JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

EN.601.482/682 Deep Learning

Introduction to Neural Radiance Fields

Mathias Unberath, PhD
Assistant Professor
Dept of Computer Science
Johns Hopkins University

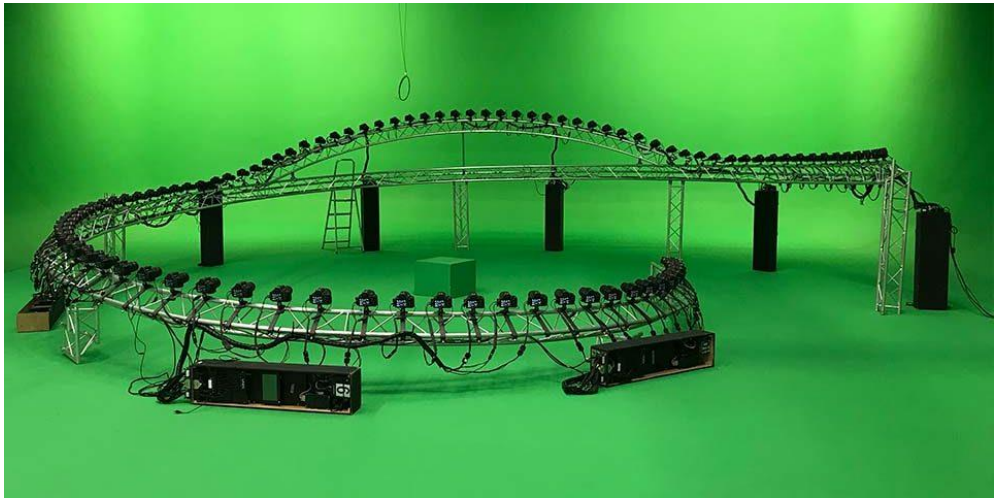
General Idea

Novel View Synthesis

- Collect a set of images of the same object
- Establish relative geometry between these views
- “Interpolate between the views” to create images from **new** viewpoints

Perhaps the First Real World Impact?

Matrix – Bullet Time



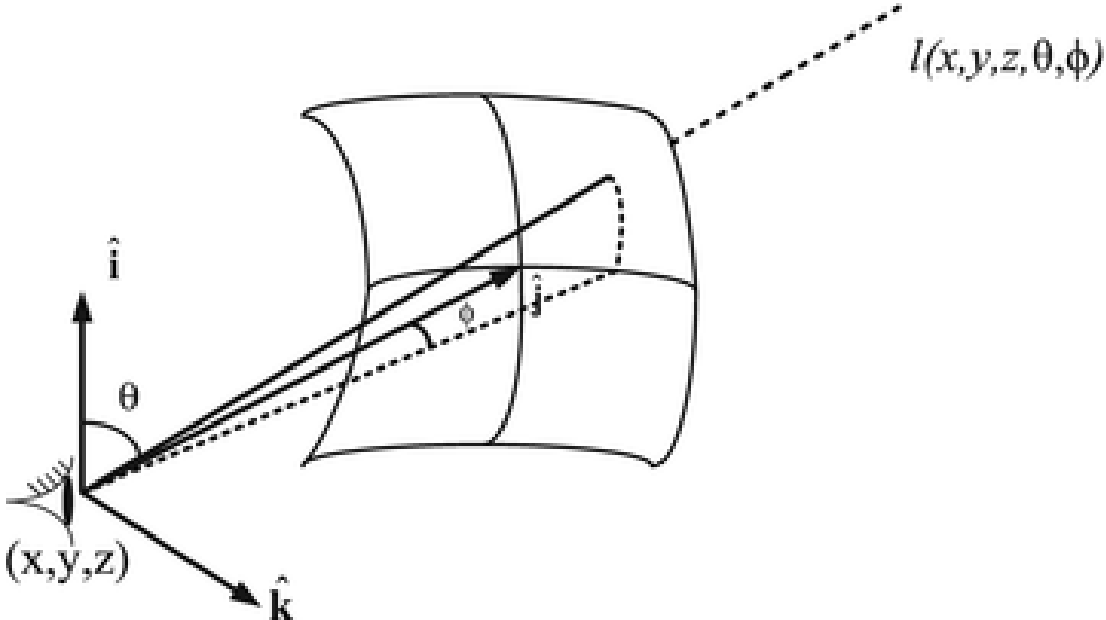
Intro NeRF

Novel View Synthesis

The Plenoptic Function

- Full parametrization of light in space
  - Every ray
  - Every wavelength
  - Every viewpoint
  - For every timepoint
  - P(θ, φ, λ, x, y, z, t) → 7D function

- Simplifications
  - Single viewpoint
  - Grayscale
  - Static scene
  - …



The Early Days

- Reduction to from 7 to 4 DoF
  - Static scene
  - Monochromatic function (3 functions for RGB)
  - Bounded object: only rays are important
- Known as the Lumigraph

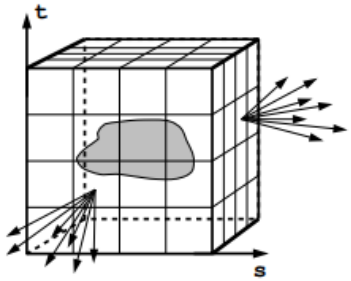
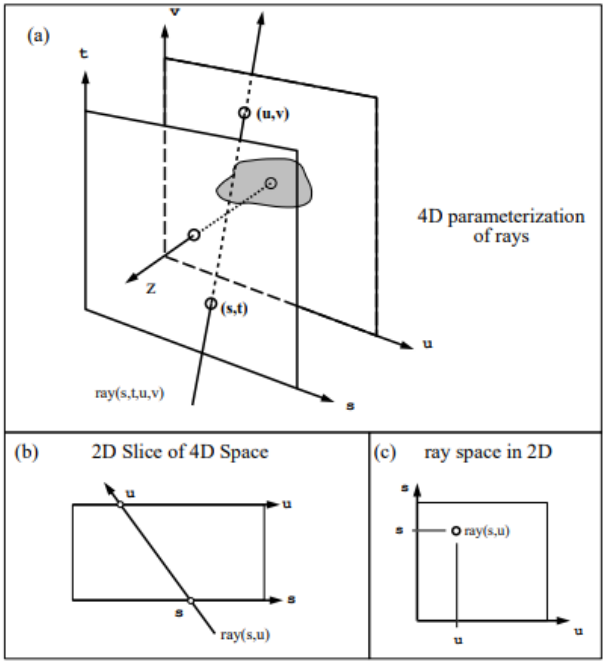


Figure 1: The surface of a cube holds all the radiance information due to the enclosed object.

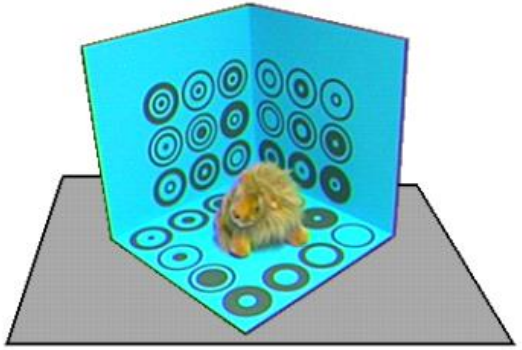


Gortler, S. J., Grzeszczuk, R., Szeliski, R., & Cohen, M. F. (1996, August). The lumigraph. Computer graphics and interactive techniques (pp. 43-54).

