NeRF

Representing Scenes as Neural Radiance Fields for View Synthesis

Ben Mildenhall* UC Berkeley
Pratul P. Srinivasan* UC Berkeley
Matthew Tancik* UC Berkeley
Jonathan T. Barron Google Research
Ravi Ramamoorthi UC San Diego
Ren Ng UC Berkeley

Jathurshan Pradeepkumar

NeRF

Representing Scenes as Neural Radiance Fields for View Synthesis

ECCV 2020 Oral - Best Paper Honorable Mention

Ben Mildenhall* — UC Berkeley
Pratul P. Srinivasan* — UC Berkeley
Matthew Tancik* — UC Berkeley
Jonathan T. Barron — Google Research
Ravi Ramamoorthi — UC San Diego
Ren Ng — UC Berkeley

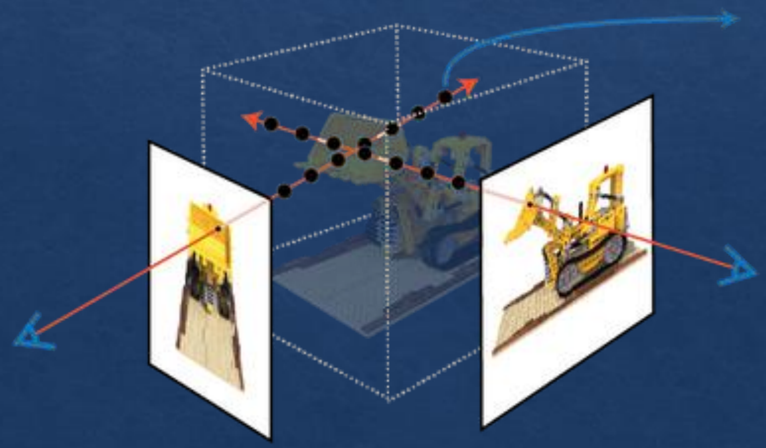
*Denotes Equal Contribution

Paper **</Code>** **Data**

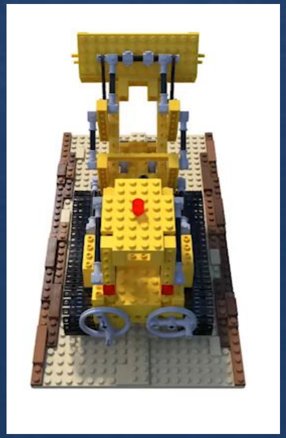
Ref: https://www.matthewtancik.com/nerf

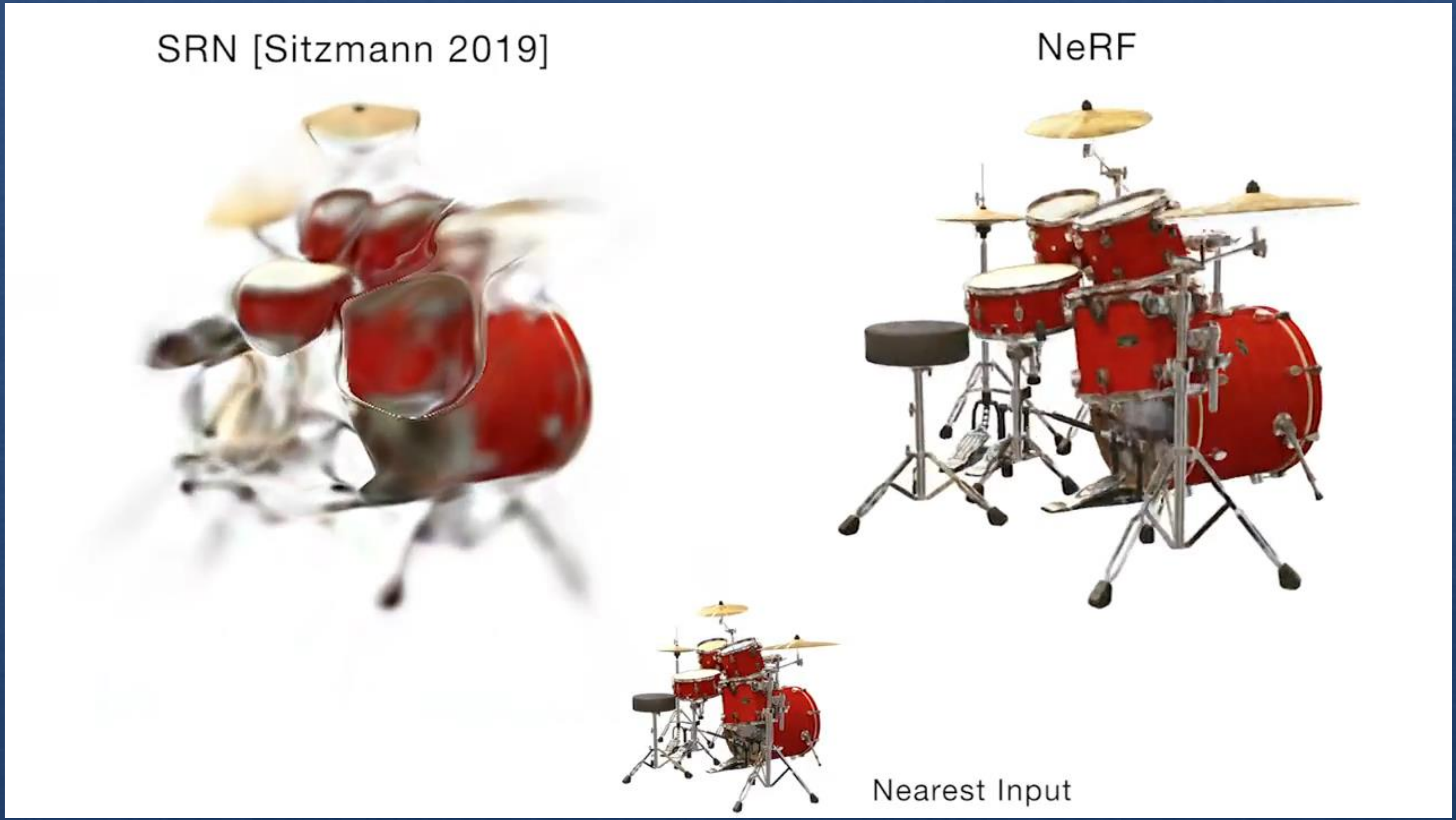
Task

- 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.



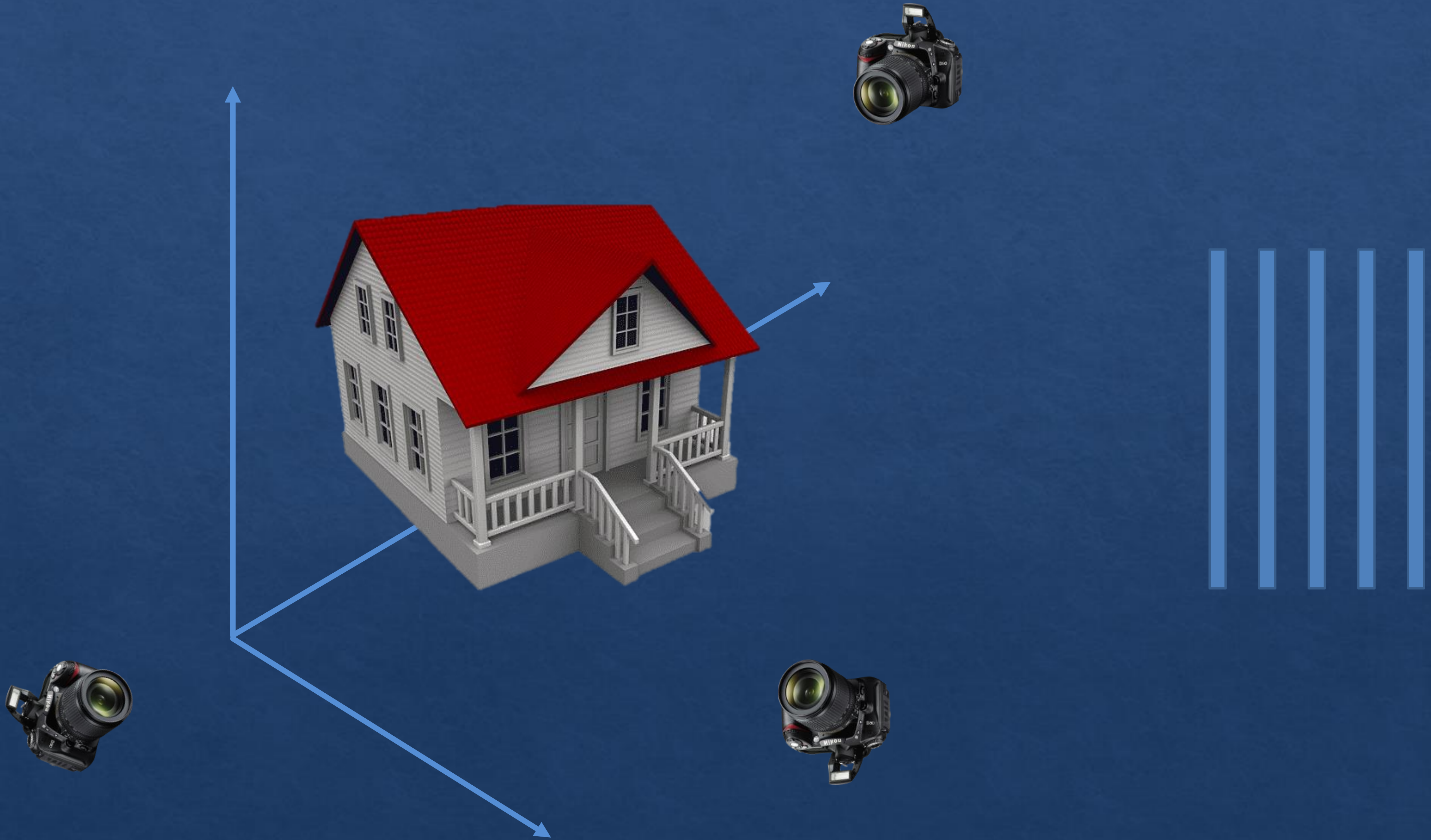
$$\longrightarrow F_{\Theta}(.) \longrightarrow$$



