NeRF-MAE:

Masked AutoEncoders for Self-Supervised 3D Representation Learning for Neural Radiance Fields

European Conference on Computer Vision, ECCV 2024
also appeared at CVPR Neural Rendering Intelligence Workshop, CVPR 2024


Zubair Irshad


Sergey Zakharov

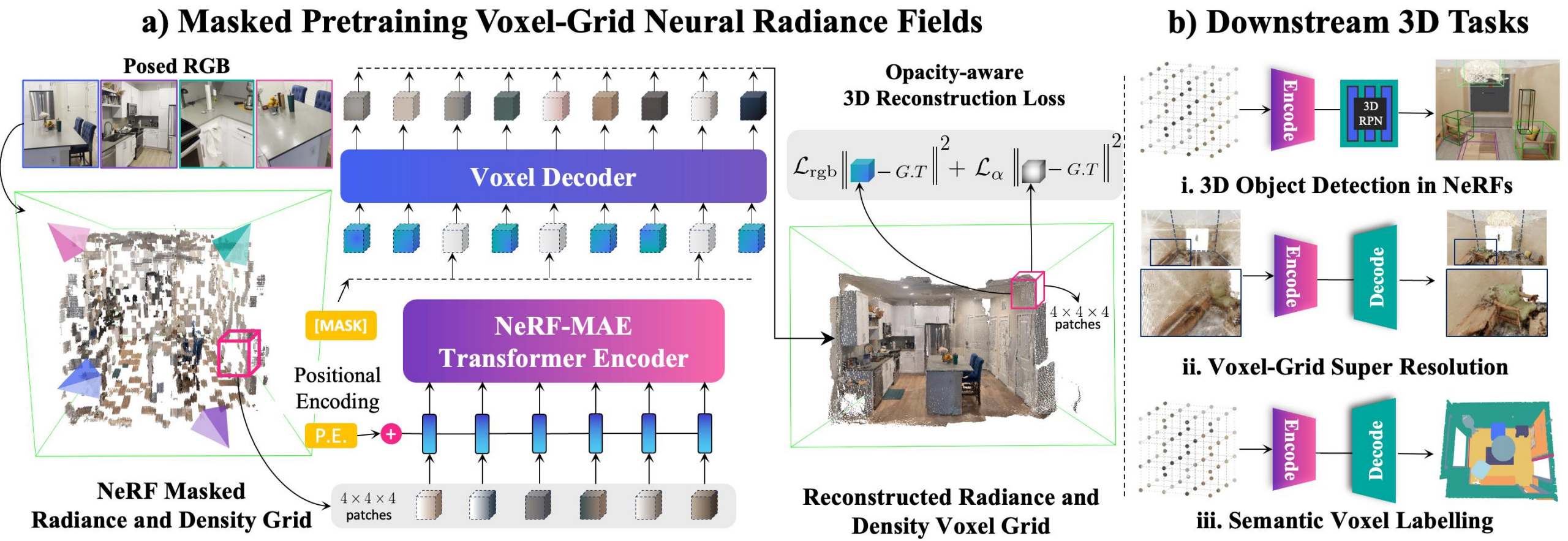

Vitor Guizilini


Adrien Gaidon

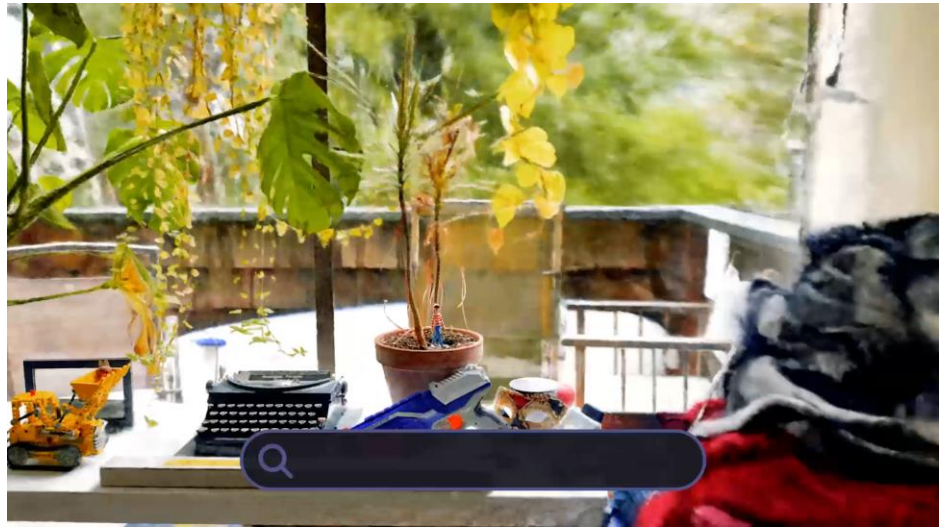

Zsolt Kira


Rares Ambrus

We propose NeRF-MAE, a framework for self-supervised 3D representation learning from NeRFs



Neural Fields beyond showcasing high-rendering quality

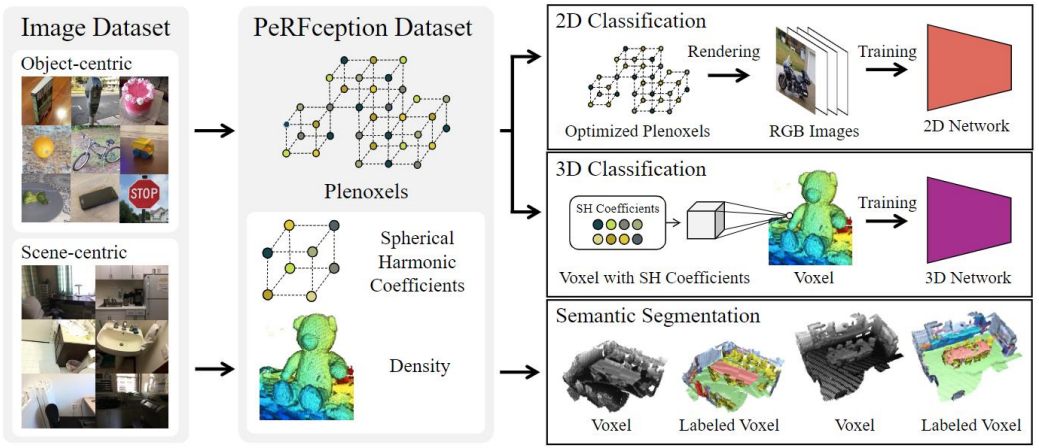


Language-Embedded Radiance Fields (LeRF, Kerr et al)

Inferring Accurate Geometry (NeRFMeshing, Rakotosaona et al)