The Early Days

- Capturing the Lumigraph
  - Acquire images on the hemisphere around object
  - Camera calibration to estimate camera poses (SfM-style algorithm)
  - Optimizing for Lumigraph grid points
    Non-trivial, because rays may not intersect with grid points

- Synthesizing images
  - Specify camera position
  - For each ray (defined by pixel idx)
  - Map textured region to input



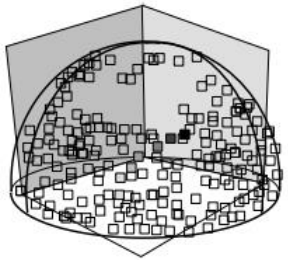


Figure 11: The user interface for the image capture stage displays the current and previous camera positions on a viewing sphere. The goal of the user is to "paint" the sphere.

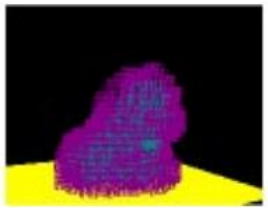


Figure 12: Segmented image plus volume construction

Gortler, S. J., Grzeszczuk, R., Szeliski, R., & Cohen, M. F. (1996, August). The lumigraph. Computer graphics and interactive techniques (pp. 43-54).

Recent Improvements

- For mixed reality, the two-plane setup is inadequate
  → Limits the possible viewing directions

- Different plenoptic function representation: Spherical Fibonacci point sets

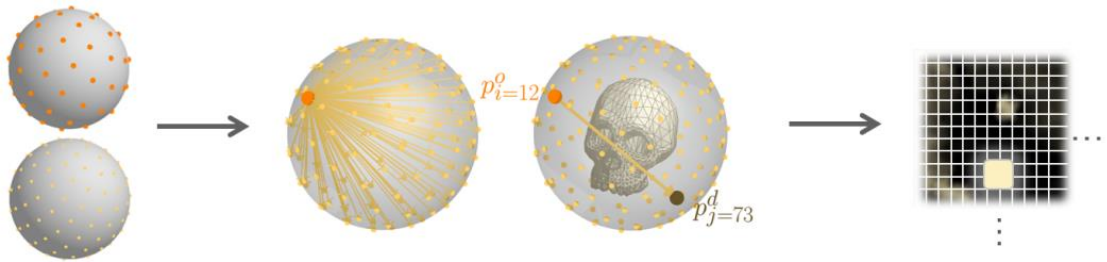
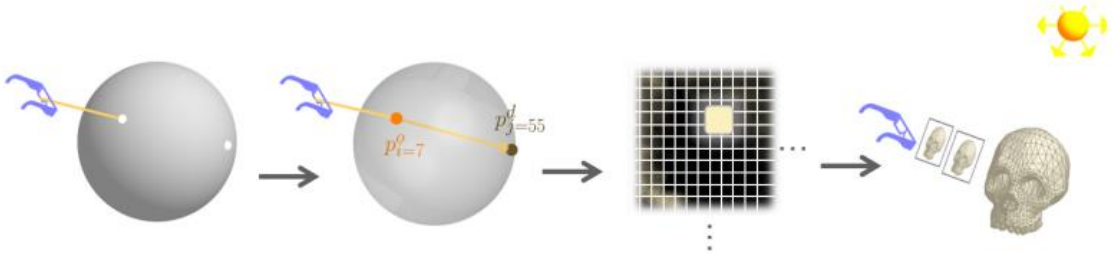


Fig. 1. $\hat{L}(i,j)$ and filling the texture. The surface of the bounding sphere **S** is discretized by the two point sets P_o^M and P_d^N. Rays are traced from each p_i^o to each p_j^d resulting in re-parameterization and discretization of the plenoptic function, referred to as $\hat{L}(i,j)$. The value of $\hat{L}(i,j)$ is written to a 2D texture at position (i,j).

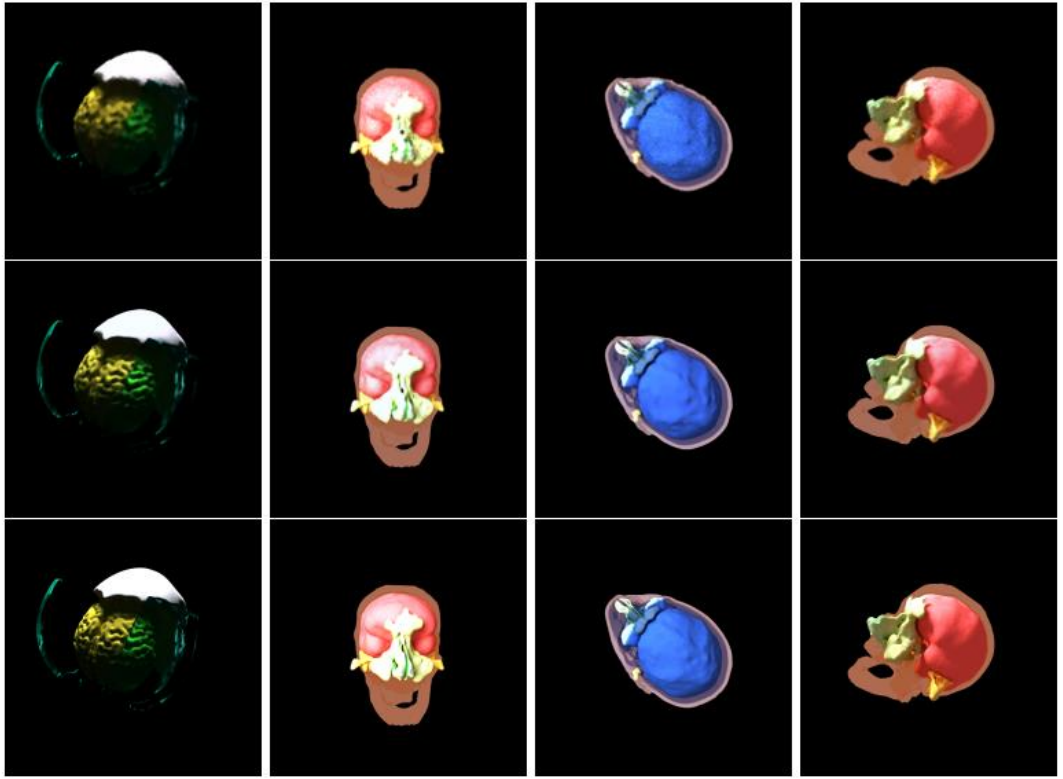
Fink, L., Lee, S. C., Wu, J. Y., Liu, X., Song, T., Velikova, Y., ... & Unberath, M. (2019). Lumipath–towards real-time physically-based rendering on embedded devices. MICCAI



Mixed reality applications!

- Limited hardware resources
  - Storage is small
  - Compute is limited
- Rendering quality is poor

→ Post-rendering correction using CNNs for sharpening



Fink, L., Lee, S. C., Wu, J. Y., Liu, X., Song, T., Velikova, Y., ... & Unberath, M. (2019). Lumipath–towards real-time physically-based rendering on embedded devices. MICCAI

The Later Days

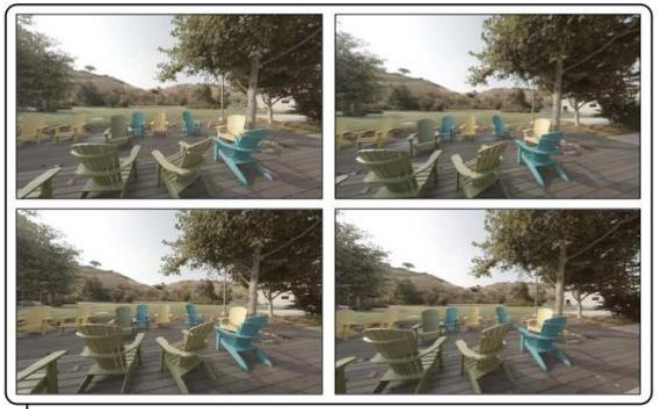
- Approximating the plenoptic function with multi-plane image (MPIs)
  - Stack of semi-transparent, colored layers
  - Arranged at various depths

