NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis

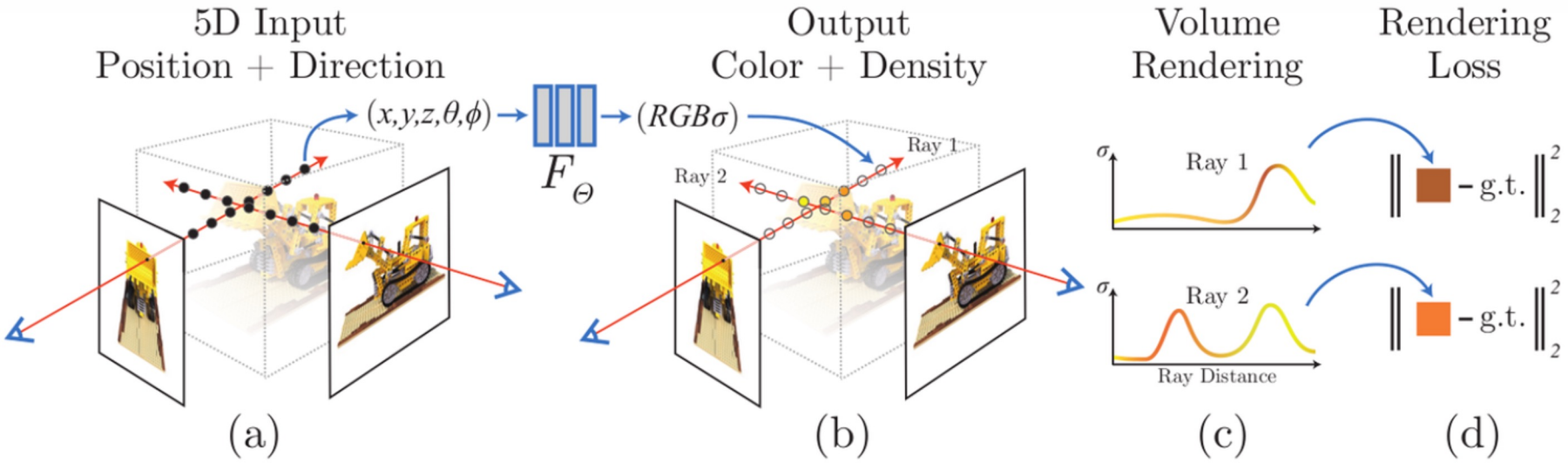
Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, Ren Ng