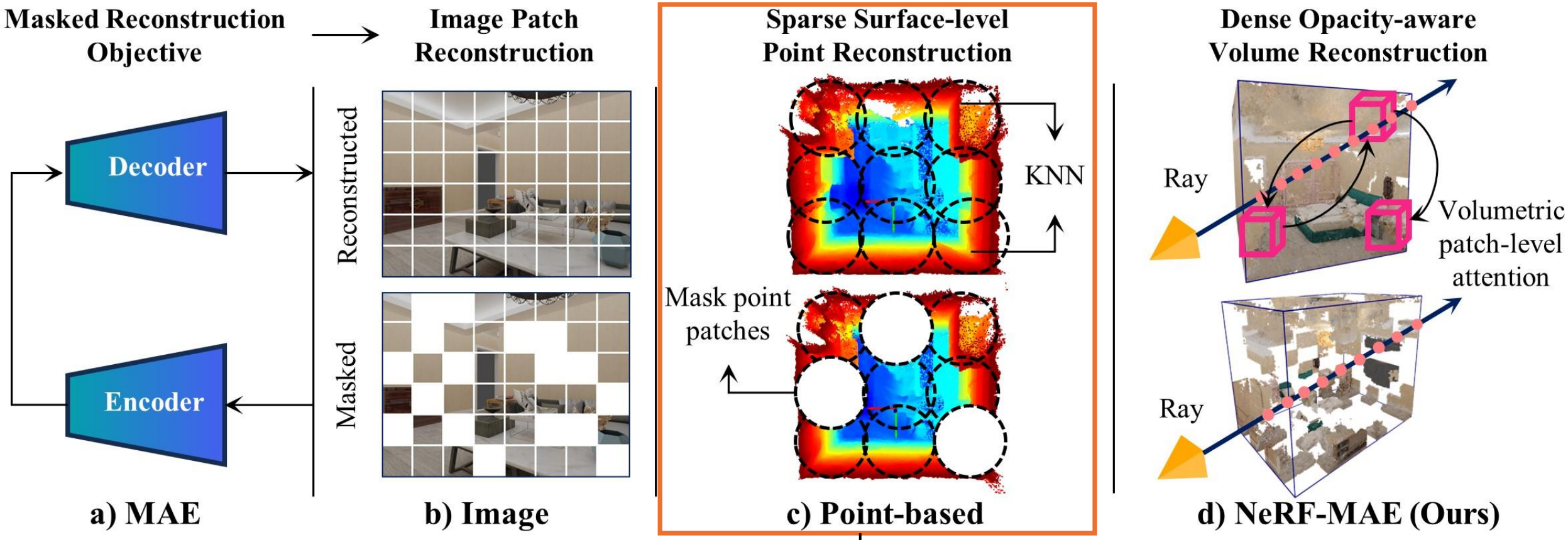
1. Scan Scene

Open-world Manipulation (F3RM, Shen et al)



Efficient Data Storage (PerFception, Jeong et al)

Existing **3D MAE Architectures** operate on pointclouds with **uneven information density** only model **surface-level information** and are **highy irregular data structures**

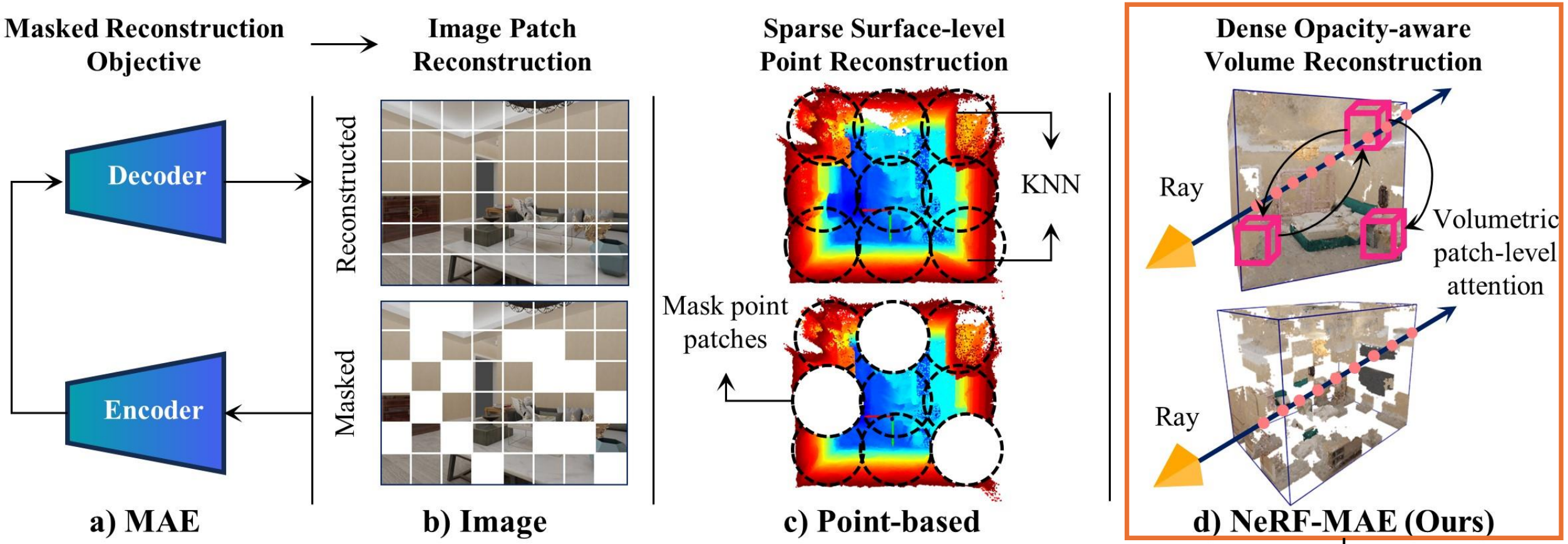


✗ Model Surface Level Information

✗ Irregular data Structures

✗ Uneven information density

Our approach uses **NeRF's dense grid** as input to the Transformer.
This makes our approach a direction extension of **image MAE to 3D**