- Reconstructing MPIs from view images is an ill-posed inverse problem
  - Similar to computed tomography reconstruction or deblurring
  - # parameters is much larger than effective number of evidence

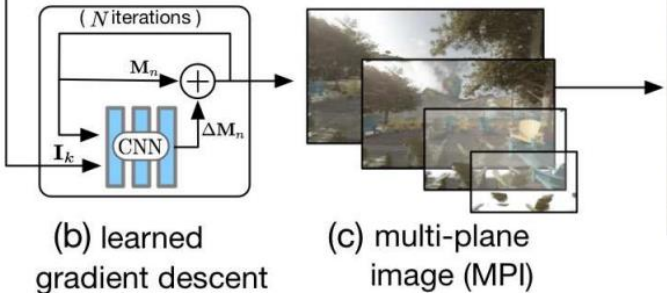
- Introduction of “learned gradient descent” to solve the inverse problem
  - Initially proposed for computed tomography reconstruction
  - Update network learns to generate representations that stay within the manifold of natural scenes
  - Larger update steps → Faster convergence

Flynn, J., Broxton, M., Debevec, P., DuVall, M., Fyffe, G., Overbeck, R., ... & Tucker, R. (2019). Deepview: View synthesis with learned gradient descent. CVPR

The Later Days

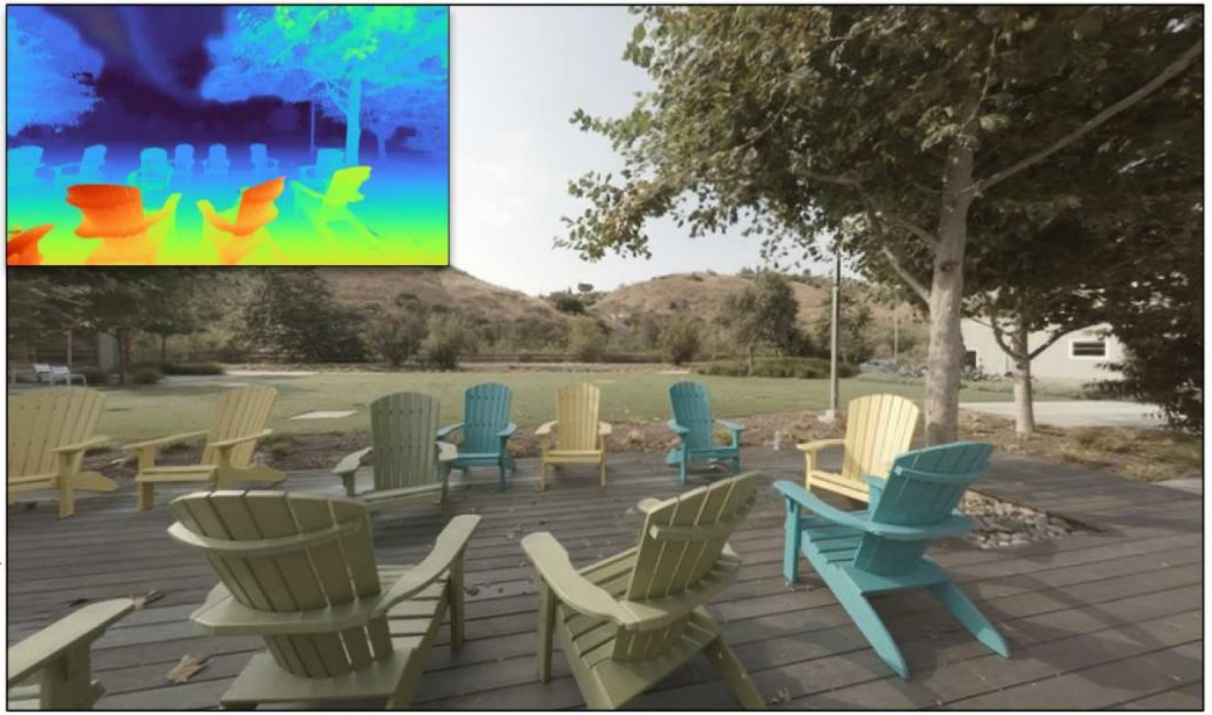


(a) sparse input views



(b) learned gradient descent

(c) multi-plane image (MPI)



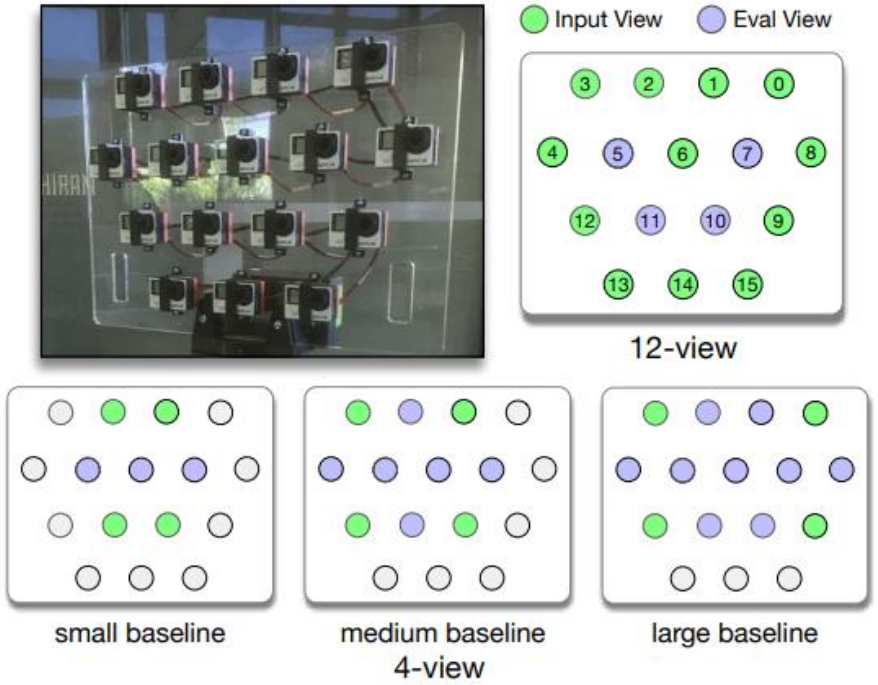
(d) novel synthesized view & depth visualization

Flynn, J., Broxton, M., Debevec, P., DuVall, M., Fyffe, G., Overbeck, R., ... & Tucker, R. (2019). Deepview: View synthesis with learned gradient descent. CVPR

The Later Days

Caveat:

MPI formulation limits viewpoint range



Flynn, J., Broxton, M., Debevec, P., DuVall, M., Fyffe, G., Overbeck, R., ... & Tucker, R. (2019). Deepview: View synthesis with learned gradient descent. CVPR

Intro NeRF

Neural Rendering

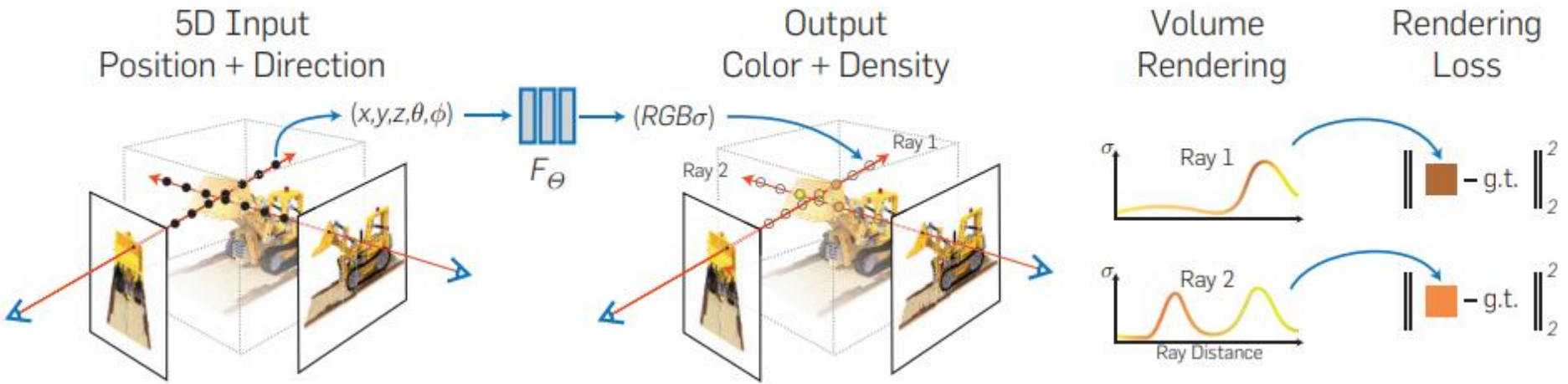
How do we achieve **THIS**?

