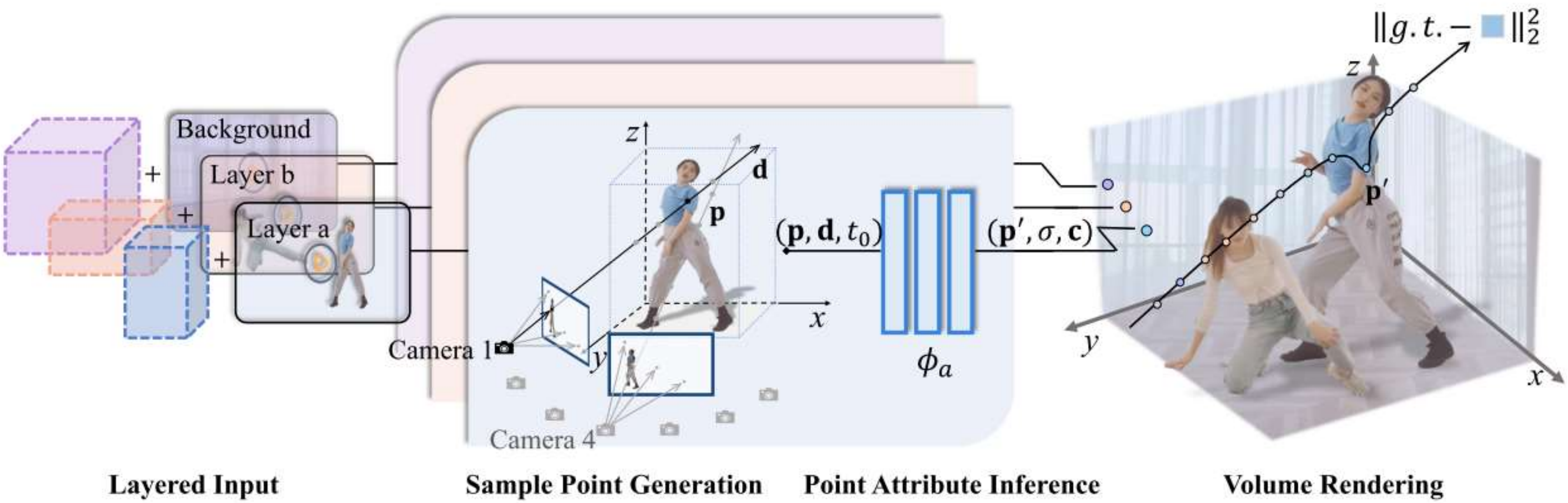
NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters



