中国电信人工智能研究院
Institute of AI, China Telecom

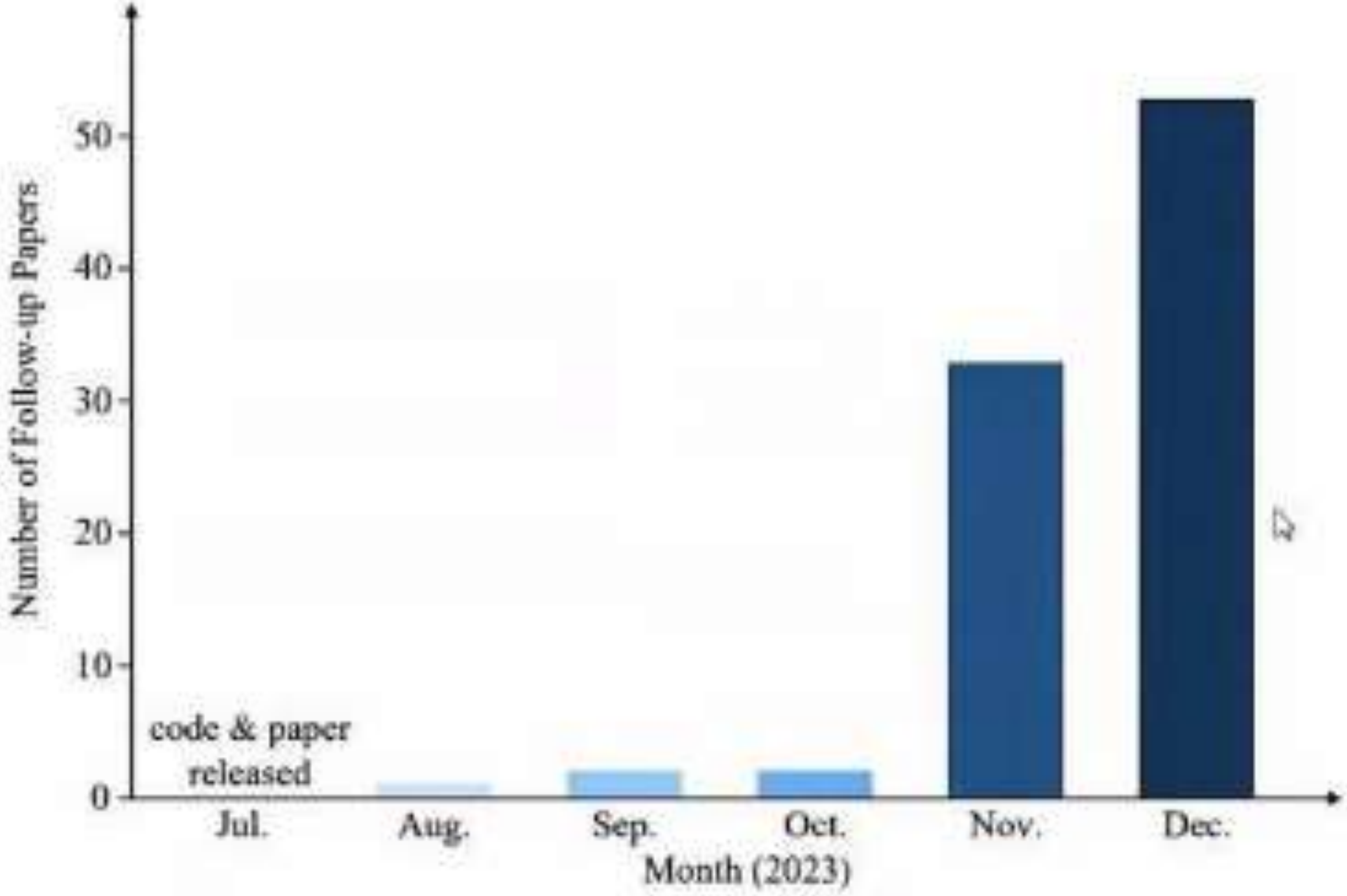
3D Gaussian Splatting

大规模&自动驾驶场景重建

汇报人：杨泽鹏

2024/10/18

Gaussian



Number of Follow-up Papers
0
10
20
30
40
50
code & paper
released
Jul.
Aug.
Sep.
Oct.
Nov.
Dec.
Month (2023)