Presenter: Wenfan Jiang

EECS ELECTRICAL ENGINEERING AND COMPUTER SCIENCE
UNIVERSITY OF MICHIGAN

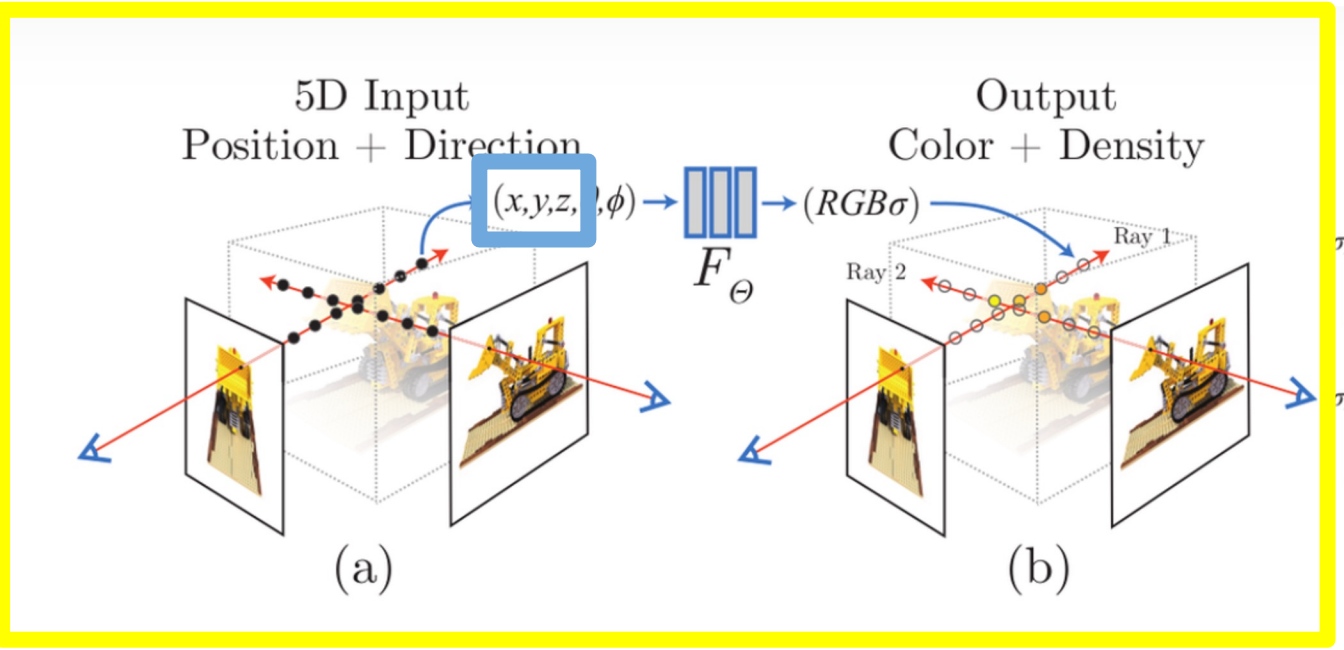
What is Neural Radiance Field (NeRF)?

“A method that achieves state-of-the-art results for synthesizing novel views of complex scenes by optimizing an underlying continuous volumetric scene function using a sparse set of input views.”



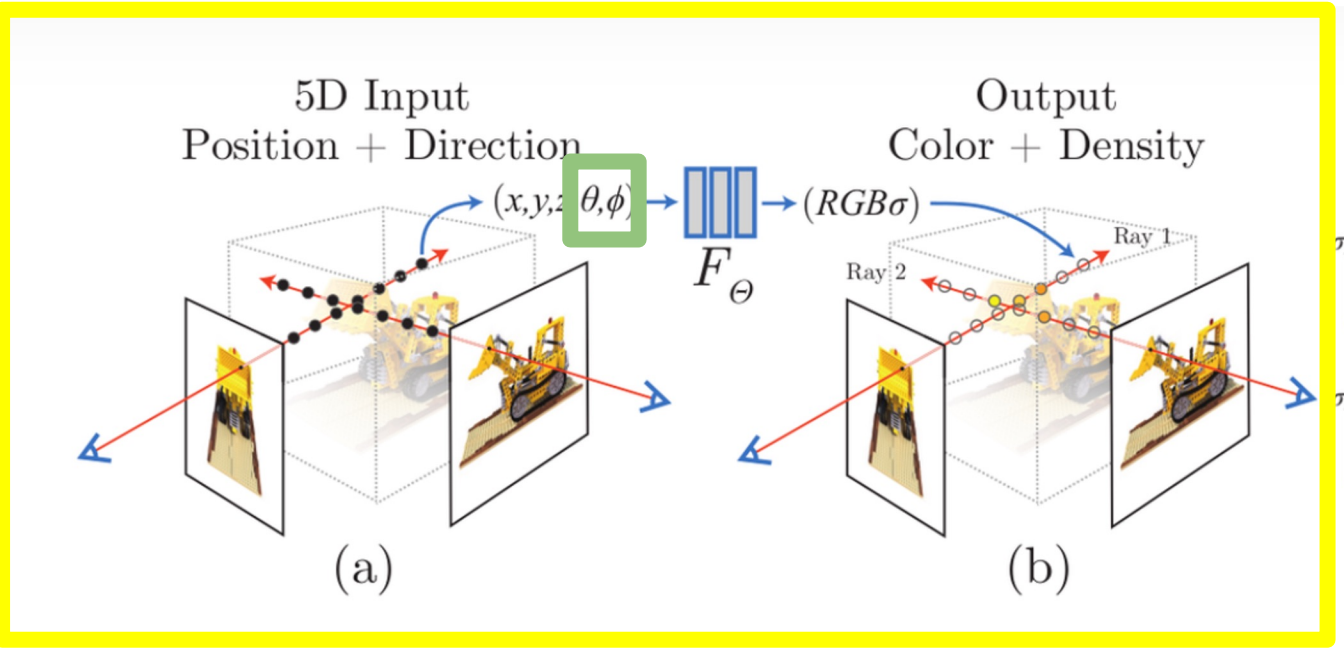
How does NeRF represent a 3D scene?