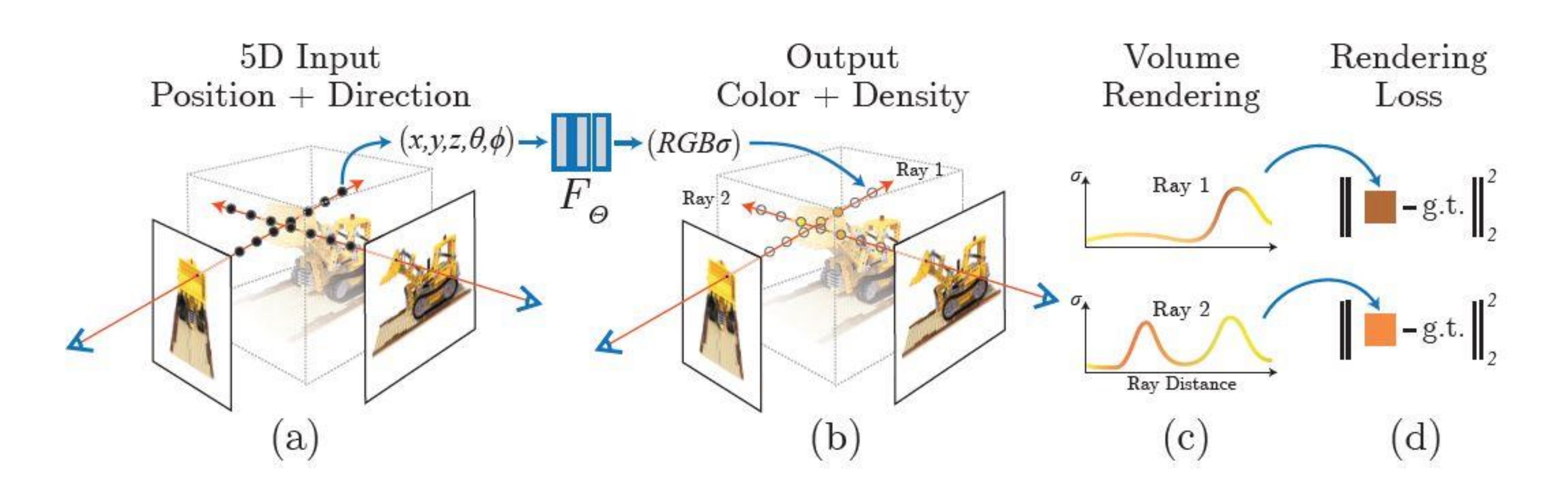


Ref: “NeRF: Neural Radiance Fields”, link: https://youtu.be/JuH79E8rdKc

How?



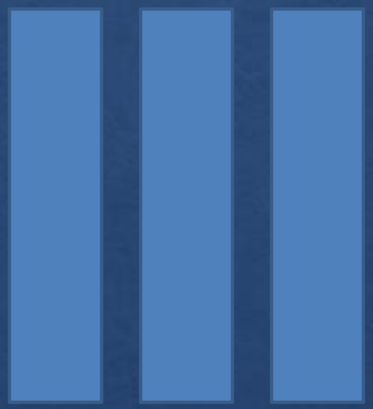
How?



5D Input
Position + Direction
(x,y,z,θ,ϕ)
F_Θ
$(RGB\sigma)$
Output
Color + Density
Ray 1
Ray 2
Volume
Rendering
Rendering
Loss
σ
Ray 1
$\| \blacksquare - \text{g.t.} \|_2^2$
σ
Ray 2
Ray Distance
$\| \blacksquare - \text{g.t.} \|_2^2$
(a)
(b)
(c)
(d)

Representing a scene as a continuous 5D function

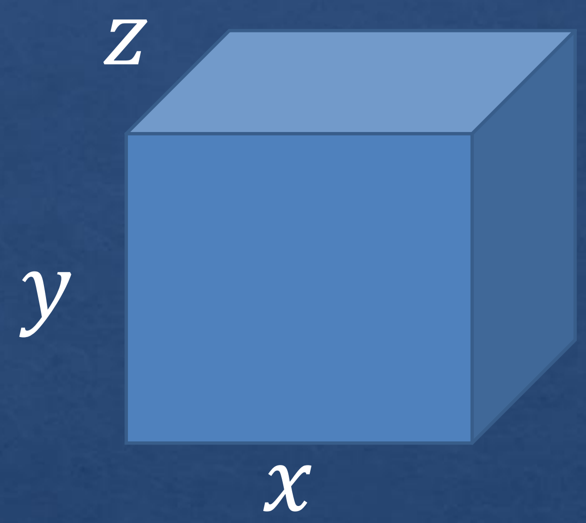
$$(\underbrace{x, y, z}_{\text{Spatial Location}}, \underbrace{\theta, \phi}_{\text{Viewing Direction}}) \rightarrow F_{\Theta} \rightarrow (\underbrace{r, g, b}_{\text{Color}}, \underbrace{\sigma}_{\text{Density}})$$



F_{Θ}

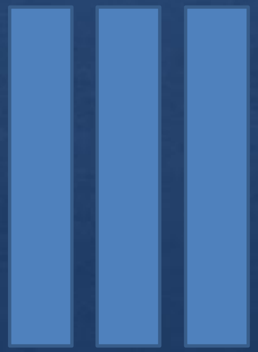
Fully connected neural network,
9 layers 256 channels

Representing a scene as a continuous 5D function



≥15GB

for one “Realistic Synthetic" scene



5MB

$$\underbrace{(x, y, z}_{\text{Spatial Location}}, \underbrace{\theta, \phi)}_{\text{Viewing Direction}} \rightarrow F_{\Theta} \rightarrow \underbrace{(r, g, b}_{\text{Color}}, \underbrace{\sigma)}_{\text{Density}}$$

