Introduction to Neural Scene Representation and Neural Rendering

Lingjie Liu

Penn
University of Pennsylvania

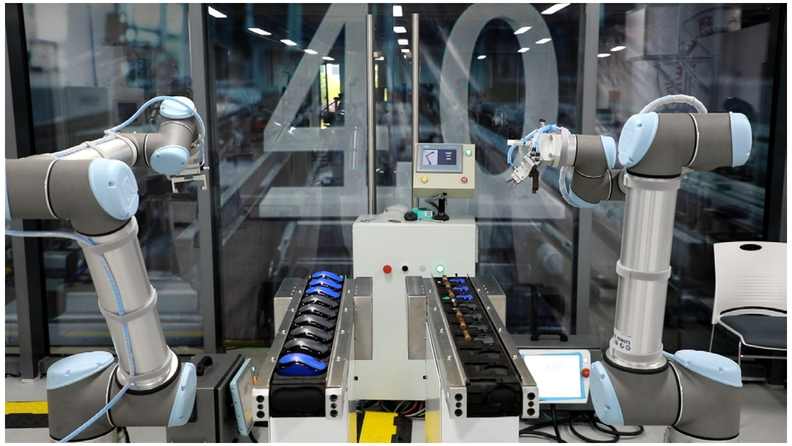
We Live in a World that is 3D and Contains Dynamics

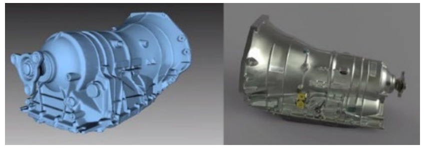
We Digitize Our World in 3D

Future AI: Towards 3D Aware

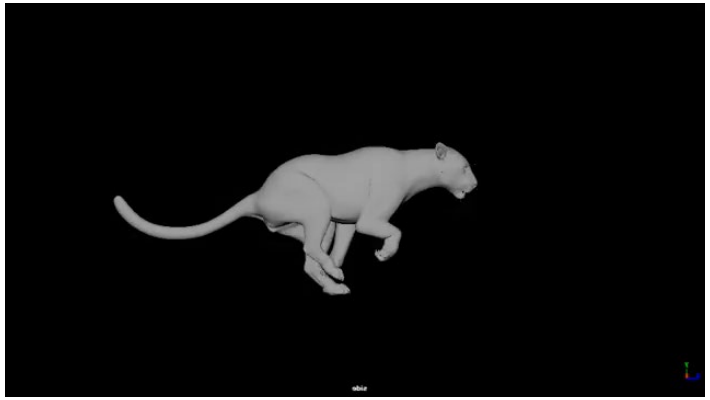
Lingjie Liu

3D Reconstruction of Real-world Scenes





Geometry
+ Appearance



Motion
+ Deformation

Photo-realistic Rendering

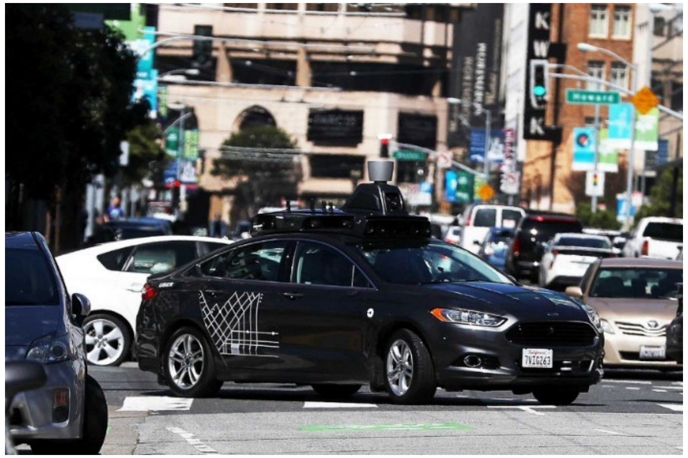
- **Image Synthesis** of Real-world Scenes with 3D Control.

Applications

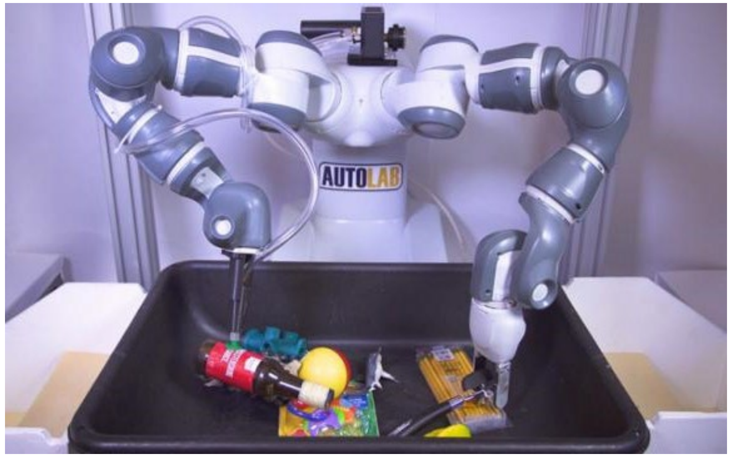
AR / VR

Gaming / Movie

Healthcare



Autonomous Driving



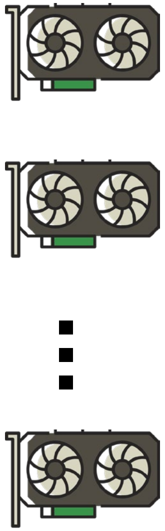
Robot Grasping

Human-robot Interaction

Why are they challenging?

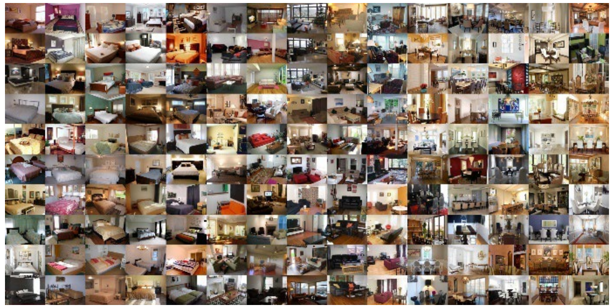
Problem formulation

Captured images → Processing → Rendering of real-world place



[Mildenhall et al., Neural Radiance Fields (NeRF), ECCV 2020]
[Wu et al., Scalable Neural Indoor Scene Rendering, SIGGRAPH 2022]

Classical Computer Graphics Pipeline



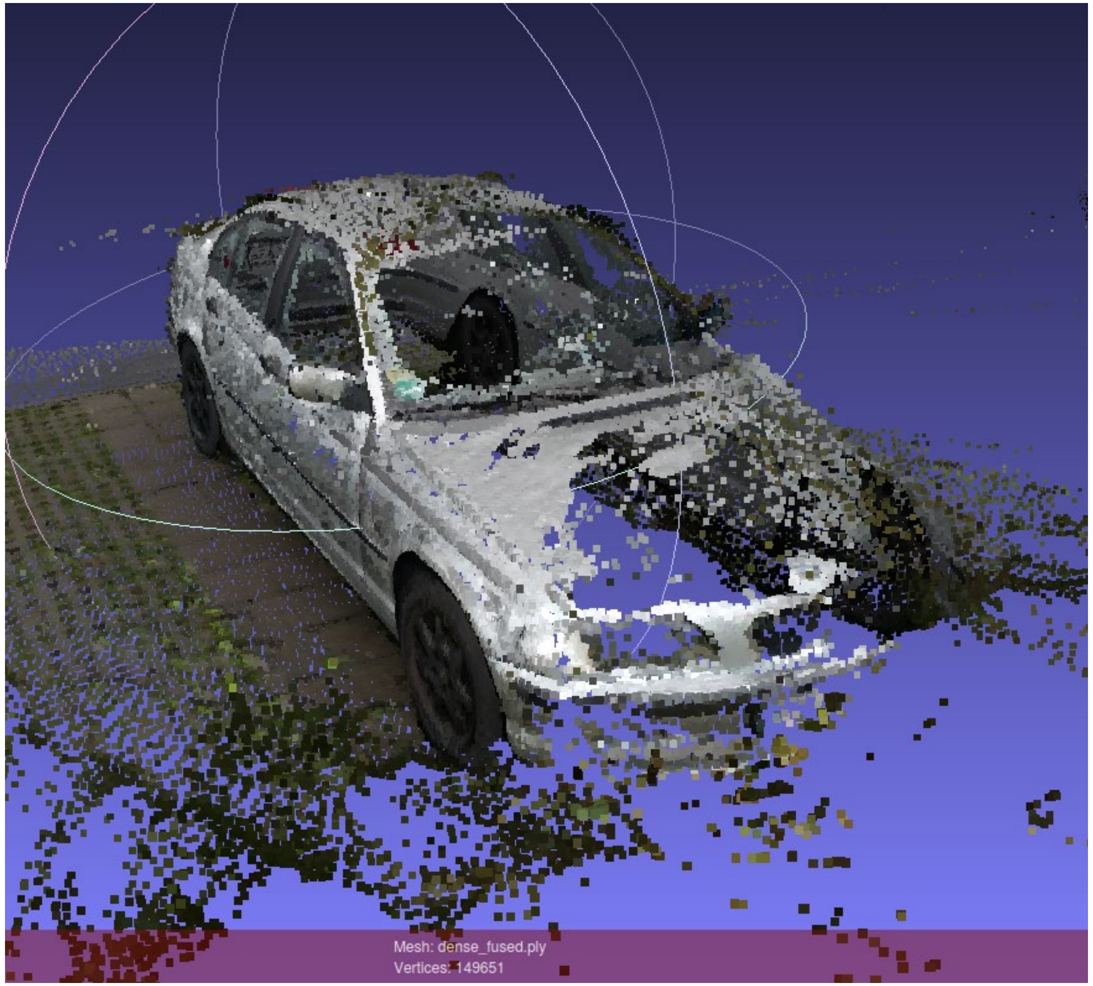
3D Reconstruction

Image Synthesis

Image-based 3D Reconstruction

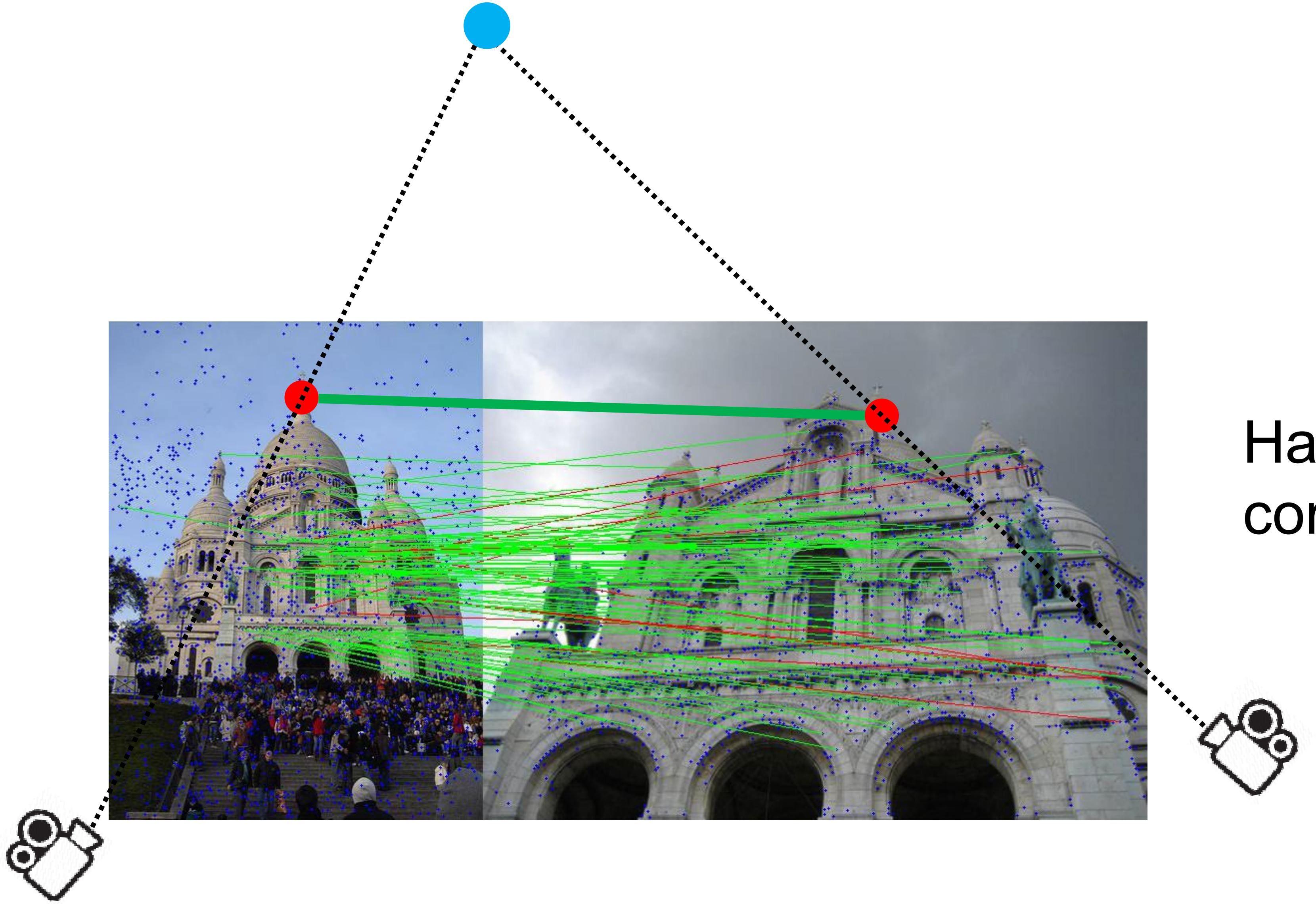
Computer Graphics Rendering

Image-based 3D Reconstruction



COLMAP [Johannes et al. 2016, Schoenberger et al. 2016]
(Input: 100 images)

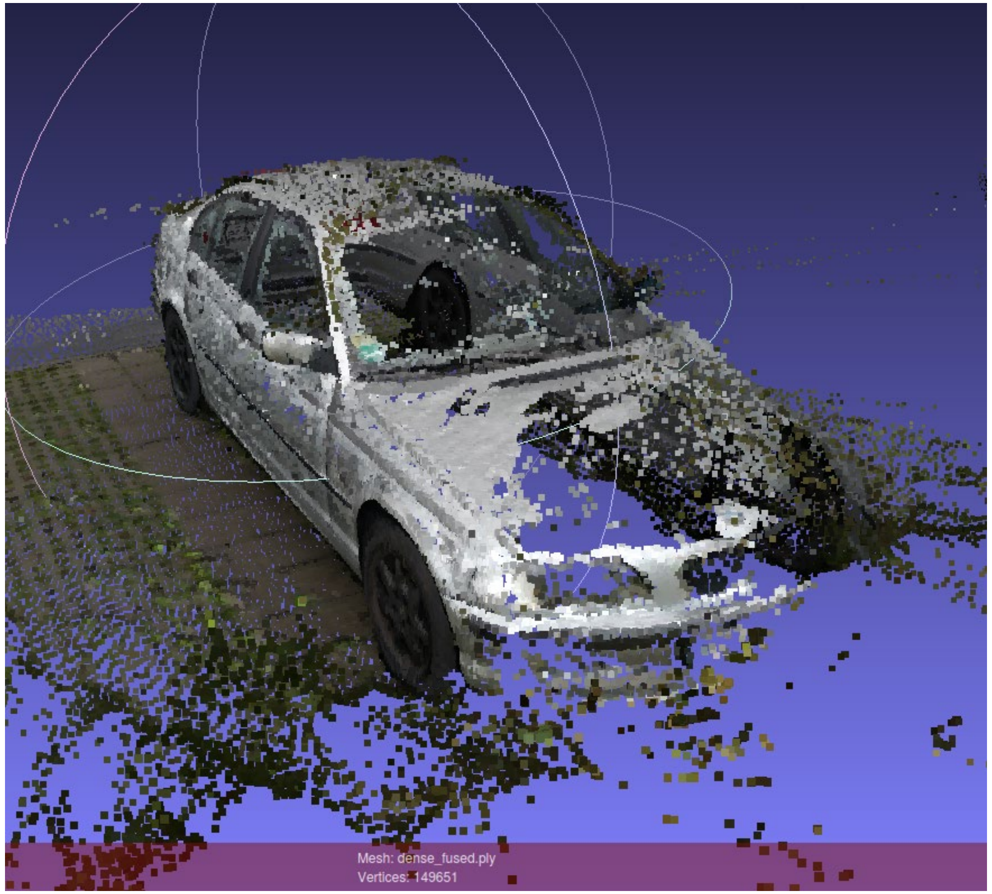
Challenges in Image-based Reconstruction



Hard to extract reliable correspondences!

Computer Graphics Rendering

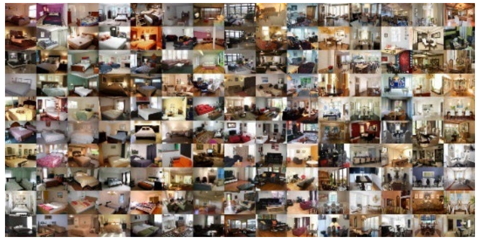
Rendering requires very high-quality 3D models



VS

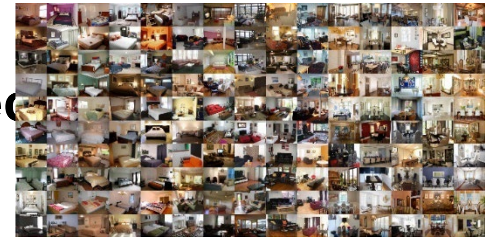
Neural Scene Representation and Neural Rendering
To the rescue

Neural Scene Representation and Neural Rendering



Scene Representation

Neural Rendering



3D Reco

Sy

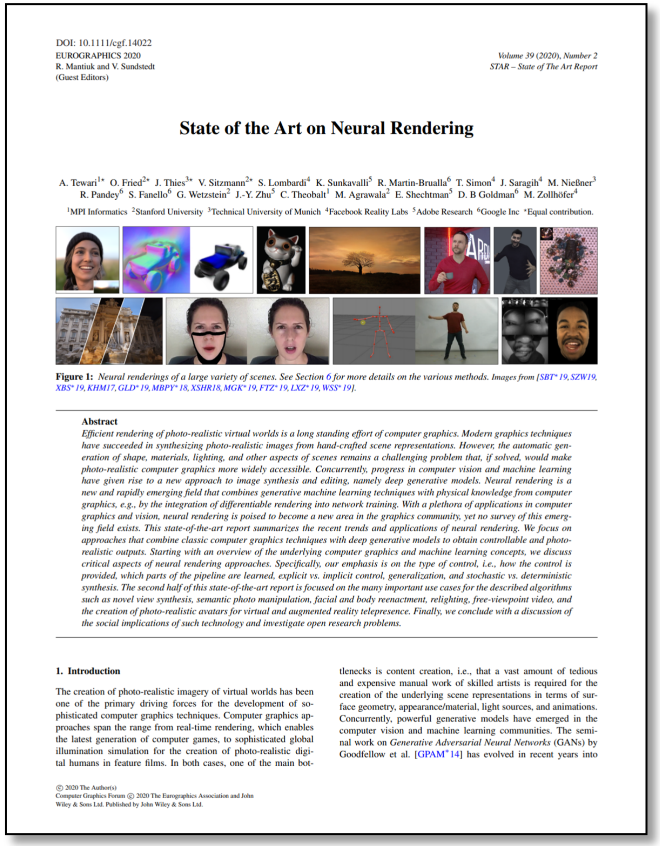
Image Loss

Co puter Graphics Rende g

Lingjie Liu

Neural Scene Representation and Neural Rendering

Neural Rendering



DOI: 10.1111/cgf.14022
EUROGRAPHICS 2020
R. Mantiuk and V. Sundstedt
(Guest Editors)

Volume 39 (2020), Number 2
STAR – State of The Art Report

State of the Art on Neural Rendering

A. Tewari[1*] O. Fried[2*] J. Thies[3*] V. Sitzmann[2*] S. Lombardi[4] K. Sunkavalli[5] R. Martin-Brualla[6] T. Simon[4] J. Saragih[4] M. Nießner[3] R. Pandey[6] S. Fanello[6] G. Wetzstein[2] J.-Y. Zhu[5] C. Theobalt[1] M. Agrawala[2] E. Shechtman[5] D. B Goldman[6] M. Zollhöfer[4]

[1]MPI Informatics [2]Stanford University [3]Technical University of Munich [4]Facebook Reality Labs [5]Adobe Research [6]Google Inc *Equal contribution.

Figure 1: *Neural renderings of a large variety of scenes. See Section 6 for more details on the various methods.* Images from [SBT* 19, SZW19, XBS* 19, KHM17, GLD* 19, MBPY* 18, XSHR18, MGK* 19, FTZ* 19, LXZ* 19, WSS* 19].

