Grid-guided Neural Radiance Fields for Large Urban Scenes

Linning Xu, Yuanbo Xiangli, Sida Peng, Xingang Pan,
Nanxuan Zhao, Christian Theobalt, Bo Dai, Dahua Lin
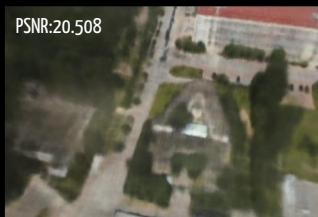
(WED-AM-007)

Grid-guided Neural Radiance Fields for Large Urban Scenes

PSNR:20.508

PSNR:21.511

PSNR:24.893

PSNR:24.970

Two-branch outputs

NeRF-based (MLP-based): Blurry, no fine detail
- <u>Cause</u>: Capacity limitation
- <u>Possible remedy</u>: BlockNeRF, Mega-NeRF
  Geographically divide scenes and assign each region a different sub-NeRF to learn in parallel
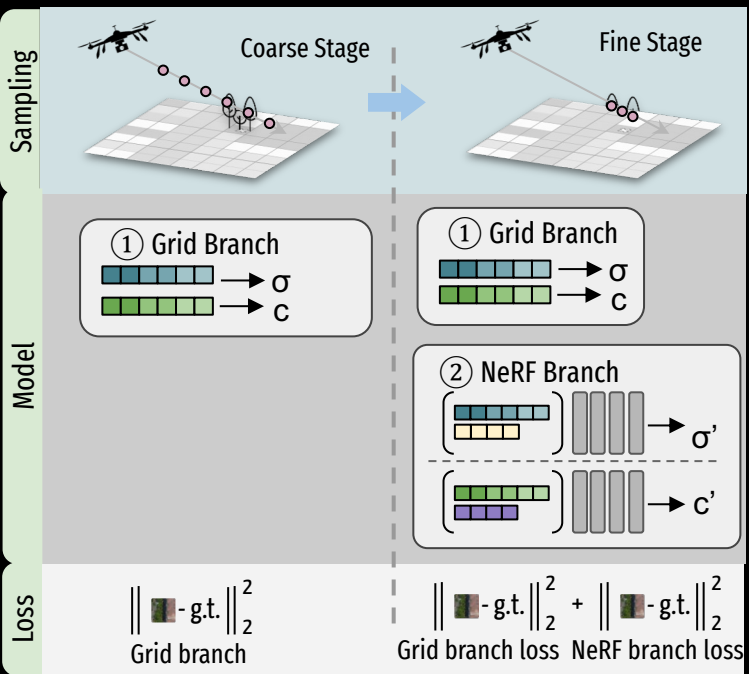- <u>Drawback</u>: Linear scaled number of sub-NeRFs