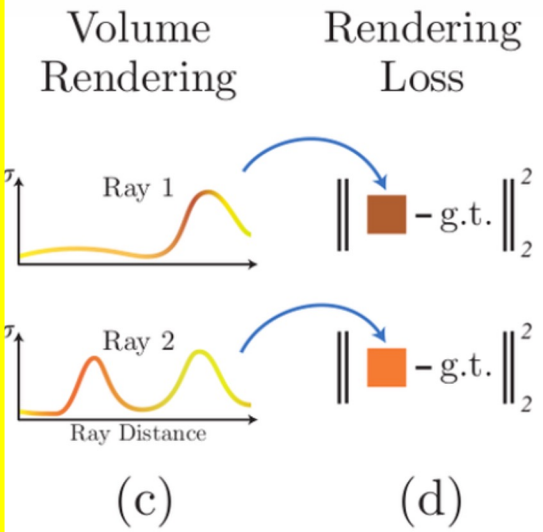
The NeRF function represents a continuous scene as a function with a 5D input vector, which includes the 3D-coordinates of a spatial point and the viewing direction (colatitude and longitude angle).



How does NeRF represent a 3D scene?

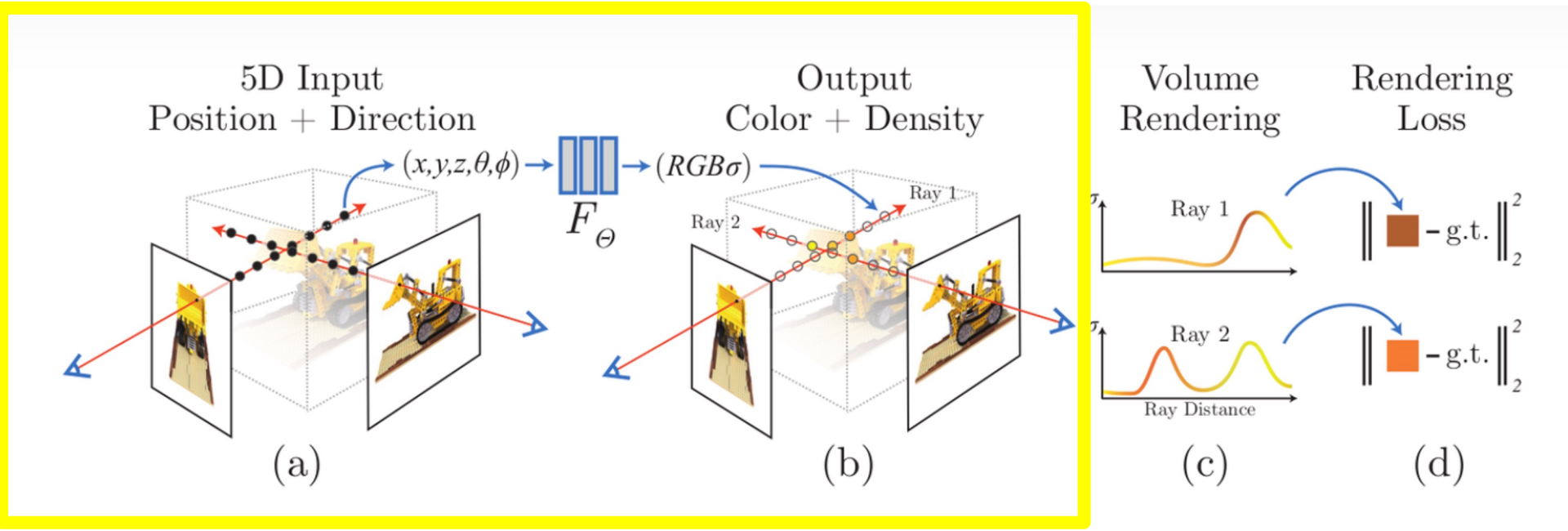
The NeRF function represents a continuous scene as a function with a 5D input vector, which includes the 3D-coordinates of a spatial point and the viewing direction (colatitude and longitude angle).