Neural Radiance Fields – NeRF

A very deep learn-y idea:

If I cannot well model the plenoptic function, …

…why not just approximate it with an MLP?



Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*.

Neural Radiance Fields – NeRF

- 3D location (x,y,z)
- Viewing direction (θ, φ) represented as Cartesian unit vector **d**

- Plenoptic function represented as MLP: $f_\theta(\mathbf{x},\mathbf{d}) \rightarrow (\mathbf{c},\sigma)$
  - **c** is (r,g,b) the emitted color
  - σ is it the volume density

- Importantly: σ should be *independent* of viewing direction **d**!
  - f_θ first processes x (8 FC layers, ReLU activation); 256 channels per layer
  - Outputs σ and 256-dim feature vector
  - Concatenated with viewing direction, followed by another FC layer → **c**

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*.

Neural Radiance Fields – NeRF

Volume rendering based on this MLP representation

- Apply principles from classical rendering
- σ(**x**) can be interpreted as probability of ray terminating at **x**
  → the expected color then is

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\sigma(\mathbf{r}(t))\mathbf{c}(\mathbf{r}(t), \mathbf{d})dt,$$

$$\text{where } T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))ds\right).$$

- T(t) is the "accumulated transmittance":
  Probability that ray traverses to t without hitting another particle

- Then, simply train NeRF by optimizing $\mathcal{L} = \sum_{\mathbf{r}\in\mathcal{R}} \left\| \hat{C}(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2$

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*.

Neural Radiance Fields – NeRF

Important tricks

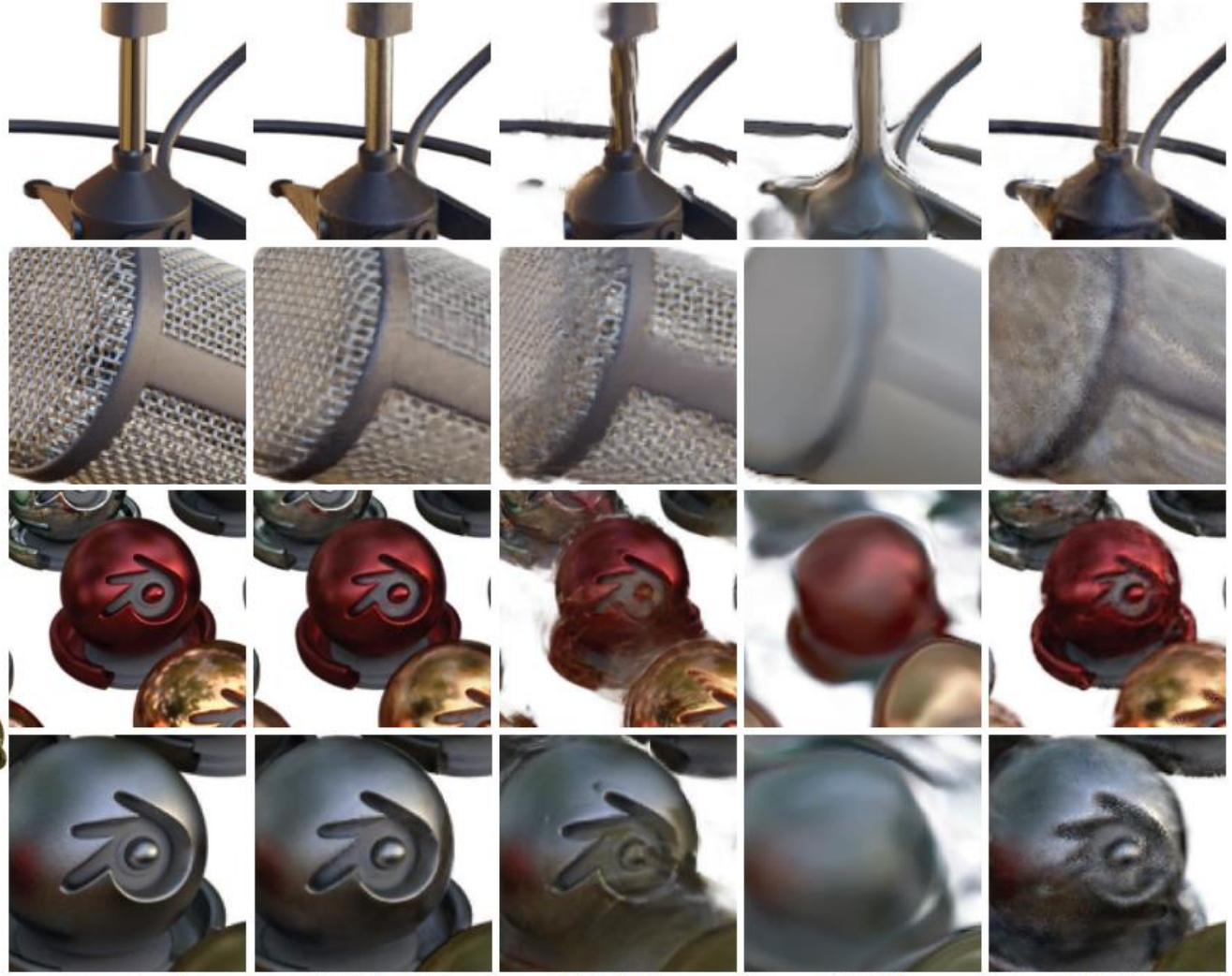
- Hierarchical sampling
  - Querying the MLP at finite sample positions along ray introduces rasterization artifacts
  - Query in steps: First coarse, then finer around high-density area

- Positional encoding
  - We want a function that can change rapidly, even with small changes in x,y,z
  - → Increase difference in representation between **x** and **x**+Δ**x**
  - → Positional encoding $\gamma(p) = \left(\sin(2^0 \pi p), \cos(2^0 \pi p), \cdots, \sin(2^{L-1} \pi p), \cos(2^{L-1} \pi p)\right)$
  - → Applied to all coordinates in **x** and **d** independently (L=10 and L=4, respectively)

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*.

Microphone

Materials



Ground Truth | NeRF (ours) | LLFF [12] | SRN [21] | NV [8]

Quite Promising Results

Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2021). Nerf: Representing scenes as neural radiance fields for view synthesis. *ECCV*.

Intro NeRF

Fast Neural Rendering

Accelerating NeRF

An important caveat:
Training NeRFs takes a day or more on conventional GPUs

Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, *41*(4), 1-15.

