Lecture 20
Implicit neural representations
Neural rendering

January 18th 2022

Meirav Galun

Based on

1. The ECCV 2022 Tutorial Neural Volumetric Rendering for Computer Vision
2. In particular, slides by Matt Tancik and Ben Mildenhall

Last time Deep image (implicit) prior

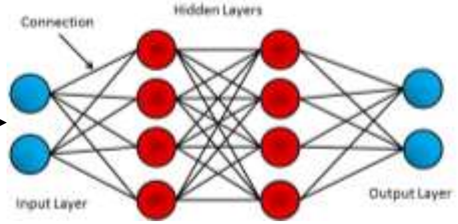
- Constructing an implicit prior by neural network

$$\min_{x} ||d(x) - \hat{x}||$$

s.t. x is an output of CNN

The network weights parametrize the restored image

z →

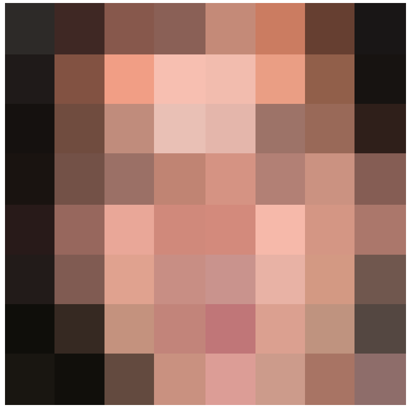


Noise

→ x

→ $||d(x) - \hat{x}||$

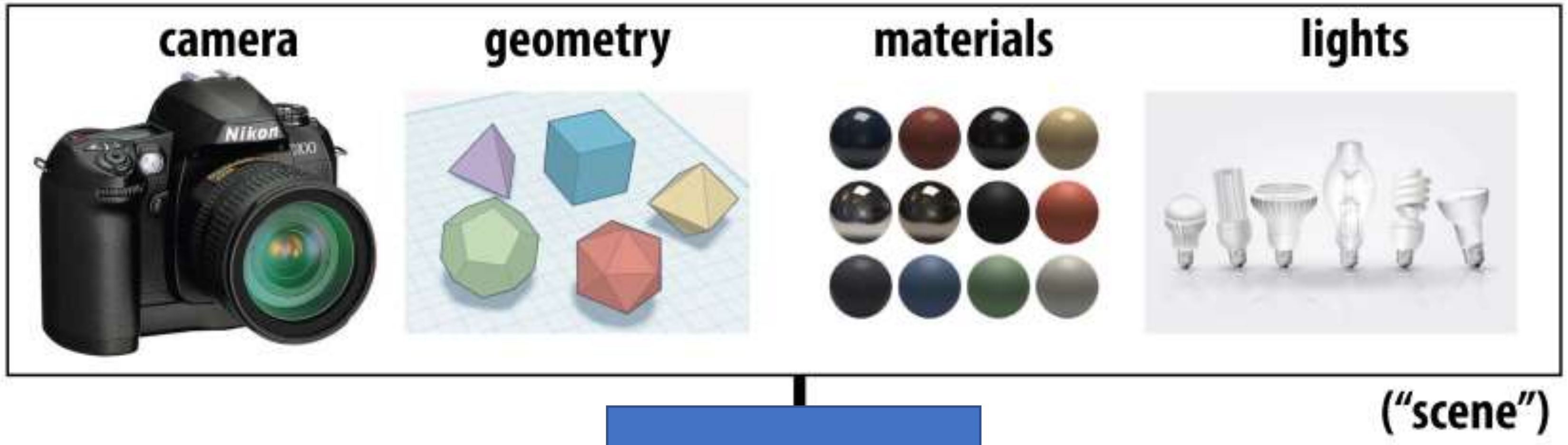
$\hat{x}$



Ulyanov, D., Vedaldi, A., & Lempitsky, V., Deep image prior, CVPR, 2018

Last time computer graphics and rendering

The process of generating a photorealistic image from a 3D model



Very challenging to solve the rendering equation

CMU 15-462/662

Today

- Neural rendering
  (Deep-based computer graphics)

- Implicit neural scene representations
  A network can parametrize
  - Geometry
  - 3D volumes
  - Continuous functions

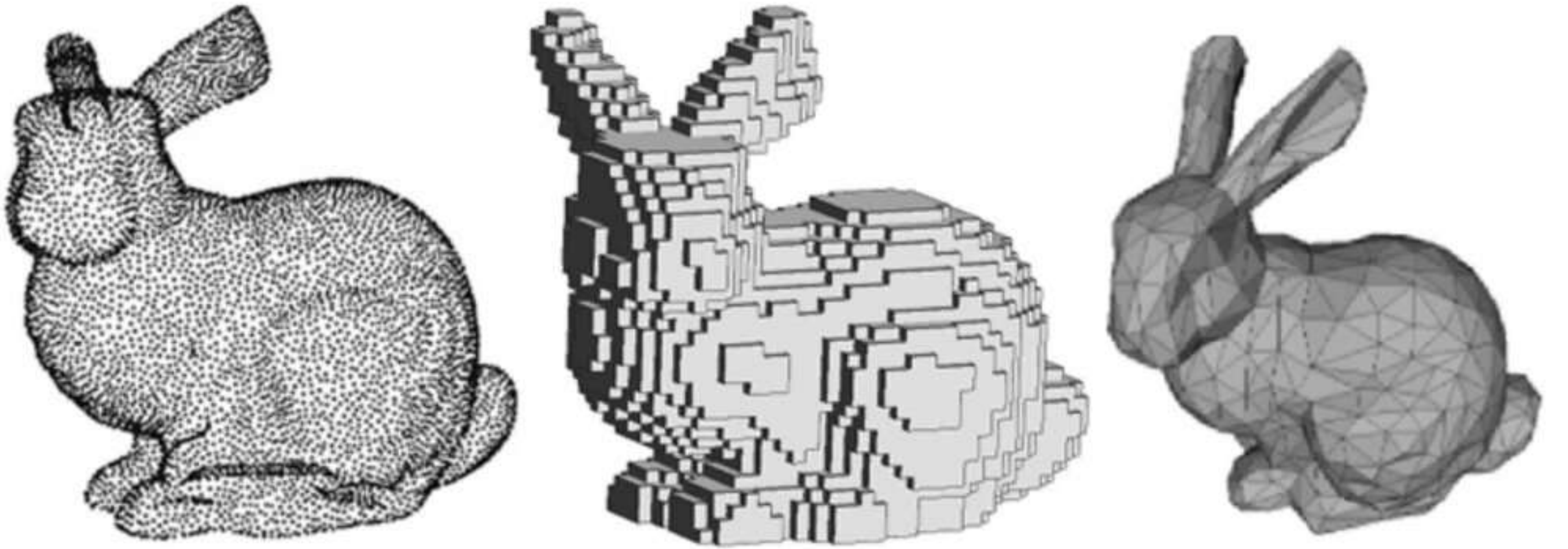
Why not explicit representation?

Geometry

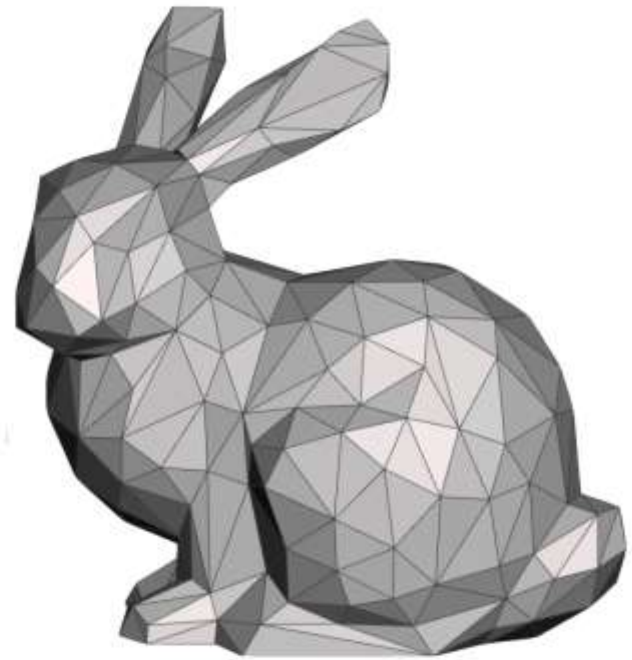
Scene representation

<u>Explicit</u> (discretization of the object geometry)