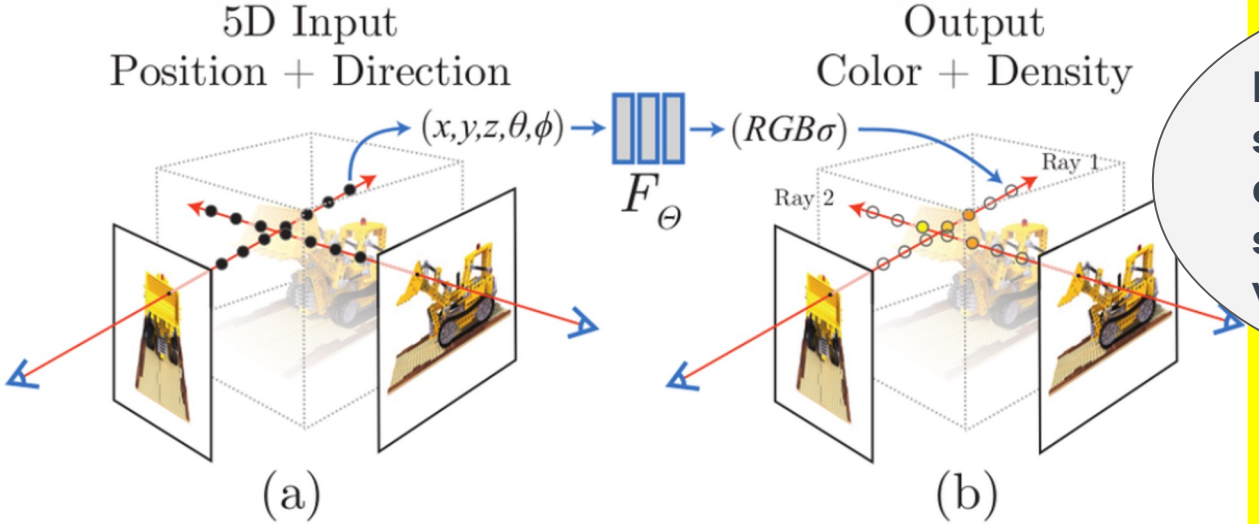
How does NeRF represent a 3D scene?

$$F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$$



How does NeRF represent a 3D scene?

