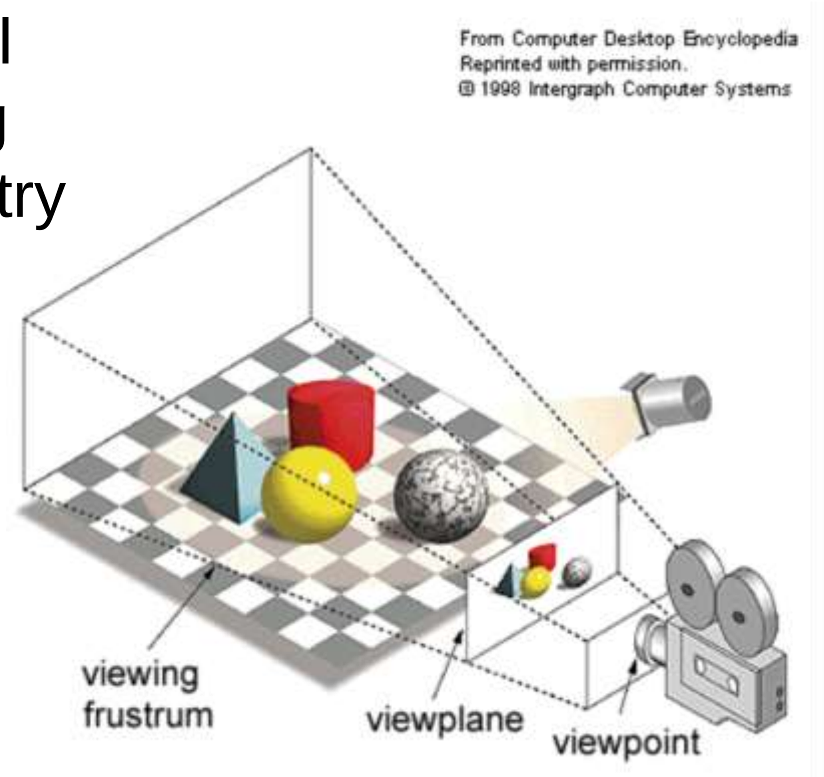
Lecture 20: Advanced Topics: Neural Rendering II

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)
- …

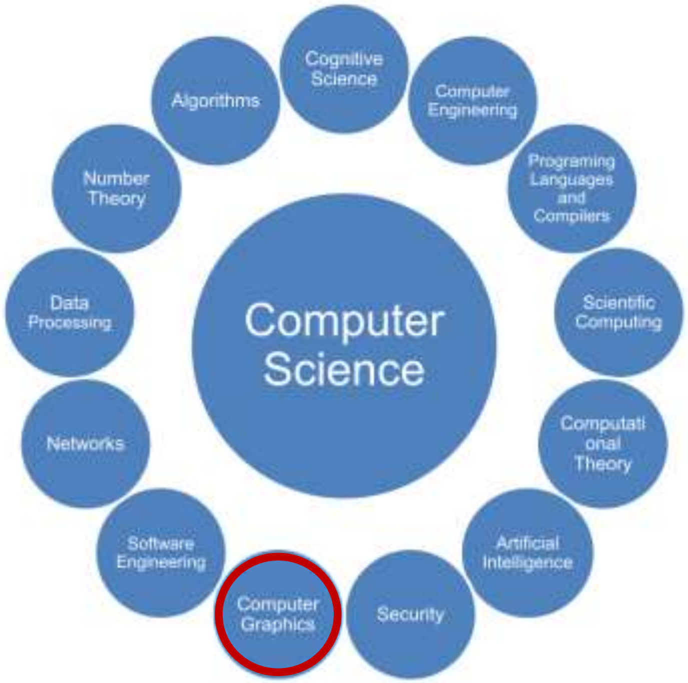
Photo-realistic Image Synthesis

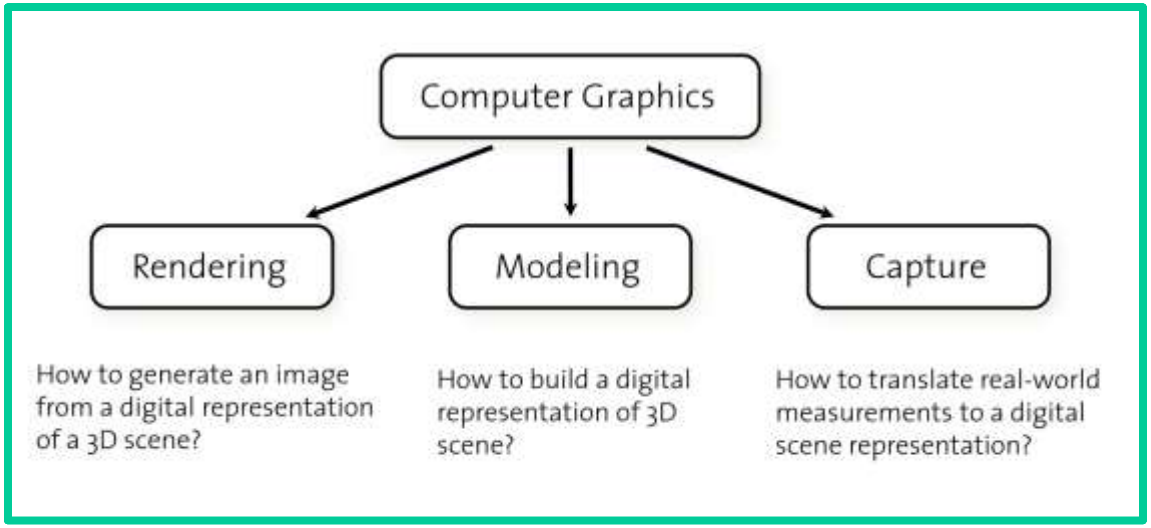
- The Rendering Equation [Kajiya 86]

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

Computer Graphics?

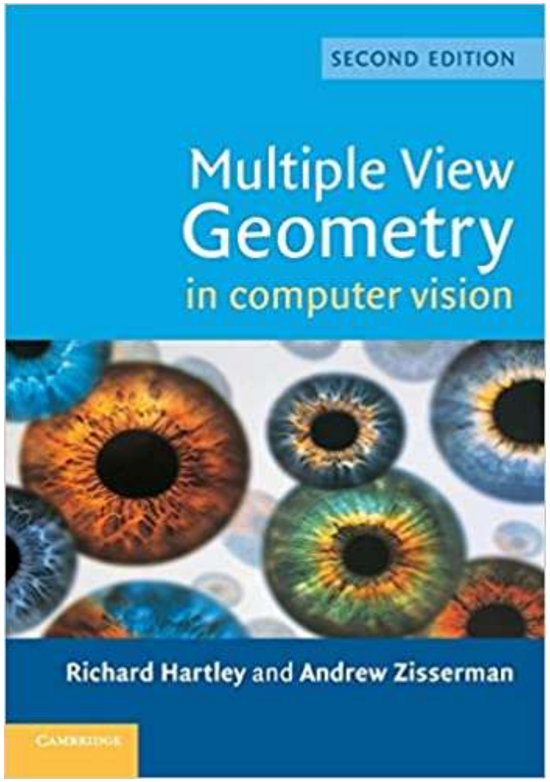
- Both inverse process and forward process
- From **real world** to **virtual representation**, then to **vivid rendering**





