VaxNeRF

Roger Marí

roger.mari@ens-paris-saclay.fr

MLBriefs presentation day, 12 december 2023

Reminder: NeRF

- What is a **Ne**ural **R**adiance **F**ield (**NeRF**) ?

  - Continuous function f that encodes a static 3D scene. **[Mildenhall et al. 2020]**

  - f is learned by one or more MLPs F_Θ .

$$f:(\boldsymbol{x}, \boldsymbol{d}) \to (\boldsymbol{c}, \sigma)$$

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

Reminder: NeRF

- What is a **Ne**ural **R**adiance **F**ield (**NeRF**) ?

  $$f{:}(\boldsymbol{x}, \boldsymbol{d}) \to (\boldsymbol{c}, \sigma)$$

  - Continuous function f that encodes a static 3D scene. **[Mildenhall et al. 2020]**

  - f is learned by one or more MLPs F_{Θ} .

  - Input: $\boxed{\boldsymbol{x} \to \text{spatial coords of a 3D point}}$ + $\boxed{\boldsymbol{d} \to \text{view-specific features (e.g. viewing dir, embedding vecs)}}$

  - Output: $\boxed{\boldsymbol{c} \to \text{RGB color}}$ + $\boxed{\sigma \to \text{geometry magnitude (e.g. volume density, sdf)}}$

  - *Training* data:

  + *N* camera models (typically pinhole)

N views

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

Reminder: NeRF

$$f:(\boldsymbol{x}, \boldsymbol{d}) \rightarrow (\boldsymbol{c}, \sigma)$$
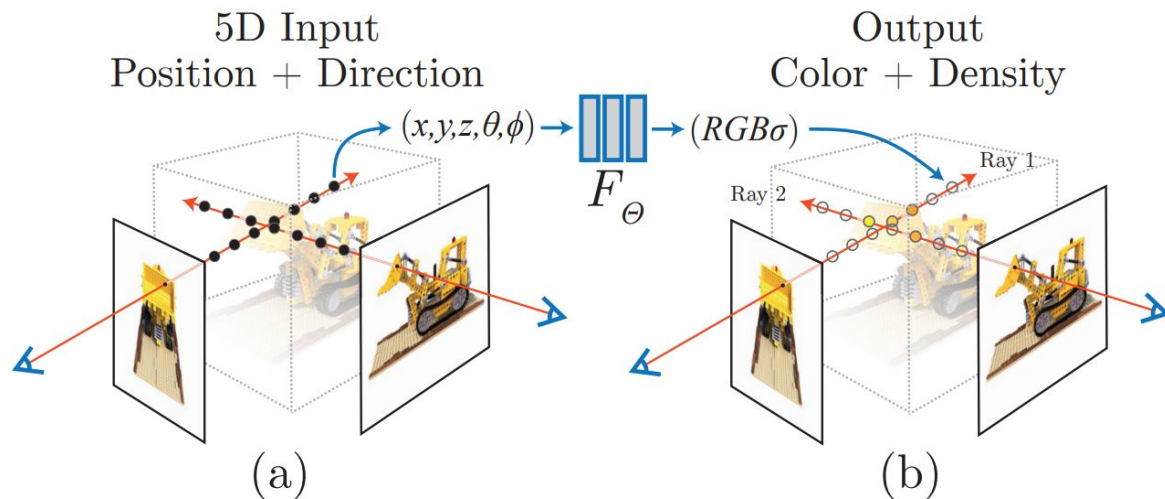
- The NeRF training strategy:

  - Render the color C($\mathbf{r}$) of a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with near and far bounds t_n and t_f:

$$C(\mathbf{r}) = \sum T_i \,\square\, \alpha_i \,\square\, \boldsymbol{c}_i$$

α = opacity = degree in which light can pass through

T = transparency: degree of visibility of the point

5D Input
Position + Direction

Output
Color + Density

$(x,y,z,\theta,\phi) \rightarrow$ F_Θ $\rightarrow (RGB\sigma)$

Ray 1

Ray 2

(a)

(b)

NeRF Reviews

ALLOCINÉ

"IMPRESSIVE QUALITY AND DETAILS"

★★★★★

*Cahiers du Cinéma
(ne pas confondre
avec "Chair du CVPR")*

OFFICIAL SELECTION
FESTIVAL DE CANNES

"TOO SLOW,
I FELL ASLEEP"

★★☆☆☆

Reviewer 2

"YOU WILL LOVE ITS
UNSUPERVISED NATURE"

★★★★☆

Reviewer 1

"UNREALISTIC.
TOO MANY POINTS OF VIEW."