Accelerating NeRF

- Some form of encoding is necessary so that MLPs can focus on learning graphic primitives
- More powerful encodings enable more efficient & smaller MLPs

All params in MLP
No params in encoding

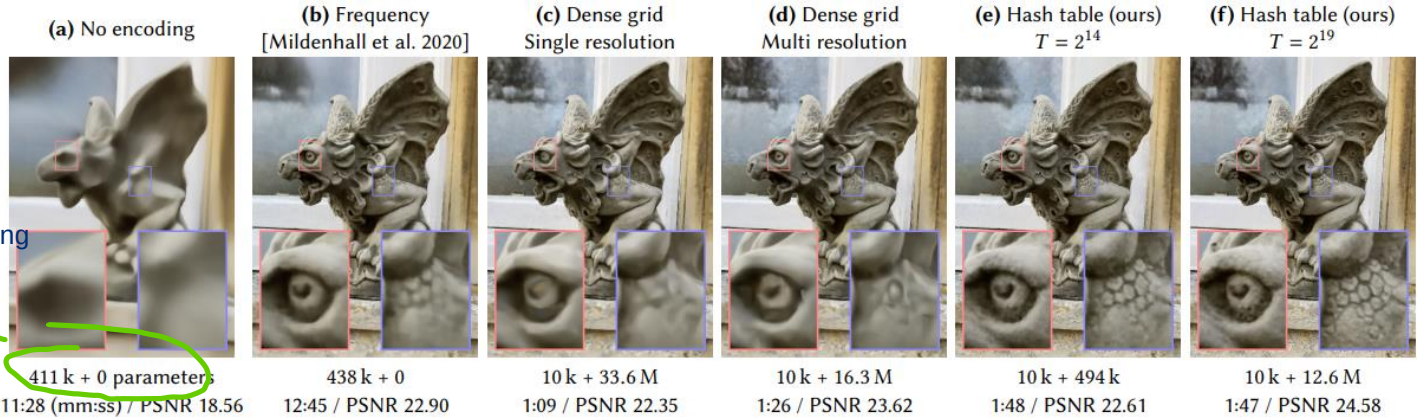
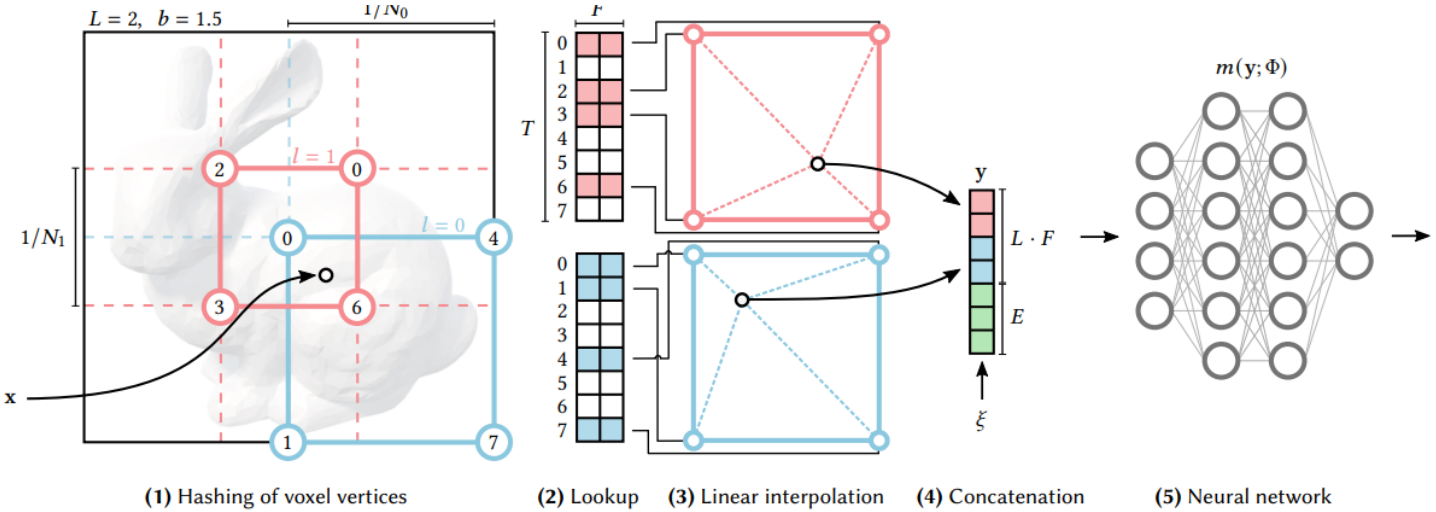


Fig. 2. A demonstration of the reconstruction quality of different encodings and parametric data structures for storing trainable feature embeddings. Each configuration was trained for 11 000 steps using our fast NeRF implementation (Section 5.4), varying only the input encoding and MLP size. The number of trainable parameters (MLP weights + encoding parameters), training time and reconstruction accuracy (PSNR) are shown below each image. Our encoding **(e)** with a similar total number of trainable parameters as the frequency encoding configuration **(b)** trains over 8× faster, due to the sparsity of updates to the parameters and smaller MLP. Increasing the number of parameters **(f)** further improves reconstruction accuracy without significantly increasing training time.

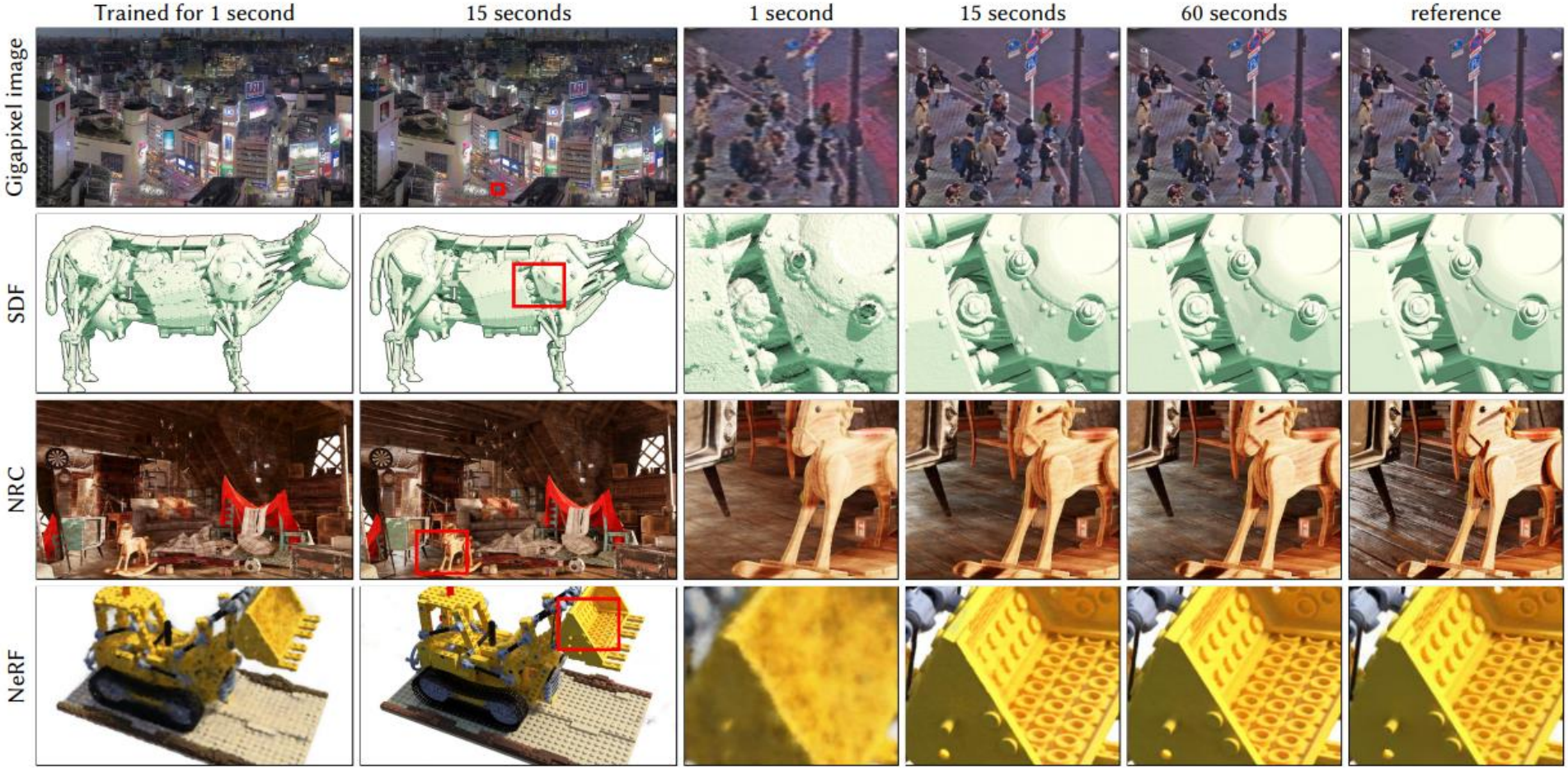
Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, *41*(4), 1-15.