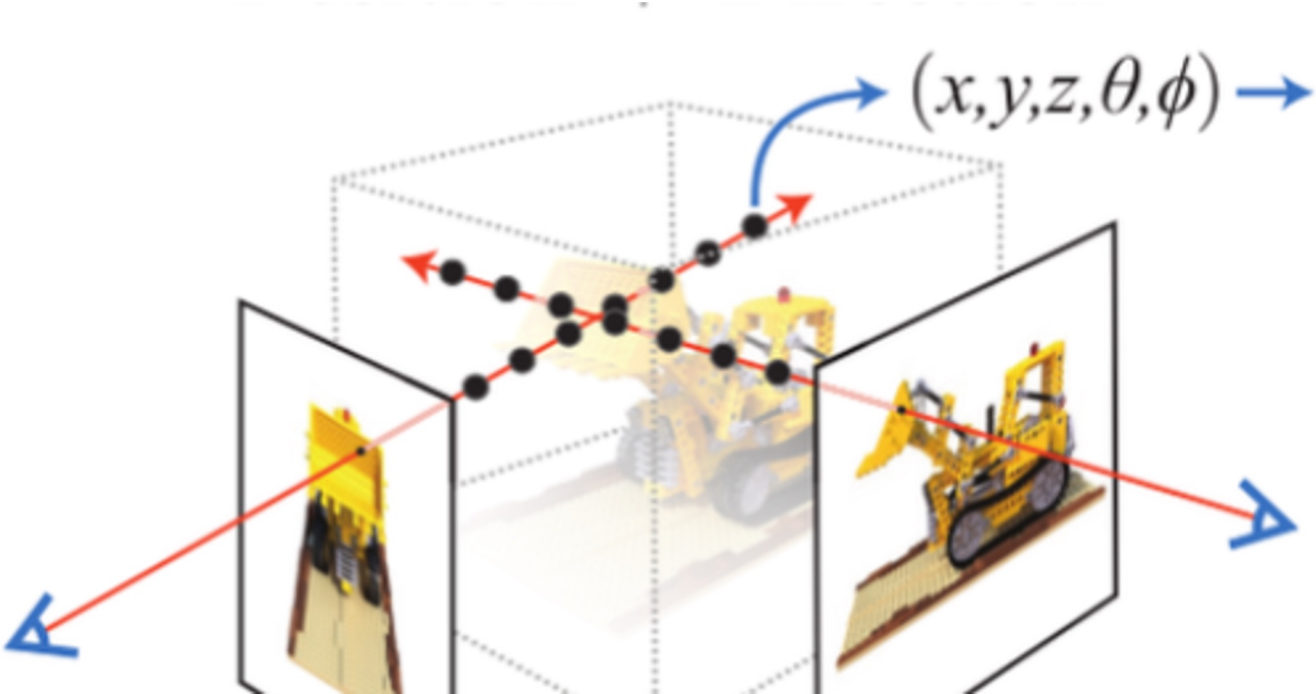
$$F_\Theta : (\mathbf{x}, \mathbf{d}) \rightarrow (\mathbf{c}, \sigma)$$



Density depends only on spatial position, while color depends on both spatial position and viewing angle.

How does NeRF render 2D images?

The NeRF function provides color and density information for a 3D point in space. When imaging a scene with a camera, a pixel in the resulting 2D image corresponds to a continuous set of spatial points along a ray originating from the camera.



Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

Color of the ray

Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

The probability of the situation that "the ray doesn't hit any particle." More maths! Check the previous lectures :)

Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

Volume density

Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

Color of the point

Volume rendering

The volume density can be interpreted as the probability of termination when a ray goes across a infinitely small particle at the spatial point. We can integral over these points to calculate the color of the ray.

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

Viewing direction

Volume rendering

Problem: In practice, it becomes computationally expensive and impractical for complex 3D scenes due to the high dimensionality of the integration, because we need to render all rays that pass through each pixel of the desired virtual camera.

Solution: Use Deterministic Quadrature (a numerical integration method). In other words, we can uniformly sample N points in the region where the integral needs to be evaluated for approximate calculation.

