Panoptic Neural Field

Group 04: Ajvi Mece, Boran Sivrikaya, Cavit Cakir, Halil Ibrahim Orhan

Supervisors: Markus Herb, Artem Savkin

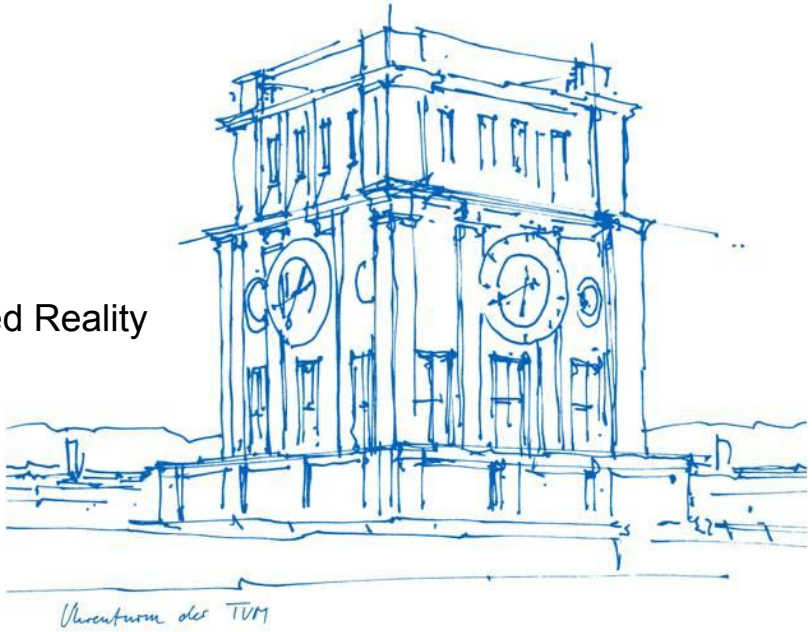
Client: Nikolas Brasch

Technische Universität München

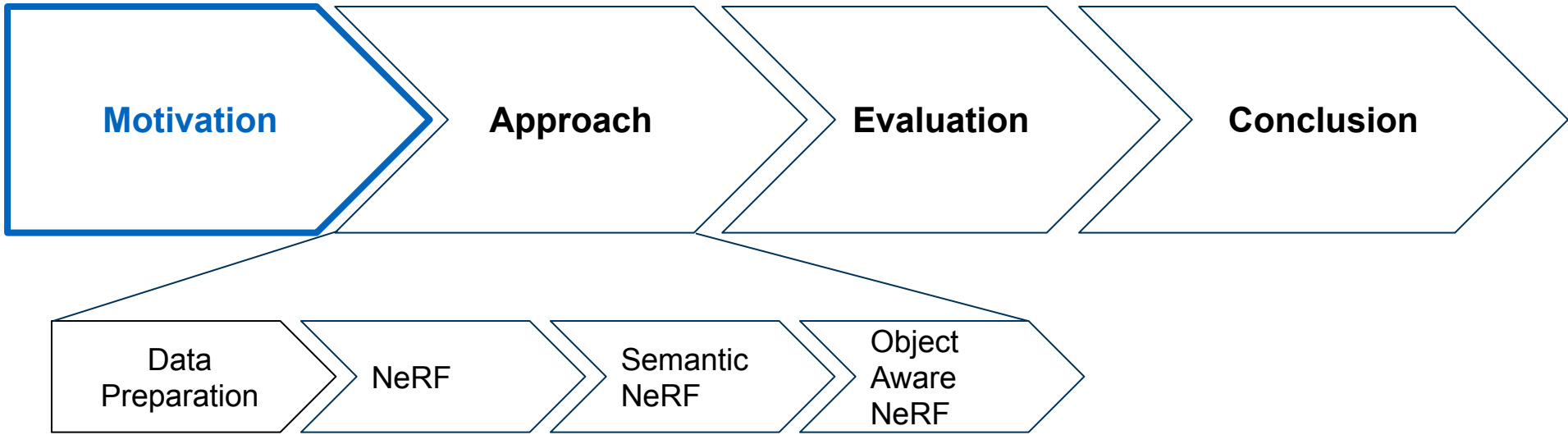
TUM School of Computation, Information and Technology

Chair for Computer Aided Medical Procedures & Augmented Reality

Munich, 03. February 2023



Agenda


Motivation
Approach
Evaluation
Conclusion
Data Preparation
NeRF
Semantic NeRF
Object Aware NeRF

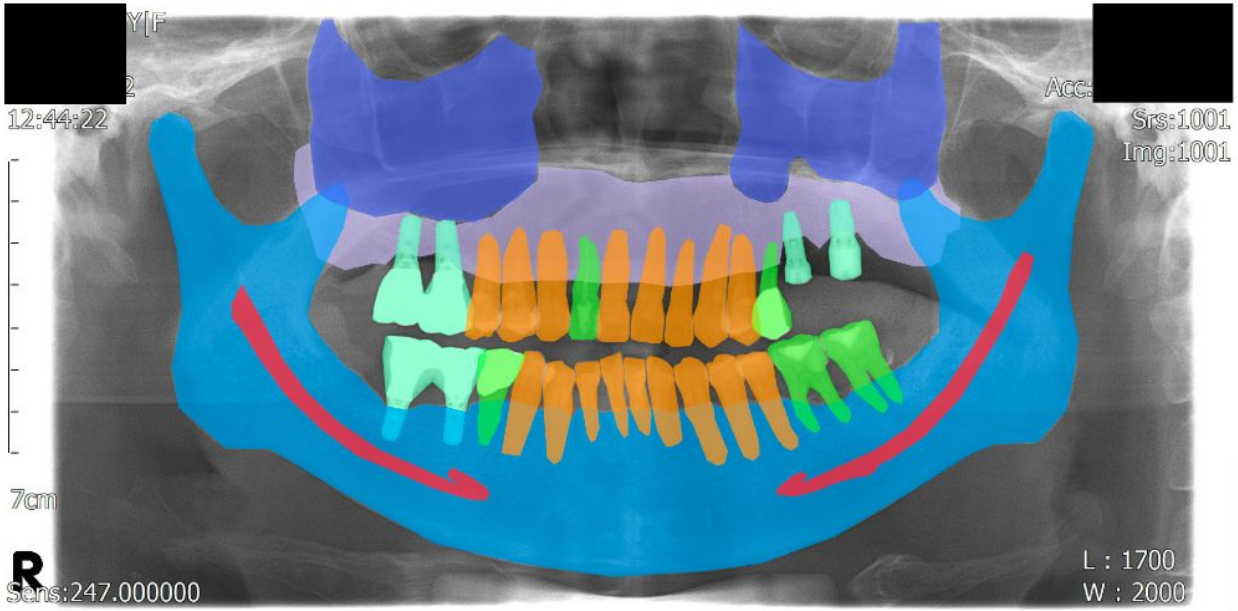
Chen, Zhiqin, et al. "Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures." (2022)

Chen, Zhiqin, et al. "Mobilenerf: Exploiting the polygon rasterization pipeline for efficient neural field rendering on mobile architectures." (2022)

Panoptic Reconstruction in Autonomous Driving

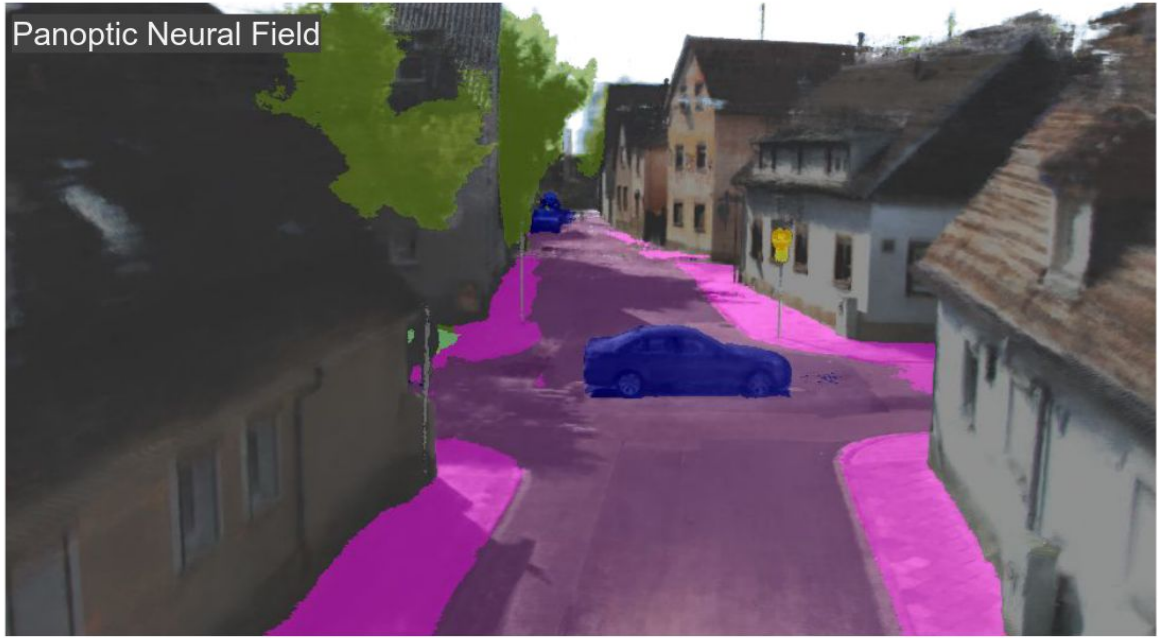
Xiong, Yuwen, et al. "Upsnet: A unified panoptic segmentation network." *CVPR* 2019.