本期主题 Topic of this video

大规模&自动驾驶场景重建

Reconstruction of Large Scale Scenes and Autonomous Driving Scenes

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction

2402.17427

Street Gaussians for Modeling Dynamic Urban Scenes

2401.01339

DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

2312.07920

Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-time Rendering

2311.18561

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction

Code (Soon) | arXiv 2402.17427

Tsinghua University | Huawei Noah's Ark Lab | Chinese Academy of Sciences

✓ Dataset: UrbanScene3D

Large Scene

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction

Code (Soon) | arXiv 2402.17427

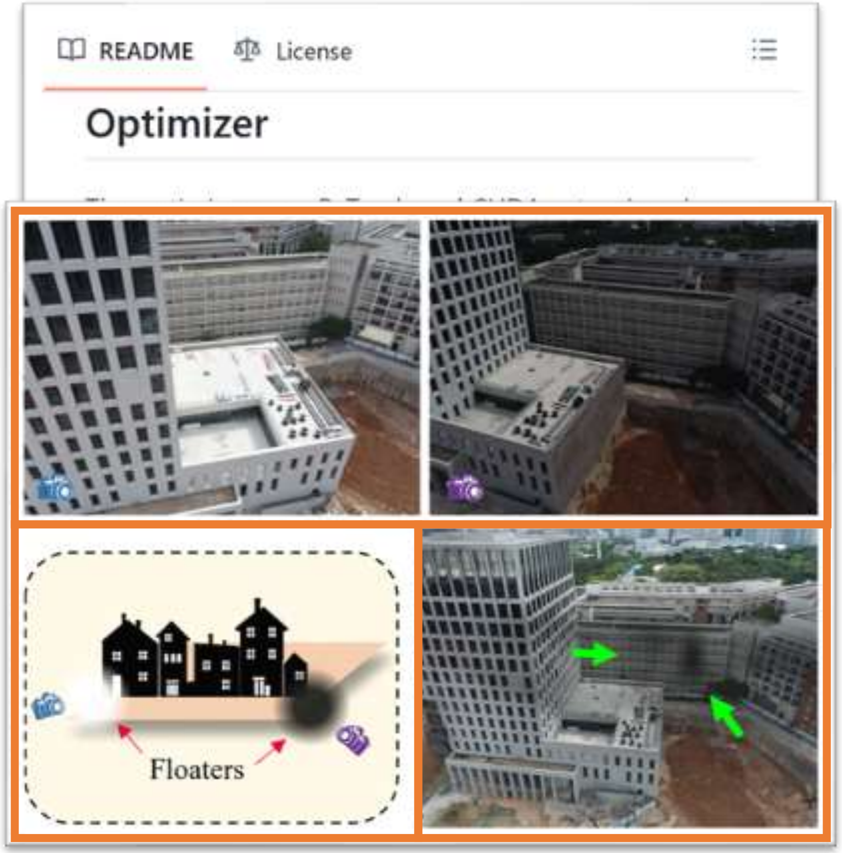
Tsinghua University | Huawei Noah's Ark Lab | Chinese Academy of Sciences