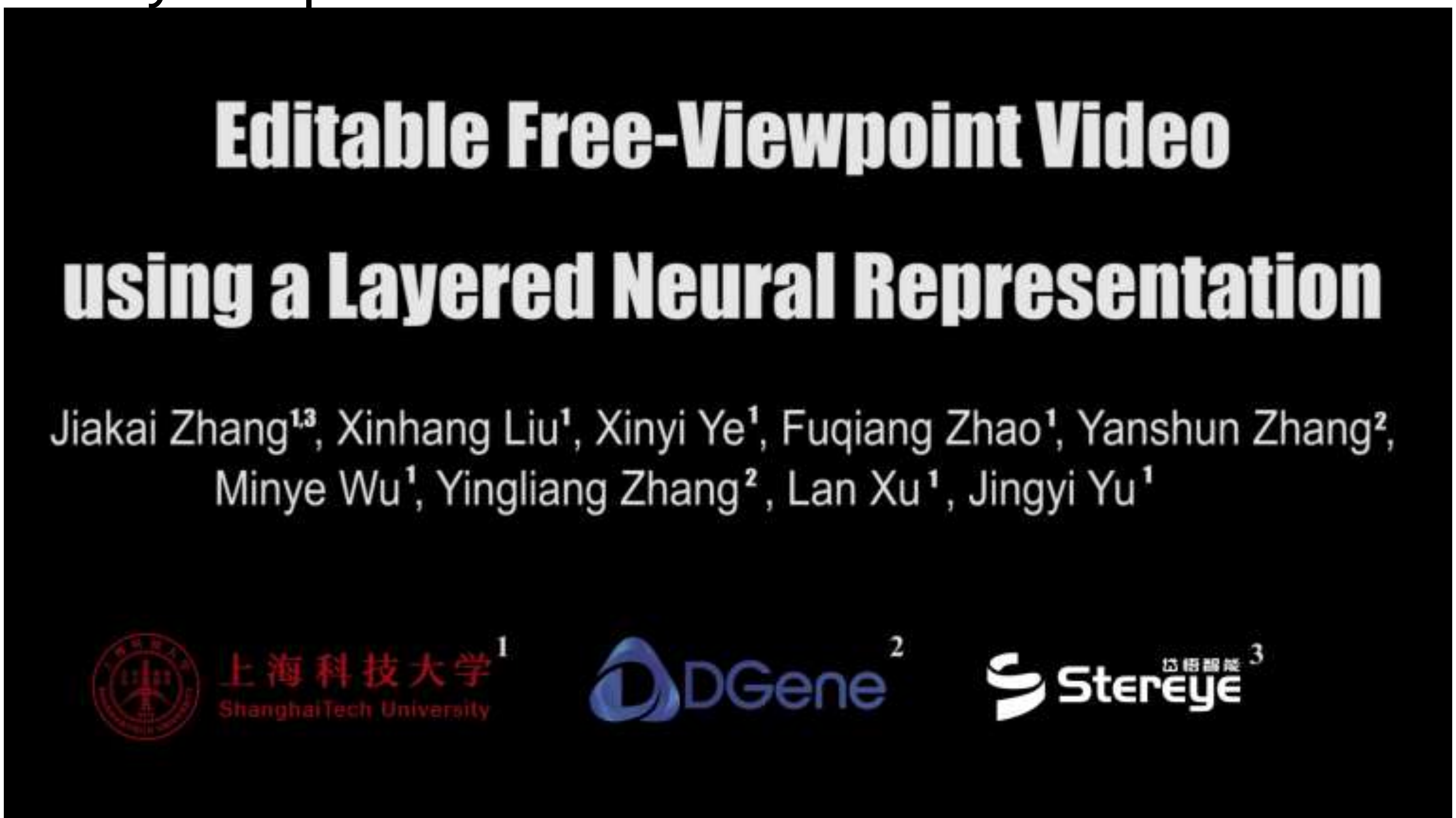
NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters
- Multi-layer representation



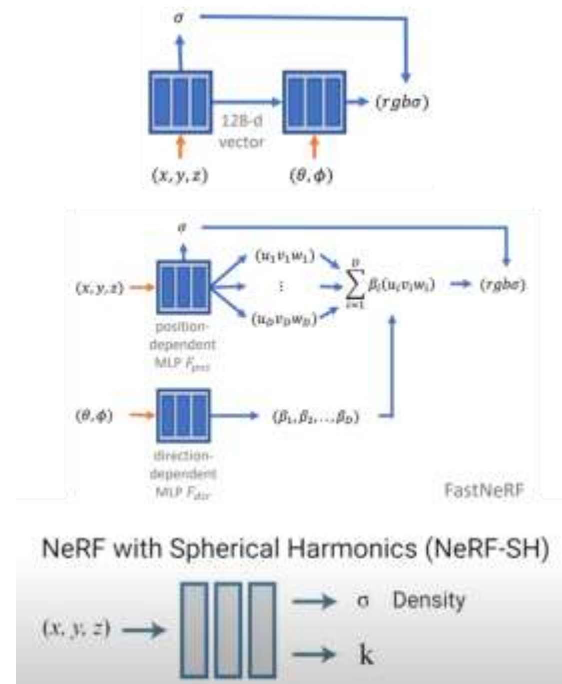
NeRF-related examples

- Static scene → Dynamic scenes
- Conditioning to time & motion parameters
- Multi-layer representation



NeRF-related examples

- Offline inference → Real-time inference
- Store the view-dependent attributes explicitly