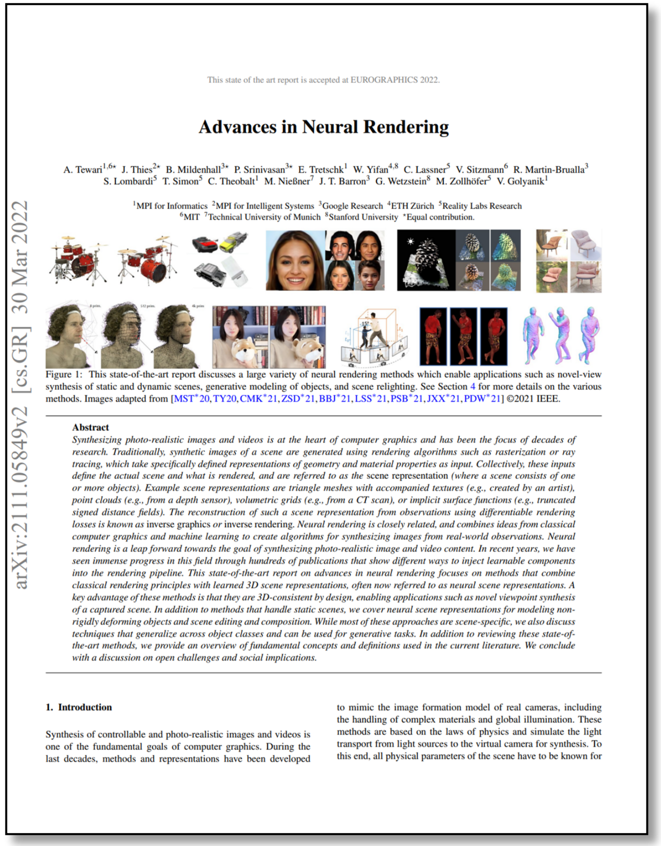
Abstract
Efficient rendering of photo-realistic virtual worlds is a long standing effort of computer graphics. Modern graphics techniques have succeeded in synthesizing photo-realistic images from hand-crafted scene representations. However, the automatic generation of shape, materials, lighting, and other aspects of scenes remains a challenging problem that, if solved, would make photo-realistic computer graphics more widely accessible. Concurrently, progress in computer vision and machine learning have given rise to a new approach to image synthesis and editing, namely deep generative models. Neural rendering is a new and rapidly emerging field that combines generative machine learning techniques with physical knowledge from computer graphics, e.g., by the integration of differentiable rendering into network training. With a plethora of applications in computer graphics and vision, neural rendering is poised to become a new area in the graphics community, yet no survey of this emerging field exists. This state-of-the-art report summarizes the recent trends and applications of neural rendering. We focus on approaches that combine classic computer graphics techniques with deep generative models to obtain controllable and photo-realistic outputs. Starting with an overview of the underlying computer graphics and machine learning concepts, we discuss critical aspects of neural rendering approaches. Specifically, our emphasis is on the type of control, i.e., how the control is provided, which parts of the pipeline are learned, explicit vs. implicit control, generalization, and stochastic vs. deterministic synthesis. The second half of this state-of-the-art report is focused on the many important use cases for the described algorithms such as novel view synthesis, semantic photo manipulation, facial and body reenactment, relighting, free-viewpoint video, and the creation of photo-realistic avatars for virtual and augmented reality telepresence. Finally, we conclude with a discussion of the social implications of such technology and investigate open research problems.

1. Introduction

The creation of photo-realistic imagery of virtual worlds has been one of the primary driving forces for the development of sophisticated computer graphics techniques. Computer graphics approaches span the range from real-time rendering, which enables the latest generation of computer games, to sophisticated global illumination simulation for the creation of photo-realistic digital humans in feature films. In both cases, one of the main bottlenecks is content creation, i.e., that a vast amount of tedious and expensive manual work of skilled artists is required for the creation of the underlying scene representations in terms of surface geometry, appearance/material, light sources, and animations. Concurrently, powerful generative models have emerged in the computer vision and machine learning communities. The seminal work on *Generative Adversarial Neural Networks* (GANs) by Goodfellow et al. [GPAM*14] has evolved in recent years into

© 2020 The Author(s)
Computer Graphics Forum © 2020 The Eurographics Association and John Wiley & Sons Ltd. Published by John Wiley & Sons Ltd.

[Tewari et al. 2020]



This state of the art report is accepted at EUROGRAPHICS 2022.

Advances in Neural Rendering

A. Tewari[1,6*] J. Thies[2*] B. Mildenhall[3*] P. Srinivasan[3*] E. Tretschk[1] W. Yifan[4,8] C. Lassner[5] V. Sitzmann[6] R. Martin-Brualla[3] S. Lombardi[5] T. Simon[5] C. Theobalt[1] M. Nießner[7] J. T. Barron[3] G. Wetzstein[8] M. Zollhöfer[5] V. Golyanik[1]

[1]MPI for Informatics [2]MPI for Intelligent Systems [3]Google Research [4]ETH Zürich [5]Reality Labs Research [6]MIT [7]Technical University of Munich [8]Stanford University *Equal contribution.

arXiv:2111.05849v2 [cs.GR] 30 Mar 2022

Figure 1: This state-of-the-art report discusses a large variety of neural rendering methods which enable applications such as novel-view synthesis of static and dynamic scenes, generative modeling of objects, and scene relighting. See Section 4 for more details on the various methods. Images adapted from [MST*20, TY20, CMK*21, ZSD*21, BBJ*21, LSS*21, PSB*21, JXX*21, PDW*21] ©2021 IEEE.

Abstract
Synthesizing photo-realistic images and videos is at the heart of computer graphics and has been the focus of decades of research. Traditionally, synthetic images of a scene are generated using rendering algorithms such as rasterization or ray tracing, which take specifically defined representations of geometry and material properties as input. Collectively, these inputs define the actual scene and what is rendered, and are referred to as the scene representation *(where a scene consists of one or more objects). Example scene representations are triangle meshes with accompanied textures (e.g., created by an artist), point clouds (e.g., from a depth sensor), volumetric grids (e.g., from a CT scan), or implicit surface functions (e.g., truncated signed distance fields). The reconstruction of such a scene representation from observations using differentiable rendering losses is known as* inverse graphics *or* inverse rendering*. Neural rendering is closely related, and combines ideas from classical computer graphics and machine learning to create algorithms for synthesizing images from real-world observations. Neural rendering is a leap forward towards the goal of synthesizing photo-realistic image and video content. In recent years, we have seen immense progress in this field through hundreds of publications that show different ways to inject learnable components into the rendering pipeline. This state-of-the-art report on advances in neural rendering focuses on methods that combine classical rendering principles with learned 3D scene representations, often now referred to as neural scene representations. A key advantage of these methods is that they are 3D-consistent by design, enabling applications such as novel viewpoint synthesis of a captured scene. In addition to methods that handle static scenes, we cover neural scene representations for modeling non-rigidly deforming objects and scene editing and composition. While most of these approaches are scene-specific, we also discuss techniques that generalize across object classes and can be used for generative tasks. In addition to reviewing these state-of-the-art methods, we provide an overview of fundamental concepts and definitions used in the current literature. We conclude with a discussion on open challenges and social implications.*

1. Introduction

Synthesis of controllable and photo-realistic images and videos is one of the fundamental goals of computer graphics. During the last decades, methods and representations have been developed to mimic the image formation model of real cameras, including the handling of complex materials and global illumination. These methods are based on the laws of physics and simulate the light transport from light sources to the virtual camera for synthesis. To this end, all physical parameters of the scene have to be known for

[Tewari et al. 2021]

Lingjie Liu

Neural Rendering - Definition

- Definition:

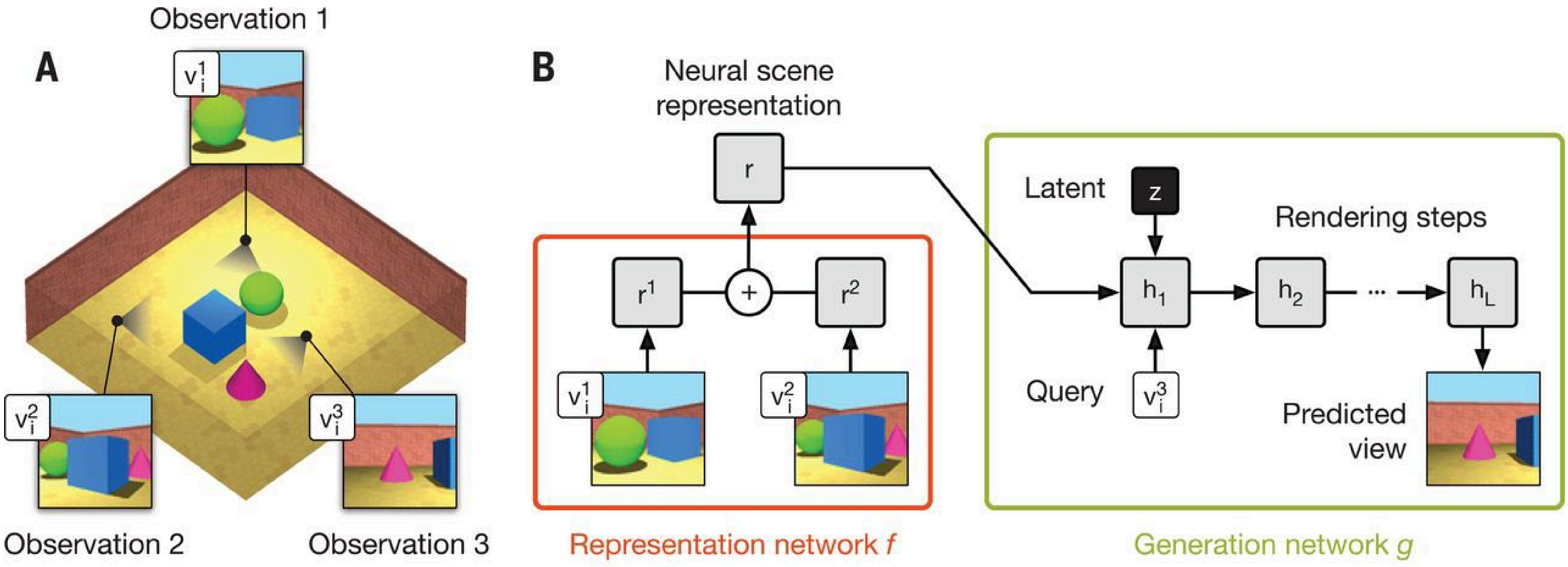
*"Deep neural networks for **image or video generation** that enable **explicit or implicit control** of **scene properties**"*

1) Generative networks that synthesis raw pixel output

2) Controllable by interpretable parameters or by video/audio input.

3) Illumination, camera, pose, geometry, appearance, or semantic structure controllable

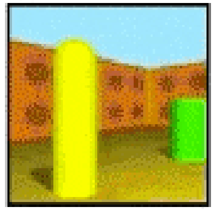
Generative Query Network (GQN)



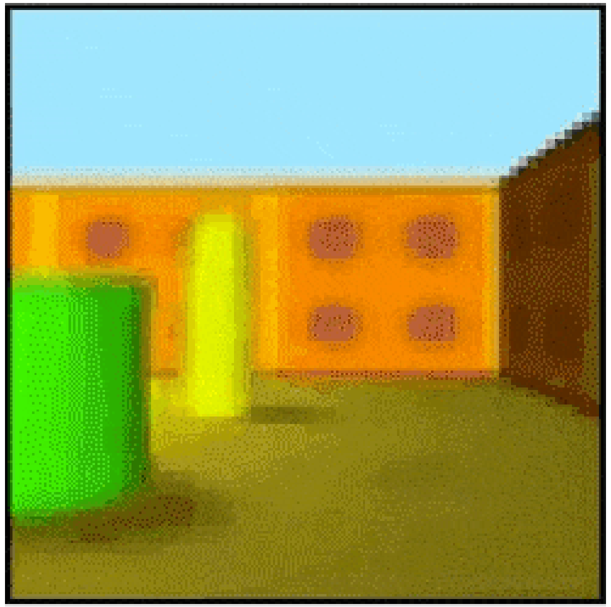
Neural scene representation and rendering, Eslami et al. 2018

Generative Query Network (GQN)

observation



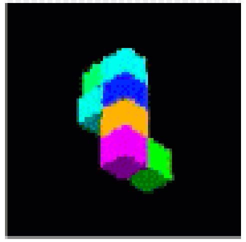
neural rendering



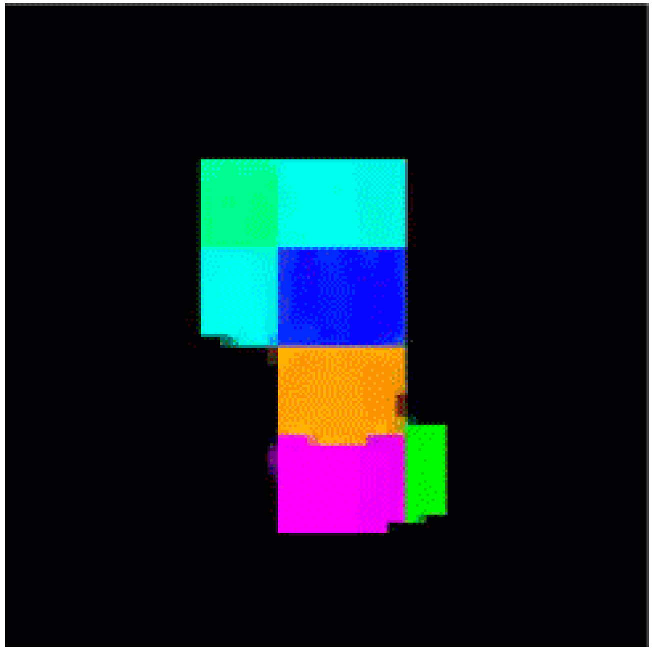
Neural scene representation and rendering, Eslami et al. 2018

Generative Query Network (GQN)

observation

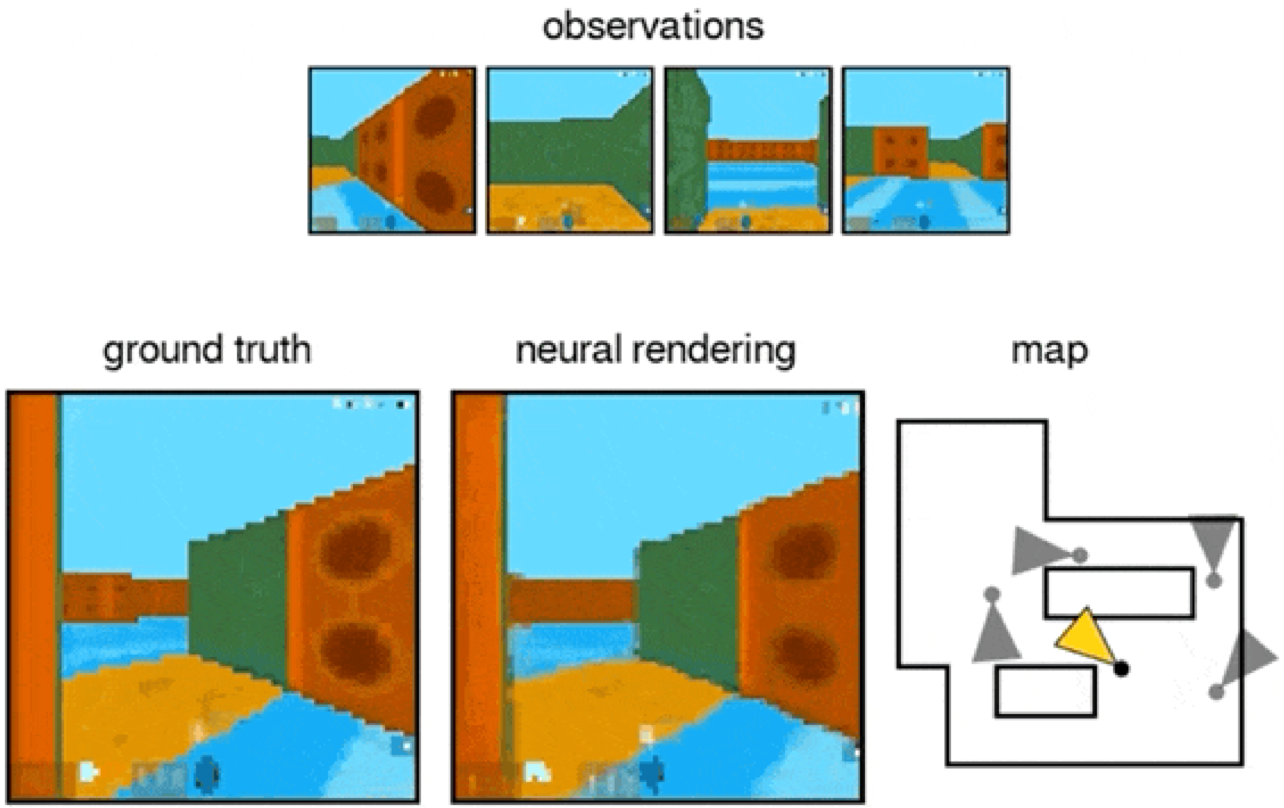


neural rendering



Neural scene representation and rendering, Eslami et al. 2018

Generative Query Network (GQN)



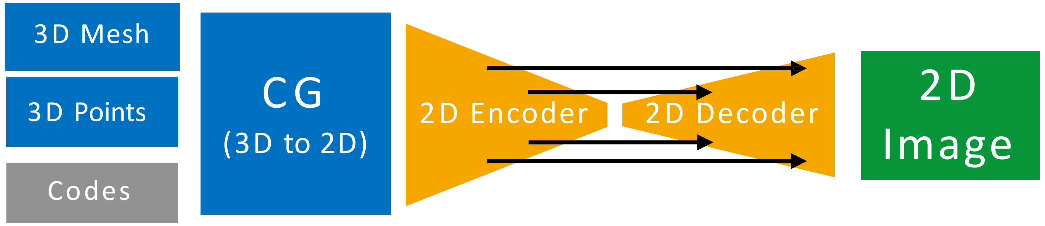
Neural scene representation and rendering, Eslami et al. 2018

Neural Rendering Zoo

“Regress it”

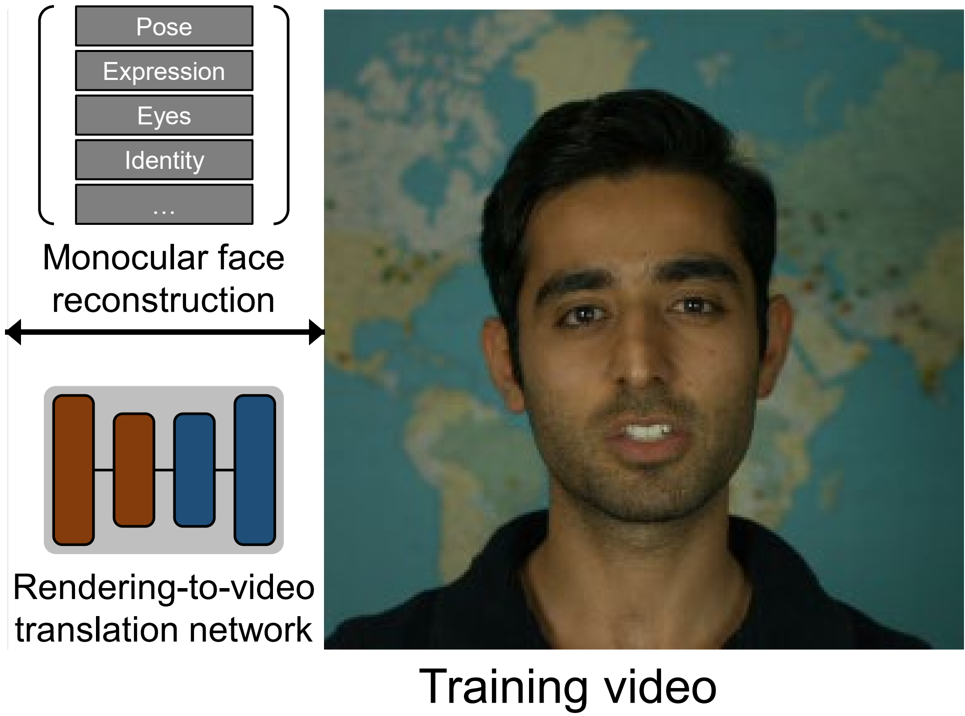
e.g., GQN

”Make it more real”