Grid-based: Noisy, lack of continuity, lack fine details
- <u>Cause</u>: independently optimized grid feature, limited grid resolution
- <u>Possible remedy</u>: instant-NGP, TensoRF
  Increase grid resolution according to scene, use small MLP renderer to translate features
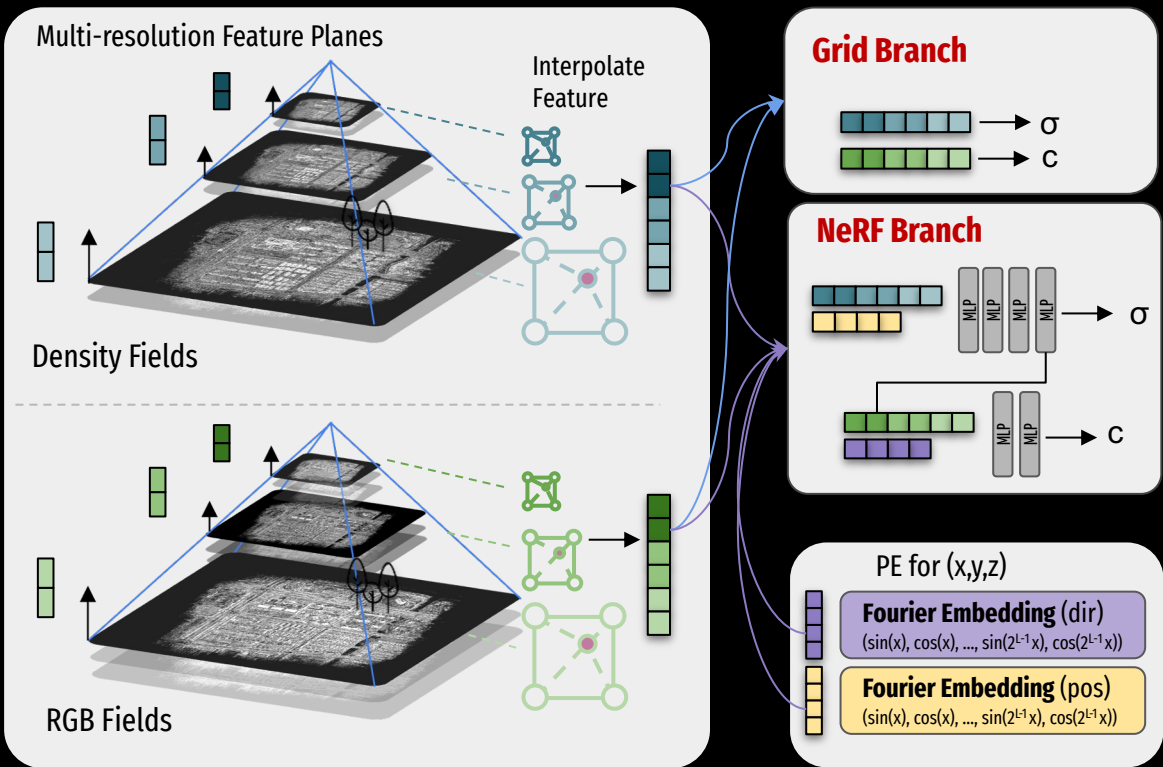- <u>Drawback</u>: heavy on memory, MLP learning ability

Grid-guided Neural Radiance Fields for Large Urban Scenes

Stage1: Pre-training Grid Branch

Stage2: Joint Learning Grid- and NeRF- Branches

Sampling

Coarse Stage   Fine Stage

Model

① Grid Branch → σ, c

① Grid Branch → σ, c

② NeRF Branch → σ', c'

Loss

$\left\| \blacksquare - g.t. \right\|_2^2$

Grid branch

$\left\| \blacksquare - g.t. \right\|_2^2 + \left\| \blacksquare - g.t. \right\|_2^2$

Grid branch loss   NeRF branch loss

(a) Two Stage Training Scheme

Multi-resolution Feature Planes

Interpolate Feature

Density Fields

RGB Fields

(b) Two Branch Model Design

Grid Branch → σ, c

NeRF Branch → σ, c

PE for (x,y,z)

Fourier Embedding (dir)
$(\sin(x), \cos(x), ..., \sin(2^{L-1}x), \cos(2^{L-1}x))$