Panoptic Reconstruction in Medical Imaging



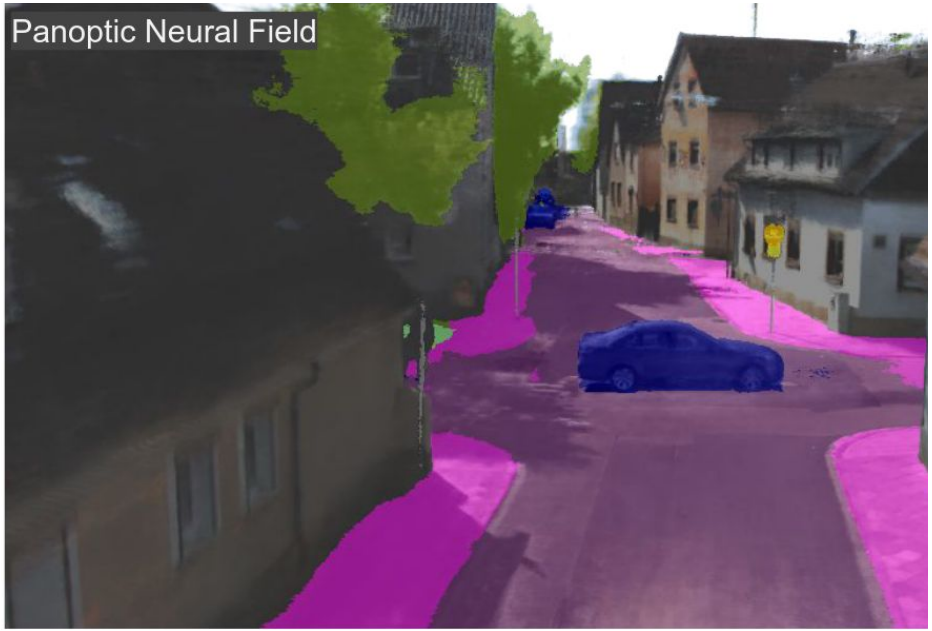
Cha, J.-Y.; Yoon, et al. “Panoptic Segmentation on Panoramic Radiographs: Deep Learning-Based Segmentation of Various Structures Including Maxillary Sinus and Mandibular Canal.” *J. Clin. Med.* **2021**

Panoptic Neural Field



Kundu, Abhijit, et al. "Panoptic neural fields: A semantic object-aware neural scene representation." *CVPR* 2022.

Panoptic Neural Fields

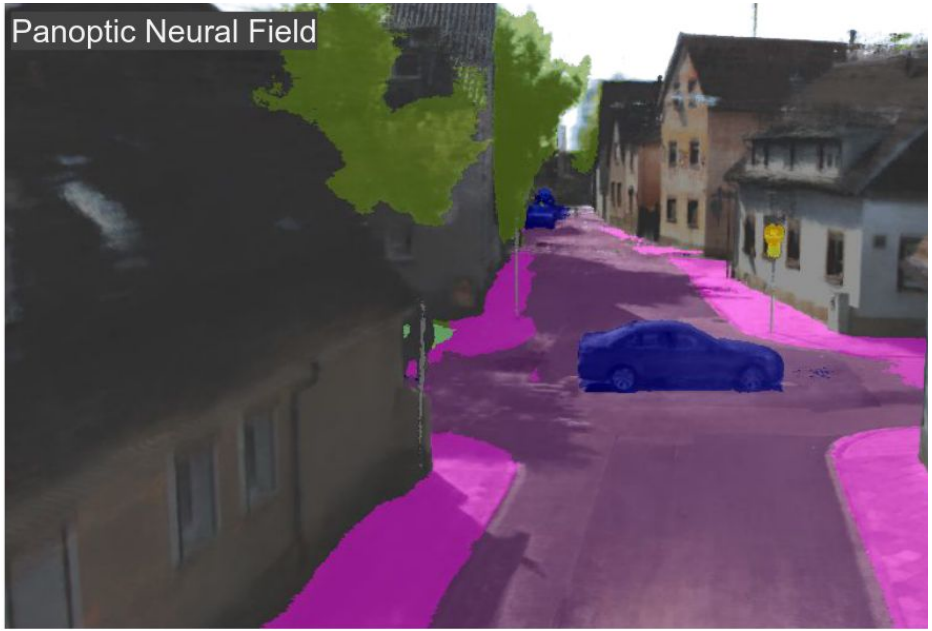


Idea

- NeRF
- NeRF with Object Decomposition
- NeRF with Semantics
- NeRF with Dynamics

Kundu, Abhijit, et al. "Panoptic neural fields: A semantic object-aware neural scene representation." *CVPR* 2022.

Panoptic Neural Fields

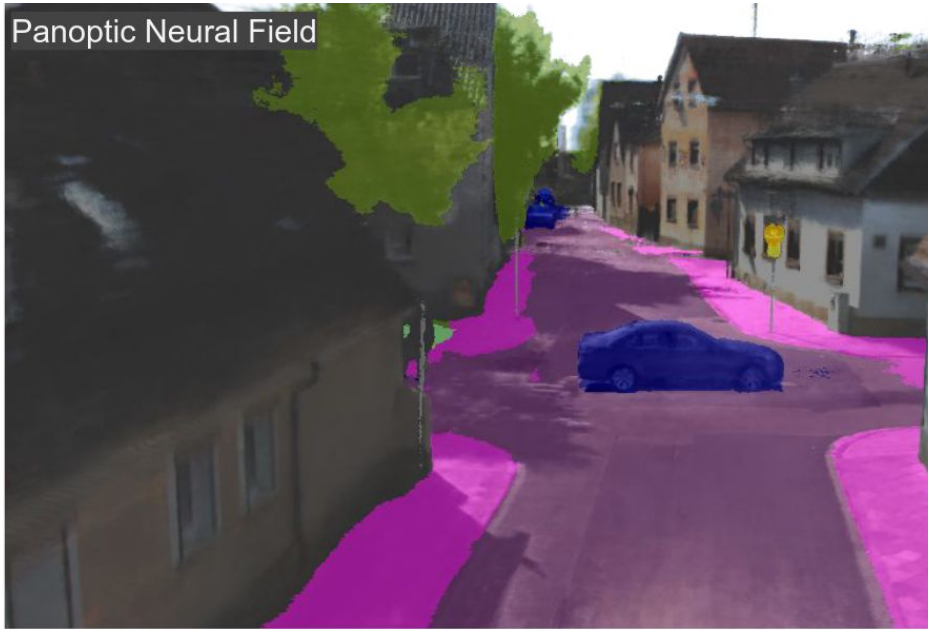


State of the Art

Paper	Sem	Obj	Pan	Dyn	Opt	Syn
MeshRCNN [16]		✓				
Total3D [39]	✓	✓				
Atlas [36]	✓					
SLAM++ [50]		✓.				
PanopticFusion [37]	✓		✓			
Kimera [48]	✓					
DynSceneGraphs [49]	✓	✓	✓	✓	✓	
SemanticNerF [66]	✓					✓
NSG [40]		✓		✓		✓
ObjectNeRF [61]		✓				✓
PNF (Ours)	✓	✓	✓	✓	✓	✓

Kundu, Abhijit, et al. "Panoptic neural fields: A semantic object-aware neural scene representation." *CVPR* 2022.

Panoptic Neural Fields

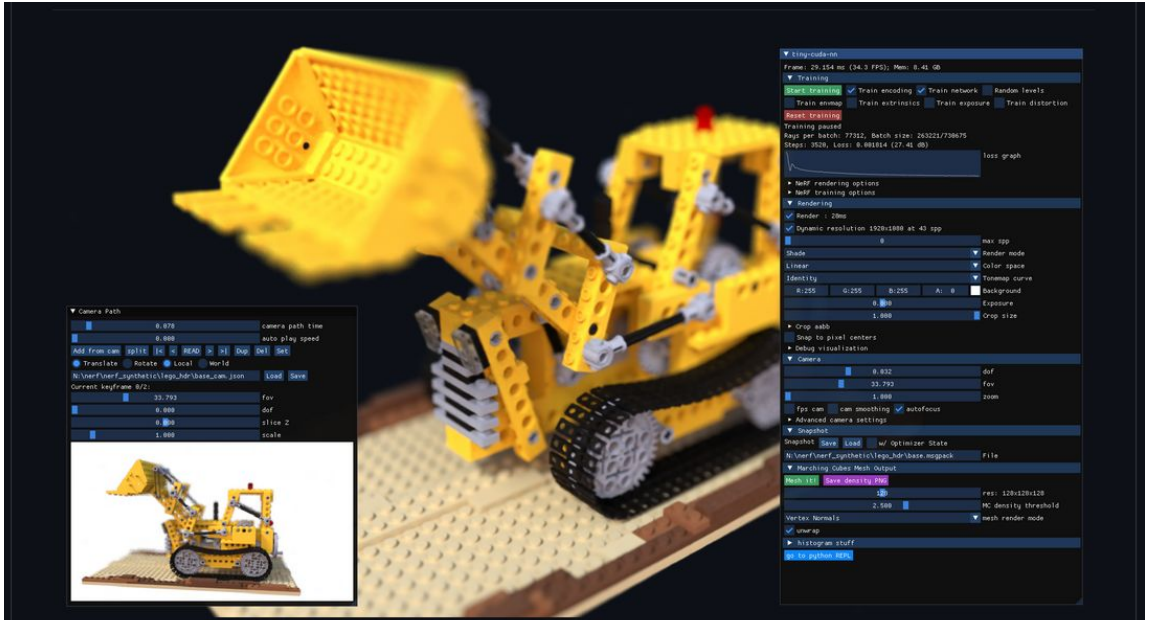


Problem

“Like most other NeRF-style methods, our model is compute-intensive and hence currently only suited for offline applications.”

Kundu, Abhijit, et al. "Panoptic neural fields: A semantic object-aware neural scene representation." *CVPR* 2022.