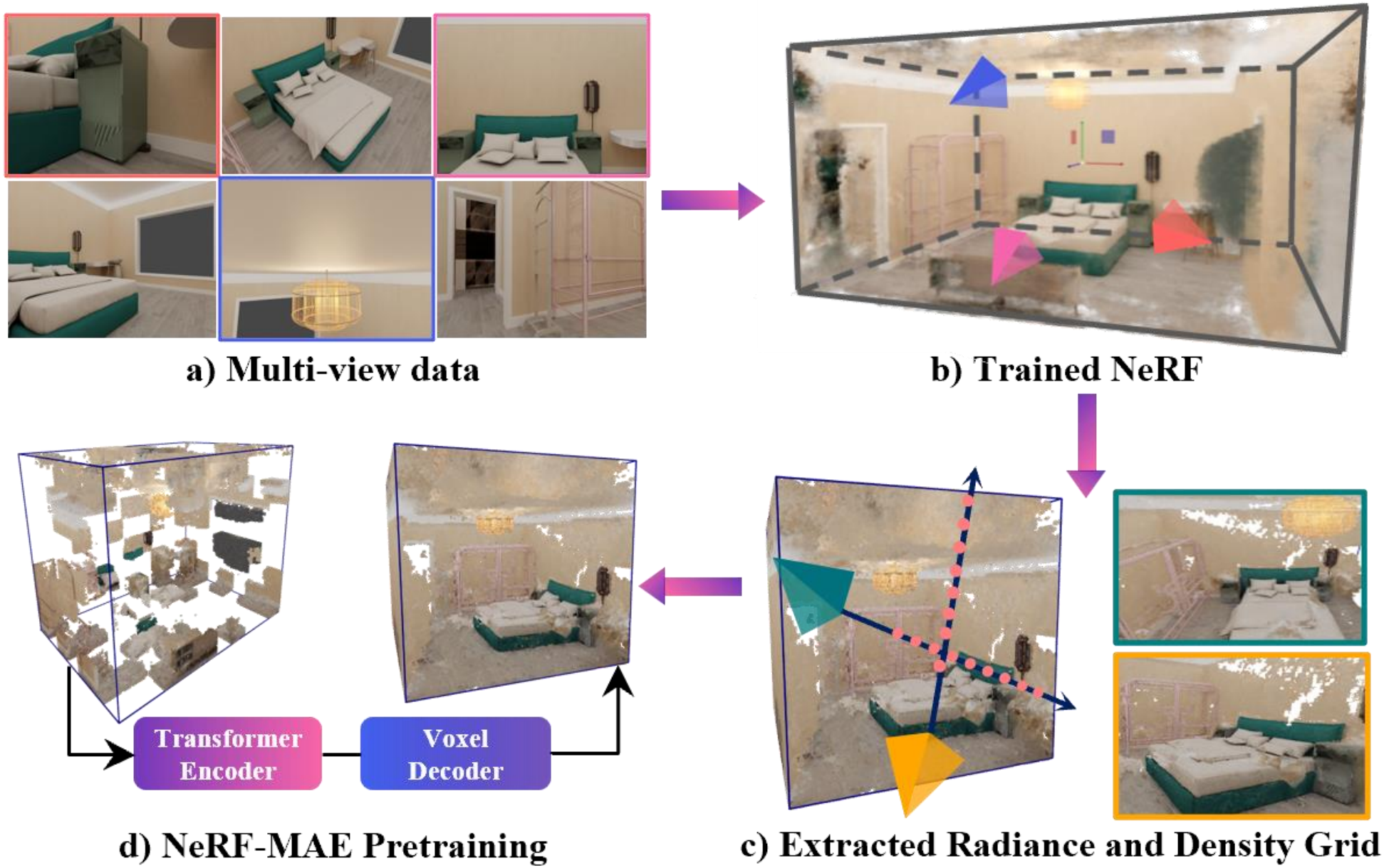


✓ High Information Density

✓ Regular unbiased Sampling

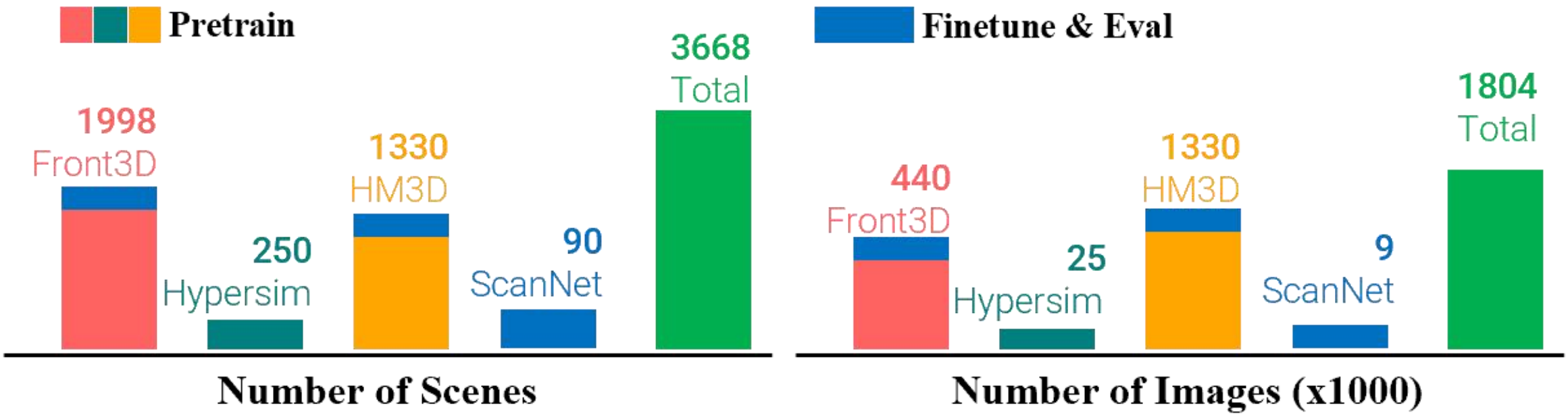
✓ Spatial data redundancy

Data Preprocessing flow for large-scale NeRF pretraining