Fourier Embedding (pos)
$(\sin(x), \cos(x), ..., \sin(2^{L-1}x), \cos(2^{L-1}x))$
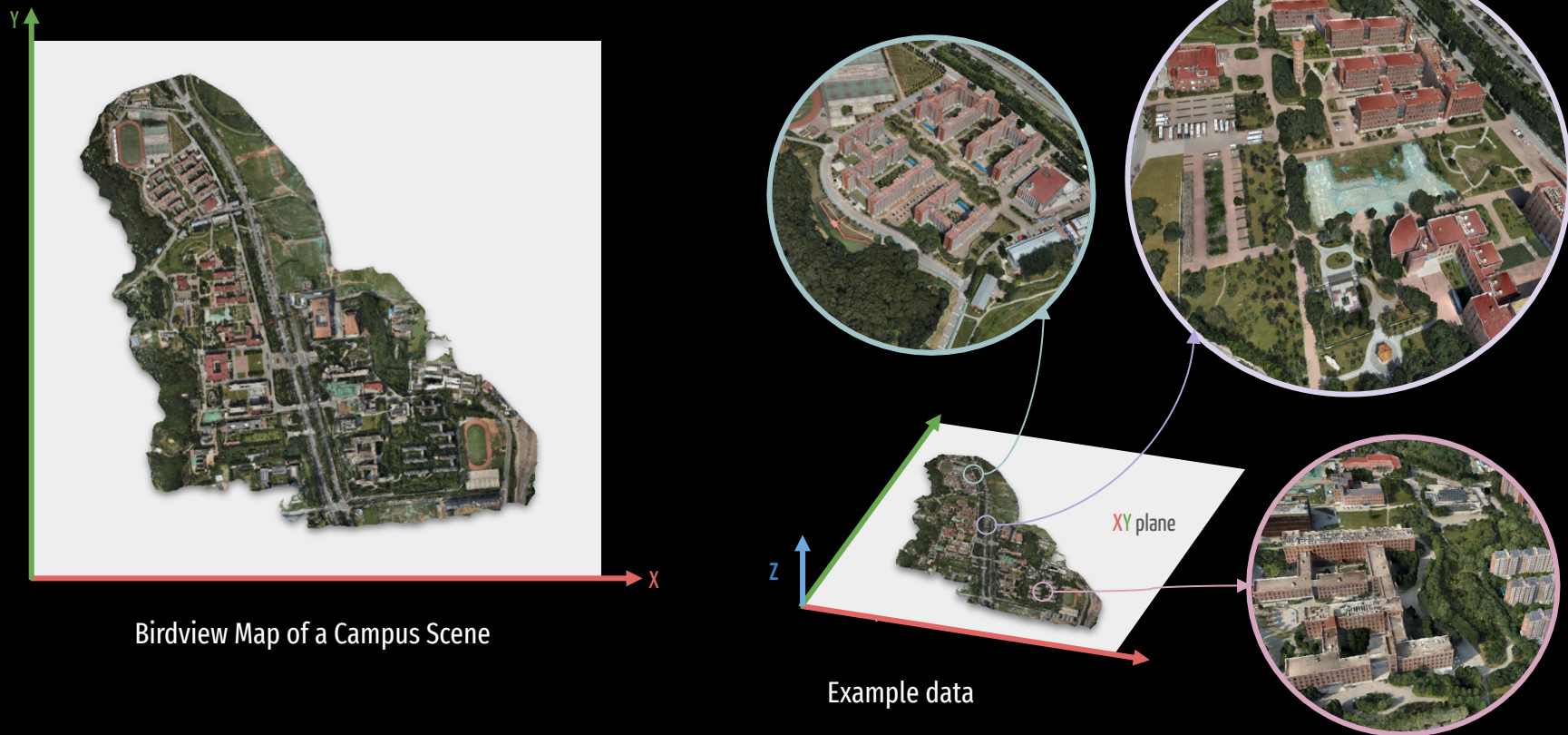
😎 Roam the City！✈️

Model view (extracted point cloud)
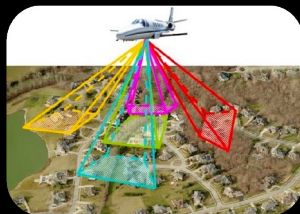
Demo Trajectory view