- triangle (polygon) mesh
- voxels
- point cloud

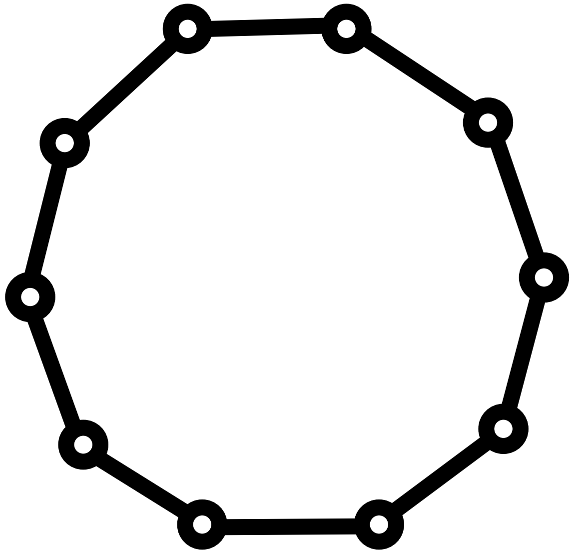


Geometry
Mesh representation

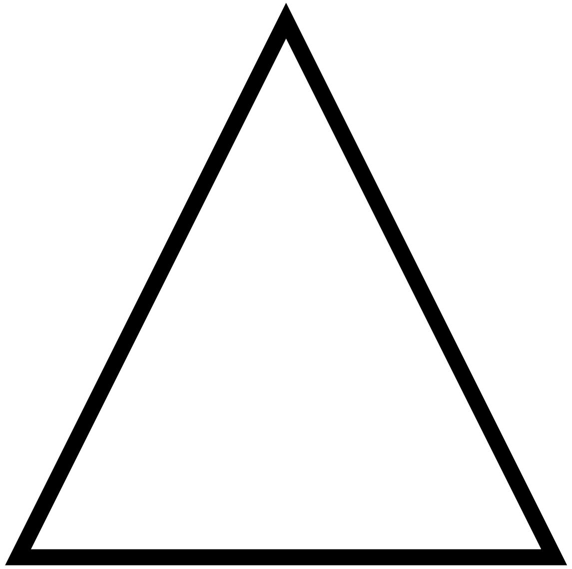


Slides on explicit geometry
by Matt Tancik

Gradient Based Optimization

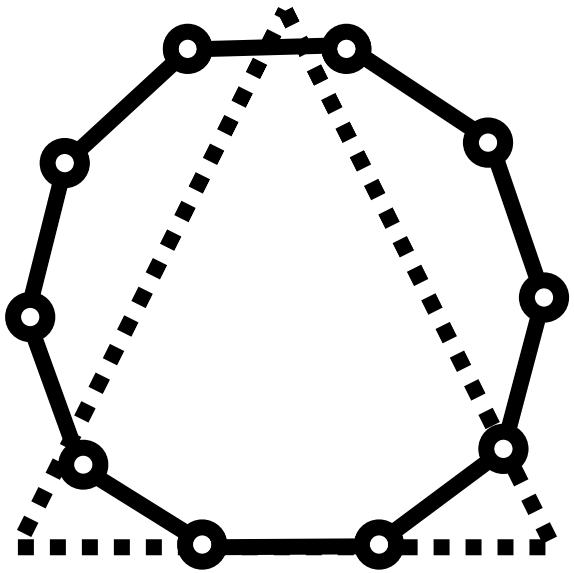


Initial Geometry

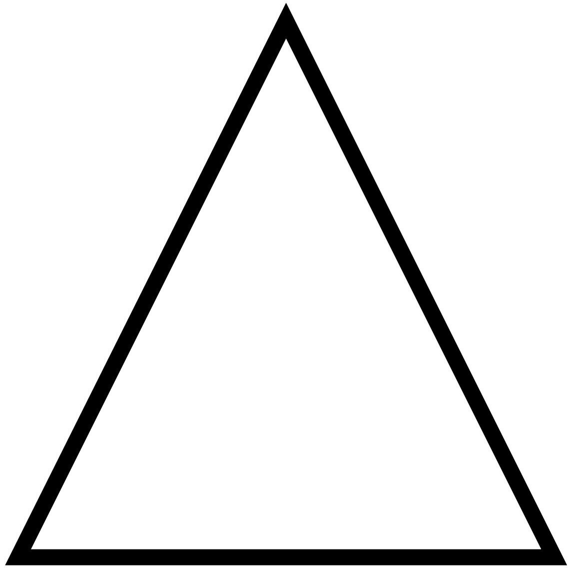


Target Geometry

Gradient Based Optimization

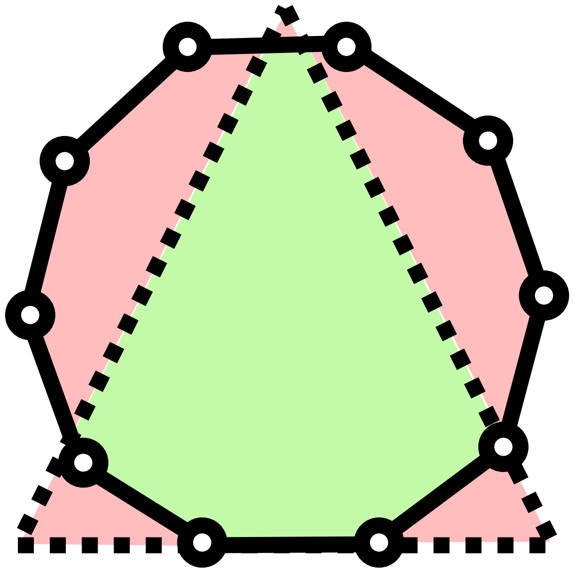


Initial Geometry

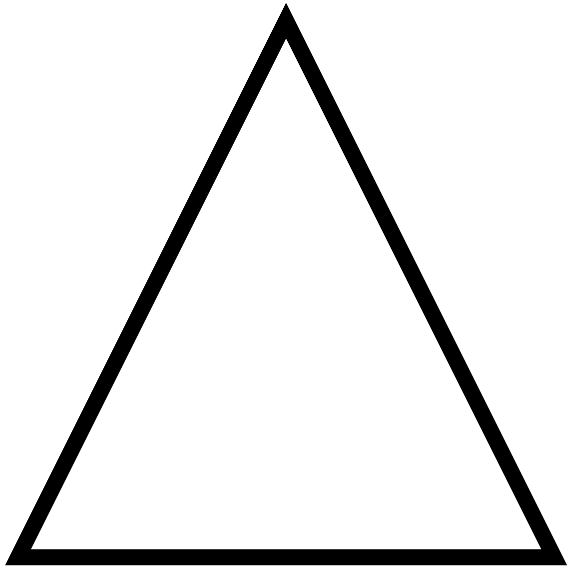


Target Geometry

Gradient Based Optimization

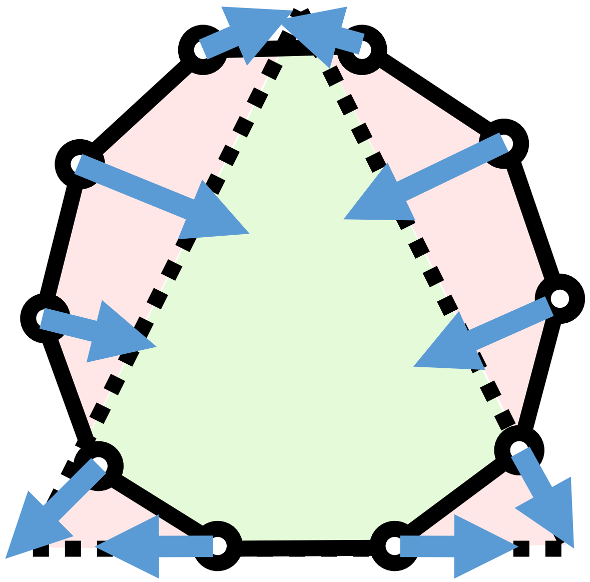


Compute Gradients

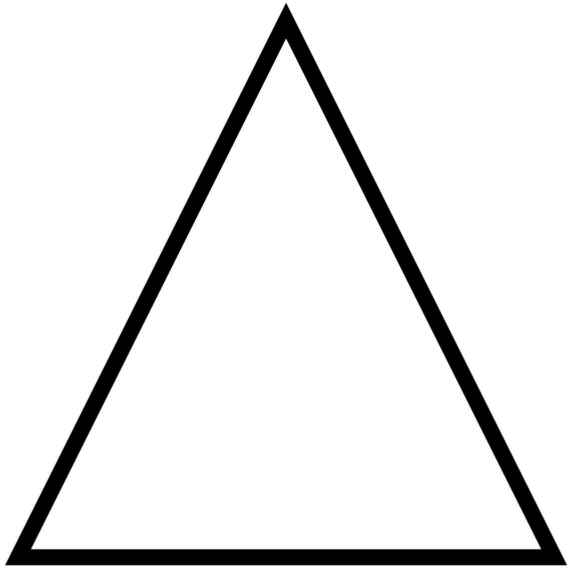


Target Geometry

Gradient Based Optimization

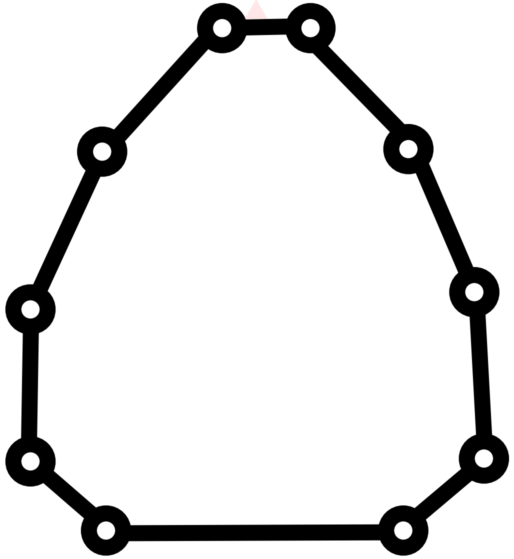


Compute Gradients

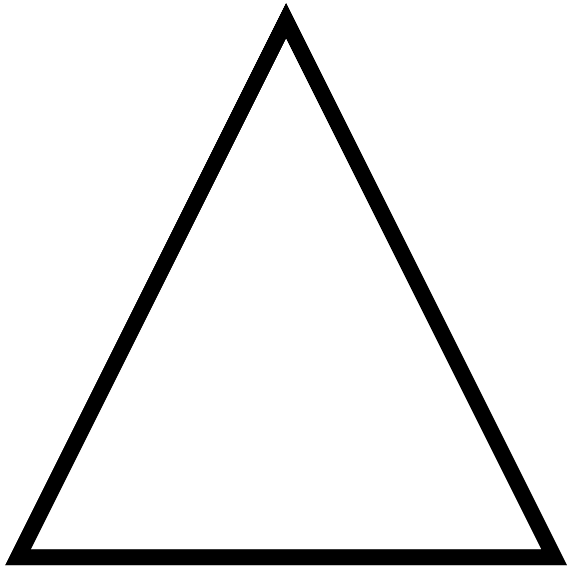


Target Geometry

Gradient Based Optimization

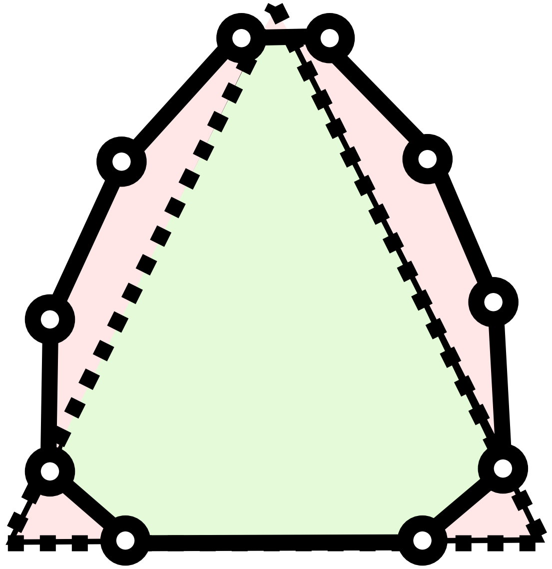


Update positions

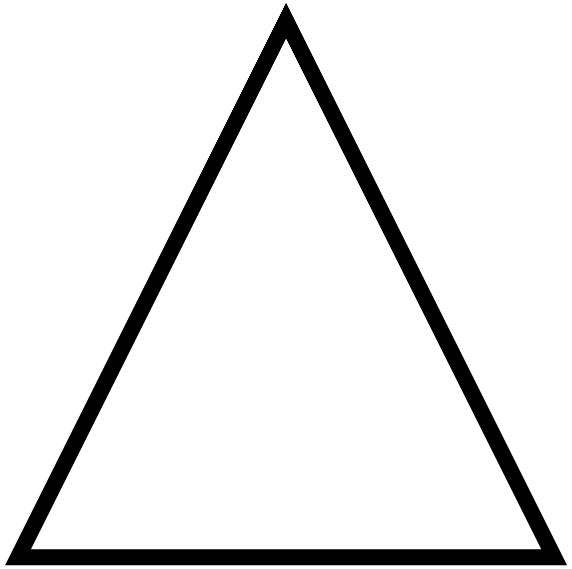


Target Geometry

Gradient Based Optimization

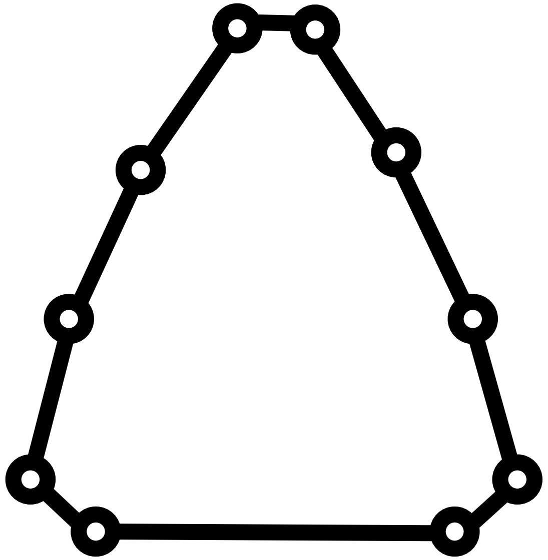


Compute New Error

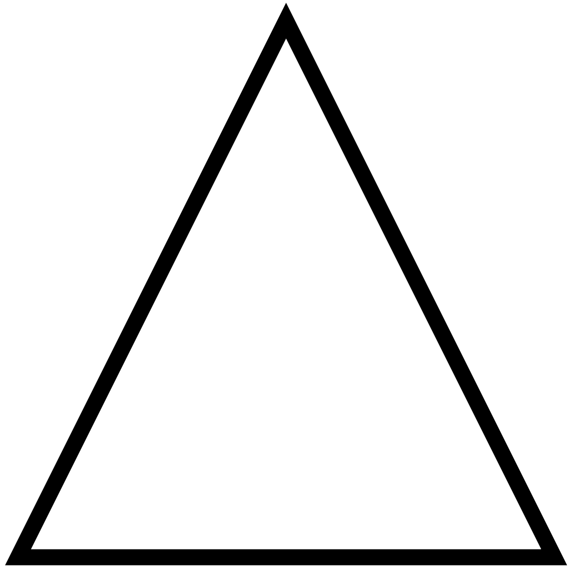


Target Geometry

Gradient Based Optimization

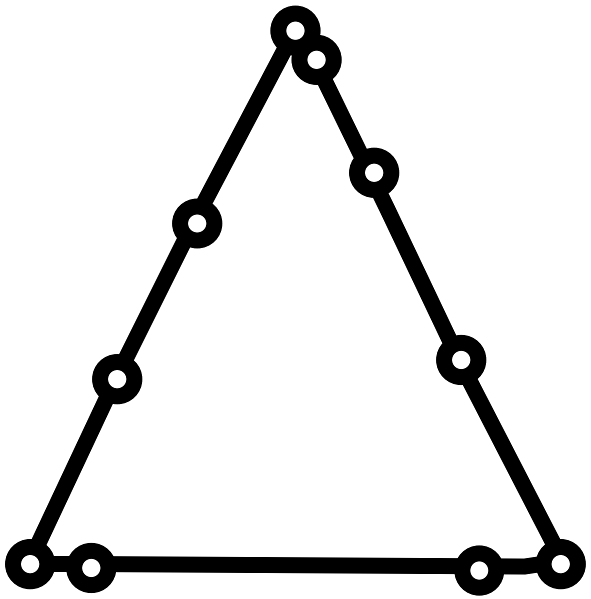


Repeat

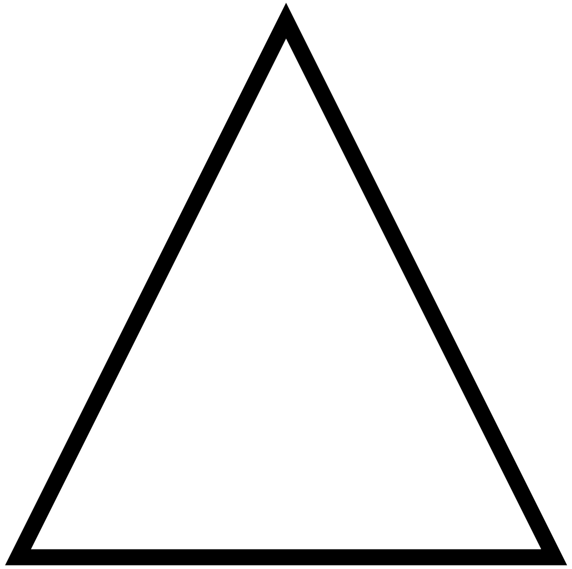


Target Geometry

Gradient Based Optimization

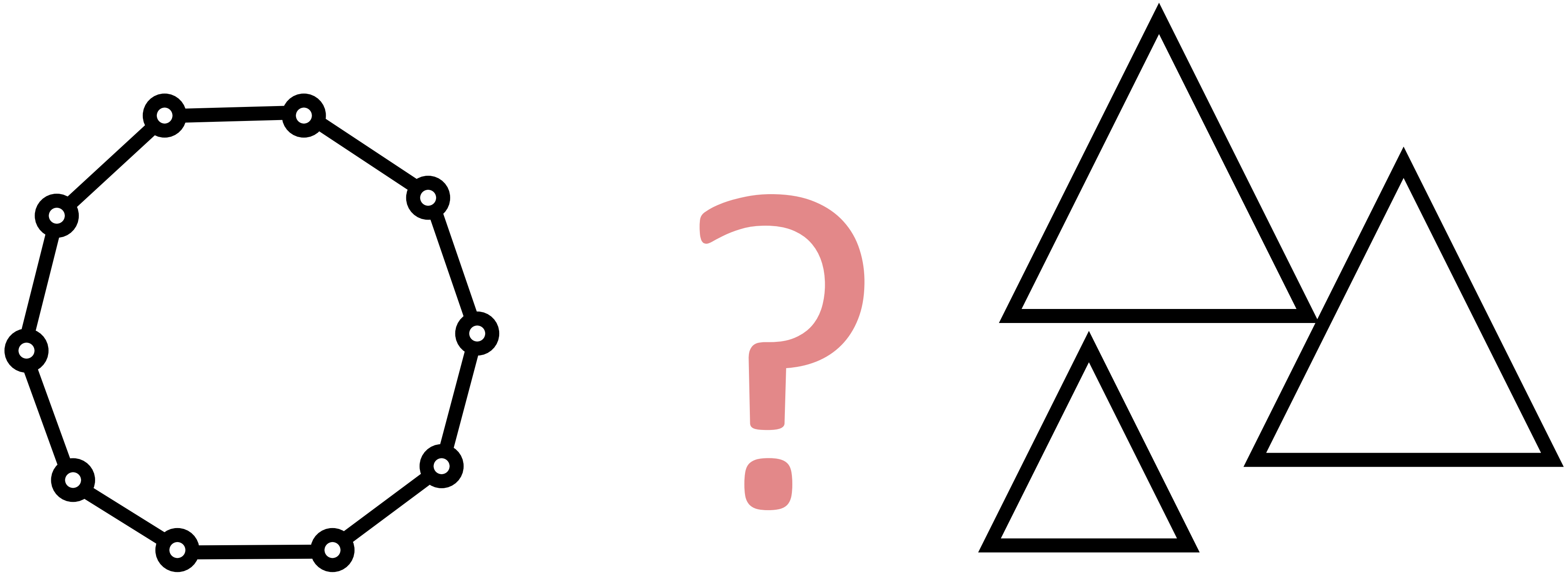


Repeat



Target Geometry

Gradient Based Optimization

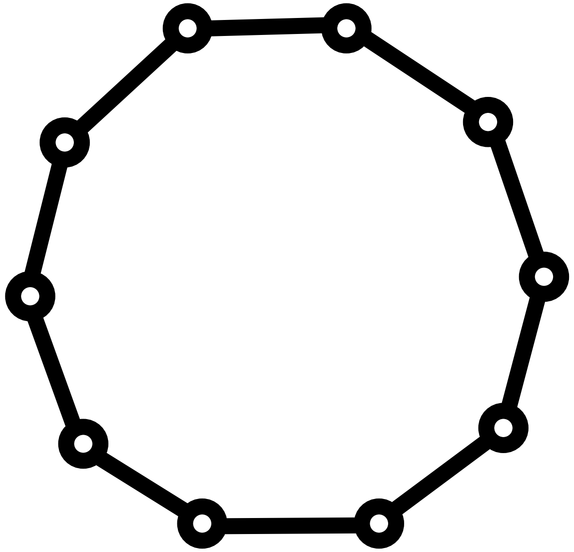


Gradient Based Optimization

Initial Geometry

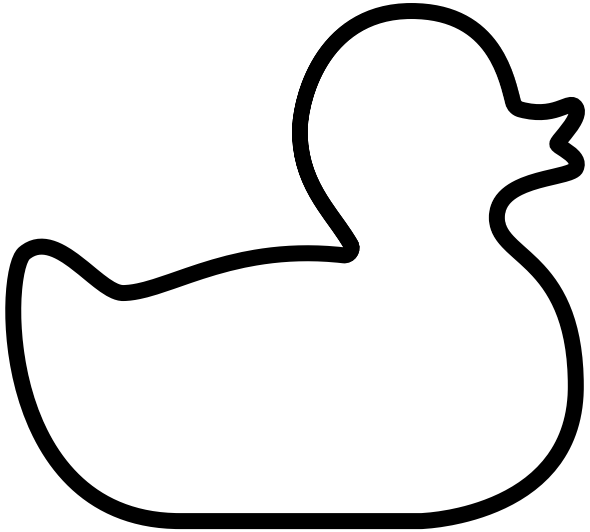
Target Geometry

Gradient Based Optimization



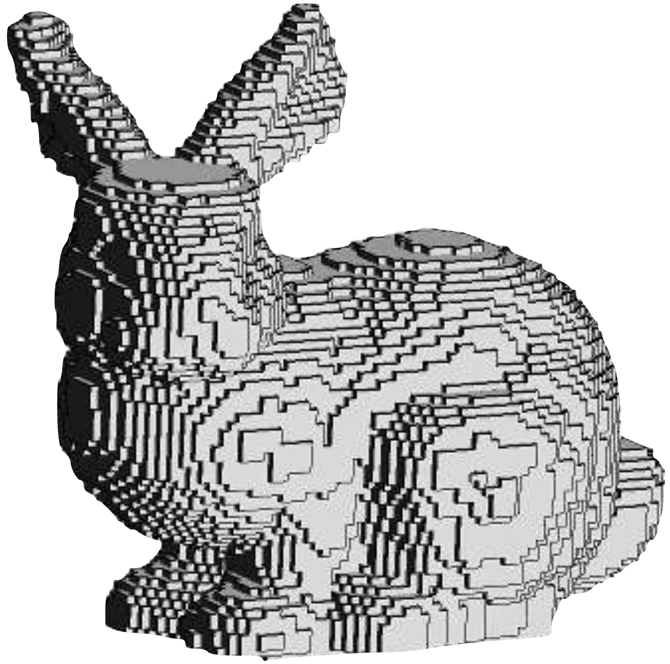
Initial Geometry

?

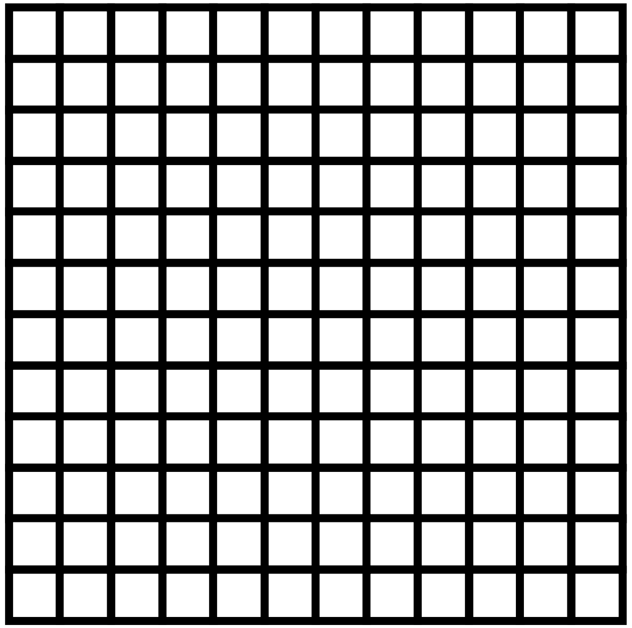


Target Geometry

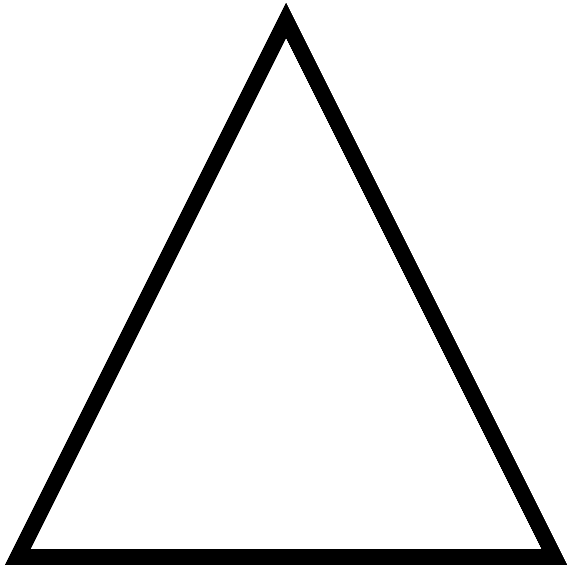
Voxel Representation



Gradient Based Optimization

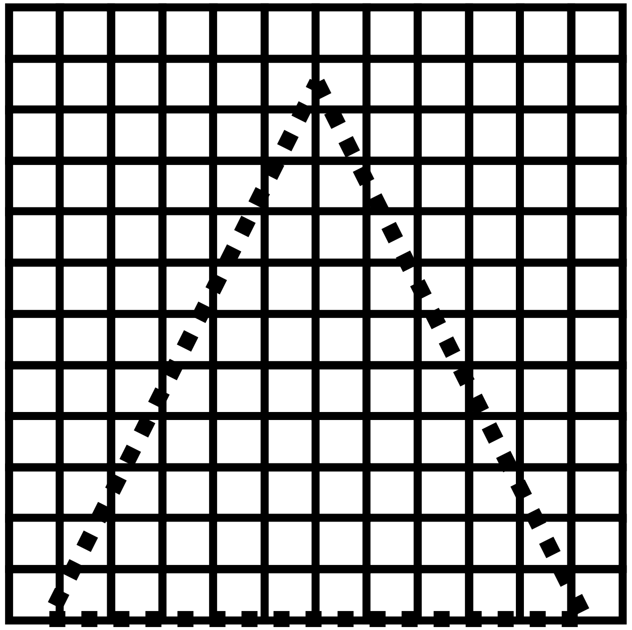


Initialized Grid

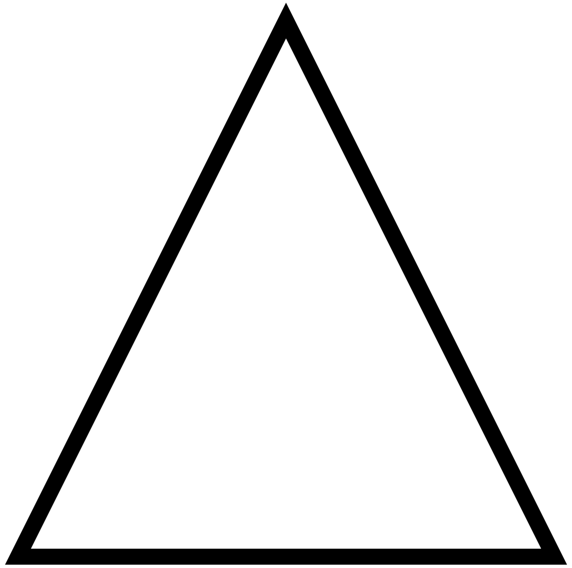


Target Geometry

Gradient Based Optimization

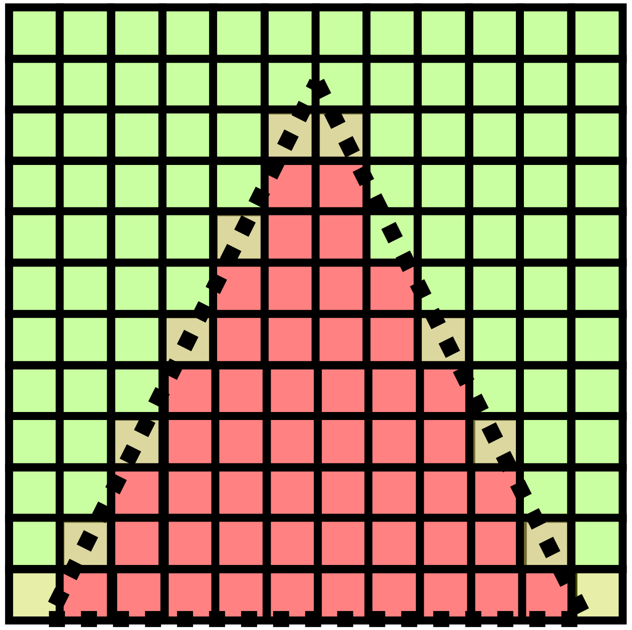


Initialized Grid

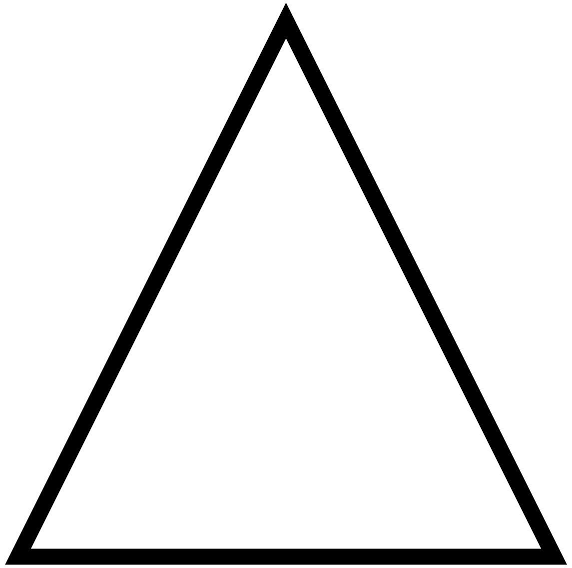


Target Geometry

Gradient Based Optimization

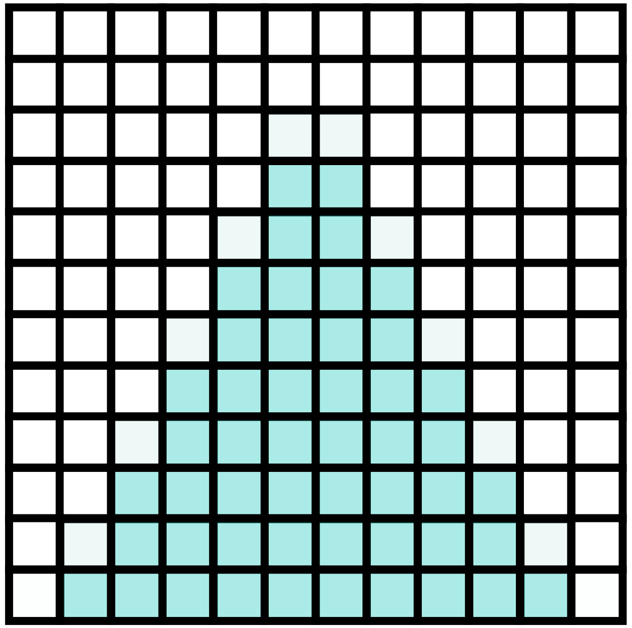


Loss

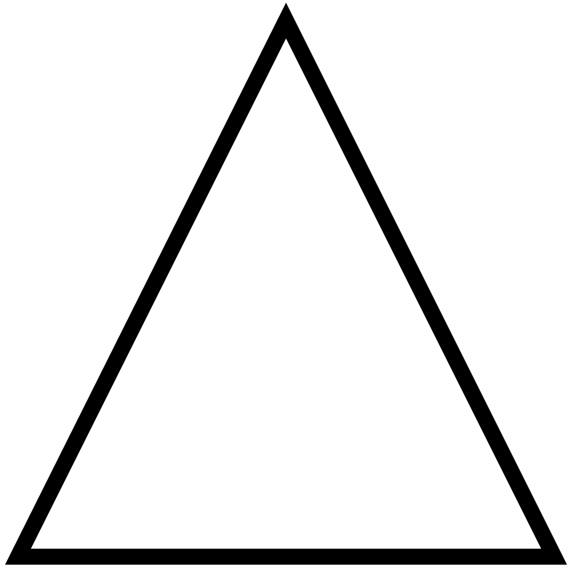


Target Geometry

Gradient Based Optimization

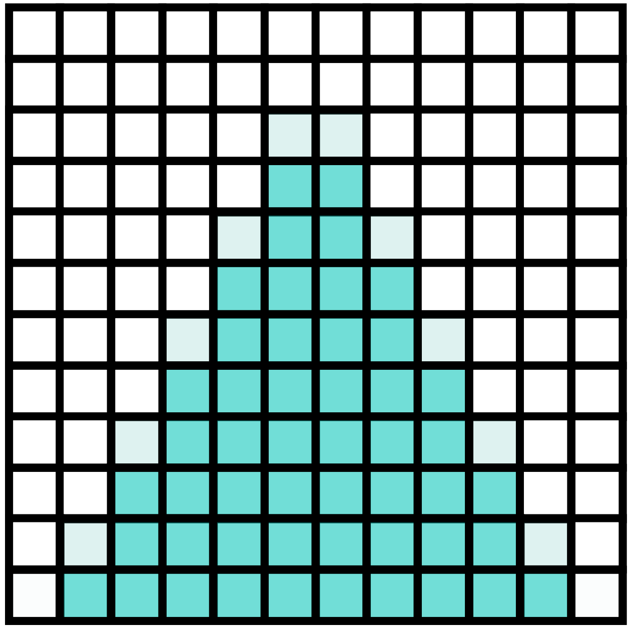


Gradient Step

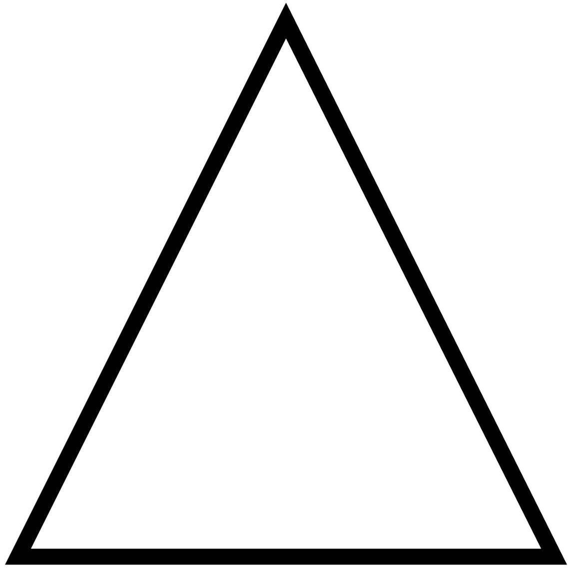


Target Geometry

Gradient Based Optimization

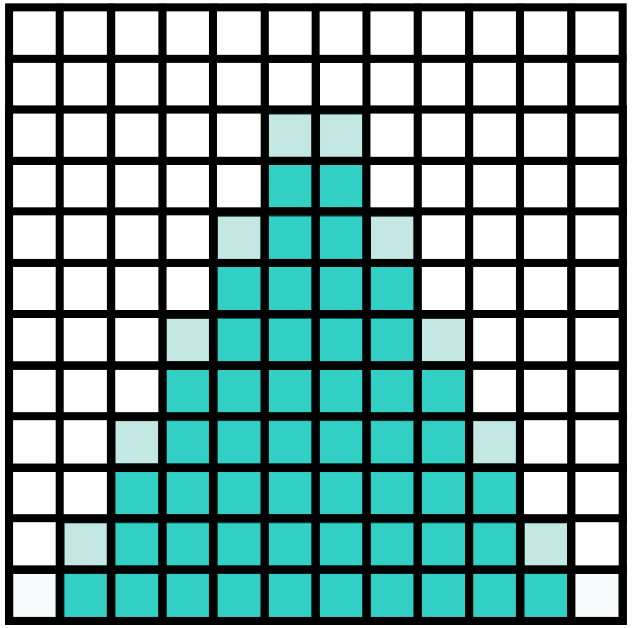


Repeat

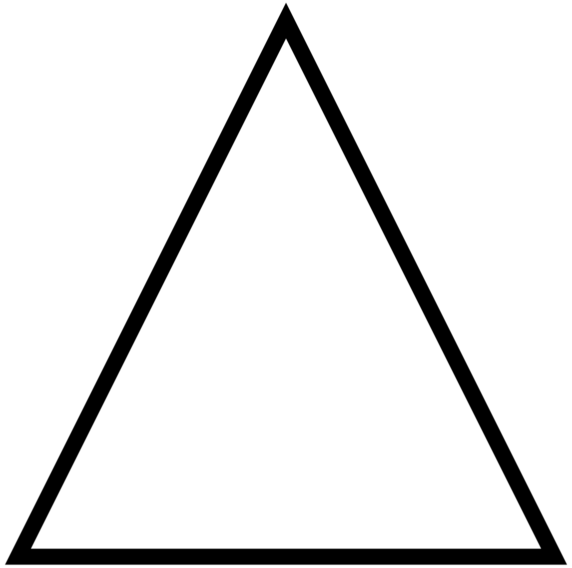


Target Geometry

Gradient Based Optimization

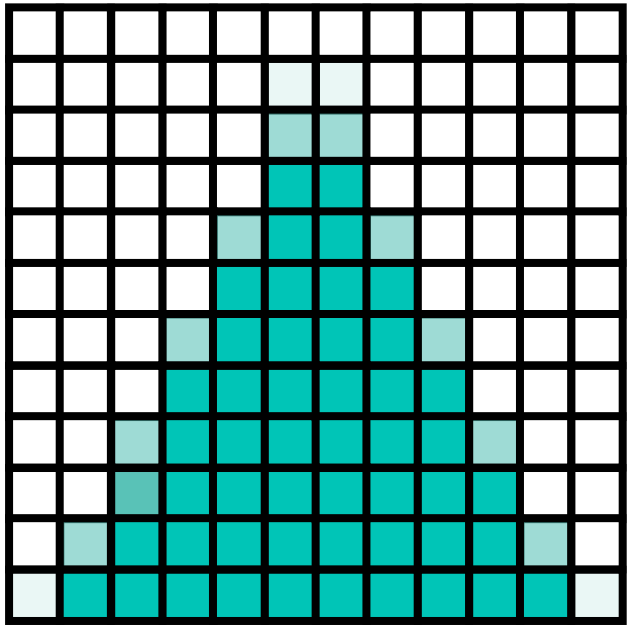


Repeat

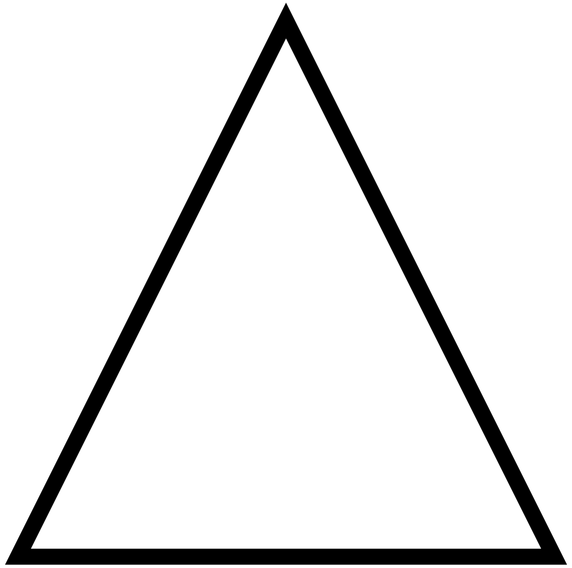


Target Geometry

Gradient Based Optimization

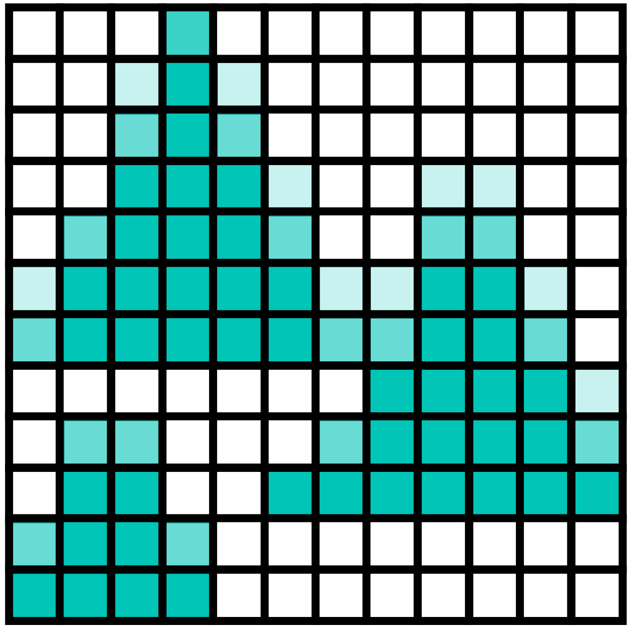


Repeat

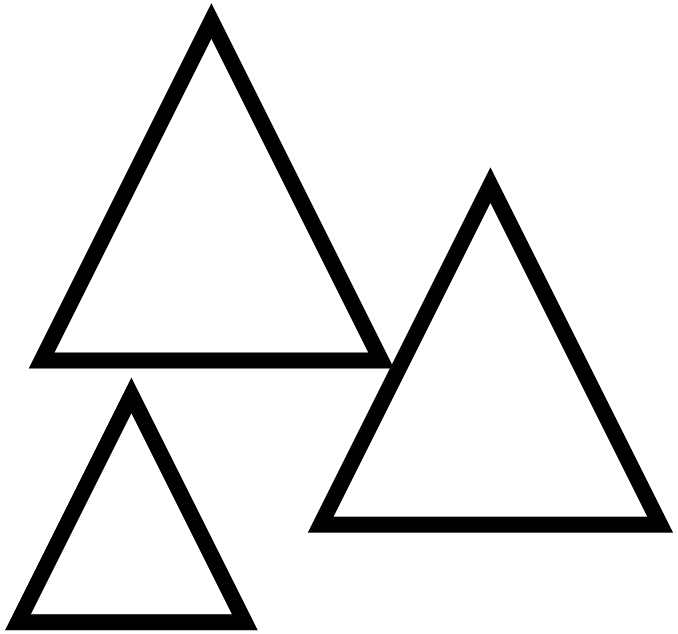


Target Geometry

Gradient Based Optimization

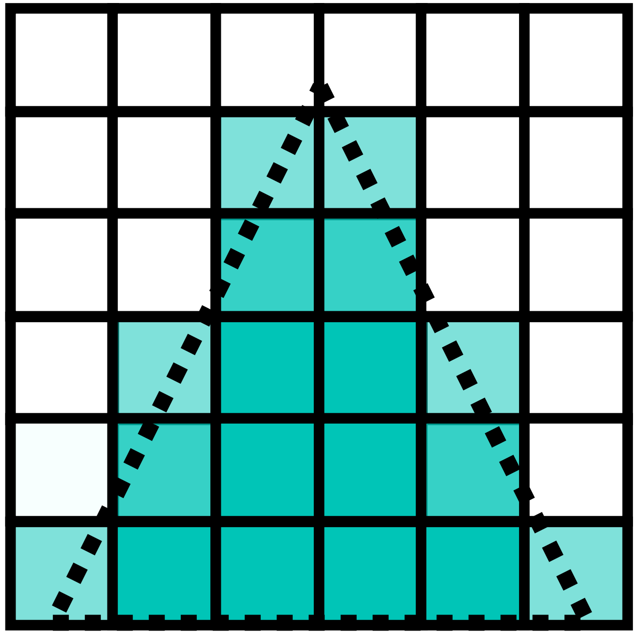


Reconstruction

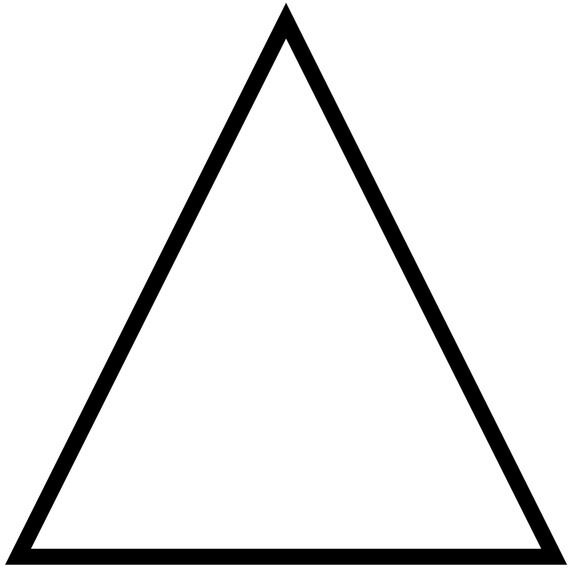


Target Geometry

Gradient Based Optimization

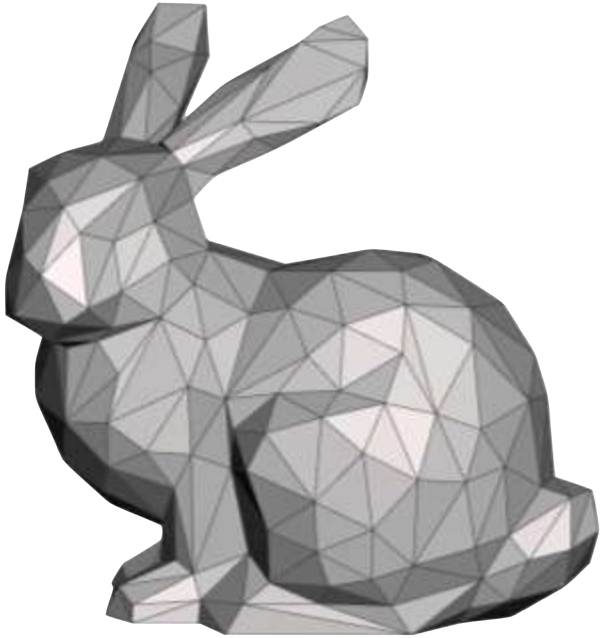


Reconstruction

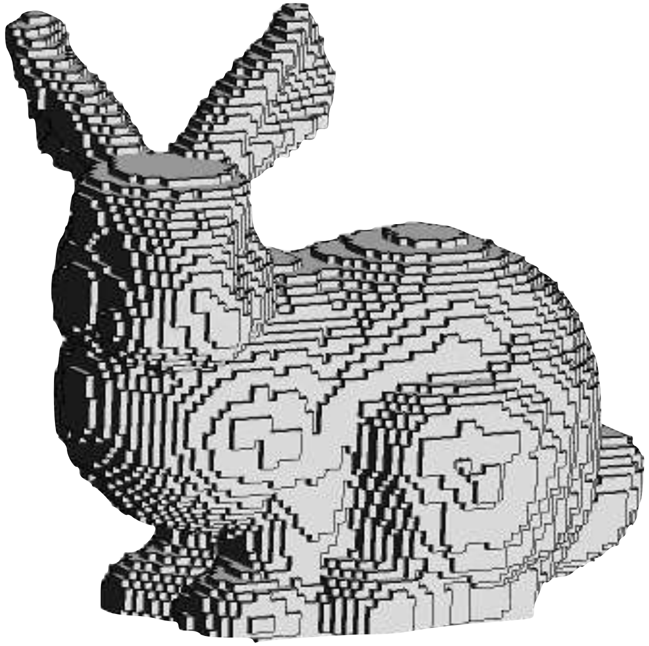


Target Geometry

Geometry Representations



Mesh Representation



Voxel Representation

Small memory footprint
Hard to optimize

Easy to optimize
Large memory footprint

Geometry
Scene representation

Implicit (continuous representation)
- algebraic surfaces
- level set $f\colon R^3 \rightarrow R,\ f(x,y,z)=0$
- more general, signed distance function