Volume rendering

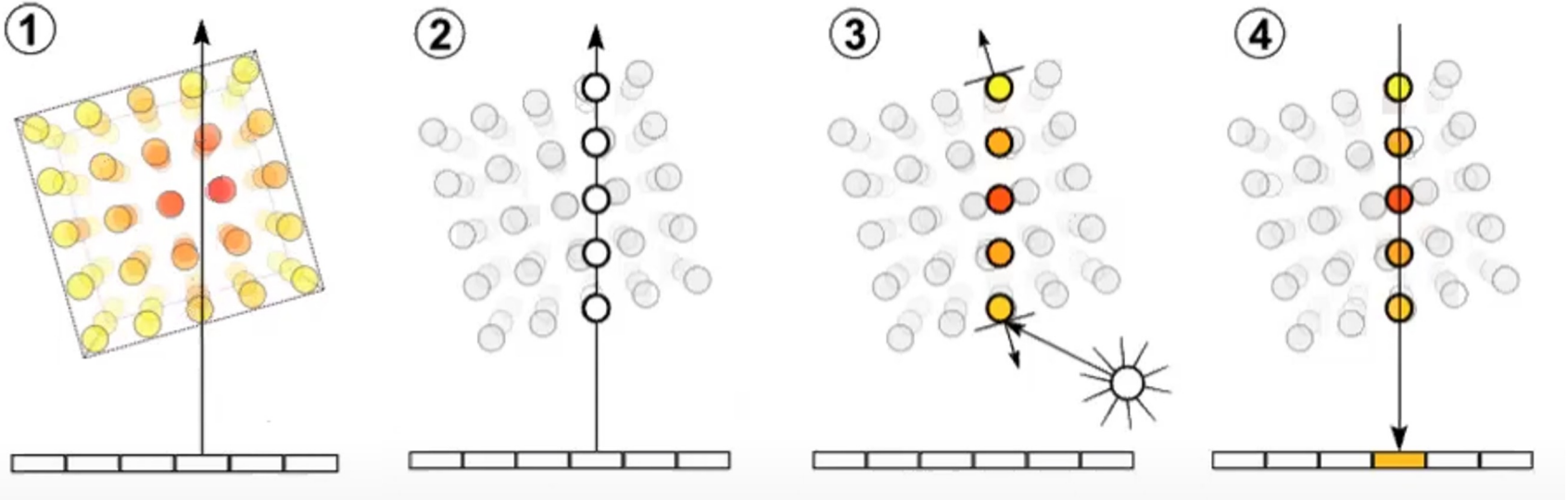
Problem: In practice, it becomes computationally expensive and impractical for complex 3D scenes due to the high dimensionality of the integration, because we need to render all rays that pass through each pixel of the desired virtual camera.

Solution: Use Deterministic Quadrature (a numerical integration method). In other words, we can uniformly sample N points in the region where the integral needs to be evaluated for approximate calculation.

Another Problem: Deterministic Quadrature would limit the resolution of the scene representation because it would only sample at a fixed, discrete set of locations, potentially missing fine details.

Stratified Sampling

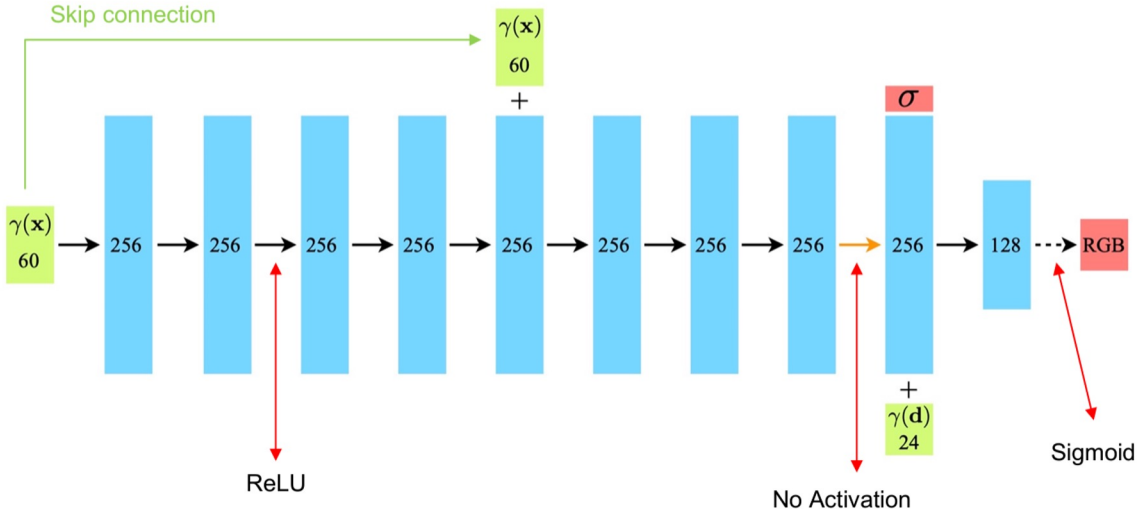
1. Divide the region along the ray that needs to be integrated into N segments.
2. Divide each segment into smaller subintervals and select points uniformly within them.
3. Combine the samples to estimate the integral over the entire range.