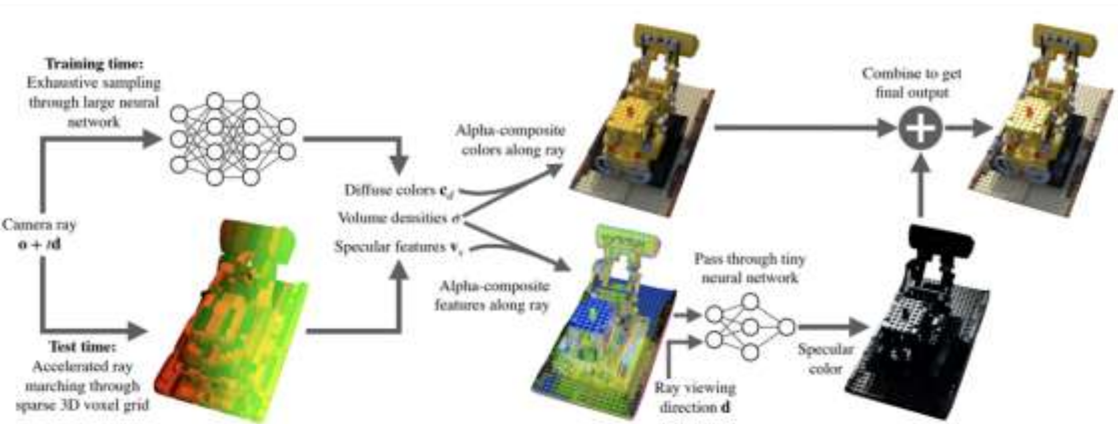


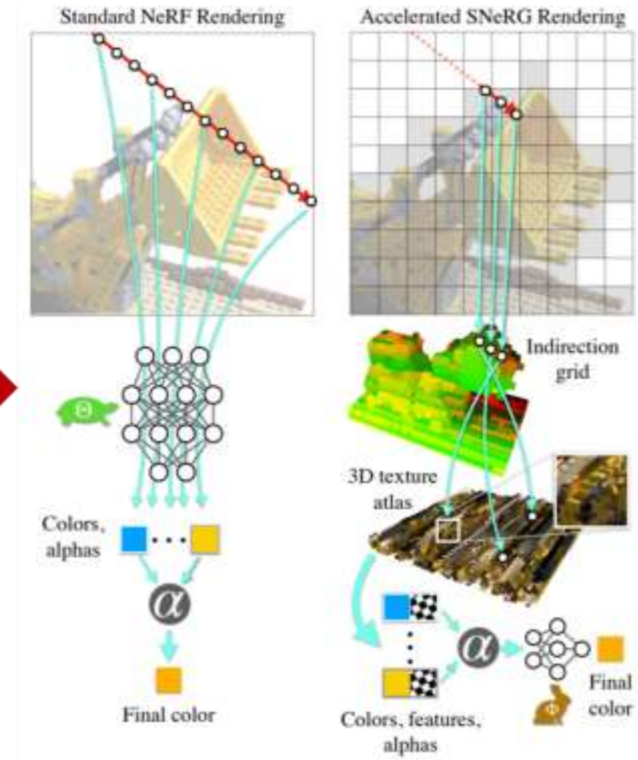
NeRF-related examples

- Offline inference → Real-time inference
- Store the view-dependent attributes explicitly

Baking Neural Radiance Fields for Real-Time View Synthesis

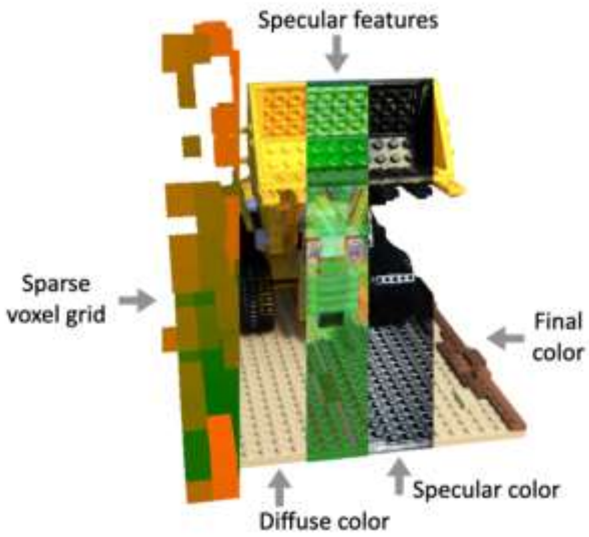
Peter Hedman, Pratul P. Srinivasan, Ben Mildenhall, Jonathan T. Barron, Paul Debevec





NeRF-related examples

- Offline inference → Real-time inference
- Store the view-dependent attributes explicitly