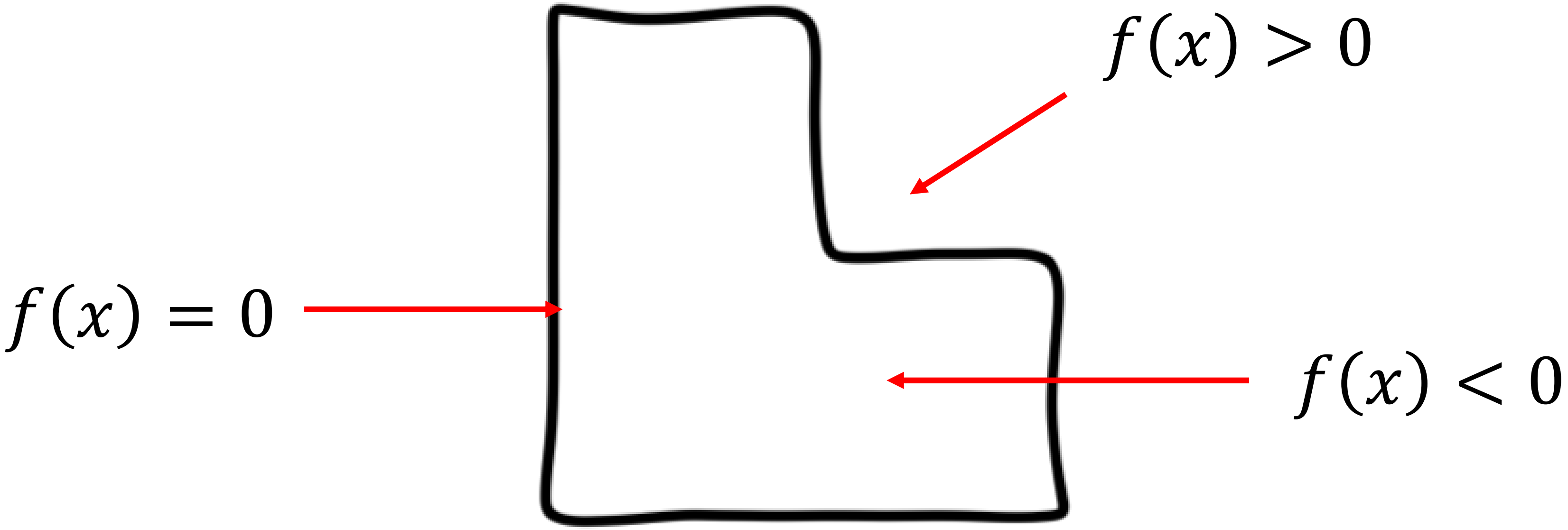
■ **Examples:**

$x^2+y^2+z^2=1$ $\qquad (R-\sqrt{x^2+y^2})^2+z^2=r^2$

Geometry

Scene representation

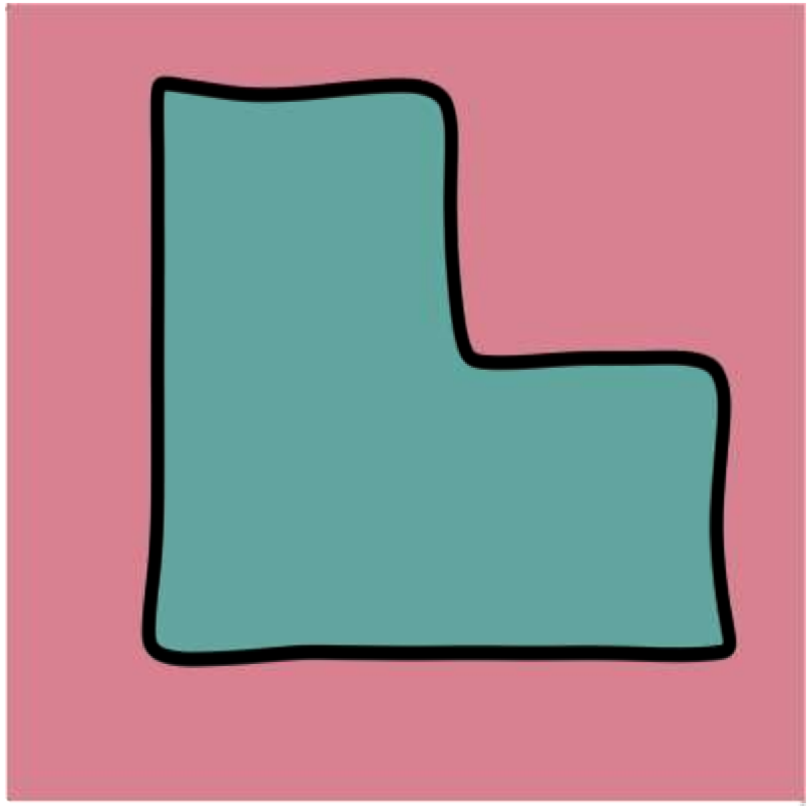
Implicit shapes



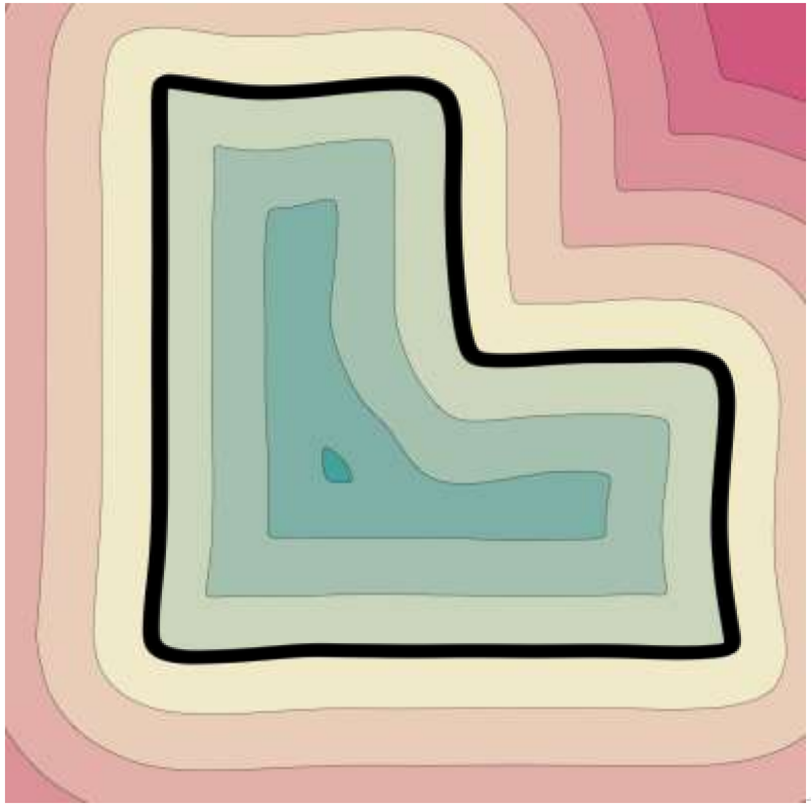
Surface represented implicitly

$$s = \{x \in \mathbb{R}^3 | f(x) = 0\}$$

Implicit shapes



Indicator / occupancy



Signed Distance Function
(SDF)

Geometry

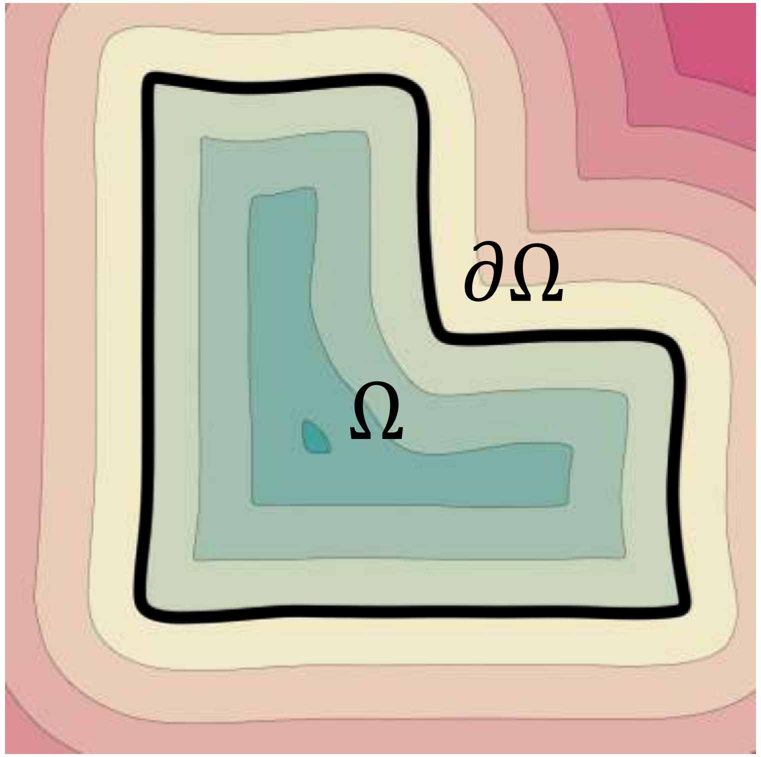
Scene representation

Implicit shapes

Eikonal equation

$$\|\nabla f(\boldsymbol{x})\| = 1, \boldsymbol{x} \in \Omega$$

$$f(\boldsymbol{x}) = 0, \boldsymbol{x} \in \partial\Omega$$



Signed distance function (SDF)

Implicit representations
Properties

- continuous representation
- can represent arbitrary topology at arbitrary resolution
- not limited by excessive memory requirements
- geometric quantities, e.g., normals
- blend well with deep learning techniques

How?

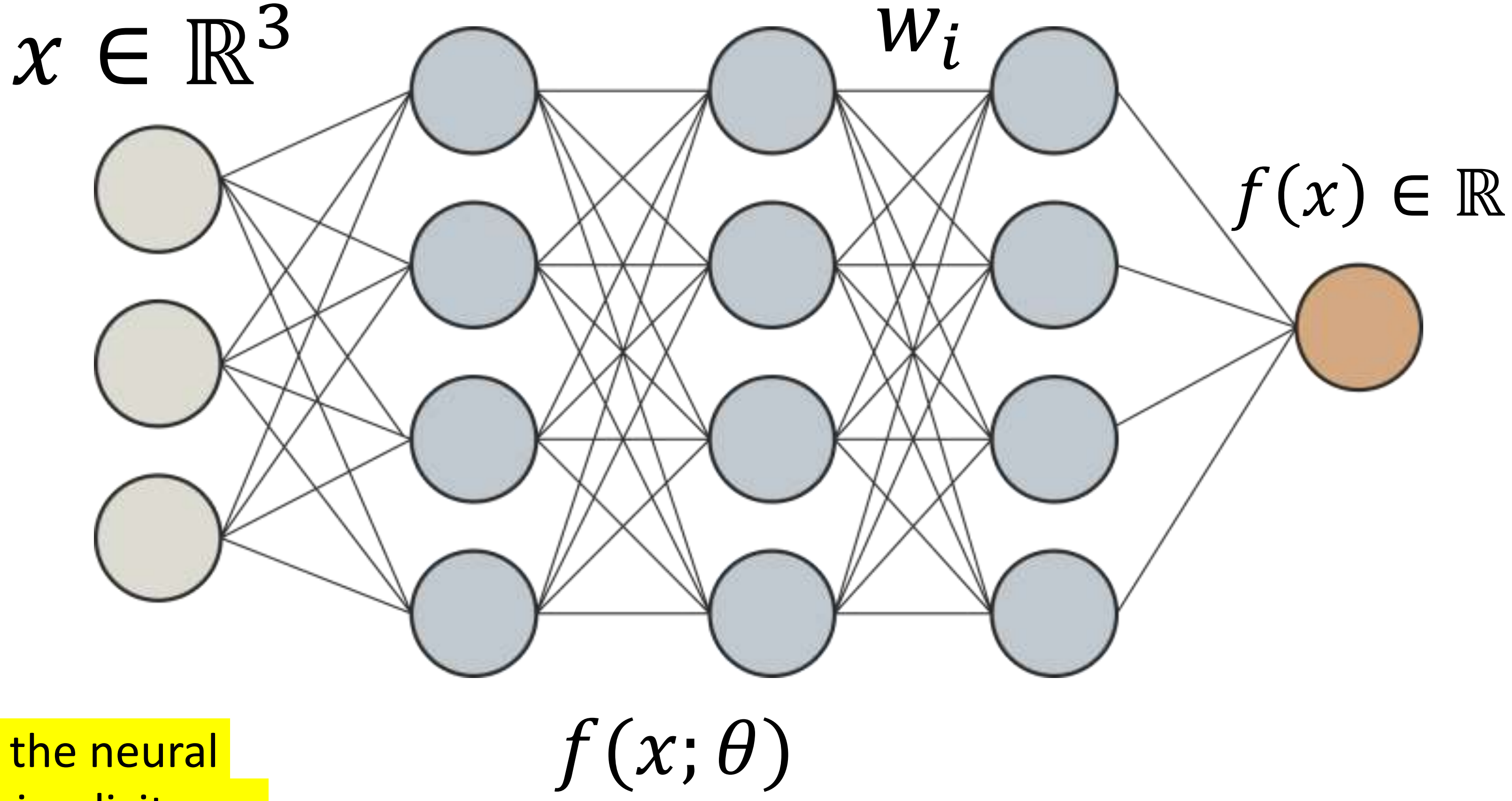
Implicit neural representations

[Park et al. 2019, Chen & Zhang 2019, Mescheder et al. 2019, Atzmon et al. 2019]

Theorem (Universality).

Any watertight piecewise linear surface can be exactly represented as the neural level set S of MLP with ReLU activations.

$$S = \{x | f(x; \theta) = 0\}$$



After training, the obtained weights in the neural net actually represent the shape, in an implicit way.

Controlling neural level sets, Atzmon et al., 2019

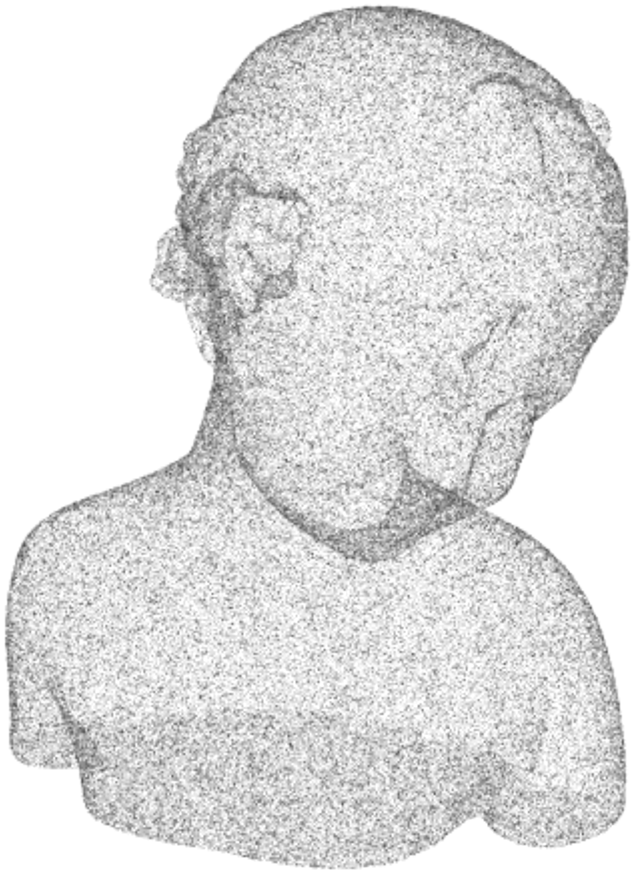
How to learn implicit neural representations?

Surface represented implicitly

$$S_\theta = \{\boldsymbol{x} | f(\boldsymbol{x};\theta) = 0\}$$

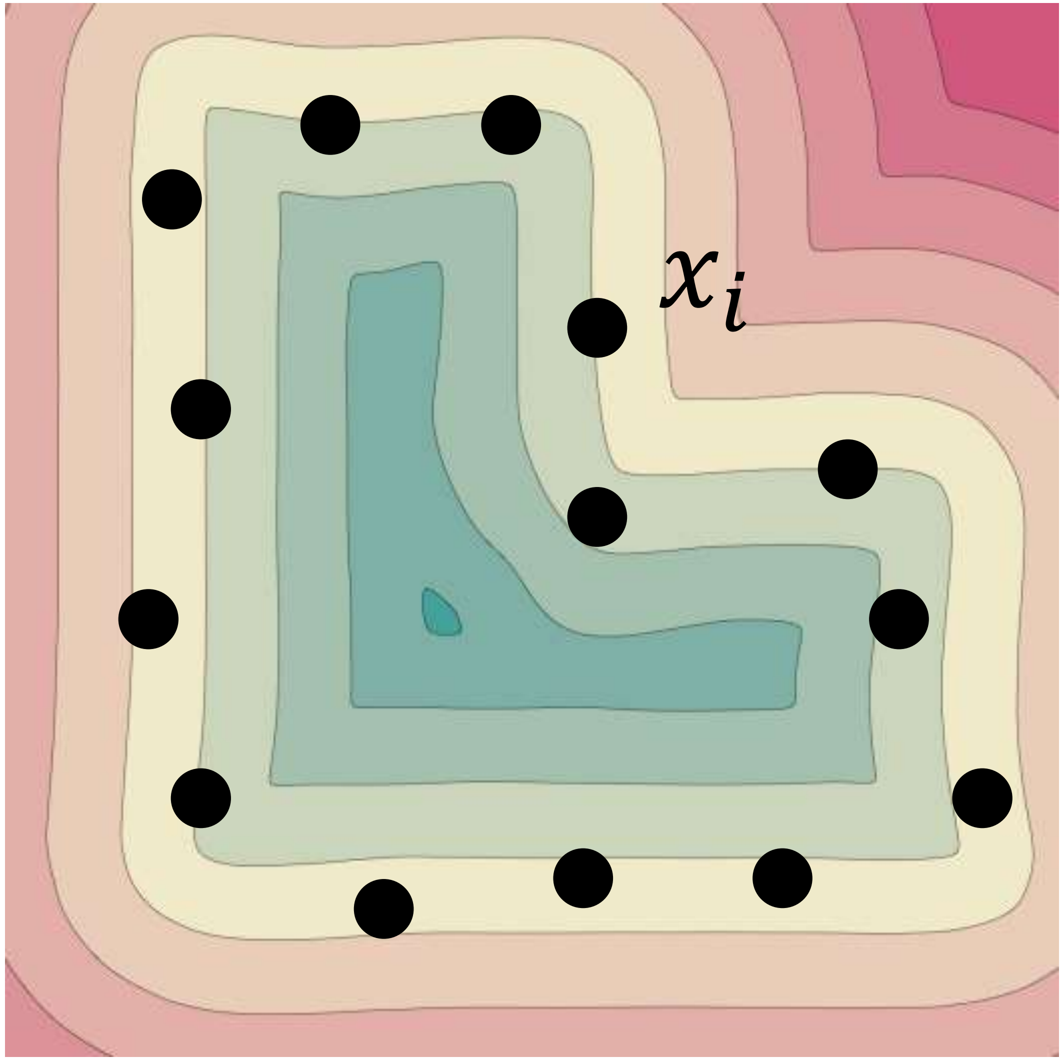
How to learn implicit neural representations?

- Full 3D supervision
- Raw data (weak supervision)



Regression with full 3D supervision

[Park et al. 2019, Chen & Zhang 2019, Mescheder et al. 2019]



$$\hat{f}$$

$$\hat{f}_i = \hat{f}(x_i)$$

$$l(f(x_i;\theta), \hat{f}_i)$$

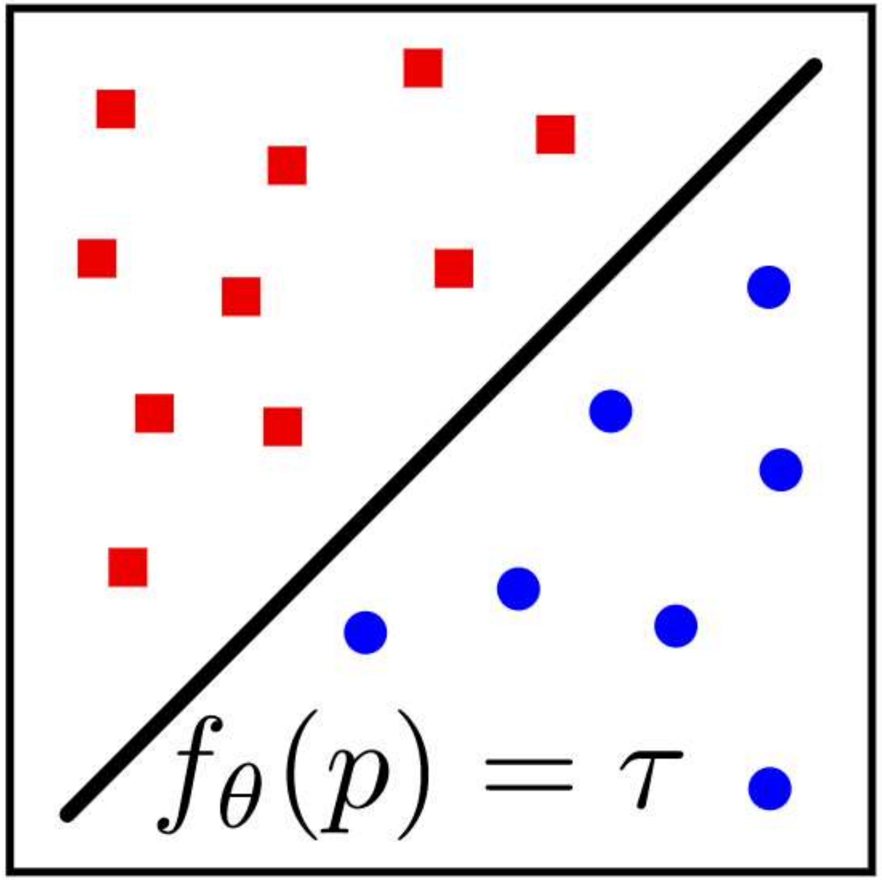
Regression with full 3D supervision

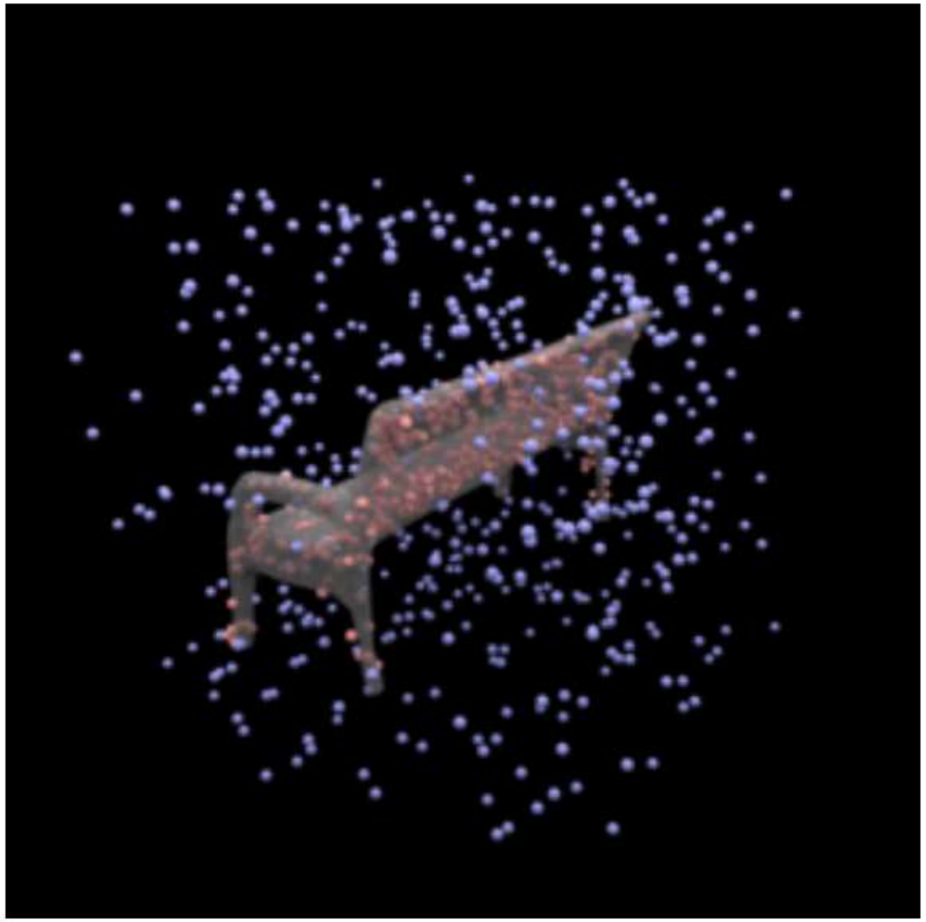
Occupancy Networks: Learning 3D Reconstruction in Function Space, Mescheder et al, 2019

- Representing the 3D geometry as the decision boundary of a classifier that *learns* to separate the object's inside from its outside
- This yields a continuous implicit surface representation
- *At inference*, queries of 3D points, allows to construct watertight meshes, by using marching cubes algorithm

Full 3D supervision

Occupancy networks, Mescheder et al., 2019





Full 3D supervision

Occupancy networks, Mescheder et al., 2019

Full 3D supervision

Occupancy networks, Mescheder et al., 2019

Occupancy network.

