NeRF在场景重建中的应用

胡晨

2022年5月

保密信息

目录

Contents

保密信息

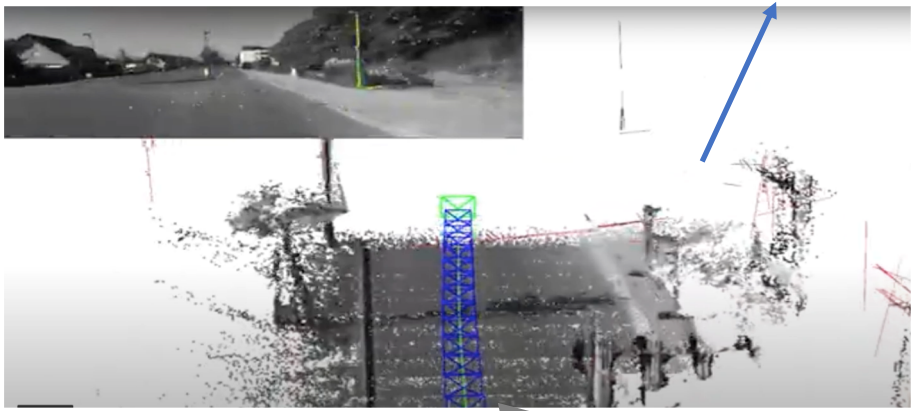
/ 01

Motivation of Scene Reconstruction

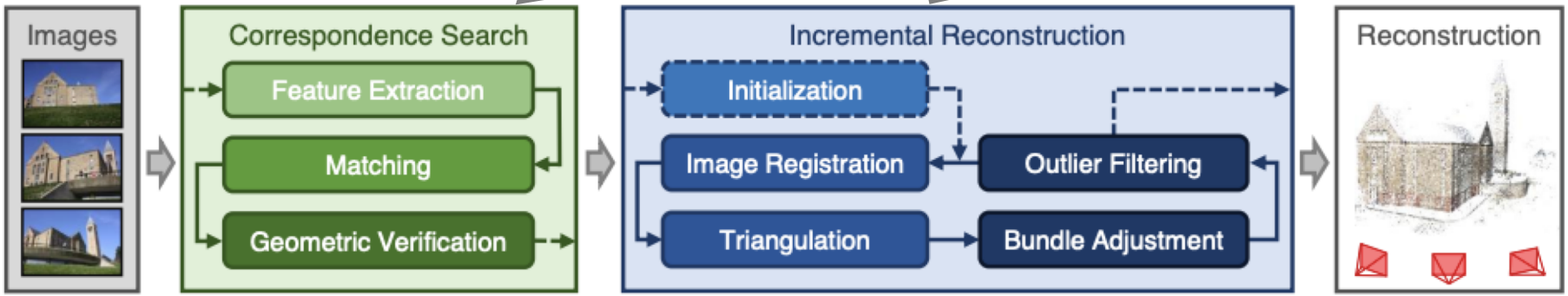
保密信息

Traditional methods

images



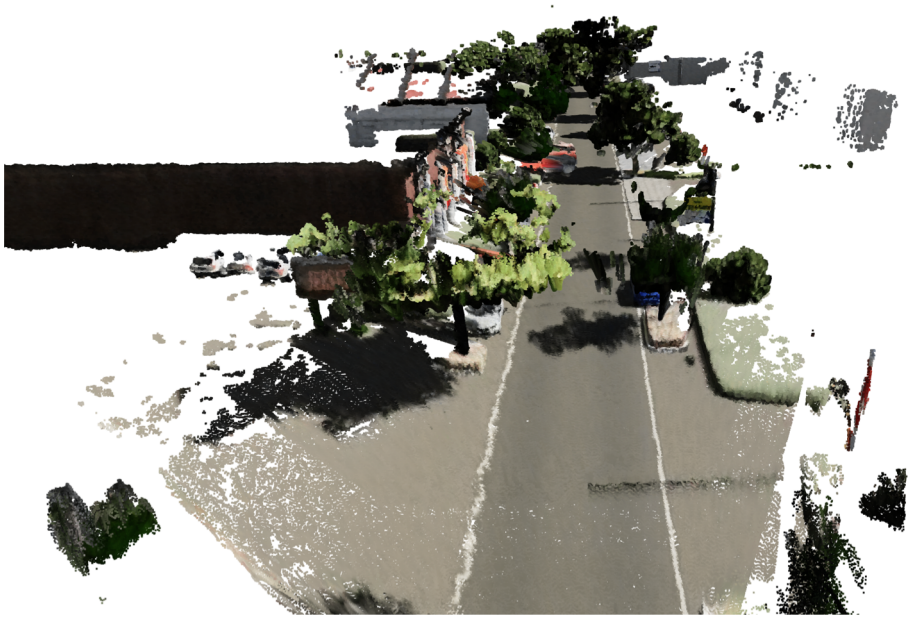
Sparse PointCloud



Schonberger J L, Frahm J M. Structure-from-motion revisited[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 4104-4113.

Scene reconstruction -- NeRF

- Simple pipeline
- Dense scene representation

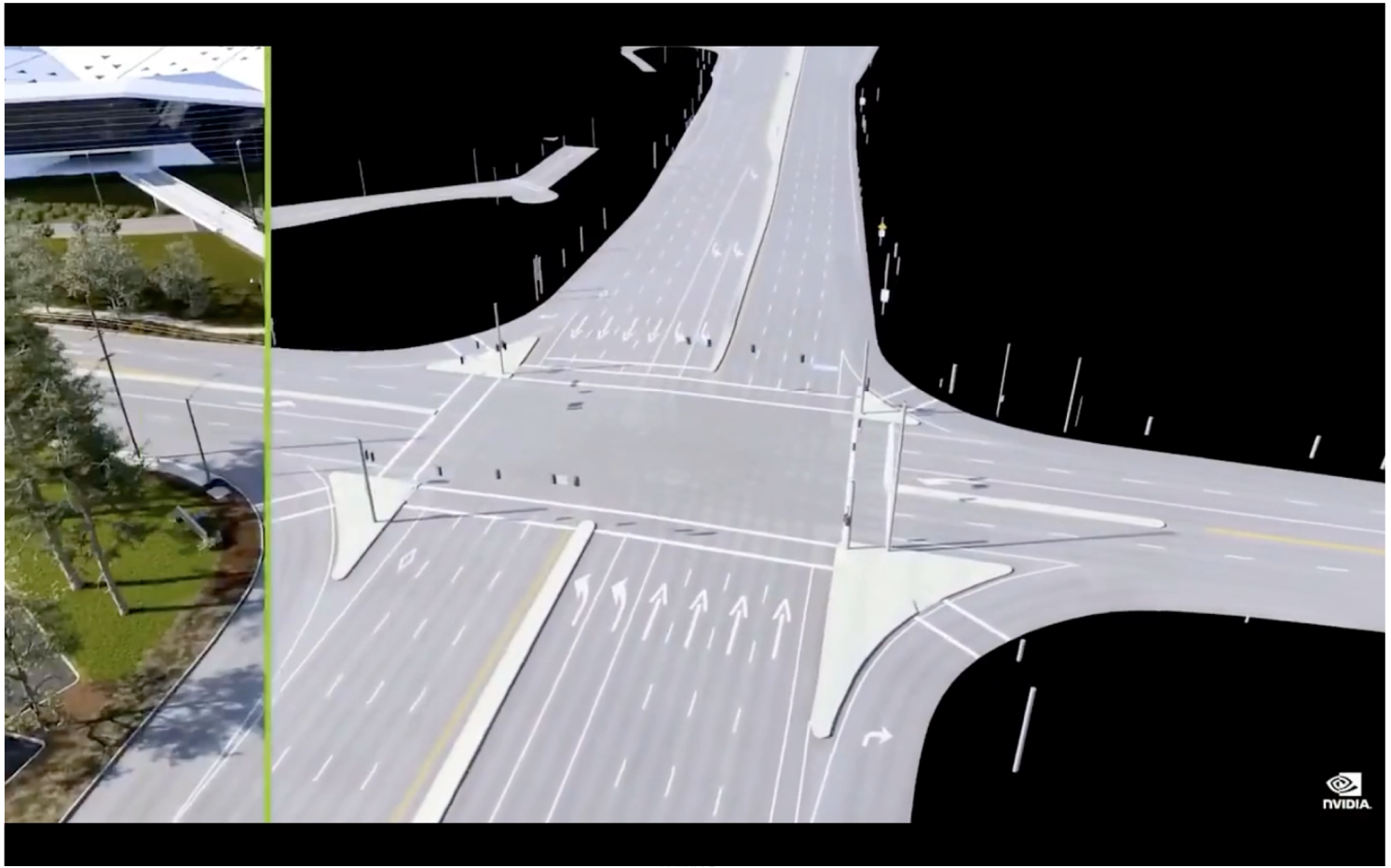


COLMAP
vSLAM
...

Pose → NeRF → Dense PointCloud
Mesh
Multi-view image

保密信息

Applications -- Autodrive



2022 Nvidia Explains Self-Driving Car Vision at GTC

Applications -- Metaverse and game

Tancik M, Casser V, Yan X, et al. Block-nerf: Scalable large scene neural view synthesis[J]. arXiv preprint arXiv:2202.05263, 2022.