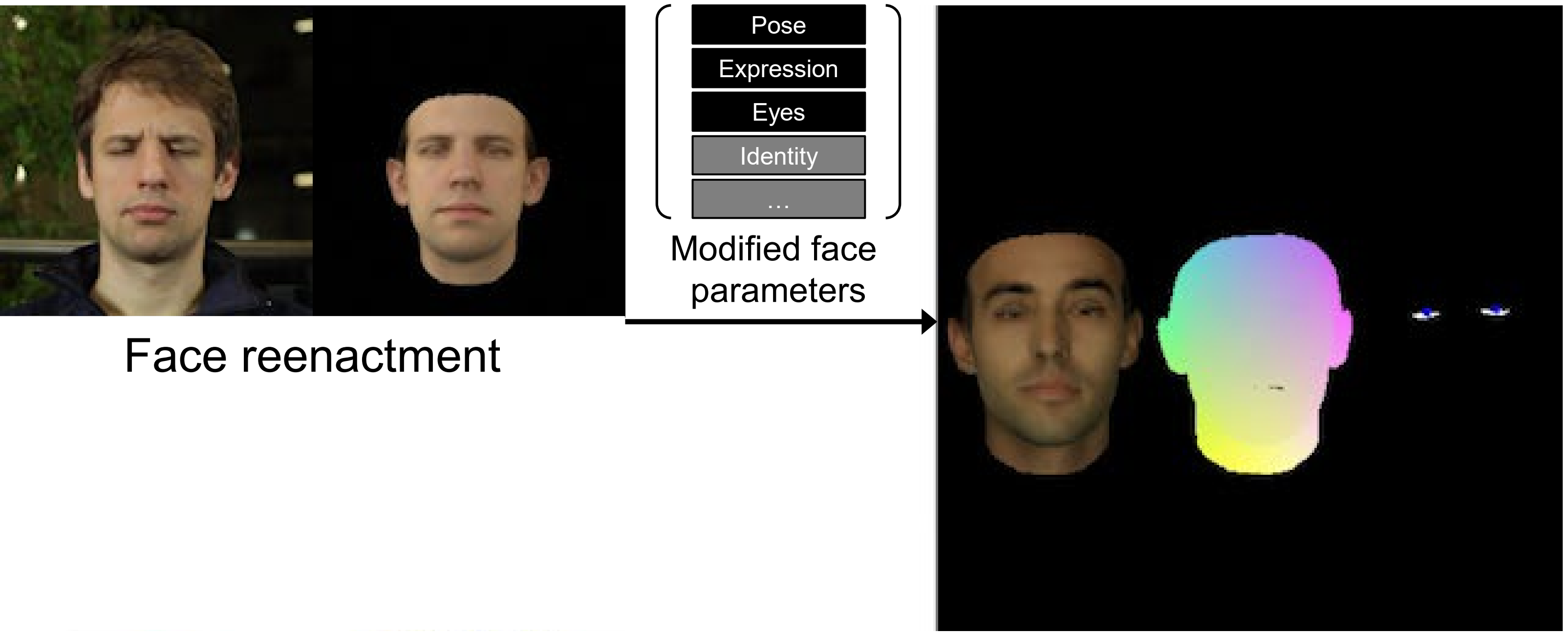
e.g., DVP or DNR

Deep Video Portraits (DVP)



Deep Video Portraits, Kim et al. 2018

Deep Video Portraits (DVP)



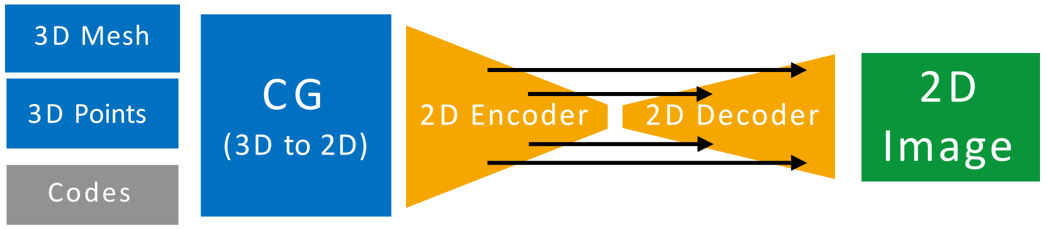
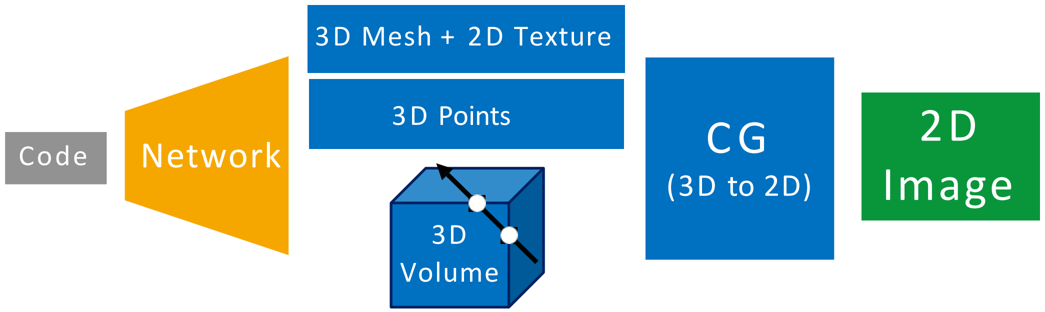
User interaction

Modified rendering

Deep Video Portraits, Kim et al. 2018

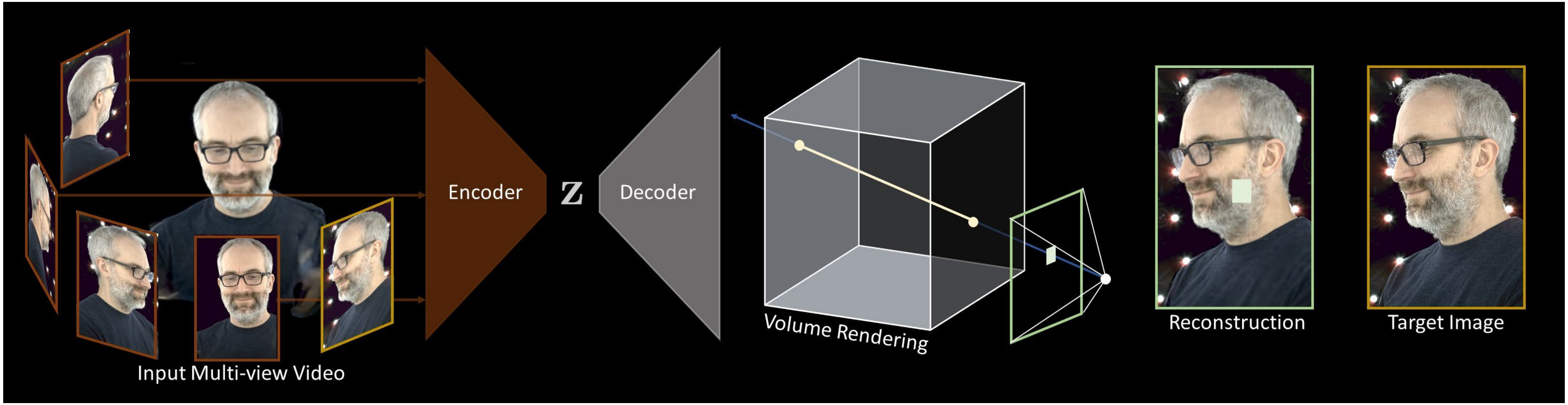
Penn
University of Pennsylvania

Neural Rendering Zoo

"Regress it"		e.g., GQN
"Make it more real"		e.g., DVP or DNR
"Regress & render"		e.g., Neural Volumes

Lingjie Liu

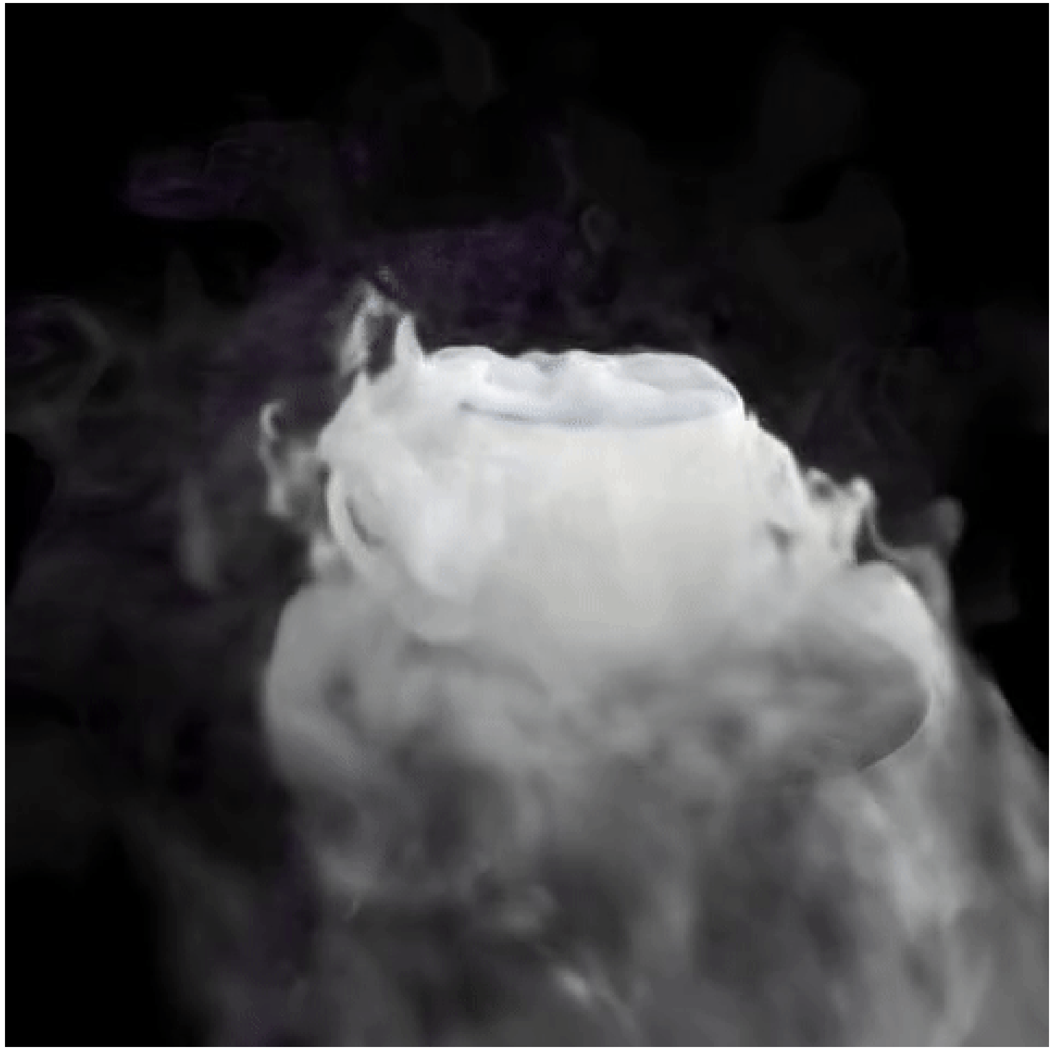
Neural Volumes



Neural Volumes: Learning Dynamic Renderable Volumes from Images, Lombardi et al. 2019

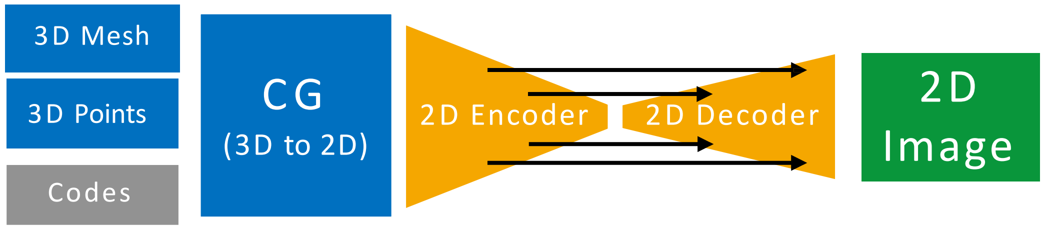
Neural Volumes

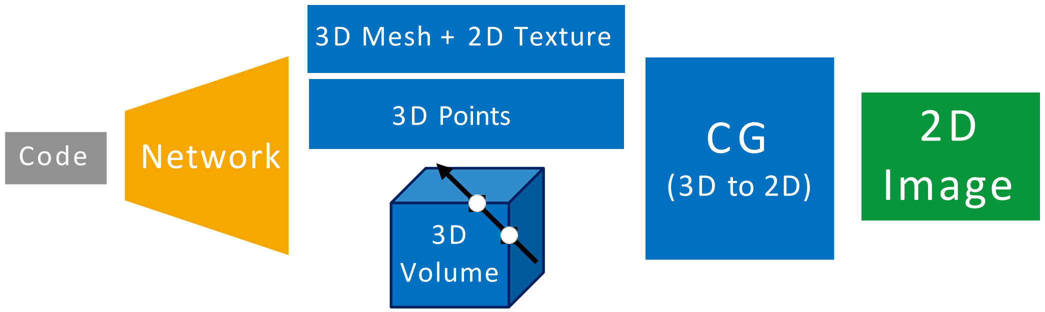
Neural Volumes



Neural Rendering Zoo

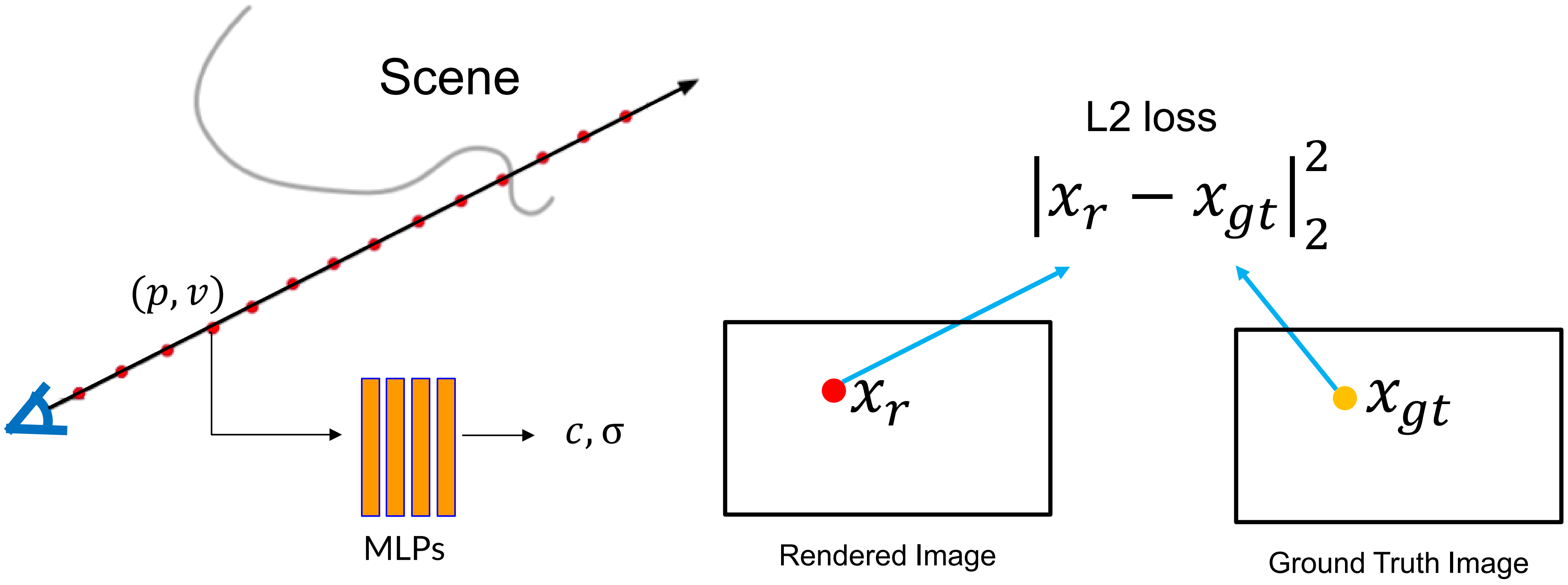
“Regress it”	e.g., GQN
”Make it more real”	e.g., DVP or DNR
“Regress & render”	e.g., Neural Volumes
”Step, sample & blend”	e.g., NeRF





Lingjie Liu

Neural Radiance Fields (NeRF)

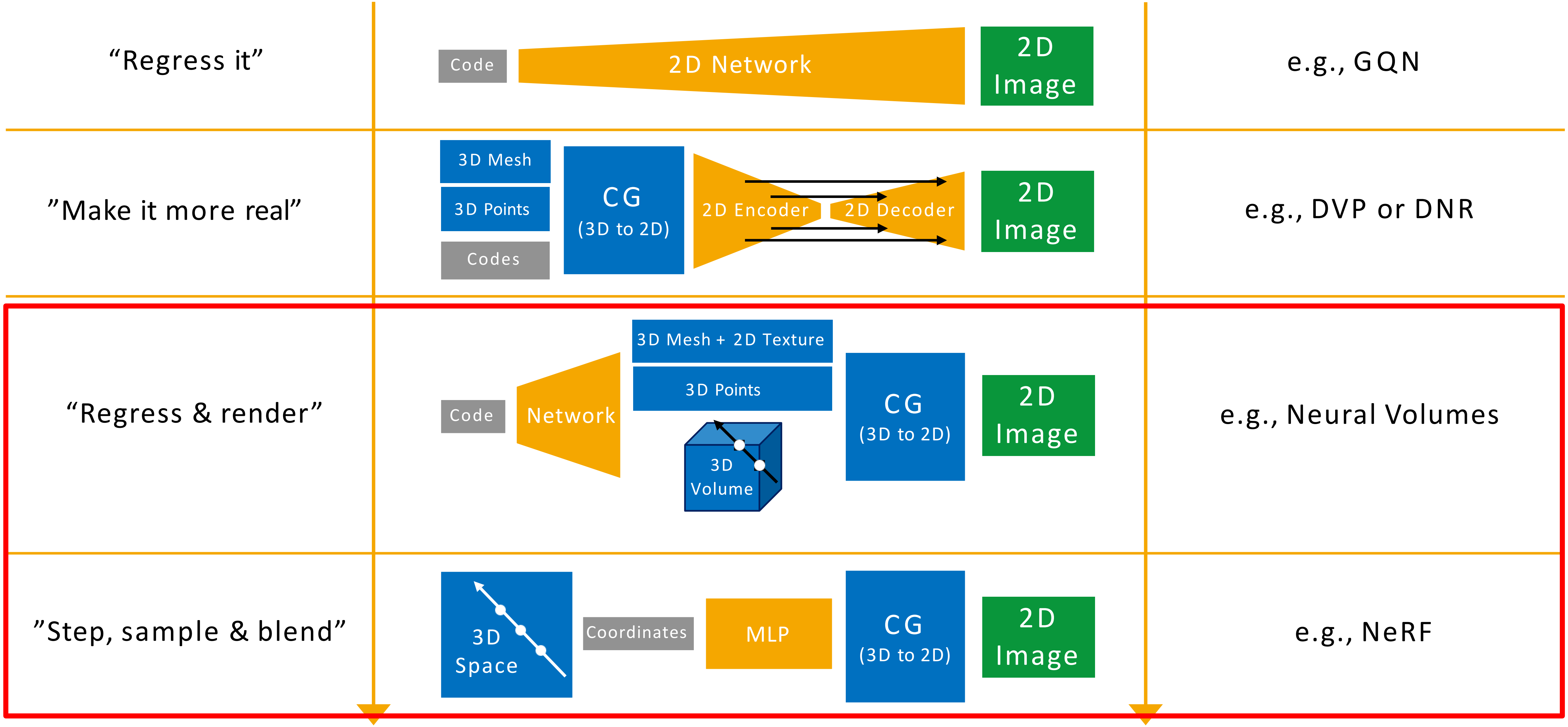


[Mildenhall et al. 2020]

Neural Radiance Fields (NeRF)

[Mildenhall et al. 2020]

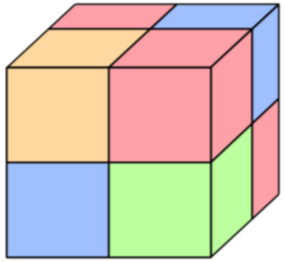
Neural Rendering Zoo



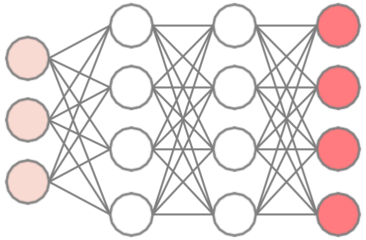
Lingjie Liu

Overview

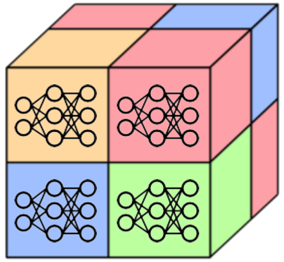
Scene Representation



Voxelgrids



Implicit Function



Hybrid Implicit/Explicit

• • •

(We'll talk more about other representations in our next class)

Renderer

Volumetric

Sphere-Tracing
Volumetric

Volumetric

Both Scene Representation and Differentiable Renderer often adapted from traditional computer graphics.