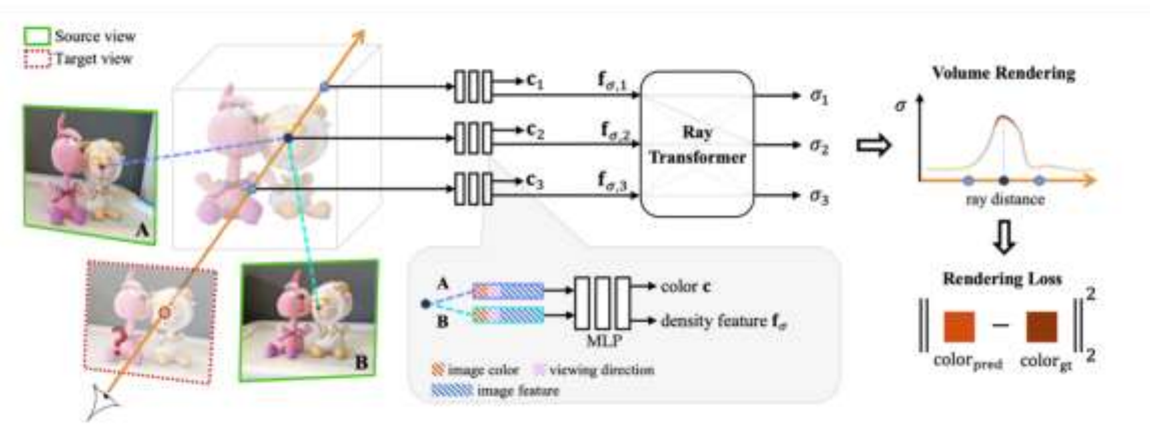


NeRF-related examples

- Per-scene training → General Scenes
- Encode category-aware features as input

IBRNet: Learning Multi-View Image-Based Rendering

Qianqian Wang, Zhicheng Wang, Kyle Genova, Pratul Srinivasan, Howard Zhou, Jonathan T. Barron, Ricardo Martin-Brualla, Noah Snavely, Thomas Funkhouser



Even works in sparse 6 views!

NeRF-related examples

- Offline inference → Real-time inference
- Store the view-dependent attributes explicitly

Recent Observation 1: from Implicit Reconstruction to Modeling/Generation

Neural Implicit Shape Modeling

- Recent Huge progress in 3D Generation (Lec-19)

Magic3D 18 Nov 2022

Point-E 21 Dec 2022

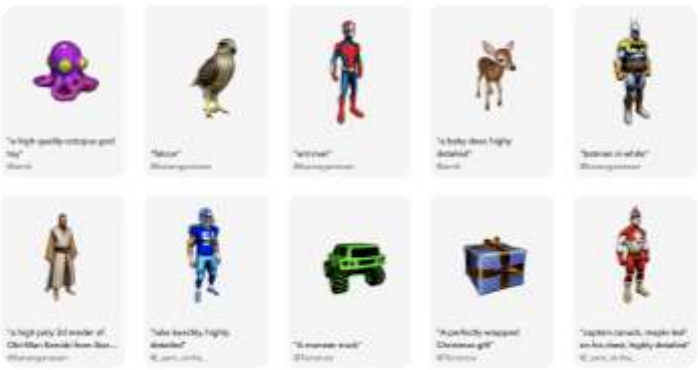
DreamFusion 29 Sep 2022

Microsoft

Luma AI

Rodin Diffusion 12 Dec 2022

Imagine 14 Dec 2022



Hard to Say...