MEGVII 旷视

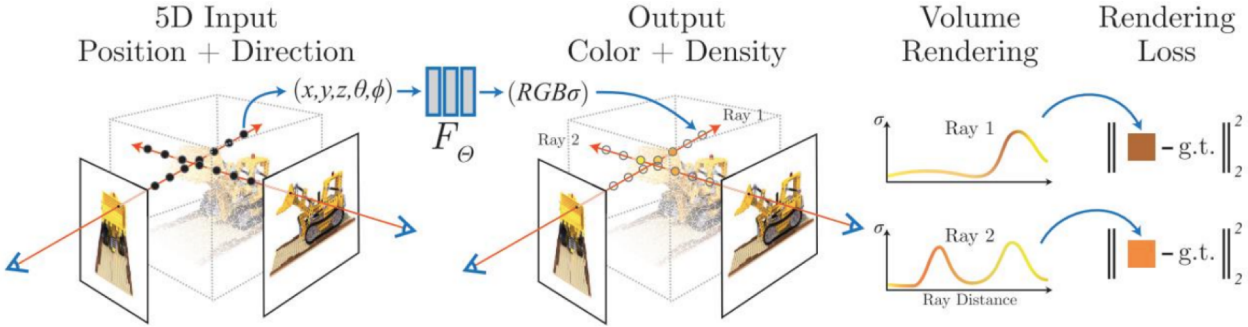
02

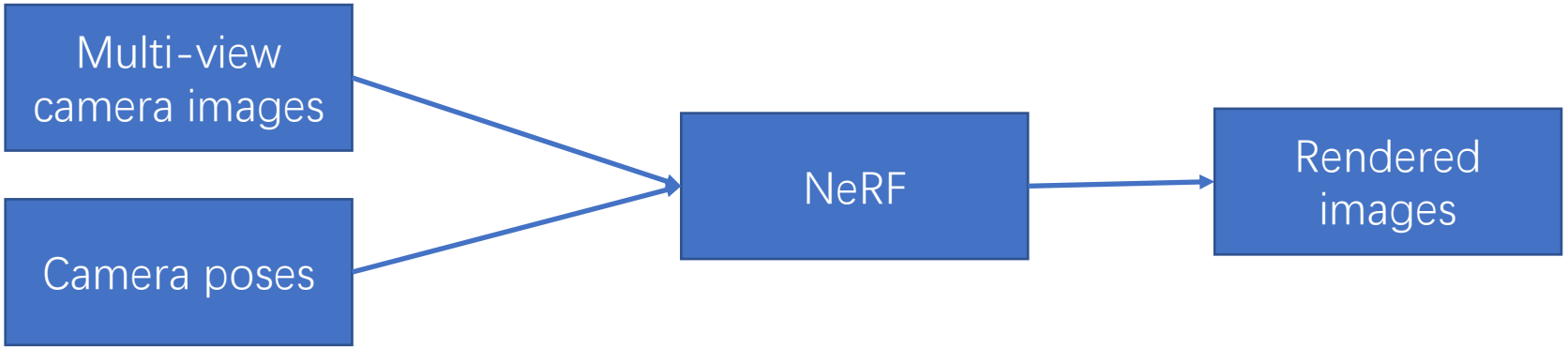
Challenges in Scene Reconstruction

保密信息

NeRF and 3D Reconstruction

- Neural Radiance Fields (NeRF)

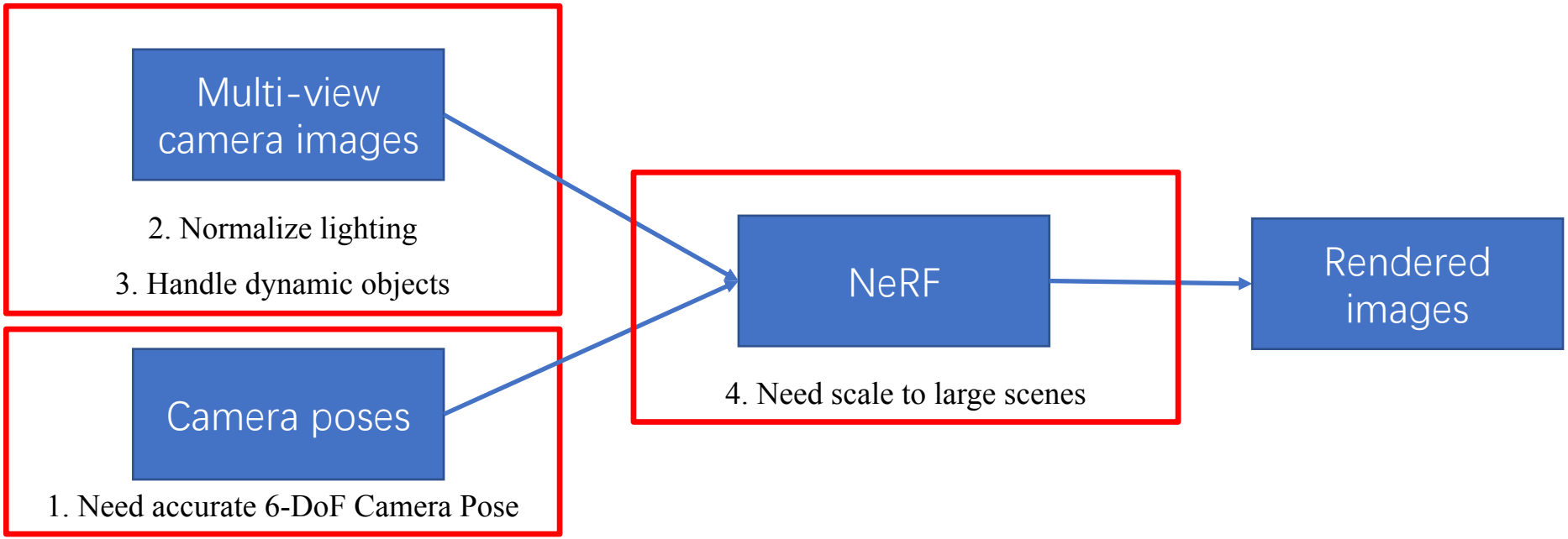




Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[C]//European conference on computer vision. Springer, Cham, 2020: 405-421.

Challenges in NeRF Reconstruction

- Neural Radiance Fields (NeRF)



Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[C]//European conference on computer vision. Springer, Cham, 2020: 405-421.

MEGVII 旷视

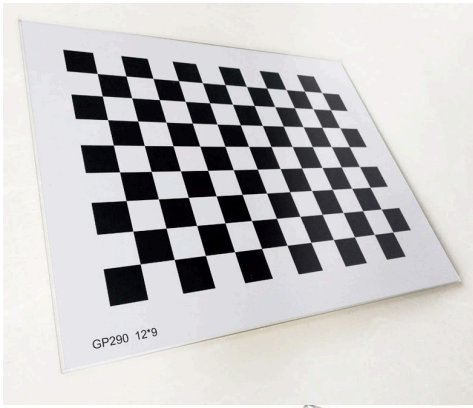
/ 03

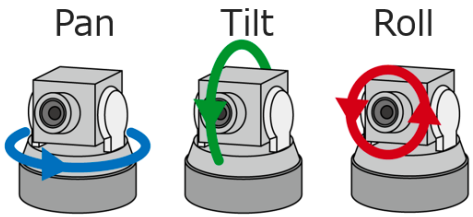
Neural approach to Scene Reconstruction

保密信息

Need accurate 6-DoF Camera Pose

- Camera Calibration Boards：
  - ArUco Board
- SLAM:
  - Visual SLAM
  - Lidar, IMU...
- SFM:
  - COLMAP
- NeRF:
  - iNeRF

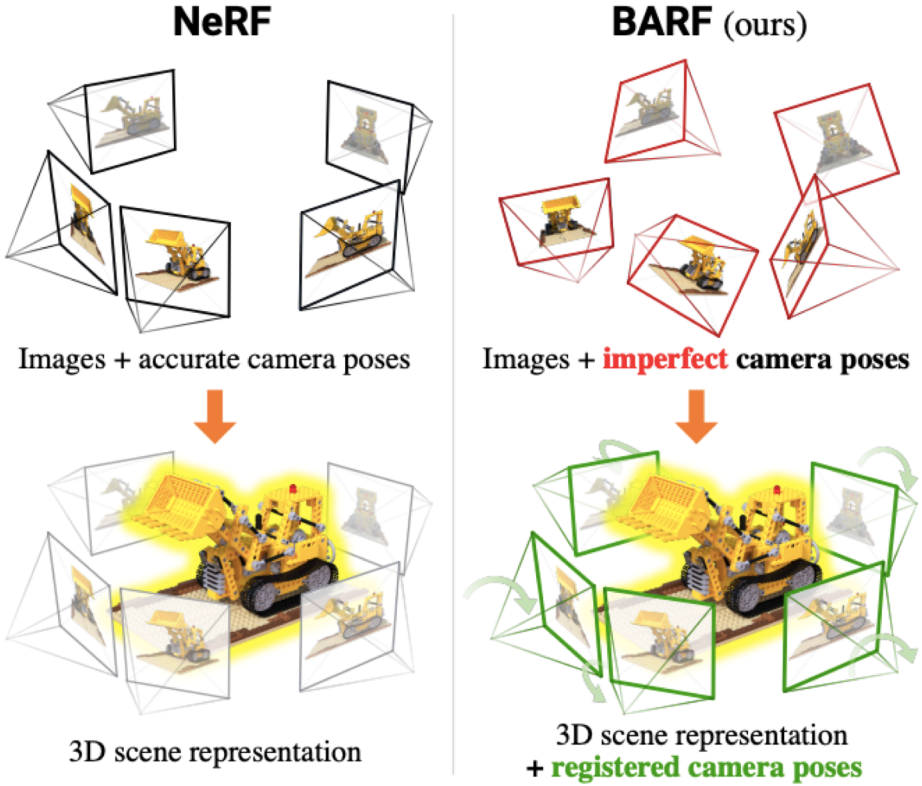




+ (x, y, z)