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

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

architecture

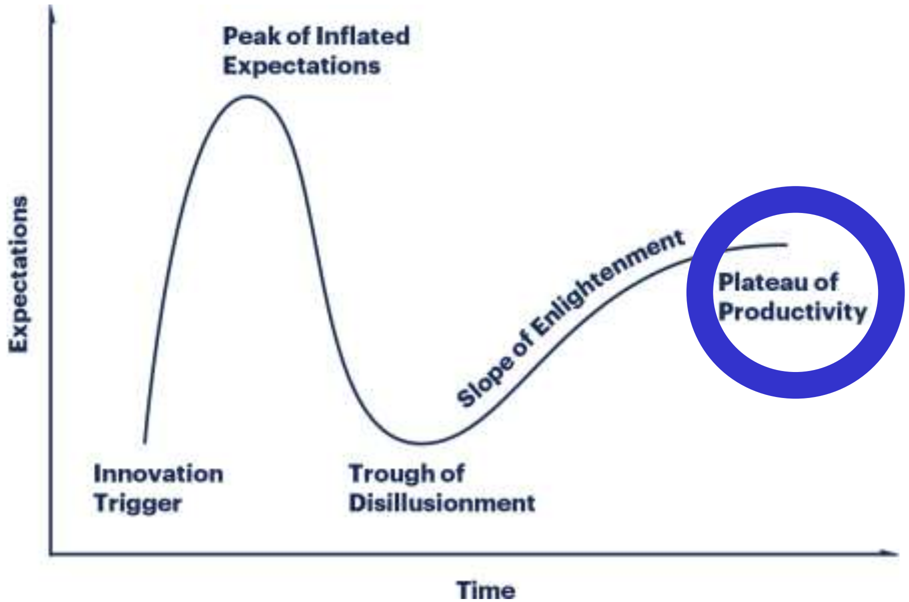
digital twin

Movie

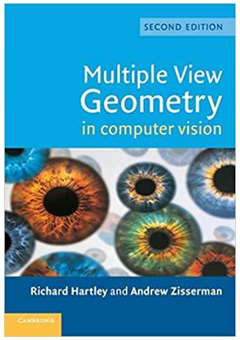
E-commerce

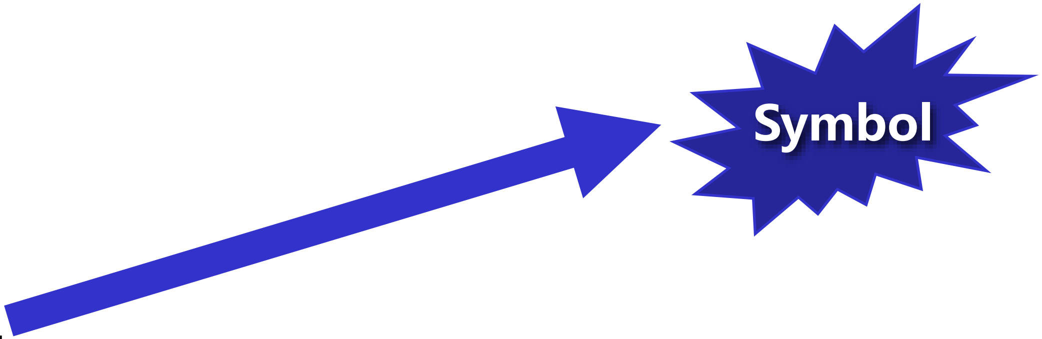
Recall traditional pipeline

- Traditional Pipeline: mature in the past decades



2021: Epic Games buys Capturing Reality

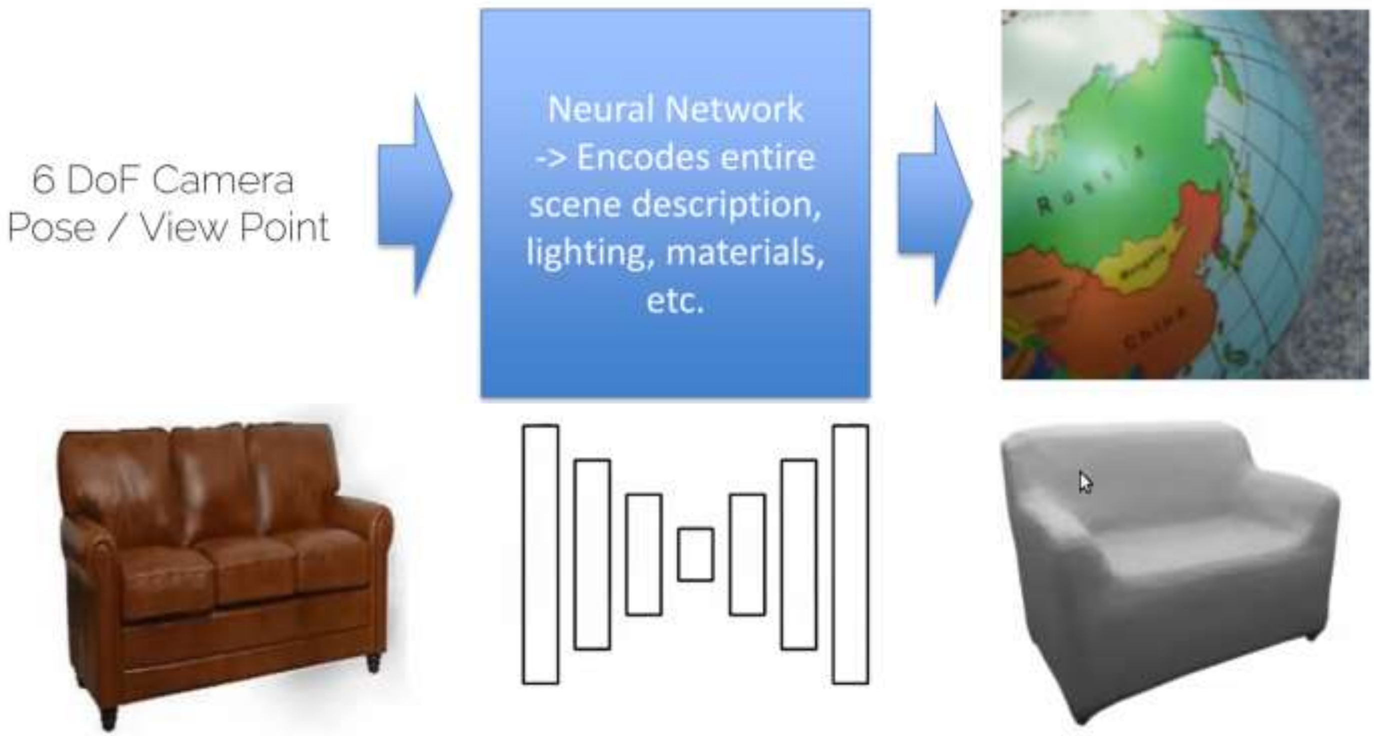




2004

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**

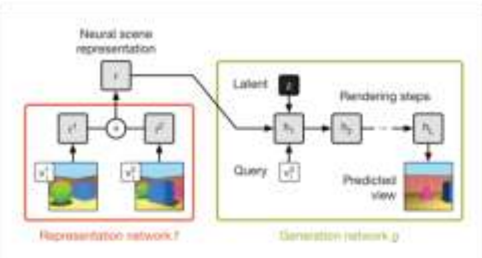
1)
Generative networks that synthesis raw pixel output

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

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

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

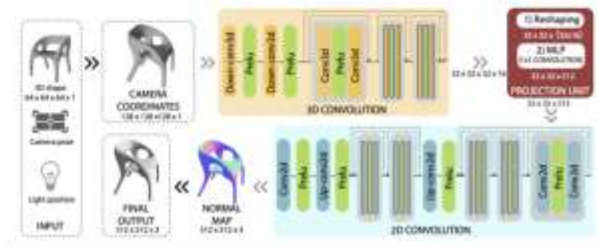
Neural Representation History



Generative Query Networks
[Eslami et al. 2018]

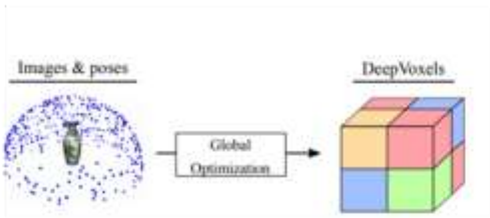
[Flynn et al., 2016; Zhou et al., 2018b; Mildenhall et al. 2019]

Multiplane Images (MPIs)

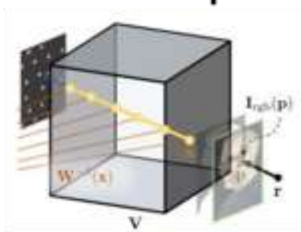


RenderNet [Nguyen-Phuoc et al. 2018]

Voxel Grids + CNN decoder



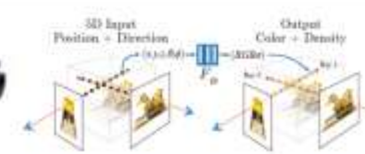
DeepVoxels
[Sitzmann et al. 2019]



Neural Volumes
[Lombardi et al. 2019]

Voxel Grids + Ray Marching

SRN
[Sitzmann et al. 2019b]



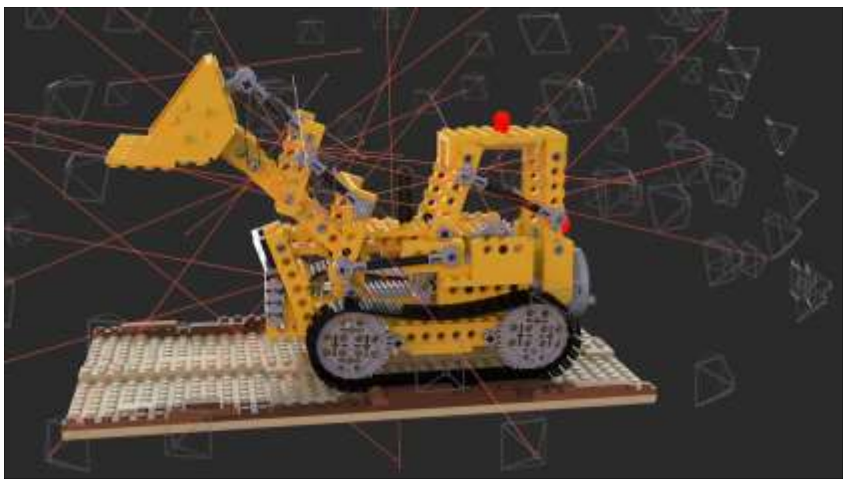
NeRF
[Mildenhall et al. 2020]

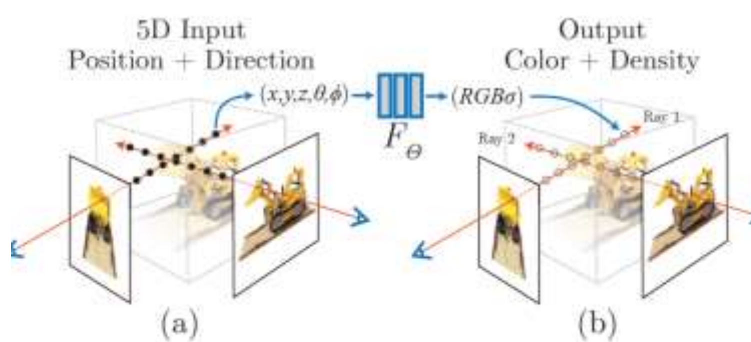
IDR
[Yariv et al. 2020]