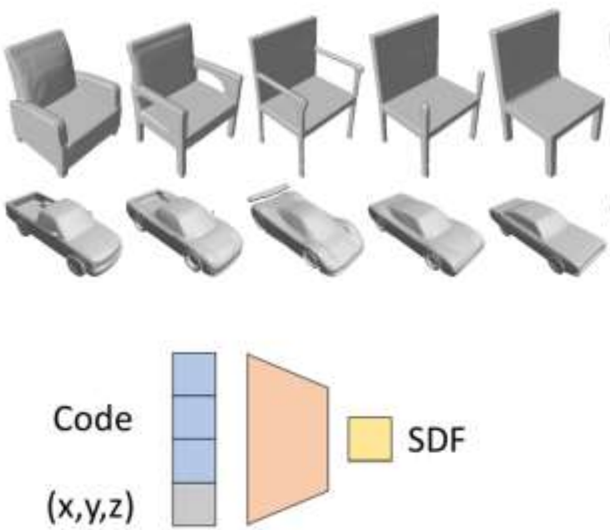
KAEDIM

3DFY.ai

Sloyd Beta

Neural Implicit Shape Modeling

- Date back to Implicit Shape Representation for Generation

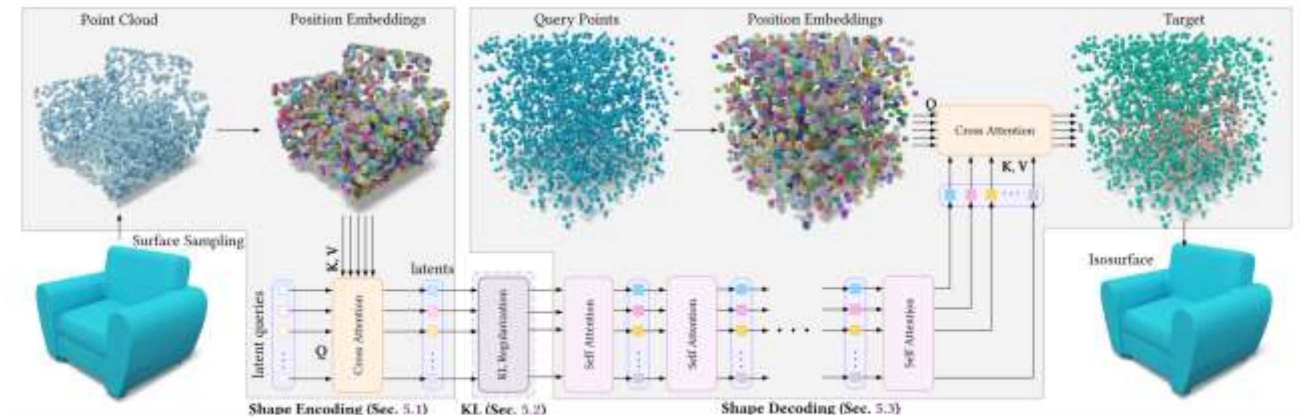


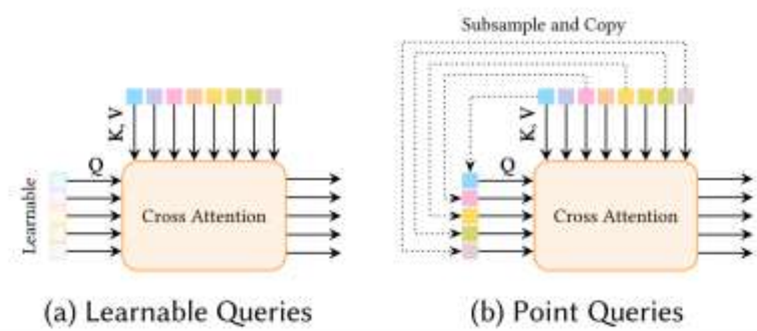
Coded Shape DeepSDF

DeepSDF: Learning Continuous Signed Distance Functions for Shape Representation, *CVPR 2019*

Neural Implicit Shape Modeling

- Recent trend: more powerful 3D shape dataset and generation model





3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diff usion Models, SIGGRAPH 2023

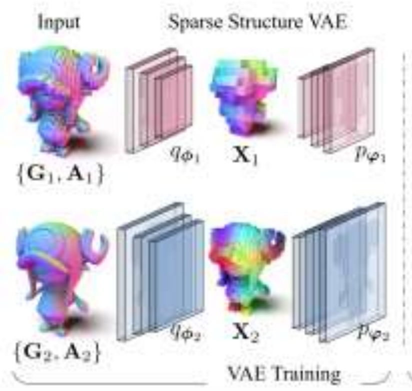
Neural Implicit Shape Modeling

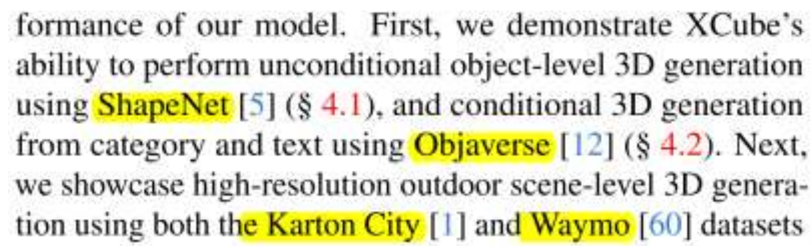
- Recent trend: more powerful 3D shape dataset and generation model