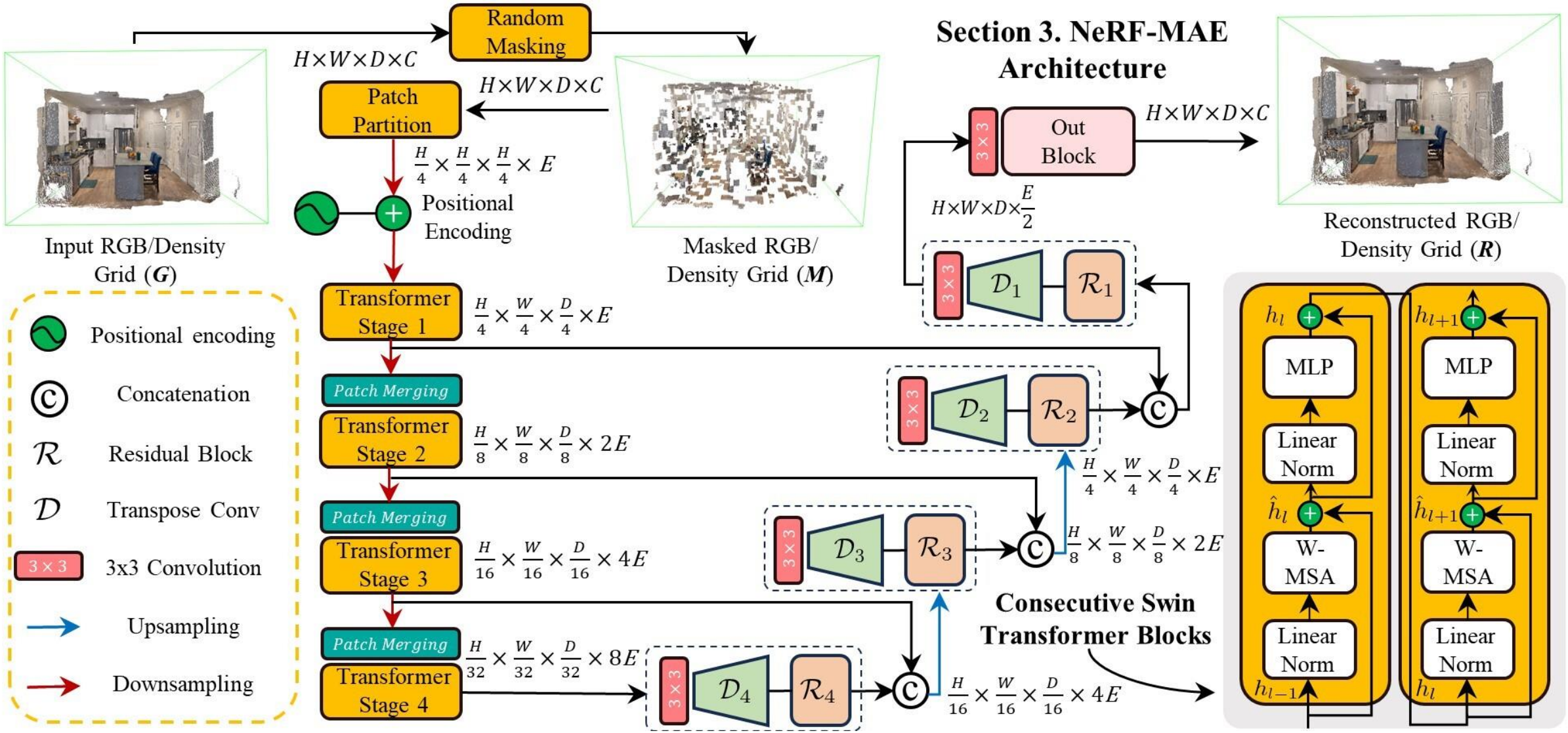


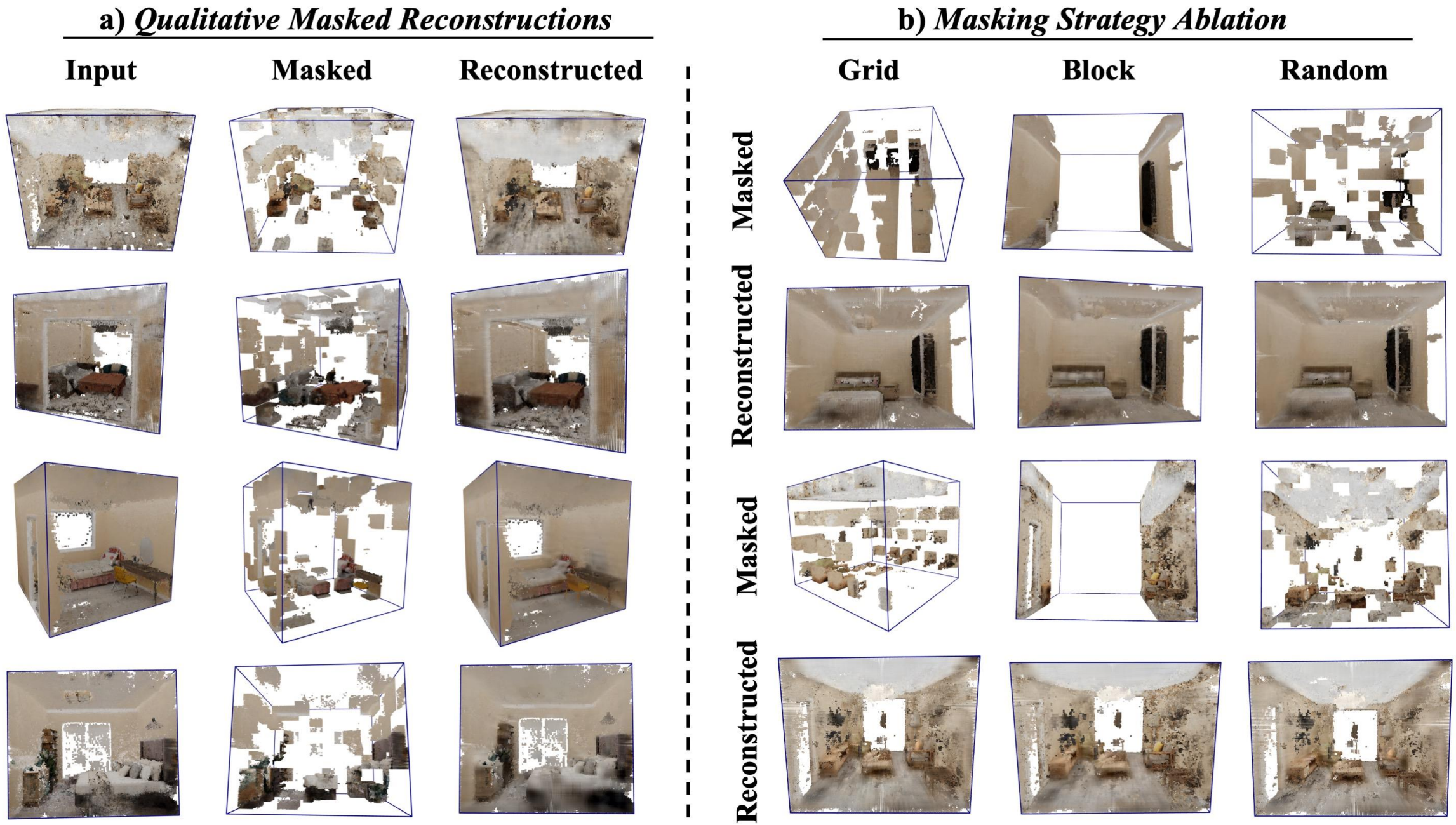
a) Multi-view Dataset Setup

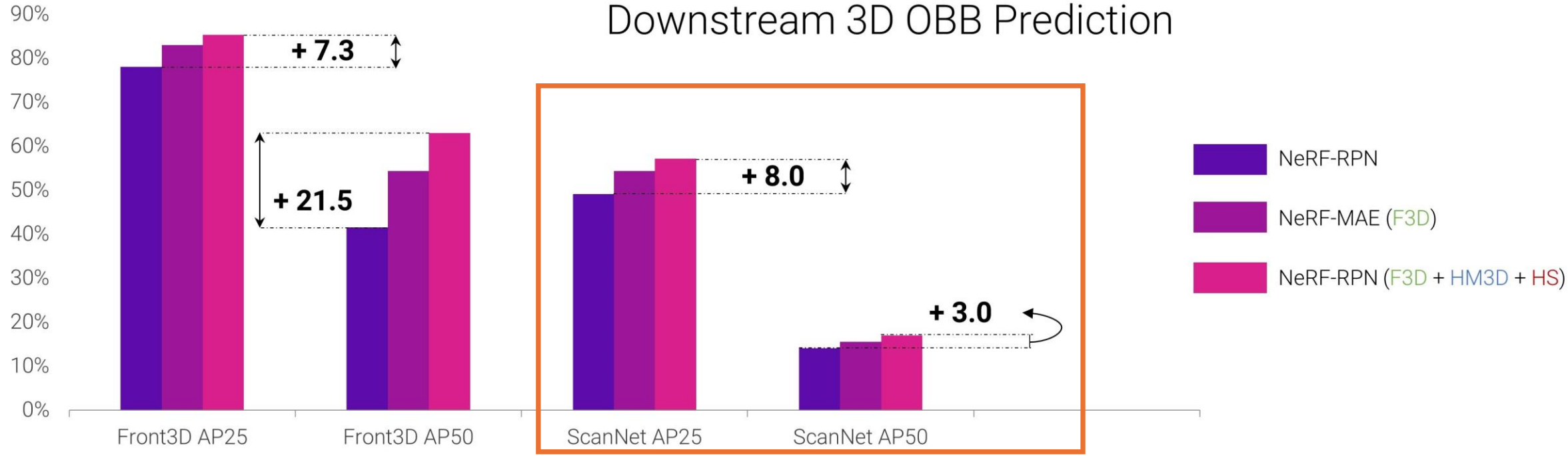
b) NeRF-MAE Data Mix & Statistics