Pose Optimization -- BARF

- Finetune image poses with Bundle Adjustment
  - Use imperfect (with noise) camera poses
  - Coarse-to-fine registration on coordinate-based scene representations

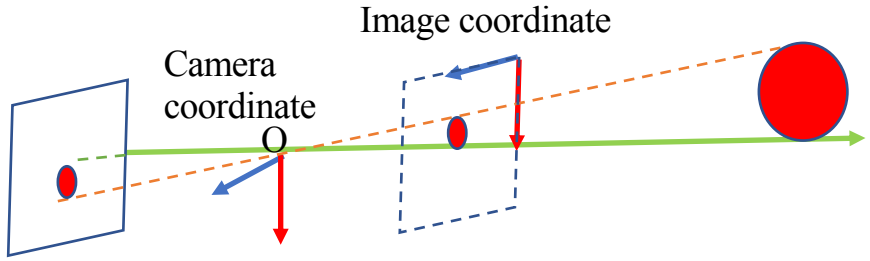


Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- BARF

Coordinate systems

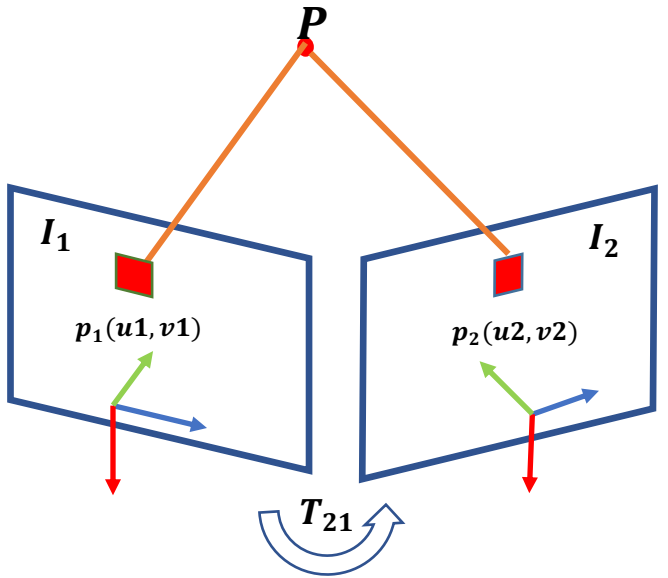
- Image coordinate system (2d)
- Camera coordinate system (3d)
- World coordinate system (3d)



CameraToImage transform:

$$p1 = \begin{bmatrix} u1 \\ v1 \\ 1 \end{bmatrix} = \frac{1}{Z_1} KP \qquad K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}$$

$$p2 = \begin{bmatrix} u2 \\ v2 \\ 1 \end{bmatrix} = \frac{1}{Z_2} K(R_{21}P + t_{21}) = \frac{1}{Z_2} K(T_{21}P)_{1:3}$$



Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- BARF

- Bundle Adjustment for pose optimization:

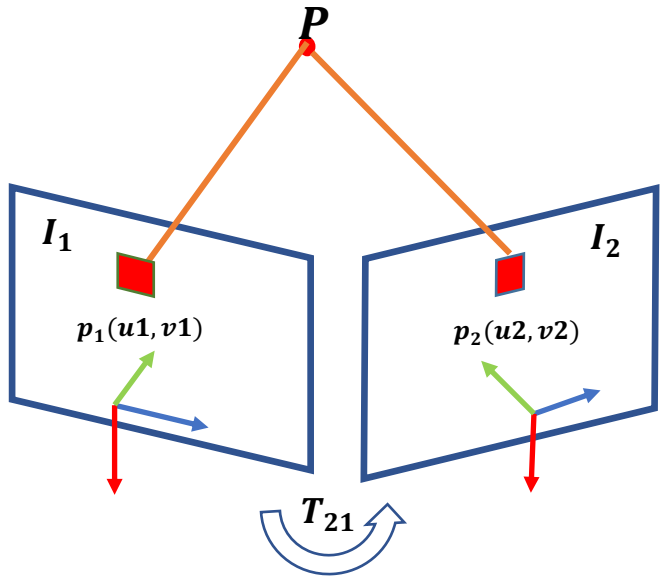
$$e_i = I_1(p_{1,i}) - I_2(p_{2,i})$$

$$min_T J(T) = ||e||^2 = \sum_{i=1}^{N} e_i^T e_i$$

$q = T_{21}P$，q为P在第二个相机坐标系的坐标

$u = \frac{1}{Z_2}Kq$，u为P在第二个相机图像坐标的像素坐标

ξ 是位姿 T 的李代数表示



$$J = \frac{\partial e}{\partial T} = \frac{\partial I_2}{\partial u}\frac{\partial u}{\partial q}\frac{\partial q}{\partial \delta\xi}\delta\xi$$

Gauss-Newton

$$H\Delta\xi = b$$

$$H = JJ^T, b = -Je$$

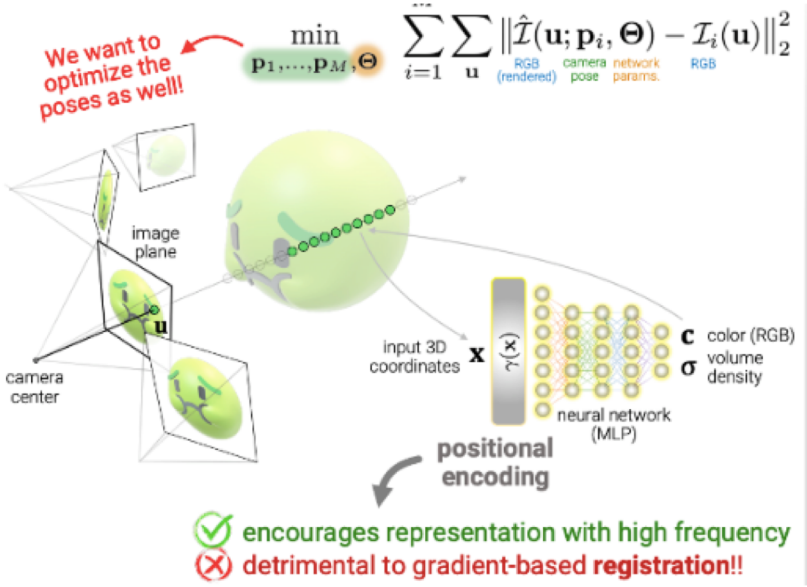
Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- BARF

- Train BARF in Torch3d/LieTorch:

```
pose_refinement = torch.nn.Embedding(n_views, 6)
torch.nn.init.zeros_(pose_refinement.weight)
```

$$P = \Delta P + P$$



Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- BARF

- Positional encoding is good for reconstruction, but ***hurts*** BARF

$$\gamma(\mathbf{x}) = \left[\mathbf{x}, \gamma_0(\mathbf{x}), \gamma_1(\mathbf{x}), \ldots, \gamma_{L-1}(\mathbf{x})\right] \in \mathbb{R}^{3+6L} ,$$

where the k-th frequency encoding $\gamma_k(\mathbf{x})$ is

$$\gamma_k(\mathbf{x}) = \left[\cos(2^k\pi\mathbf{x}), \sin(2^k\pi\mathbf{x})\right] \in \mathbb{R}^6 ,$$

$$\frac{\partial\gamma_k(\mathbf{x})}{\partial\mathbf{x}} = 2^k\pi \cdot \left[-\sin(2^k\pi\mathbf{x}), \cos(2^k\pi\mathbf{x})\right]$$

(x, y, z)

$x,\ y,\ z$

$\sin(2^0\pi x),\ \sin(2^0\pi y),\sin(2^0\pi z)$,

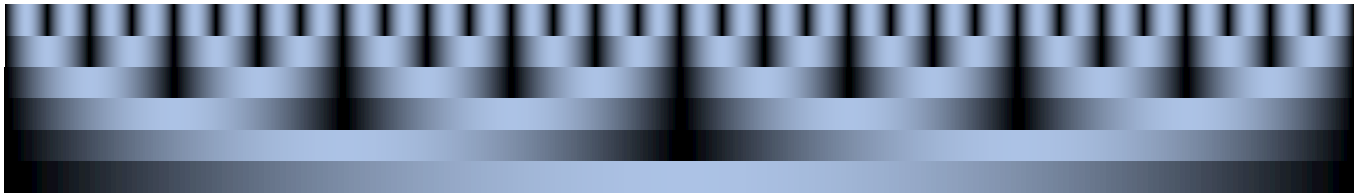
$\cos(2^0\pi x),\ \cos(2^0\pi y),\ \cos(2^0\pi z)$,