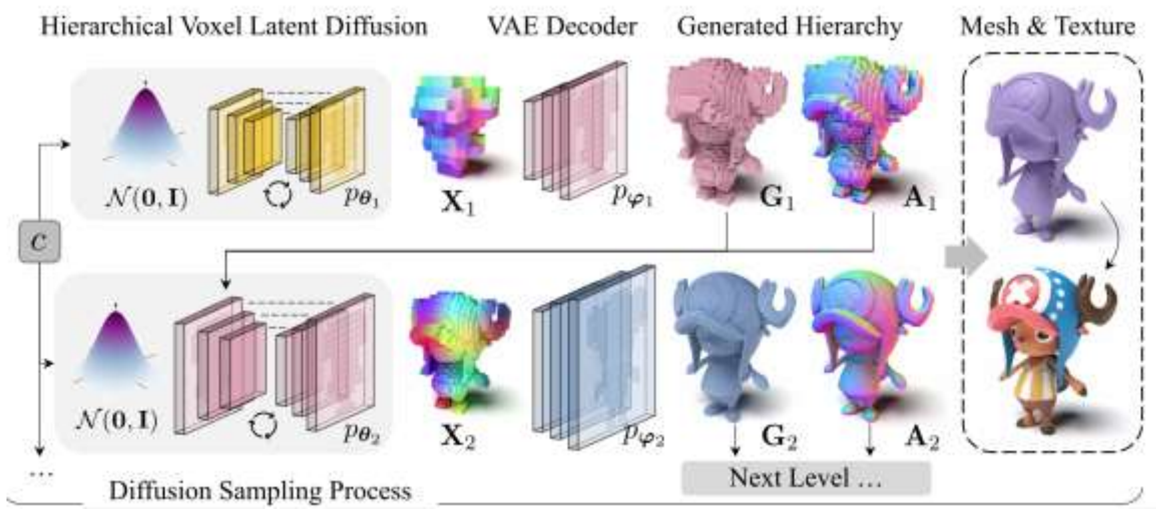
3DShape2VecSet: A 3D Shape Representation for Neural Fields and Generative Diff usion Models, SIGGRAPH 2023

Neural Implicit Shape Modeling

- Recent trend: more powerful 3D shape dataset and generation model




formance of our model. First, we demonstrate XCube's ability to perform unconditional object-level 3D generation using ShapeNet [5] (§ 4.1), and conditional 3D generation from category and text using Objaverse [12] (§ 4.2). Next, we showcase high-resolution outdoor scene-level 3D generation using both the Karton City [1] and Waymo [60] datasets