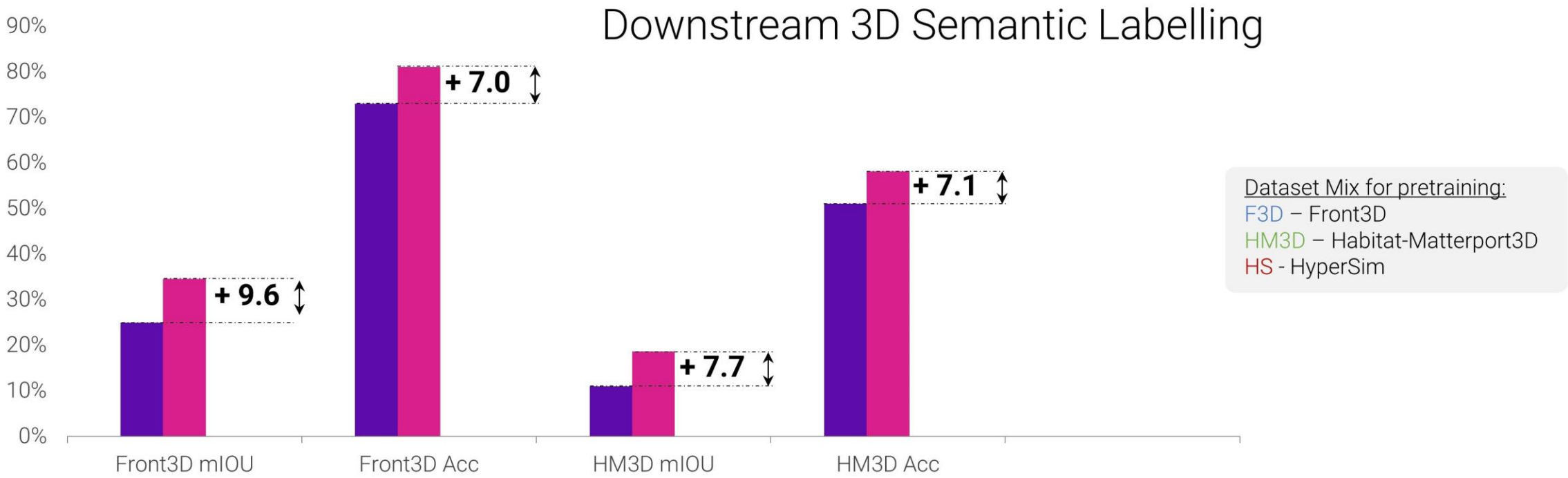


We utilize a **UNet architecture** employing SwinTransformer as encoder lightweight voxel decoder to enforce **mask reconstruction objective in 3D**