$\sin(2^1\pi x),\ \sin(2^1\pi y),\ \sin(2^1\pi z)$,

$\cos(2^1\pi x),\ \cos(2^1\pi y),\cos(2^1\pi z)$,

$\sin(2^{L-1}\pi x),\ \sin(2^{L-1}\pi y),\sin(2^{L-1}\pi z)$,

$\cos(2^{L-1}\pi x),\ \cos(2^{L-1}\pi y),\cos(2^{L-1}\pi z)$,



Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- BARF

- Positional encoding have some negative effect when using BARF

$$\gamma_k(\mathbf{x};\alpha) = w_k(\alpha)\cdot\big[\cos(2^k\pi\mathbf{x}),\sin(2^k\pi\mathbf{x})\big]\ ,$$

where the weight w_k is defined as

$$w_k(\alpha) = \begin{cases} 0 & \text{if } \alpha < k \\ \dfrac{1-\cos((\alpha-k)\pi)}{2} & \text{if } 0 \le \alpha - k < 1 \\ 1 & \text{if } \alpha - k \ge 1 \end{cases}$$

$$\frac{\partial\gamma_k(\mathbf{x};\alpha)}{\partial\mathbf{x}} = w_k(\alpha)\cdot 2^k\pi\cdot\big[-\sin(2^k\pi\mathbf{x}),\cos(2^k\pi\mathbf{x})\big]$$

(x, y, z)

$x,\ y,\ z$

$\sin(2^0\pi x),\ \sin(2^0\pi y),\sin(2^0\pi z)\ ,$

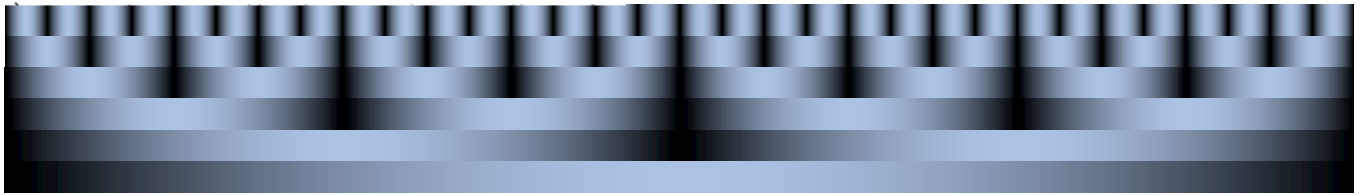
$\cos(2^0\pi x),\ \cos(2^0\pi y),\ \cos(2^0\pi z),$

$\sin(2^1\pi x),\ \sin(2^1\pi y),\ \sin(2^1\pi z),$

$\cos(2^1\pi x),\ \cos(2^1\pi y),\cos(2^1\pi z),$

$\sin(2^{L-1}\pi x),\ \sin(2^{L-1}\pi y),\sin(2^{L-1}\pi z),$

$\cos(2^{L-1}\pi x),\ \cos(2^{L-1}\pi y),\cos(2^{L-1}\pi z),$



Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization – BARF in VKITTI and KITTI

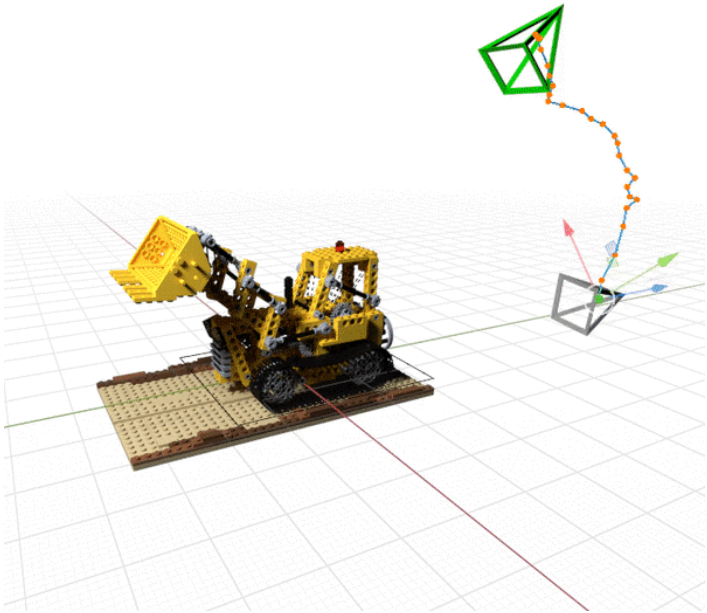
Without BARF

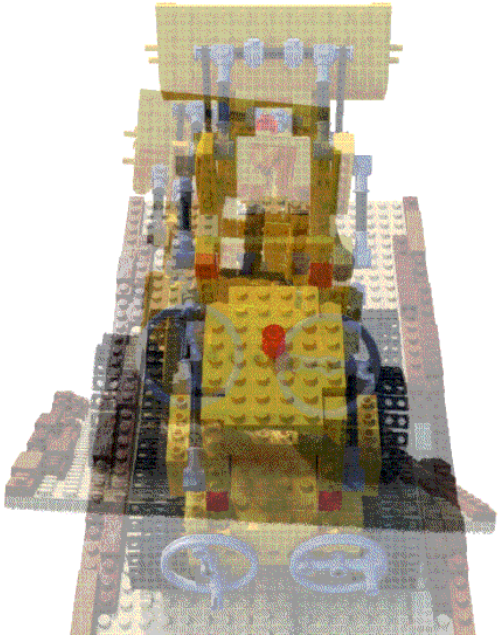
With BARF

Lin C H, Ma W C, Torralba A, et al. Barf: Bundle-adjusting neural radiance fields[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 5741-5751.

Pose Optimization -- iNeRF

- Perform pose estimation when the objects or scenes are represented as trained Neural Radiance Fields





GB/T 7714Yen-Chen L, Florence P, Barron J T, et al. inerf: Inverting neural radiance fields for pose estimation[C]//2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2021: 1323-1330.

Normalize lighting

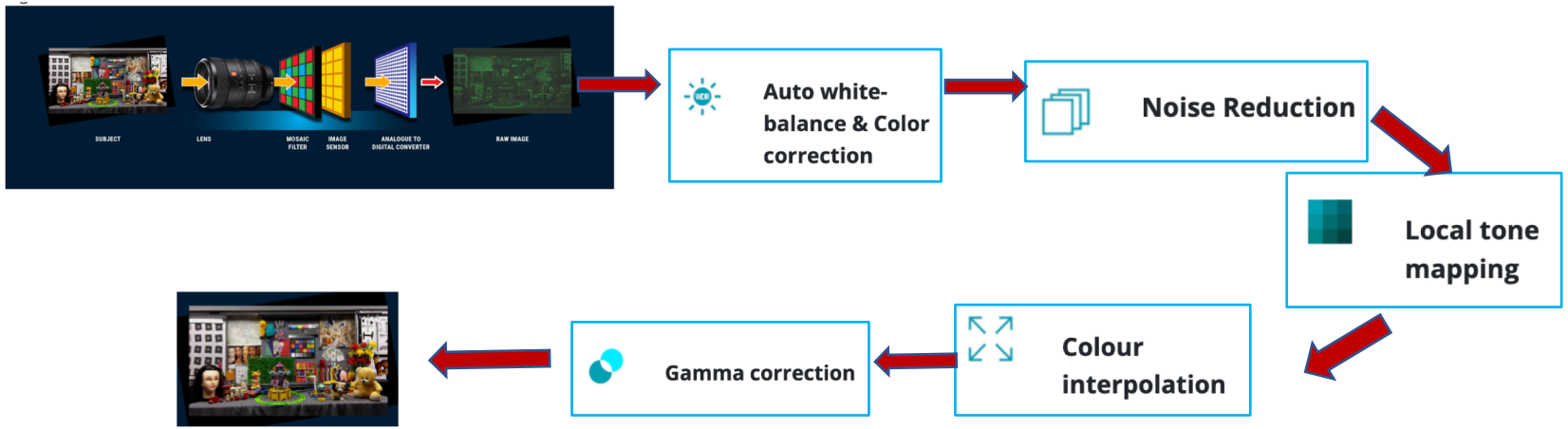
- Same view with different camera settings (exposure, HDR, etc.)
- Same view with different time (Day and night)

(a) Photos (b) Renderings

Martin-Brualla R, Radwan N, Sajjadi M S M, et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7210-7219.