Implicit Fields

Neural Implicit Representation

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

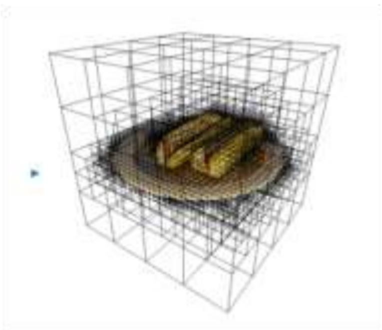




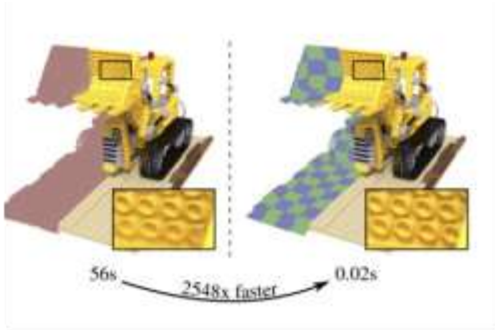
Representing Scenes as Neural Radiance Fields for View Synthesis, Mildenhall et al., *ECCV 2020 Oral - Best Paper Honorable Mention*

Powerful NeRF everywhere

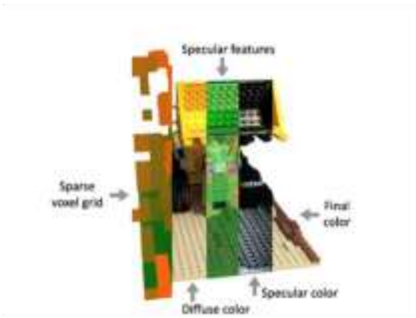
- Fast Rendering and Fast Training



Yu et. al, 2021



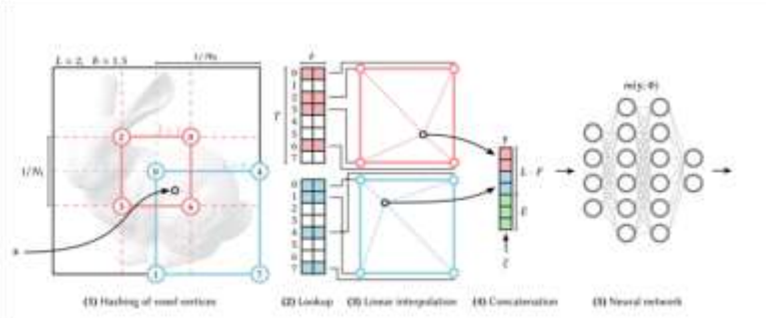
Reiser et. al, 2021



Hedman et. al, 2021

Garbin et. al, 2021

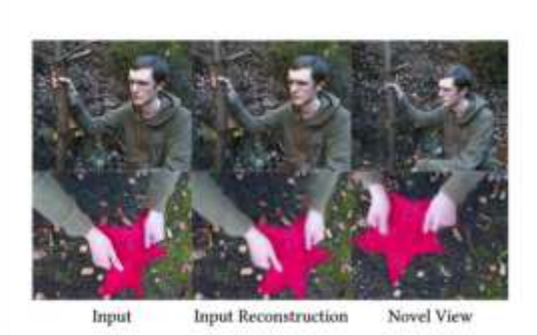
Wang et. al, 2022



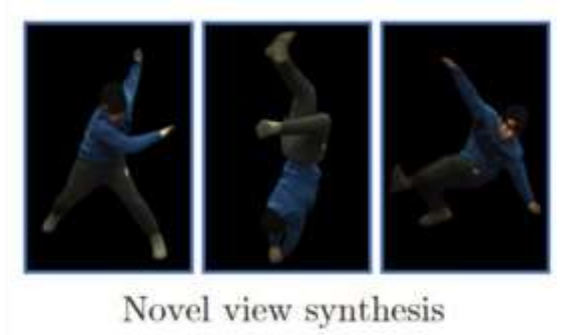
Müller, et. al, 2022

Powerful NeRF everywhere

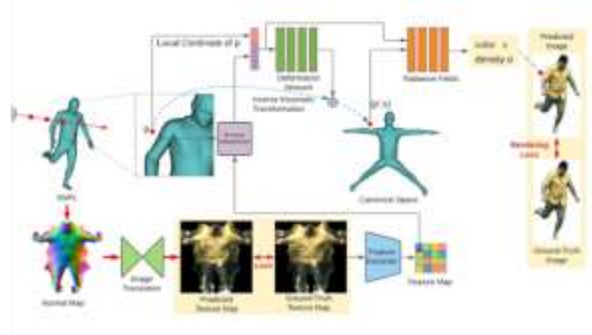
- Dynamic Modeling



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

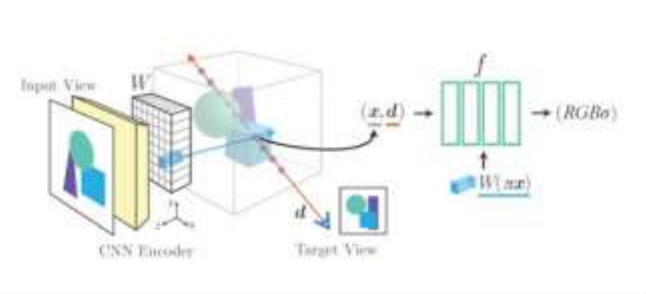
Zheng et al. 2022

HumanNeRF [Zhao et al. 2022]

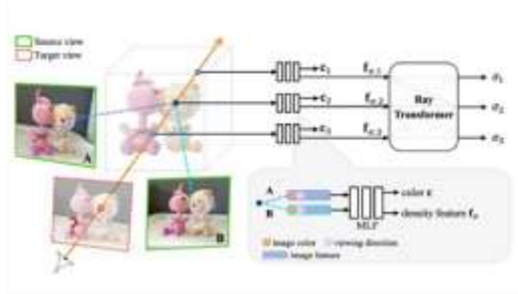
Artemis [Luo et al. 2022]

Powerful NeRF everywhere

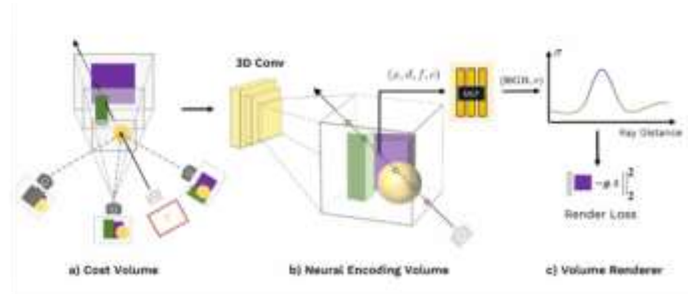
- Generalization



PixelNeRF [Yu et al. 2021]



IBRNet [Wang et al. 2021]



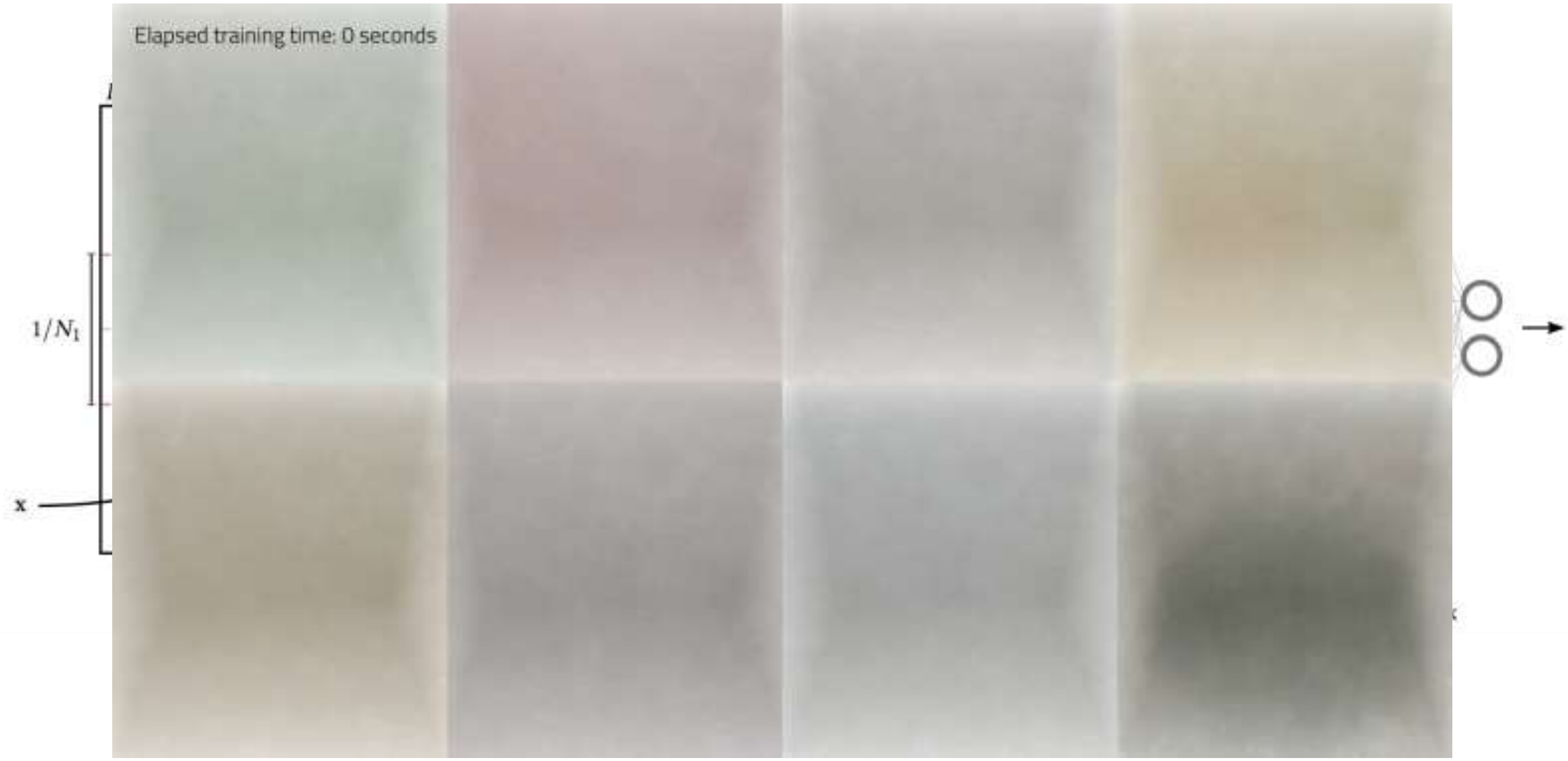
MVSNeRF [Chen et al. 2021]

- Pose estimation
- Relighting
- Editing and Composition

…….