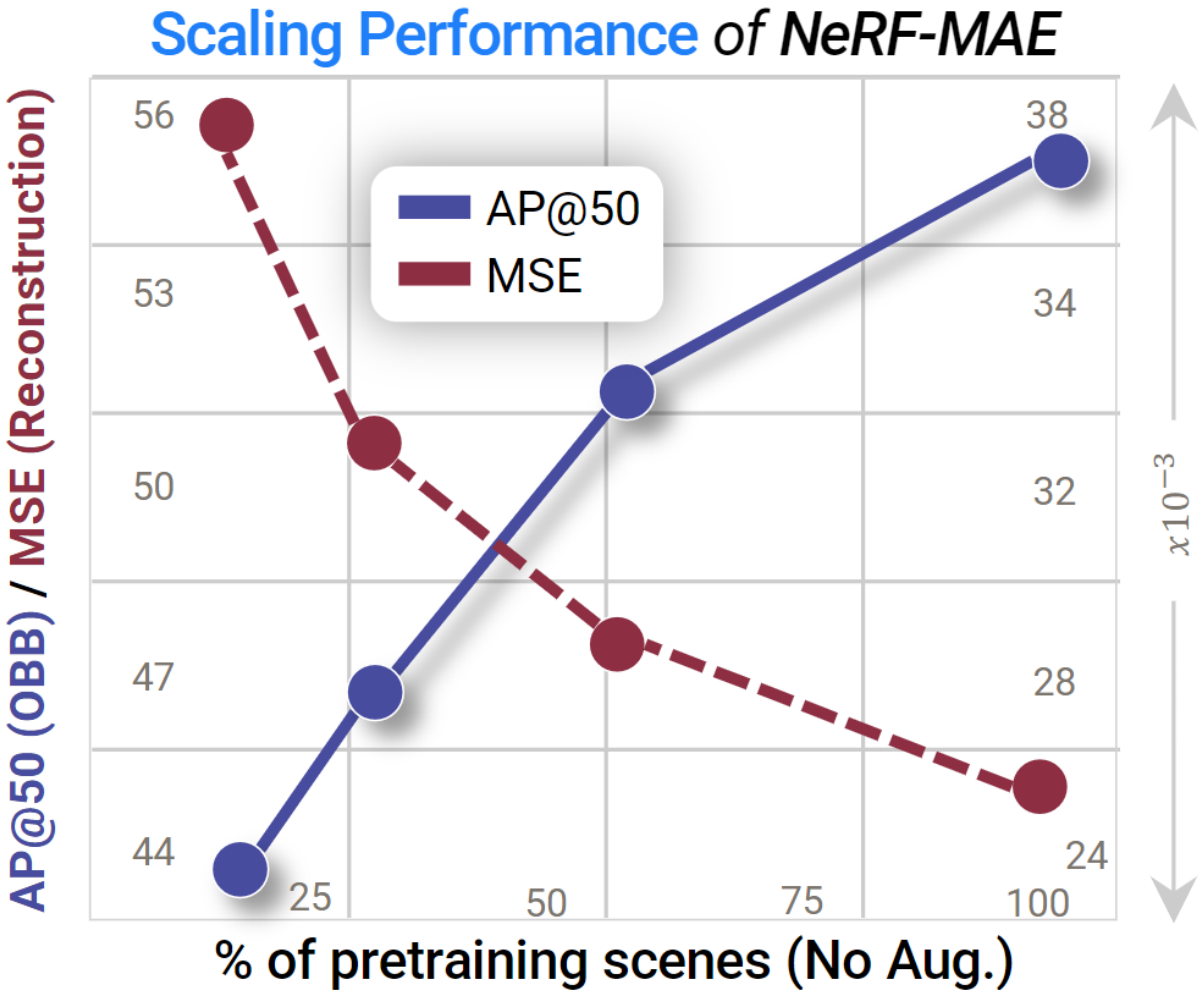


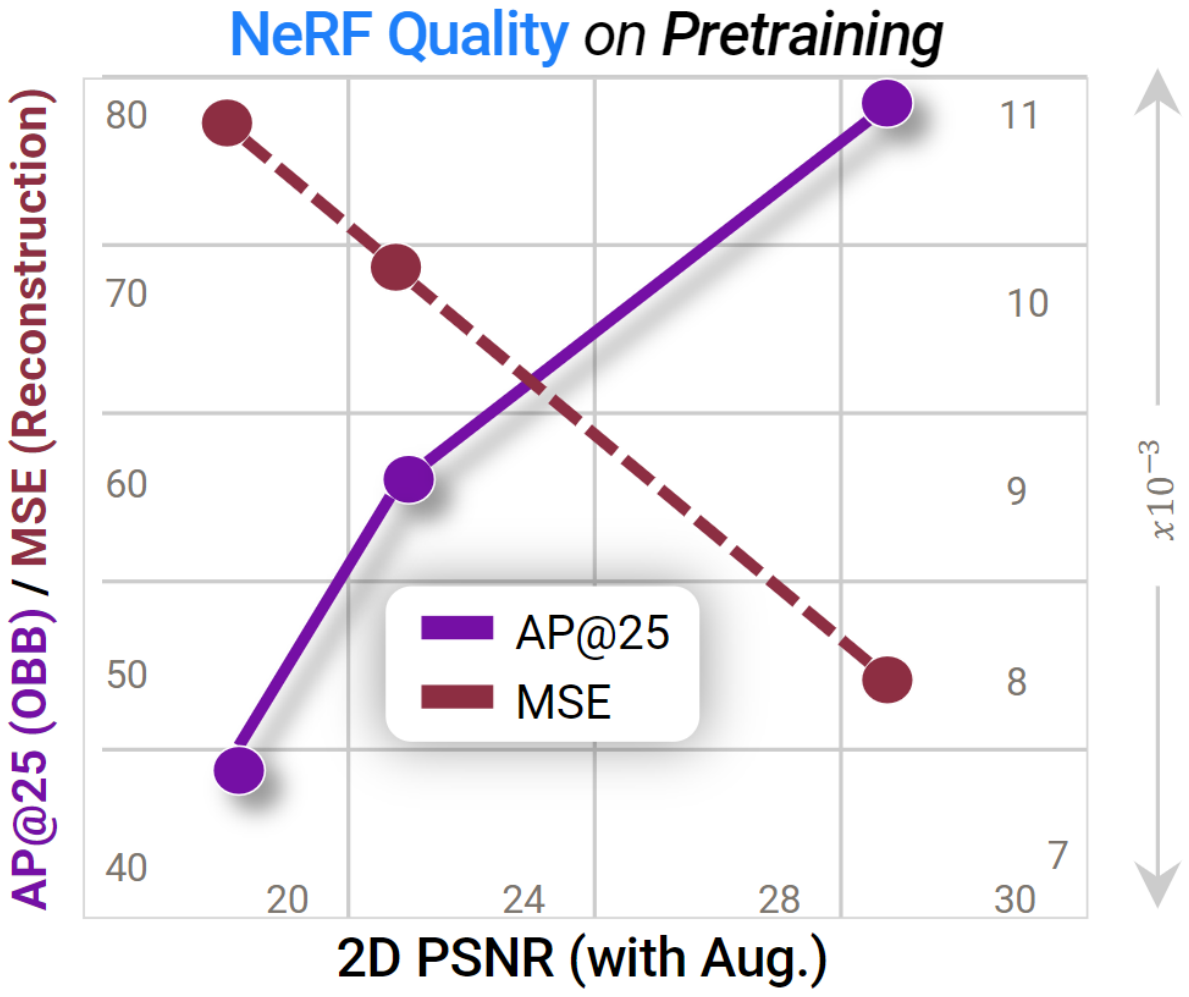

a) Qualitative Masked Reconstructions
Input
Masked
Reconstructed
b) Masking Strategy Ablation
Grid
Block
Random
Masked
Reconstructed
Masked
Reconstructed


Downstream 3D OBB Prediction
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
+ 7.3
+ 21.5
+ 8.0
+ 3.0
Front3D AP25
Front3D AP50
ScanNet AP25
ScanNet AP50
NeRF-RPN
NeRF-MAE (F3D)
NeRF-RPN (F3D + HM3D + HS)