Accelerating NeRF

- Multi-resolution hash grid encoding
- Interpolation at all resolution levels, then concatenation of representation
- Efficiency: Hash tables are O(1) and map well to modern GPUs
  - There are, however, many more small tricks to make this as fast as it is



Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, *41*(4), 1-15.



Müller, T., Evans, A., Schied, C., & Keller, A. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (ToG)*, *41*(4), 1-15.

Intro NeRF

NeRF in the Wild

Dealing with Real World Data

- Crowd-sourced images are often messy
  - Variable occluders
  - Lighting conditions
  - …

Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. CVPR

Dealing with Real World Data

- Crowd-sourced images are often messy
  - Variable occluders
  - Lighting conditions
  - …

- Idea: Decompose scene into static and transient content

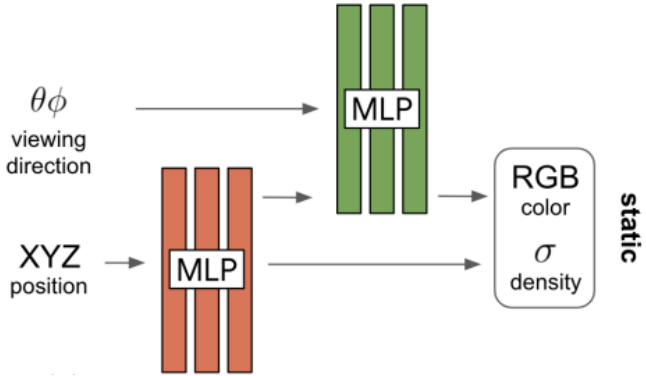
(a) Static (b) Transient (c) Composite (d) Image (e) Uncertainty

Figure 4: NeRF-W separately renders the static (a) and transient (b) elements of the scene, and then composites them (c). Training minimizes the difference between the composite and the true image (d) weighted by uncertainty (e), which is simultaneously optimized to identify and discount anomalous image regions. Photo by Flickr user vasnic64 / CC BY.

Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. CVPR

The NeRF-W Model

- Two MLPs
  - One for density $[\sigma(t), \mathbf{z}(t)] = \mathrm{MLP}_{\theta_1}(\gamma_{\mathbf{x}}(\mathbf{r}(t)))\,,$
  - One for color $\mathbf{c}(t) = \mathrm{MLP}_{\theta_2}(\mathbf{z}(t), \gamma_{\mathbf{d}}(\mathbf{d}))$



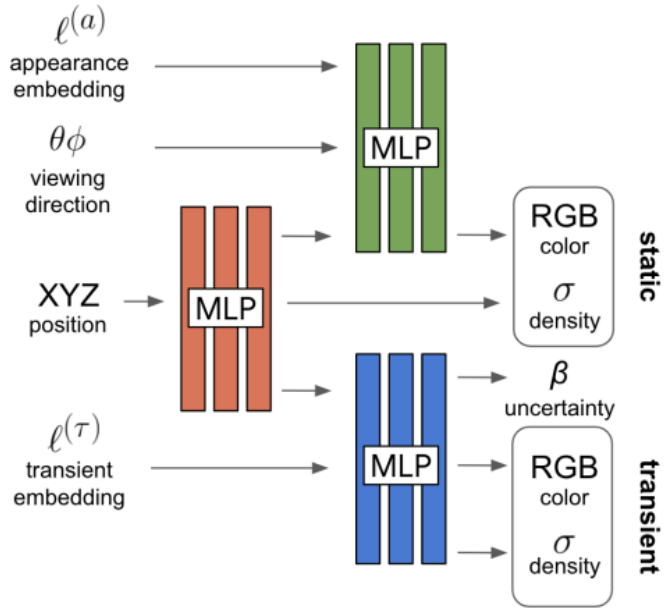
Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. CVPR

The NeRF-W Model

- Two MLPs
  - One for density $[\sigma(t), \mathbf{z}(t)] = \mathrm{MLP}_{\theta_1}(\gamma_{\mathbf{x}}(\mathbf{r}(t)))$,
  - One for color $\mathbf{c}(t) = \mathrm{MLP}_{\theta_2}(\mathbf{z}(t), \gamma_{\mathbf{d}}(\mathbf{d}))$

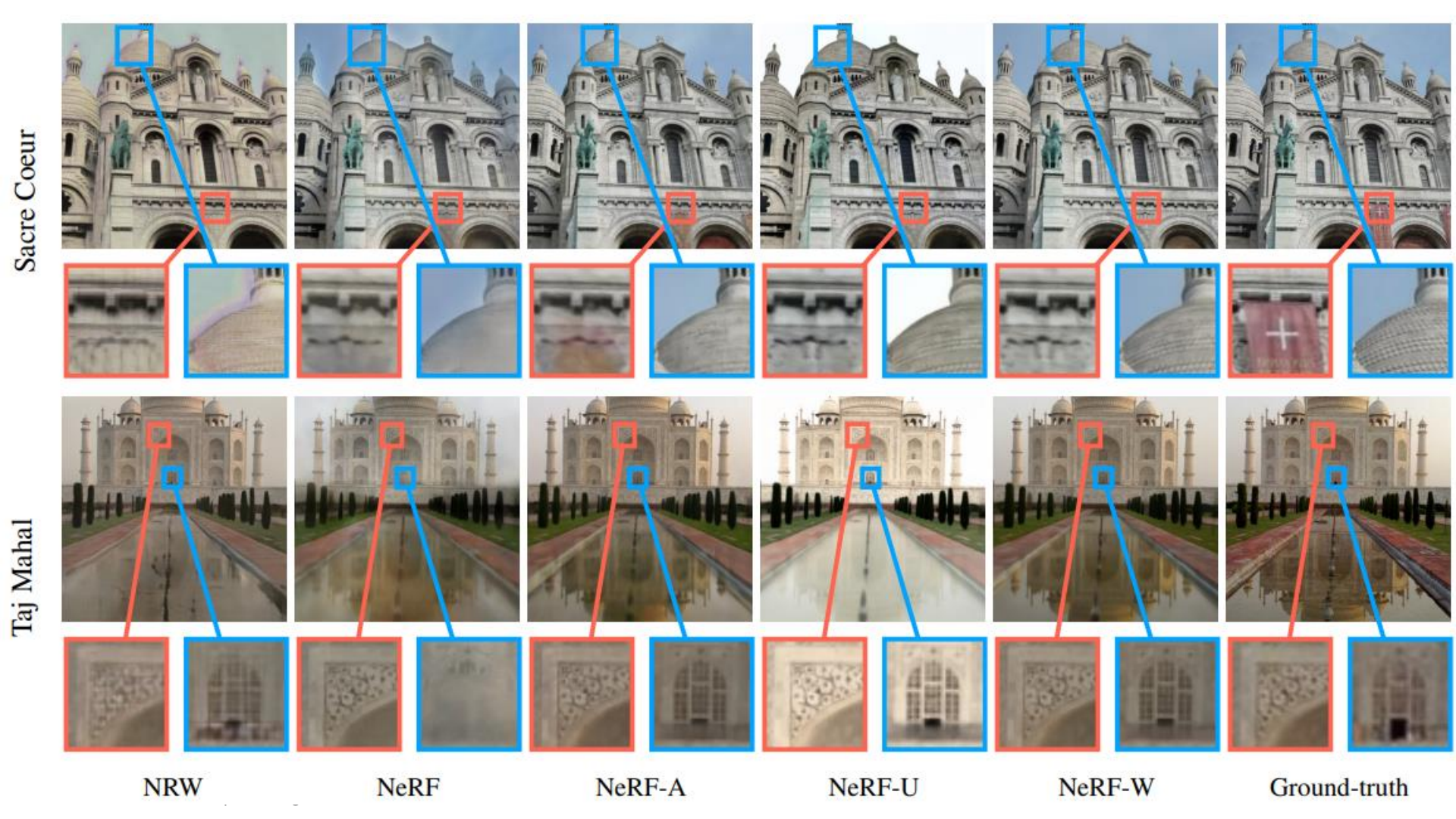
- Photometric variation
  - Embedding l^a to account for variable light
  - Results in image-dependent radiance **c_i**