Instant NGP for Fast NeRF Training



Müller, Thomas, et al. "Instant neural graphics primitives with a multiresolution hash encoding."
ACM ToG 2022

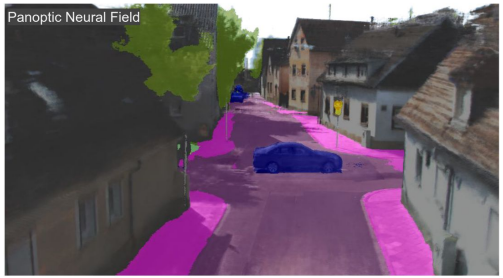
Goal of the Project

★ Use **Instant NGP** to implement a **Neural Panoptic Field**

- **Given**
  - Groundtruth Images
  - Groundtruth Poses
  - Groundtruth Semantic Labels
  - Groundtruth Bounding Boxes
- **Tools**
  - Instant NGP
  - Kaolin WISP

Approach

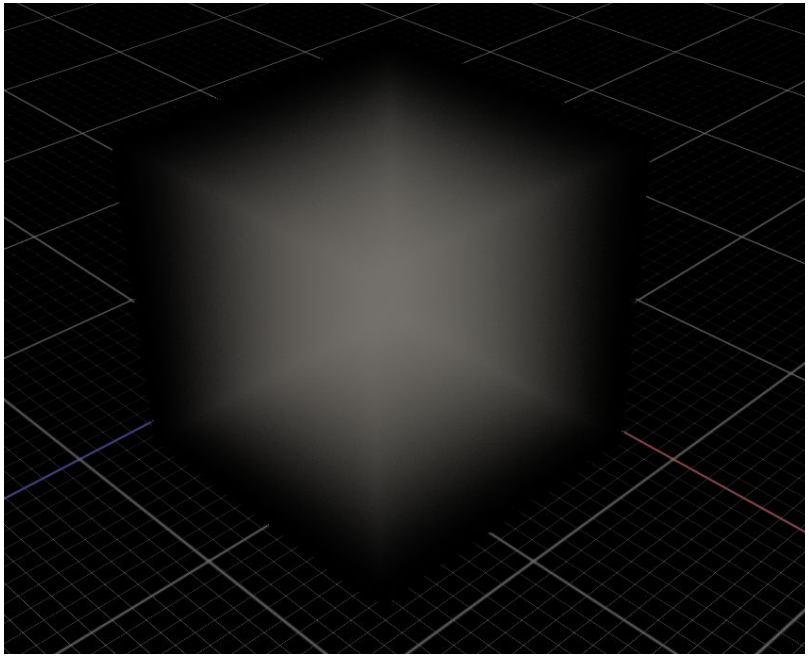

Data Preparation
NeRF
Semantic NeRF
Object Aware NeRF


Panoptic Neural Field

Data Preparation

- **Given**
  - Groundtruth Images
  - Groundtruth Poses
  - Groundtruth Semantic Labels
  - Groundtruth Bounding Boxes
- **Tool**
  - **Instant NGP**

Load them according to
the assumptions of the framework!

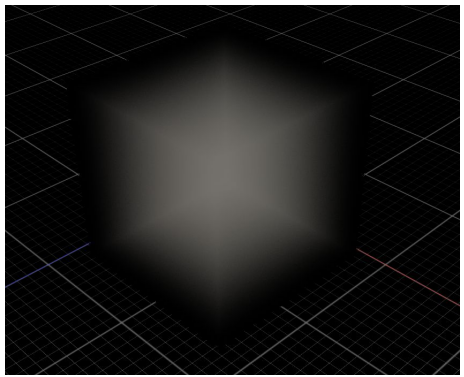


Unit Cube where the scene should be centered

Data Preparation

- **Problem**
  - We have to scale and offset the scene onto the unit cube

➡ Offset and scale **the poses**



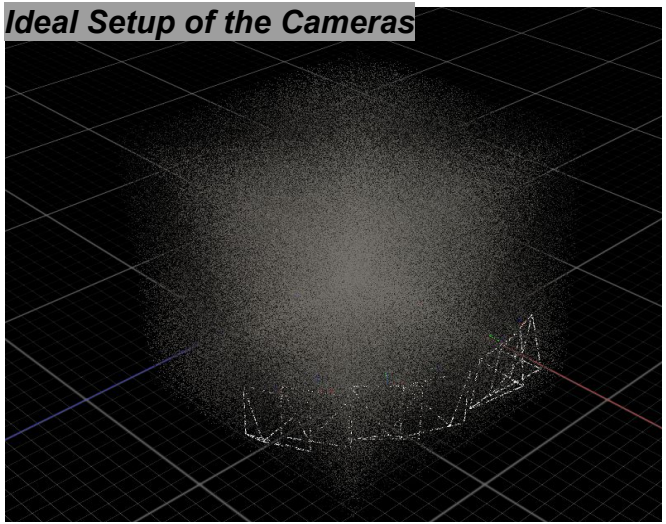
Unit Cube

Our scene

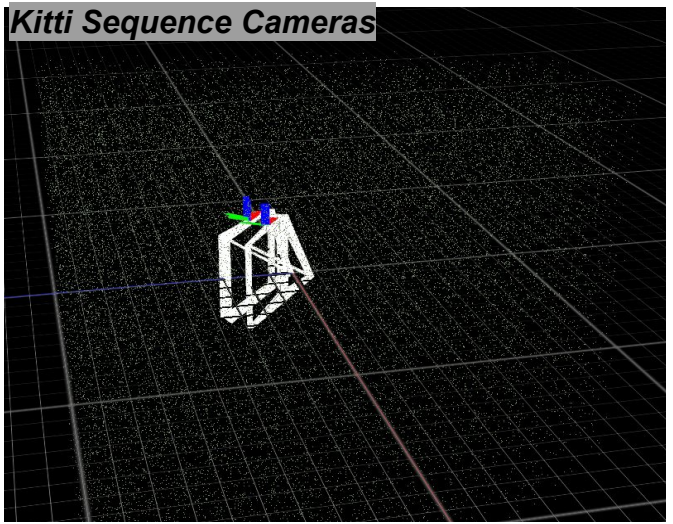
Data Preparation: Poses

- **Challenges**
  - Cameras only moving in one direction

Ideal Setup of the Cameras

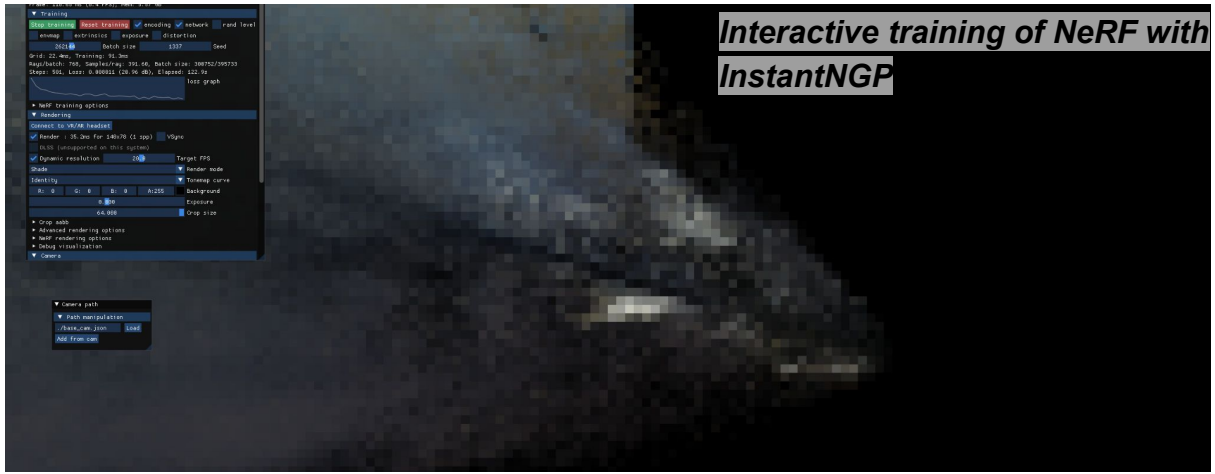


Kitti Sequence Cameras



Data Preparation: Poses

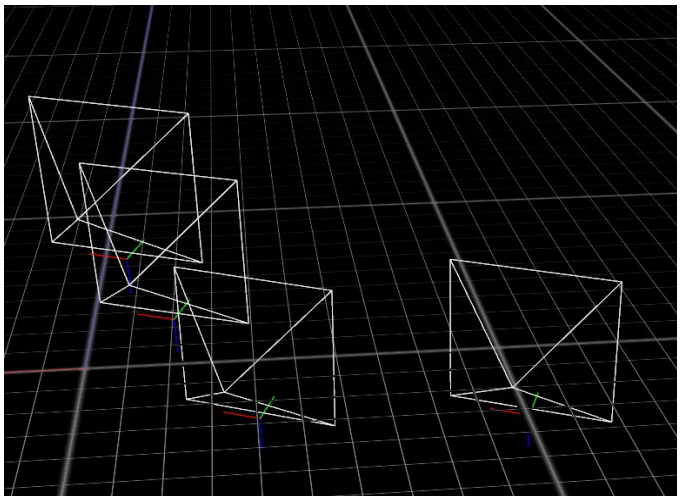
- **Challenges**
  - Cameras only moving in one direction



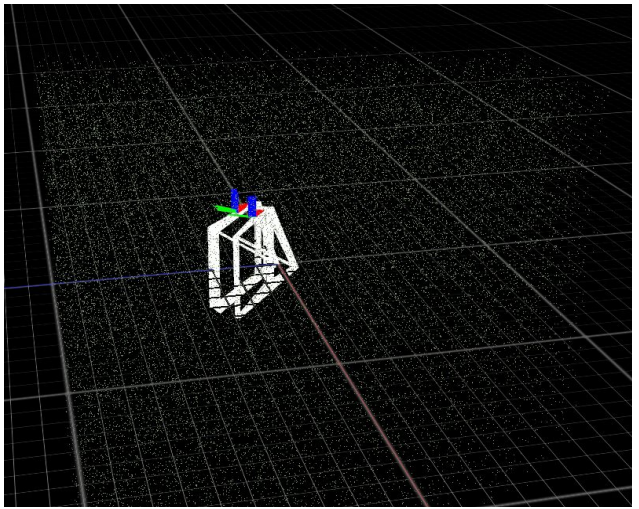
Data Preparation: Poses

- **Challenges**
  - How to test it?