★★★☆☆

Reviewer 3

NeRF Accelerations

1. Convert a NeRF MLP to a different data structure

- *Only improves inference time.*
- *E.g. sparse voxel octrees (PlenOctrees, Neural Sparse Voxel Fields).*

2. Subdivide the scene (2021)

- *Each region of space is assigned an individual MLP*
- *E.g. Uniform or hierarchical subdivisions (KiloNeRF, Recursive-NeRF)*
- *Learnable subdivisions (DeRF, non-regular Voronoi decomposition)*

3. Voxel grids that cache density + learnable features (2022)

- *DVGO or Plenoxels optimize neural features or spherical harmonics for each voxel.*
- *Shallow MLP or none at all.*

4. Multi-resolution hash grids (2022)

- *Instant-NGP. Also for aerial imagery in GridNeRF or Neuralangelo.*

5. Gaussian splatting (2023)

Other tools

JaxNeRF

NerfAcc

Nerfstudio

The VaxNeRF Method (Voxel-Accelerated NeRF)

The NeRF search space can be reduced by exploiting the **visual hull**.

Initial volume

Visual hull

Silhouette back-projection

Camera frustum intersections

foreground

background
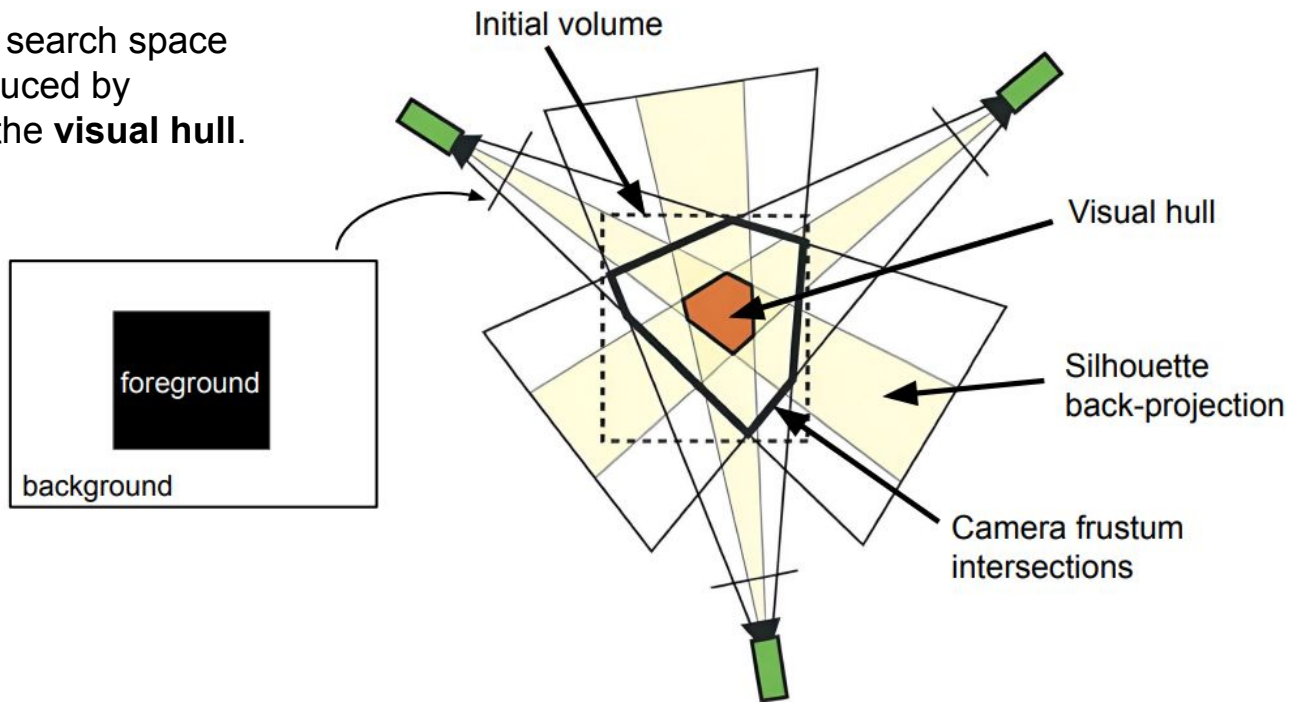
The visual hull is the intersection of the set of foreground silhouettes back-projected into the 3D space. Toy example using 3 input cameras.
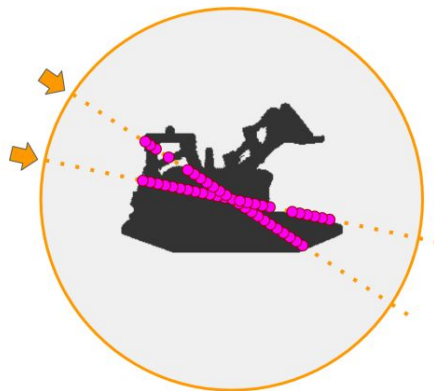
The VaxNeRF Method (Voxel-Accelerated NeRF)