- Transient objects
  - Embedding l^T to account for occluders
  - Transient MLP head



- Final luminance for ray is then a composite of static and transient components

Martin-Brualla, R., Radwan, N., Sajjadi, M. S., Barron, J. T., Dosovitskiy, A., & Duckworth, D. (2021). Nerf in the wild: Neural radiance fields for unconstrained photo collections. CVPR


Sacre Coeur
Taj Mahal
NRW
NeRF
NeRF-A
NeRF-U
NeRF-W
Ground-truth

Intro NeRF

Scene Reconstruction using Neural Fields

NeRF for Scene Reconstruction

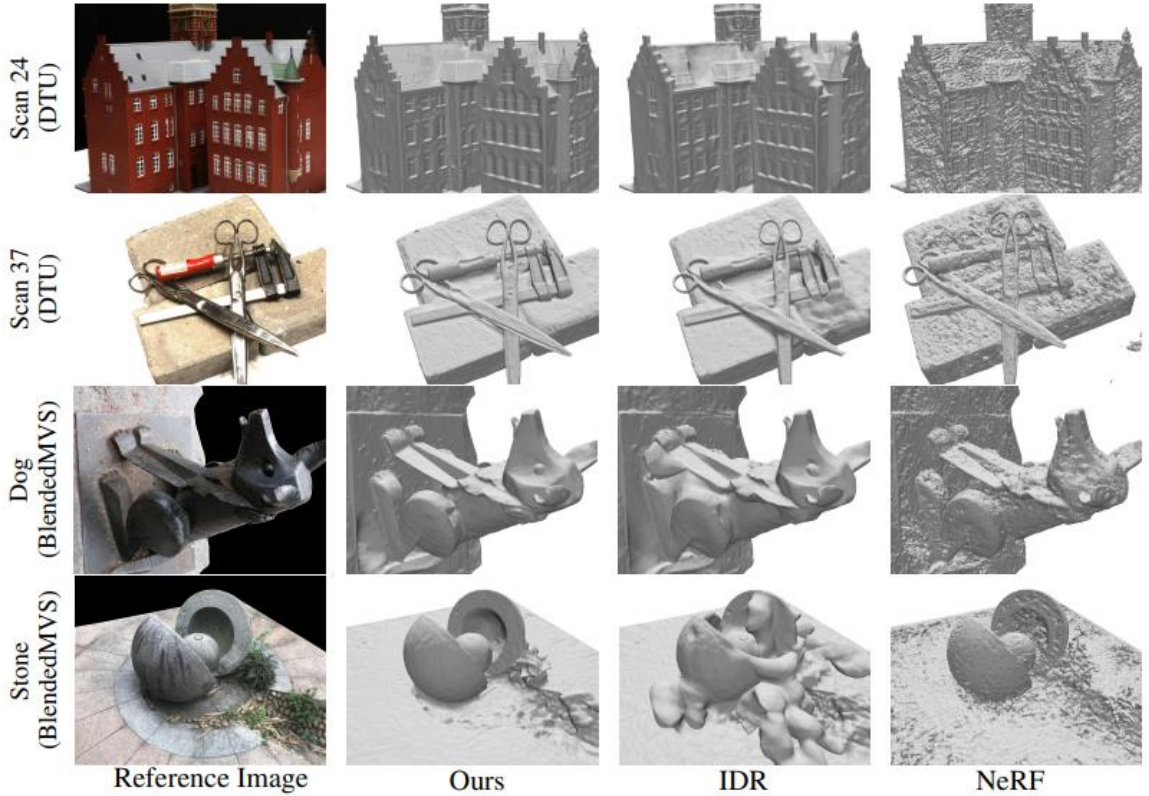
- Turns out:
  - Due to the volume density representation, NeRF is not precise in reconstruction
  - Further, because density is arbitrary, need to select threshold for every surface extraction
  - Surface is encoded in density, but implicitly

- Encode surface explicitly: Neural Distance Fields
  - Two functions encoded as MLPs
    - One to map spatial position to signed distance to object
    - One to encode color associated with point and direction
  - Surface then is the zero-level set of the SDF

- Rendering: Convert SDF to density and follow previous approach

Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., & Wang, W. (2021). Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. arXiv preprint arXiv:2106.10689.

NeRF for Scene Reconstruction



Wang, P., Liu, L., Liu, Y., Theobalt, C., Komura, T., & Wang, W. (2021). Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. arXiv:2106.10689.

Neuralangelo: Multi-resolution Hash with Neural SDF

- Combine explicit learning of the surface with multi-resolution hash grids

→ Benefits from increased representation of hash encodings
→ Capitalizes on the explicit definition of the surface SDF

- Training is not trivial
  - Hash grids provide high resolution
  - However, analytical gradient would only provide information for local
  - Use numerical derivatives to ensure smoothness of computation