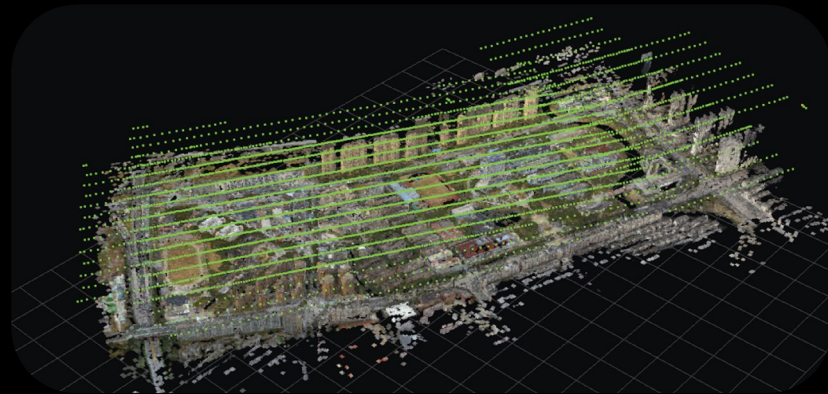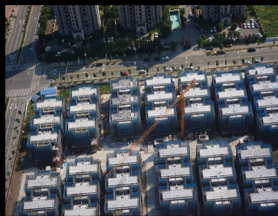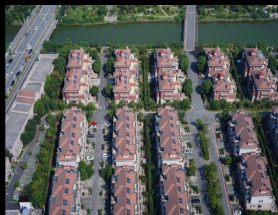# Target Scenario: Large-scale Urban Scenes

Birdview Map of a Campus Scene

Example data

Oblique Photography

Set of 5 images captured per frame
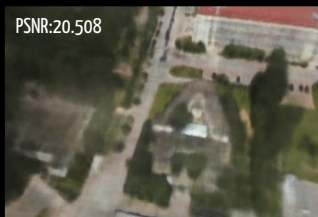

Example camera trajectories

Grid-guided Neural Radiance Fields for Large Urban Scenes

Example View:	NeRF: over-smooth+slow	Feature Grid: noisy+memory cost	Ours: photorealistic

PSNR:20.508
PSNR:21.511
PSNR:24.893
PSNR:24.970

&

NeRF-based (MLP-based): Blurry, no fine detail
- <u>Cause</u>: Capacity limitation
- <u>Possible remedy</u>: BlockNeRF, Mega-NeRF
  Geographically divide scenes and assign each region a different sub-NeRF to learn in parallel
- <u>Drawback</u>: Linear scaled number of sub-NeRFs

Grid-based: Noisy, lack of continuity, lack fine details
- <u>Cause</u>: independently optimized grid feature, limited grid resolution
- <u>Possible remedy</u>: instant-NGP, TensoRF
  Increase grid resolution according to scene, use small MLP renderer to translate features
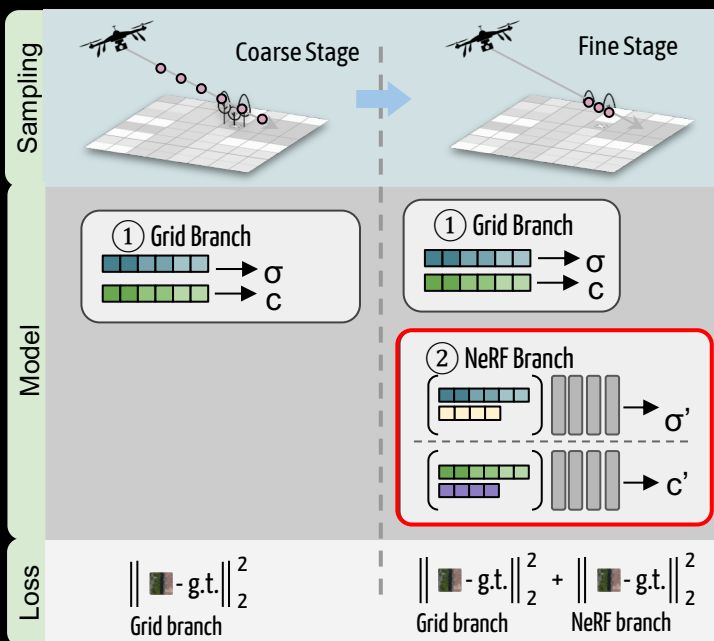- <u>Drawback</u>: heavy on memory, MLP learning ability