Instant Neural Graphics Primitives

- Multi-resolution hash encoding
- Shallow MLP, CUDA implementation



Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, Müller et al., *ACM Transactions on Graphics (SIGGRAPH 2022)*

Instant Neural Graphics Primitives

- Multi-resolution hash encoding
- Shallow MLP

INSTANT NEURAL GRAPHICS PRIMITIVES WITH A MULTIRESOLUTION HASH ENCODING

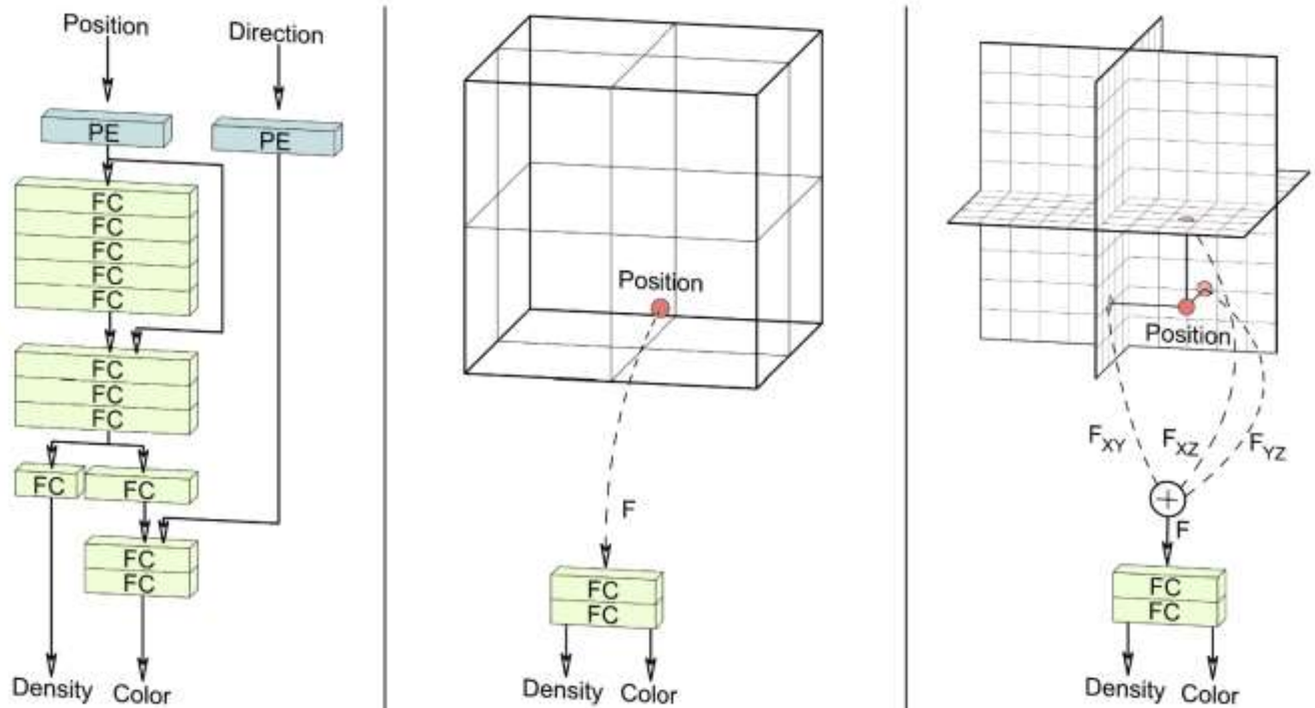
Thomas Müller Alex Evans Christoph Schied Alexander Keller

https://nvlabs.github.io/instant-ngp

Instant Neural Graphics Primitives with a Multiresolution Hash Encoding, Müller et al., *ACM Transactions on Graphics (SIGGRAPH 2022)*

Tri-plane feature representation

- Pack the continue feature manifold into planes

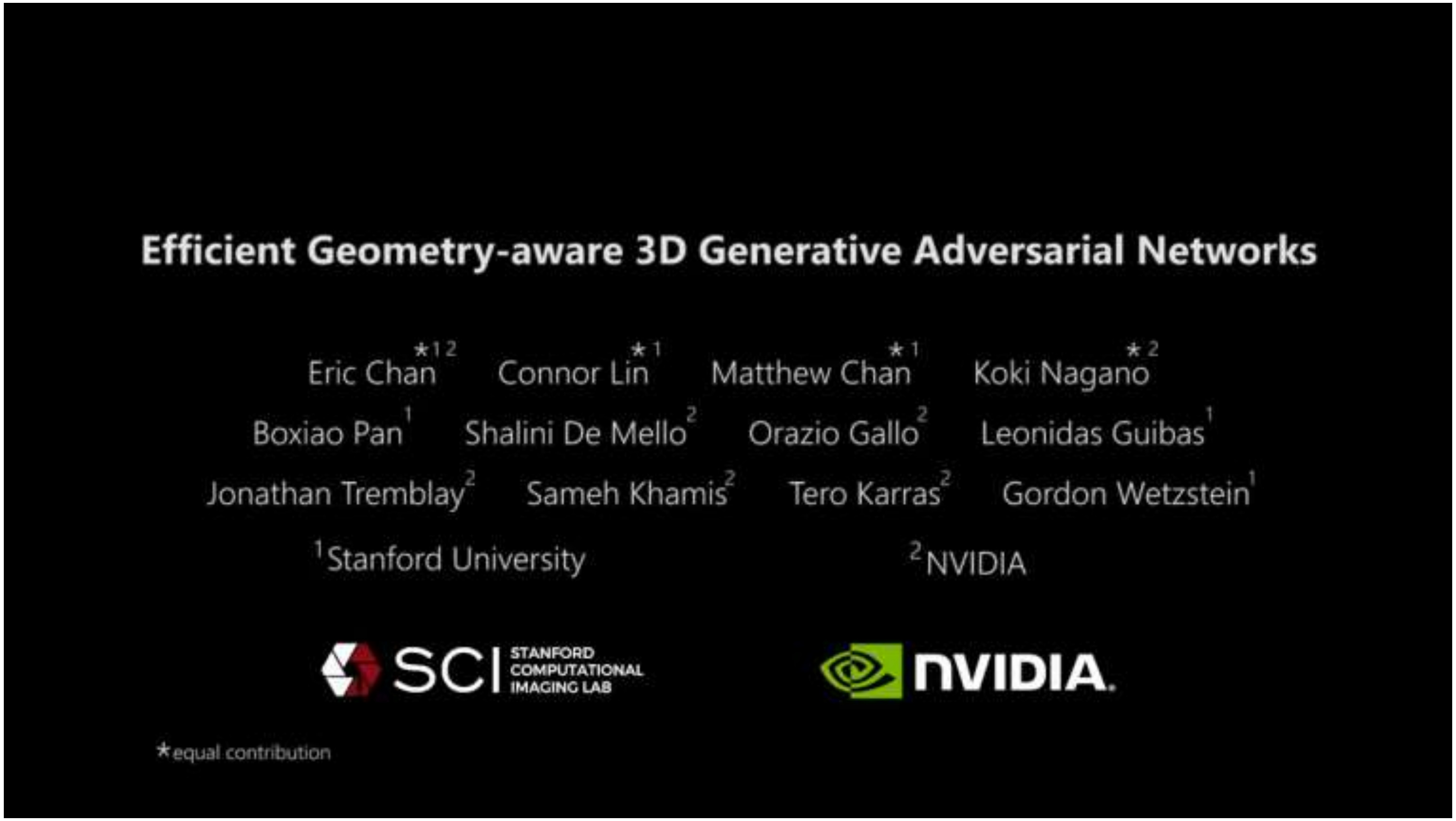


(a) NeRF (Implicit) (b) Voxels (Explicit or Hybrid) (c) Ours (Hybrid)

EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks, Chan et al., *CVPR 2022*

Tri-plane feature representation

- Pack the continue feature manifold into planes



EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks, Chan et al., *CVPR 2022*

Tri-plane feature representation

- Pack the continue feature manifold into planes

EG3D: Efficient Geometry-aware 3D Generative Adversarial Networks, Chan et al., *CVPR 2022*