Loaded Groundtruth Poses

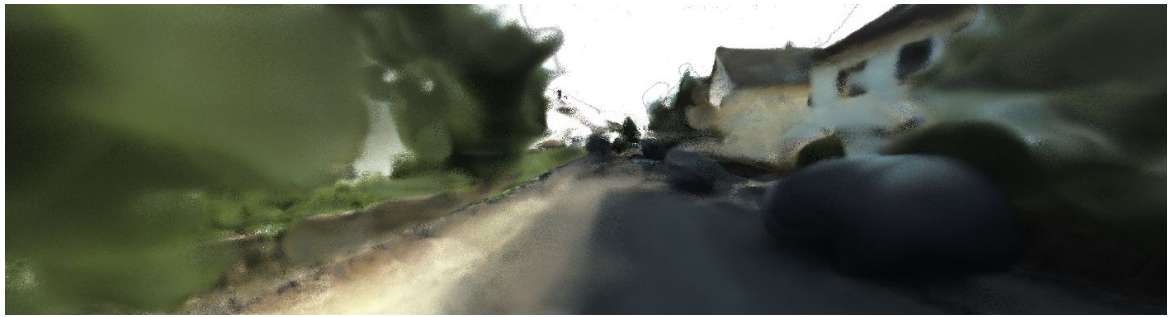


Colmap's Output



Data Preparation: Poses

- **Challenges**
  - How to test it?
    - ➡ Output of NeRF?

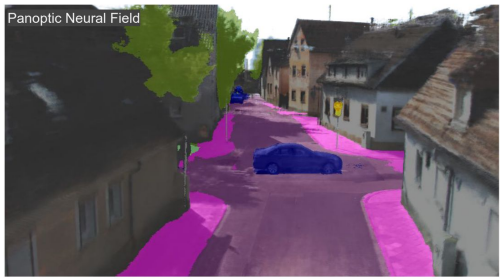


Colmap's Output

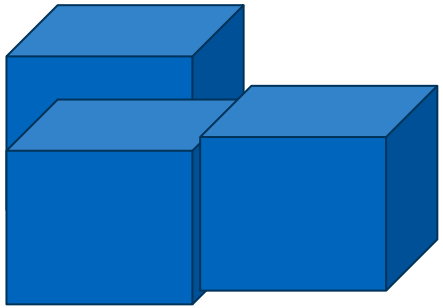
Loaded Groundtruth Poses

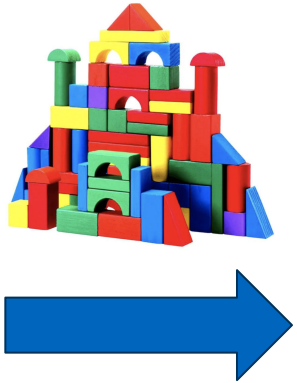
Approach

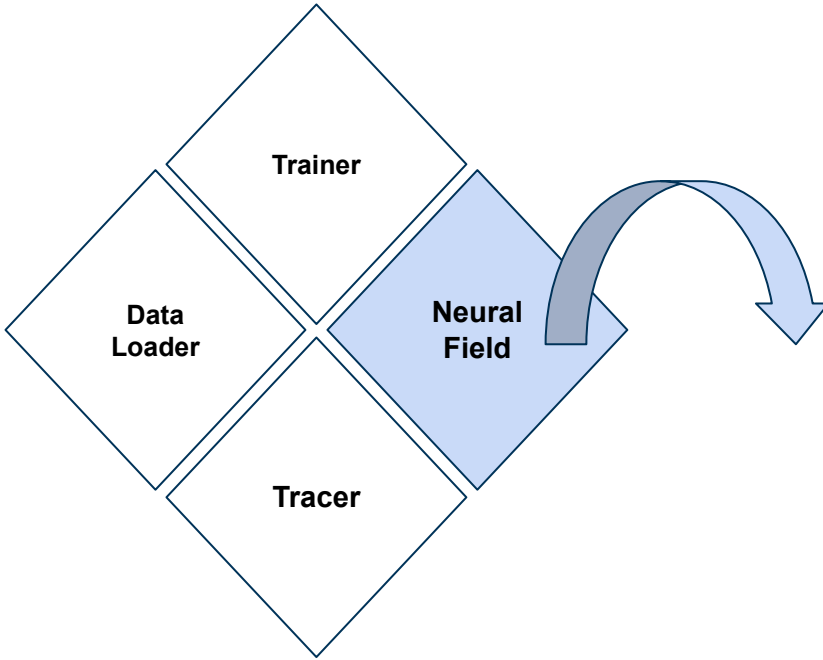

Data Preparation
NeRF
Semantic NeRF
Object Aware NeRF


Panoptic Neural Field

NeRF with Kaolin WISP

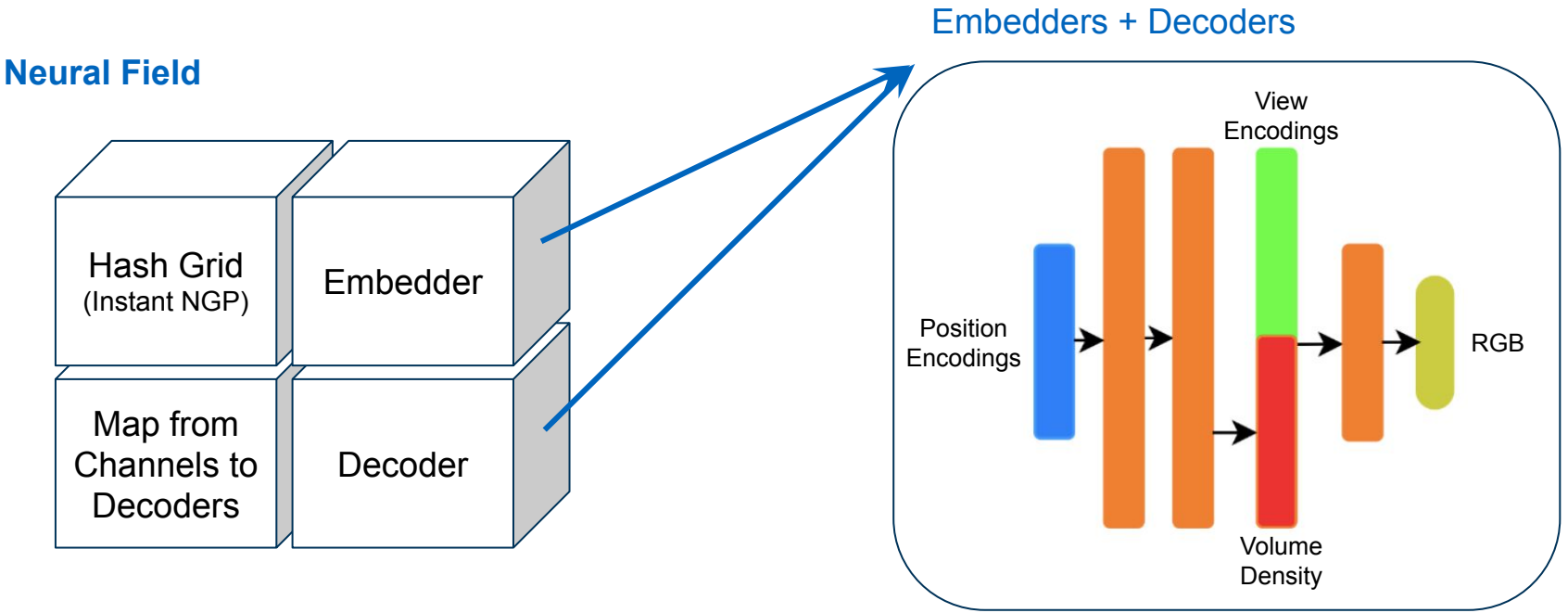




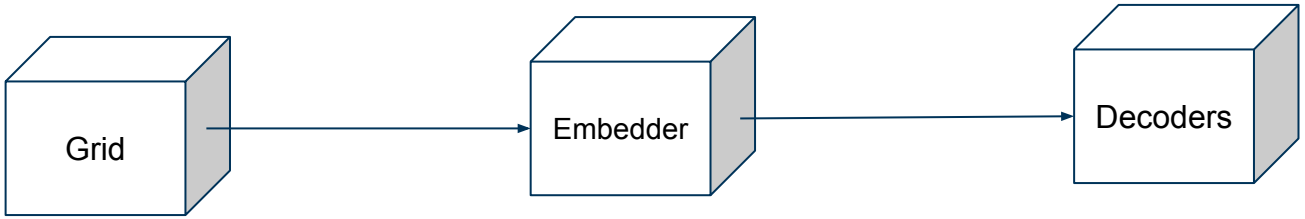


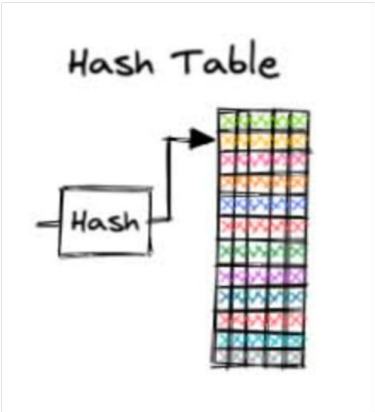
https://cdn.shopify.com/s/files/1/0035/0715/9129/products/BuildingBlocks-WoodenStackingBlocksForKids-90219_1500x.jpg?v=1629795416