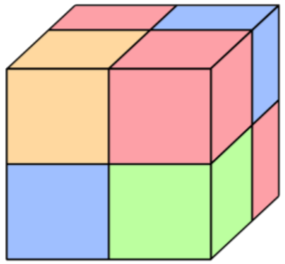
Requirements

Scene Representation

Renderer

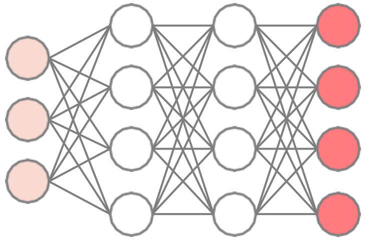
Pros

Cons



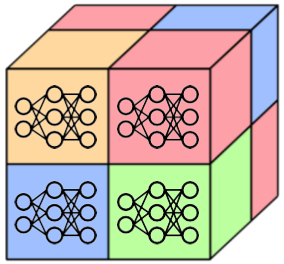
Voxelgrids

Volumetric



Implicit Function

Sphere-Tracing
Volumetric



Hybrid
Implicit/Explicit

Volumetric

• • •

Lingjie Liu

Voxel-based methods

DeepVoxels

Sitzmann et al., CVPR 2018

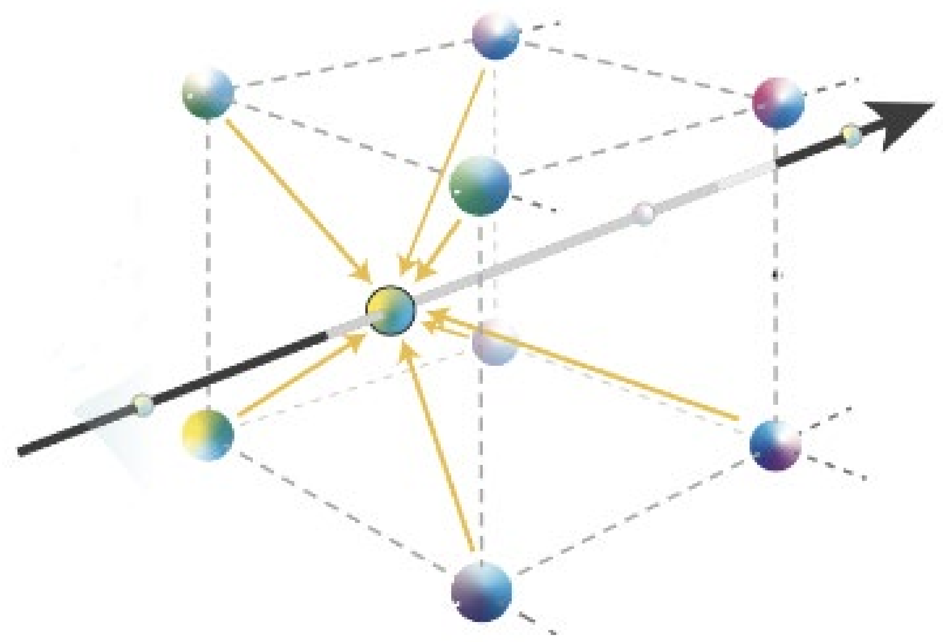
Neural Volumes

Lombardi et al., SIGGRAPH 2019

HoloGAN

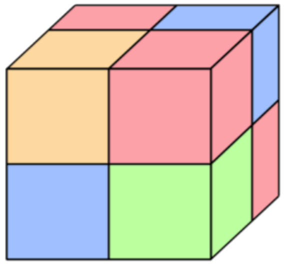
Phuoc et al., ICCV 2019

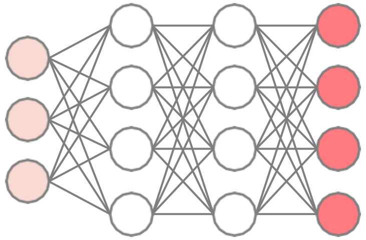
Voxel-based methods

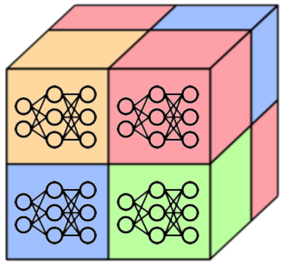


Trilinear Interpolation

Requirements



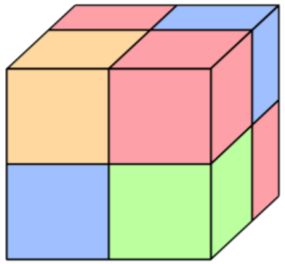
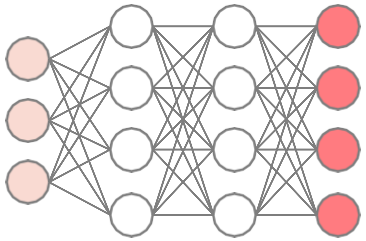
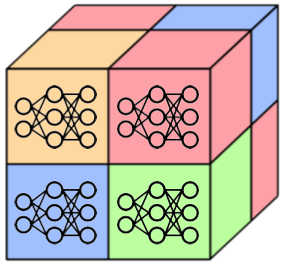




Scene Representation	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	• • •
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric	
Pros	Fast rendering			
Cons	Memory $O(n^3)$ Limited spatial resolution			

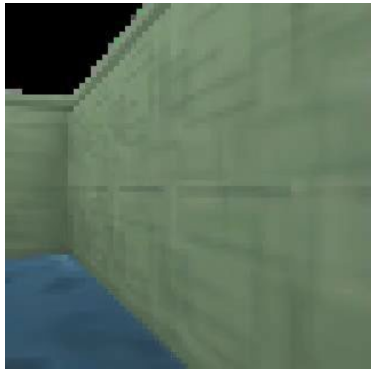
Lingjie Liu

Requirements

	 Voxelgrids	 Implicit Function	 Hybrid Implicit/Explicit	• • •
Scene Representation	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric	
Pros	Fast rendering			
Cons	Memory $O(n^3)$ Limited spatial resolution			

Lingjie Liu

Neural Implicit Approaches

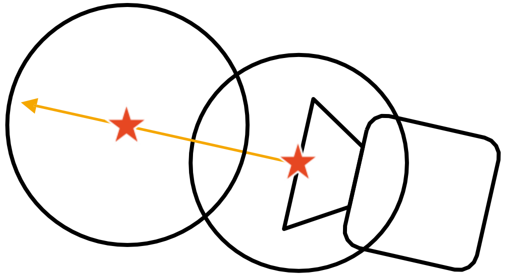


Scene Representation Networks
Generalizes across scenes
Sitzmann et al., NeurIPS 2019

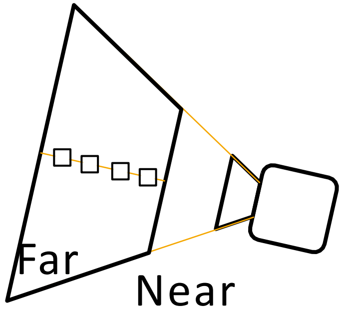
Differentiable Volumetric Rendering
Generalizes across scenes
Niemeyer et al., CVPR 2020

NeRF
Single-scene
Mildenhall et al., ECCV 2020

Implicit Differentiable Renderer
Single-scene
Yariv et al., NeurIPS 2020



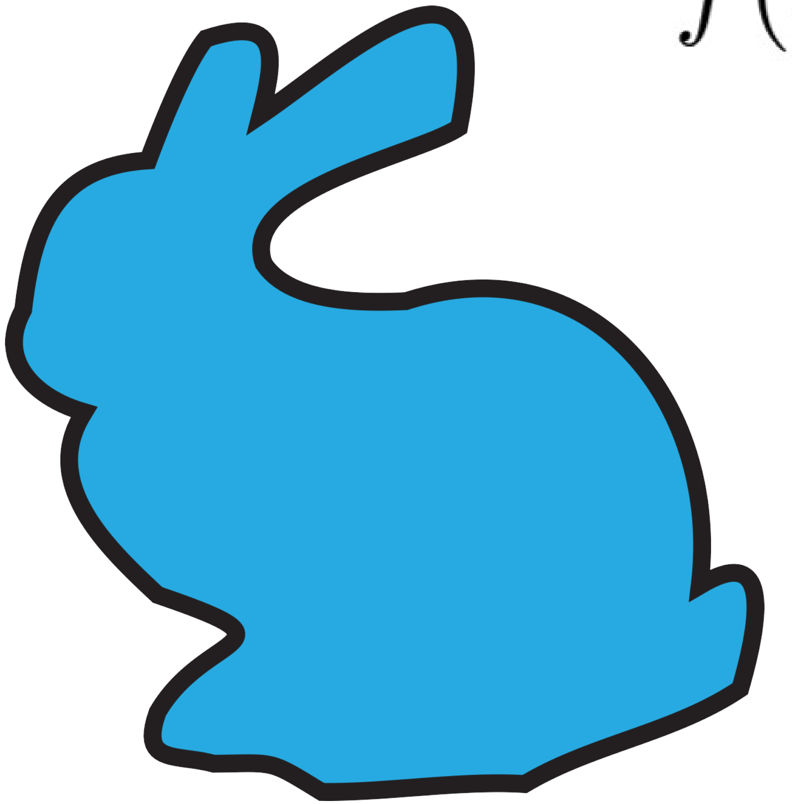
Sphere tracing



Volumetric

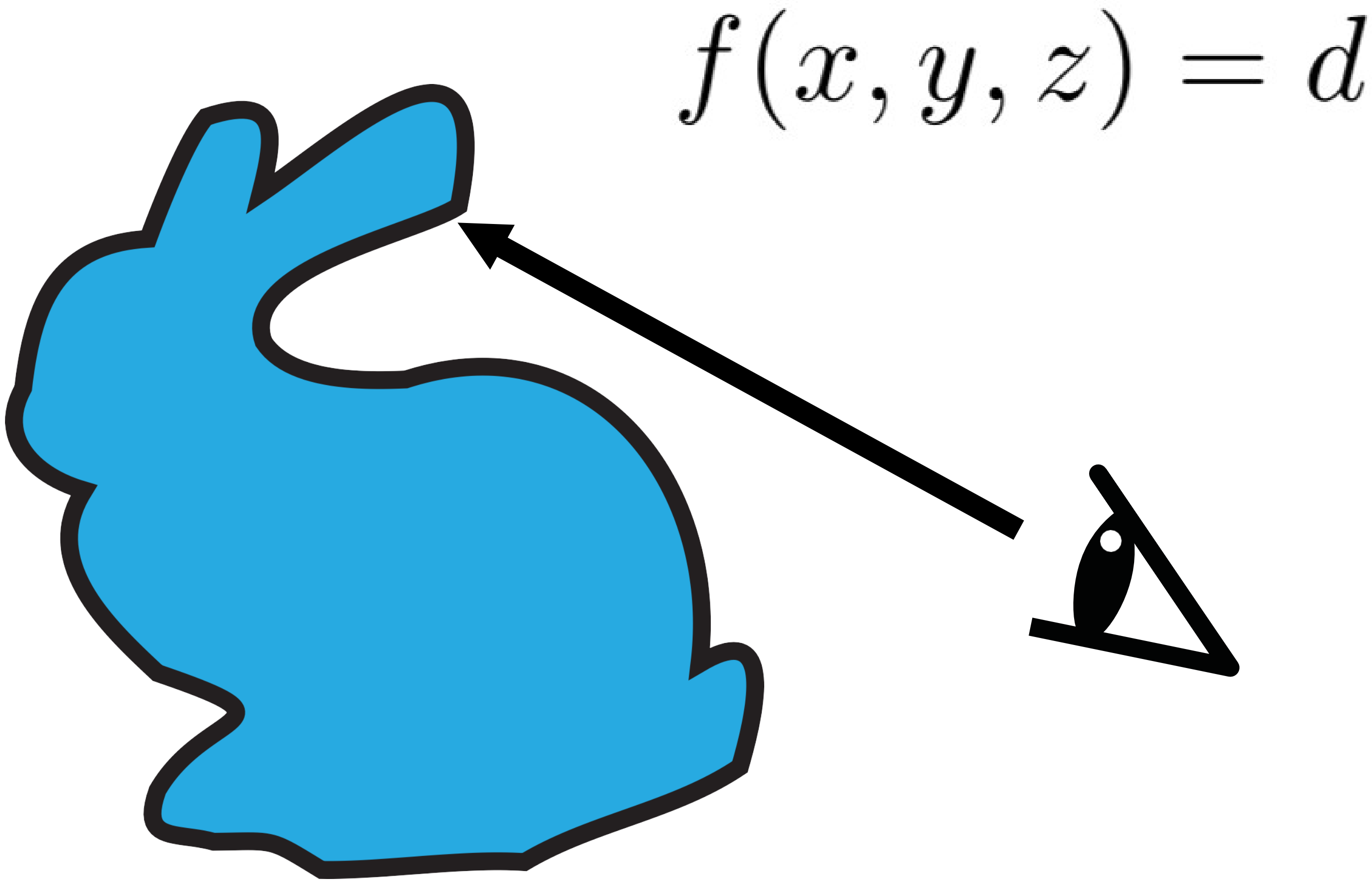
Sphere Tracing

$$f(x, y, z) = d$$



[Source: Takikawa et al]

Sphere Tracing

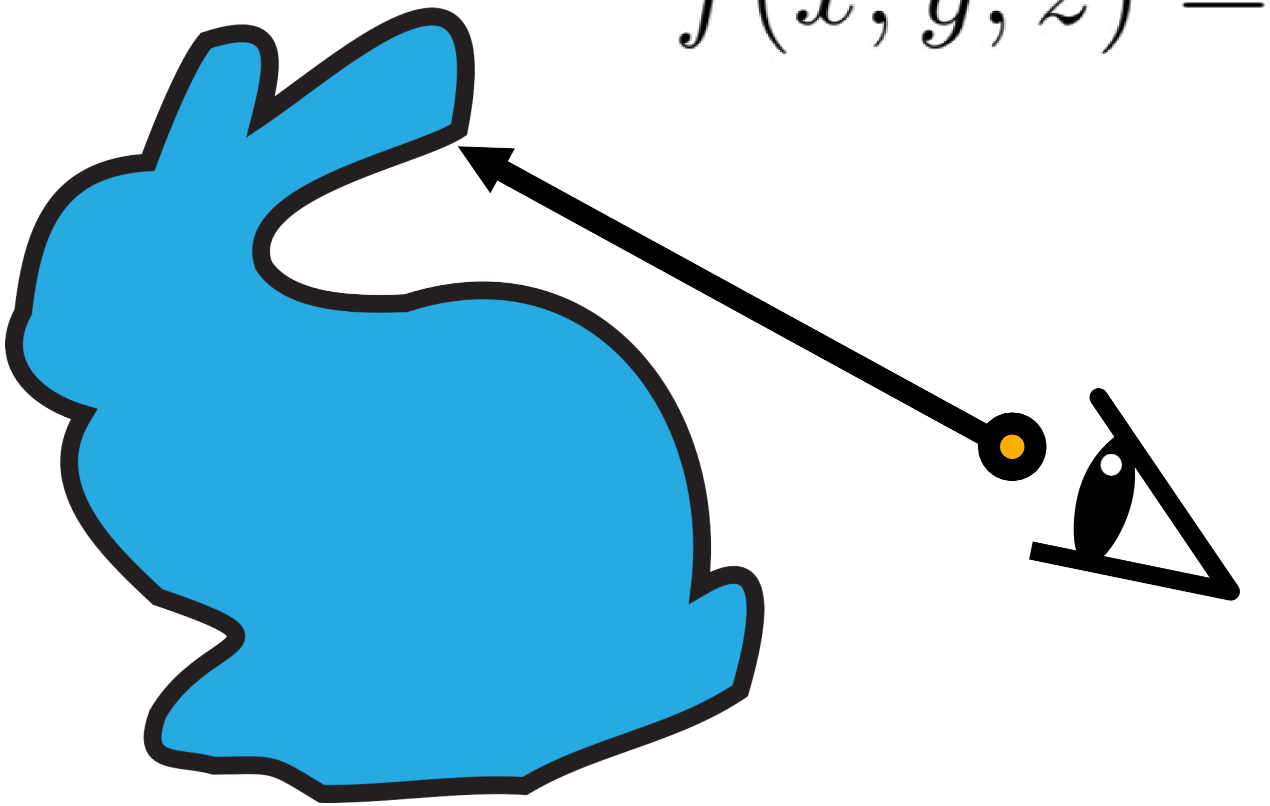


[Source: Takikawa et al]

Sphere Tracing

Penn
University of Pennsylvania

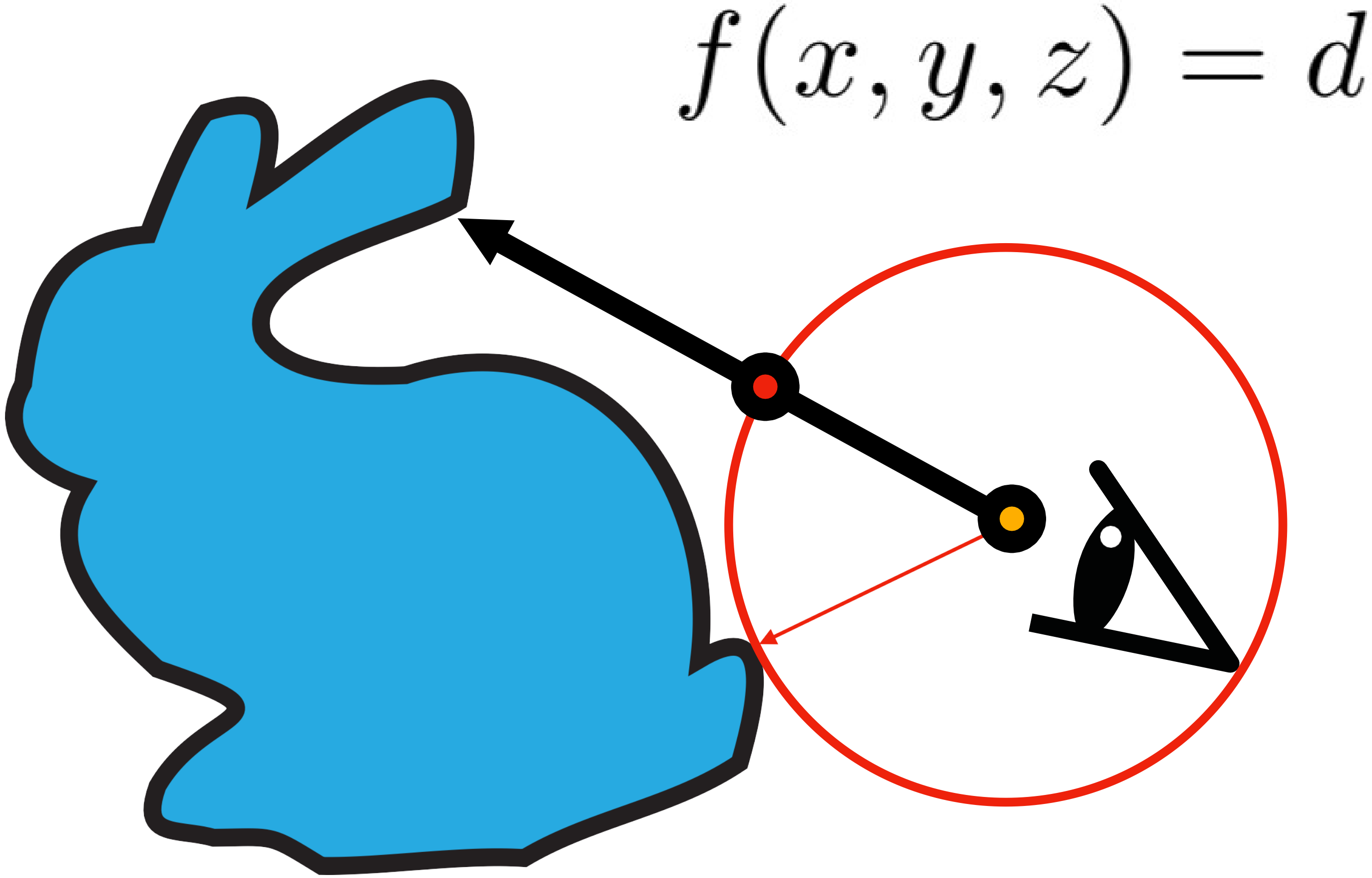
$$f(x, y, z) = d$$



[Source: Takikawa et al]