Learning non-linear function

$$f_\theta: \mathbb{R}^3 \to [0,1]$$

Input: $\boldsymbol{p} \in \mathbb{R}^3$
Output: probability of occupancy

The decision boundary, $f_\theta(\boldsymbol{p}) = \tau,\ (\tau = 0.5)$, represents the surface of the reconstructed shape

Full 3D supervision

Occupancy networks, Mescheder et al., 2019

- After *training* the weights of the neural net represent the surface.
- *At inference*, queries of 3D points, allows to construct watertight meshes, by using marching cubes algorithm
- **Caveat.** Full 3D supervision demands in some sense surface reconstruction

Learning implicit representation

By weak supervision, from the raw data

Point clouds

- given an input point cloud $\chi = \{x_i\}_{i \in I} \subset \mathbb{R}^3$
- our goal is to compute θ
- $f(x;\theta)$ is approximately the signed distance function to a plausible surface $\mathcal{M}$ defined by χ
- without any additional supervised data preparation

Learning implicit representation

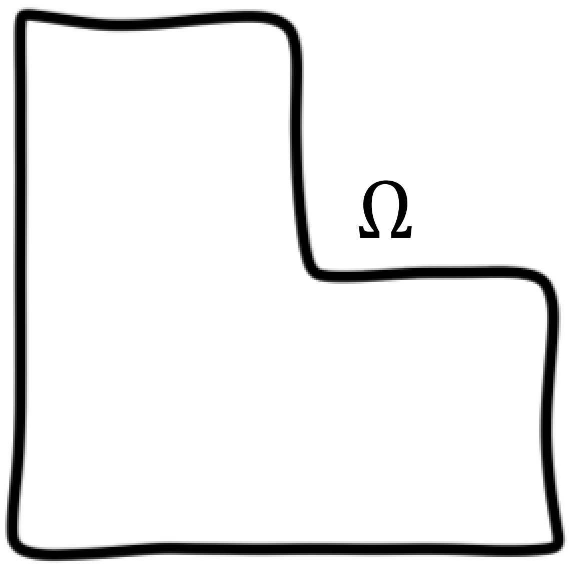
By weak supervision, from the raw data

Implicit geometric regularization **(IGR)** by Gropp, Yariv, Haim, Atzmon and Lipman 2020

Eikonal PDE

$$\|\nabla f(\boldsymbol{x})\| = 1$$

$$f(\boldsymbol{x}) = 0, \boldsymbol{x} \in \Omega$$



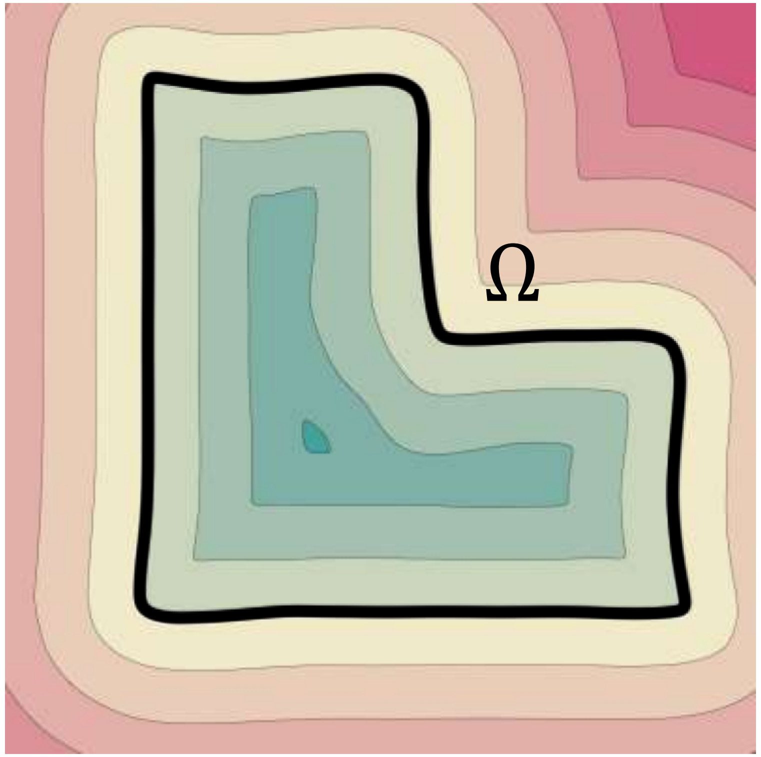
Weak supervision

Implicit geometric regularization **(IGR)**, Gropp et al., 2020

Eikonal PDE

$$\|\nabla f(\boldsymbol{x})\| = 1$$

$$f(\boldsymbol{x}) = 0, \boldsymbol{x} \in \Omega$$



Signed distance function (SDF)

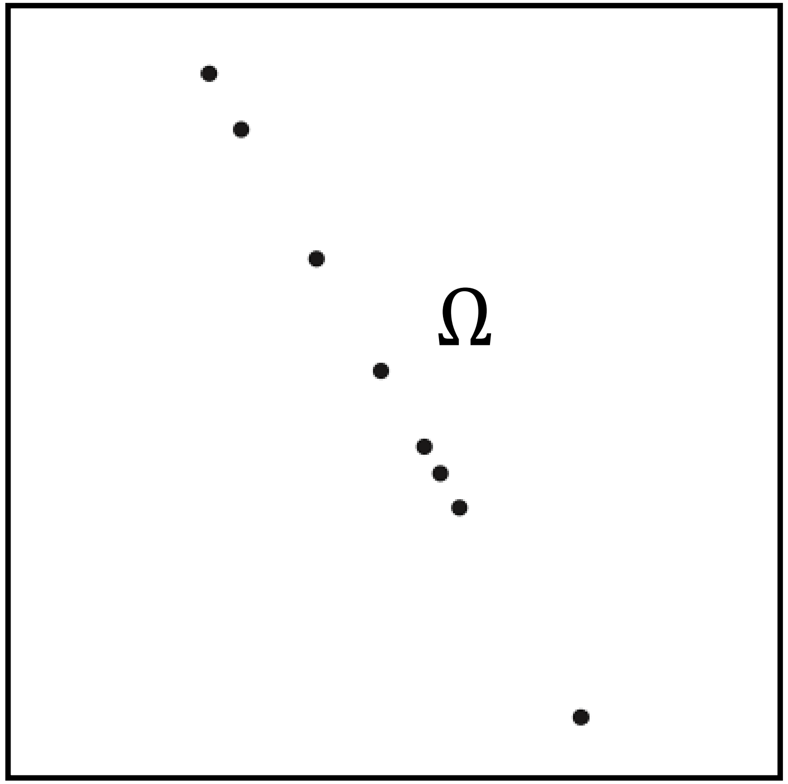
Weak supervision

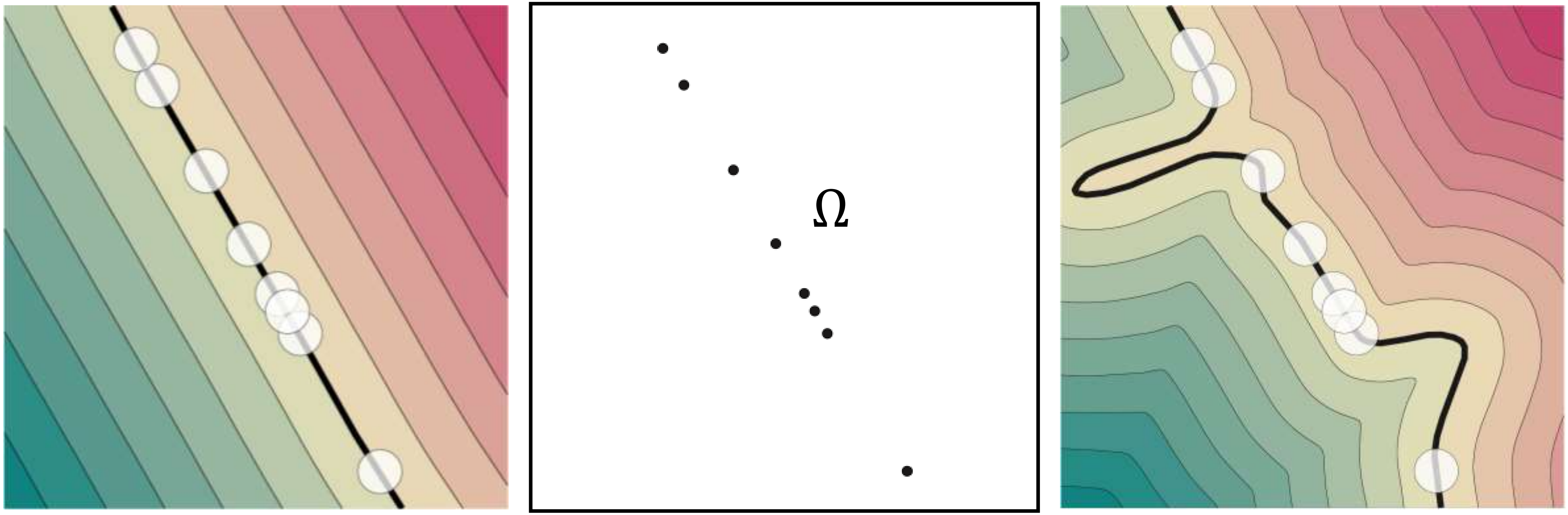
Implicit geometric regularization **(IGR)**, Gropp et al., 2020

Eikonal PDE

$$\|\nabla f(\boldsymbol{x})\| = 1$$

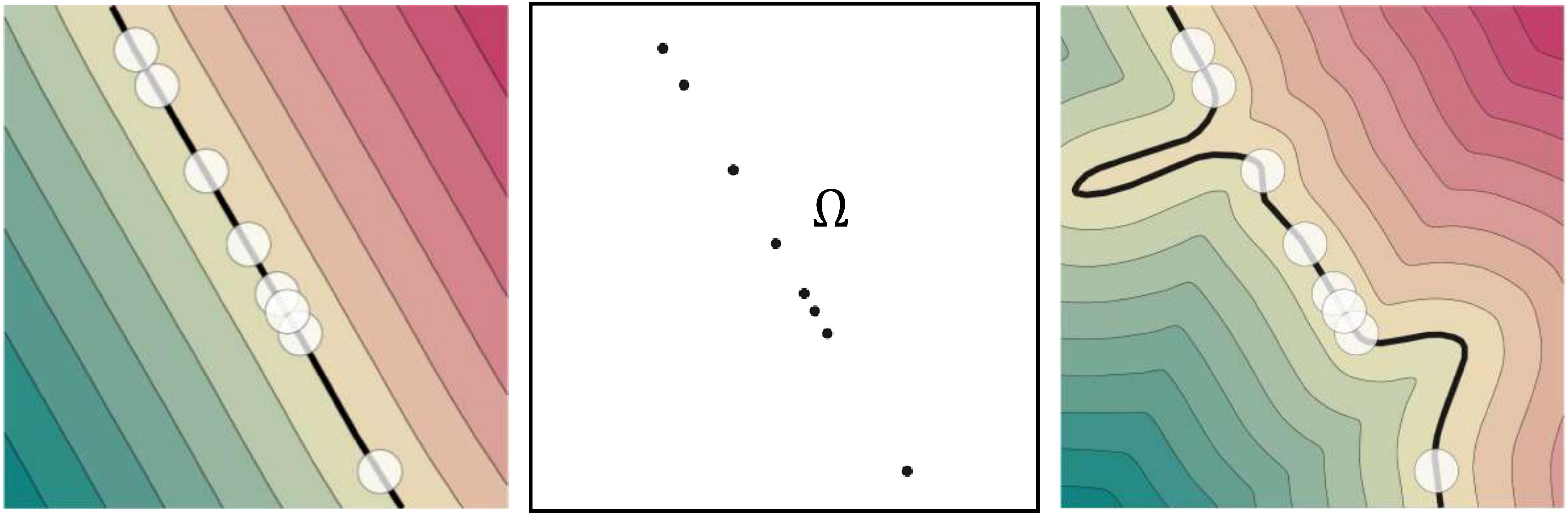
$$f(\boldsymbol{x}) = 0, \boldsymbol{x} \in \Omega$$




Ω

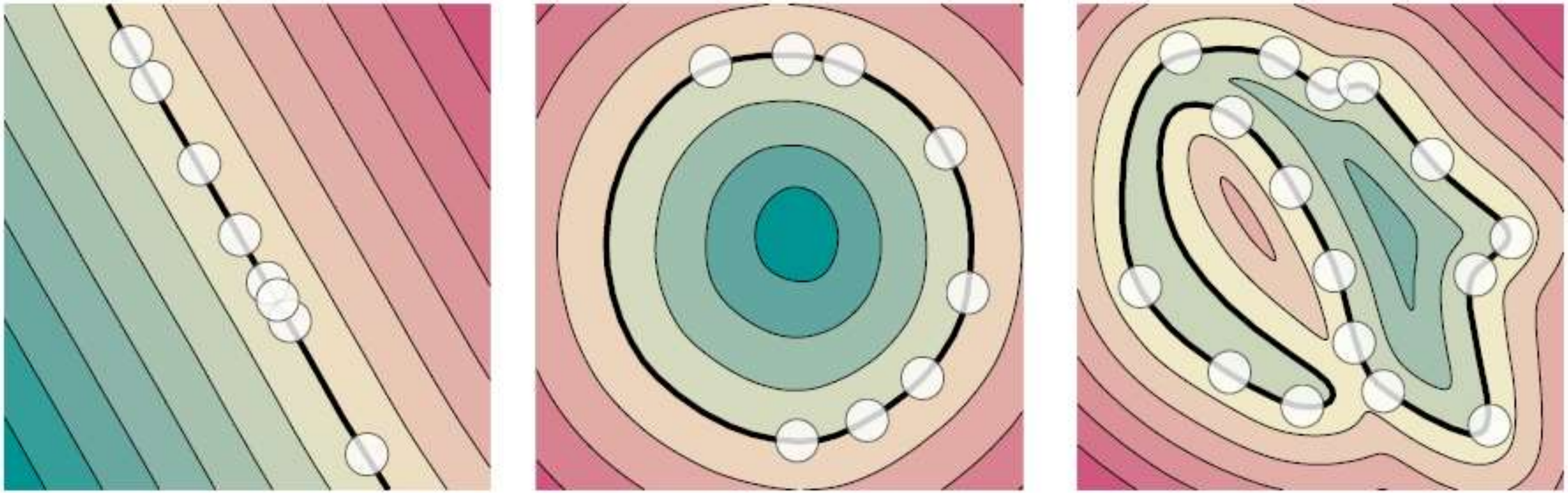
$$\text{loss}(\theta) = \sum_{i \in I} |f(x_i;\theta)|^2 + \lambda \mathbb{E}_x (\|\nabla_x f(x;\theta)\| - 1)^2$$

vanish — Eikonal



$$\text{loss}(\theta) = \sum_{i \in I} |f(x_i;\theta)|^2 + \lambda \mathbb{E}_x(\|\nabla_x f(x;\theta)\| - 1)^2$$

vanish — Eikonal



Weak supervision

Implicit geometric regularization **(IGR)**, Gropp et al., 2020

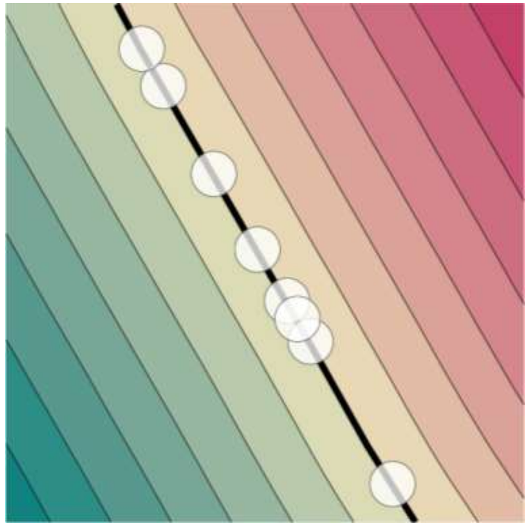
Weak supervision

Implicit geometric regularization **(IGR)**, Gropp et al., 2020

Inductive bias

Theorem (Convergence and linear reproduction)
Gradient descent of the linear model with random initialization converges with probability 1 to the reproducing plane

$$\text{loss}(\theta) = \sum_{i \in I} \left(w^T x_i\right)^2 + \lambda\left(\|w\|^2 - 1\right)^2$$

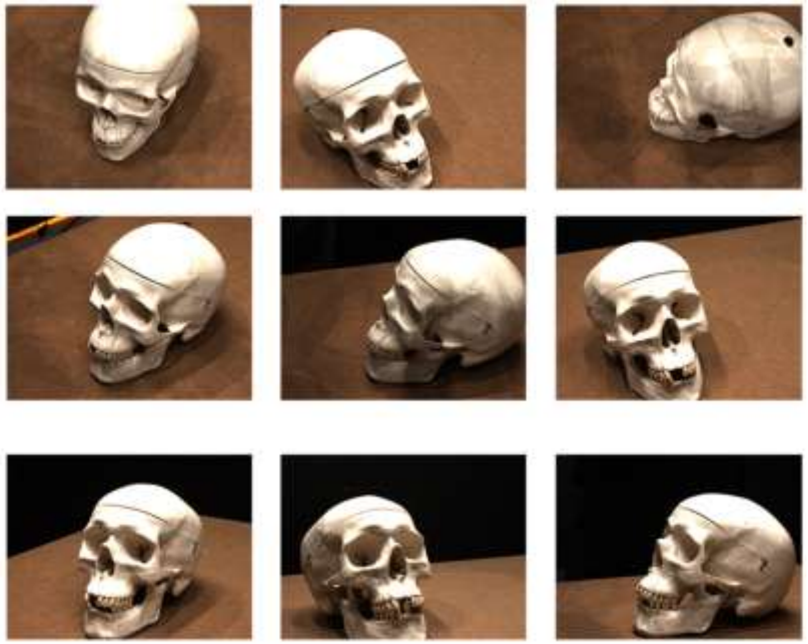


Learning implicit neural representation

By weak supervision, from the raw data

Point clouds

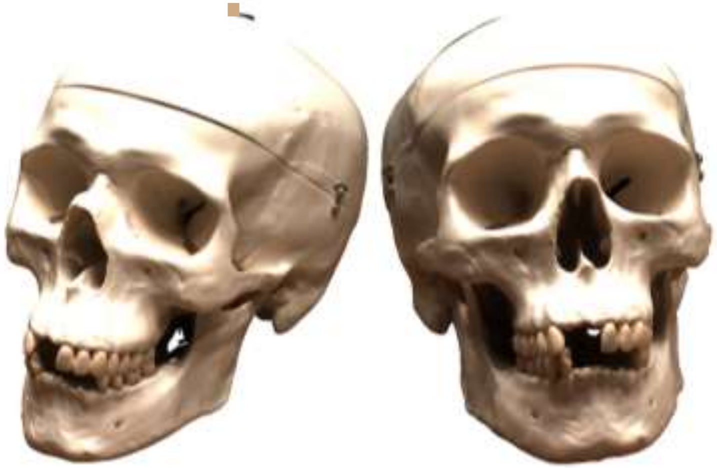
Images



Neural rendering

Geometry reconstruction

Render new views



Cameras

Slide by Lior Yariv

Neural rendering

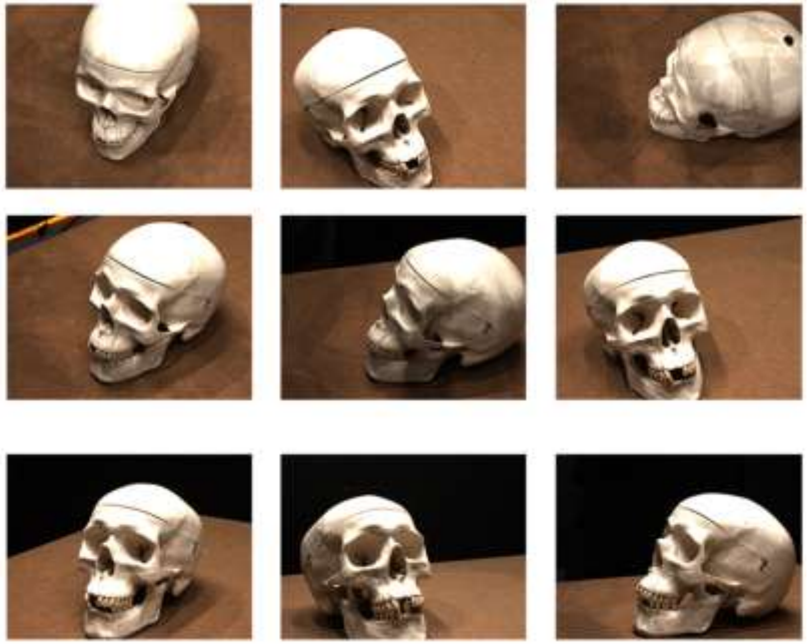
Deep image or video generation approaches that enable explicit or implicit control of scene properties such as illumination, camera parameters, pose, geometry, appearance and semantic structure

Neural rendering brings the promise of addressing both *reconstruction* and *rendering* by using deep networks to learn complex mappings from captured images to novel images

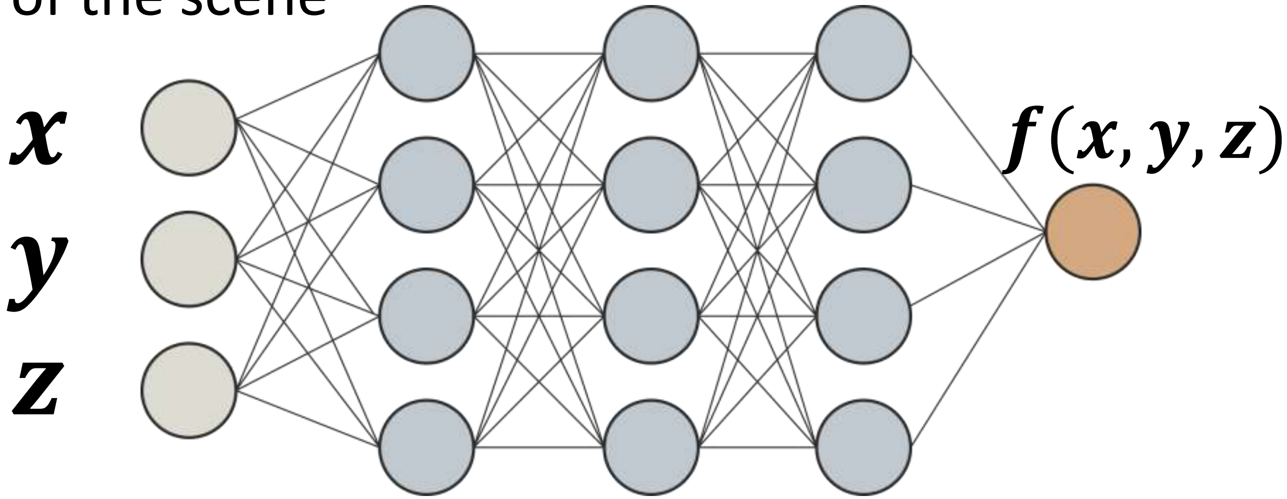
State of the Art on Neural Rendering, A. Tewari et al., 2020

Neural rendering

- Learning from raw data (weak supervision)

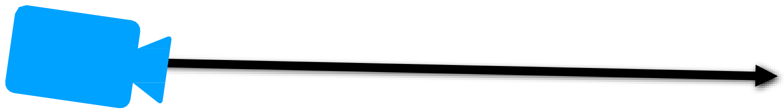


- Building (implicit) neural representation of the scene



Neural Volumetric Rendering

computing color along
rays through 3D space

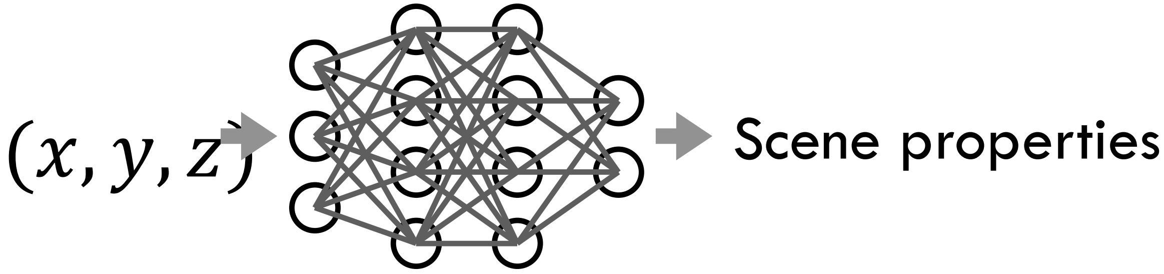


What color is this pixel?