Appearance Optimization

- Add appearance embedding in RGB color space
- Accurately model the postprocess of image (exposure, tone-mapping, focus..)
- Reconstruction directly from RAW image

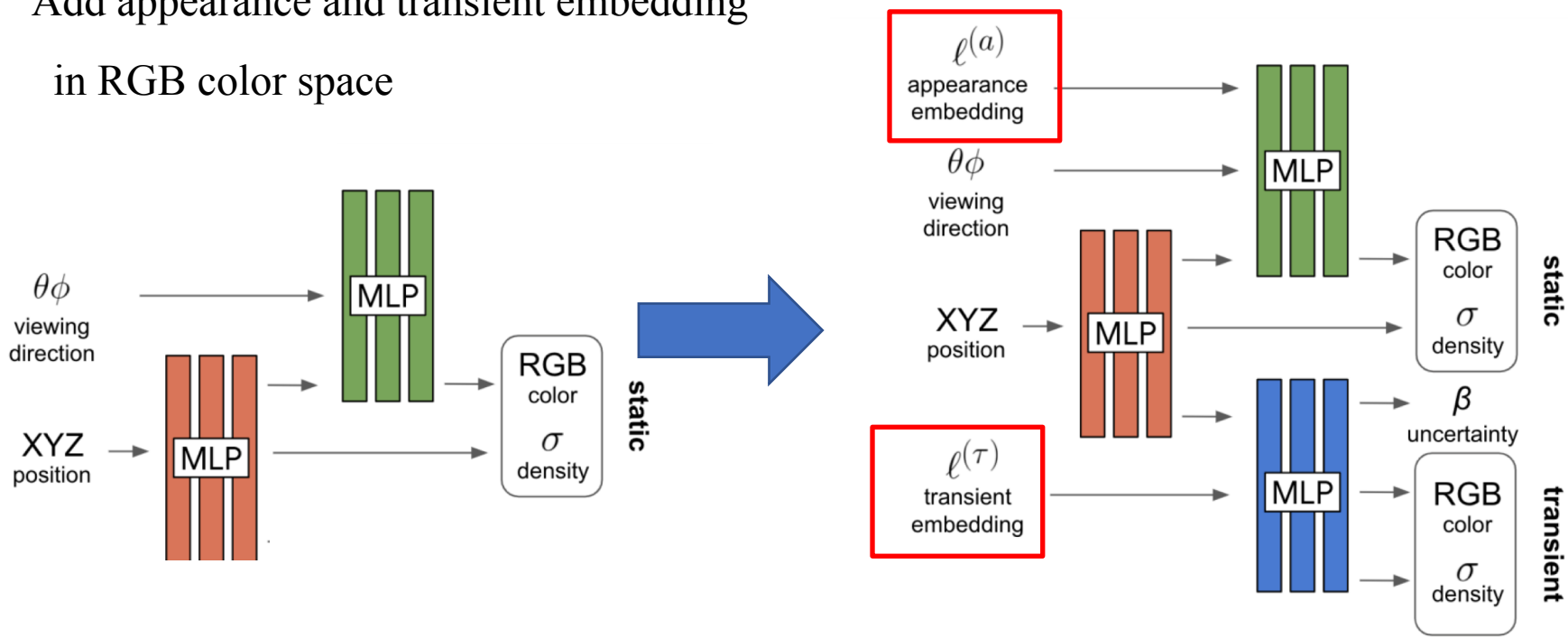


Mildenhall B, Hedman P, Martin-Brualla R, et al. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images[J]. arXiv preprint arXiv:2111.13679, 2021.

Appearance Optimization – NeRF in the wild

- Add appearance and transient embedding

  in RGB color space



Martin-Brualla R, Radwan N, Sajjadi M S M, et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7210-7219.

Appearance Optimization – NeRF in the wild

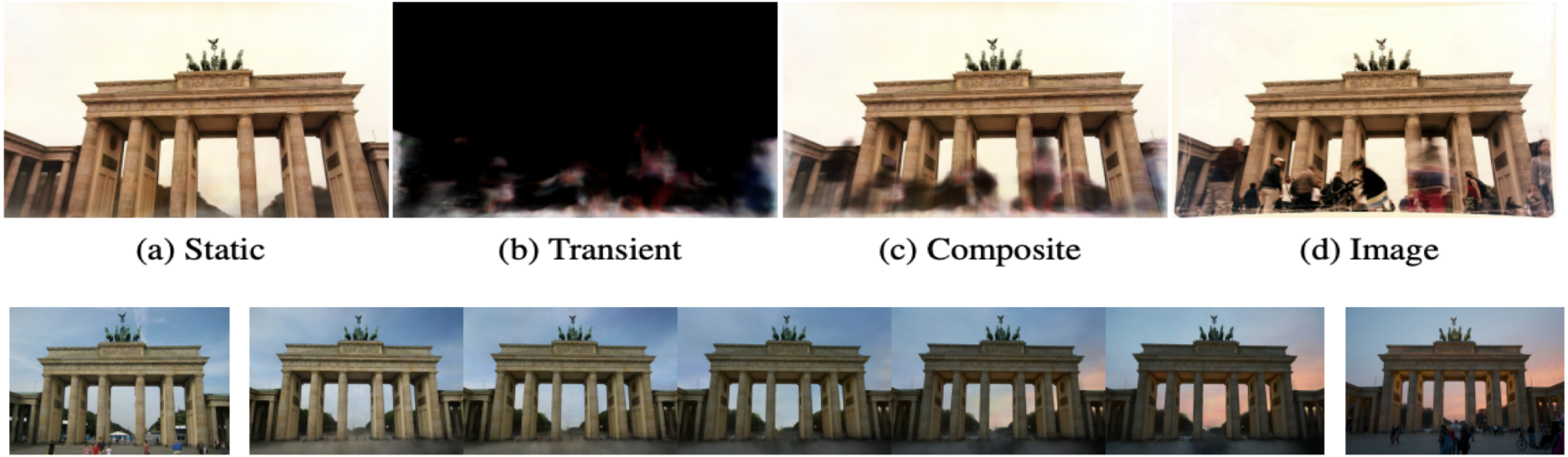
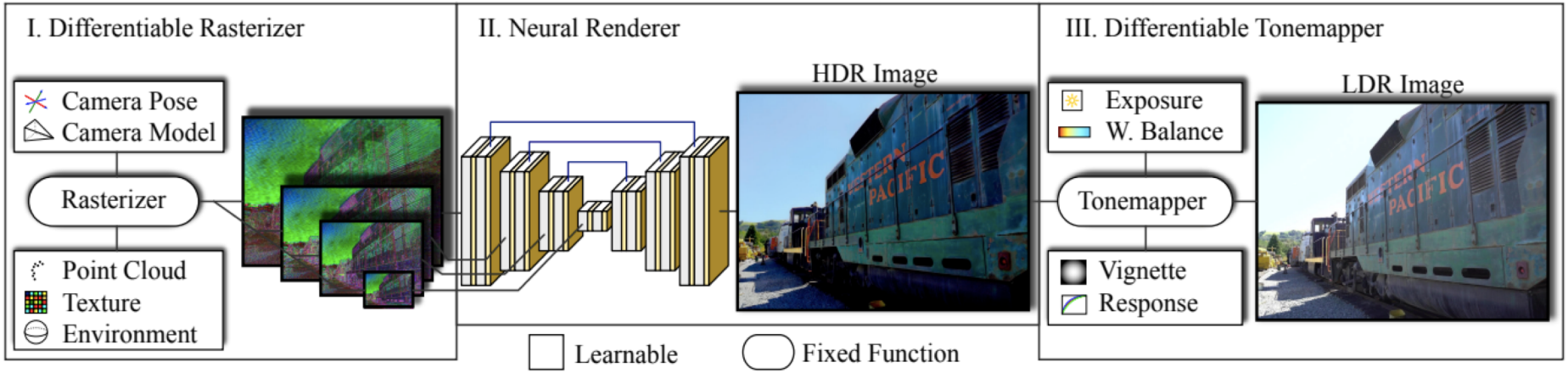


Figure 8: Interpolations between the appearance embeddings $\boldsymbol{\ell}^{(a)}$ of two training images (left, right), which results in renderings (middle) where color and illumination are interpolated but geometry is fixed. Note that the training images contain people (left) and lights (right) that do not appear in the renderings. Photos by Flickr users mightyohm, blatez / CC BY.

Martin-Brualla R, Radwan N, Sajjadi M S M, et al. Nerf in the wild: Neural radiance fields for unconstrained photo collections[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 7210-7219.

Appearance Optimization – ADOP

- Accurately model the ISP of camera (exposure, tone-mapping, out-of-focus)



Rückert D, Franke L, Stamminger M. Adop: Approximate differentiable one-pixel point rendering[J]. arXiv preprint arXiv:2110.06635, 2021.