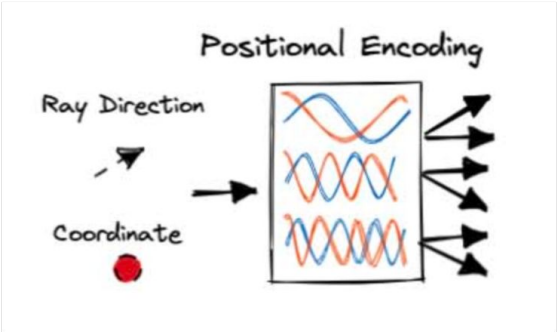
NeRF with Instant NGP

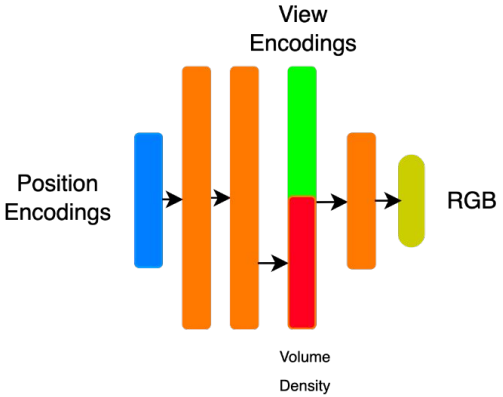


NeRF


Grid
Embedder
Decoders

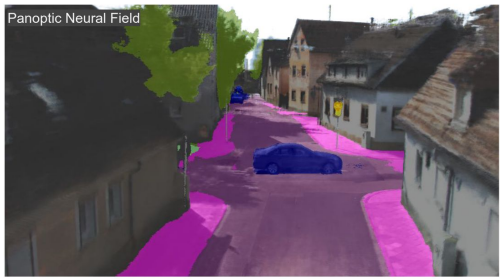

Hash Table
Hash


Positional Encoding
Ray Direction
Coordinate


View
Encodings
Position
Encodings
RGB
Volume
Density

Approach

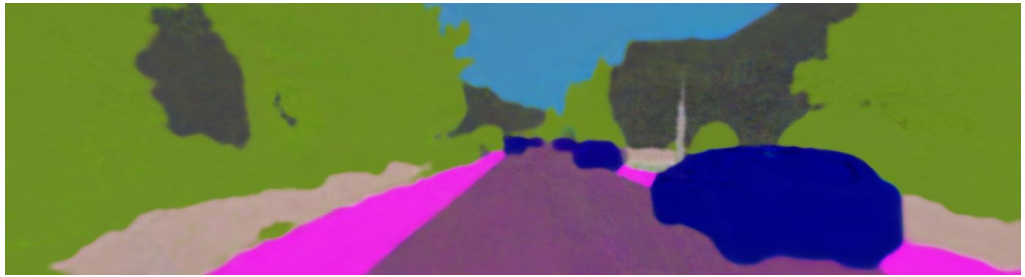

Data Preparation
NeRF
Semantic NeRF
Object Aware NeRF


Panoptic Neural Field

Semantic NeRF

- Using semantic images as RGB images
- Using labels from KITTI Dataset
  - Using 46 semantic labels
  - Using 8 category labels

With RGB

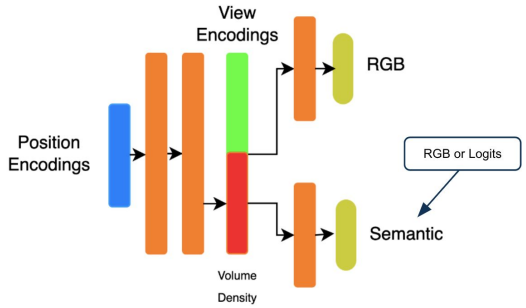


With logits (8 category labels)

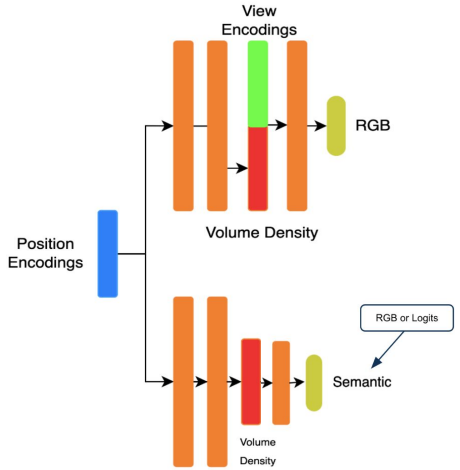
Semantic NeRF

- Using semantic images as RGB images
- Using labels from KITTI Dataset
  - Using 46 semantic labels
  - Using 8 category labels
- Single density decoder
- Separate density decoders

Single Density Decoder



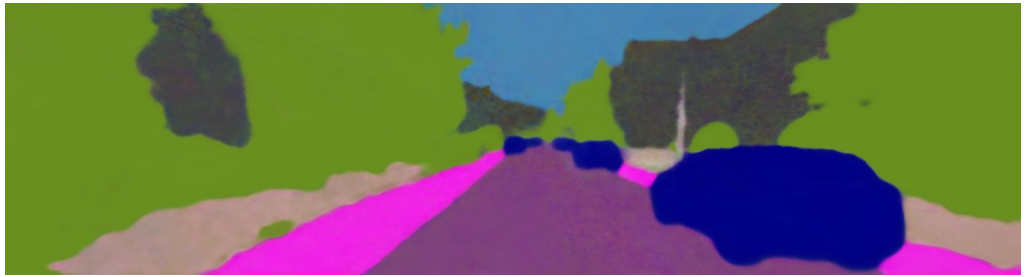
Two Density Decoders



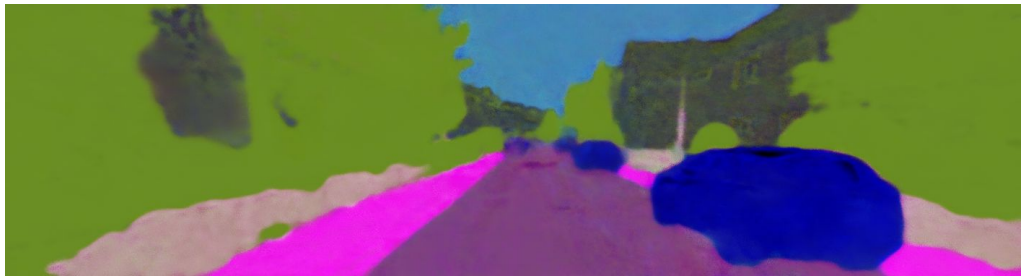
Semantic NeRF

- Using semantic images as RGB images
- Using labels from KITTI Dataset
  - Using 46 semantic labels
  - Using 8 category labels
- Single density decoder
- Separate density decoders
- Single pass through density decoder
- Double pass through density decoder

Single pass through density decoder

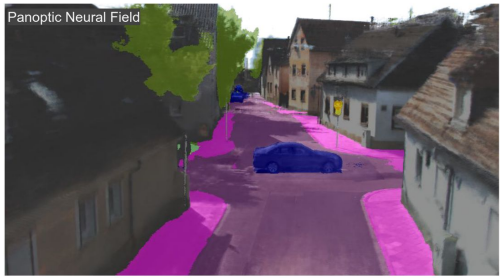


Double pass through density decoder

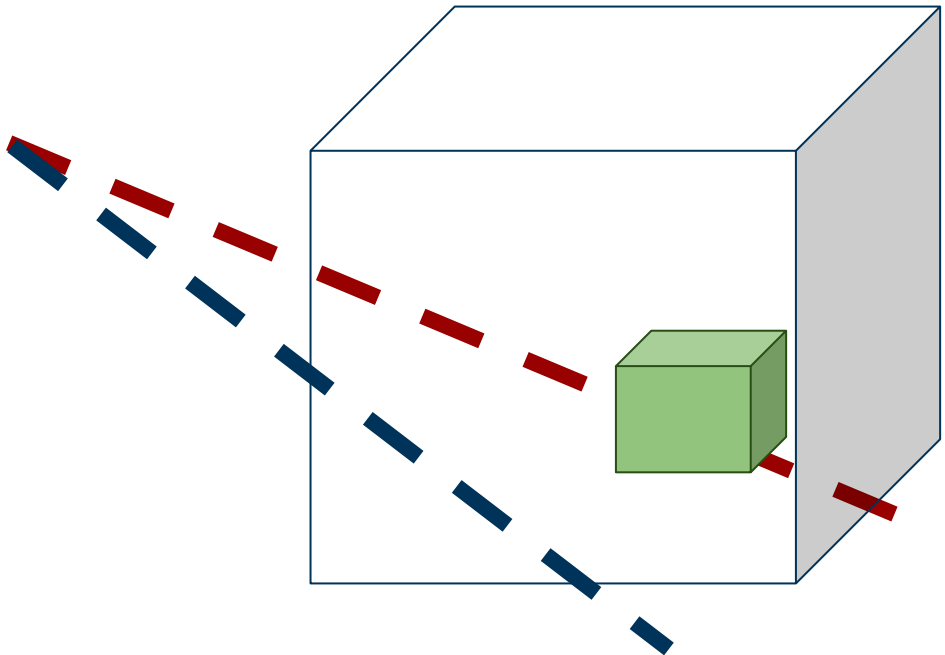


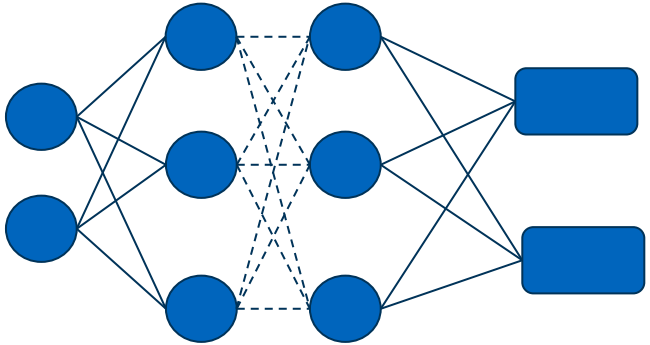
Approach

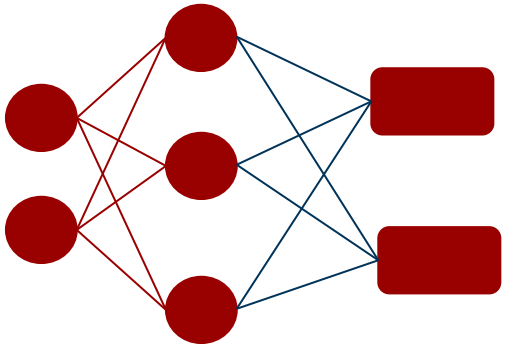
Data Preparation → NeRF → Semantic NeRF → Object Aware NeRF