Positional Encoding

Mapping to a higher dimensional space using high frequency functions before passing them to the network enables better fitting of data that contains high frequency variation.

$$\gamma(p) = \left(\sin\left(2^0\pi p\right), \cos\left(2^0\pi p\right), \cdots, \sin\left(2^{L-1}\pi p\right), \cos\left(2^{L-1}\pi p\right)\right)$$



Positional Encoding

Mapping to a higher dimensional space using high frequency functions before passing them to the network enables better fitting of data that contains high frequency variation.

$$\gamma(p) = \left(\sin\left(2^0\pi p\right), \cos\left(2^0\pi p\right), \cdots, \sin\left(2^{L-1}\pi p\right), \cos\left(2^{L-1}\pi p\right)\right)$$

Difference from Transformers:

Transformers use it for a different goal of providing the discrete positions of tokens in a sequence as input to an architecture that does not contain any notion of order.

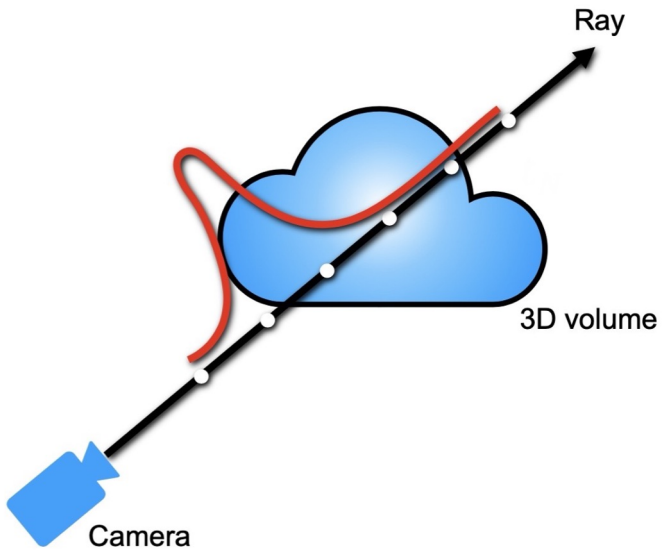
Hierarchical Volume Sampling

Notice that the rendering process of NeRF is computationally intensive, and for each ray, we must sample many points. But in fact, most areas on a ray are empty areas or occluded areas, which doesn't need to be accounted for the output color.

Solution: "Coarse to Fine" networks

Allocate samples proportionally to their expected effect on the final rendering:

1. Sample a set of locations with stratified sampling and evaluate a "coarse" network along a ray to produce a probability density function (PDF) along the ray.
2. Evaluate a "fine" network on the same locations + another a set of locations sampled from the PDF.
3. Compute the final rendered color of the ray using both sets of samples.