VaxNeRF uses this insight to dramatically reduce the optimization time. **[Kondo et al. 2021]**

(a) Input   (b) Visual hull $\mathcal{V}$ construction   (c) Output
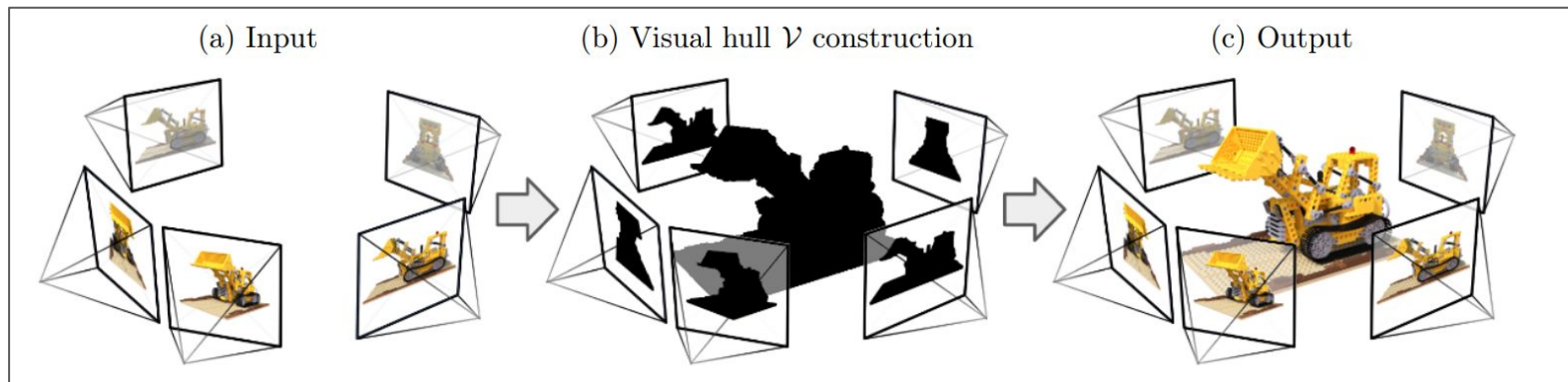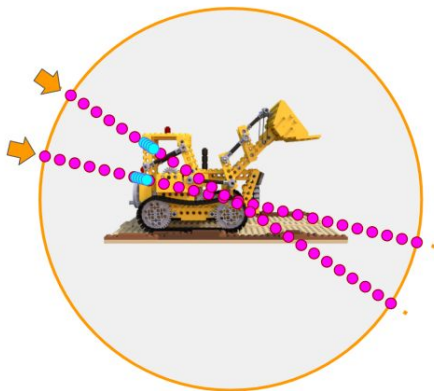
NeRF
(2020)

VaxNeRF
(2021)

Kondo, Naruya, et al. "Vaxnerf: Revisiting the classic for voxel-accelerated neural radiance field." *arXiv preprint* (2021).