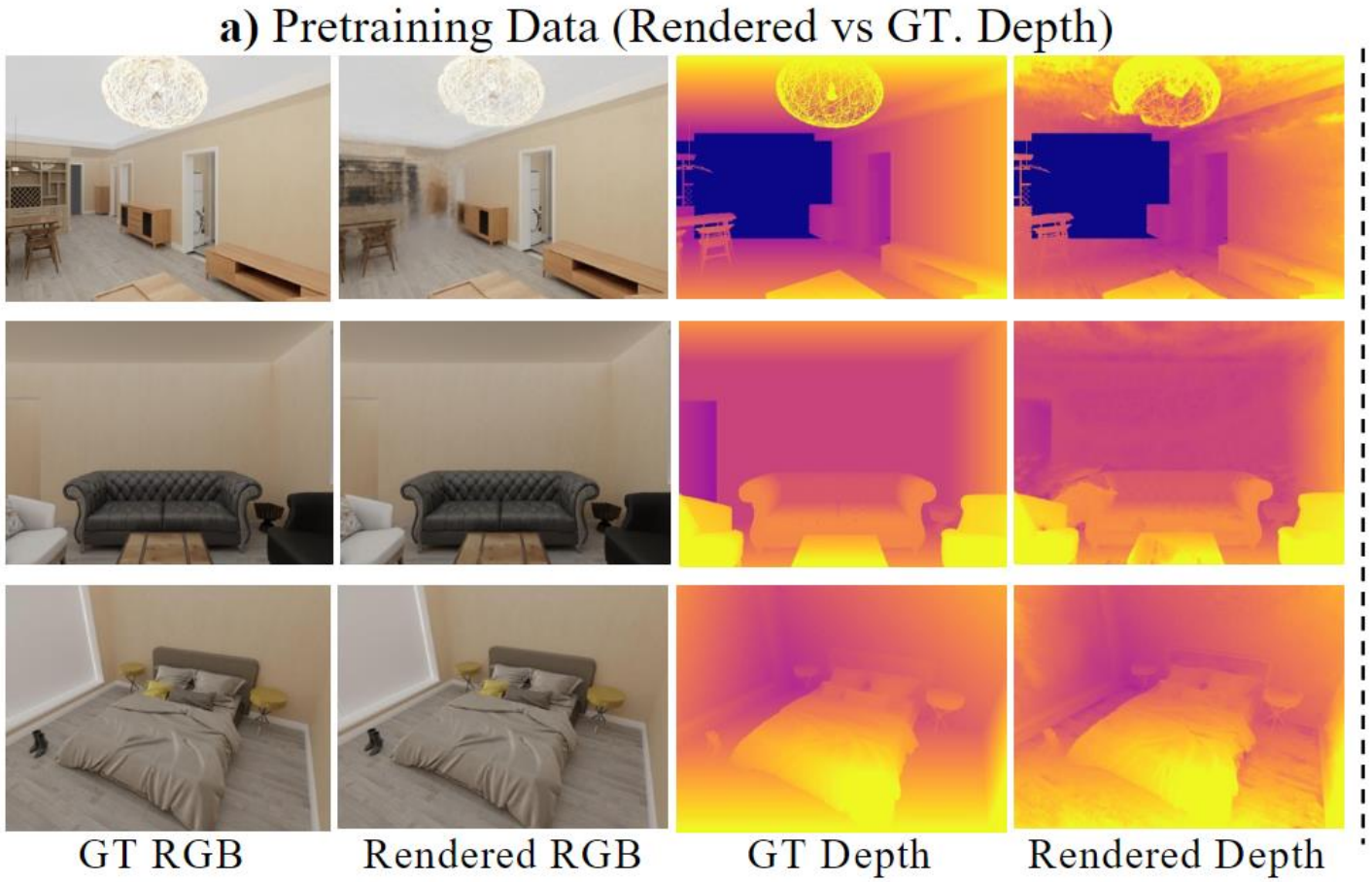

Downstream 3D Semantic Labelling
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
+ 9.6
+ 7.0
+ 7.7
+ 7.1
Front3D mIOU
Front3D Acc
HM3D mIOU
HM3D Acc
Dataset Mix for pretraining:
F3D – Front3D
HM3D – Habitat-Matterport3D
HS - HyperSim