Tensorial Radiance Fields

- Similar Plane-based feature representation
- Adopt tensor factorization

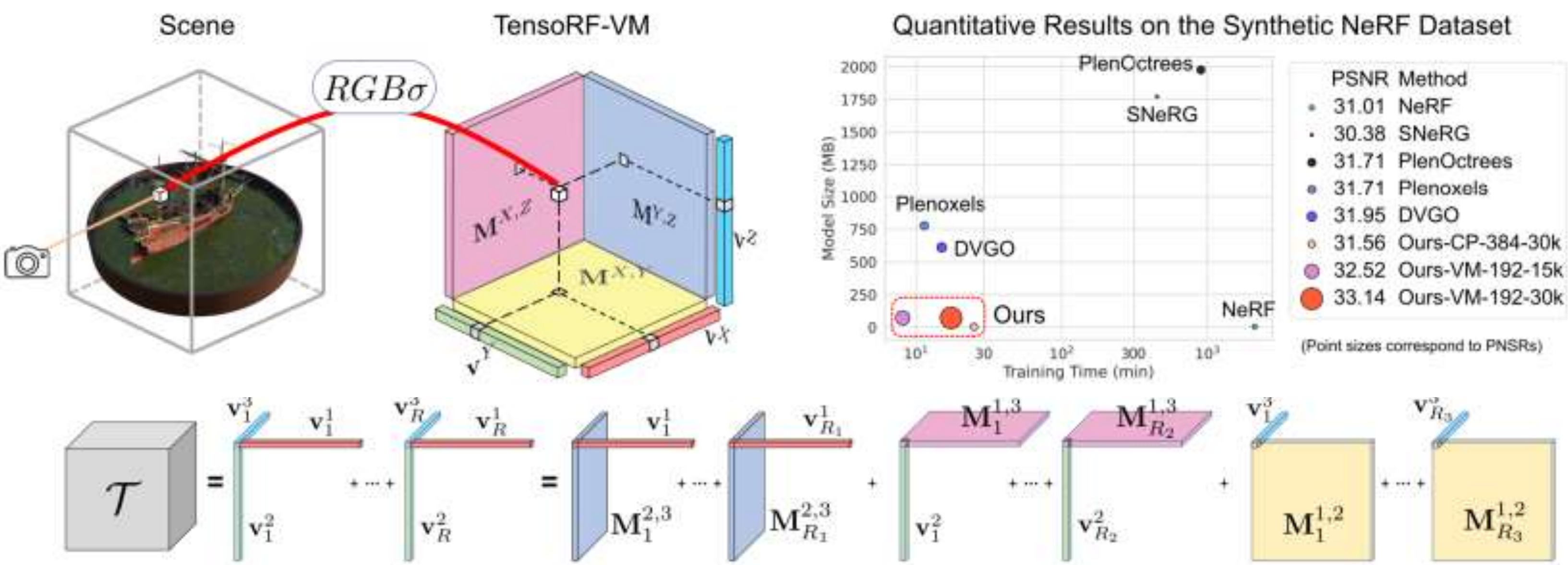


Fig. 2: Tensor factorization. Left: CP decomposition (Eqn. 1), which factorizes a tensor as a sum of vector outer products. Right: our vector-matrix decomposition (Eqn. 3), which factorizes a tensor as a sum of vector-matrix outer products.

TensoRF: Tensorial Radiance Fields, Chen et al., *ECCV 2022*

Tensorial Radiance Fields

- Similar Plane-based feature representation
- Adopt tensor factorization

TensoRF: Tensorial Radiance Fields

Anpei Chen* — ShanghaiTech University

Zexiang Xu* — Adobe Research

Andreas Geiger — University of Tübingen, MPI-IS, Tübingen

Jingyi Yu — ShanghaiTech University

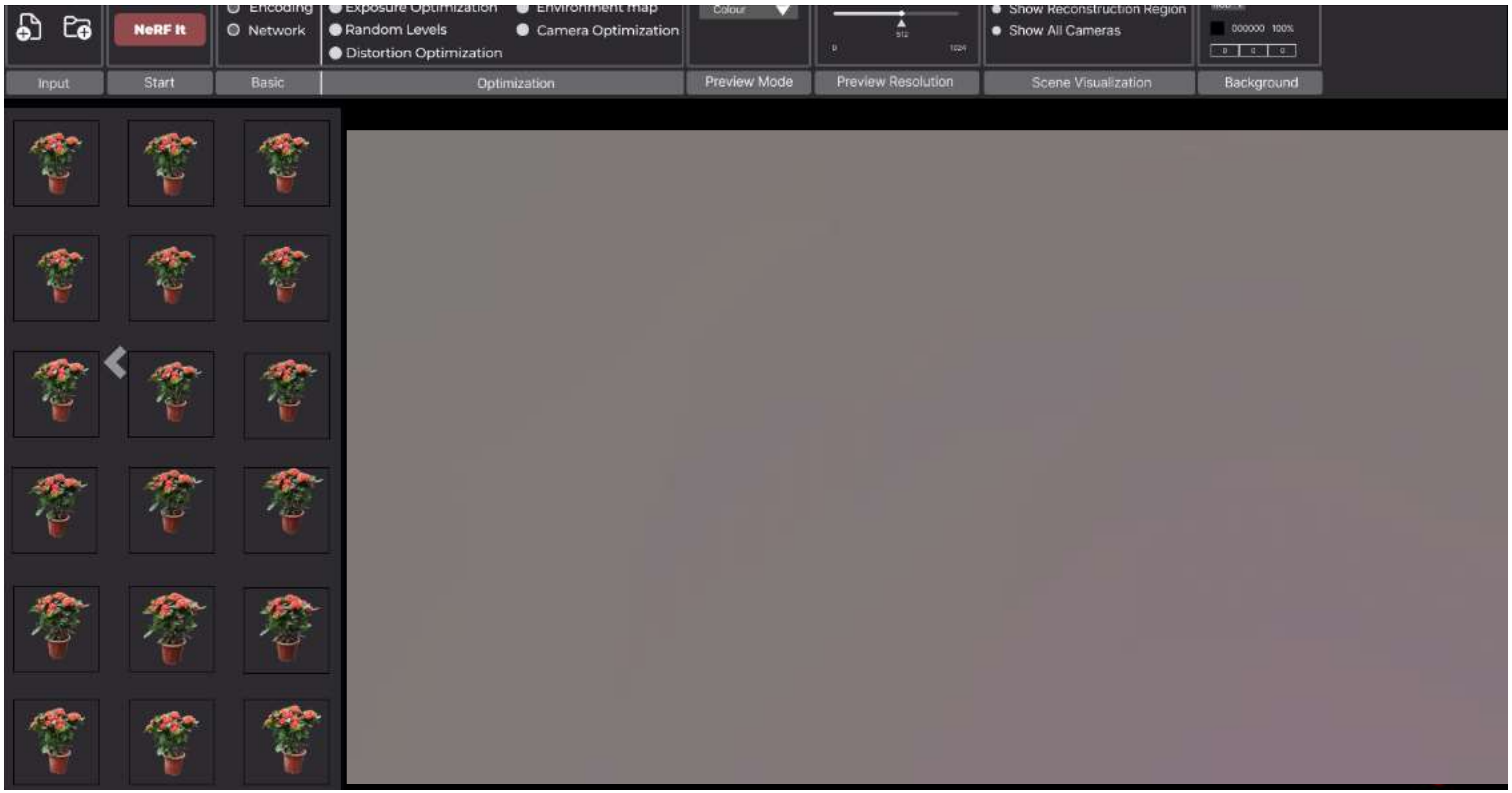
Hao Su — UC San Diego

*Denotes Equal contribution

TensoRF: Tensorial Radiance Fields, Chen et al., *ECCV 2022*

Recall: Neural Engine for Static Scene

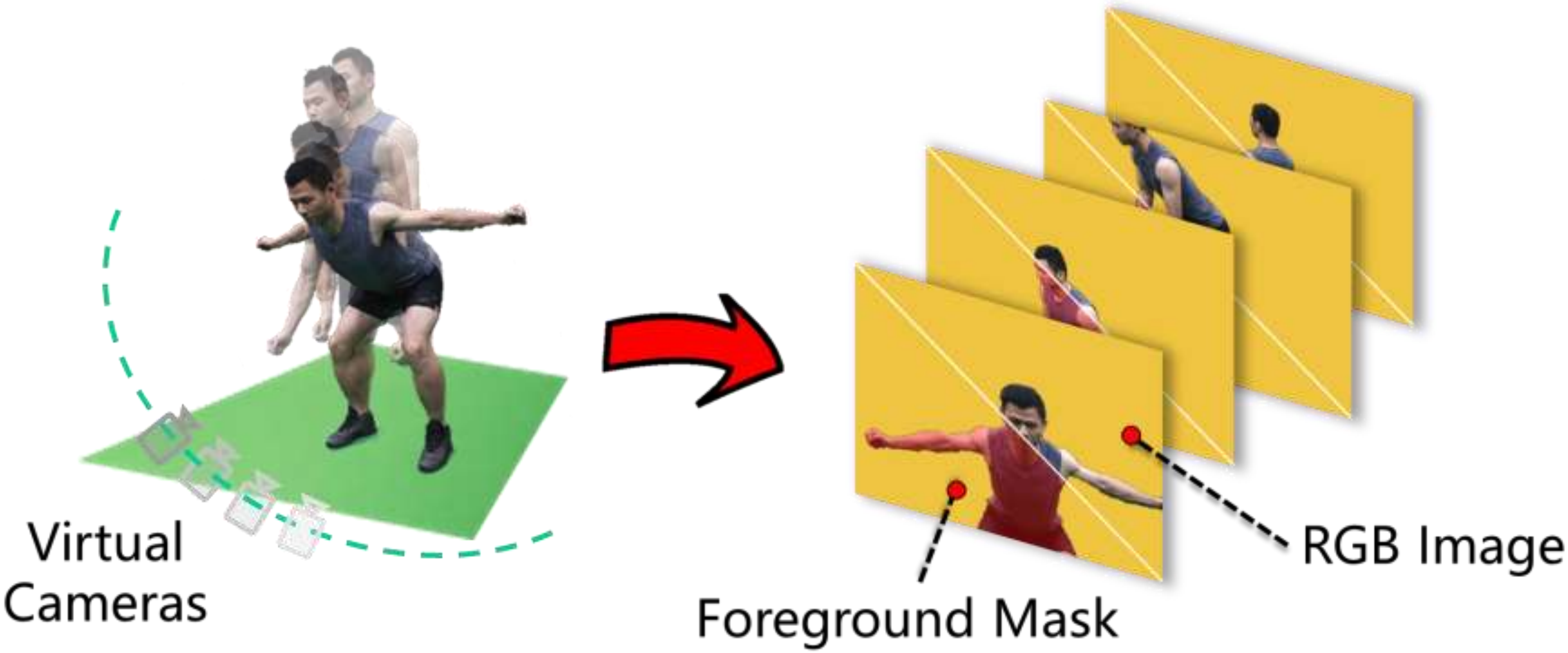
- Various attributes: appearance, geometry, etc.