Fully connected neural network,
9 layers 256 channels

View generation using volume rendering with radiance field

$$Rendering\ model\ of\ camera\ ray:\ r(t) = o + td$$

Colors

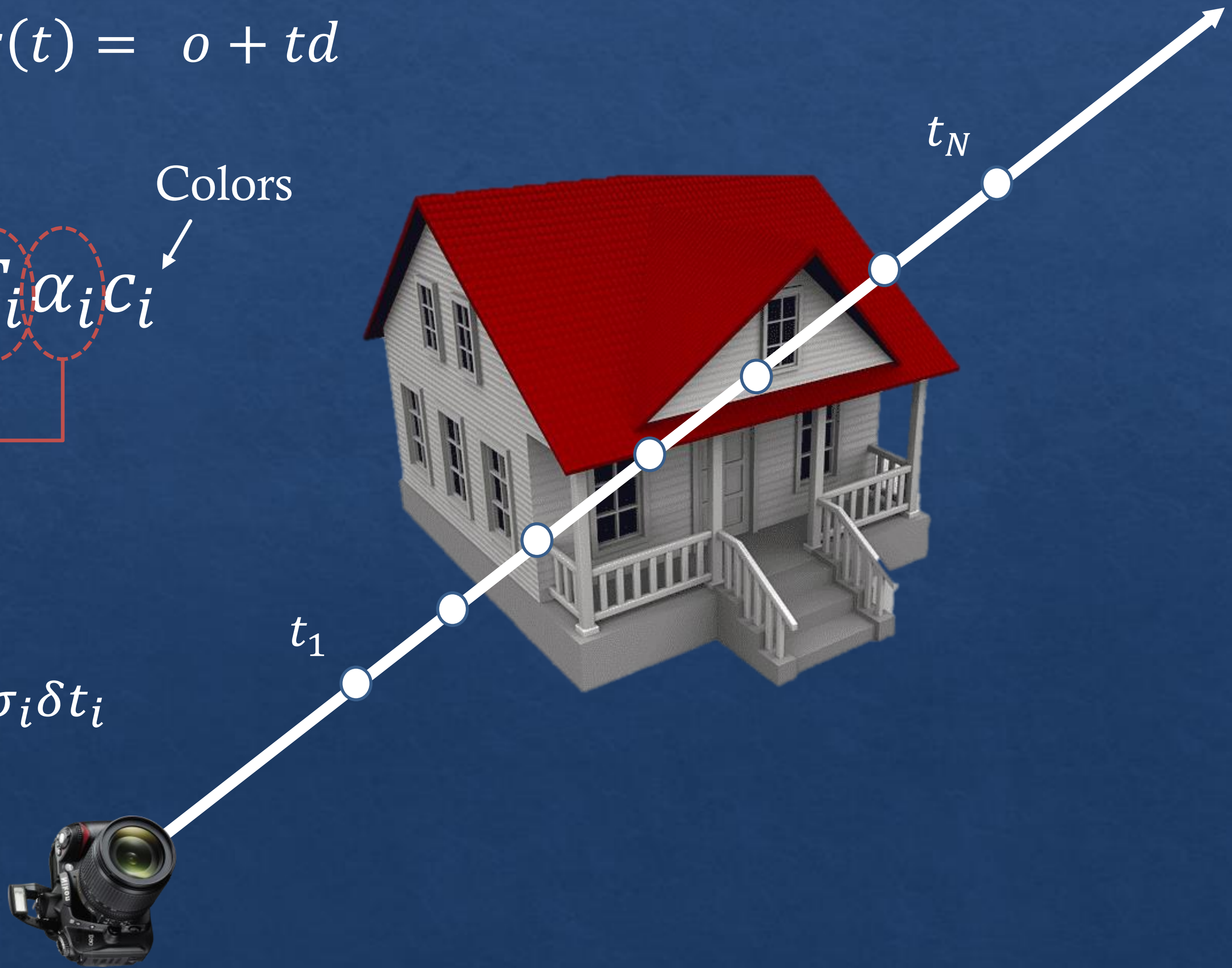
$$Expected\ Color:\ C \approx \sum_{i=1}^{N} T_i \alpha_i c_i$$

Amount of light blocked earlier:

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

Amount of light by ray segment i:

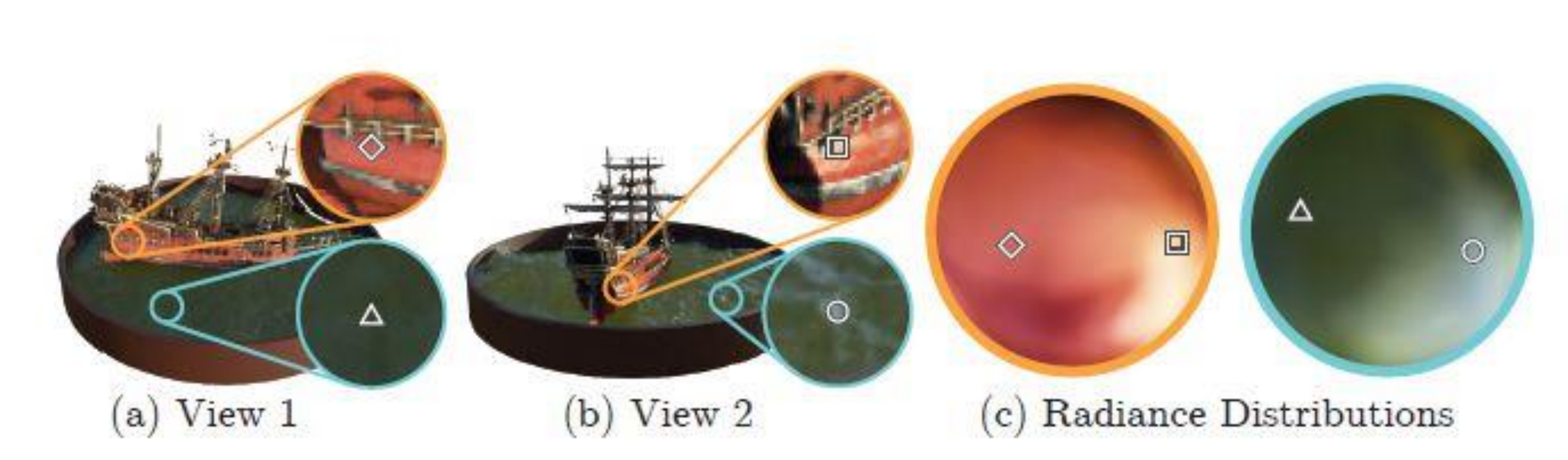
$$\alpha_i = 1 - e^{-\sigma_i \delta t_i}$$



View dependent effects

Directions as input :

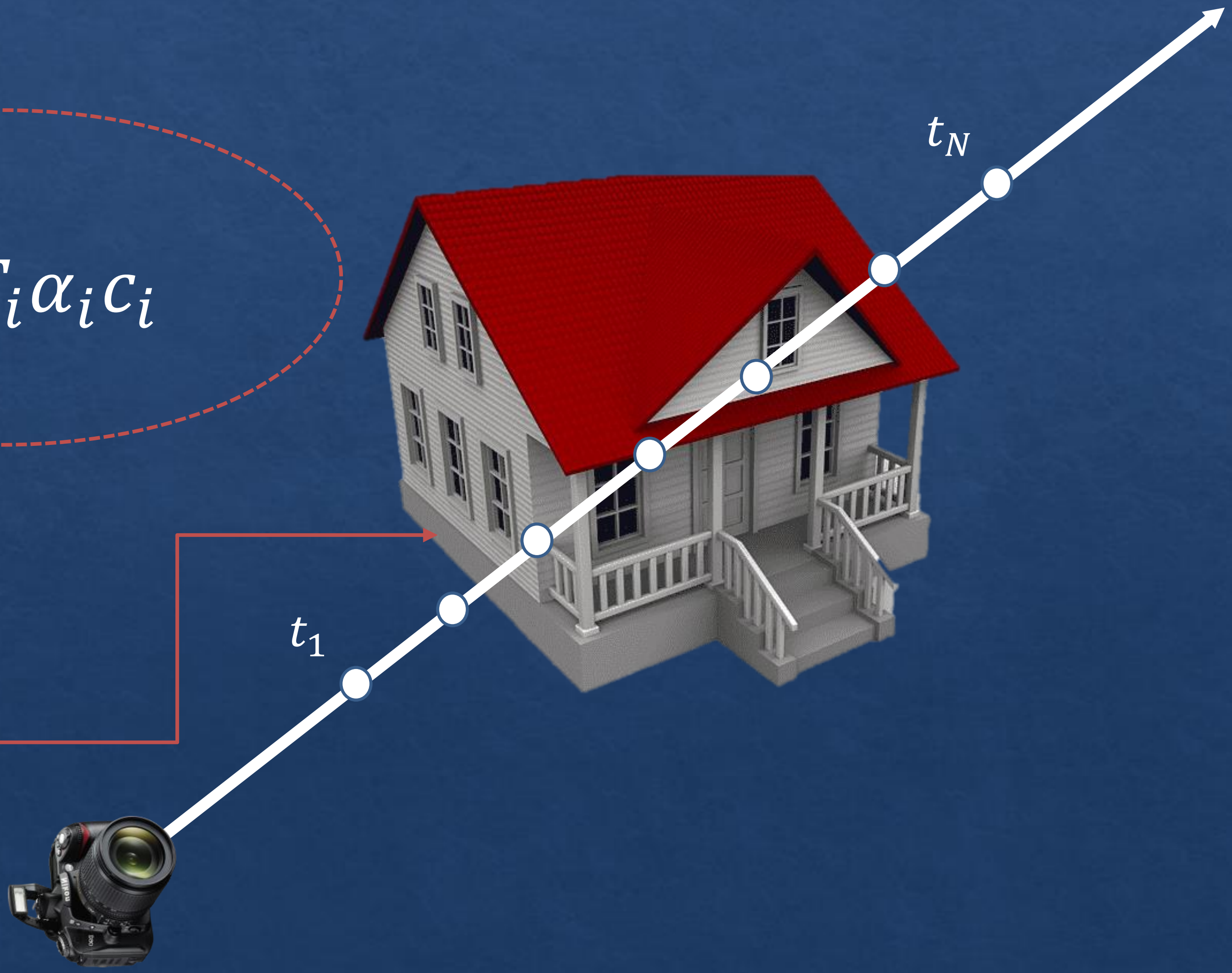
Change (θ,ϕ) to visualize view dependent effects