保密信息

Appearance Optimization

- Many different approaches exist.
- The general idea is to use additional embedding to allow learning to adapt to diverse lightings

rendered appearance image

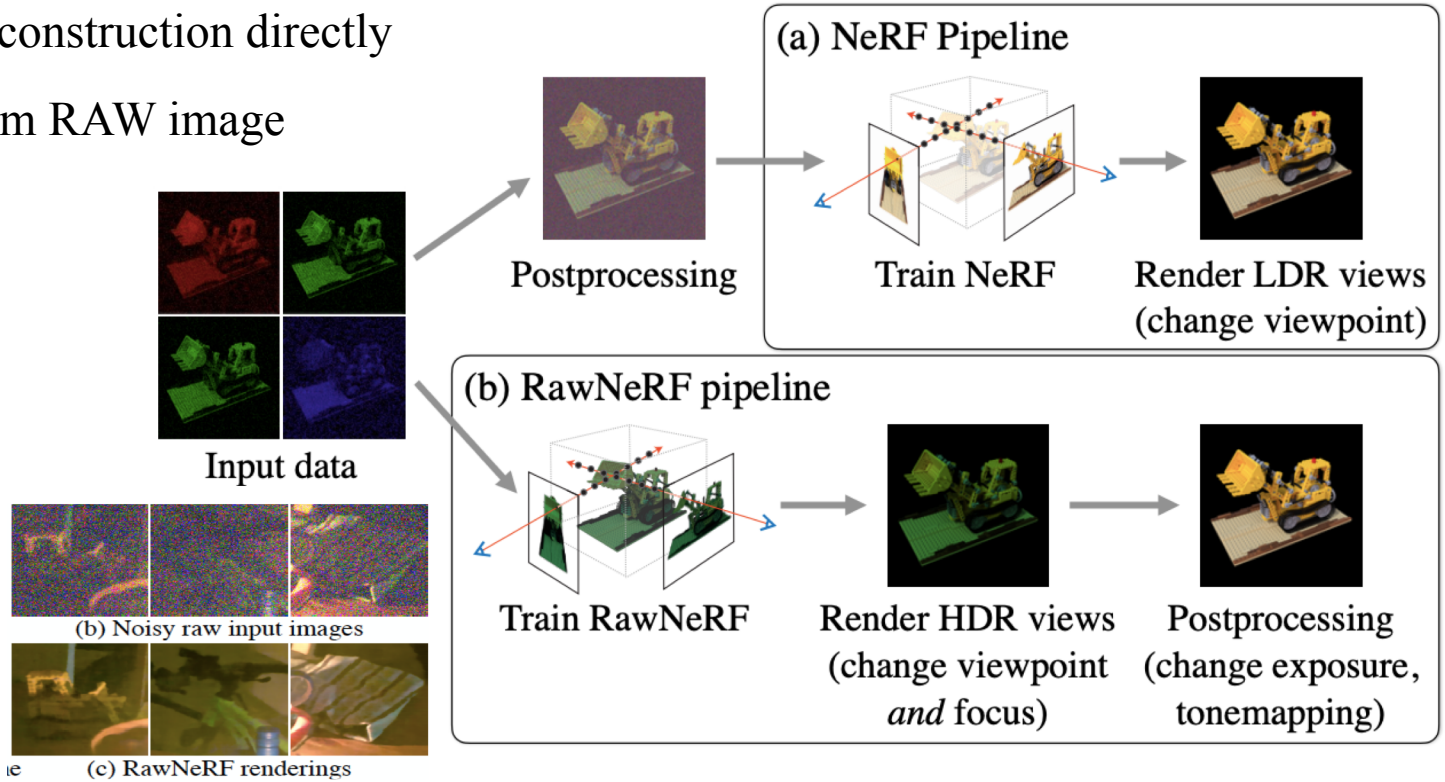
$$Img = \mathrm{Img} * \mathrm{affine}_{1x3}$$

GT

保密信息

Appearance Optimization – NeRF in the dark

- Reconstruction directly from RAW image

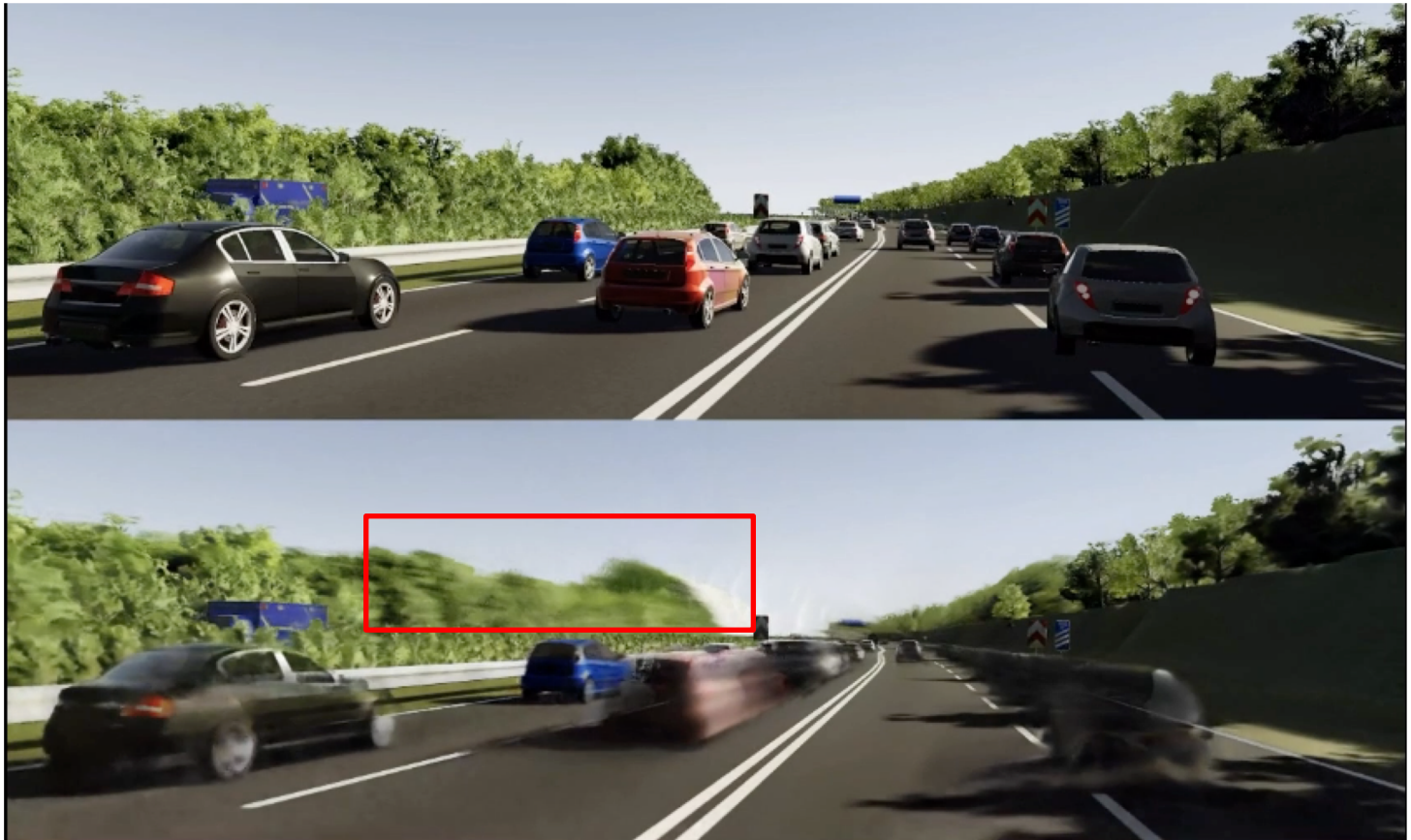


Mildenhall B, Hedman P, Martin-Brualla R, et al. NeRF in the Dark: High Dynamic Range View Synthesis from Noisy Raw Images[J]. arXiv preprint arXiv:2111.13679, 2021.