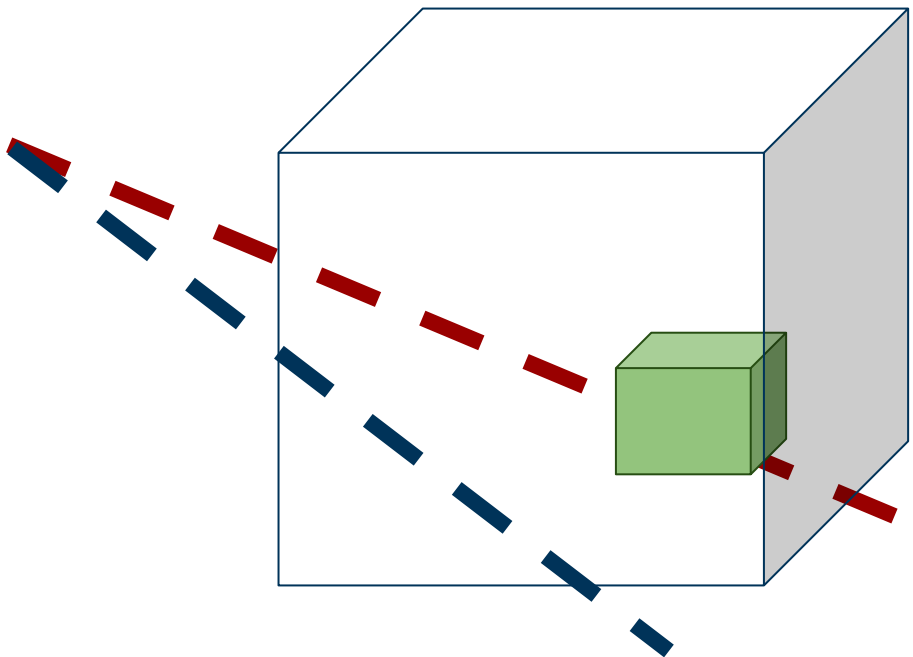
Object-Aware NeRF







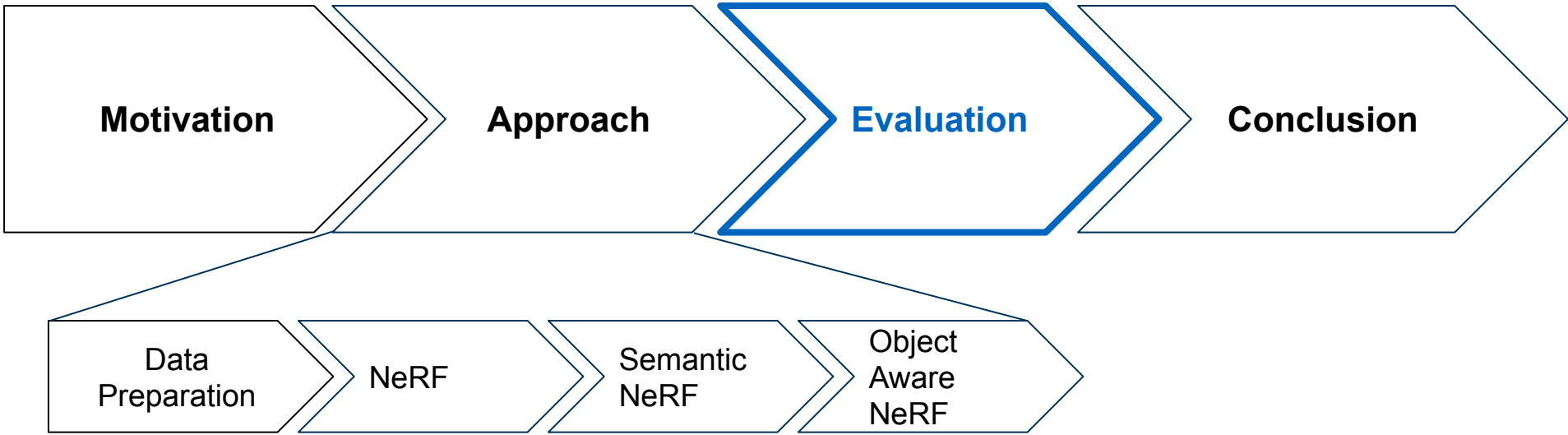
Object-Aware NeRF



Problems:

- Too slow!
- Not enough time for given technical knowledge

Agenda


Motivation
Approach
Evaluation
Conclusion
Data Preparation
NeRF
Semantic NeRF
Object Aware NeRF

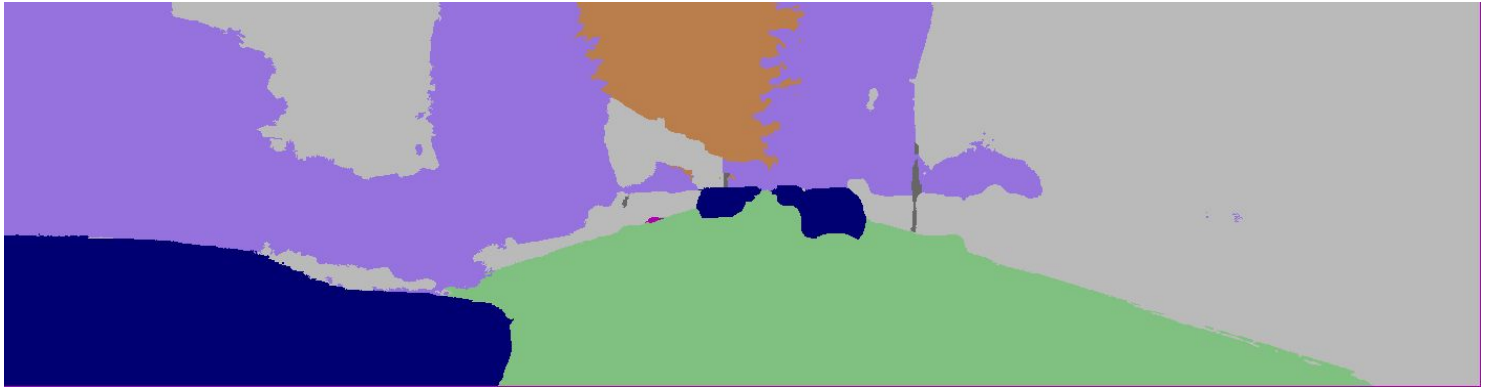
Qualitative

Ground Truth

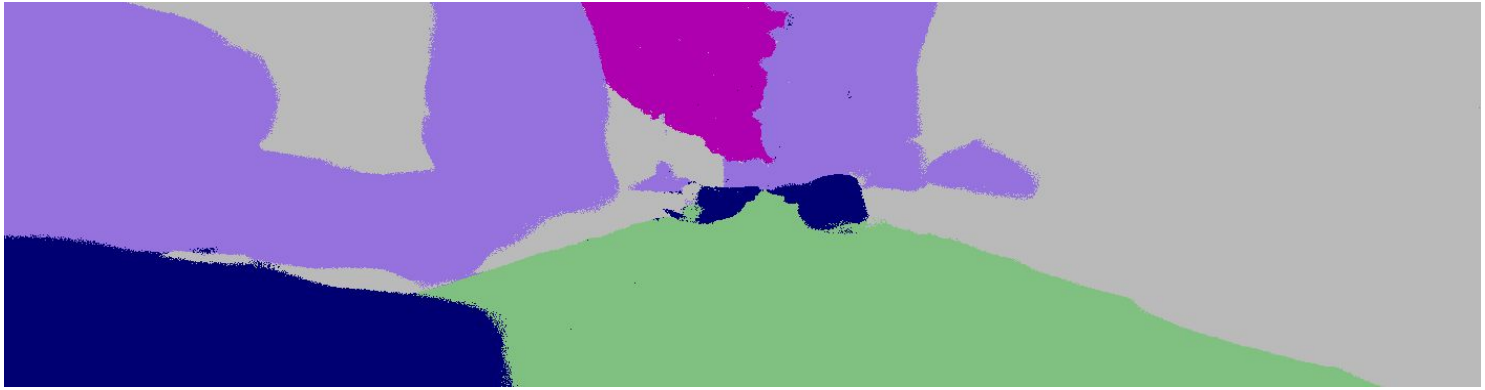
NeRF

- **2 Layers**
- **256 Neurons**
- **0.001 Learning Rate**
- **1.0 Semantic Weight**
- **19 Codebook Size**

Qualitative: Semantic NeRF



Ground Truth



Semantic Nerf

- **2 Layers**
- **32 Neurons**
- **0.001 Learning Rate**
- **1 Semantic Weight**
- **8 Category Label**
- **19 Codebook Size**

Qualitative

- **2 Layers**
- **32 Neurons**
- **0.001 Learning Rate**
- **19 Codebook Size**

Semantic Loss Weight 0.2

Semantic Loss Weight 1.0

Quantitative Evaluation

Metrics:

- **mIoU**
  - Mean Intersection over Union

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

- **PSNR**
  - Peak Signal to Noise Ratio

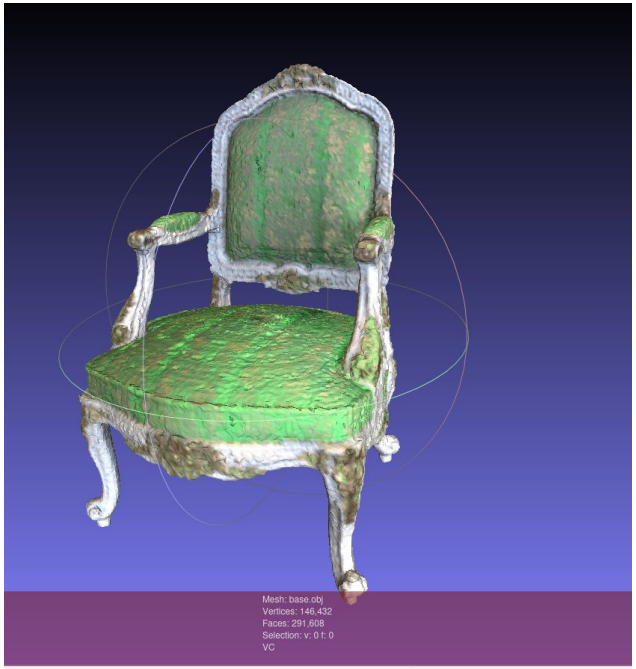
$$PSNR = 10 * \log_{10}(\frac{1}{MSE})$$

	Semantics mIoU	**Appearance PSNR**
PNF	74.28	21.91
ours	53.29	16.82

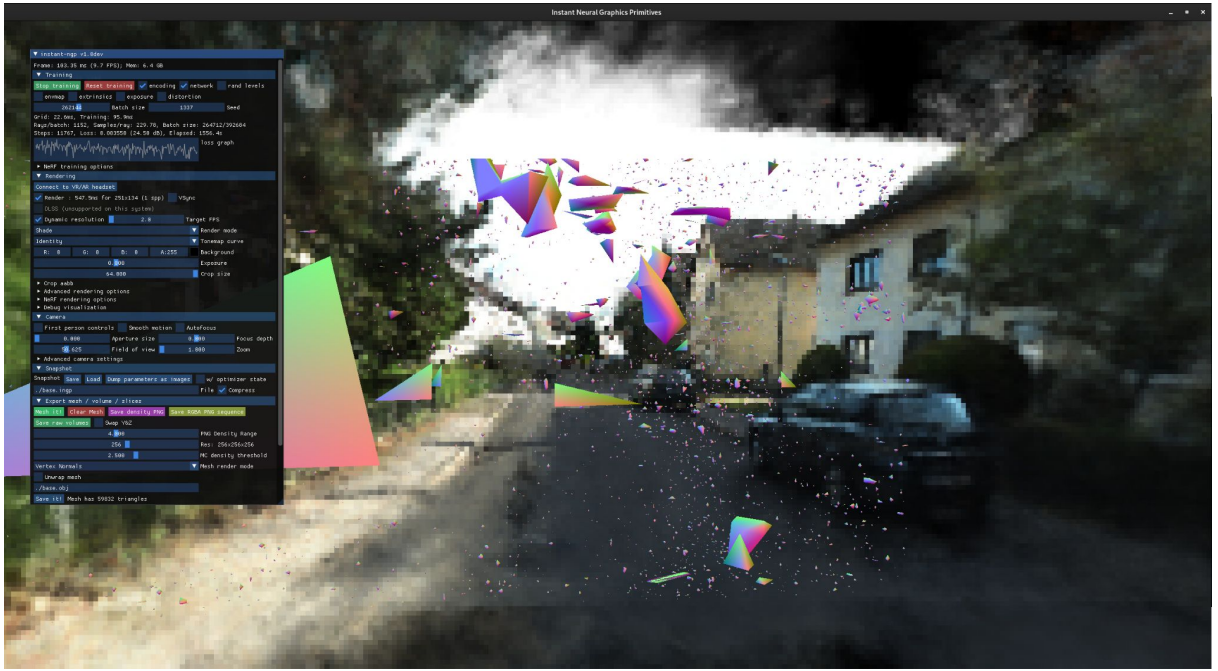
Quantitative Evaluation

Hidden Dim	Pass Count	Decoder	Prediction	Pixel Accuracy
256	1	1	RGB	0.7526
256	1	2	RGB	0.7541
128	2	1	RGB	0.7455
128	1	1	RGB	0.7540
64	1	2	RGB	0.7476
32	1	1	Logits	0.8219