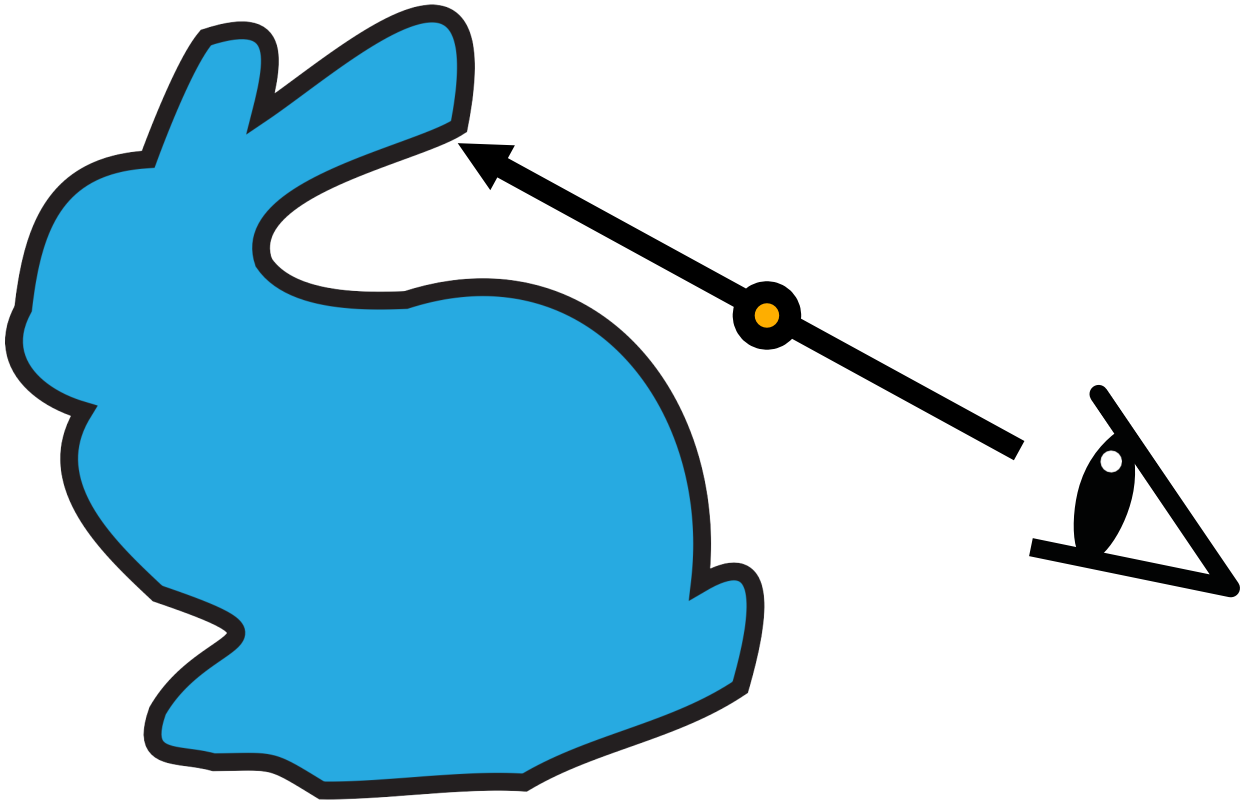
Lingjie Liu

Sphere Tracing



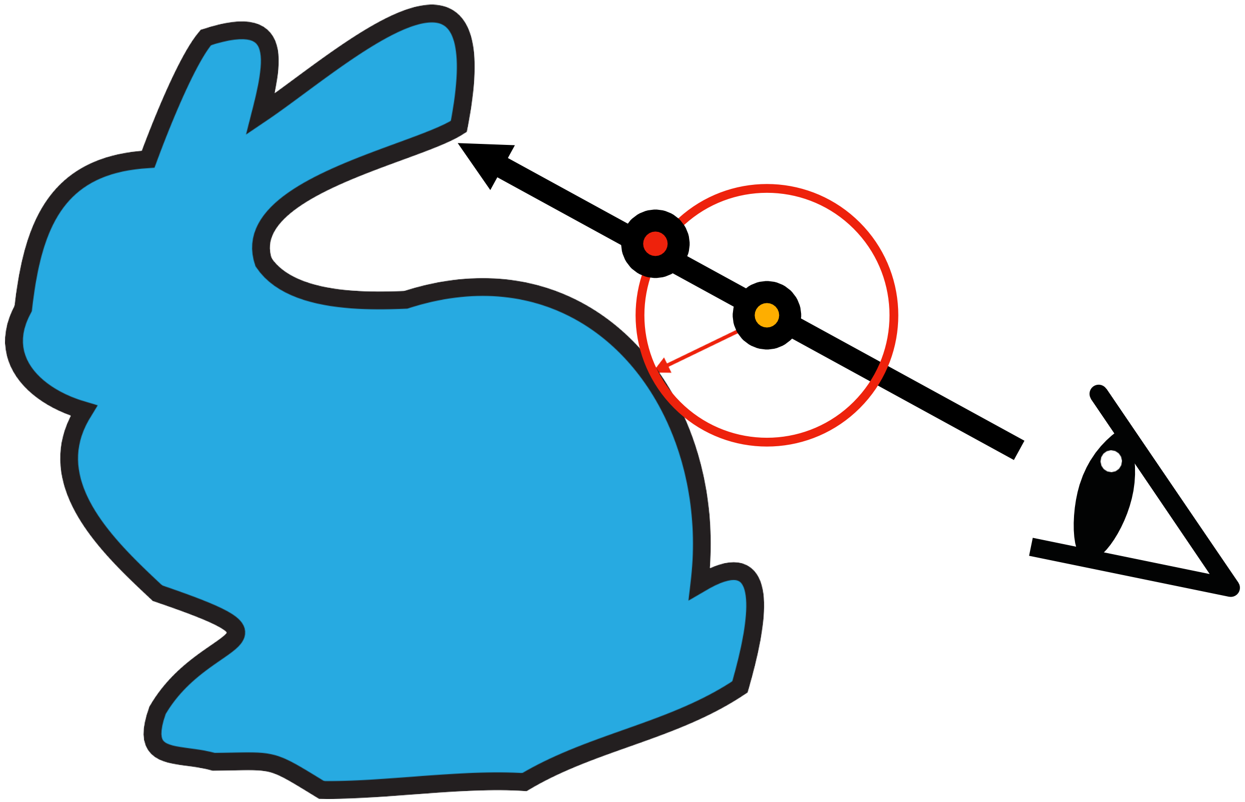
[Source: Takikawa et al]

Sphere Tracing



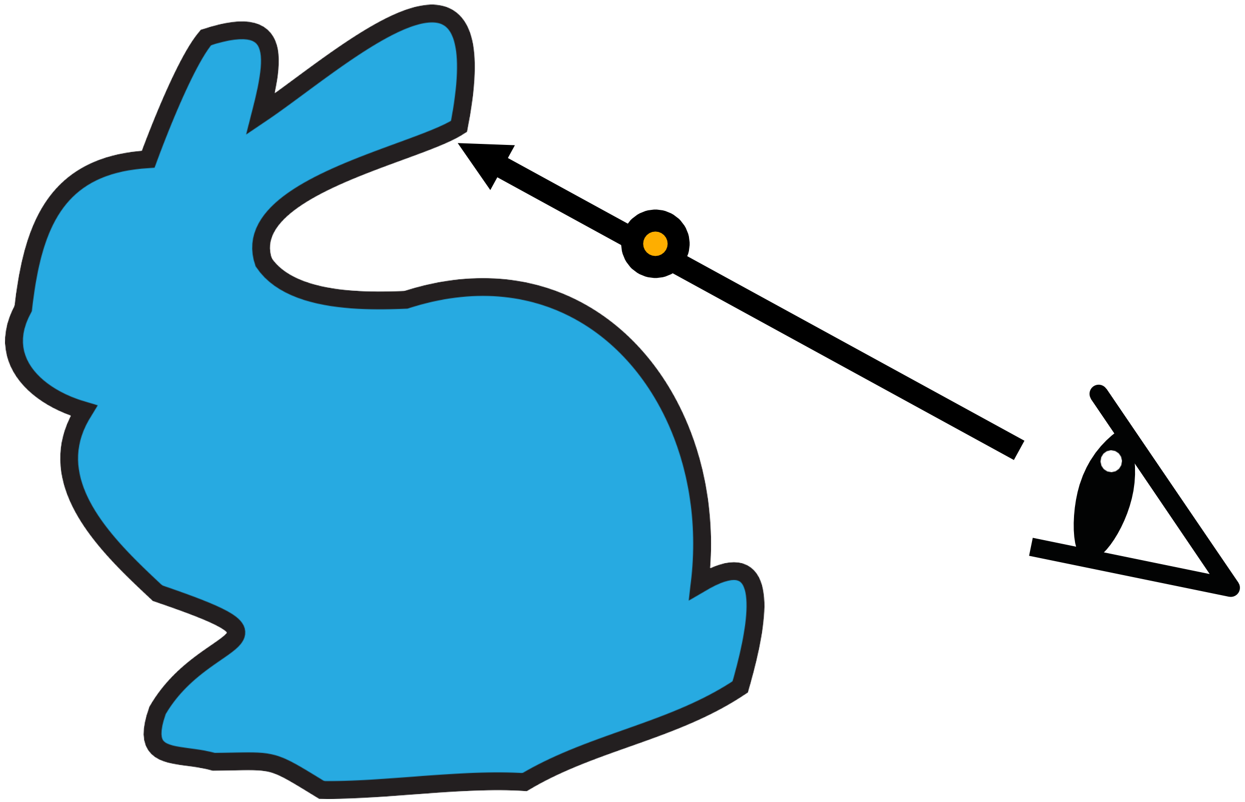
[Source: Takikawa et al]

Sphere Tracing



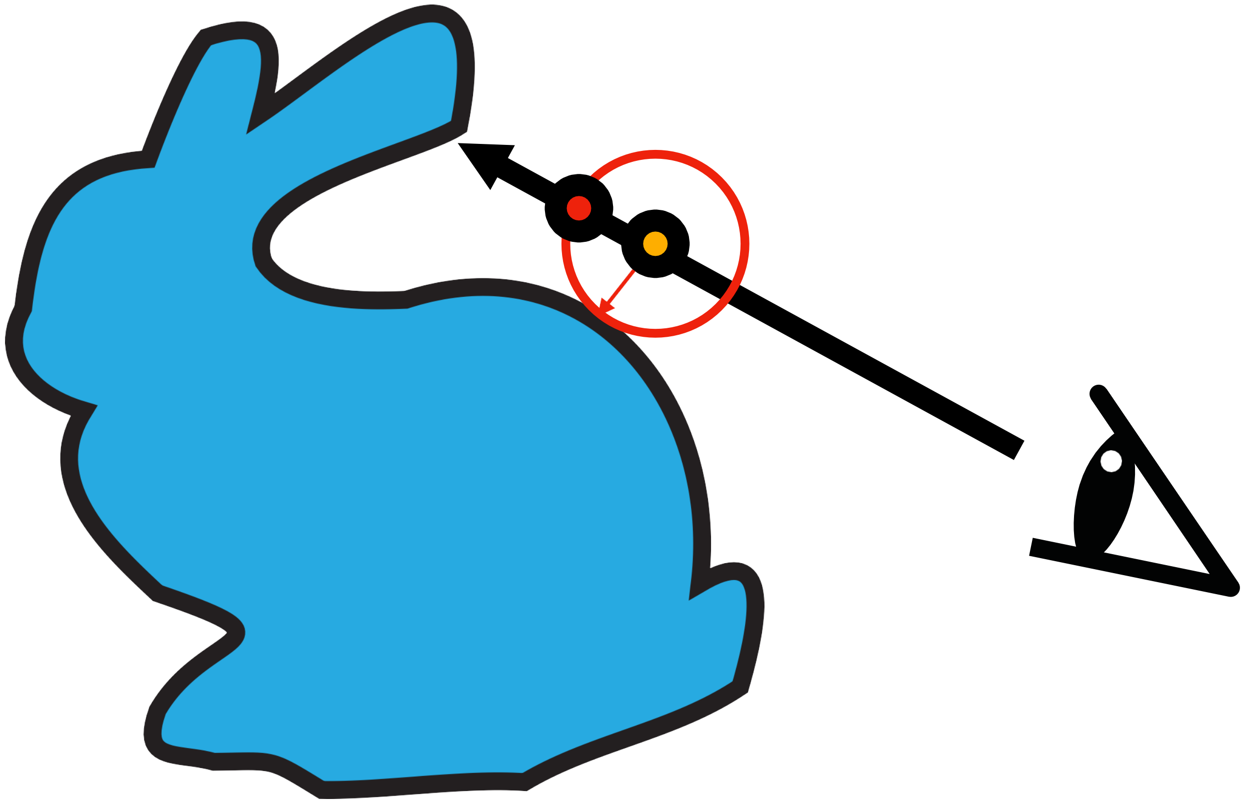
[Source: Takikawa et al]

Sphere Tracing



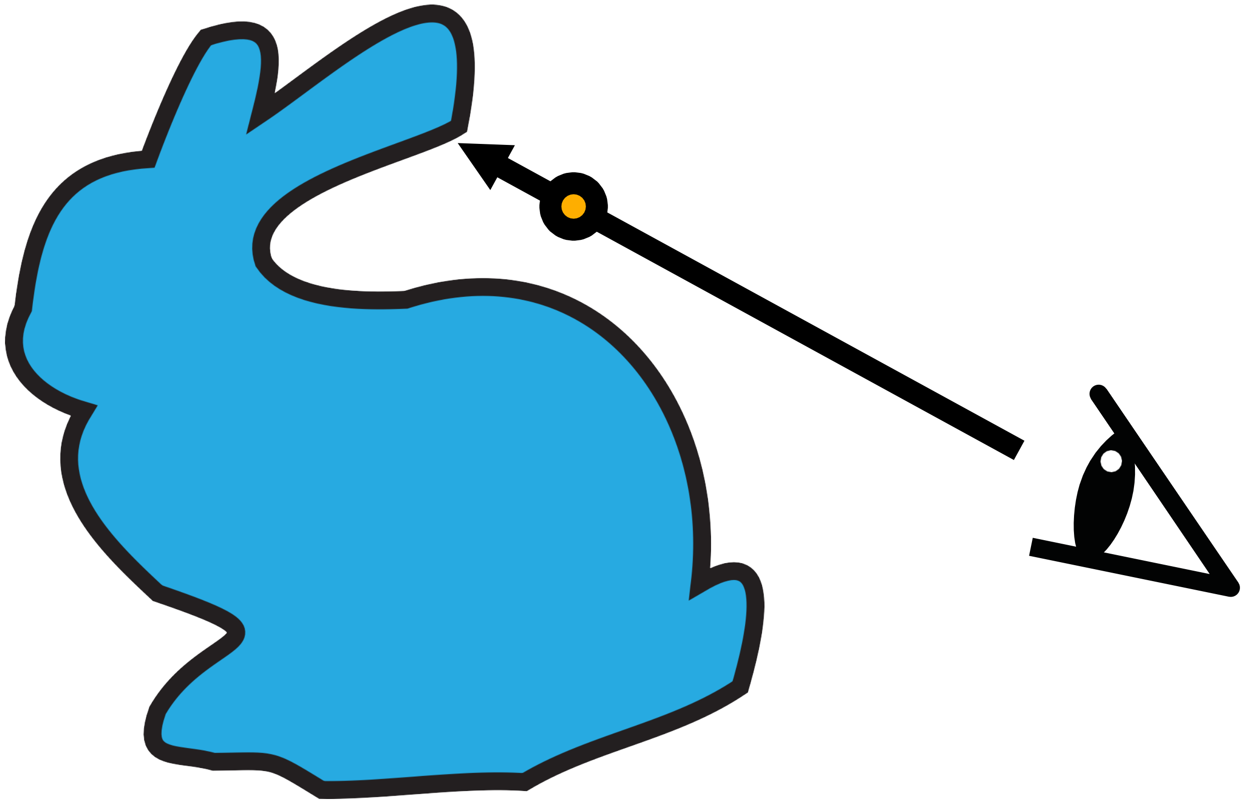
[Source: Takikawa et al]

Sphere Tracing



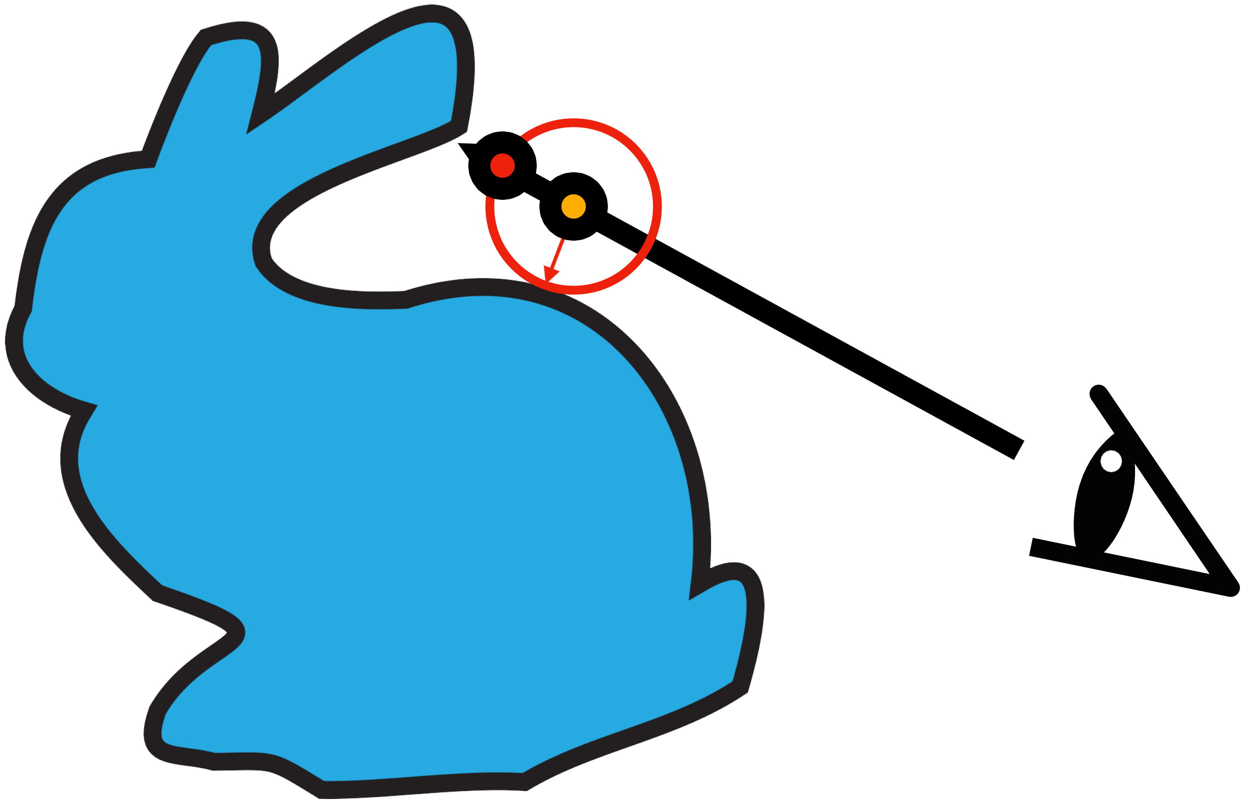
[Source: Takikawa et al]

Sphere Tracing



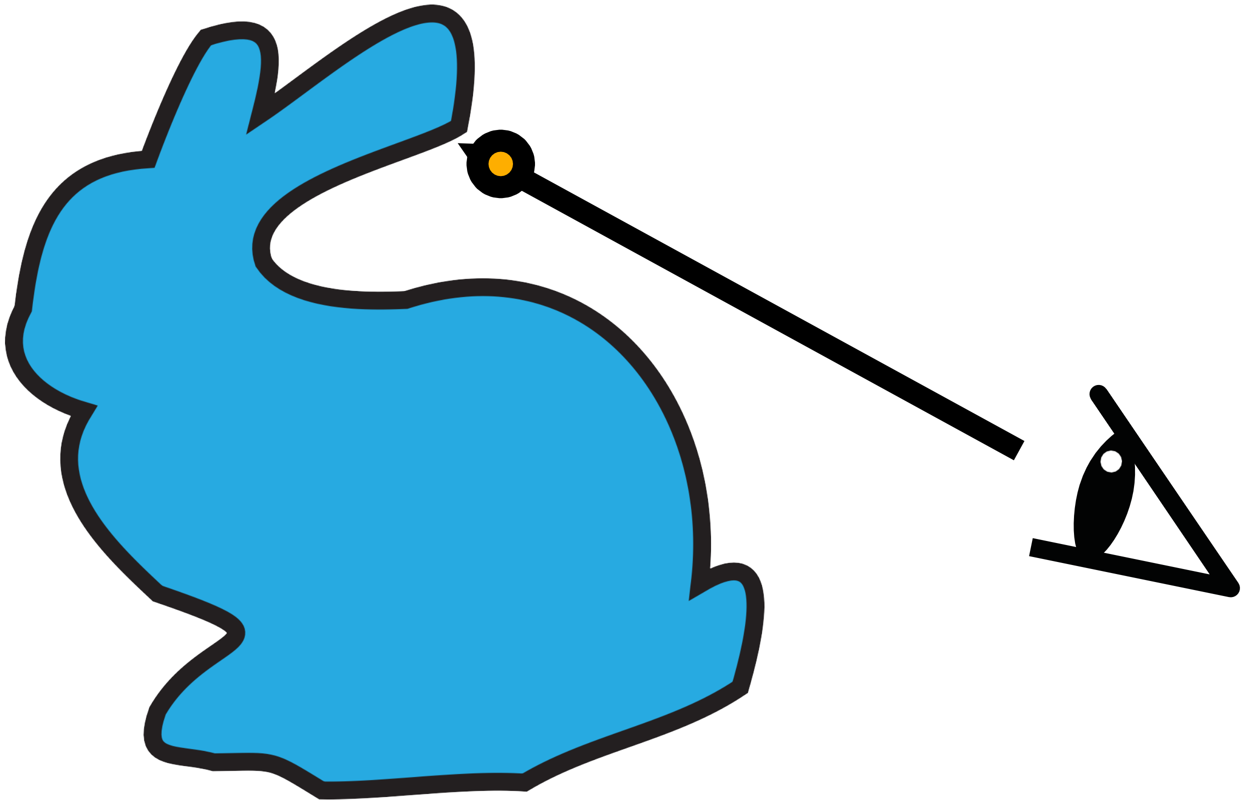
[Source: Takikawa et al]