Human Performance Modeling and Rendering via Neural Animated Mesh, Zhao et al., *ACM Transactions on Graphics (SIGGRAPH ASIA 2022)*

Neural Engine for Dynamic Scene

- Photo-realistic Neural Human Rendering
- Inherent Attribute Modeling



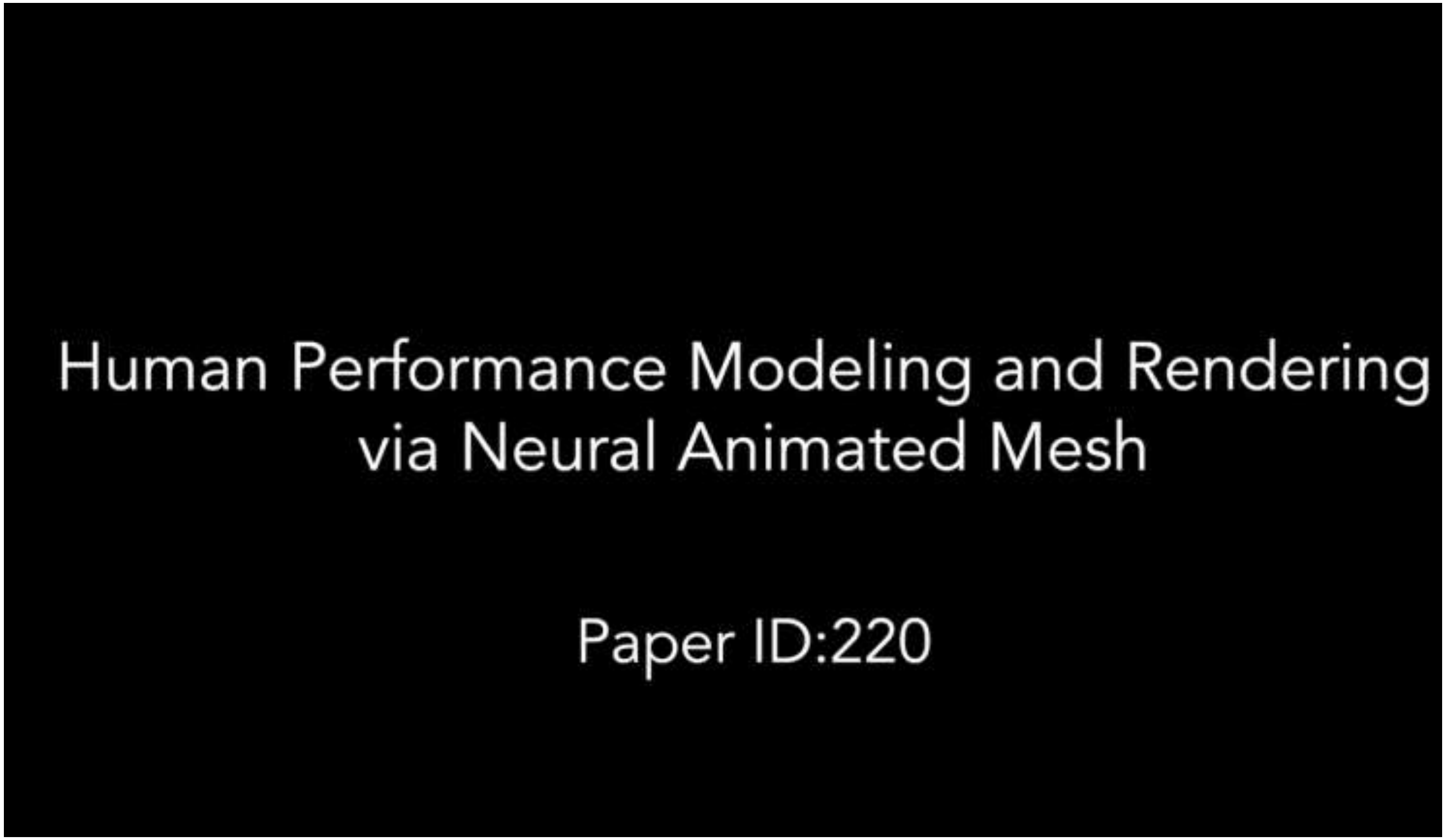
Neural Engine for Dynamic Scene

- Why neural: element editing; custom scene design

Editable Free-Viewpoint Video using a Layered Neural Representation, Zhang et al., *ACM Transactions on Graphics (SIGGRAPH 2021)*

Neural Engine for Dynamic Scene

- Why neural: ultra-fast, per-frame static to dynamic



Human Performance Modeling and Rendering via Neural Animated Mesh, Zhao et al., *ACM Transactions on Graphics (SIGGRAPH ASIA 2022)*

Neural Engine for Dynamic Scene

- Even Sparse-view and generalizable setting

NeuralHumanFVV: Real-Time Neural Volumetric Human Performance Rendering using RGB Cameras, Suo et al., *IEEE CVPR 2021*

Neural Engine for Dynamic Scene

- Even generalize to multi-person/ human-object interactions

NeuralHOFusion: Neural Volumetric Rendering under Human-object Interactions, Jiang et al., *IEEE CVPR 2022*

Neural Engine for Dynamic Scene

- Even generalize to multi-person/ human-object interactions

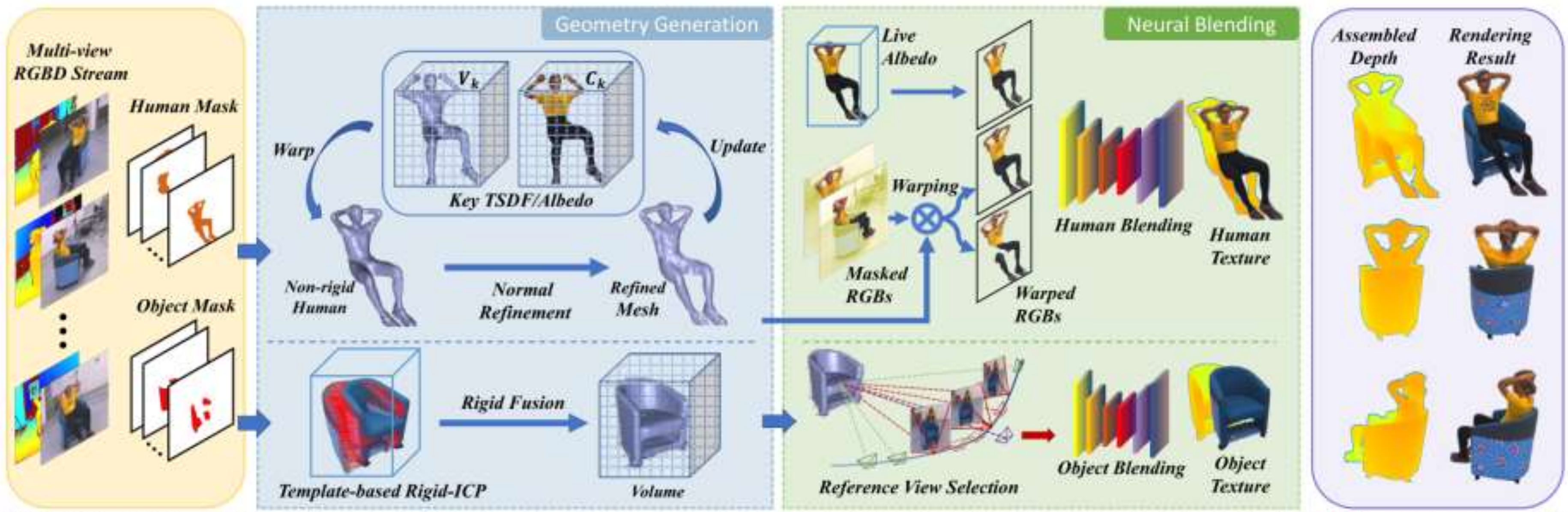


Figure 2. Our approach consists of twos stages. The geometry module includes neural human reconstruction(Sec. 4.1) and template-aid object fusion (Sec. 4.2), and the blending module includes neural human blending (Sec. 4.3) and temporal neural object blending(Sec. 4.4).