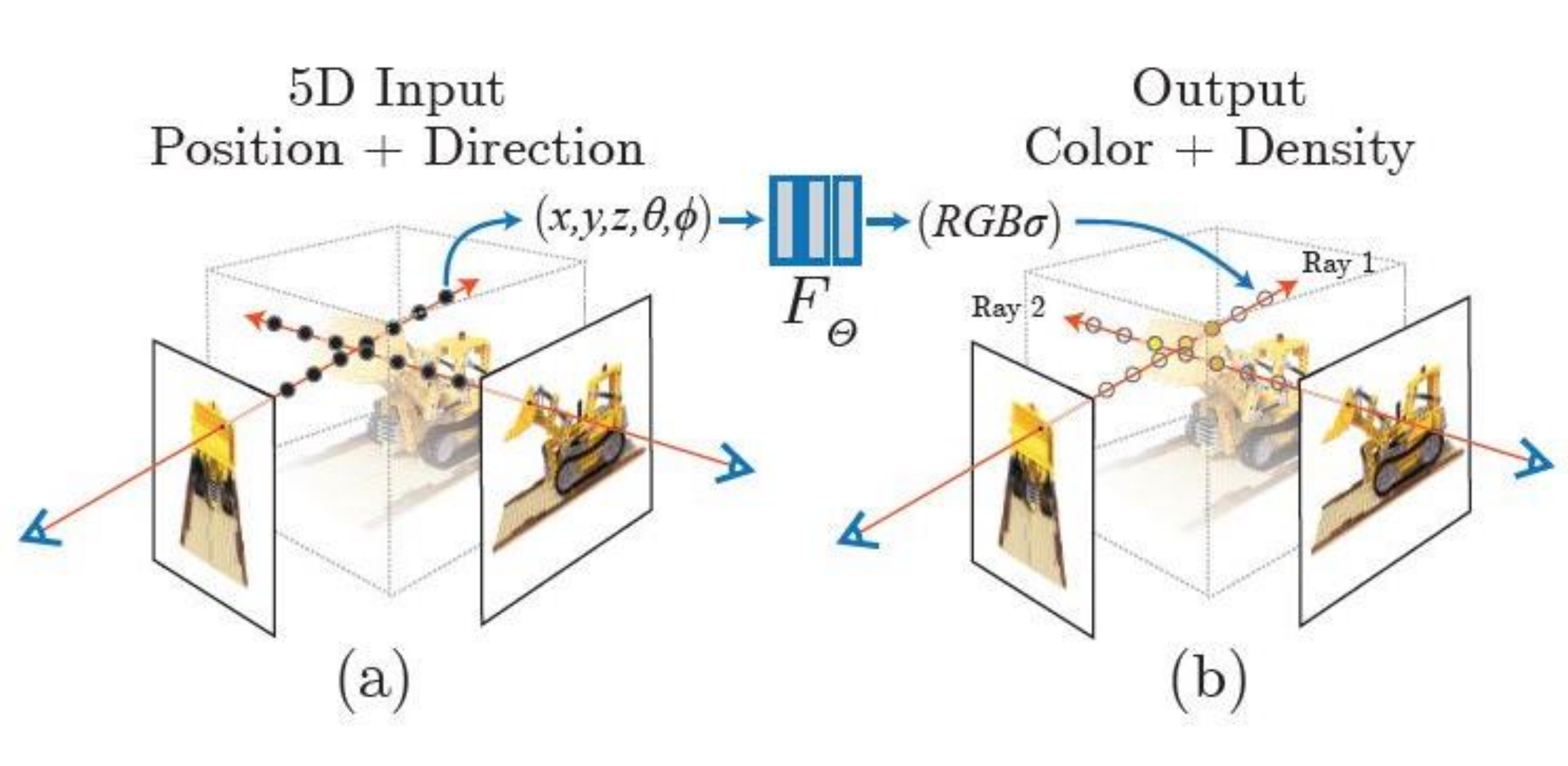


Optimizing using rendering loss

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

Trivially Differentiable
w.r.t color and volume
density





$$min_\theta \sum_i \left\| render^{(i)}(F_\theta) - I_{gt}^{(i)} \right\|^2$$

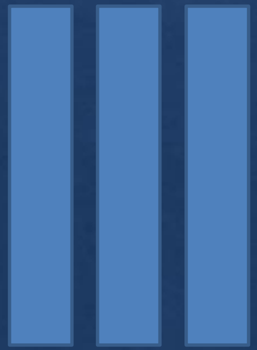
Positional encoding to recover high frequency details