Slides on volume rendering formulation
by Ben Mildenhall

Neural Volumetric Rendering

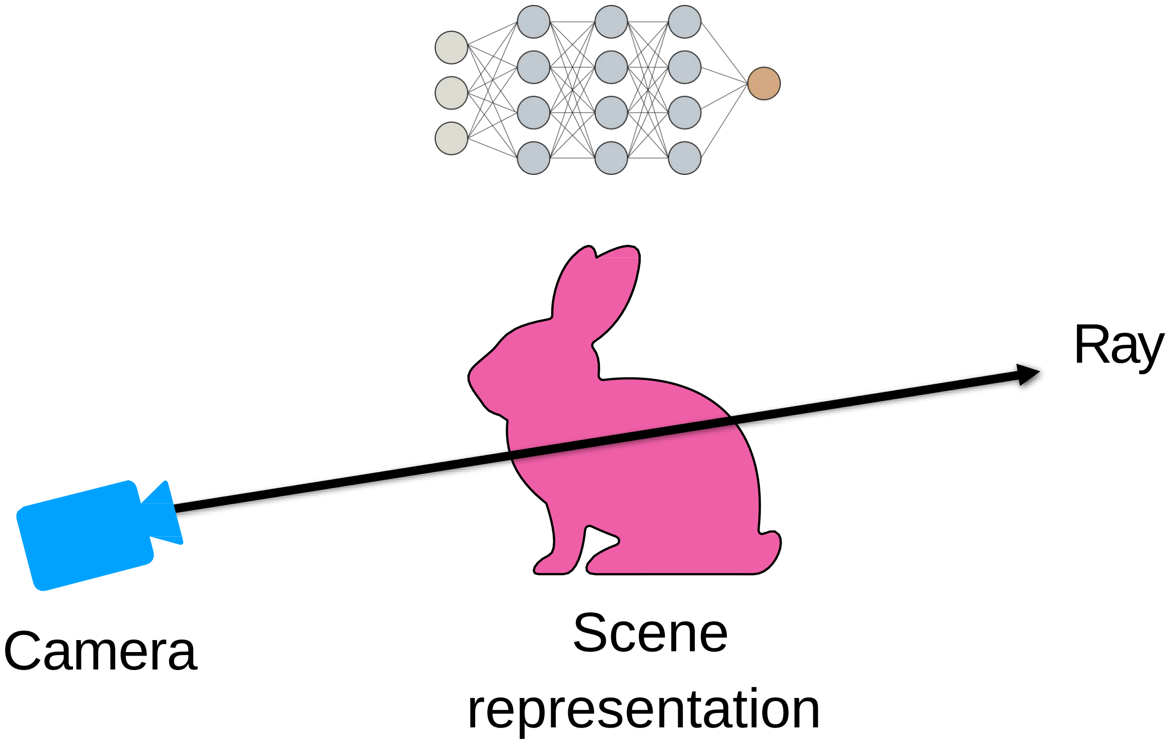
using a neural network as a
scene representation, rather than
a voxel grid of data



Neural Volumetric Rendering

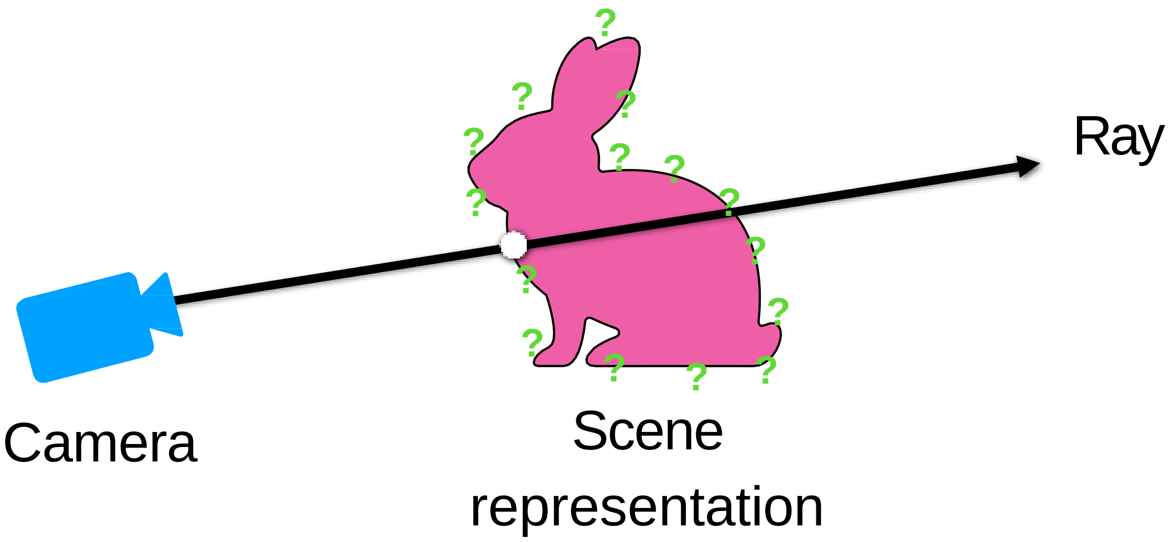
continuous, differentiable
rendering model without
concrete ray/surface intersections

Neural rendering



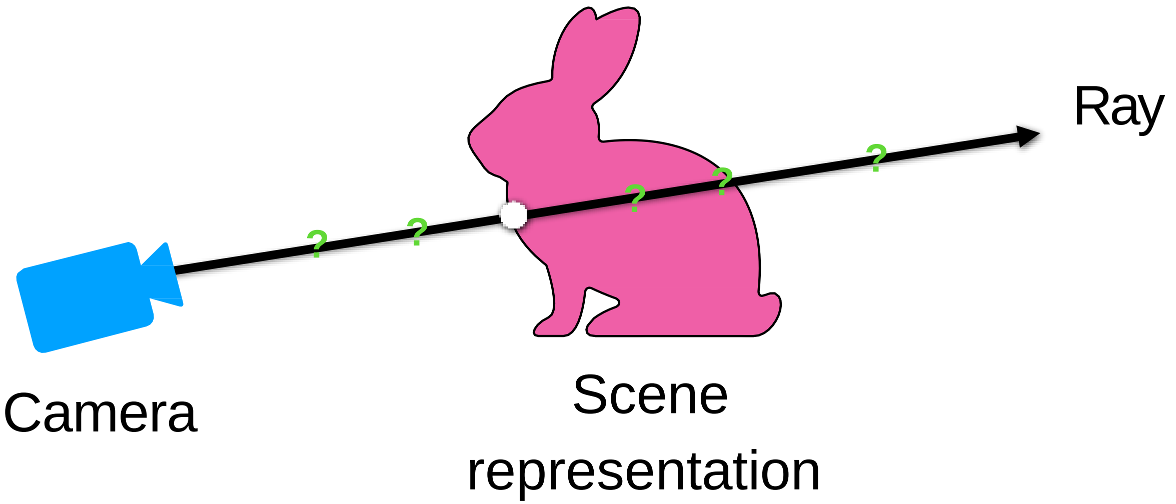
Want to know how ray interacts with scene

Neural rendering - surface vs. volume rendering



Surface rendering — loop over geometry, check for ray hits

Neural rendering - surface vs. volume rendering



Volume rendering — loop over ray points, query geometry

Neural Volumetric Rendering

NeRF

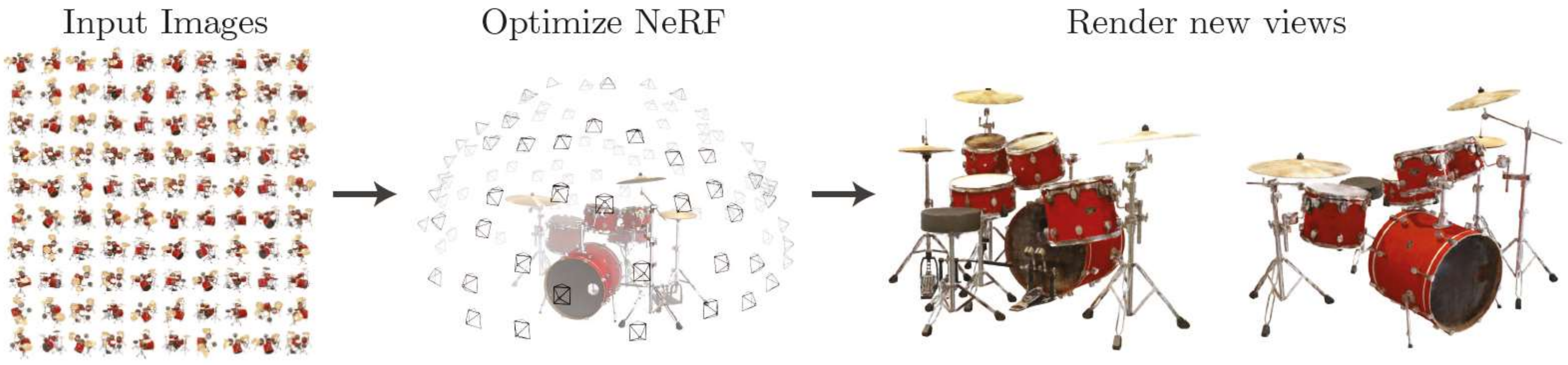
Representing Scenes as **Ne**ural **R**adiance **F**ields for View Synthesis

By Mildenhall, Srinivasan, Tancik, Barron, Ramamoorth and Ng, 2020

NeRF

Representing Scenes as Neural Radiance Fields for View Synthesis

By Mildenhall, Srinivasan, Tancik, Barron, Ramamoorth and Ng, 2020



A NeRf stores a volumetric ***scene representation as the weights of an MLP***, trained on many images with known pose

Slides on NeRF are based on slides of Yoni Kasten and Dolev Ofri

NeRF

Inference

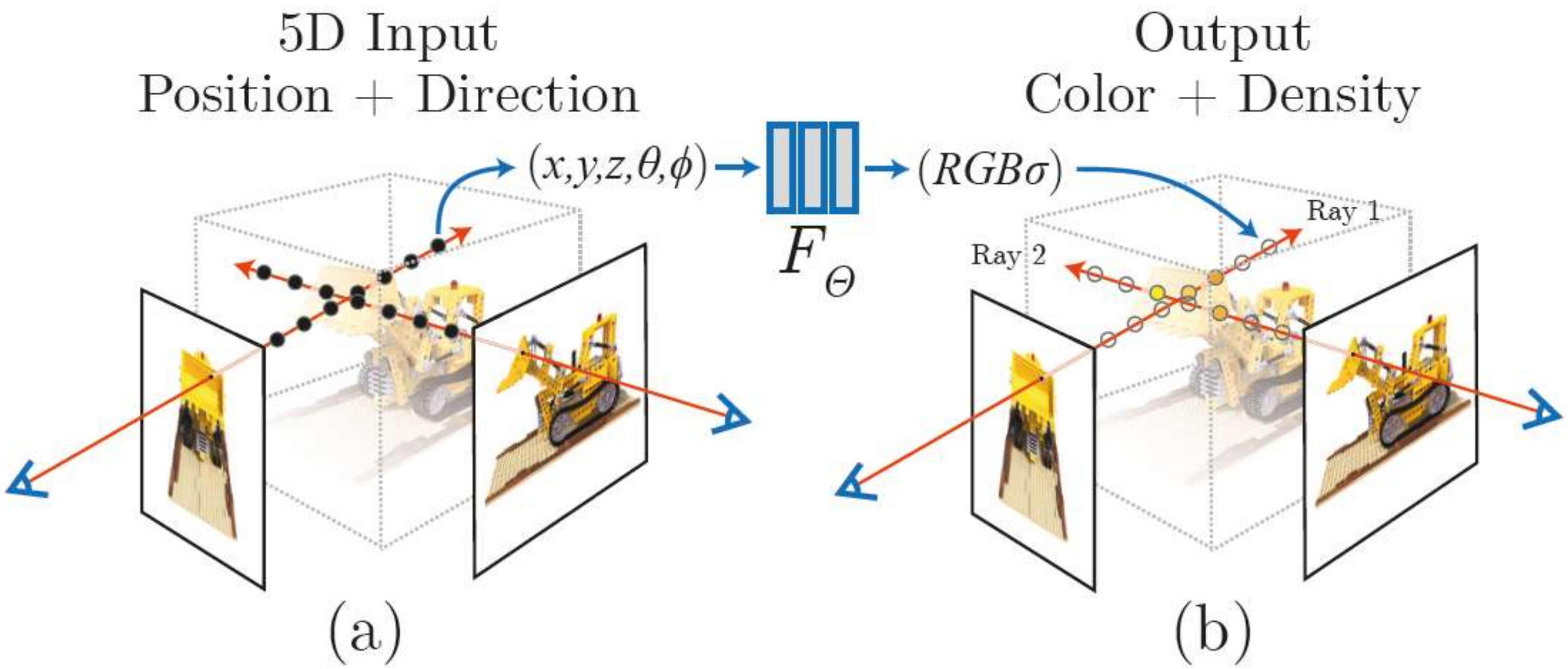
The scene is represented by MLP

Input: spatial location (x, y, z) and viewing direction (θ, ϕ)

Output: volume density (opacity), radiance emitted at direction (θ, ϕ) at point (x, y, z)

NeRF

Inference: ***render new photorealistic images from the learned scene***



New views are rendered by integrating the density and color at regular intervals along each viewing ray (volume rendering)

NeRF

Training

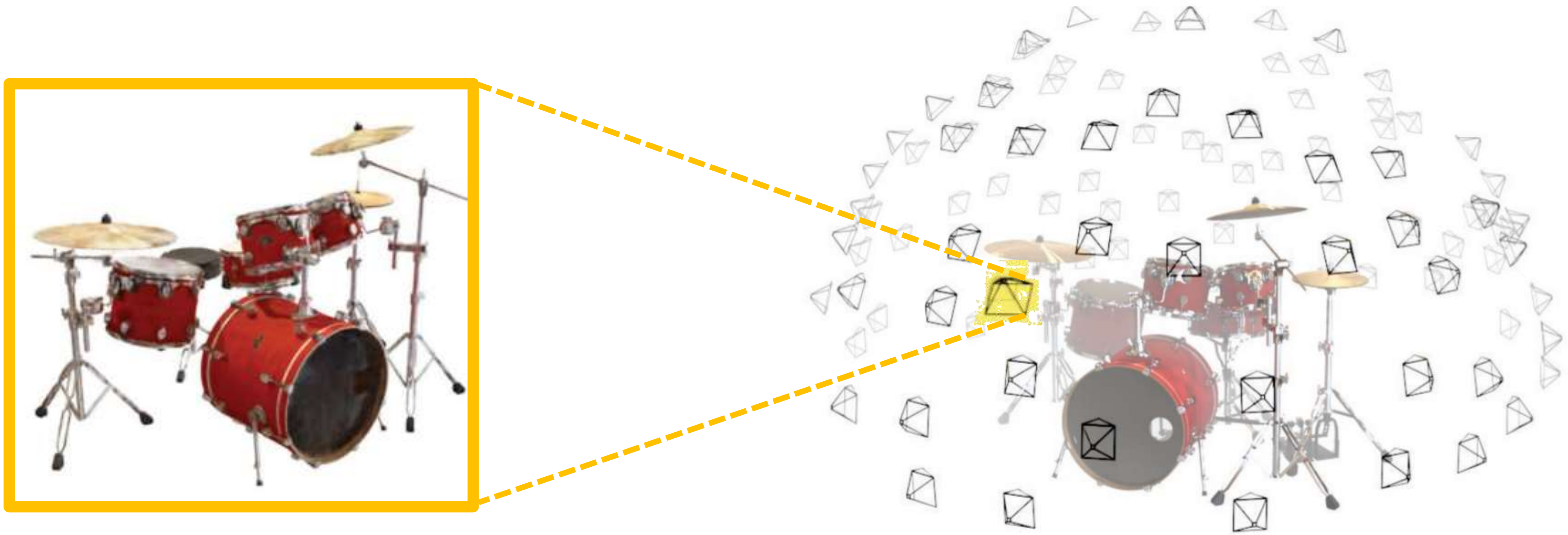
Objective: reconstruct all training views by volume rendering

Multiview Images of a single scene

Camera poses

NeRF

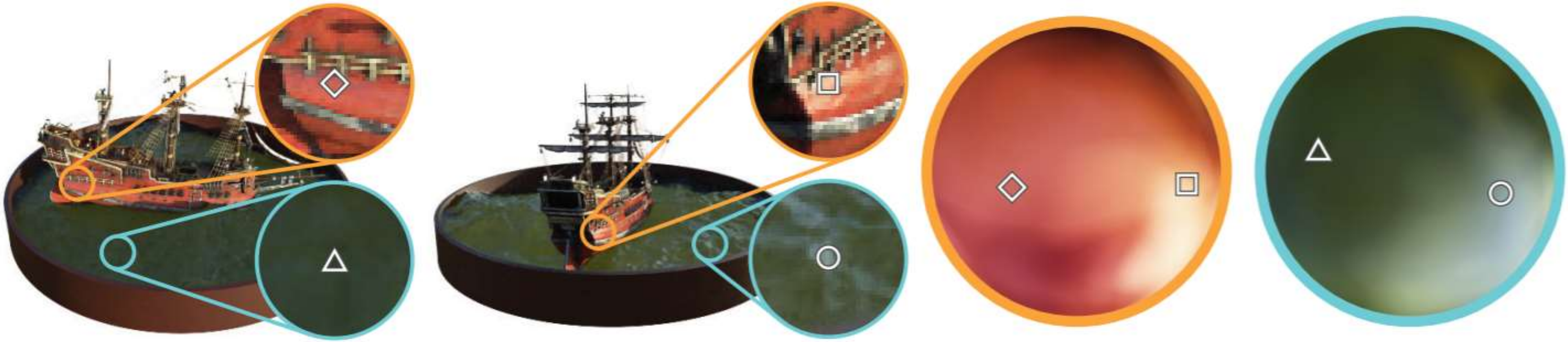
Rendering novel views



NeRF

Scene representation

Looking at materials from different point of views



(a) View 1 (b) View 2 (c) Radiance Distributions

NeRF

Neural volume rendering

Neural volume rendering refers to methods that generate images by tracing a ray into the scene and taking an integral over the length of the ray

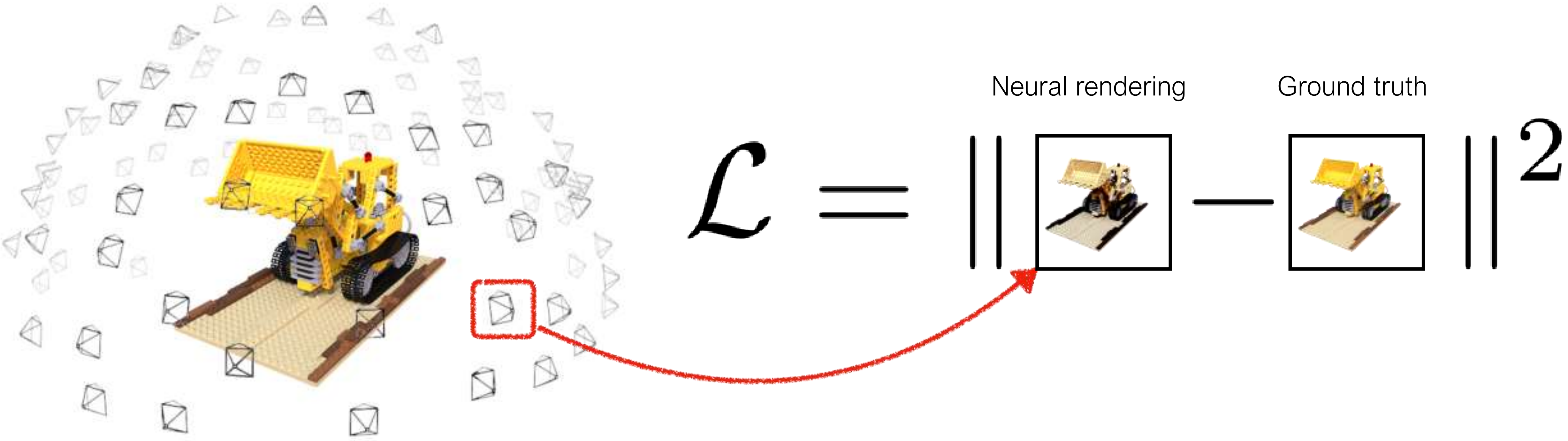
A neural network (MLP) encodes a function from the 3D coordinates on the ray to quantities like density and color, which are integrated to yield an image

Two key properties:

- Integration over the ray
- Coordinate-based scene representation

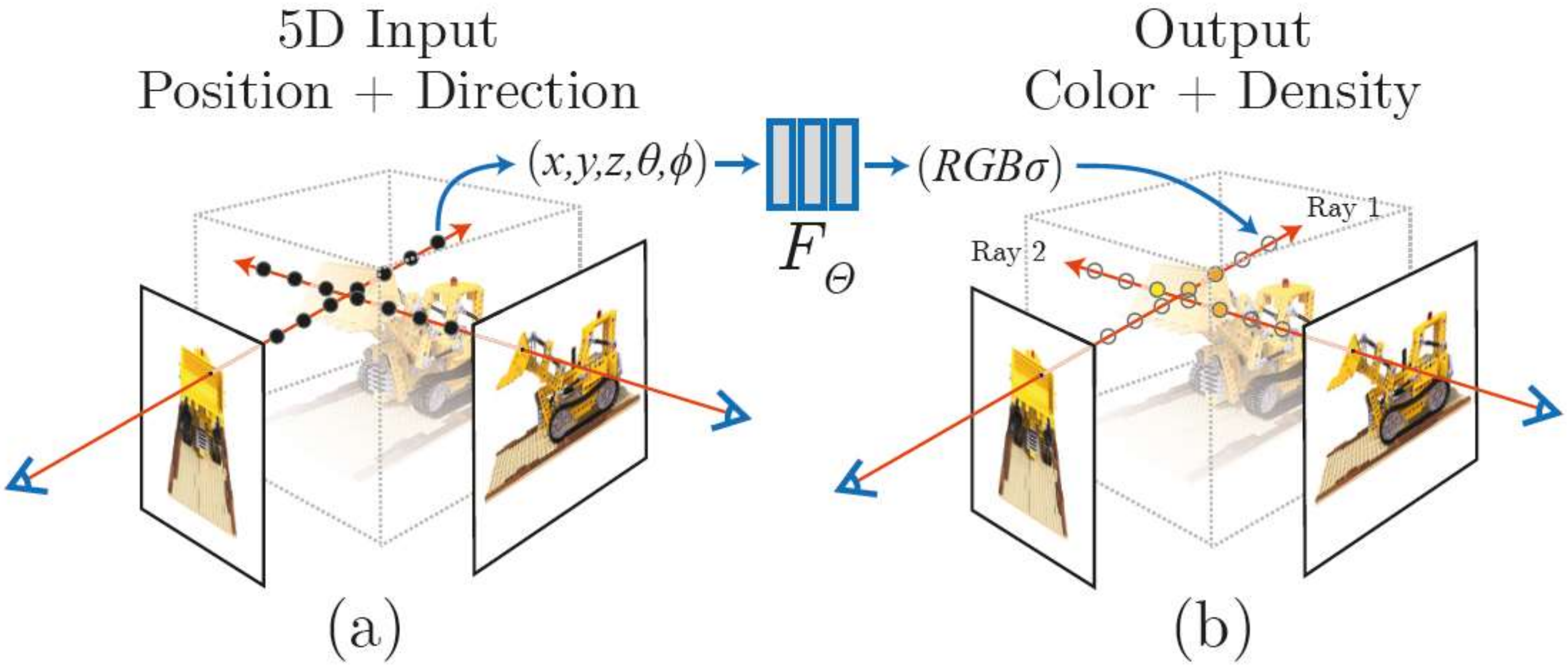
NeRF

Simulate the rendering of a learned neural scene representation in a differentiable way, and minimize:



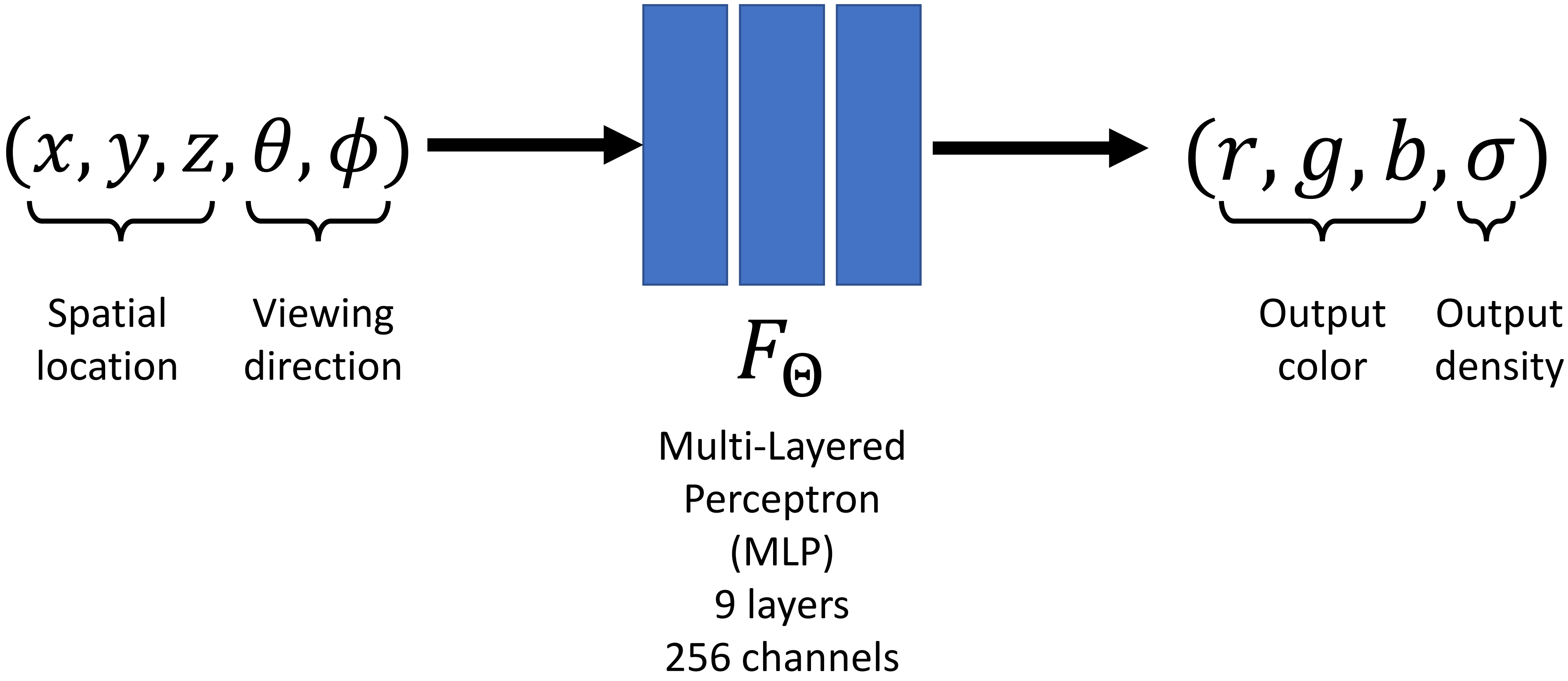
NeRF

Objective at training: reconstruct all training views by differentiable volume rendering



NeRF

Scene representation



Slide credit: Jon Barron's talk

NeRF

Scene representation

After training the scene is encoded in the weights of the neural weights

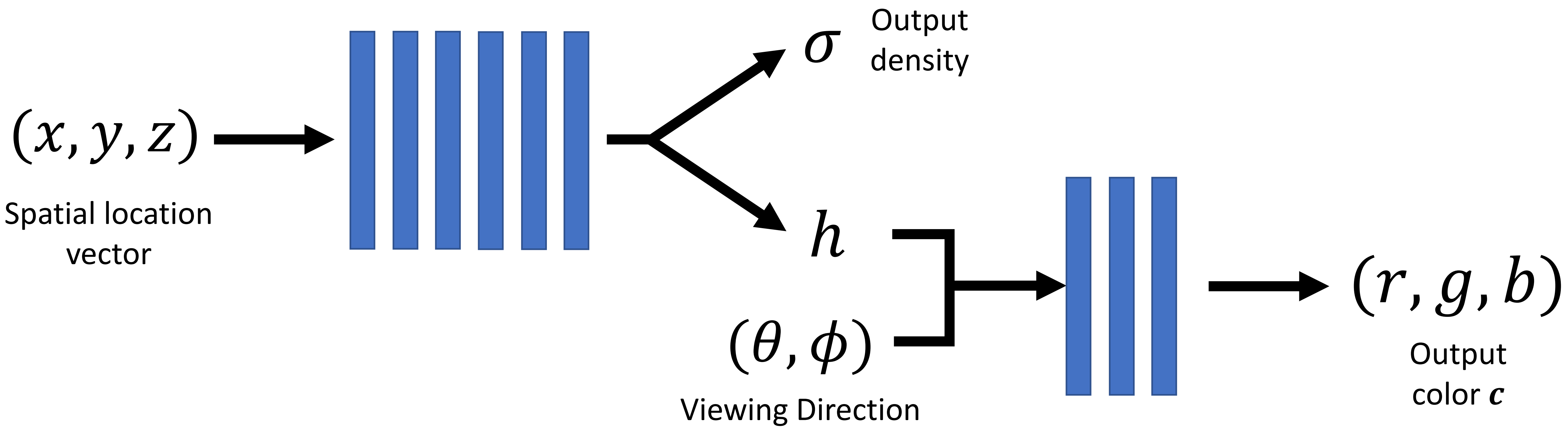
$$(\underbrace{x, y, z}_{\text{Spatial location}}, \underbrace{\theta, \phi}_{\text{Viewing direction}}) \rightarrow F_{\Theta} \rightarrow (\underbrace{r, g, b}_{\text{Output color}}, \underbrace{\sigma}_{\text{Output density}})$$