🤔

Grid-guided Neural Radiance Fields for Large Urban Scenes

- Multi-resolution Feature Grid Pre-train
- Grid-guided Neural Radiance Field (NeRF-branch)
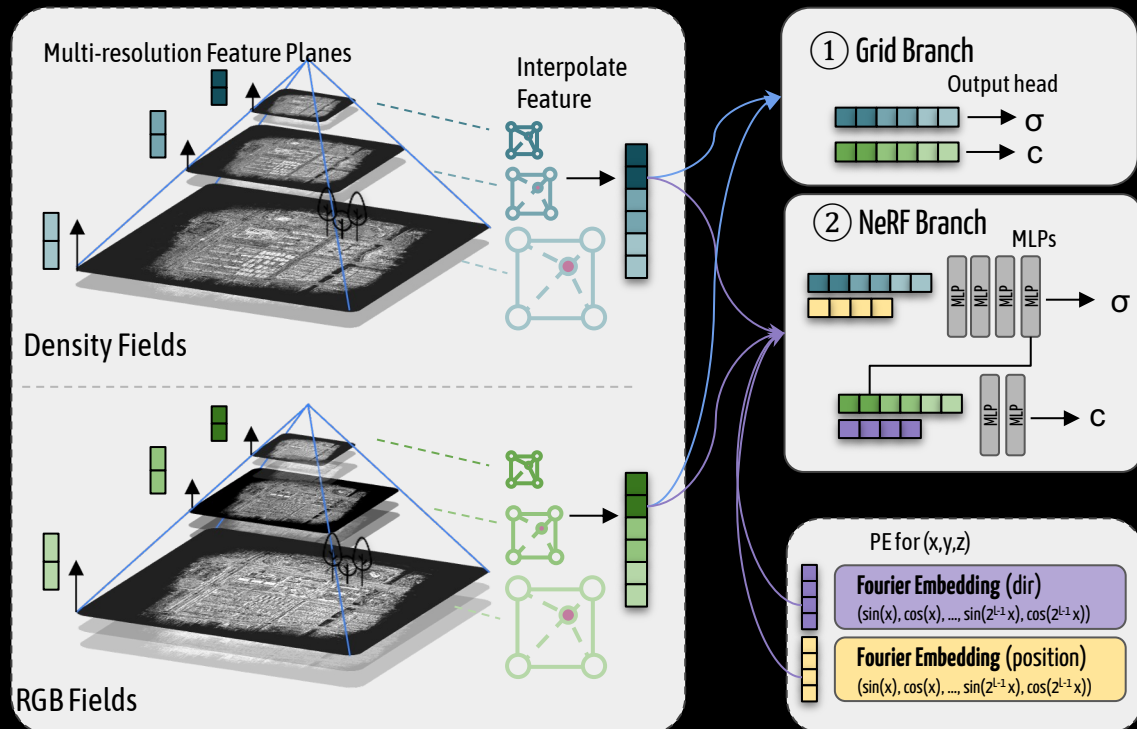- Refined Grid Feature from NeRF (Grid-branch)

$$\sigma', c' = F'(\mathcal{G}_\sigma(X), \mathcal{G}_c(X), \text{PE}(X), \text{PE}(d))$$

(a) Two Stage Training Scheme

(b) Two Branch Model Design

Grid-guided Neural Radiance Fields for Large Urban Scenes

Grid-guided Neural Radiance Field (NeRF-branch)
- Grid density guide NeRF sampling: compress sampling space to surface
- Enrich PE with intermediate features of scene content (multi-res feature)