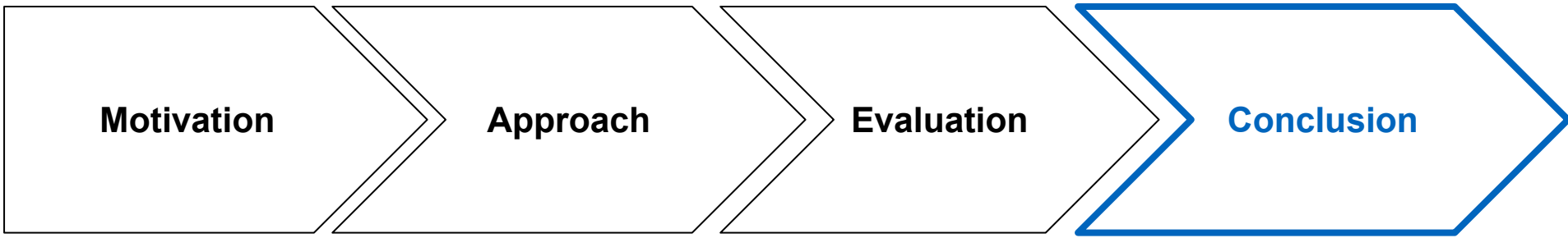
3D Reconstruction



Mesh of the chair



Instant NGP on KITTI

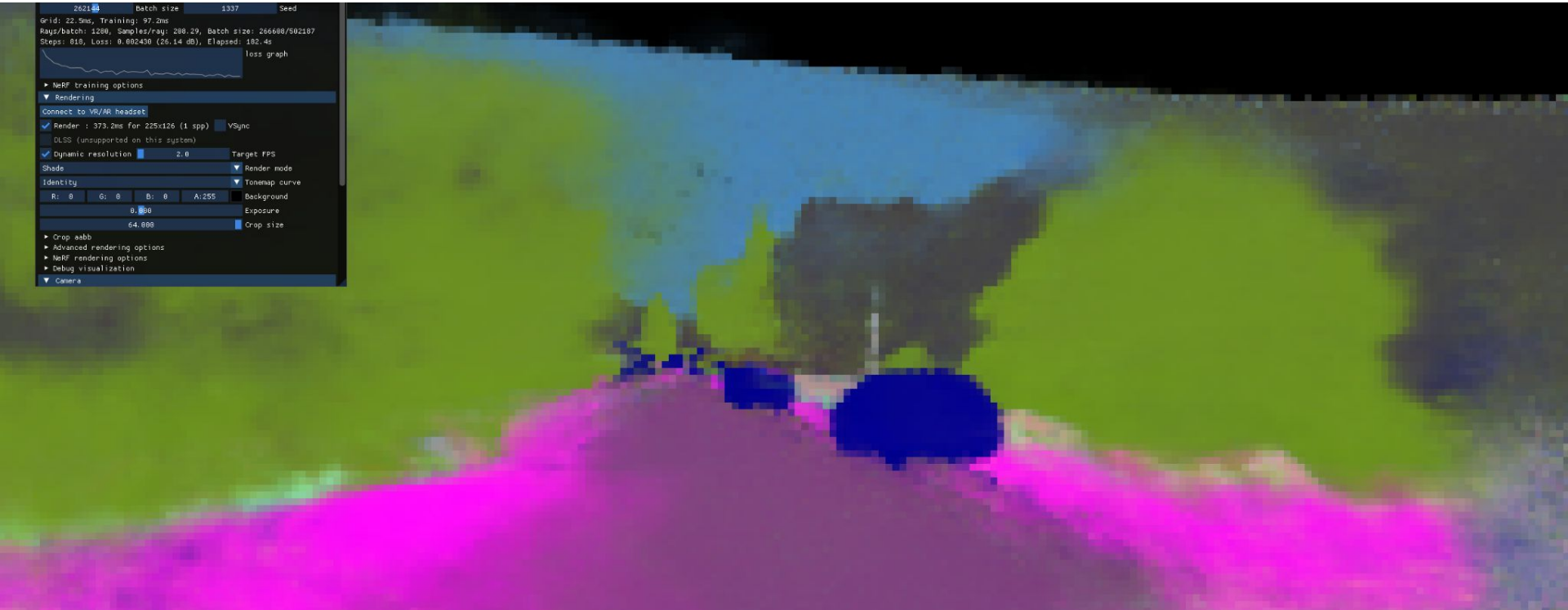

Motivation
Approach
Evaluation
Conclusion

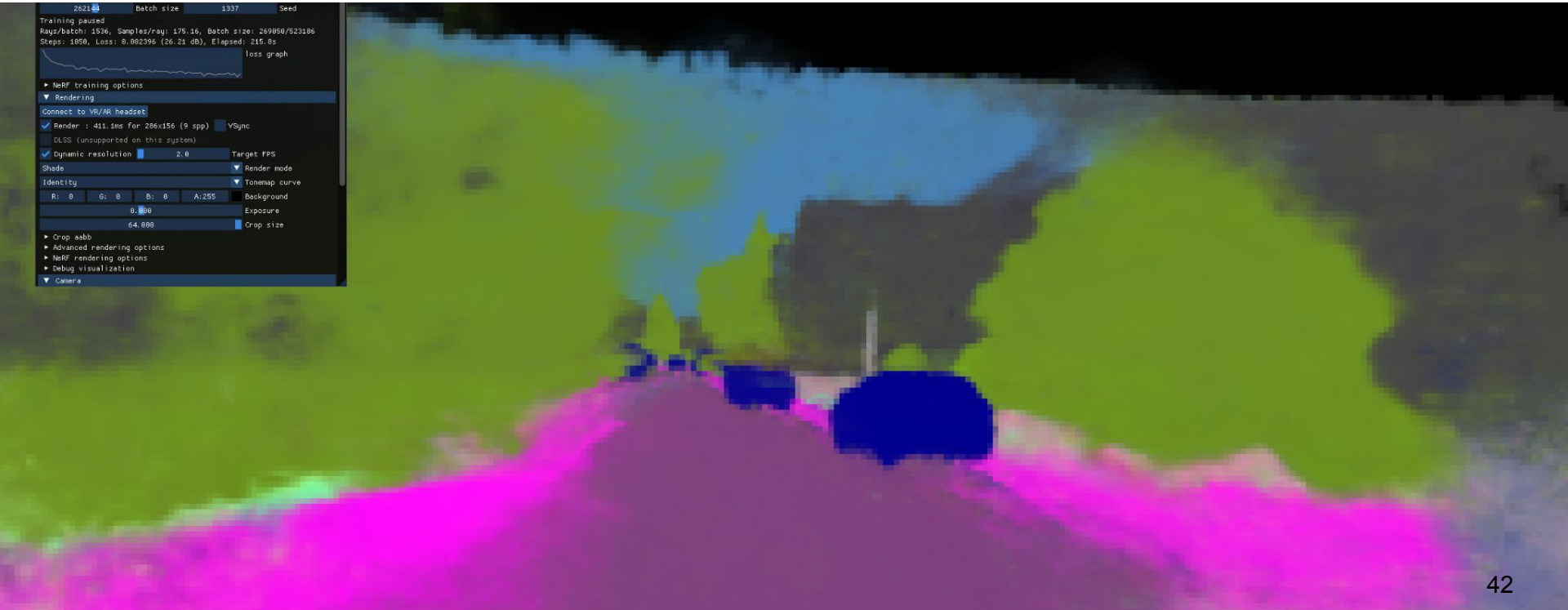
Conclusion

- **The devil lies in the detail!**

- We managed to build a **Neural Radiance Field** with **Semantic Category Logits**
- Still so much to learn about NeRF:
  - Impact in quality of One Density Decoder vs Two Density Decoders
- Preparing the input to your system and serving the output of your NeRF are no trivial tasks
- Efficient processing of the points is very crucial! No for-loops allowed!

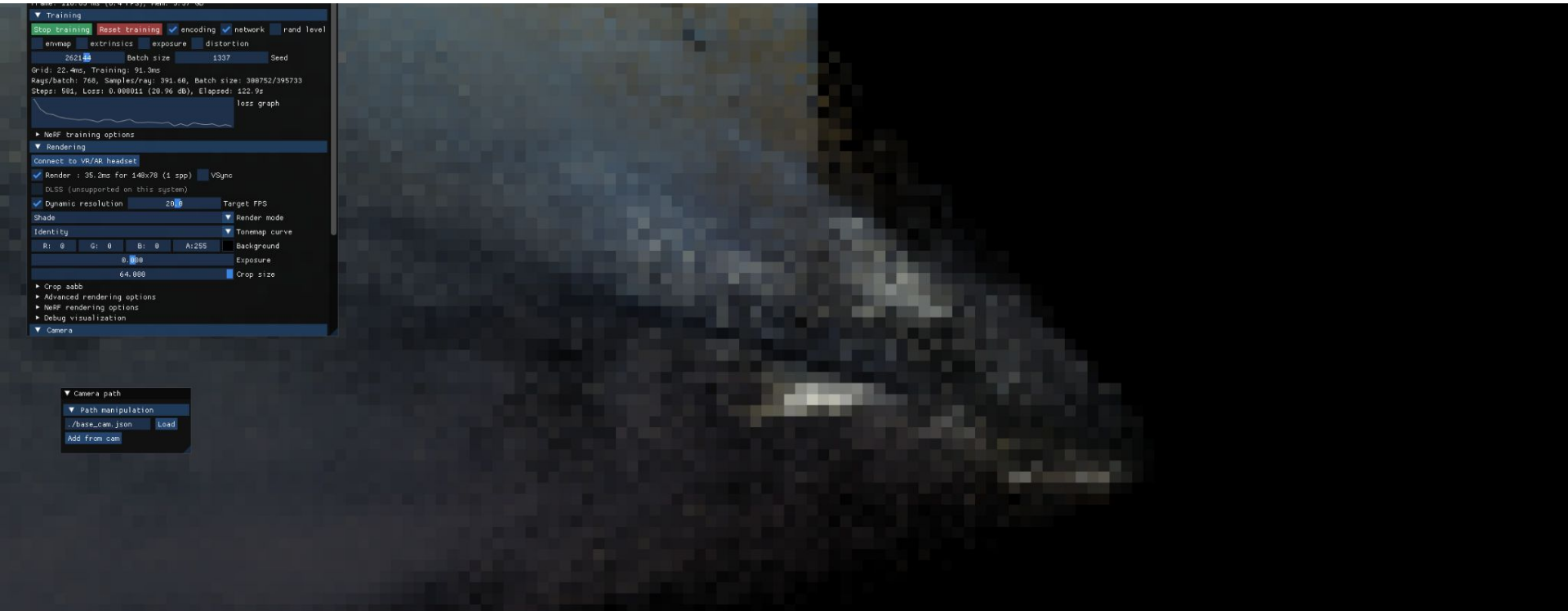
Thank you!


262144
Batch size
1337
Seed
Grid: 22.5ms, Training: 97.2ms
Rays/batch: 1280, Samples/ray: 280.29, Batch size: 266608/502187
Steps: 010, Loss: 0.002430 (26.14 dB), Elapsed: 102.4s
loss graph
NeRF training options
Rendering
Connect to VR/AR headset
Render : 373.2ms for 225x126 (1 spp)
VSync
DLSS (unsupported on this system)
Dynamic resolution
2.0
Target FPS
Shade
Render mode
Identity
Tonemap curve
R: 0
G: 0
B: 0
A:255
Background
0.000
Exposure
64.000
Crop size
Crop aabb
Advanced rendering options
NeRF rendering options
Debug visualization
Camera


262144
Batch size
1337
Seed
Training paused
Rays/batch: 1536, Samples/ray: 175.16, Batch size: 269050/523186
Steps: 1050, Loss: 0.002396 (26.21 dB), Elapsed: 215.0s
loss graph
NeRF training options
Rendering
Connect to VR/AR headset
Render : 411.1ms for 286x156 (9 spp)
VSync
DLSS (unsupported on this system)
Dynamic resolution
2.0
Target FPS
Shade
Render mode
Identity
Tonemap curve
R: 0
G: 0
B: 0
A:255
Background
0.000
Exposure
64.000
Crop size
Crop aabb
Advanced rendering options
NeRF rendering options
Debug visualization
Camera


Training
Start training
Reset training
encoding
network
rand leve
envmap
extrinsics
exposure
distortion
262144
Batch size
1337
Seed
Training paused
Rays/batch: 1024, Samples/ray: 291.78, Batch size: 298787/466282
Steps: 1037, Loss: 0.006191 (22.08 dB), Elapsed: 371.6s
loss graph
NeRF training options
Rendering
Connect to VR/AR headset
Render : 381.3ms for 199x109 (42 spp)
VSync
DLSS (unsupported on this system)
Dynamic resolution
2.0
Target FPS
Shade
Render mode
Identity
Tonemap curve
R: 0
G: 0
B: 0
A:255
Background
0.000
Exposure
64.000
Crop size
Crop aabb
Advanced rendering options
NeRF rendering options
Debug visualization
Camera
Camera path
Path manipulation
./base_cam.json
Load
Add from cam


Training
Start training
Reset training
encoding
network
rand leve
envmap
extrinsics
exposure
distortion
Batch size
1337
Seed
Training paused
Rays/batch: 1024, Samples/ray: 291.78, Batch size: 298787/466282
Steps: 1837, Loss: 0.006191 (22.08 dB), Elapsed: 423.1s
loss graph
NeRF training options
Rendering
Connect to VR/AR headset
Render : 335.0ms for 199x109 (23 spp)
VSync
DLSS (unsupported on this system)
Dynamic resolution
2.0
Target FPS
Shade
Render mode
Identity
Tonemap curve
R: 0
G: 0
B: 0
A:255
Background
0.000
Exposure
64.000
Crop size
Crop aabb
Advanced rendering options
NeRF rendering options
Debug visualization
Camera
Camera path
Path manipulation
./base_cam.json
Load
Add from cam


Training
Stop training
Reset training
encoding
network
rand level
envmap
extrinsics
exposure
distortion
262144
Batch size
1337
Seed
Grid: 22.4ms, Training: 91.3ms
Rays/batch: 768, Samples/ray: 391.60, Batch size: 300752/395733
Steps: 501, Loss: 0.000011 (20.96 dB), Elapsed: 122.9s
loss graph
NeRF training options
Rendering
Connect to VR/AR headset
Render : 35.2ms for 140x70 (1 spp)
VSync
DLSS (unsupported on this system)
Dynamic resolution
2070
Target FPS
Shade
Render mode
Identity
Tonemap curve
R: 0
G: 0
B: 0
A:255
Background
0.000
Exposure
64.000
Crop size
Crop aabb
Advanced rendering options
NeRF rendering options
Debug visualization
Camera
Camera path
Path manipulation
./base_cam.json
Load
Add from cam