The VaxNeRF Method (Voxel-Accelerated NeRF)

VaxNeRF Advantages

- *Integration is very simple. Requires minimal code modifications.*

- *Training is 2-8x faster with respect to the original NeRF.*

- *No loss of quality in the output.*

- *Outperforms coarse-fine sampling strategies using 2 MLPs.*

VaxNeRF Disadvantages

- *Foreground-background segmentation may not be feasible or straight-forward.*

- *A 3D array to represent a visual hull requires more memory than an MLP.
  (e.g. 400x400x400 voxels ~70MB w.r.t. ~7MB for an MLP)*

- *Gain of speed depends on the visual hull size.*

Misleading name issue: the name voxel-accelerated NeRF (VaxNeRF) may not be the most descriptive. The visual hull could be represented in a format other than a voxel grid.

Results

PSNR vs optimization time
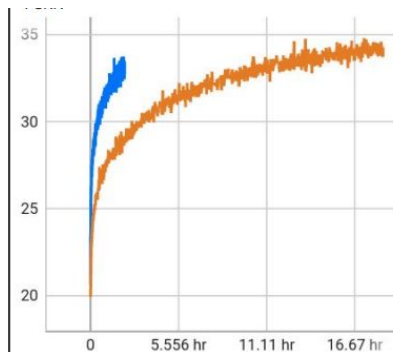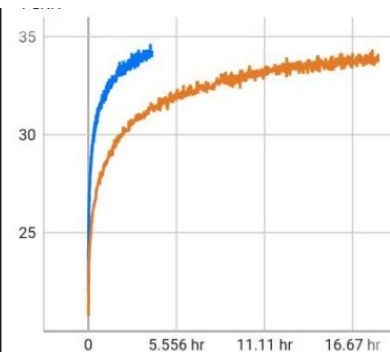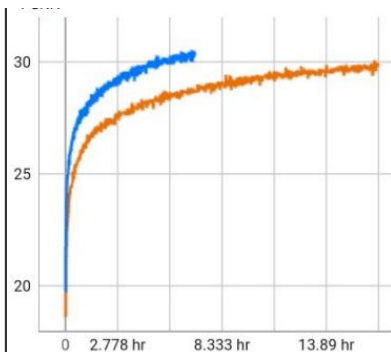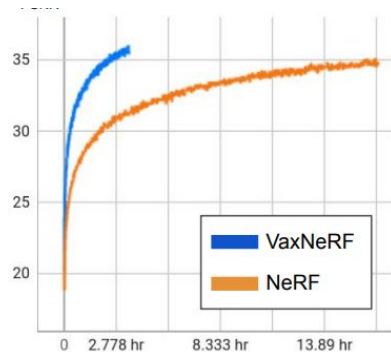
Visual hull
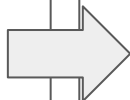(~10 s to
compute)

Results

"Lego" scene: 100 training views of 800x800px

Visual hull:

< 1 min

VaxNeRF:

5 min

Results

"Ficus" scene: 100 training views of 800x800px

Visual hull:

< 1 min

VaxNeRF:
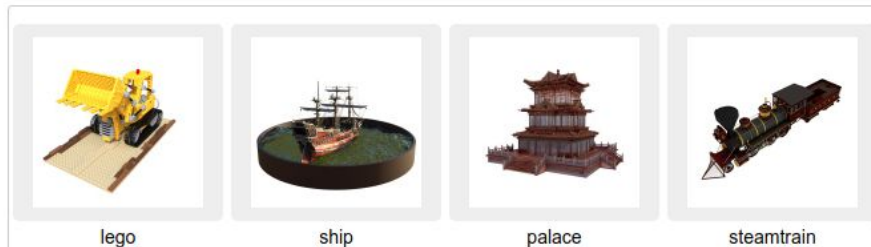

5 min

Results

Visual Hull

0.5min   1min   5min

NeRF

VaxNeRF

IPOL Demo

Generate your own custom novel views from pretrained models

14