保密信息

Handle dynamic objects

- Misalignment of dynamic objects can cause blurred renderings（Even w/ GT pose）

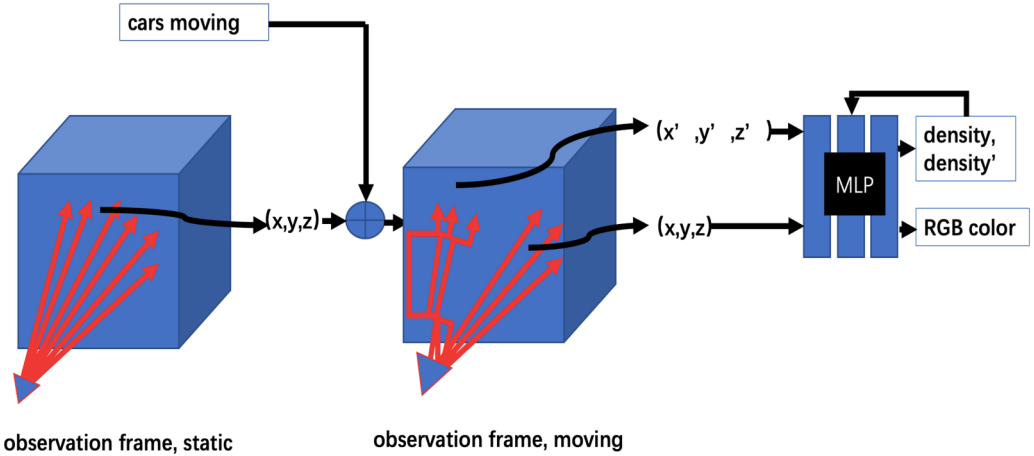


GT

Rendered image

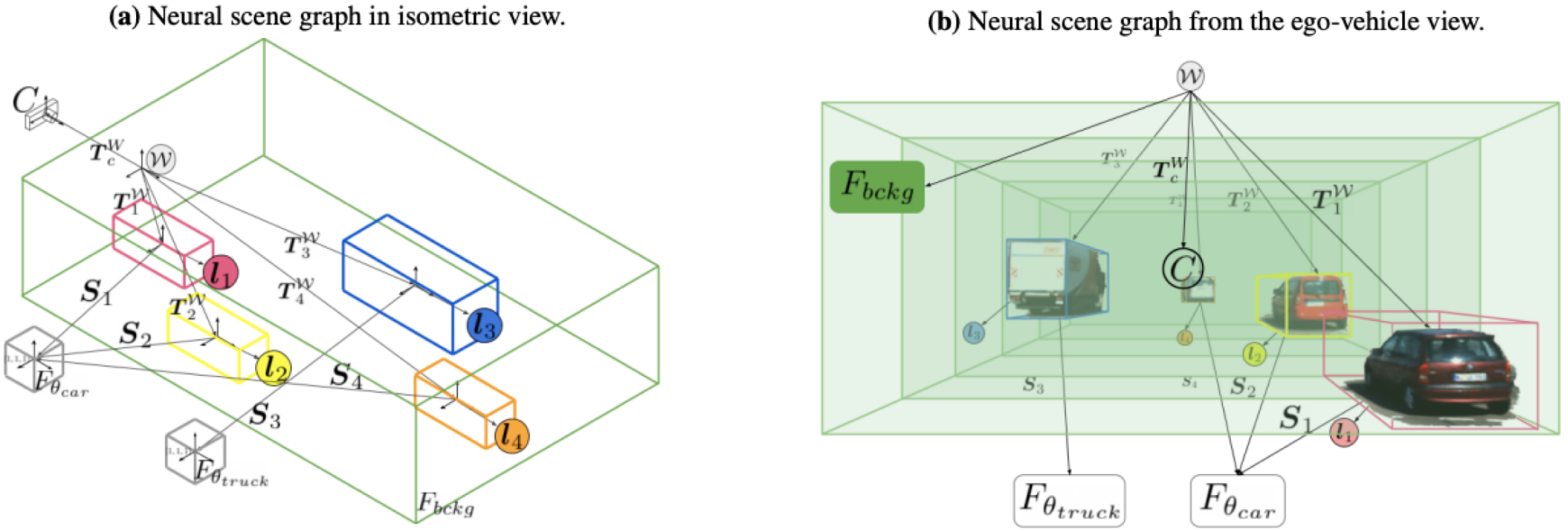
Handle for dynamic objects

- Can simply ignore those using sematic segmentation mask
- Or, use ray transformation for dynamic objects



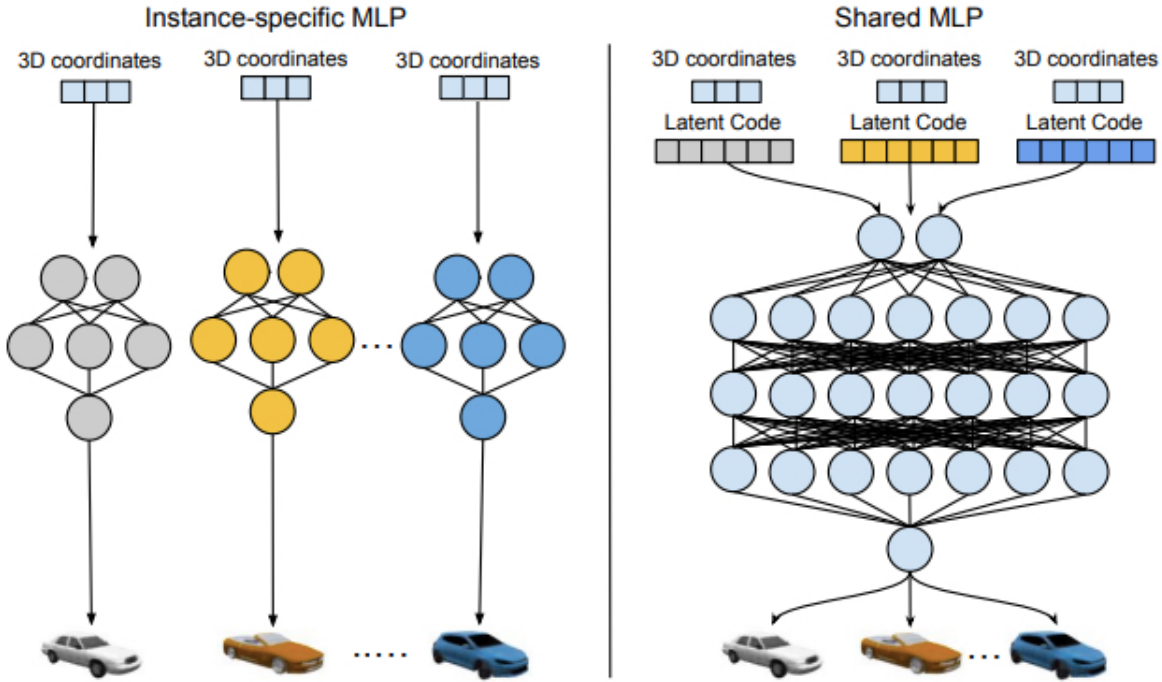
Dynamic scenes optimization -- Neural Scene Graphs

- Formulate dynamic, multi-object scenes into a learned scene graph
- Model objects using a shared volumetric representation



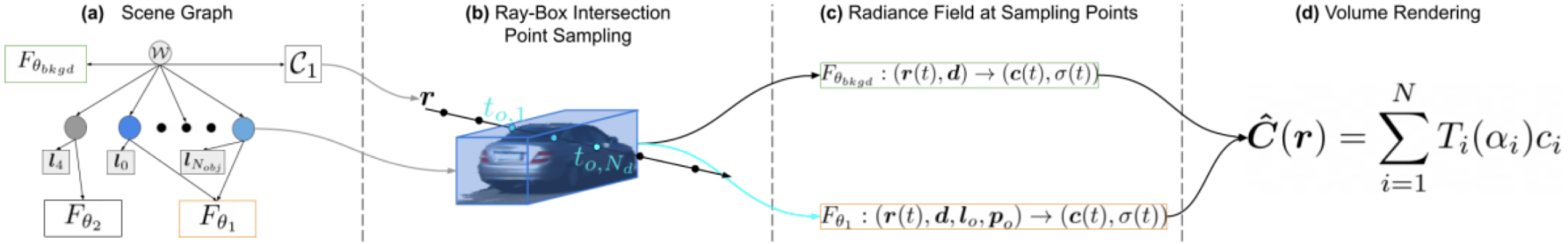
Ost J, Mannan F, Thuerey N, et al. Neural scene graphs for dynamic scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2856-2865.

Dynamic objects optimization – Panoptic Neural Fields vs Neural Scene Graph



Kundu A, Genova K, Yin X, et al. Panoptic Neural Fields: A Semantic Object-Aware Neural Scene Representation[J]. arXiv preprint arXiv:2205.04334, 2022.