NeuralHOFusion: Neural Volumetric Rendering under Human-object Interactions, Jiang et al., *IEEE CVPR 2022*

Neural Engine for Dynamic Scene

- Even for monocular RGB and Human-object setting

Instant-NVR: Instant Neural Volumetric Rendering for Human-object Interaction from Monocular RGBD Stream

Yuheng Jiang[1,2*] Kaixin Yao[1,2*] Zhuo Su[3] Zhehao Shen[1] Haimin Luo[1] Lan Xu[1]
[1]ShanghaiTech University [2]NeuDim [3]Pico IDL,ByteDance

Instant-NVR: Instant Neural Volumetric Rendering for Human-object Interactions from Monocular RGBD Stream, Jiang et al., *IEEE CVPR 2023*

Neural Engine for Dynamic Scene

- Even for monocular RGB and Human-object setting
- Tracking-(mapping)-rendering

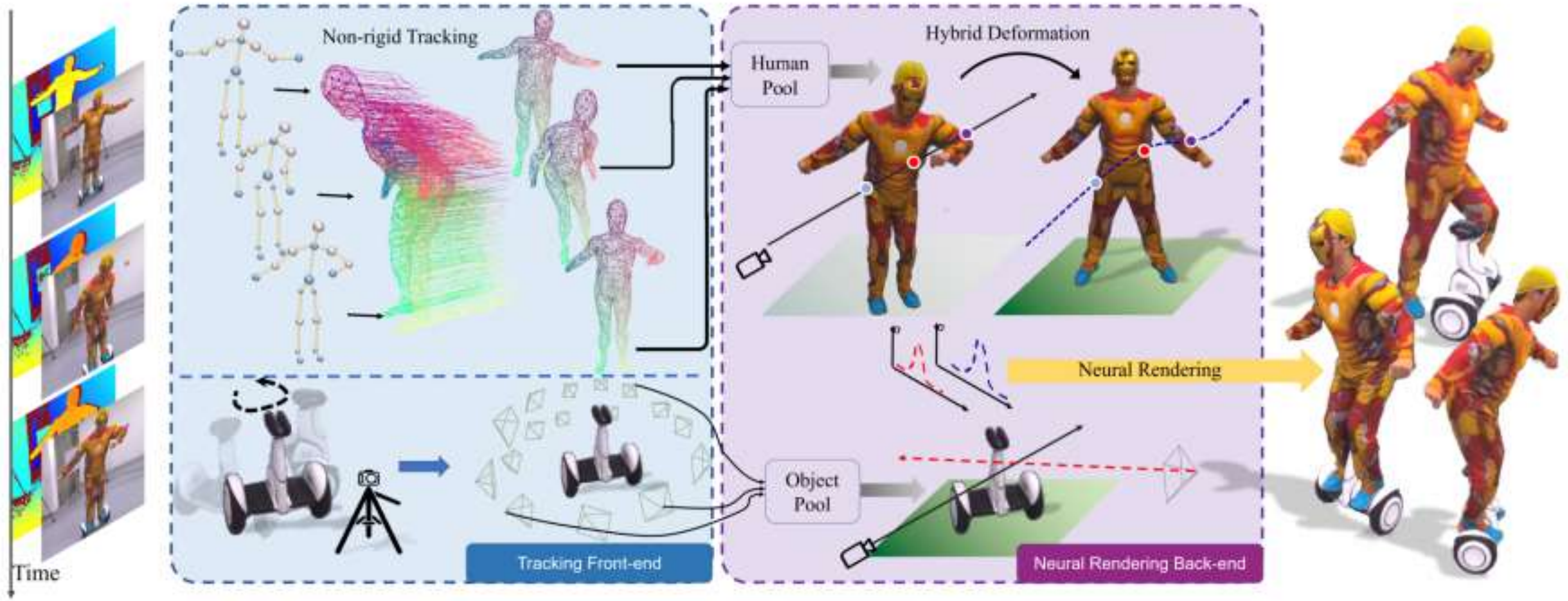
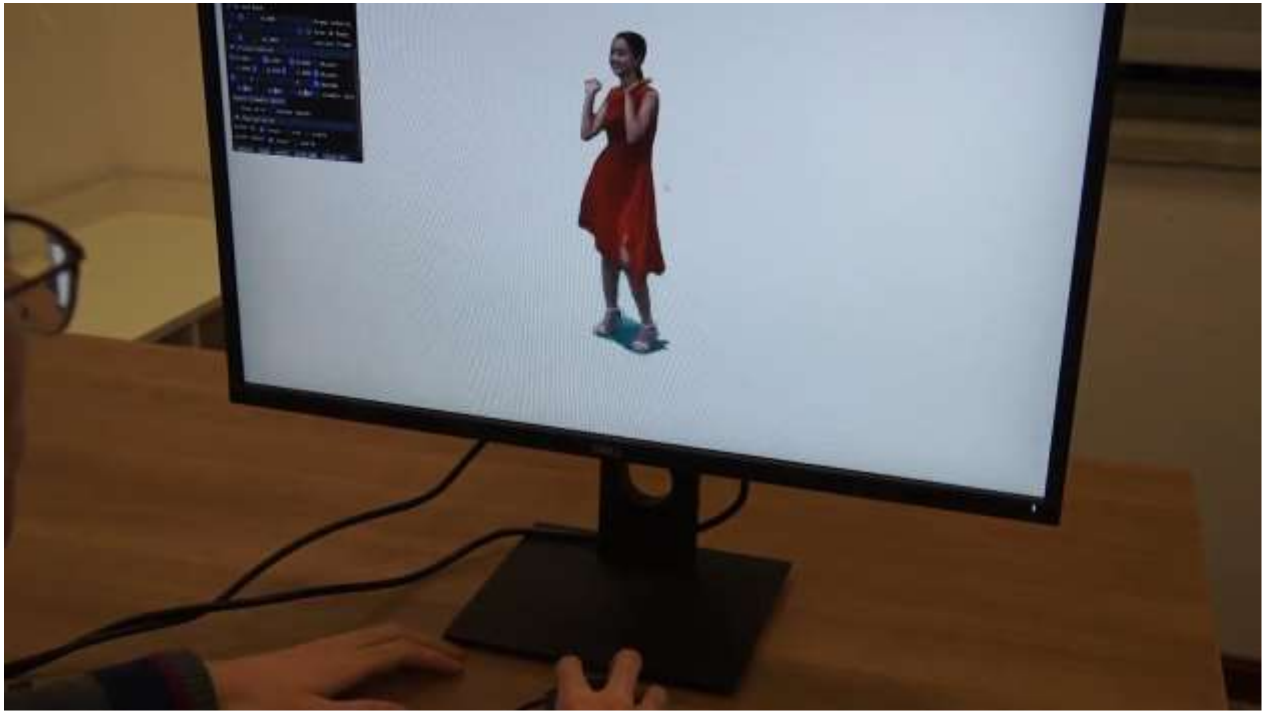


Figure 2. Our approach consists of two stages. The tracking front-end (Sec. 4.1) captures human and object motions, while the rendering back-end (Sec. 4.2) separately reconstructs the human-object radiance fields on-the-fly, for instant novel view synthesis with photo-realism.

Instant-NVR: Instant Neural Volumetric Rendering for Human-object Interactions from Monocular RGBD Stream, Jiang et al., *IEEE CVPR 2023*

Neural Engine for Dynamic Scene

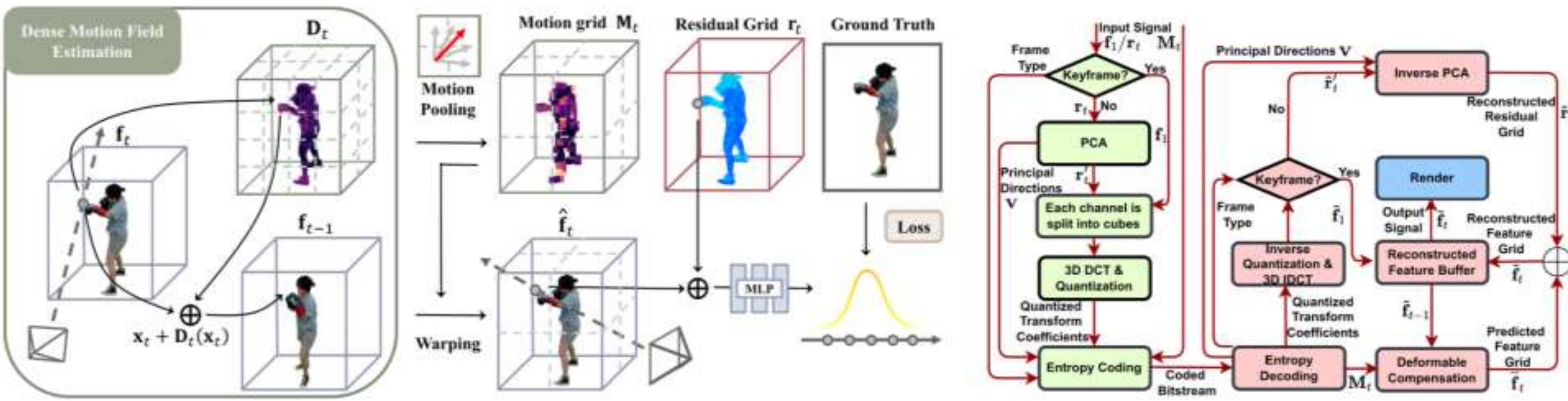
- Key idea: how to view the 4D spatial-temporally continue feature manifolds
- Fourier PlenOctree: 4D feature compression



Fourier PlenOctrees for Dynamic Radiance Field Rendering in Real-time, Wang et al., *IEEE CVPR 2022*

Neural Engine for Dynamic Scene

- Key idea: how to view the 4D spatial-temporally continue feature manifolds
  - Streamable feature compression
  - Residual Radiance Field with codec for arbitrary long sequences



Neural Residual Radiance Fields for Streamable Free-Viewpoint Videos, Wang et al., *IEEE CVPR 2023*

Neural Engine for Dynamic Scene

- Key idea: how to view the 4D spatial-temporally continue feature manifolds
  - Streamable feature compression
  - Residual Radiance Field with codec for arbitrary long sequences

Neural Residual Radiance Fields for Streamable Free-Viewpoint Videos, Wang et al., *IEEE CVPR 2023*