Refined Grid Feature from NeRF (Grid-branch)
- <u>Recall</u>: feature grid uses bilinear interpolation to infer features of points in a voxel, not enough incentive from merely pixel-wise MSE loss
NeRF to enhance supervision signal with point-wise guidance
- <u>Recall</u>: grid features are independently optimized, lack of spatial continuity & semantic similarity -> noisy visual
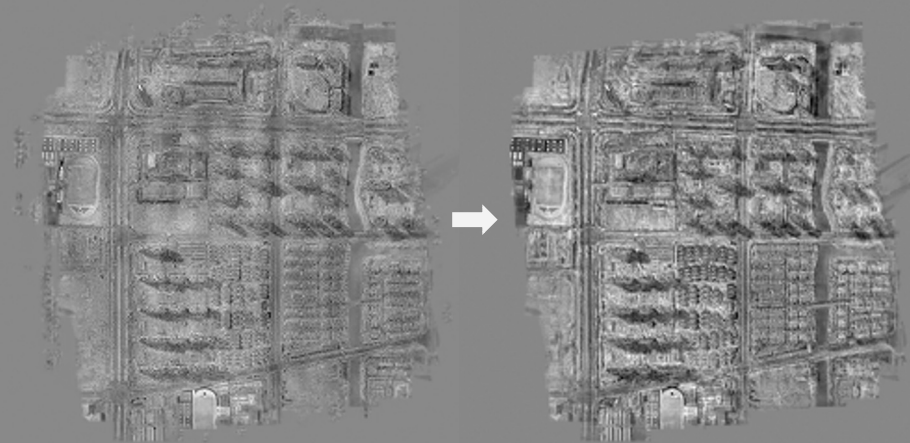NeRF provides global regularization

Design Efficacy:

Refined grid feature: less noisy, sharper edges, and regular shapes of objects; appearance feature can capture shadows
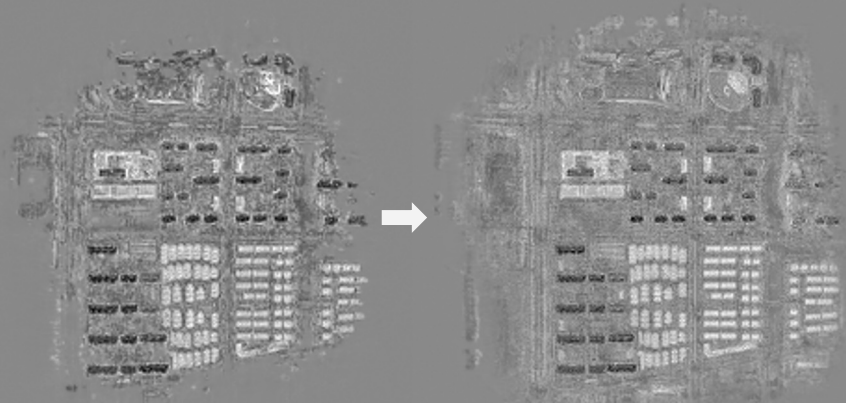
(a) Density feature plane (pre-trained -> refined)

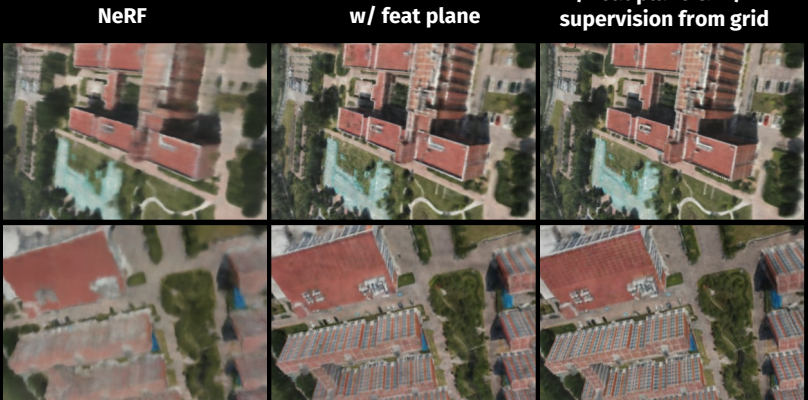(b) Appearance feature plane (pre-trained -> refined)

Design Efficacy:

Refined grid feature: less noisy, sharper edges, and regular shapes of objects; appearance feature can capture shadows
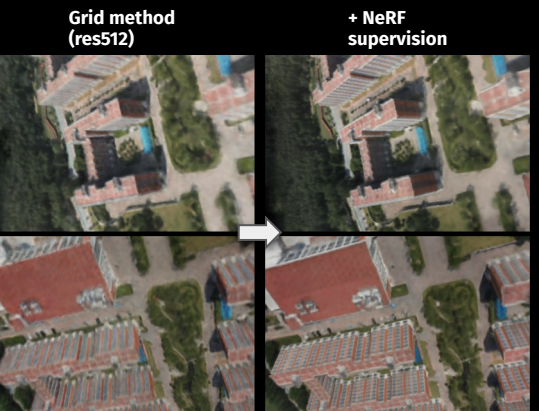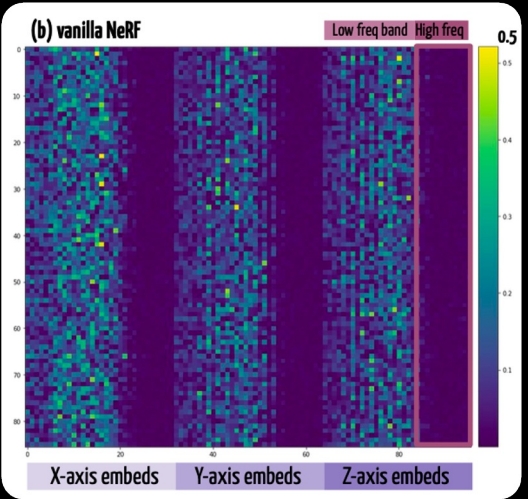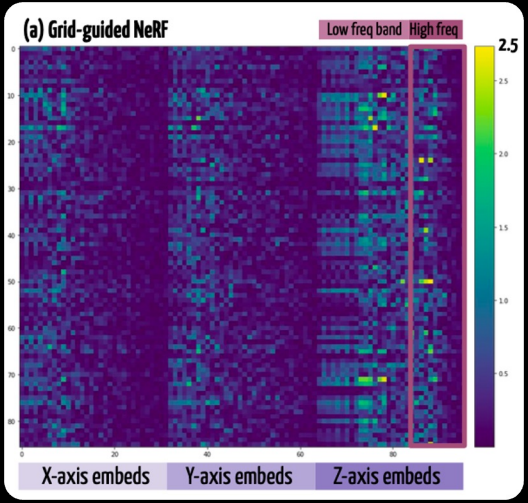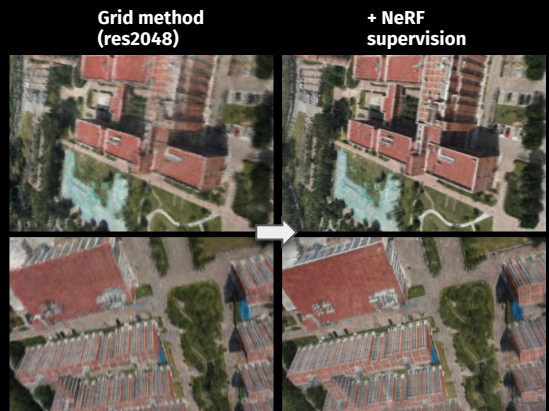
Design Efficacy:

Grid feature enriched NeRF:

NeRF | w/ feat plane | w/ feat plane & w/ supervision from grid

NeRF picks up fine details in voxel:

Grid method (res512) | + NeRF supervision

NeRF adds global regularization:

Grid method (res2048) | + NeRF supervision

(a) Grid-guided NeRF — Low freq band / High freq
X-axis embeds | Y-axis embeds | Z-axis embeds

(b) vanilla NeRF — Low freq band / High freq
X-axis embeds | Y-axis embeds | Z-axis embeds

Demo Novel Views

Two branch comparison

😎 Roam the City !