Dynamic scenes optimization -- Neural scene graphs



(a) Reference

(b) Learned Object Nodes

(c) Learned Background

(d) View Reconstruction

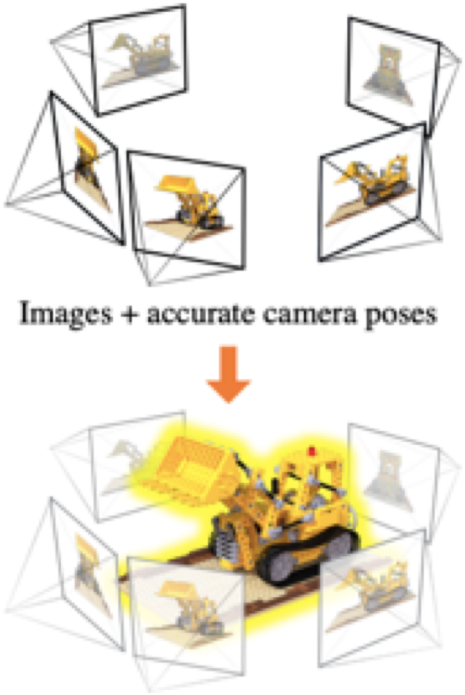
(e) Novel Scene

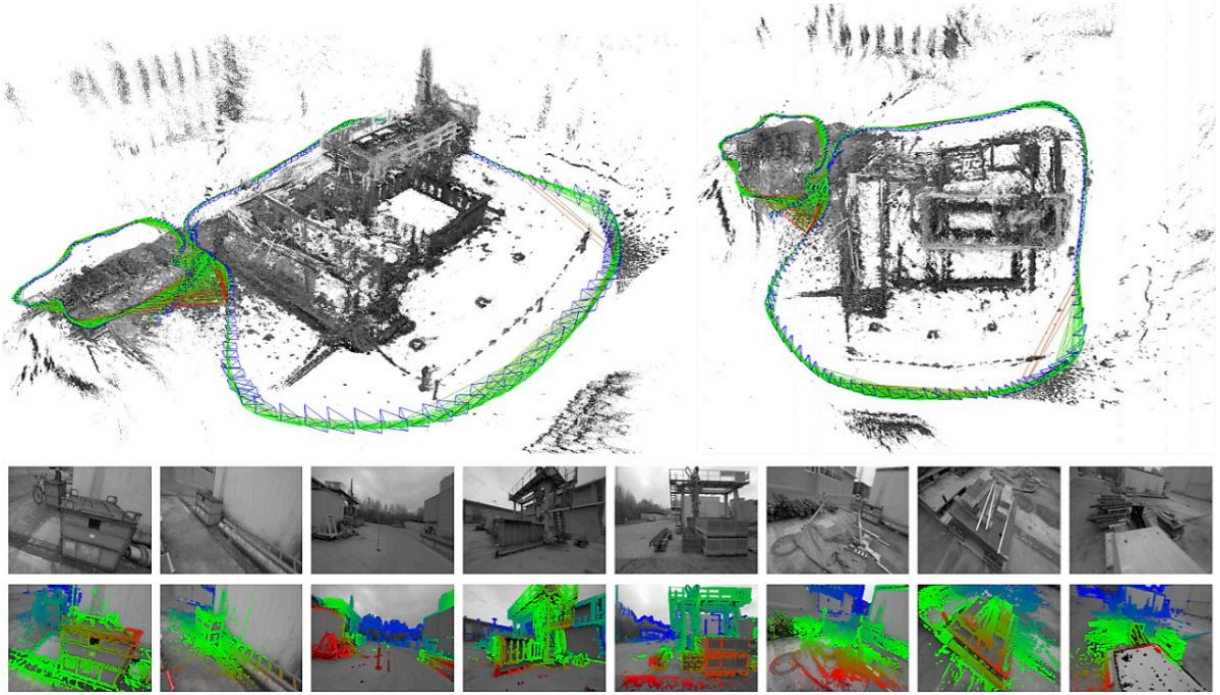
(f) Densely Populated Novel Scene

Ost J, Mannan F, Thuerey N, et al. Neural scene graphs for dynamic scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 2856-2865.

Specify reconstruction volume

- Should scale to large scenes

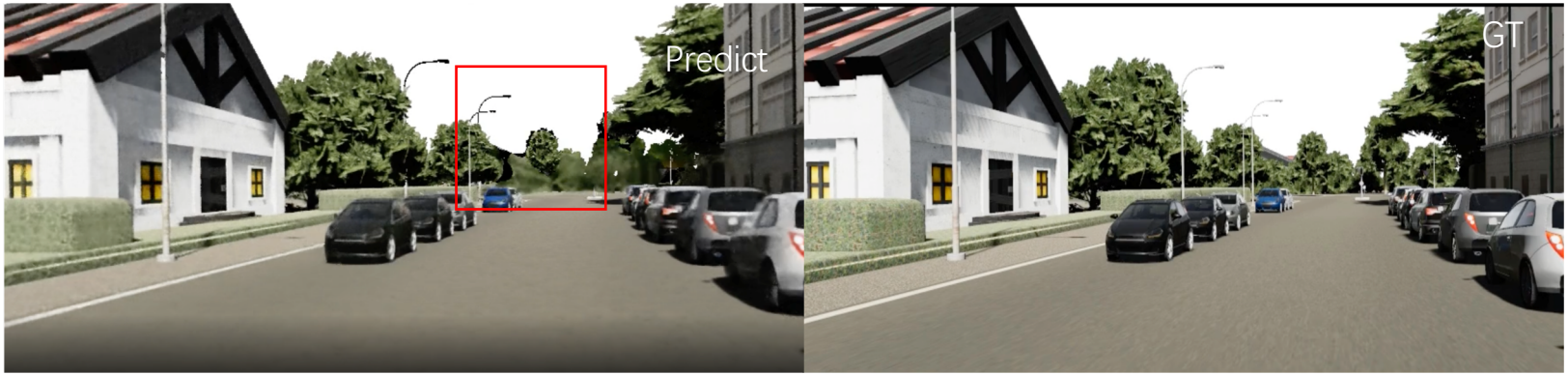




Specify reconstruction volume

- Should specify Bounding box range

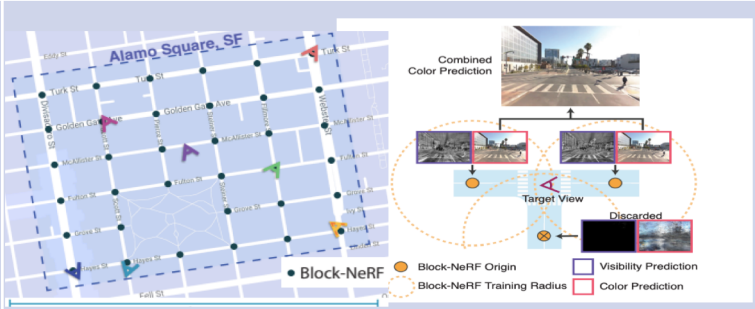
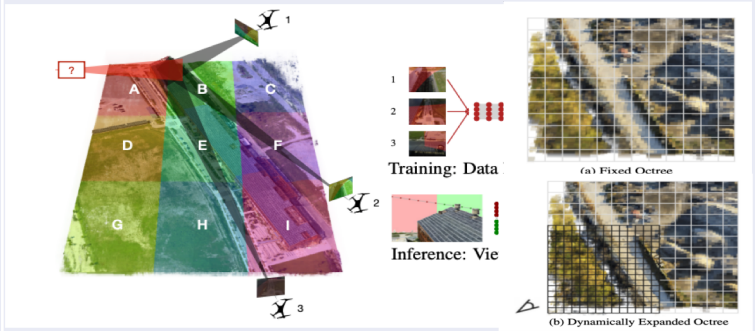
tartanair



vkitti

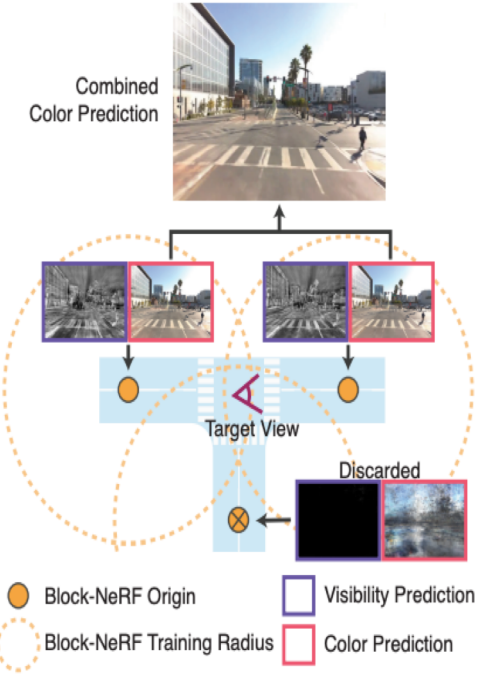
保密信息

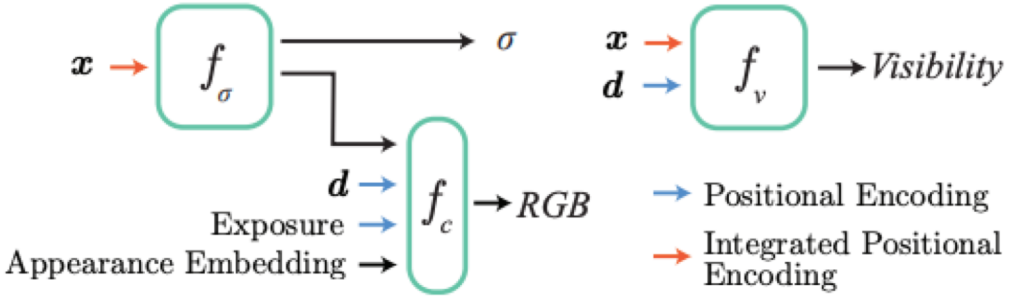
Large scene optimization

methods	dataset	Key idea	
Block nerf	Street view	**Scene => blocks with overlap** Use multi-nerf inference and combine	
Mega nerf	UAV dataset	**Scene => pixel clustering** Datastruct: Mega-NeRF-Plenoctree Mega-NeRF-KiloNeRF	

保密信息

Scene Reconstruction -- Block NeRF





Tancik M, Casser V, Yan X, et al. Block-nerf: Scalable large scene neural view synthesis[J]. arXiv preprint arXiv:2202.05263, 2022

Conclusion

NeRF for Scene Reconstruction

- Need accurate 6-DoF Camera Pose => BARF
- Normalize lighting => Appearance embeddings
- Handle dynamic objects => Ignore or ray transformation
- Need scale to large scenes => Split into blocks

保密信息

MEGVII 旷视

/ 04

Exercise

保密信息

Exercise

- 注意：作业答辩从 6月2日 改成 6月9日
- 实现大场景分块和融合的策略
  - 框架不限，可以基于release的代码基（基于torch-NGP）实现，
  - 提交代码压缩包和可视化效果（图片、视频）
  - 提交到 kangjia@megvii.com

保密信息

THANK YOU

保密信息