Sphere Tracing



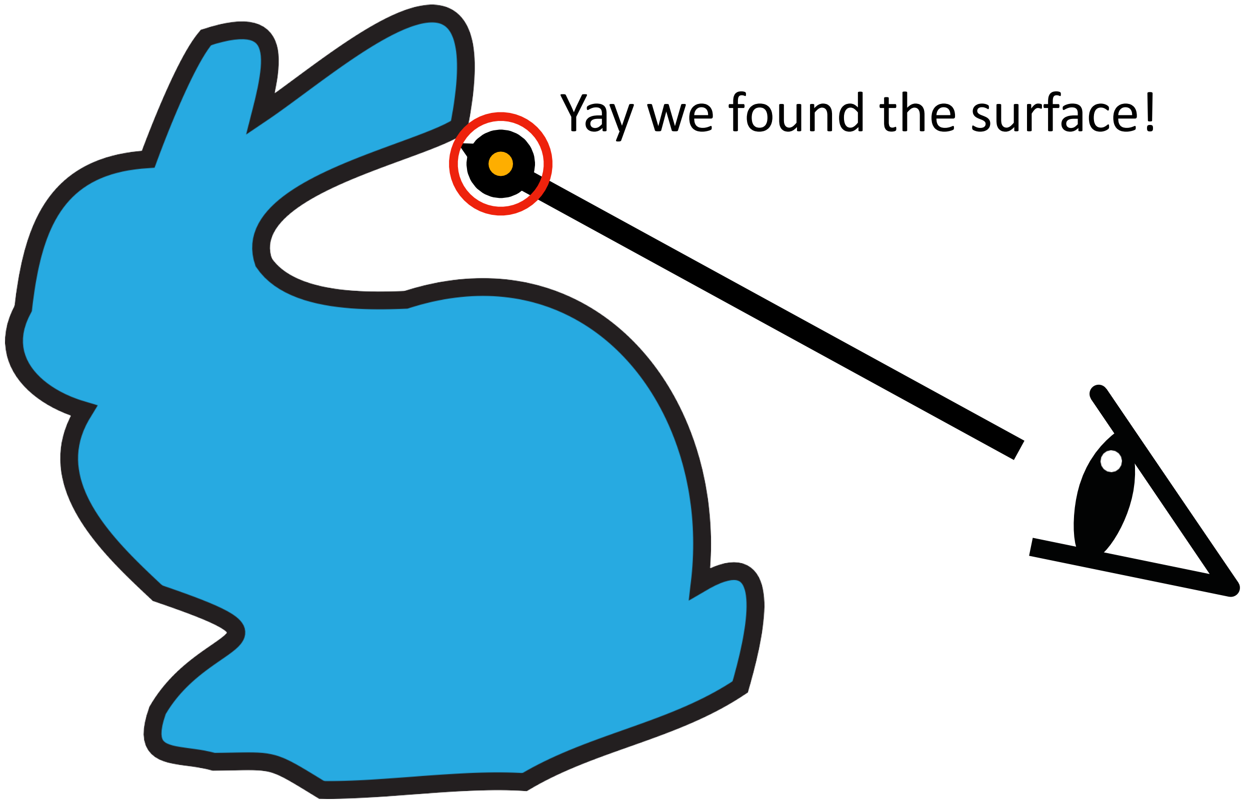
[Source: Takikawa et al]

Sphere Tracing



[Source: Takikawa et al]

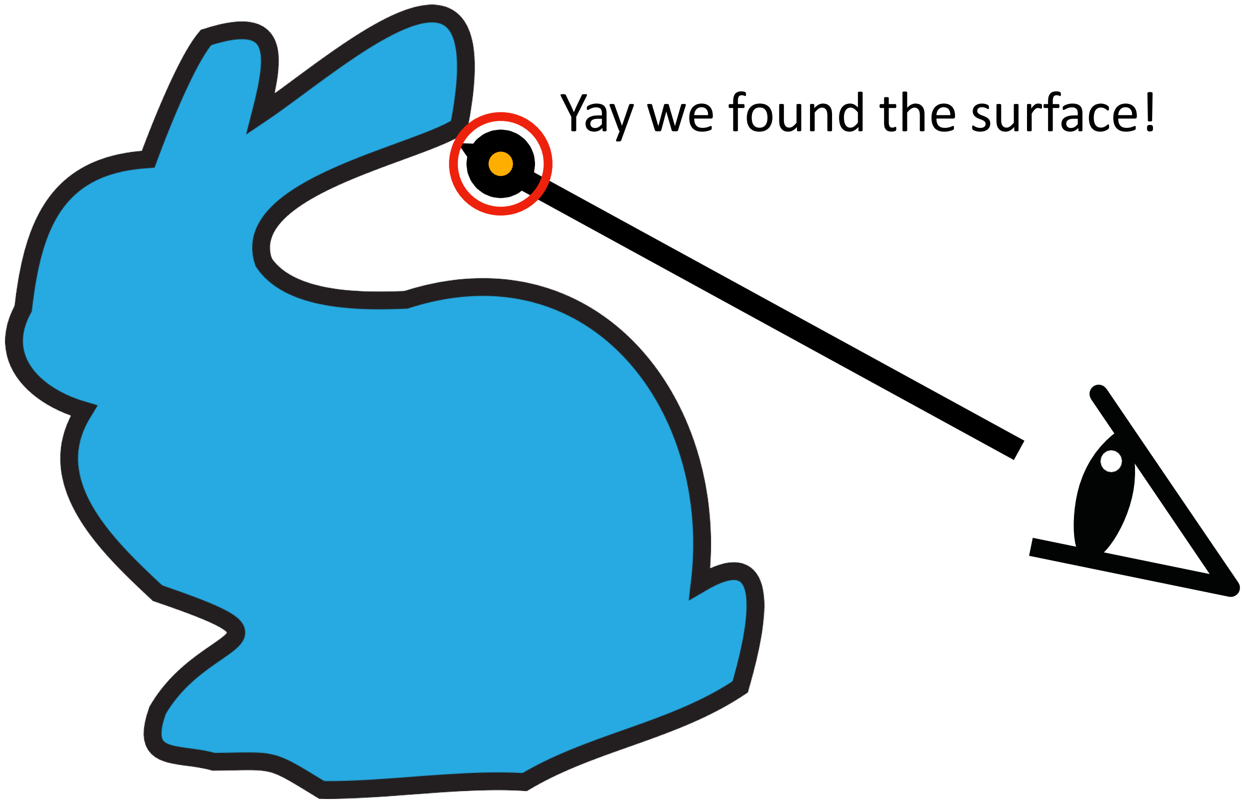
Sphere Tracing



[Source: Takikawa et al]

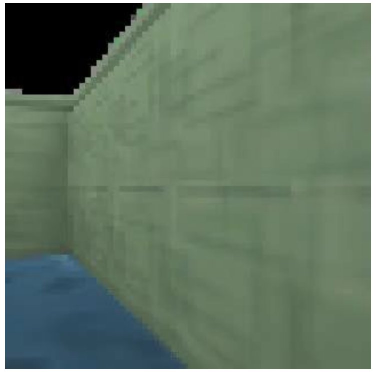
Sphere Tracing

$$f(x, y, z) = d$$



[Source: Takikawa et al]

Neural Implicit Approaches

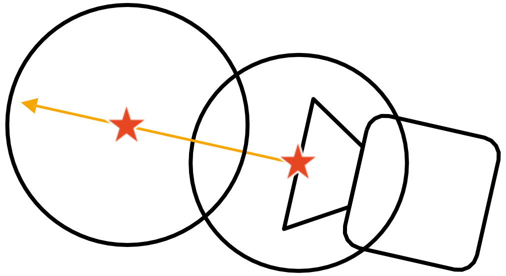


Scene Representation Networks
Generalizes across scenes
Sitzmann et al., NeurIPS 2019

Differentiable Volumetric Rendering
Generalizes across scenes
Niemeyer et al., CVPR 2020

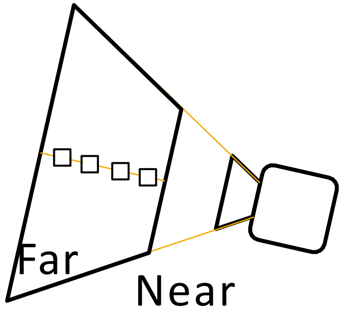
NeRF
Single-scene
Mildenhall et al., ECCV 2020

Implicit Differentiable Renderer
Single-scene
Yariv et al., NeurIPS 2020



Sphere tracing

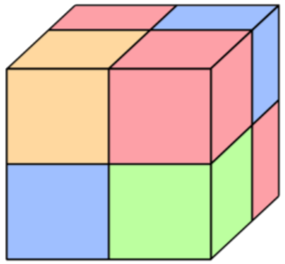
- Faster
- Fewer network evaluations
- Convergence more difficult

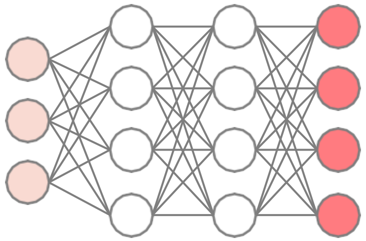


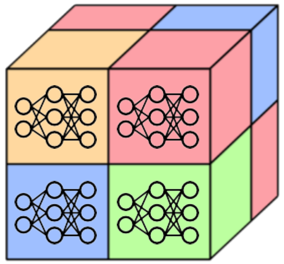
Volumetric

- Higher Quality
- Easy convergence
- Very expensive

Requirements



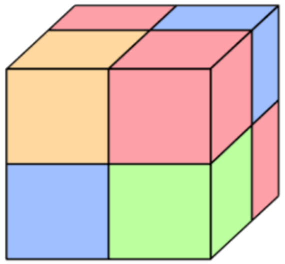
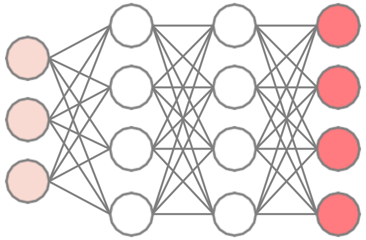
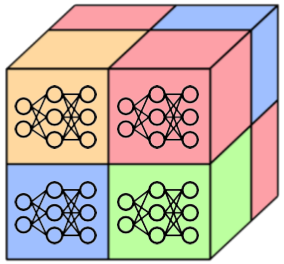




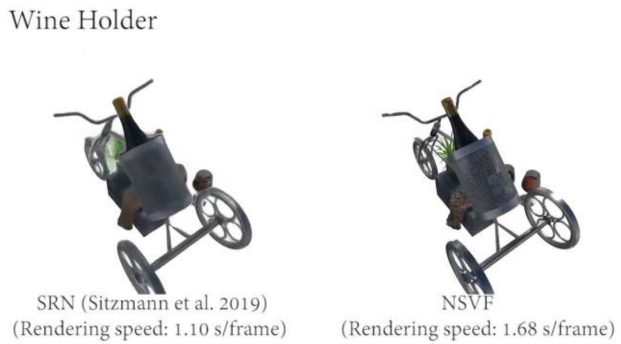
	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	• • •
Scene Representation	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric	
Pros	Fast rendering	High quality Compact Admits *global* priors		
Cons	Memory $O(n^3)$ Limited spatial resolution	Extremely expensive, slow rendering		

Lingjie Liu

Requirements

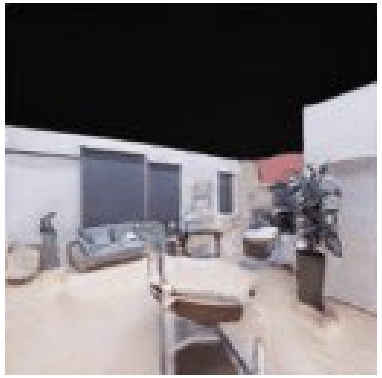
				...
Scene Representation	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric	
Pros	Fast rendering	High quality Compact Admits *global* priors		
Cons	Memory $O(n^3)$ Limited spatial resolution	Extremely expensive, slow rendering		

Hybrid Implicit / Explicit



Neural Sparse Voxel Fields,
Liu et. al., NeurIPS 2020

PiFU, Saito et al., ICCV 2019
GRF, Trevithick et al., arXiv 2020
pixelNeRF, Yu et. al., CVPR 2021
MVSNerf, Chen et al., arXiv 2021
Learn *local* (image patch-based) priors



Unconstrained Scene Generation with
Locally Conditioned Radiance Fields,
DeVries et al., arXiv 2021

Neural Sparse Voxel Fields (NSVF)

- Avoid sampling points in empty space as much as possible.

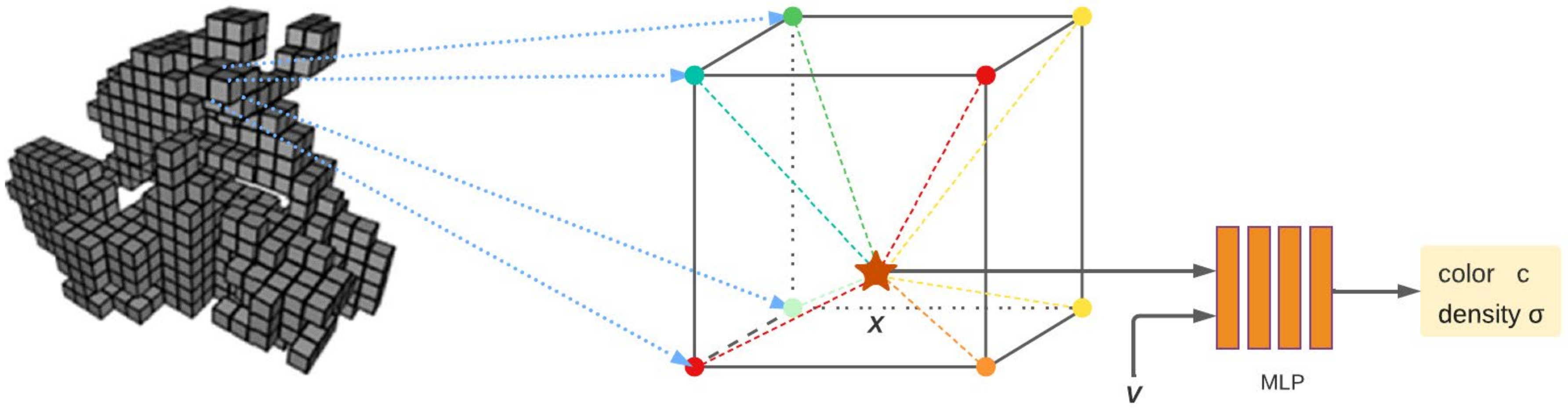


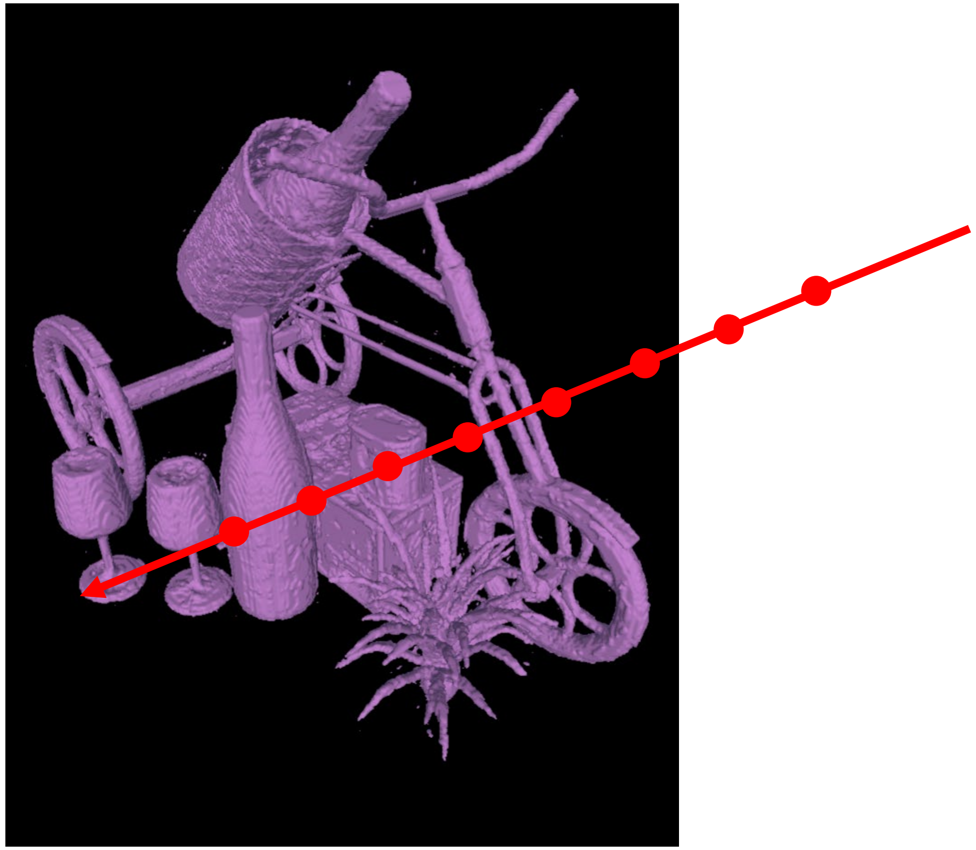
Illustration of Sparse Voxels

Illustration of a voxel-bounded neural field

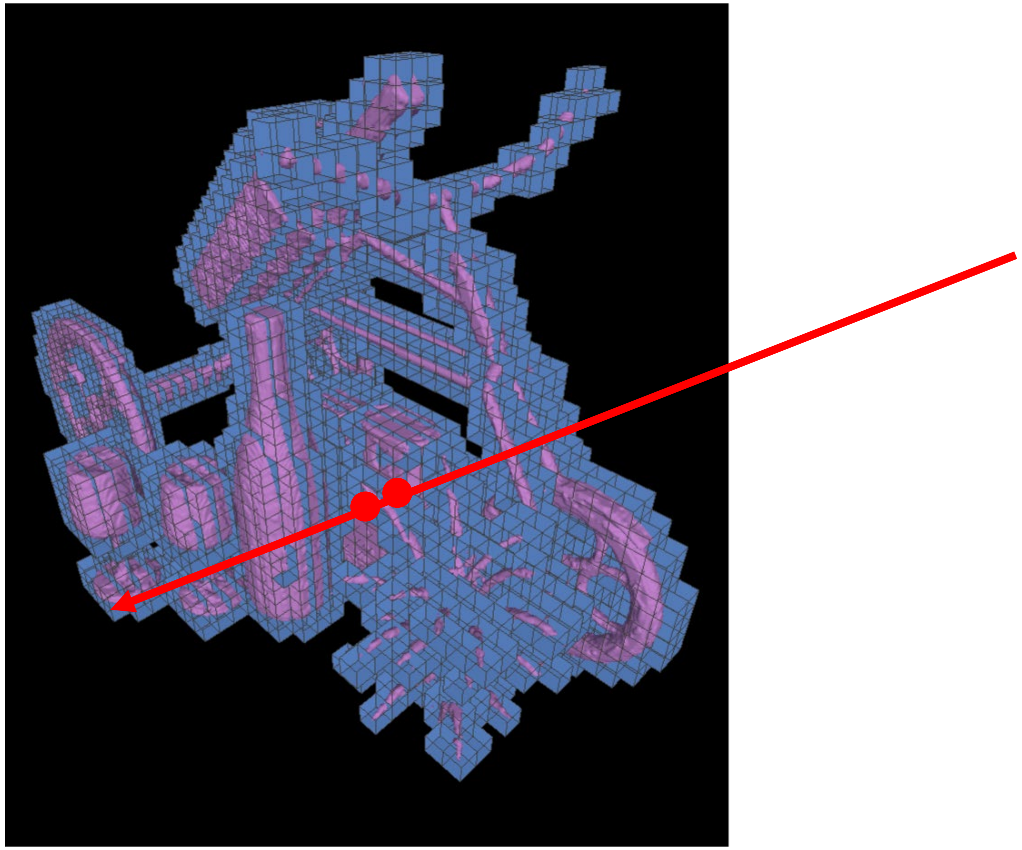
Neural Sparse Voxel Fields, Liu et al. 2020

Neural Sparse Voxel Fields (NSVF)

- Avoid sampling points in empty space as much as possible.