F_{Θ}

Multi-Layered Perceptron (MLP)
9 layers
256 channels

NeRF

Scene representation

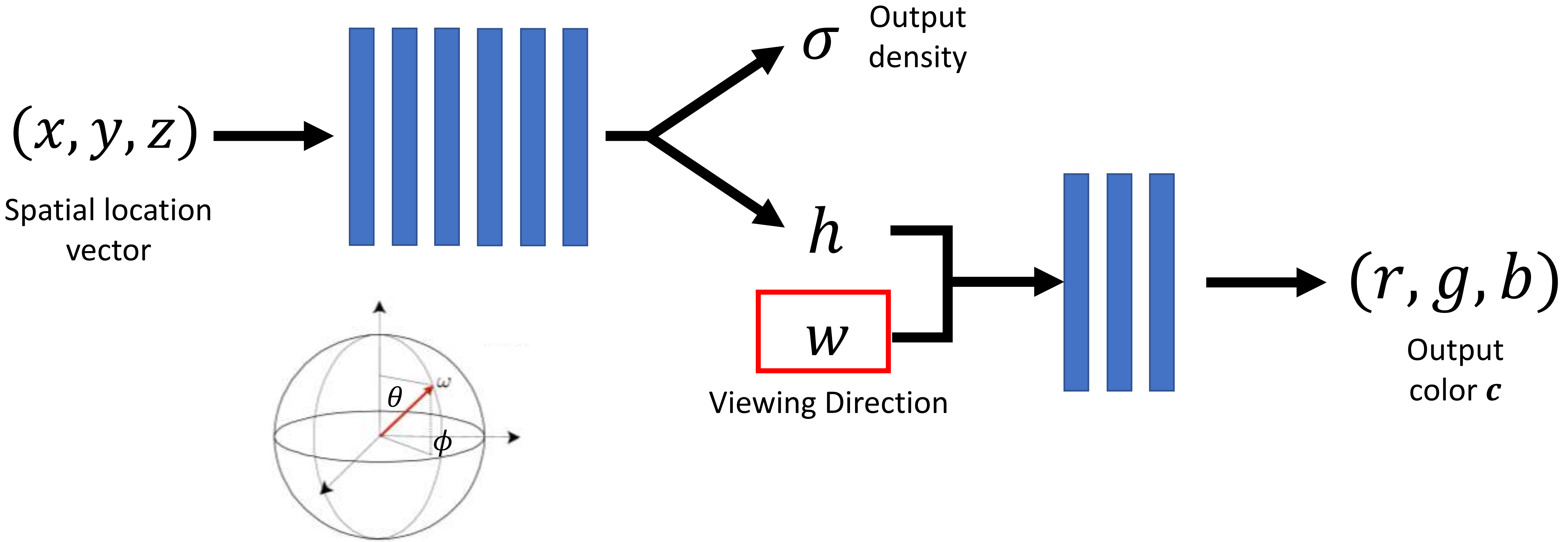


σ (spatial location)

$\boldsymbol{c}$ (spatial location, viewing direction)

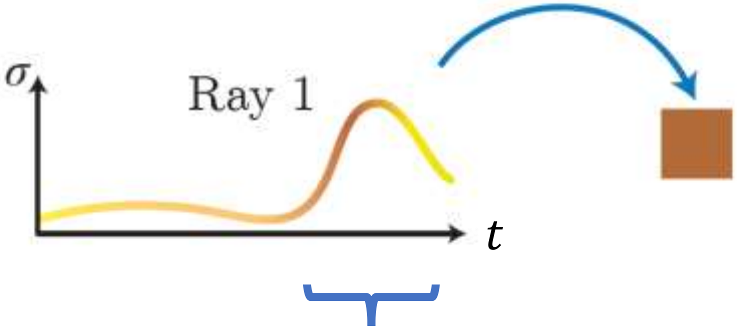
NeRF

Scene representation

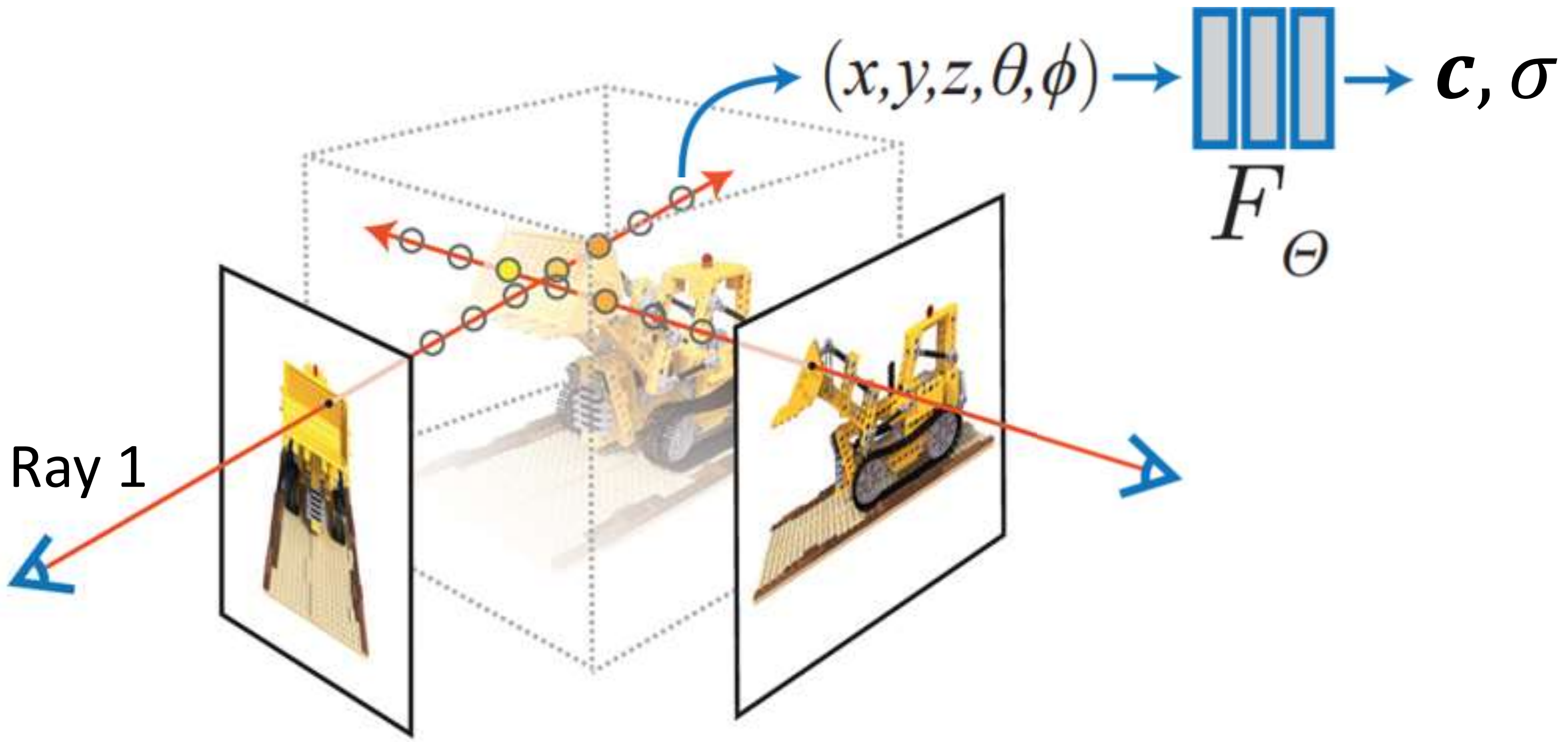


NeRF

Volume rendering



The ray hit only once

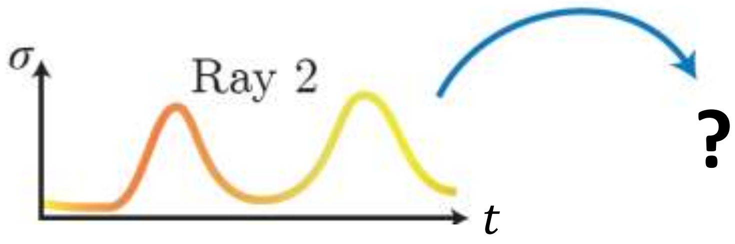


$\boldsymbol{r}(t)$ – camera ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$

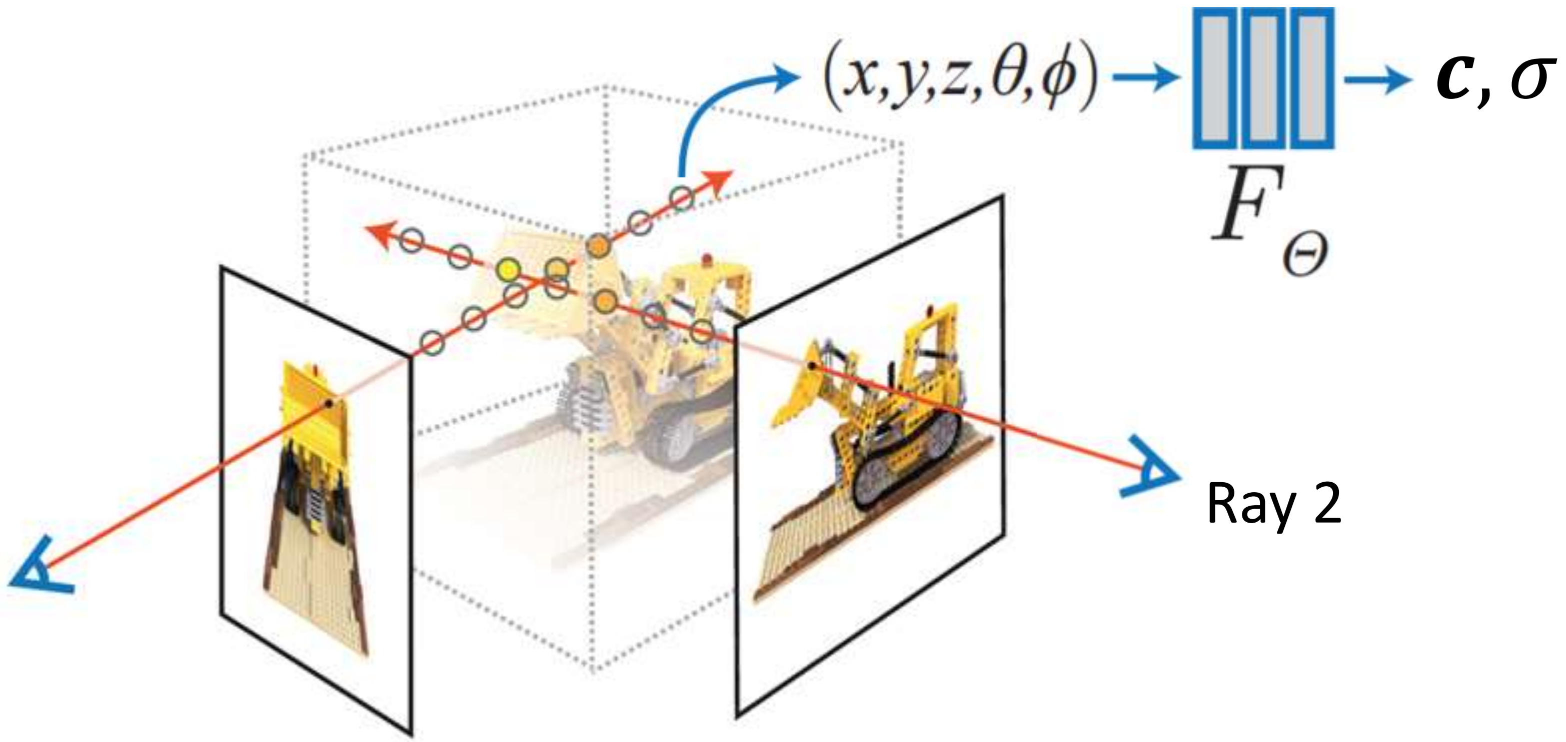
σ – volume density

NeRF

Volume rendering



?



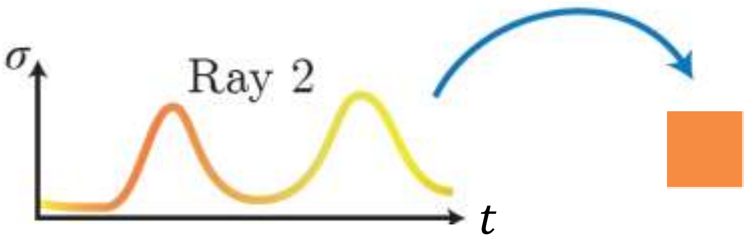
$\boldsymbol{r}(t)$ – camera ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$

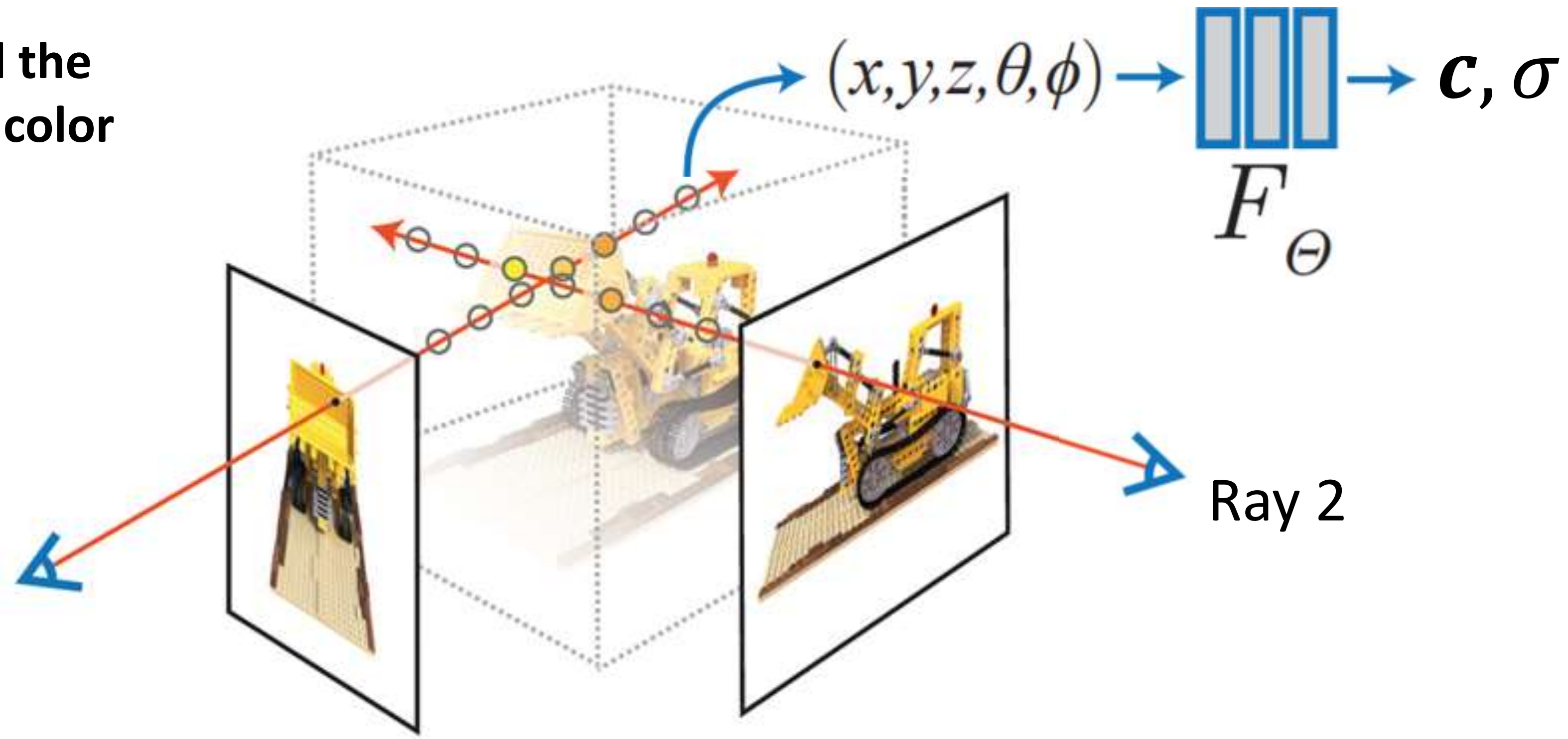
σ – volume density

NeRF

Volume rendering

We need to combine the density and the visibility in order to get the required color



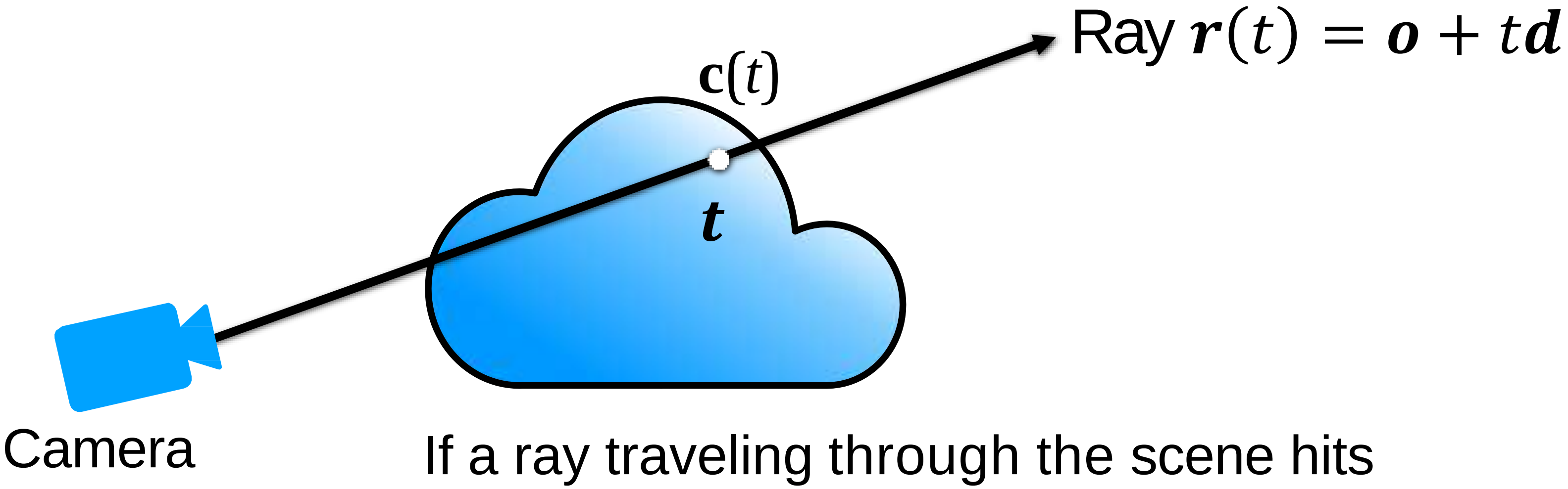


$\boldsymbol{r}(t)$ – camera ray $\boldsymbol{r}(t) = \boldsymbol{o} + t\boldsymbol{d}$

σ – volume density

NeRF

Volume rendering formulation

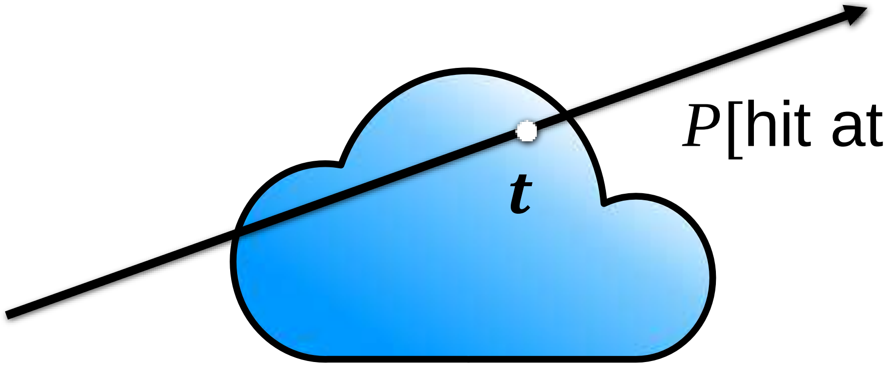


If a ray traveling through the scene hits a particle at distance t along the ray, we return its color $\mathbf{c}(t)$

NeRF

Volume rendering formulation

What does it mean for a ray to “hit” the volume?



$$P[\text{hit at } t] = \sigma(t)\, dt$$

This notion is *probabilistic:* chance that ray hits a particle in a small interval around t is $\sigma(t)\, dt$.

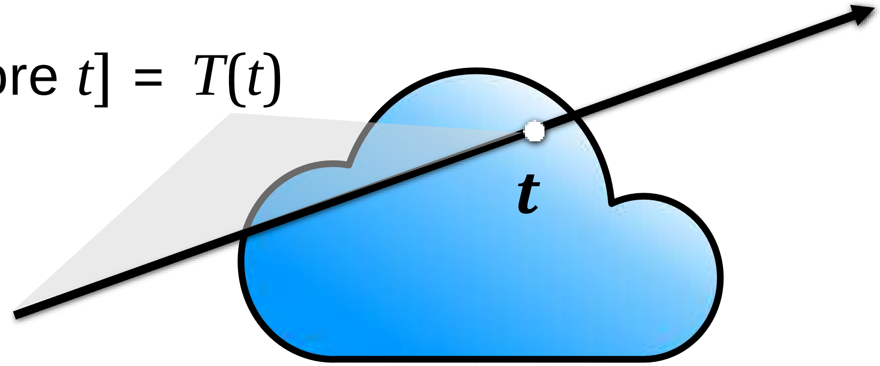
σ is called the “volume density”

NeRF

Volume rendering formulation

Probabilistic interpretation

P[no hits before t] = $T(t)$



To determine if t is the *first* hit along the ray, need to know $T(t)$: the probability that the ray makes it through the volume up to t.

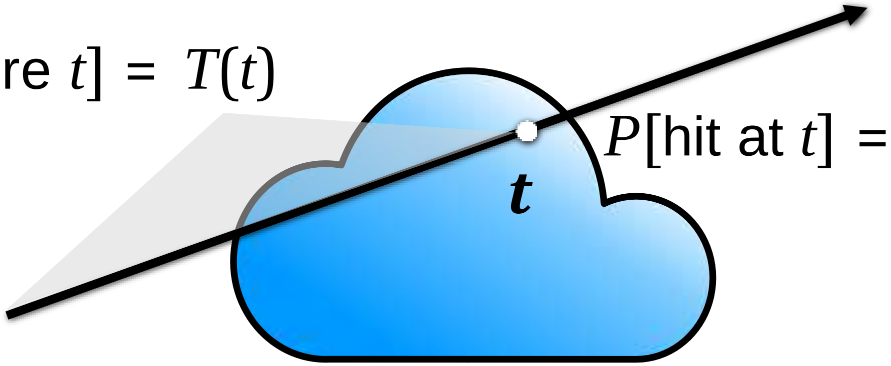
$T(t)$ is called "transmittance"

NeRF

Volume rendering formulation

Probabilistic interpretation

$P[\text{no hits before } t] = T(t)$



$P[\text{hit at } t] = \sigma(t)\,dt$

The product of these probabilities tells us

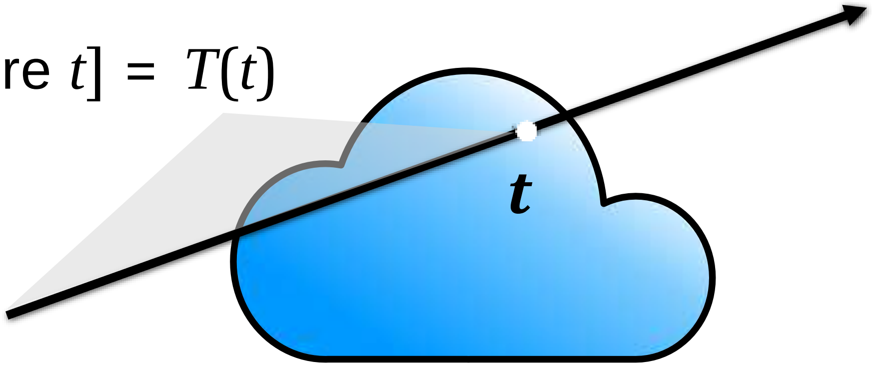
$$P[\text{first hit at } t] = P[\text{no hit before } t] \times P[\text{hit at } t] = T(t)\sigma(t)dt$$

NeRF

Volume rendering formulation

Calculating T given σ

P[no hits before t] = $T(t)$

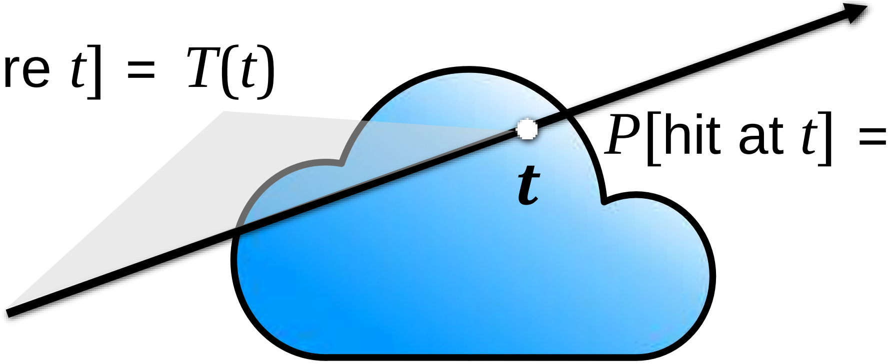


$$T(t) = \exp\left(-\int_{t_0}^{t} \sigma(s)\,ds\right)$$

NeRF

Volume rendering formulation

Probabilistic interpretation



Finally, we can write the probability that a ray terminates at $\boldsymbol{t}$ as a function of only the density σ

$$P[\text{first hit at } t] = P[\text{no hit before } t] \times P[\text{hit at } t]$$

$$= T(t)\sigma(t)dt$$

$$= \exp\left(-\int_{t_0}^{t} \sigma(s)ds\right)\sigma(t)\,dt$$

NeRF

Volume rendering formulation

Expected value of color along ray

This means the expected color returned by the ray will be

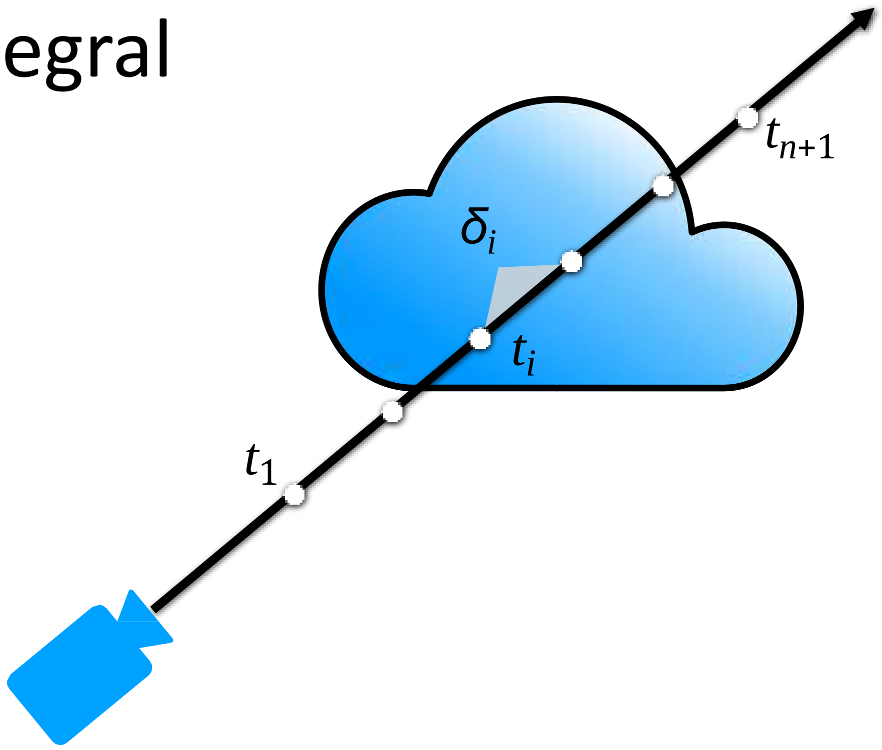
$$\int_{t_s}^{t_e} T(t)\sigma(t)c(t)dt$$

Note the nested integral!