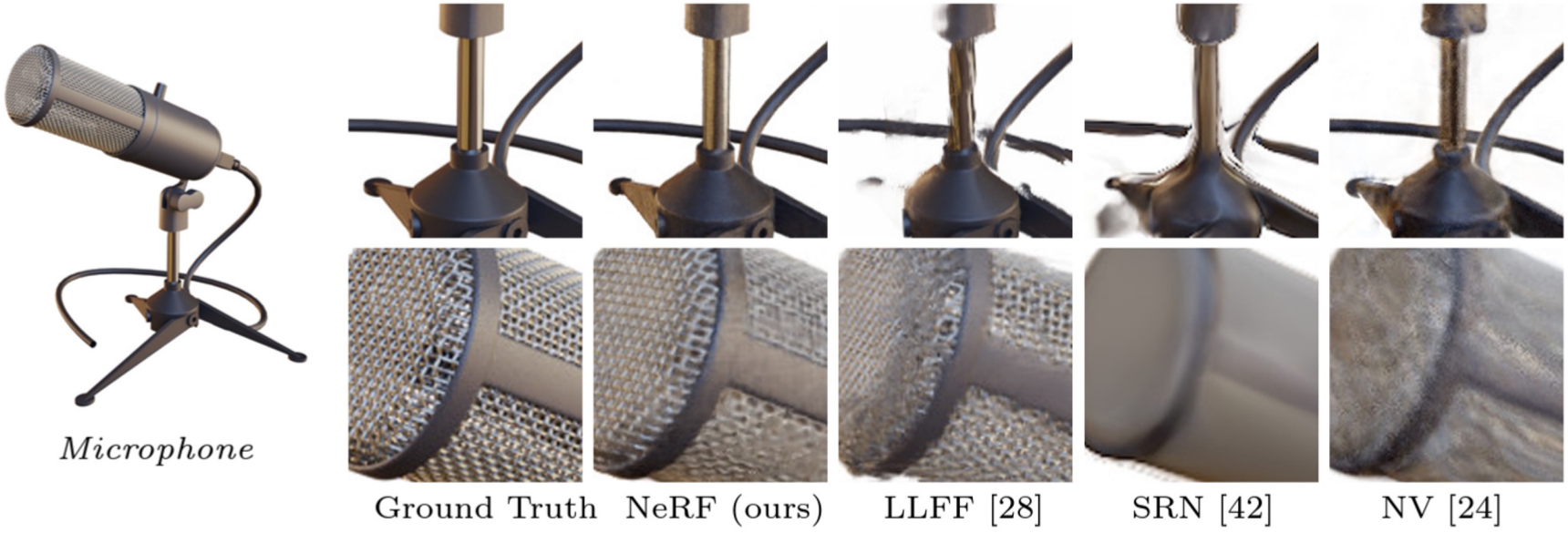
Loss Function

The loss function is defined as the total squared error between the rendered and true pixel colors for both the coarse and fine renderings.

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right]$$

Result

Result



This method is able to recover fine details in both geometry and appearance, such as the Microphone's shiny stand and mesh grille.

Conclusion

NeRF is a novel view synthesis method for reconstructing detailed 3D scenes from a collection of 2D images taken from different viewpoints.

Key idea:
Use MLP to model the volumetric scene as a continuous function that maps 3D points (5D coordinates) to RGB colors and densities.

Key components:
- Volume rendering with stratified sampling
- Positional encoding
- Hierarchical Volume sampling

Future Directions:
- More efficient optimization
- Interpretation of why some modes can fail

Questions?

Piazza Question

How can NeRF be applied in CT imaging/biomedical imaging field?

1. Difficulty of preserving geometric and topological structures in biomedical image segmentation.
2. Difficulty of obtaining high-quality surfaces.
3. High memory requirements for organ segmentation on high resolution scans.
4. Limitation of medical data such as:
   - Capturing the data as discrete grids.
   - Labeling noise.
   - Incomplete borders or occlusions.
   - Lack of large datasets which is limiting the deep learning experiments

Source:
https://collab.dvb.bayern/display/TUMdlma/NeRF+Applications+in+Medical+Imaging#:~:text=Difficulty%20of%20preserving%20geometric%20and,high%20level%20detail%20surface%20reconstruction.

Quiz

What are the inputs and outputs of NeRF?

Quiz

What are the inputs and outputs of NeRF?

Inputs:
A 5D coordinate consisting of spatial points and their corresponding viewing direction.

Outputs:
The directional emitted color and volume density of those points.

Quiz

What strategies did the authors used to improve the resolution of the rendering view?

Quiz

What strategies did the authors used to improve the resolution of the rendering view?

Quality improvement:

1. Stratified sampling
2. Positional encoding

Efficiency improvement:
Hierarchical Volume Sampling

Thank you!

References

Mildenhall, Ben, et al. "Nerf: Representing scenes as neural radiance fields for view synthesis." Communications of the ACM 65.1 (2021): 99-106.

Kim, Youwang. "[Seminar] NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis." YouTube, 14 Jan 2021, www.youtube.com/watch?v=FSG5bCkNWWo.