$$(x, \mathrm{d}) \longrightarrow$$

$$\longrightarrow (c, \sigma)$$

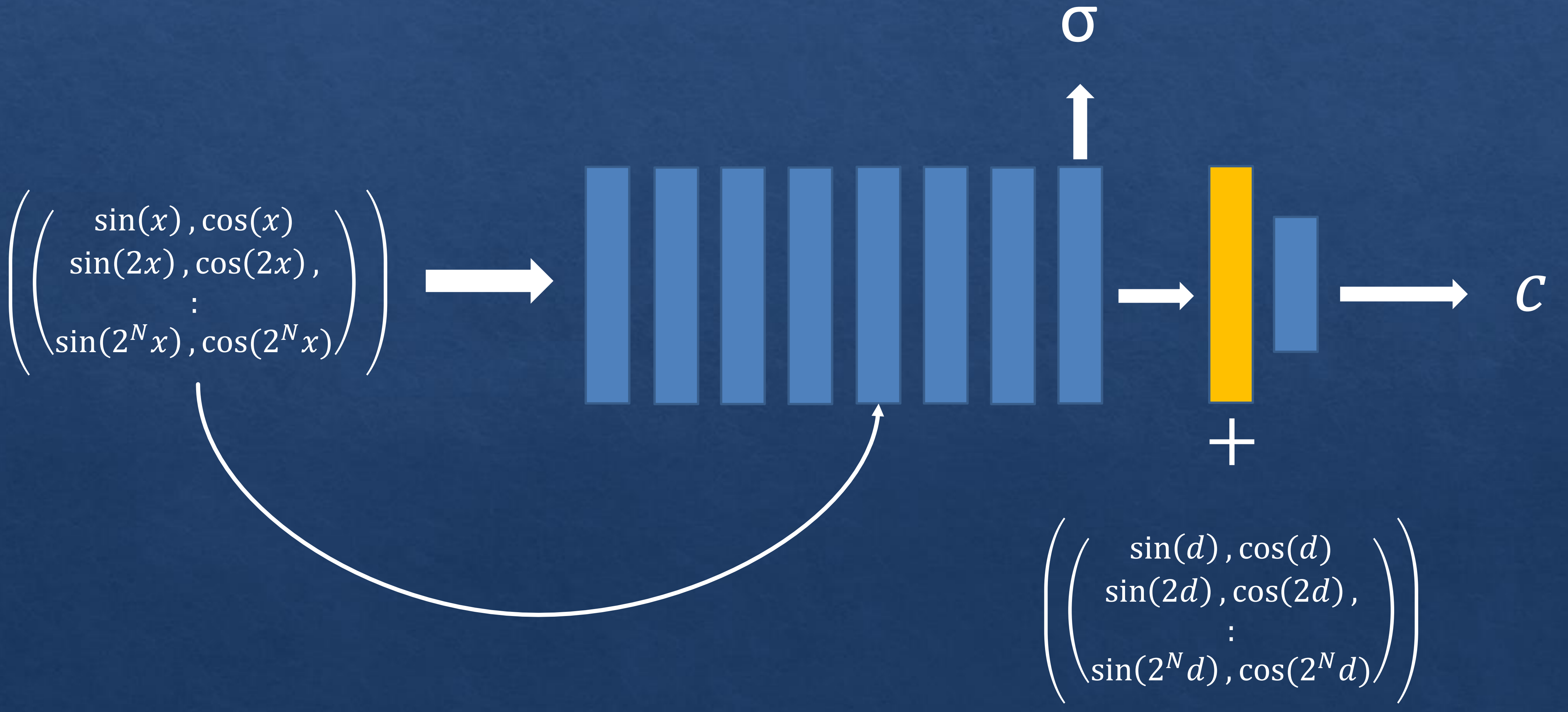
High frequency embedding of input coordinates

$$\left(\begin{pmatrix} \sin(x), \cos(x) \\ \sin(2x), \cos(2x), \\ \vdots \\ \sin(2^N x), \cos(2^N x) \end{pmatrix}, \begin{pmatrix} \sin(d), \cos(d) \\ \sin(2d), \cos(2d), \\ \vdots \\ \sin(2^N d), \cos(2^N d) \end{pmatrix} \right) \longrightarrow$$