NeRF

Volume rendering formulation

Approximating the integral



Approximate the nested integral,
splitting the ray up into n segments with endpoints $\{t_1, t_2, \ldots, t_{n+1}\}$
with lengths $\delta_i = t_{i+1} - t_i$

NeRF

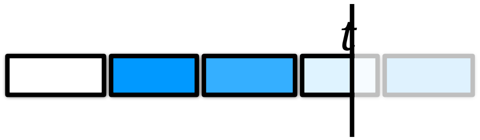
Volume rendering formulation

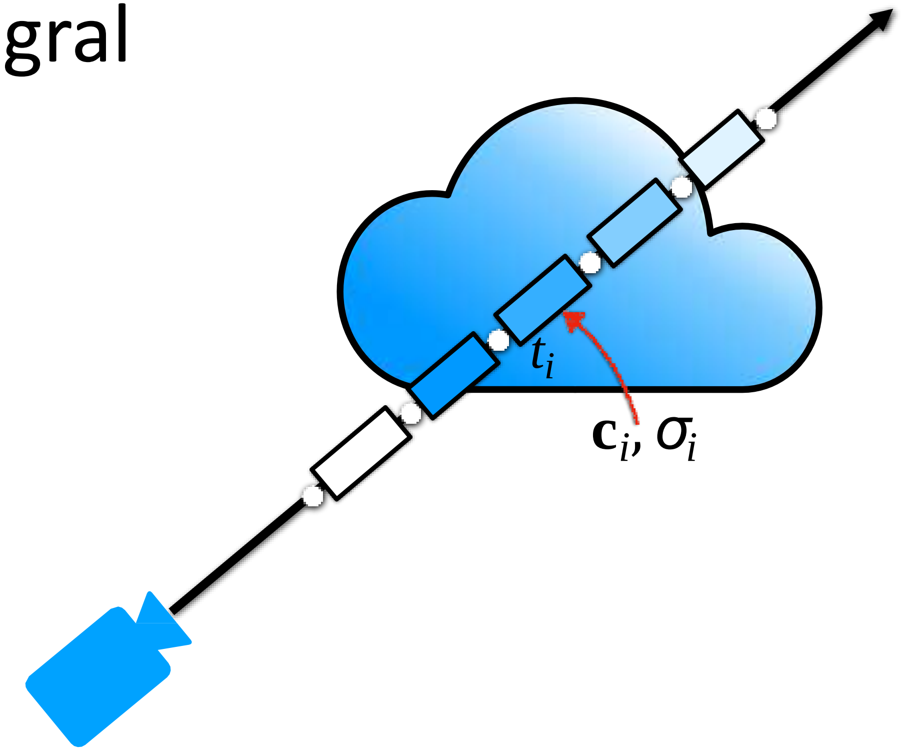
Approximating the integral

Caveat: piecewise constant density and color **do not** imply constant transmittance $T(t)$!

Important to account for how early part of a segment blocks later part when σ_i is high

We need to evaluate at continuous t values that can lie *partway through* an interval





We assume volume density and color are roughly constant within each interval

NeRF

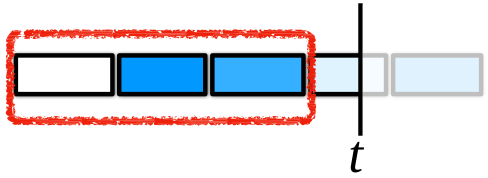
Volume rendering formulation

Evaluating T for piecewise constant density σ

For $t \in [t_i, t_{i+1}]$, $T(t) = \exp\left(-\int_{t_1}^{t_i} \sigma_i \, ds\right) \exp\left(-\int_{t_i}^{t} \sigma_i \, ds\right)$

$$\exp\left(-\sum_{j=1}^{i-1} \sigma_i \delta_i\right) = T_i$$

"How much light is blocked by all previous segments?"



NeRF

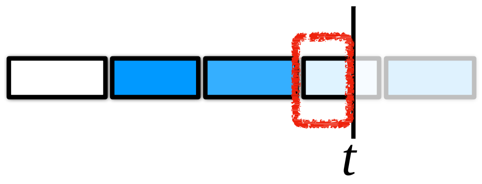
Volume rendering formulation

Evaluating T for piecewise constant density σ

For $t \in [t_i, t_{i+1}]$, $T(t) = \exp\left(-\int_{t_1}^{t_i} \sigma_i \, ds\right) \exp\left(-\int_{t_i}^{t} \sigma_i \, ds\right)$

"How much light is blocked partway through the current segment?"

$\exp(-\sigma_i(t - t_i))$



NeRF

Volume rendering formulation

Approximating the integral

$$\int T(t)\sigma(t)c(t)dt \approx \sum_{i=1}^{n} \int_{t_i}^{t_{i+1}} T(t)\sigma_i c_i dt = \sum_{i=1}^{n} T_i c_i \underbrace{(1 - \exp(-\sigma_i \delta_i))}_{\text{segment opacity } \alpha_i}$$

segment opacity α_i

$$T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_i \delta_i\right) = \prod_{j=1}^{i-1} (1 - \alpha_j)$$

$$\mathrm{C_{ray}} = \sum_{i=1}^{n} T_i \alpha_i c_i$$

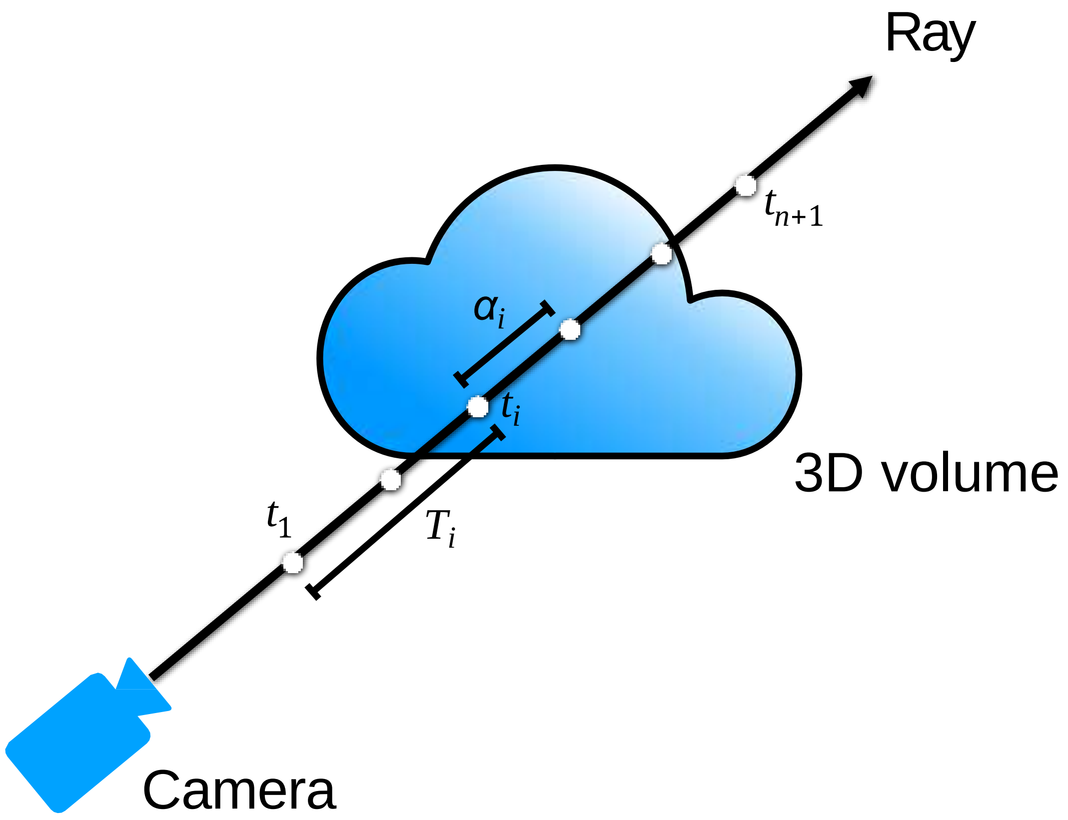
NeRF

Volume rendering formulation

Rendering formulation summary for ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$

$$C_{\text{ray}} = \sum_{i=1}^{n} T_i \alpha_i c_i$$

weights ($T_i \alpha_i$), colors (c_i)



How much light is transmitted earlier along ray:

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

How much light is contributed by ray segment *i*:

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i)$$

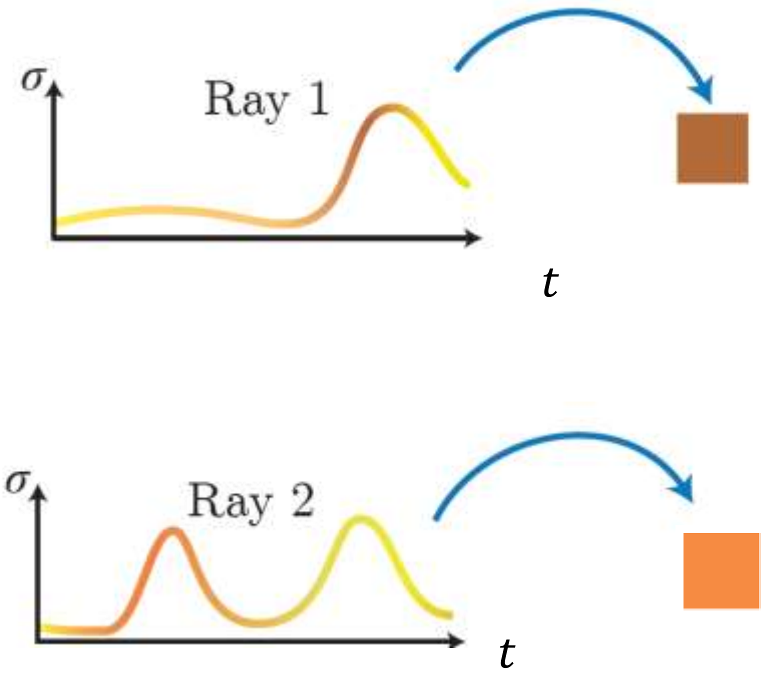
$$\mathrm{C}_{\mathrm{ray}} = \sum_{i=1}^{n} T_i \alpha_i c_i$$

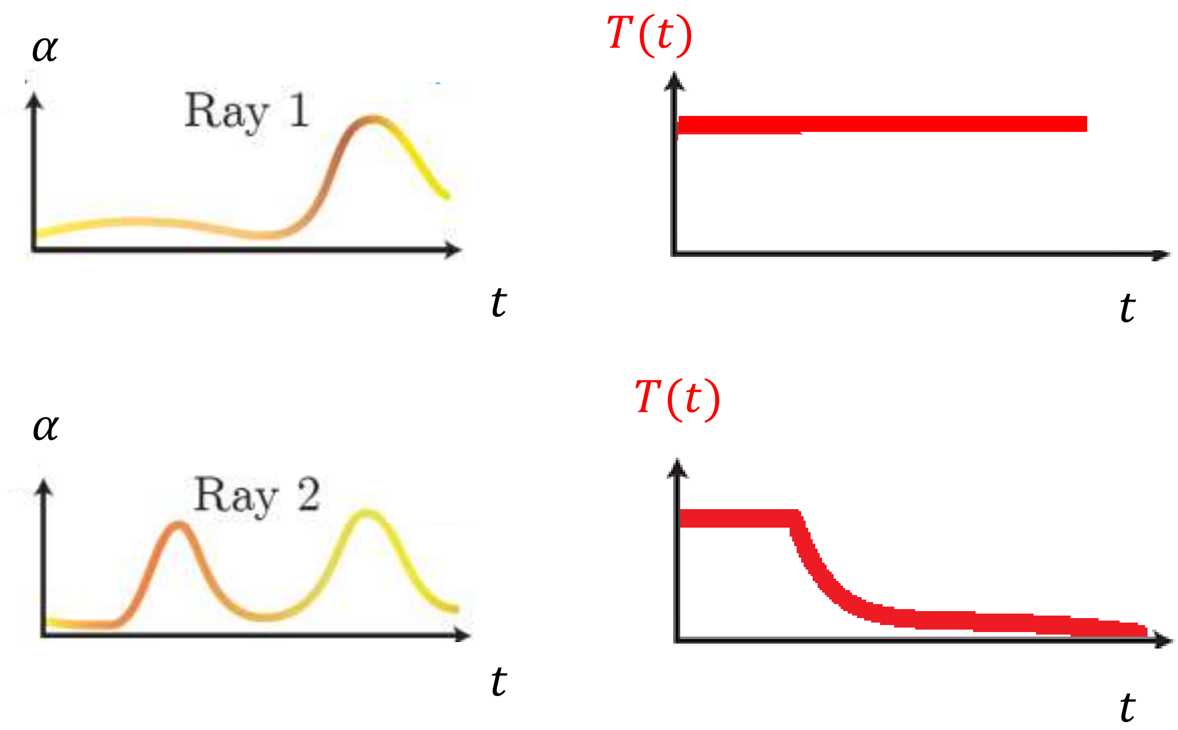
How much light is transmitted earlier along ray:

$$\boldsymbol{T_i} = \prod_{j=1}^{i-1} (\mathbf{1} - \boldsymbol{\alpha_j})$$

How much light is contributed by ray segment ***i***:

$$\alpha_i = 1 - \exp(-\sigma_i \delta_i)$$



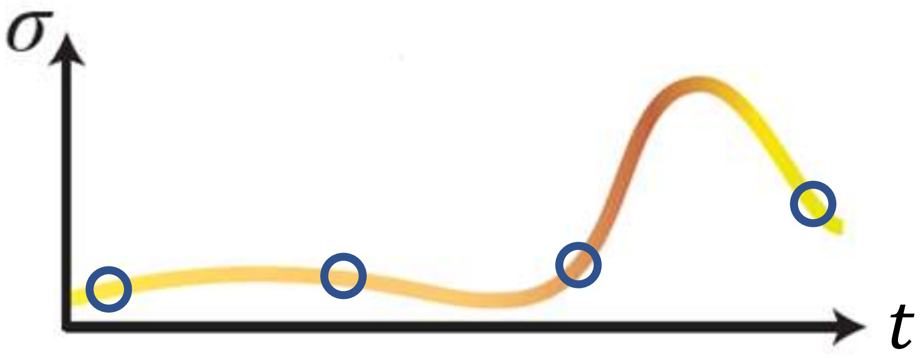


NeRF

Sampling along the ray

Sparse uniform sampling

→ Low accuracy

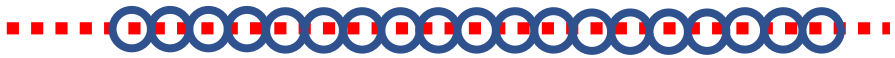


NeRF

Sampling along the ray

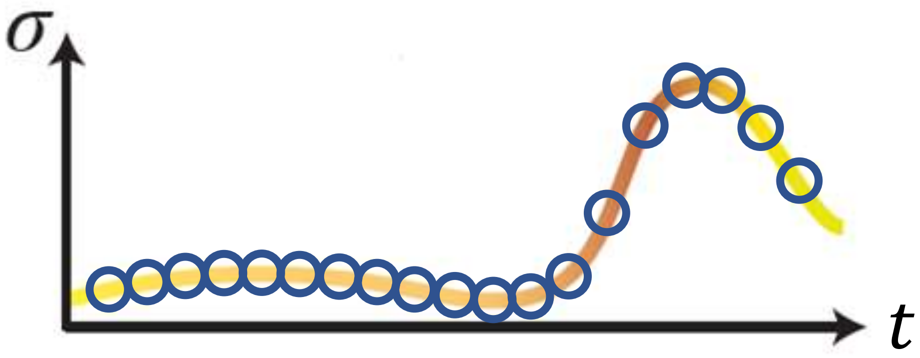
Dense uniform sampling

→ Inefficient



<u>Uniform sampling:</u>

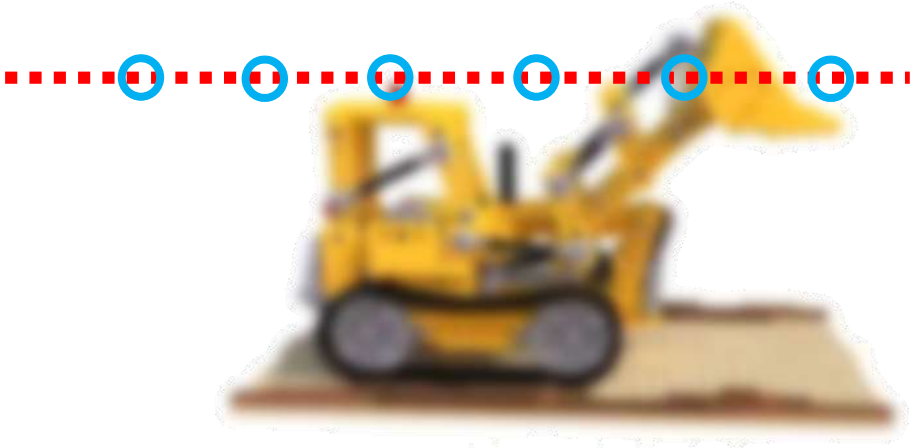
free space and occluded regions that do not contribute to the rendered image are still sampled equally



NeRF

Fine and coarse sampling along the ray

Uniform samples

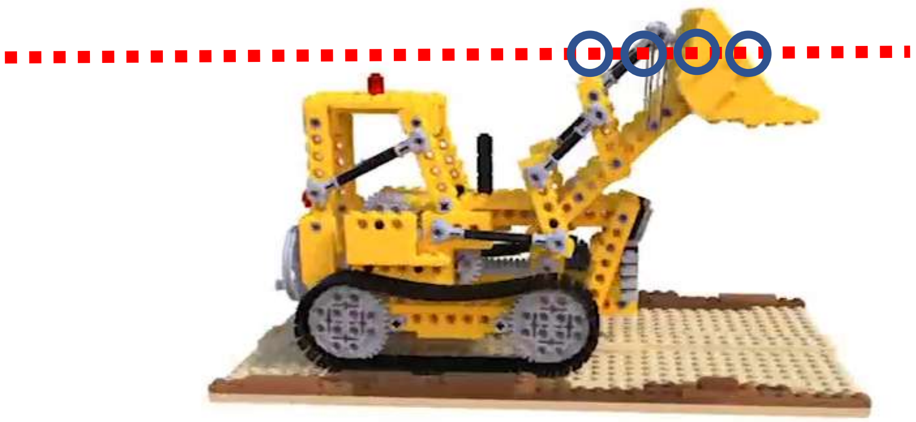


$$(x,y,z,\theta,\phi) \rightarrow [\;] \rightarrow \hat{C}_c, \sigma$$

$F_{\Theta c}$

Coarse NeRF

Non-uniform samples



$$(x,y,z,\theta,\phi) \rightarrow [\;] \rightarrow \hat{C}_f, \sigma$$

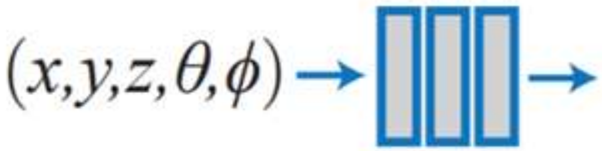
$F_{\Theta f}$

Fine NeRF

Nerf

Fine and coarse sampling along the ray

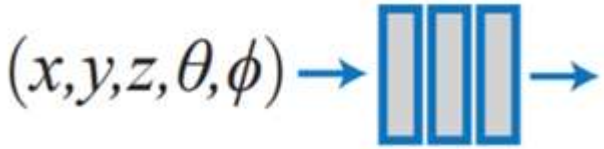
Train two networks



$\hat{C}_c, \sigma$

$F_{\Theta c}$

Coarse NeRF



$\hat{C}_f, \sigma$

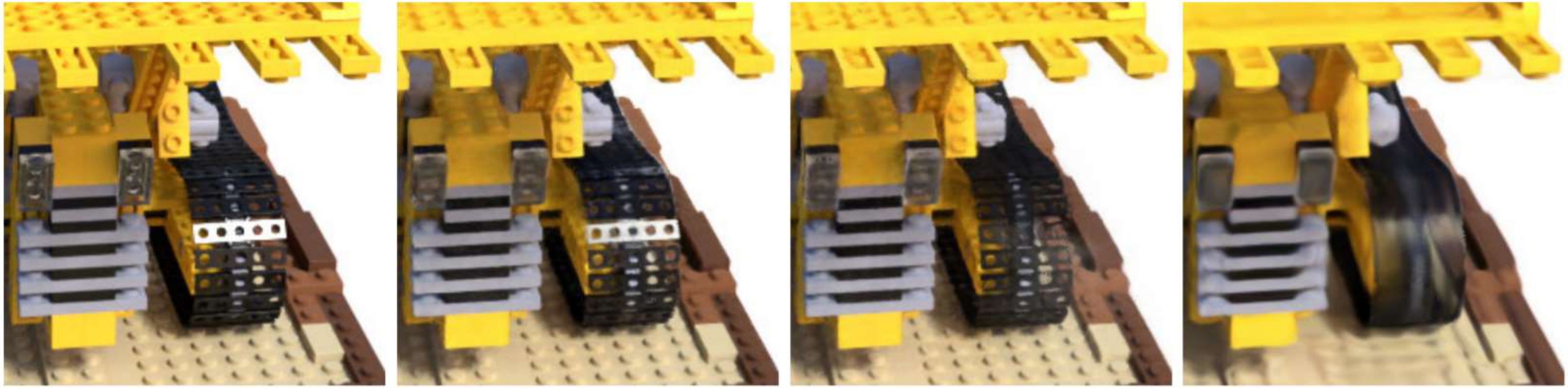
$F_{\Theta f}$

Fine NeRF

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

NeRF

Ablation study



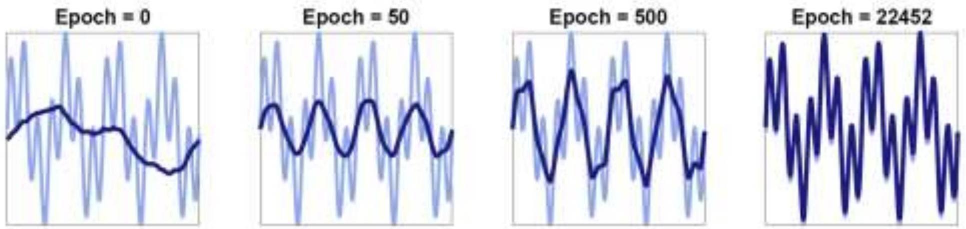
Ground Truth | Complete Model | No View Dependence | No Positional Encoding

NeRF

Positional encoding

Challenge

How to get MLPs converged faster on high-frequency target functions?

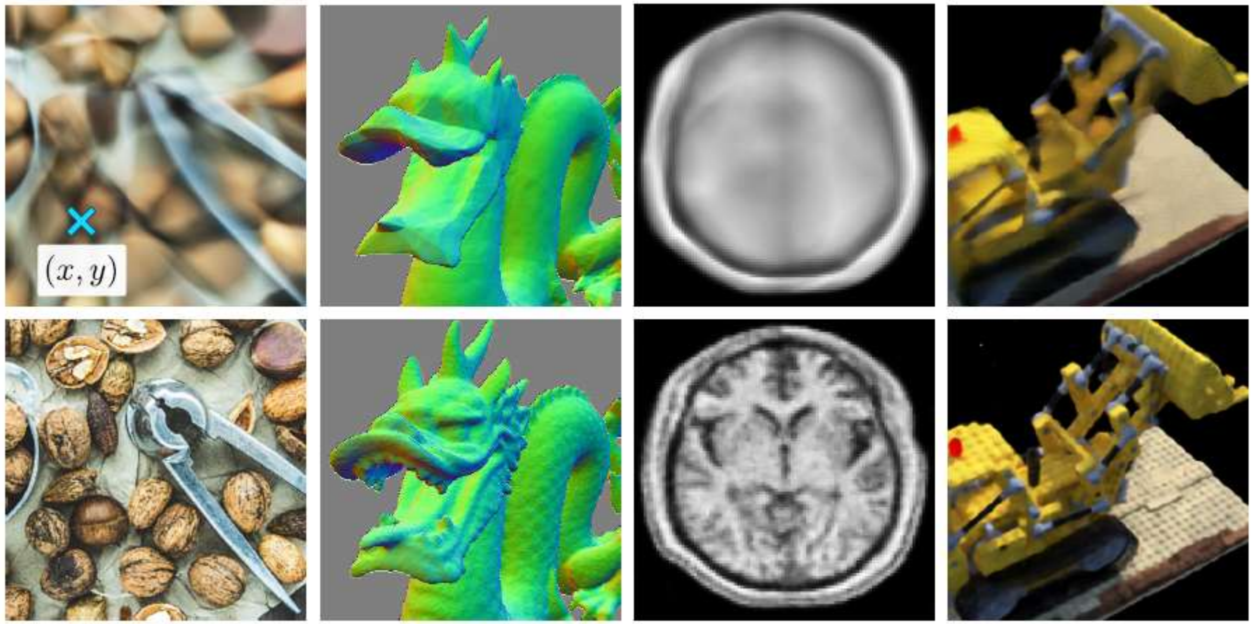


Basri et al., NeurIPS 2019

Tancik et al., NeurIPS 2020

Spectral Bias

FC network fits the lower frequency component of the target function faster than the higher frequencies



(b) Image regression $(x,y) \rightarrow$ RGB

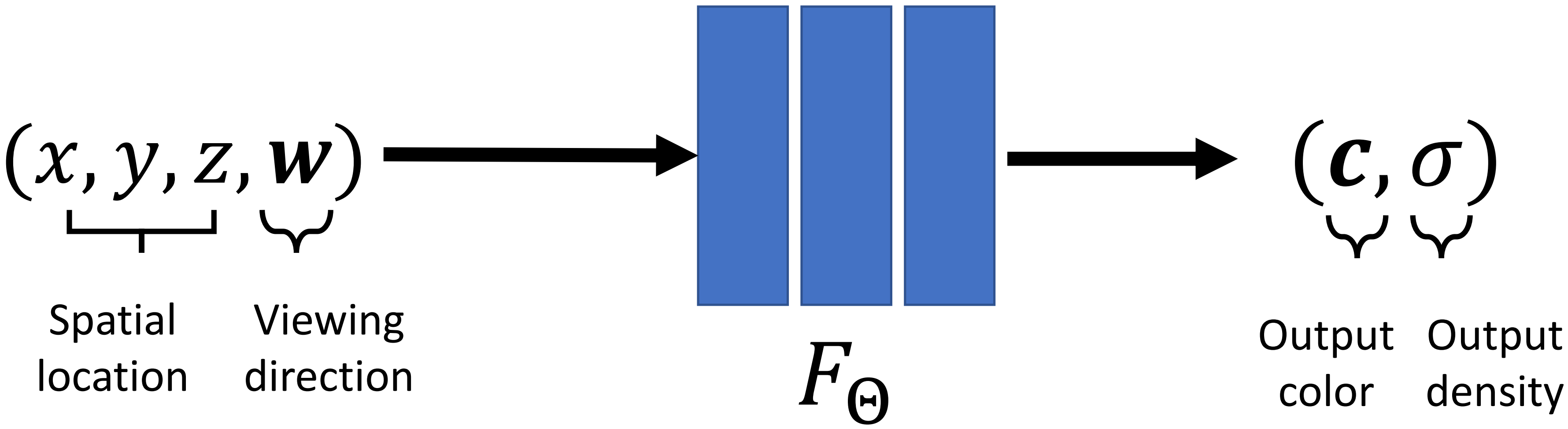
(c) 3D shape regression $(x,y,z) \rightarrow$ occupancy

(d) MRI reconstruction $(x,y,z) \rightarrow$ density

(e) Inverse rendering $(x,y,z) \rightarrow$ RGB, density

NeRF

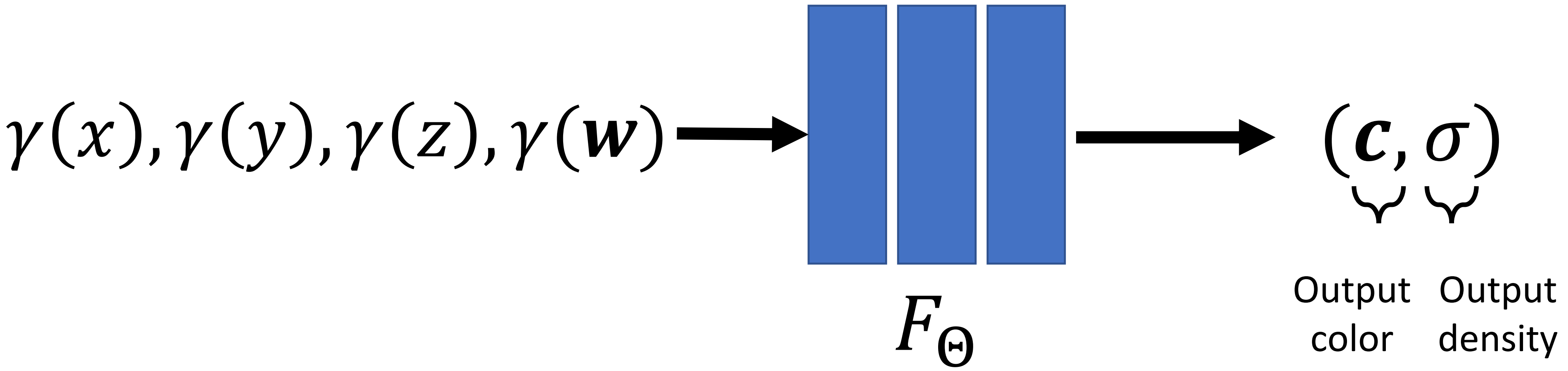
Positional encoding



NeRF

Positional encoding

Introducing positional encoding



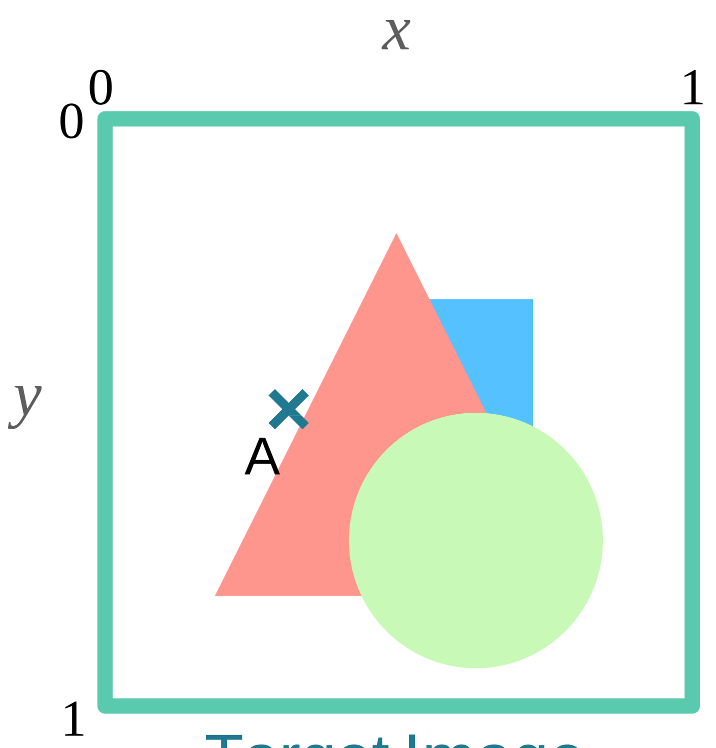
$$^{*}\,\gamma(x) = \left(\sin\left(2^0 \pi x\right), \cos\left(2^0 \pi x\right), \dots, \sin\left(2^{L-1} \pi x\right), \cos\left(2^{L-1} \pi x\right)\right)$$

Why does positional encoding help?