- Multi-resolution optimization

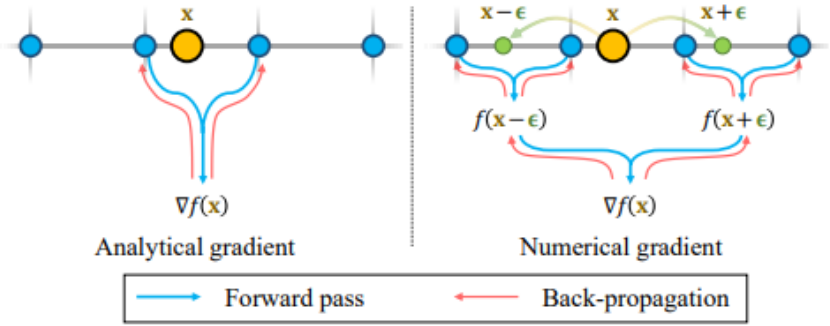


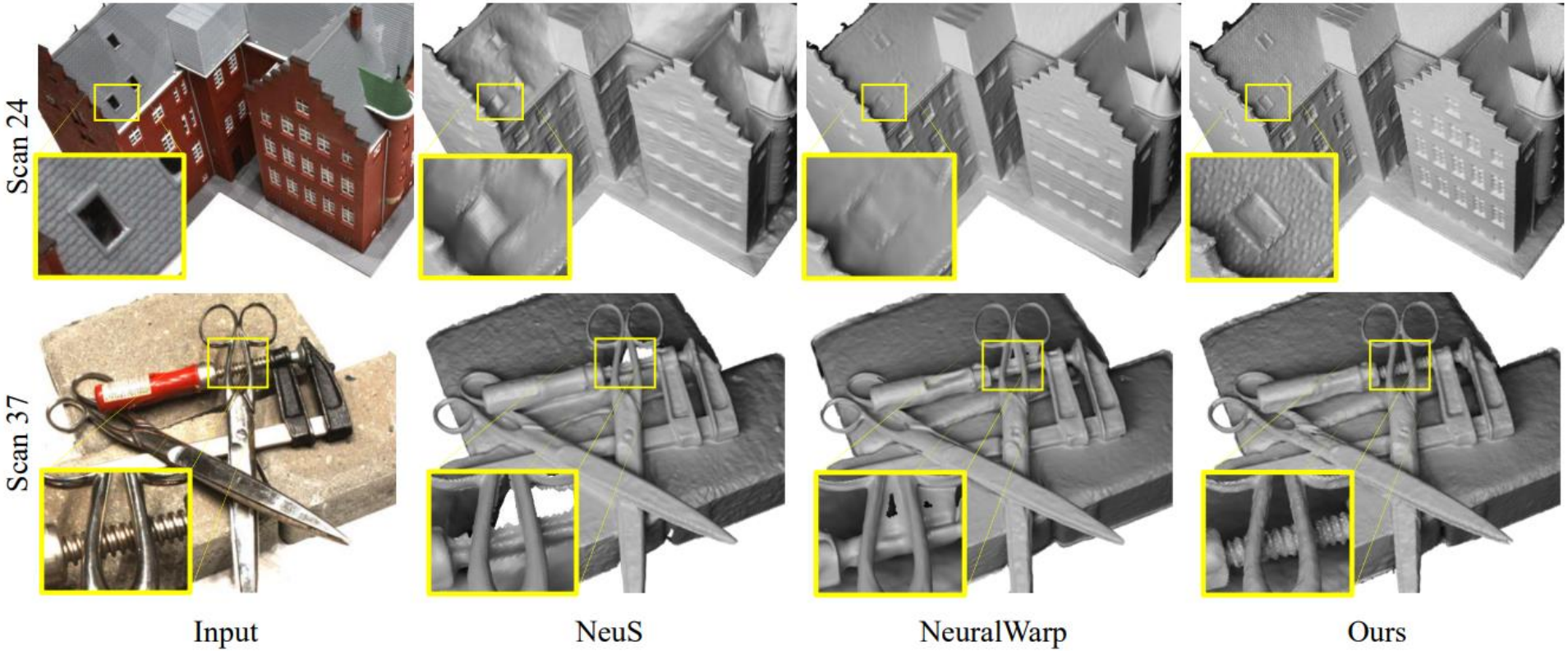
Figure 2. Using **numerical gradients** for higher-order derivatives distributes the back-propagation updates beyond the local hash grid cell, thus becoming a smoothed version of **analytical gradients**.

Li. Z, Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H. (2023) Neuralangelo: High-Fidelity Neural Surface Reconstruction. CVPR

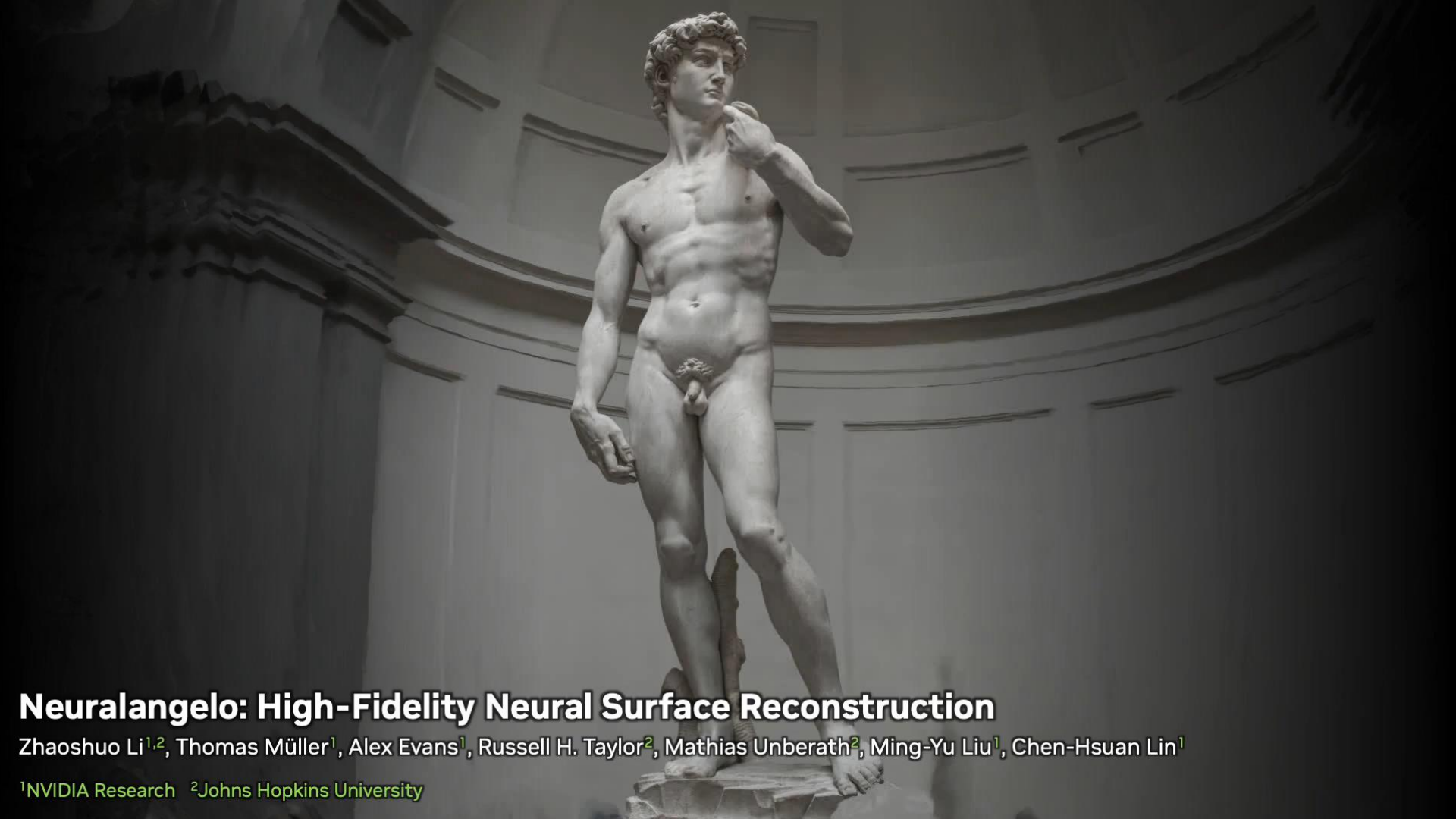
Neuralangelo: Multi-resolution Hash with Neural SDF

Li. Z, Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H. (2023) Neuralangelo: High-Fidelity Neural Surface Reconstruction. CVPR

Neuralangelo: Multi-resolution Hash with Neural SDF



Li. Z, Müller, T., Evans, A., Taylor, R.H., Unberath, M., Liu, M.Y., Lin, C.H. (2023) Neuralangelo: High-Fidelity Neural Surface Reconstruction. CVPR



Neuralangelo: High-Fidelity Neural Surface Reconstruction

Zhaoshuo Li[1,2], Thomas Müller[1], Alex Evans[1], Russell H. Taylor[2], Mathias Unberath[2], Ming-Yu Liu[1], Chen-Hsuan Lin[1]

[1]NVIDIA Research [2]Johns Hopkins University

Intro NeRF

Questions?