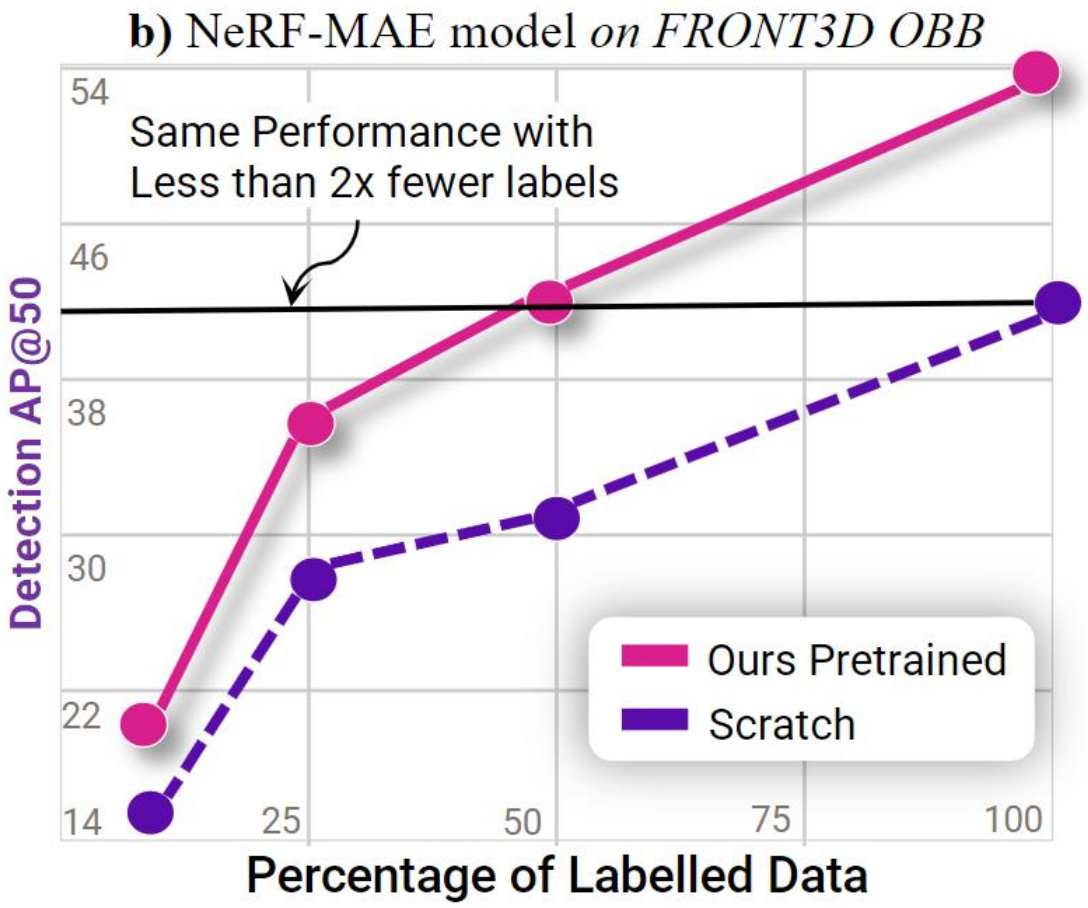
Quantitative Results



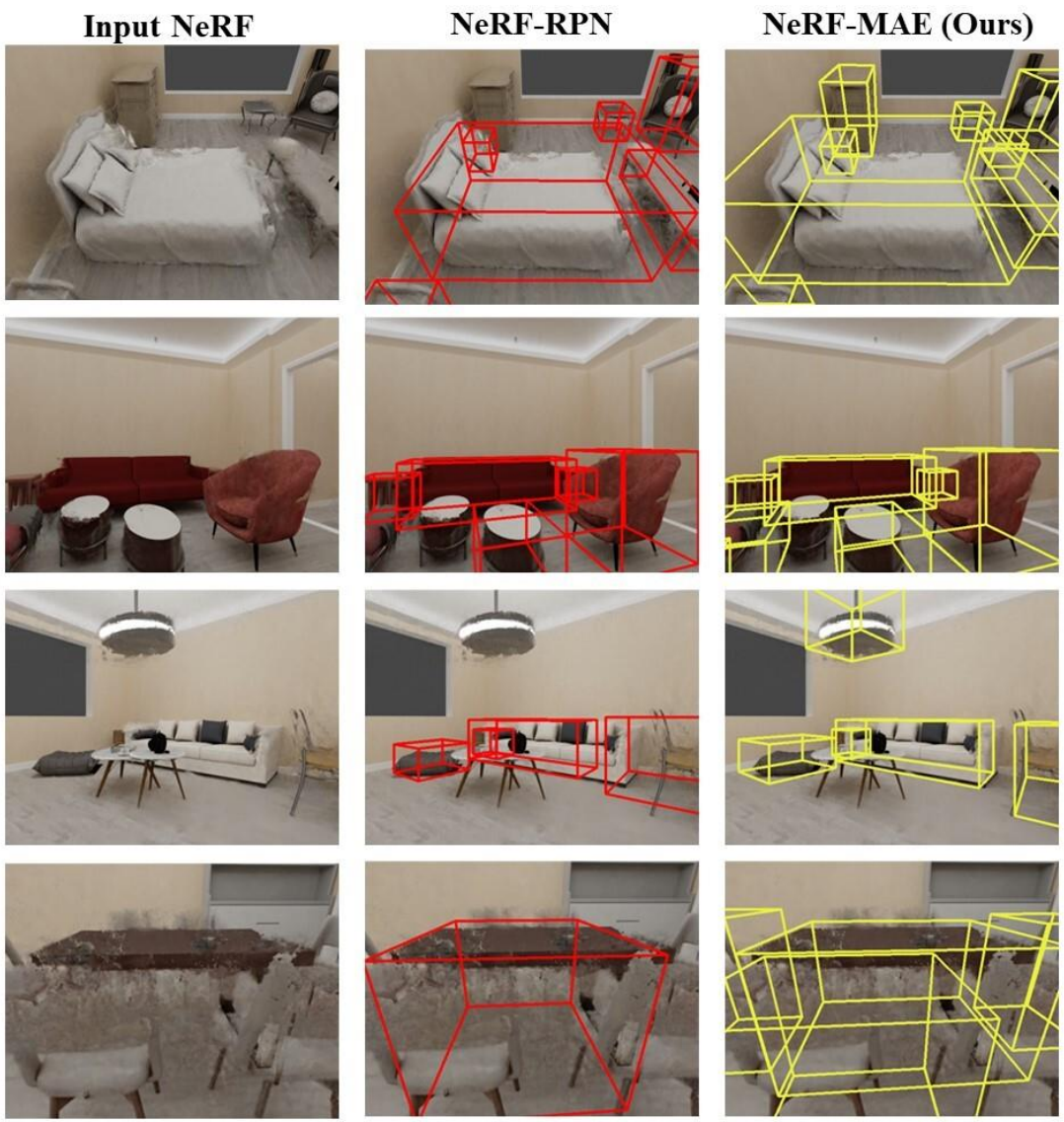


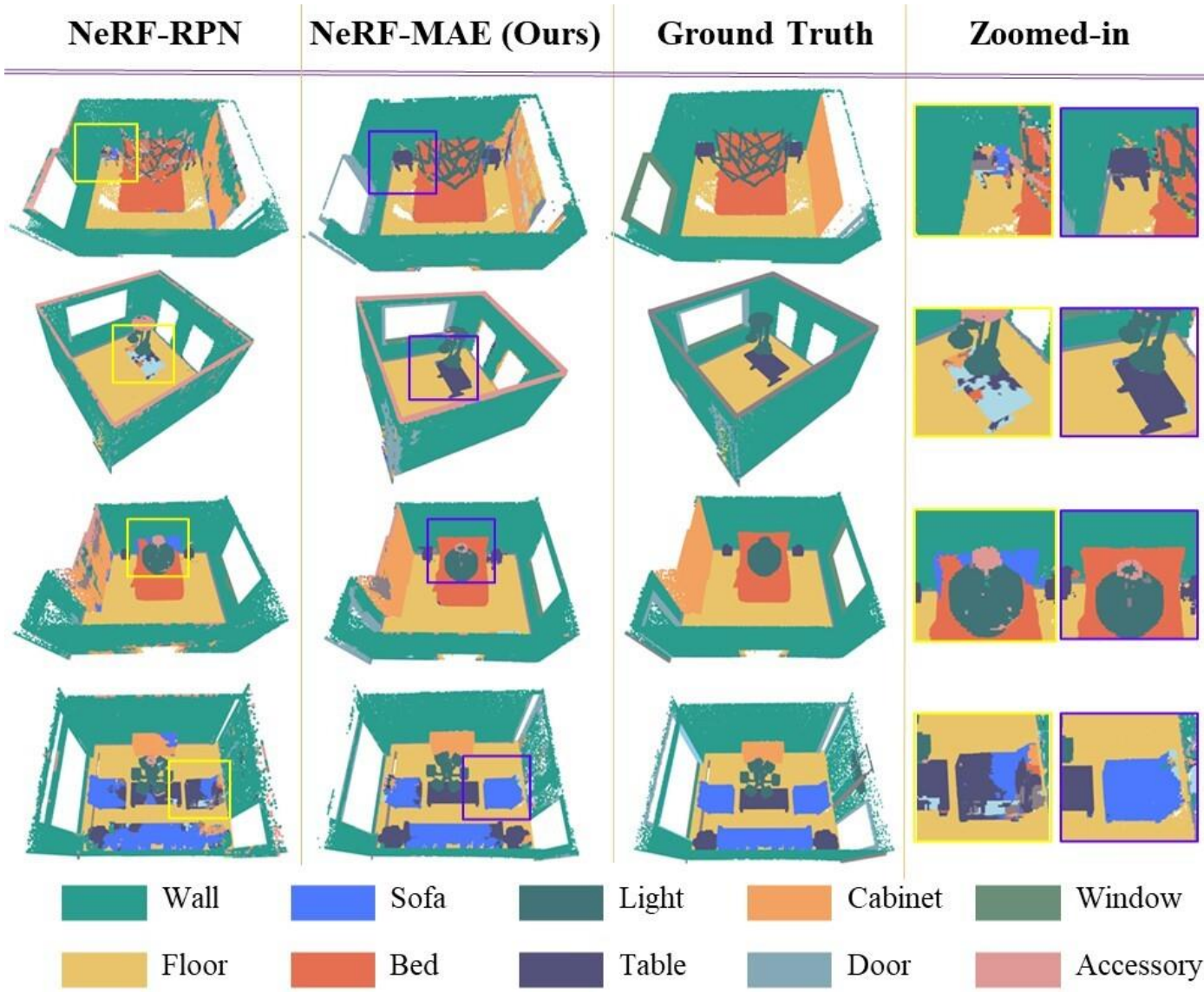
Results Analysis





Qualitative Results





NeRF-MAE:

Masked AutoEncoders for Self-Supervised 3D Representation Learning for Neural Radiance Fields

European Conference on Computer Vision, ECCV 2024
also appeared at CVPR Neural Rendering Intelligence Workshop, CVPR 2024