Target Image

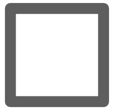
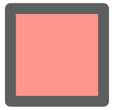
Why does positional encoding help?

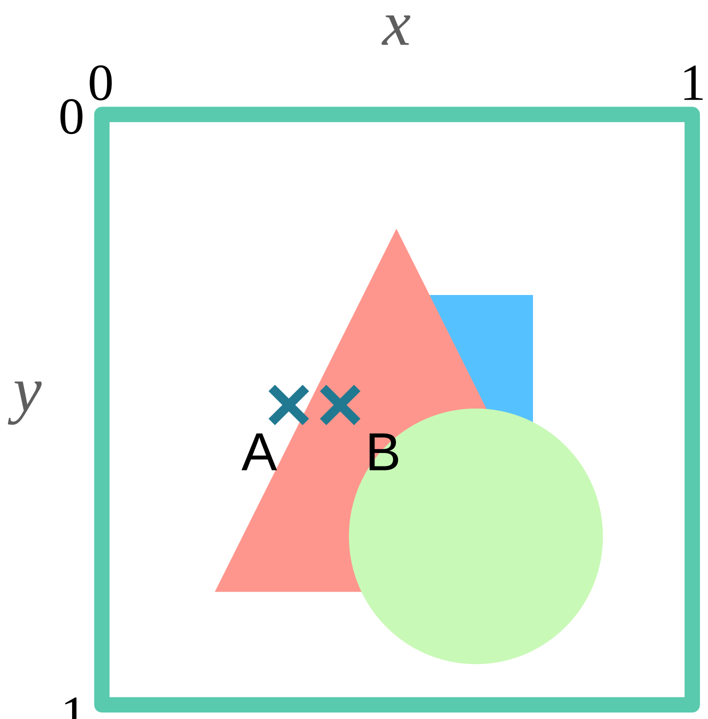
	Input		Target
	x	y	
A	.36	.5	□



Target Image

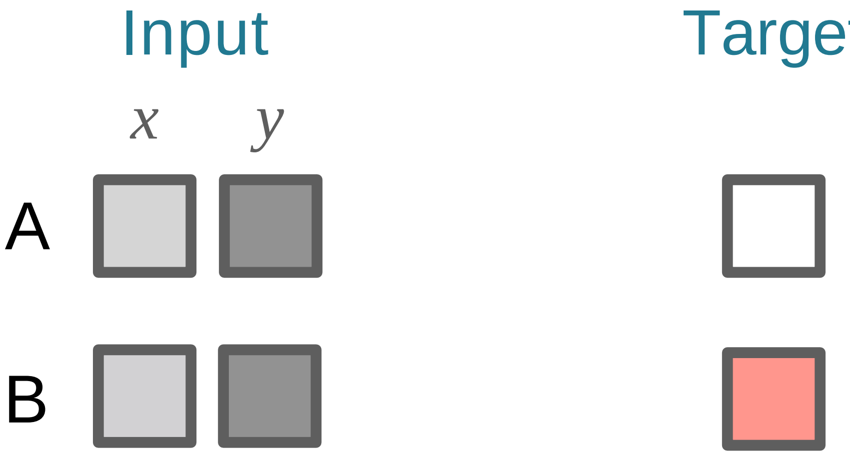
Why does positional encoding help?

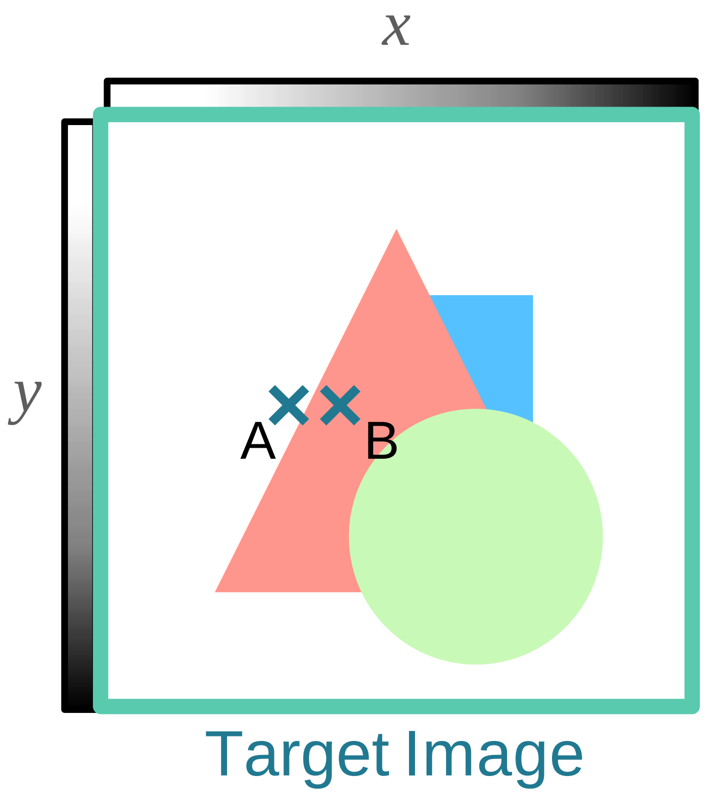
	Input x	Input y	Target
A	.36	.5	
B	.38	.5	



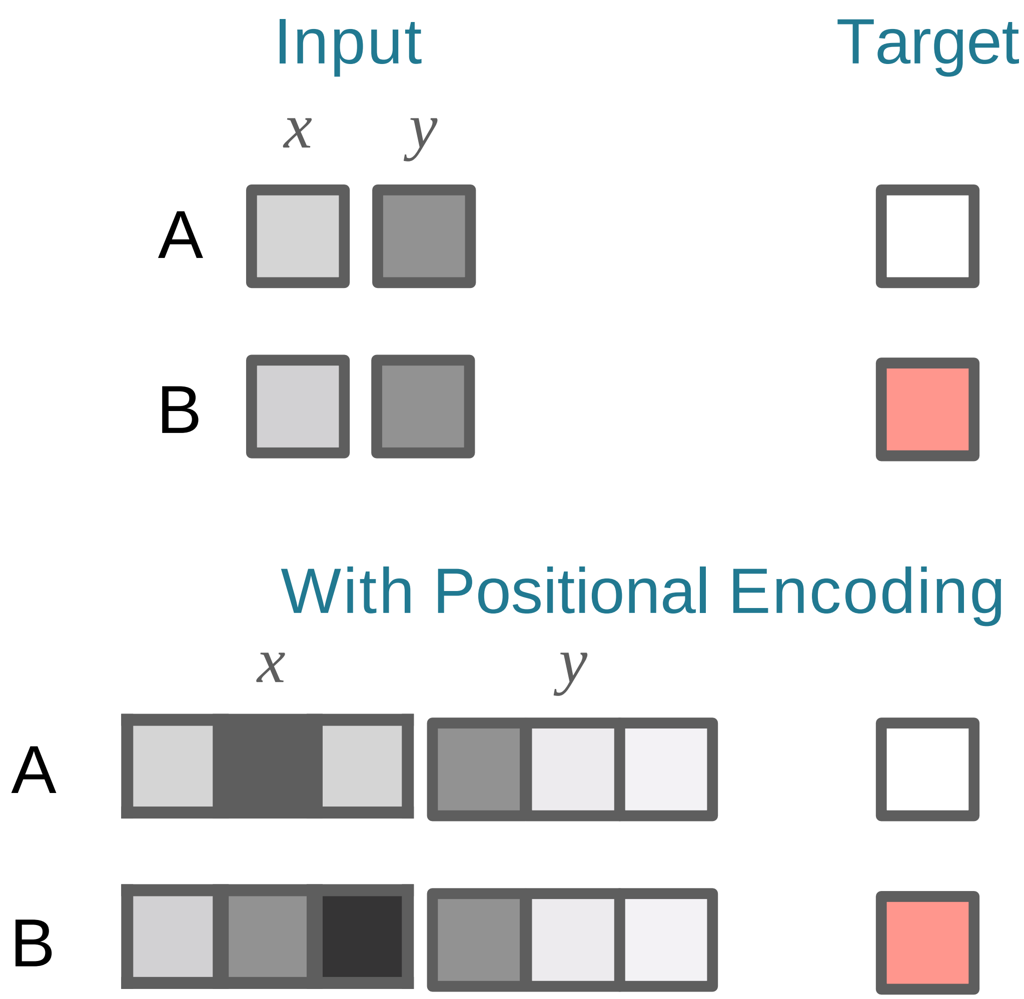
Target Image

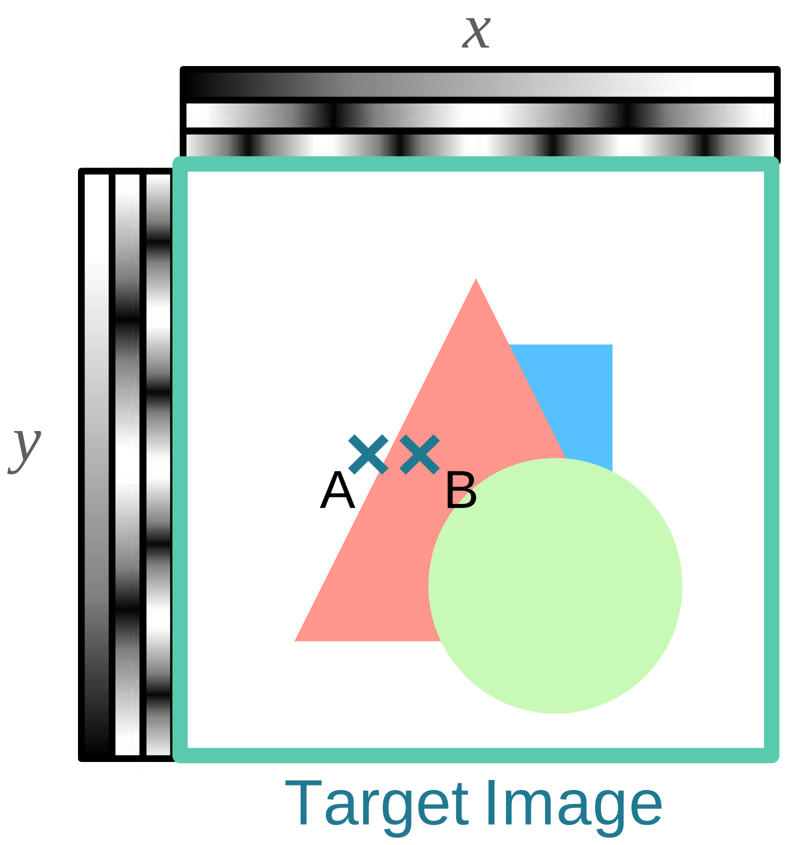
Why does positional encoding help?





Why does positional encoding help?





NeRF

Synthetic scenes

NeRF

Real scenes

SRN [Sitzmann 2019]

NeRF

Nearest Input

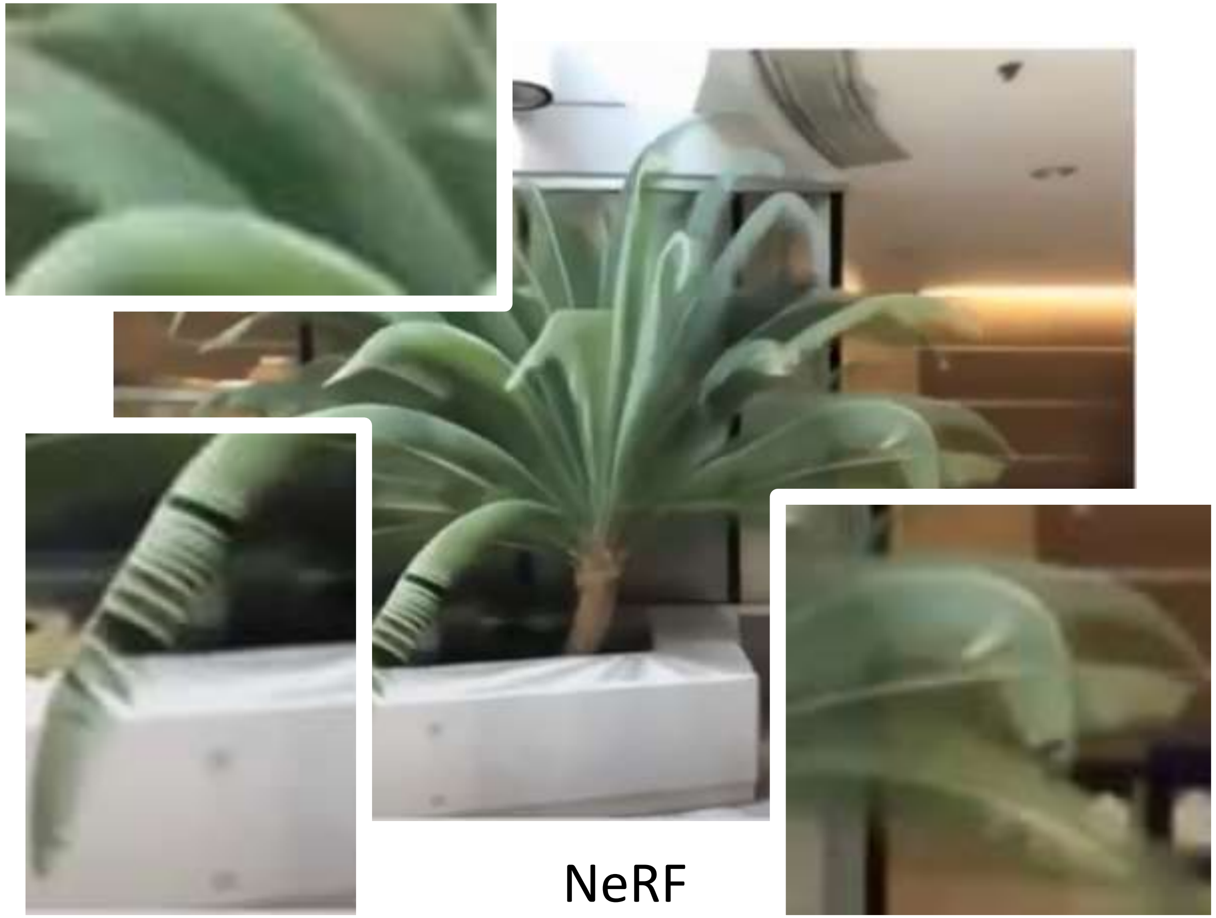
Nerf

NeRF
No positional encoding

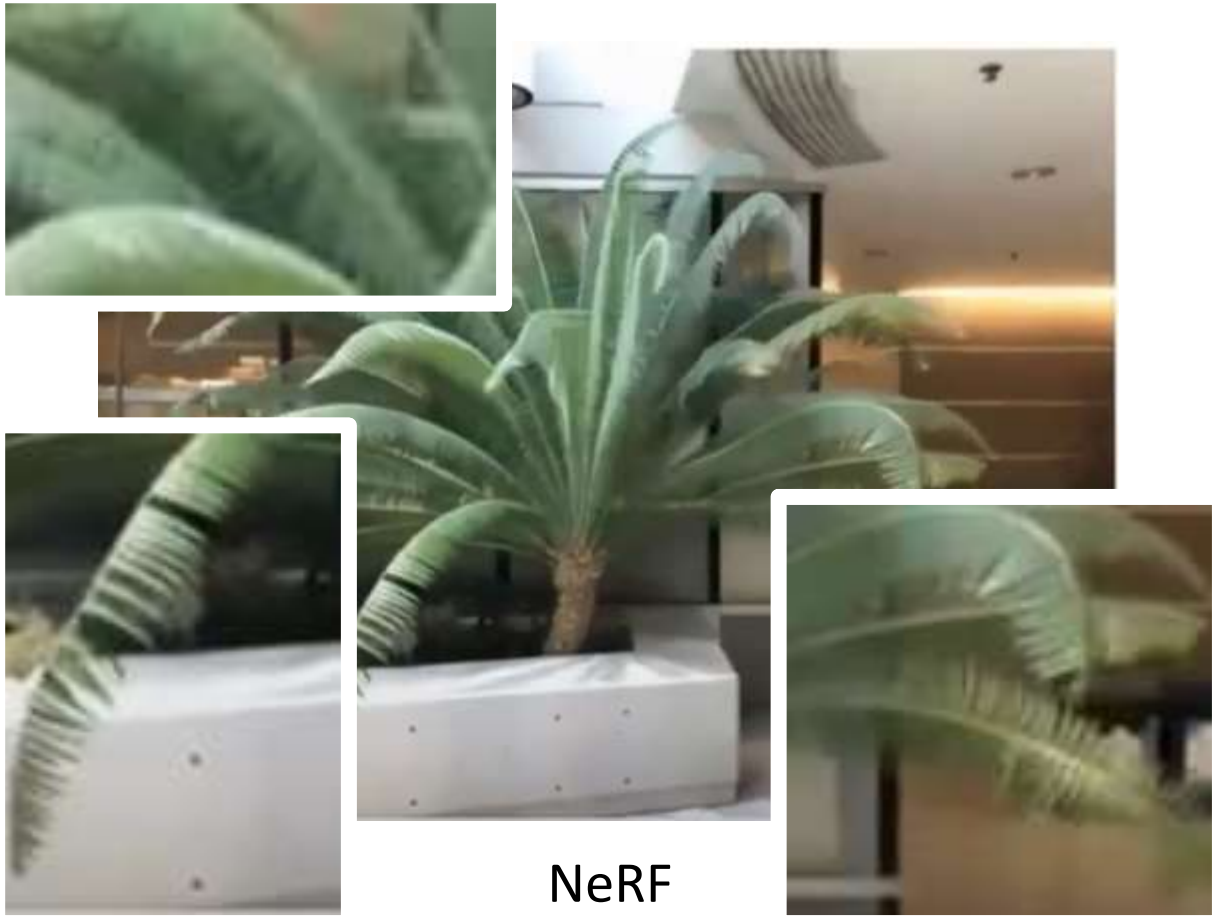
NeRF
With positional encoding

Nerf

Importance of positional encoding



NeRF
No positional encoding

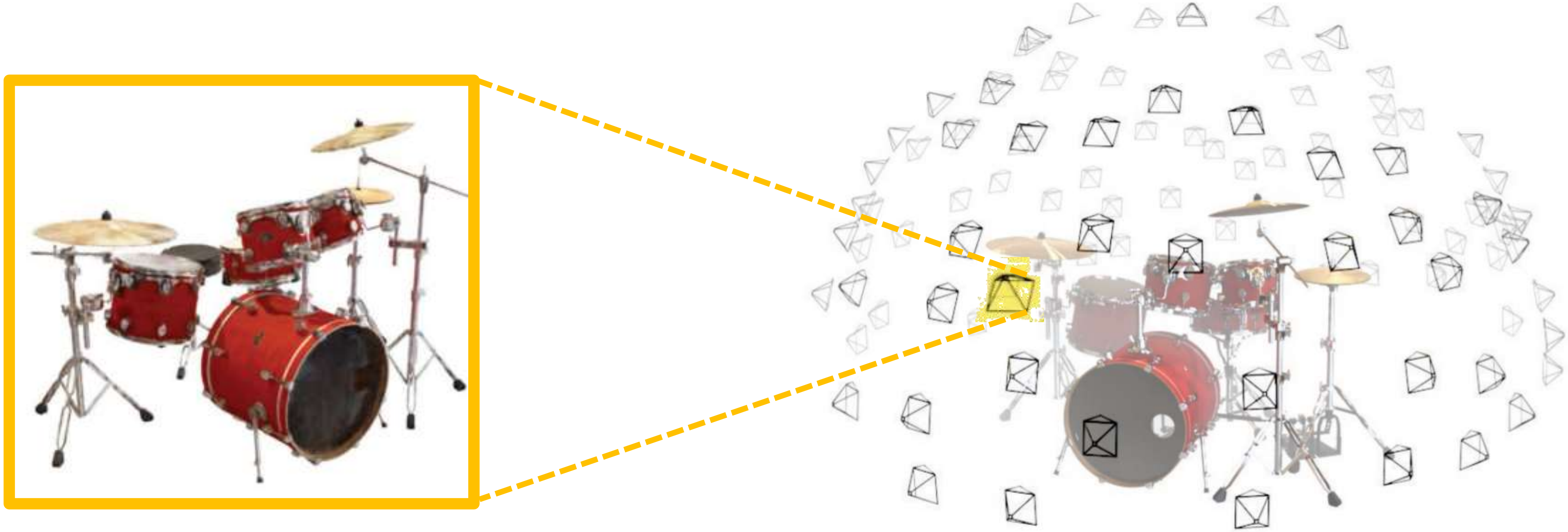


NeRF
With positional encoding

NeRF

Summary

- Novel view synthesis by volume rendering (ray integration)
- Coordinate-base scene representation
- The viewing direction is taken into account
- Encoding the scene in the MLP weights



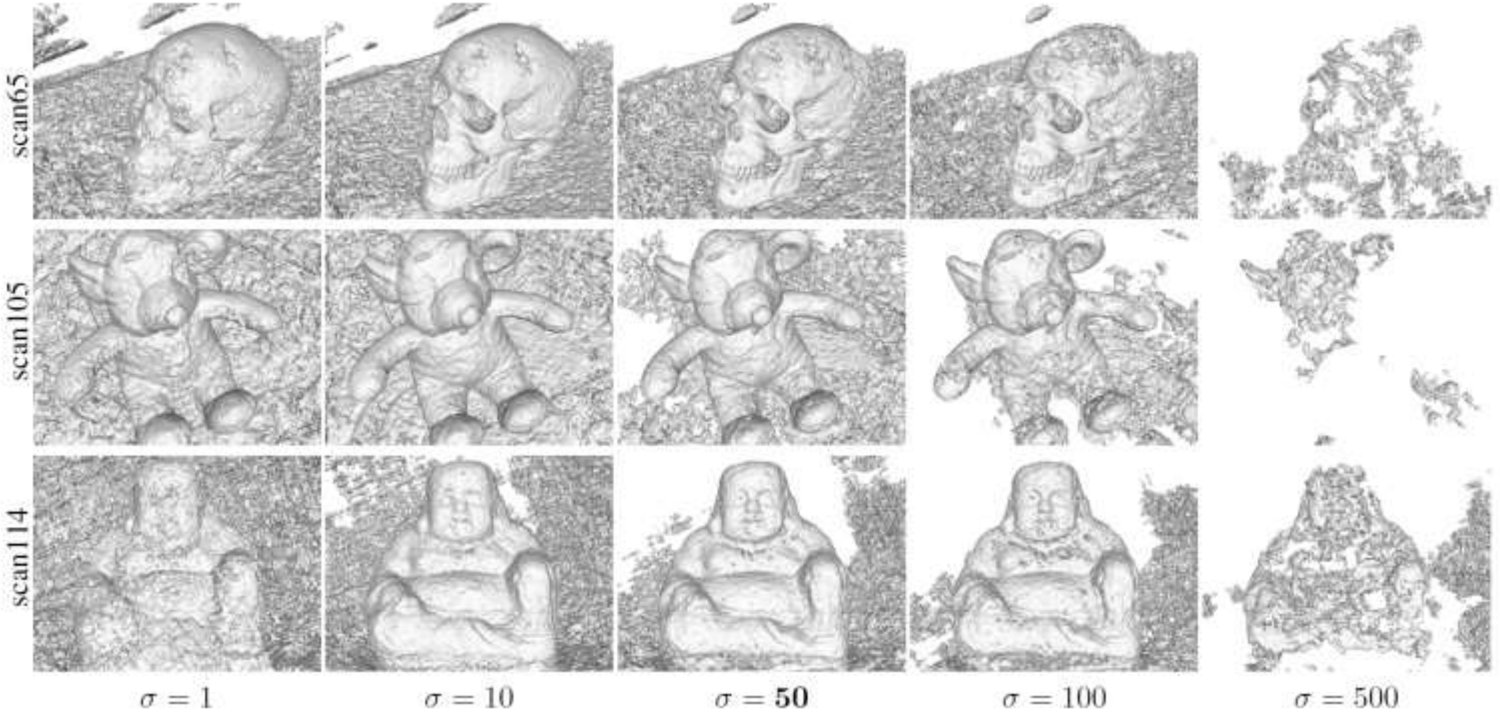
NeRF

Drawbacks / Future directions

- Trained per scene, not generalizable
- Limited by the distribution of cameras on the hemisphere
- Glossy surfaces are not modelled well
- Density, itself, is not enough to represent geometry

Neural rendering

- Representing the surface itself, why?
- Volume rendering or estimation of volume density does not admit accurate surface reconstruction



Volume density thresholds of NeRF

UNISURF: Unifying Neural Implicit Surfaces and Radiance Fields for Multi-View Reconstruction, Oechsle et al., 2021