Sample in the whole space



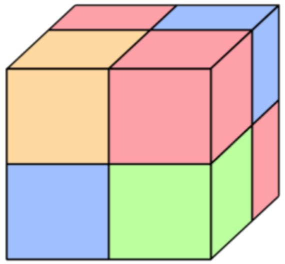
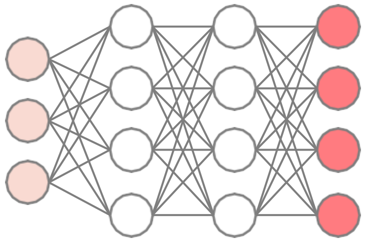
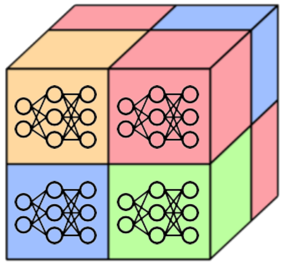
Only sample inside the sparse-voxels

Comparison

NeRF (Mildenhall et al. 2020)
(Rendering speed: 100 s/frame)

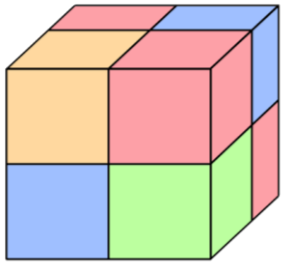
Ours (NSVF)
(Rendering speed: 2.62 s/frame)

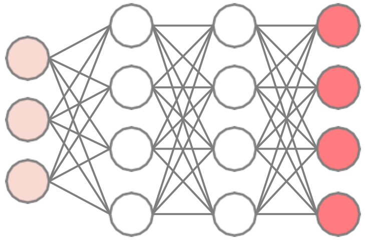
Requirements

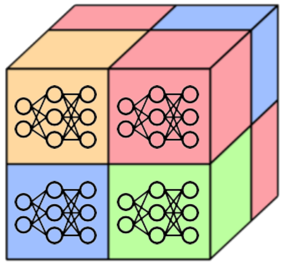
				• • •
Scene Representation	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit	
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric	
Pros	Fast rendering	High quality Compact Admits *global* priors	Significant Speedup Admits *local* priors	
Cons	Memory $O(n^3)$ Limited spatial resolution	Extremely expensive, slow rendering	No compact representation No *global* priors	

Neural Scene Representation and Neural Rendering

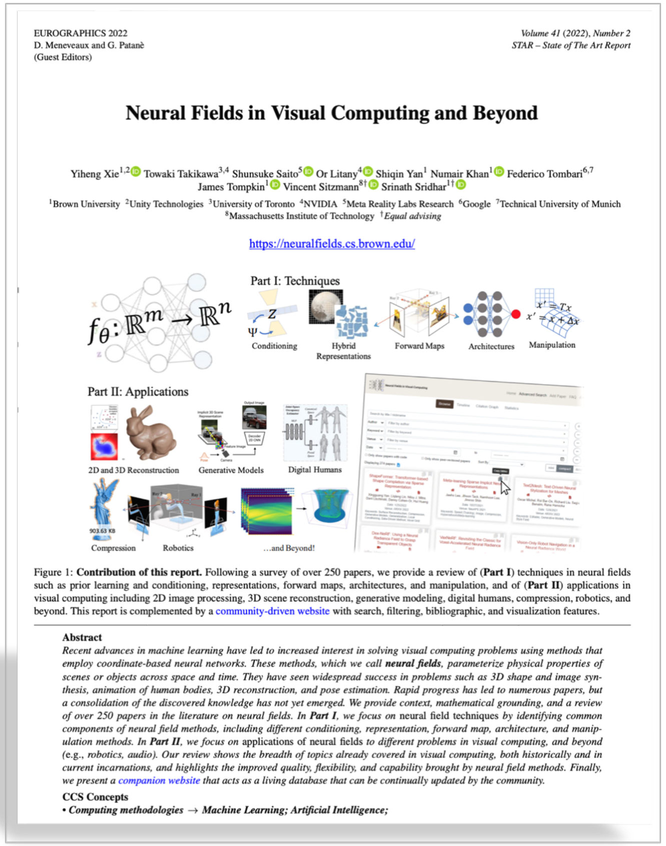
Neural Fields







	Voxelgrids	Implicit Function	Hybrid Implicit/Explicit
Scene Representation			
Renderer	Volumetric	Sphere-Tracing Volumetric	Volumetric
Pros	Fast rendering	High quality Compact Admits *global* priors	Significant Speedup Admits *local* priors
Cons	Memory $O(n^3)$ Limited spatial resolution	Extremely expensive, slow rendering	No compact representation No *global* priors



EUROGRAPHICS 2022
D. Meneveaux and G. Patanè
(Guest Editors)

Volume 41 (2022), Number 2
STAR – State of The Art Report

Neural Fields in Visual Computing and Beyond

Yiheng Xie[1,2] Towaki Takikawa[3,4] Shunsuke Saito[5] Or Litany[4] Shiqin Yan[1] Numair Khan[1] Federico Tombari[6,7]
James Tompkin[1] Vincent Sitzmann[8†] Srinath Sridhar[1†]

[1]Brown University [2]Unity Technologies [3]University of Toronto [4]NVIDIA [5]Meta Reality Labs Research [6]Google [7]Technical University of Munich
[8]Massachusetts Institute of Technology †*Equal advising*

https://neuralfields.cs.brown.edu/

Figure 1: **Contribution of this report.** Following a survey of over 250 papers, we provide a review of (**Part I**) techniques in neural fields such as prior learning and conditioning, representations, forward maps, architectures, and manipulation, and of (**Part II**) applications in visual computing including 2D image processing, 3D scene reconstruction, generative modeling, digital humans, compression, robotics, and beyond. This report is complemented by a community-driven website with search, filtering, bibliographic, and visualization features.

Abstract
Recent advances in machine learning have led to increased interest in solving visual computing problems using methods that employ coordinate-based neural networks. These methods, which we call ***neural fields****, parameterize physical properties of scenes or objects across space and time. They have seen widespread success in problems such as 3D shape and image synthesis, animation of human bodies, 3D reconstruction, and pose estimation. Rapid progress has led to numerous papers, but a consolidation of the discovered knowledge has not yet emerged. We provide context, mathematical grounding, and a review of over 250 papers in the literature on neural fields. In* ***Part I****, we focus on* neural field techniques *by identifying common components of neural field methods, including different conditioning, representation, forward map, architecture, and manipulation methods. In* ***Part II****, we focus on* applications of neural fields *to different problems in visual computing, and beyond (e.g., robotics, audio). Our review shows the breadth of topics already covered in visual computing, both historically and in current incarnations, and highlights the improved quality, flexibility, and capability brought by neural field methods. Finally, we present a companion website that acts as a living database that can be continually updated by the community.*

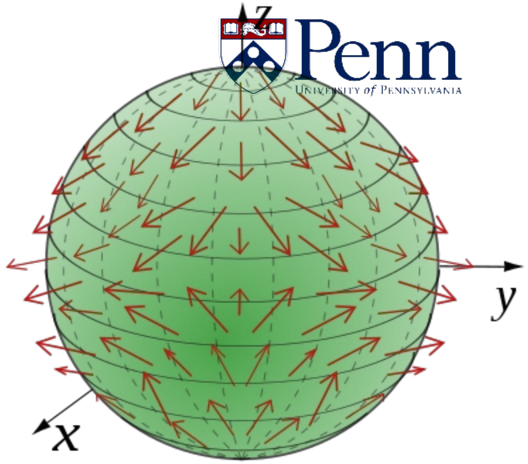
CCS Concepts
• ***Computing methodologies → Machine Learning; Artificial Intelligence;***

Lingjie Liu

Definition of Fields

A *field* is a quantity defined for all spatial and / or temporal coordinates.

Examples of Fields



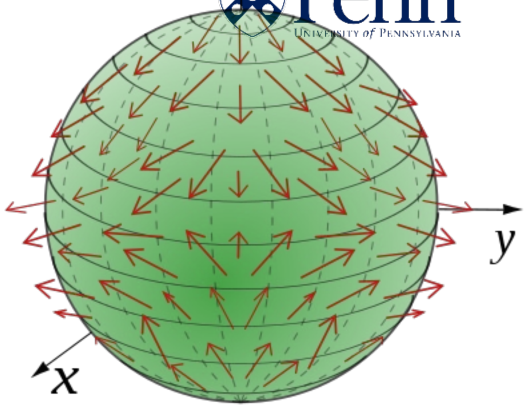
Vector Field

Fields

[Source: Wikipedia]

Examples of Fields

Image

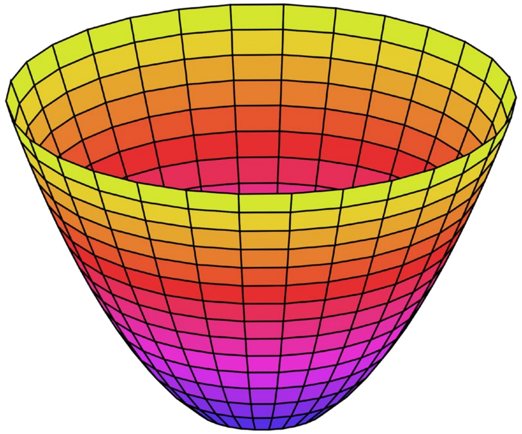


Vector Field

Fields

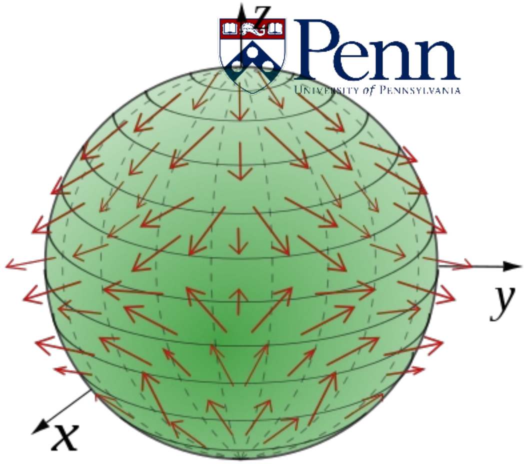
[Source: Wikipedia]

Examples of Fields



3D Parabola
(Explicit Surface)

Image

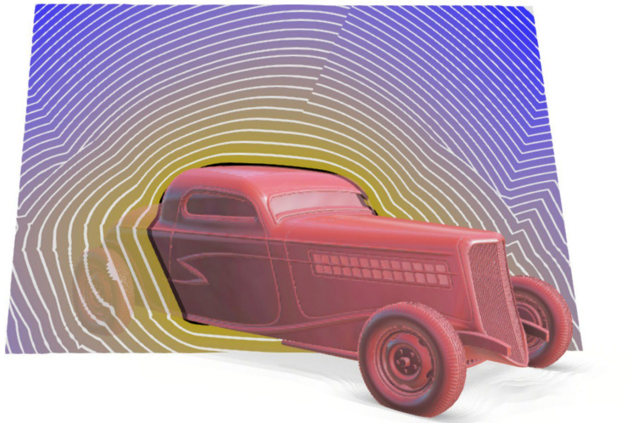


Vector Field

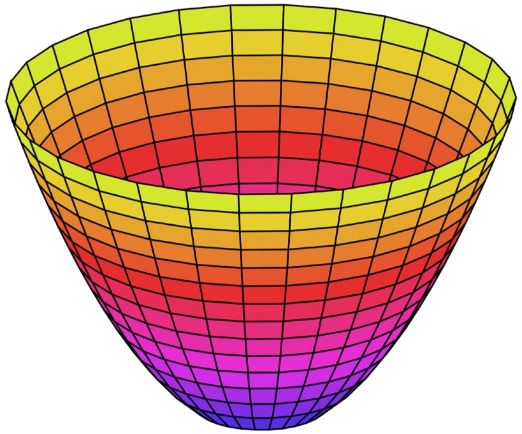
Fields

[Source: Wikipedia]

Examples of Fields

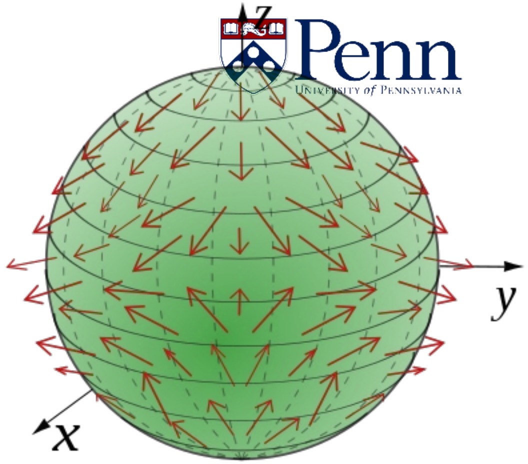


3D Signed Distance Fields
(Implicit Surface)



3D Parabola
(Explicit Surface)

Image

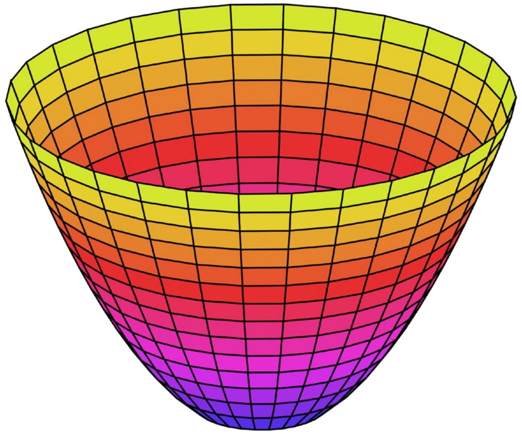


Vector Field

Fields

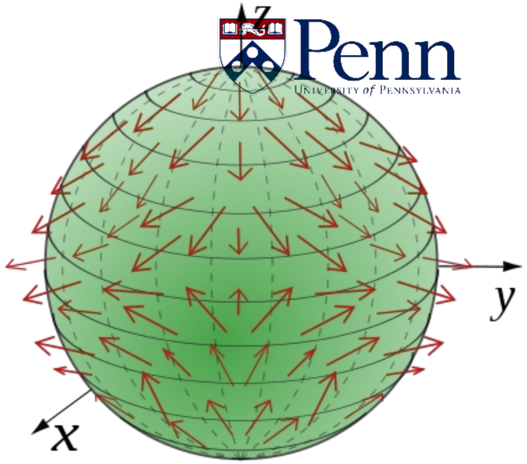
[Source: Wikipedia]

Examples of Fields

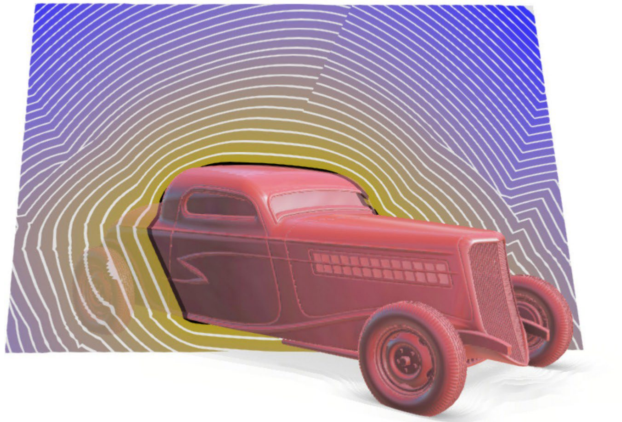


3D Parabola
(Explicit Surface)

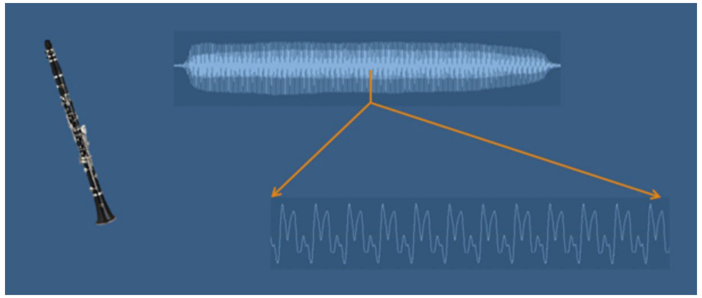
Image



Vector Field



3D Signed Distance Fields
(Implicit Surface)



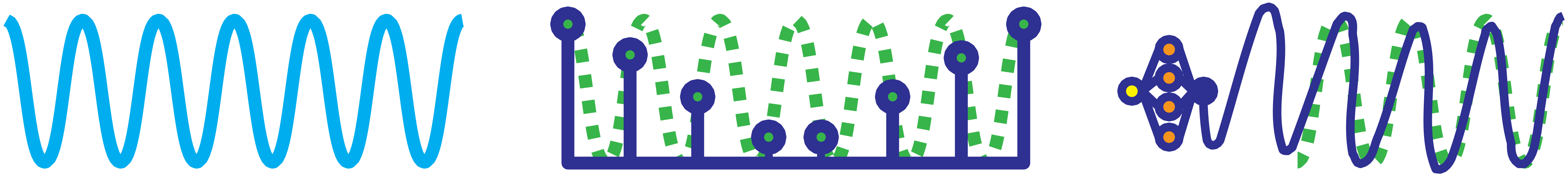
Audio

Fields

[Source: Wikipedia]

What are neural fields?

Fields / signals can be represented in many ways.

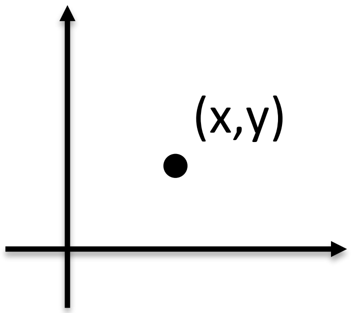


Continuous

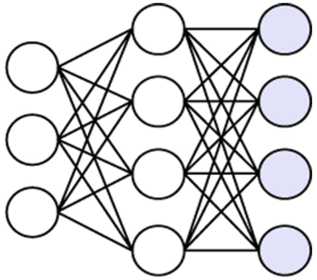
Discrete

Neural

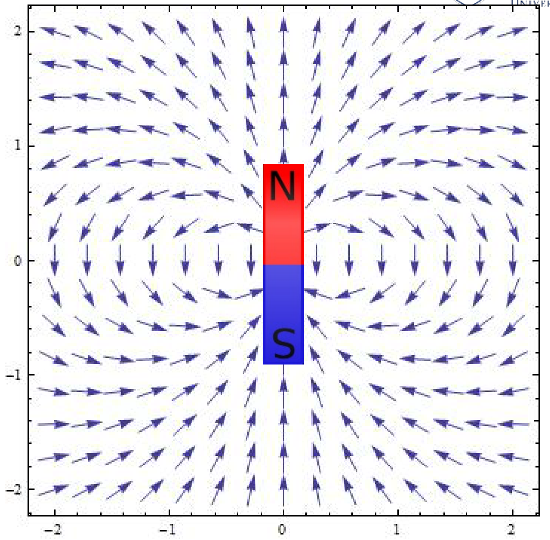
What are neural fields?



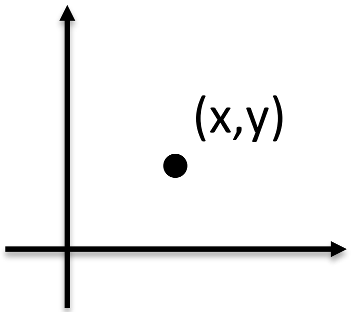
$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$



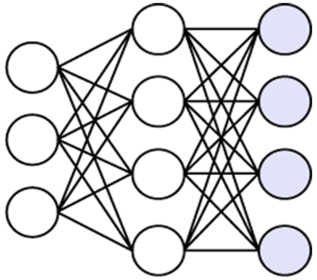
Neural Network (Φ)



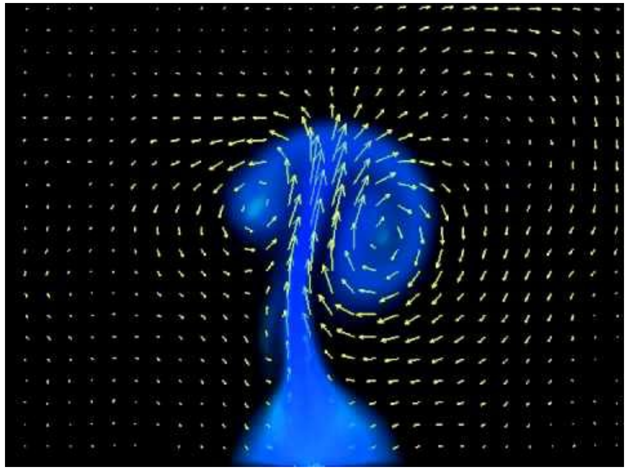
Magnetic Field



$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2$



Neural Network (Φ)

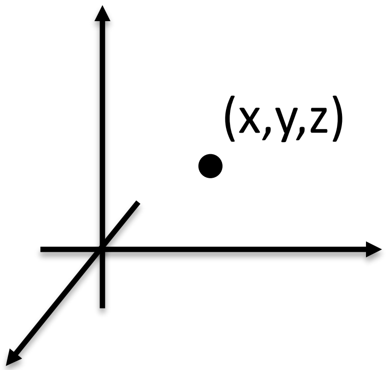


Eulerian Flow Field of a Fluid

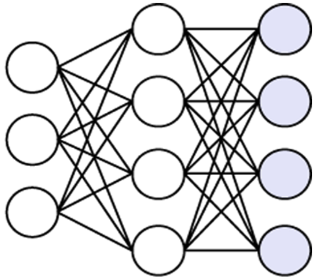
[Koldora CC]

Lingjie Liu

What are neural fields?