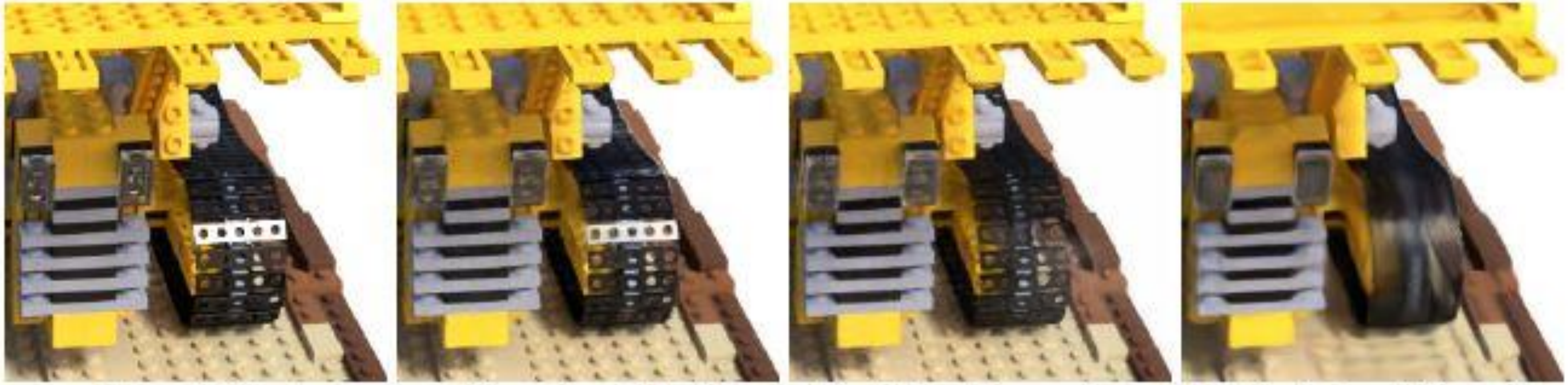


$$\longrightarrow (c, \sigma)$$

Positional encoding to recover high frequency details



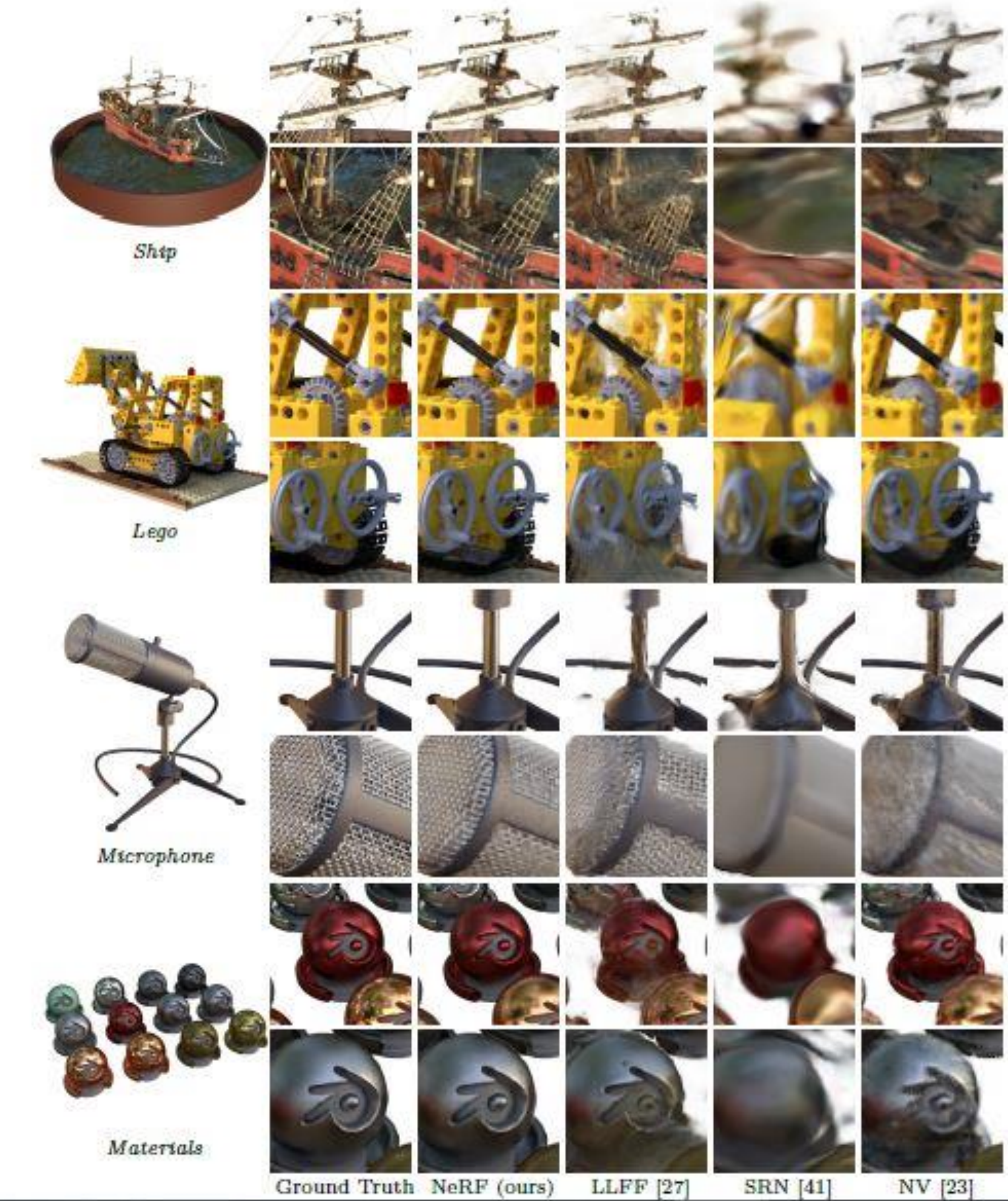
Results



Ground Truth | Complete Model | No View Dependence | No Positional Encoding

Results

Method	Diffuse Synthetic 360° [40]			Realistic Synthetic 360°			Real Forward-Facing [27]		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
SRN [41]	33.20	0.963	0.073	22.26	0.846	0.170	22.84	0.668	0.378
NV [23]	29.62	0.929	0.099	26.05	0.893	0.160	-	-	-
LLFF [27]	34.38	0.985	0.048	24.88	0.911	0.114	24.13	0.798	**0.212**
Ours	**40.15**	**0.991**	**0.023**	**31.01**	**0.947**	**0.081**	**26.50**	**0.811**	0.250



Summary

- Continuous neural network as a volumetric scene representation (5D = position + direction)
- Use volume rendering model to synthesize new views.
- Optimization using rendering loss for one scene (no prior training)
- Apply positional encoding before passing coordinates into network to recover high frequency details.

Reference

- Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. ECCV.

Project website

- Link: https://www.matthewtancik.com/nerf

Useful Youtube videos

- Link 1: "NeRF: Neural Radiance Fields", https://youtu.be/JuH79E8rdKc
- Link 2: "[ECCV 2020] NeRF: Neural Radiance Fields (10 min talk)", https://www.youtube.com/watch?v=LRAqeM8EjOo