Cube^3: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies, Arxiv 2023

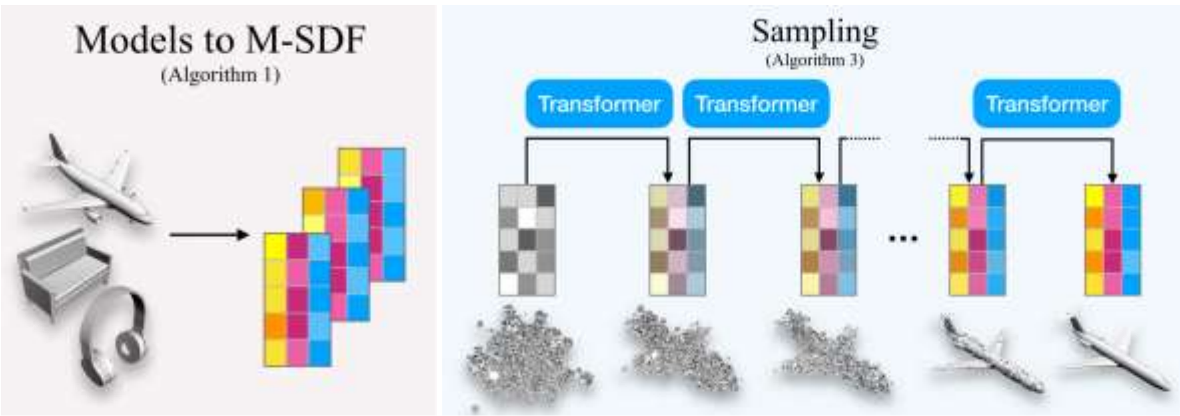
Neural Implicit Shape Modeling

- Recent trend: more powerful 3D shape dataset and generation model

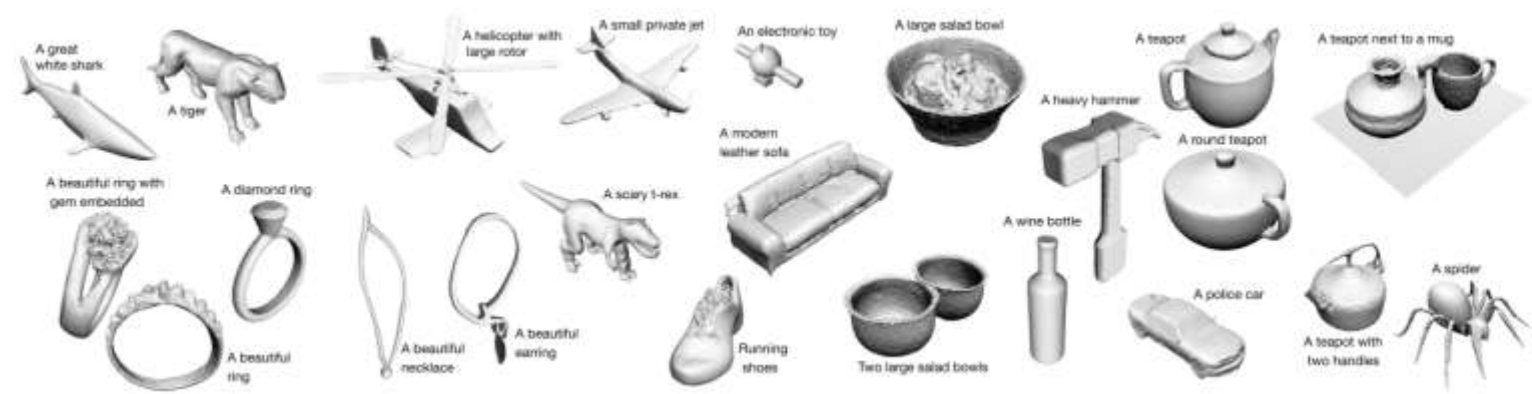
Cube^3: Large-Scale 3D Generative Modeling using Sparse Voxel Hierarchies, Arxiv 2023

Neural Implicit Shape Modeling

- Recent trend: more powerful 3D shape dataset and generation model



ing c. Our transformer is built with 24 layers with 16 heads and 1024 hidden dimension, which result in a 328M parameter model. We train U^{θ} for 500K iterations with batch size of 1024 using the ADAM optimizer [16] and learning rate of 1e−4 with initial warm-up of 5K iterations. We additionally perform EMA (Exponential Moving Average) to the transfomer's weights. Both training were done on 8 nodes of 8 NVIDIA A100 GPUs, which takes around a week.



Mosaic-SDF for 3D Generative Models, arXiv 2023

Recent Key Observation 2: from Implicit to Explicit to Hybrid Representations

Neural Shape Modeling

- The recent Gaussian-splating strikes back to explicit scenarios
- More explicit feature primitives

"3D Gaussian Splatting for Real-Time Radiance Field Rendering", SIGGRAPH 2023

Neural Shape Modeling

- Explicit non-rigid tracking and rendering via 4D-GS

HiFi4G: High-Fidelity
Human Performance Rendering
via Compact Gaussian Splatting

Yuheng Jiang[1,2] Zhehao Shen[1] Penghao Wang[1]
Zhuo Su[3] Yu Hong[1] Yingliang Zhang[4]
Jingyi Yu[1] Lan Xu[1]

[1]ShanghaiTech University [2]NeuDim [3]ByteDance [4]DGene

HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting, Arxiv 2023

Neural Shape Modeling

- Embed into SLAM systems: tracking-mapping-splating

SplaTAM from CMU/MIT

GS-SLAM from ICL

Neural Shape Modeling

- Embed into SLAM systems: tracking-mapping-splating
- Another one from Lab. of AI and Robotics (LAIR), UNIST