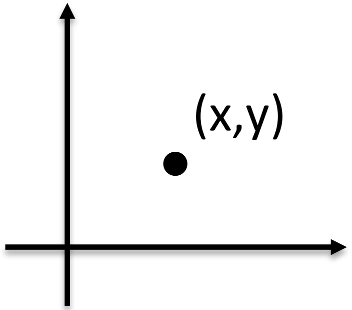


$\Phi: \mathbb{R}^n \rightarrow \mathbb{R}$

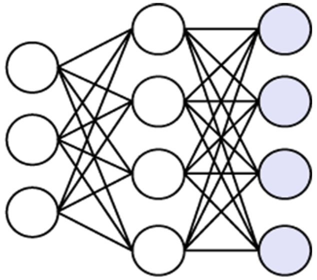


Neural Network (Φ)

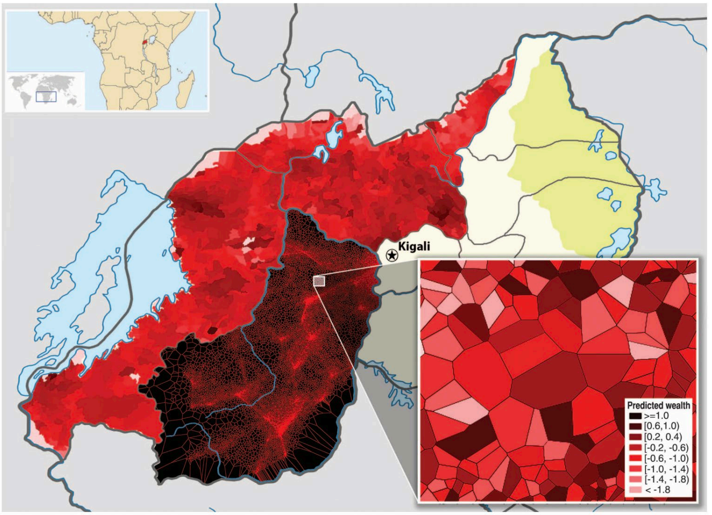
Signed Distance Function (SDF)



$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^n$



Neural Network (Φ)

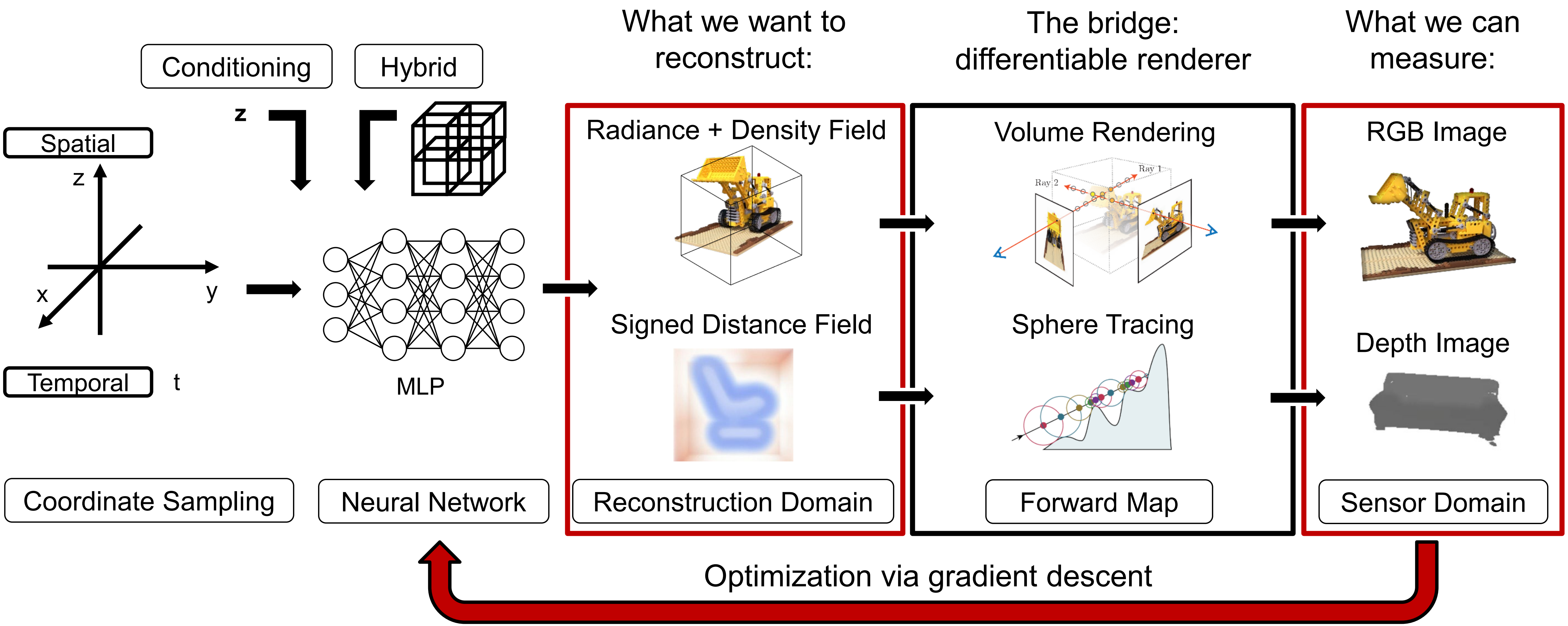


Geospatial Data

[Blumenstock et al. 2015]

Lingjie Liu

Neural Fields General Framework



Lingjie Liu

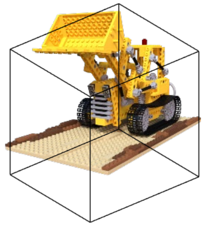
Differentiable Rendering

Reconstruction

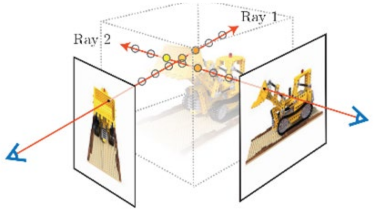
Forward Map

Sensor Domain

Radiance Field

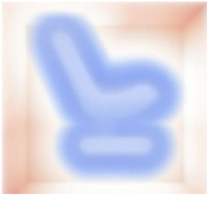


Volume Rendering

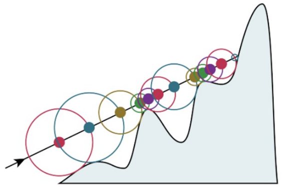


RGB Image Depth

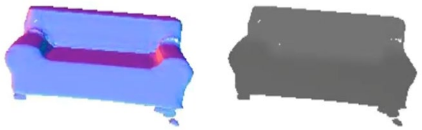
SDF

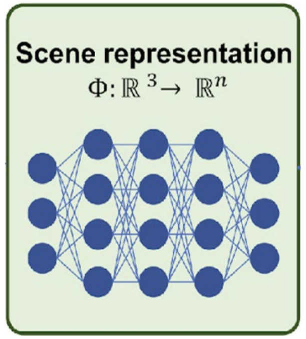


Sphere Tracing



Normal Depth





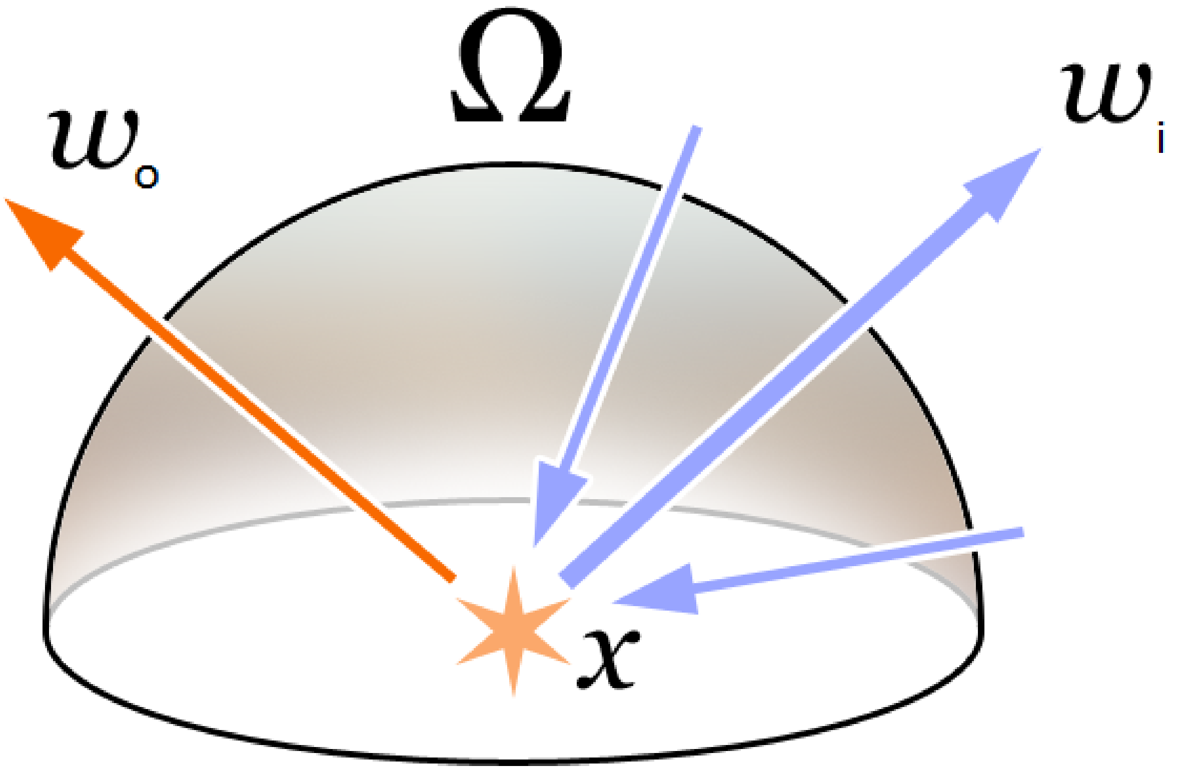
Neural Renderer

Output Rendering

Figures adapted from:
Mildenhall et al. 2020 (NeRF)
Sitzmann et al. 2019 (SRN)

Lingjie Liu

BRDF Shading



$$L(\mathbf{x},\vec{\omega}_o)=L_e(\mathbf{x},\vec{\omega}_o)+\int_S f_r(\mathbf{x},\vec{\omega}_i\rightarrow\vec{\omega}_o)L(\mathbf{x}',\vec{\omega}_i)G(\mathbf{x},\mathbf{x}')V(\mathbf{x},\mathbf{x}')d\omega_i$$

Lingjie Liu

Course Link:
https://neural-representation-2025.github.io/index.html

TAs

Abhimanyu Suthar
Email: absutharus@gmail.com

Chuhao Chen
Email: chuhaoc@seas.upenn.edu

Tentative Syllabus

	Presentation	Paper Presenting Schedule
1	*Sep 2 (Tue)*	Introduction
2	*Sep 4*	Introduction 2
3	*Sep 9 (Tue)*	Presentation Round 1 Start ->
4	*Sep 11*	
5	*Sep 16 (Tue)*	
6	*Sep 18*	
7	*Sep 23 (Tue)*	
8	*Sep 25*	
9	*Sep 30 (Tue)*	
10	*Oct 2*	Presentation Round 1 End <-
11	*Oct 7 (Tue)*	Guest Talk 1
	Oct 9	No Class
12	*Oct 14 (Tue)*	Guest Talk 2

13	*Oct 16*	Presentation Round 2 Start ->
	Oct 21 - Oct 23	No Class
14	*Oct 28 (Tue)*	
15	*Oct 30*	
16	*Nov 4 (Tue)*	
17	*Nov 6*	
18	*Nov 11 (Tue)*	
19	*Nov 13*	
20	*Nov 18 (Tue)*	Presentation Round 2 End <-
21	*Nov 20*	Favorite Paper + New Ideas
22	*Nov 25 (Tue)*	Practice Lecture + Summary
	Nov 27	No Class
23	*Dec 2*	No Class (Tentative)
24	*Dec 4*	No Class (Tentative)

Next Class

1. Present some pioneering works in this field, e.g., NeRF, SRN, Neural Volumes, Gaussian Splatting, NeuS, …
2. Fundamentals of Classical 3D Representations and Rendering in Computer Graphics

Lingjie Liu

2	**(Remarks: the papers within each blue block are similar, choosing one of them to share would be enough, doing a brief summary of others would be greatly welcome)**		
3	Fast Training	Taming 3DGS: High-Quality Radiance Fields with Limited Resources	Sep 9 (Tue)
4		DashGaussian: Optimizing 3D Gaussian Splatting in 200 Seconds	
5	High-quality Rendering	Volumetrically Consistent 3D Gaussian Rasterization	Sep [illegible]e)
6	Generalisable 3D Reconstruction	MVSplat: Efficient 3D Gaussian Splatting from Sparse Multi-View Images	Sep 11 (T[illegible]
7		Flash3D: Feed-Forward Generalisable 3D Scene Reconstruction from a Single Image	Sep 11 (Thu)
8	Fast Visual Geometry	DUSt3R: Geometric 3D Vision Made Easy	Sep 16 (Tue)
9		Spann3R: 3D Reconstruction with Spatial Memory	
10		Fast3R: Towards 3D Reconstruction of 1000+ Images in One Forward Pass	
11		VGGSfM: Visual Geometry Grounded Deep Structure From Motion	Sep 16 (Tue)
12		VGGT: Visual Geometry Grounded Transformer	Sep 18 (Thu)
13		Pi^3: Scalable Permutation-Equivariant Visual Geometry Learning	Sep 18 (Thu)
14	Transformer-based Neural Rendering	LVSM: A Large View Synthesis Model with Minimal 3D Inductive Bias	Sep 23 (Tue)
15		RenderFormer: Transformer-based Neural Rendering of Triangle Meshes with Global Illumination	Sep 23 (Tue)
16	3D Generation	MeshFormer: High-Quality Mesh Generation with 3D-Guided Reconstruction Model	Sep 25 (Thu)
17		Structured 3D Latents for Scalable and Versatile 3D Generation (TRELLIS)	Sep 25 (T[illegible]
18	4D Reconstruction & Generation	SV4D: Dynamic 3D Content Generation with Multi-Frame and Multi-View Consistency	Sep 30 (Tue)
19		SV4D 2.0: Enhancing Spatio-Temporal Consistency in Multi-View Video Diffusion for High-Quality 4D Generation	
20		CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models	
21		MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D Motion Scaffolds	Sep 30 (Tue)
22		Shape of Motion: 4D Reconstruction from a Single Video	Oct 2 (Thu)
23		DIMO: Diverse 3D Motion Generation for Arbitrary Objects	Oct 2 (Thu)

Note: For a paper bundle, you only need to present one of the papers in the bundle according to their preference, but you are encouraged to discuss the connections between the papers in the bundle.

Physics-grounded Reconstruction & Generation	Reconstruction and Simulation of Elastic Objects with Spring-Mass 3D Gaussians	Oct 16 (Tue)
	PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos	
	Vid2Sim: Generalizable, Video-based Reconstruction of Appearance, Geometry and Physics for Mesh-free Simulation	Oct 16 (Tue)
	Gaussian Splashing: Unified Particles for Versatile Motion Synthesis and Rendering	Oct 28 (Tue)
	PhysMotion: Physics-Grounded Dynamics From a Single Image	Oct 28 (Tue)
	WonderPlay: Dynamic 3D Scene Generation from a Single Image and Actions	
Multi-modality	Dr. Splat: Directly Referring 3D Gaussian Splatting via Direct Language Embedding Registration	Oct 30 (Thu)
3D Editing	EditSplat: Multi-View Fusion and Attention-Guided Optimization for View-Consistent 3D Scene Editing with 3D Gaussian Splatting	Oct 30 (Thu)
Robotics	MASt3R-SLAM: Real-Time Dense SLAM with 3D Reconstruction Priors.	Nov 4 (Tue)
	GWM: Towards Scalable Gaussian World Models for Robotic Manipulation	Nov 4 (Tue)
	Splat-Nav: Safe Real-Time Robot Navigation in Gaussian Splatting Maps	Nov 6 (Thu)
	Pre-training Auto-regressive Robotic Models with 4D Representations	Nov 6 (Thu)
Surface Reconstruction	Objects as volumes: A stochastic geometry view of opaque solids	Nov 11 (Tue)
	Geometry Field Splatting with Gaussian Surfels	Nov 11 (Tue)
3D Segmentation & Part-aware Generation	Part123: Part-aware 3D Reconstruction from a Single-view Image	Nov 13 (Thu)
	HoloPart: Generative 3D Part Amodal Segmentation	Nov 13 (Thu)
	PartField: Learning 3D Feature Fields for Part Segmentation and Beyond	
Video Generation	GEN3C: 3D-Informed World-Consistent Video Generation with Precise Camera Control	Nov 18 (Tue)
	Voyaging into Perpetual Dynamic Scenes from a Single View	Nov 18 (Tue)

Before the seminar

- Read the papers of the week.

- Submit at least two questions for discussion before the seminar to a Google form (https://docs.google.com/forms/d/e/1FAIpQLSfSxryv_JO9Ffbd7iKCIqnczqPWJUqv3OGFI6K-2sAKOJmBYQ/viewform). This is important – your contribution will be marked. The deadline for submitting questions is one hour before each class session (so Tuesday 2:30 PM and Thursday 2:30 PM).

During the seminar (Starting from Sept 9, two rounds)

- Overview (10 minutes)
  - The instructor or TAs give a brief introduction on the topic.

- 2x Presentations (each 25 minutes, 25 % of grade):
  - Two pre-assigned participants present the paper of their choice.
  - 5 minutes on motivation, background and related work.
  - 20 minutes of presentation of the paper.

- Discussion and Feedback (30 minutes, 25% of grade across weeks):
  - One participant is assigned at random at the beginning of the seminar to lead the discussion. Everyone leads the discussion at least once in the seminar series.
  - The discussion leader receives a digest of the submitted questions just before the seminar.
  - The discussion leader raises questions appropriately throughout the discussion, covers future work aspects, and finally provides a summary of the strengths and weaknesses of the techniques and of the discipline.
  - The students provide feedback to the presenting student on their presentation with respect to what has worked well, and what could be improved and how.

Grading Criteria

Form (30%)
To time? Verbal speed & clarity? Body posture? Engagement with audience?

Content (50%)
Structure/storyline? Main points? Paper connections? Valid conclusions?

Answers (20%)
Good answers to questions? Knowledgeable?

2x Presentations
(50% of grade)

Moderation (30%)
Integrates questions well? Pushes forward discussion?
Good summary? Strengths and weaknesses of paper?

Questions (70%)
One question per paper (two questions per class) should be submitted at least one hour before the class during which the paper will be presented. (However, students are permitted to submit questions late (but before the discussion), up to two occurrences, without facing any penalties)

Discussion
(25% of grade)

Practice Lecture (30%)
Listen attentively to the lecture? Engage in small coding exercises?

Discussion on your favorite papers (35%)
Each person has 5 minutes to present their favorite paper. Is your presentation clear? Explain clearly why you chose it and what you like about it?

Brainstorming (35%)
Actively participate in the discussion? Contribute your own ideas or opinions?

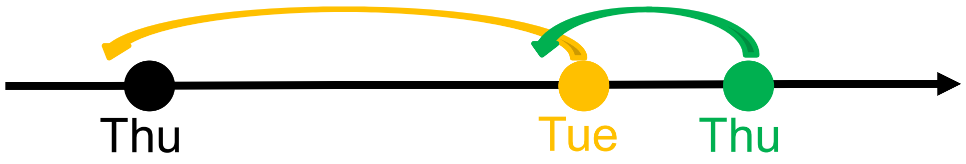
Other Activity Participation
(25% of grade)

TODOs After this Class

1. ***Paper Selection and Registration***: **[Important! Deadline: Sept 9]**
Please select and register for the two papers you would like to present using the following Excel link:
https://docs.google.com/spreadsheets/d/1_FJueXqWnKWoYOGZTNiwp2qRmSP0u1H6ayEYE5j3lb0/edit?gid=0#gid=0

2. ***Presentation Preparation***:

- Ensure you are fully prepared **one class before your scheduled class for presentation**.
- Upload your slides to the Google folder (https://drive.google.com/drive/folders/1OIKCn562KA3sPUGh-9x8mFuqazU92Cyk) **at least one hour before the class prior to your assigned class for presentation**. This is important in case of an emergency requiring us to reschedule your talk.
- For example, if you're presenting on Tuesday, upload your slides by the previous Thursday at 2:30 PM. If presenting on Thursday, upload by Tuesday at 2:30 PM.



3. ***Class Participation***:
•Before each class, please read the papers that will be discussed and submit two questions **at least one hour before the class** using the following link:
https://docs.google.com/forms/d/e/1FAIpQLSeVtMP-_PzgtJ9ZthAisBP5MZZ5JqBDrUQ4qqNkusmqVrnquQ/viewform?usp=sharing&ouid=101671697877703922510

TODOs NOW

Find the first four speakers for the first two classes next week (Sept 9 and Sept 11).

Bonus: You can earn 3 bonus point for Tuesday's class and 2 bonus points for Thursday's class toward your final score!

Acknowledgments

- Advances in Neural Rendering
- Neural Fields in Visual Computing and Beyond
- awesome-NeRF: a curated list of awesome neural radiance fields papers
- MPII Summer Semester 2023: Computer Vision and Machine Learning for Computer Graphics

Any Questions?