Neural Engine for Dynamic Scene

- Key idea: how to view the 4D spatial-temporally continue feature manifolds
  - Streamable feature compression
  - Residual Radiance Field with codec for arbitrary long sequences

Neural Residual Radiance Fields for Streamably Free-Viewpoint Videos

Liao Wang[1,3], Qiang Hu[1], Qihan He[1,4], Ziyu Wang[1], Jingyi Yu[1]
Tinne Tuytelaars[2], Lan Xu[1], Minye Wu[2]

CVPR 2023

[1]ShanghaiTech University, [2]KU Leuven, [3]NeuDim, [4]DGene

上海科技大学 ShanghaiTech University KU LEUVEN NeuDim DGene

Neural Residual Radiance Fields for Streamable Free-Viewpoint Videos, Wang et al., *IEEE CVPR 2023*

Neural Engine for Dynamic Scene

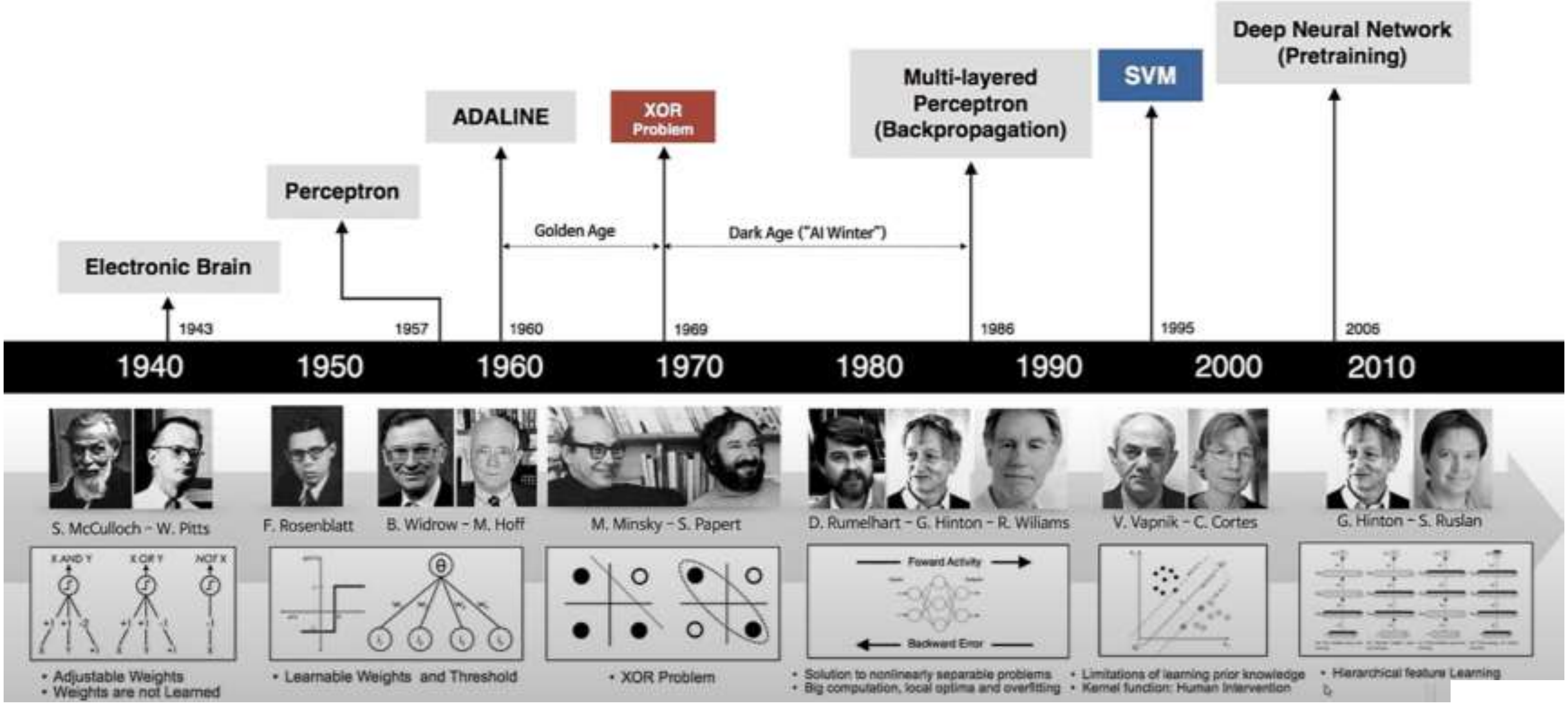
- Key idea: how to view the 4D spatial-temporally continue feature manifolds
  - Streamable feature even on mobile devices?

VideoRF: Rendering Dynamic Radiance Fields as 2D Feature Video Streams, Wang et al., *Arxiv 2023*

A quick summary……

Summary: Why deep learning?

- A long story with a huge recent success



Summary: Why deep learning?

- A long story with a huge recent success

ACM Turing Award 2019 (Nobel Prize of Computing)
Yann LeCun, Geoffrey Hinton, and Yoshua Bengio

Summary: Why deep learning?

- Huge recent success
- with **chaos**, **peer-pressure**, or even **misunderstandings**

This guy didn't know about neural networks

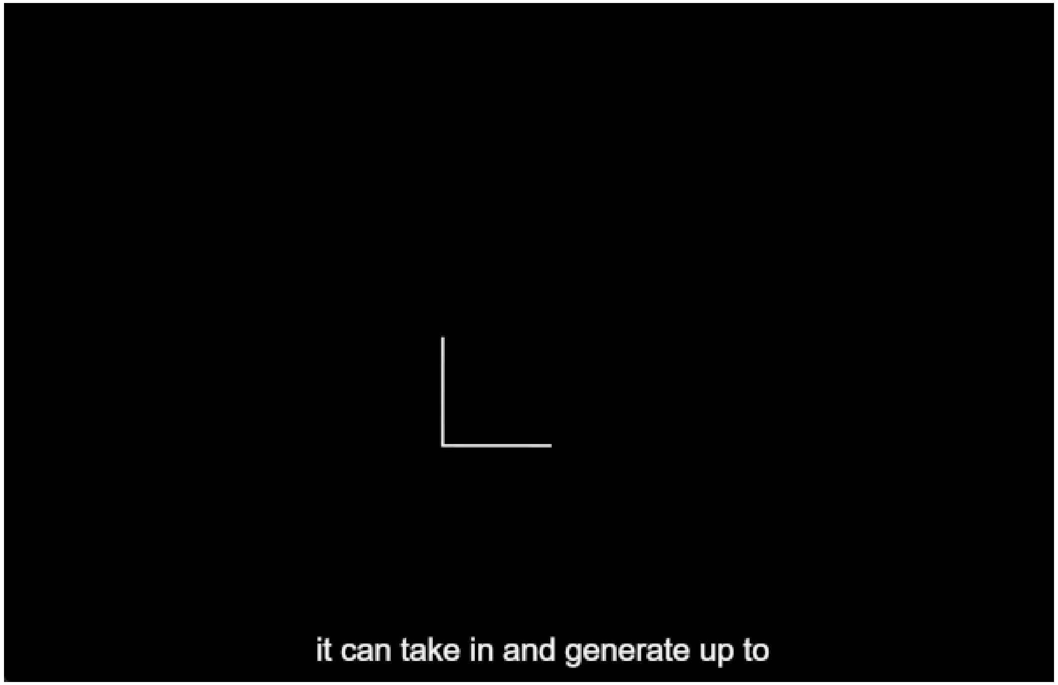
This guy learned about neural networks

Summary: Why deep learning?

- Huge recent success
- with **chaos**, **peer-pressure**, or even **misunderstandings**

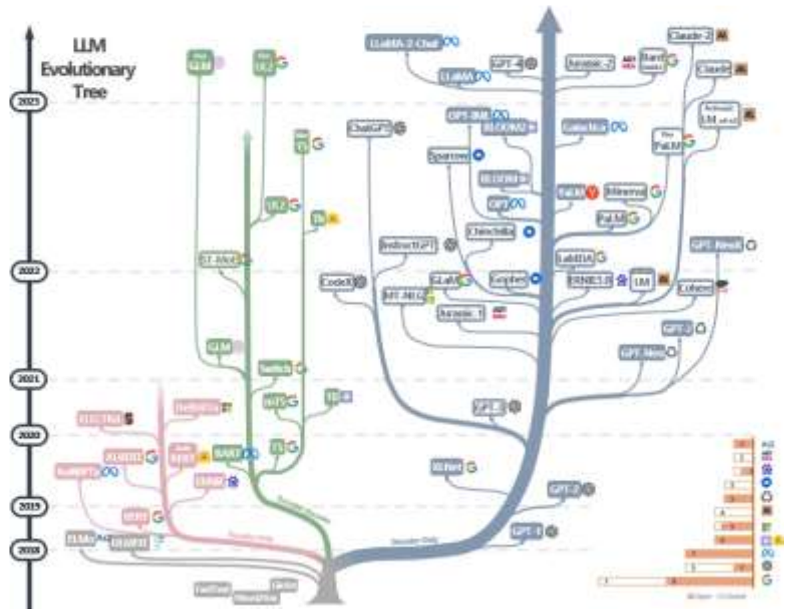
Summary: Why deep learning?

- New paradigm IS happening!
- Large model, Large computing, Large data
- **Move fast** and be **open-minded** to the new party



Summary: Why deep learning?

- New paradigm IS happening!
- Large model, Large computing, Large data
- **Move fast** and be **open-minded** to the new party



Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond

DALL-E 3 integrates with **ChatGPT** so users don't have to think of prompts anymore

Keep moving, my friends ……