将 3D GS 扩展到大场景的难点：

1. Memory Limitation：受限于内存，3D GS在大规模场景下只能重建出低质量模型；

2. Long Optimization Time：需要把整个大场景视为整体进行充分迭代优化，很耗时，而且没有好的约束的情况下不稳定；

3. Appearance Variations：大场景光照不均匀。



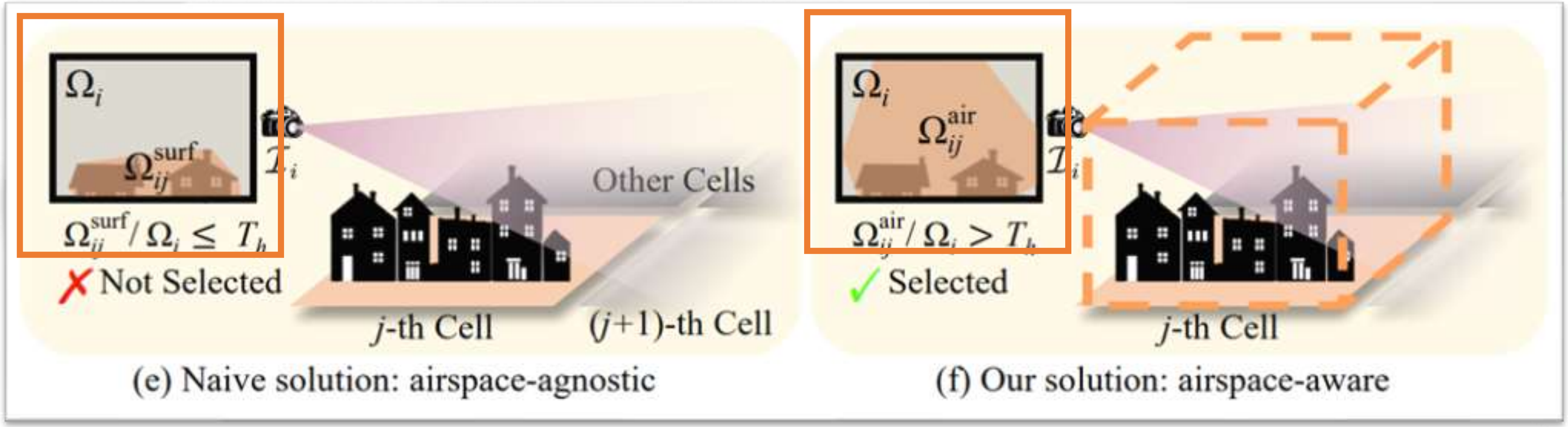
Large Scene

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction

Code (Soon) | arXiv 2402.17427

Tsinghua University | Huawei Noah's Ark Lab | Chinese Academy of Sciences

1 Progressive Data Partitioning



Large Scene

VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction

Code (Soon) | arXiv 2402.17427

Tsinghua University | Huawei Noah's Ark Lab | Chinese Academy of Sciences

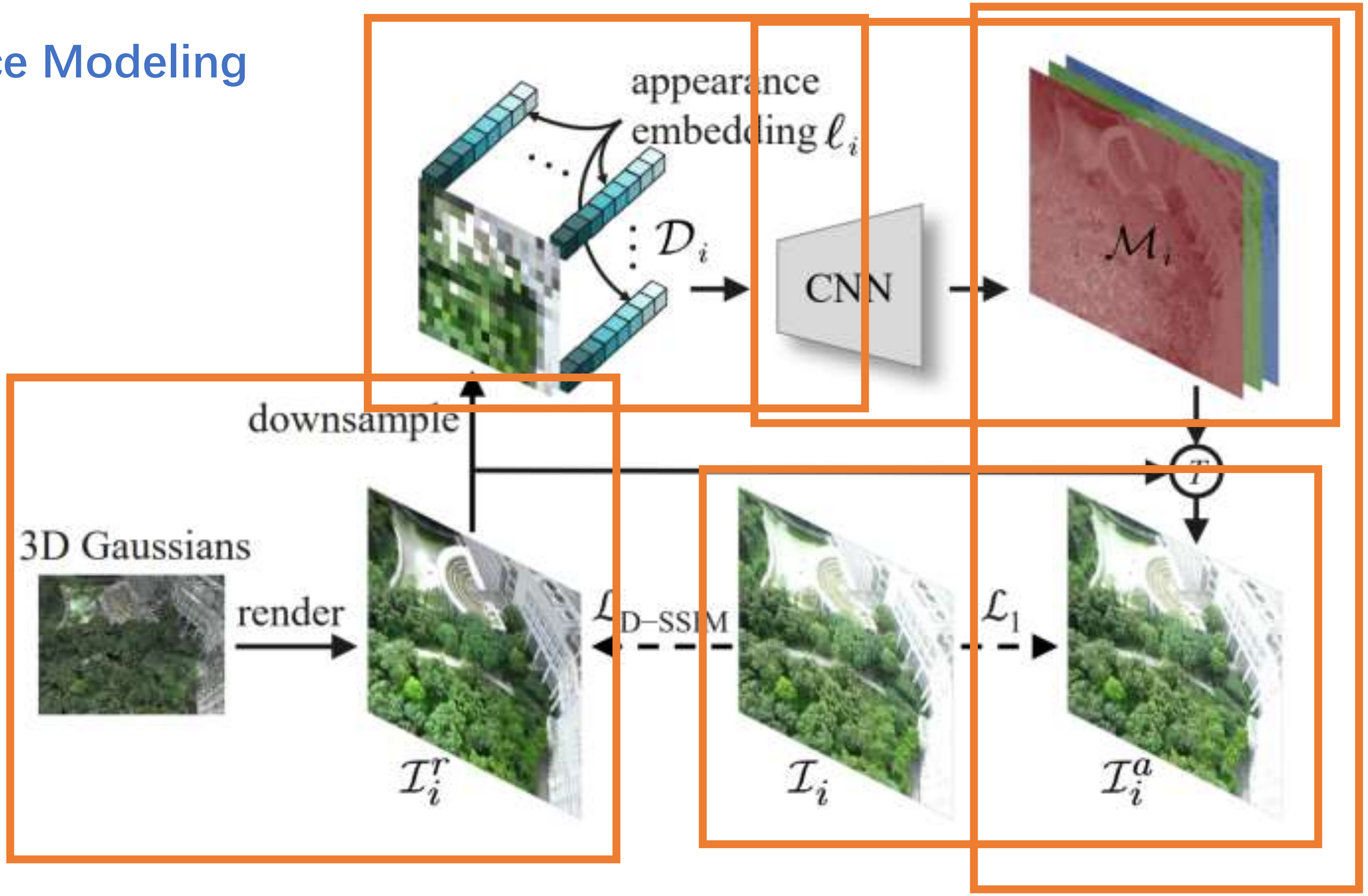
2 Decoupled Appearance Modeling

a) 下采样

b) 像素绑定 embedding

c) 特征图上采样

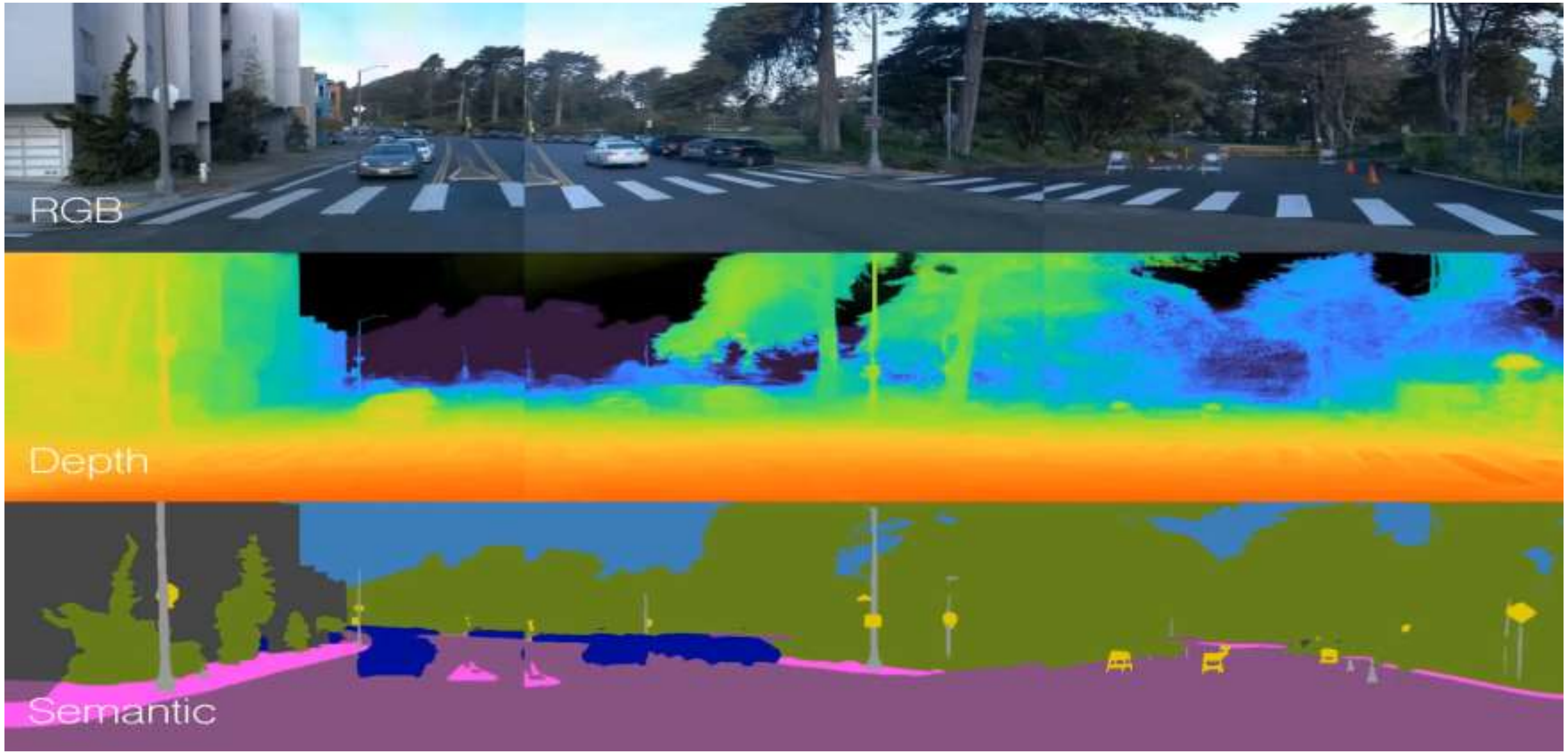
d) 调整光照



Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-time Rendering

Code (Soon) | arXiv 2311.18561

Fudan University | University of Surrey



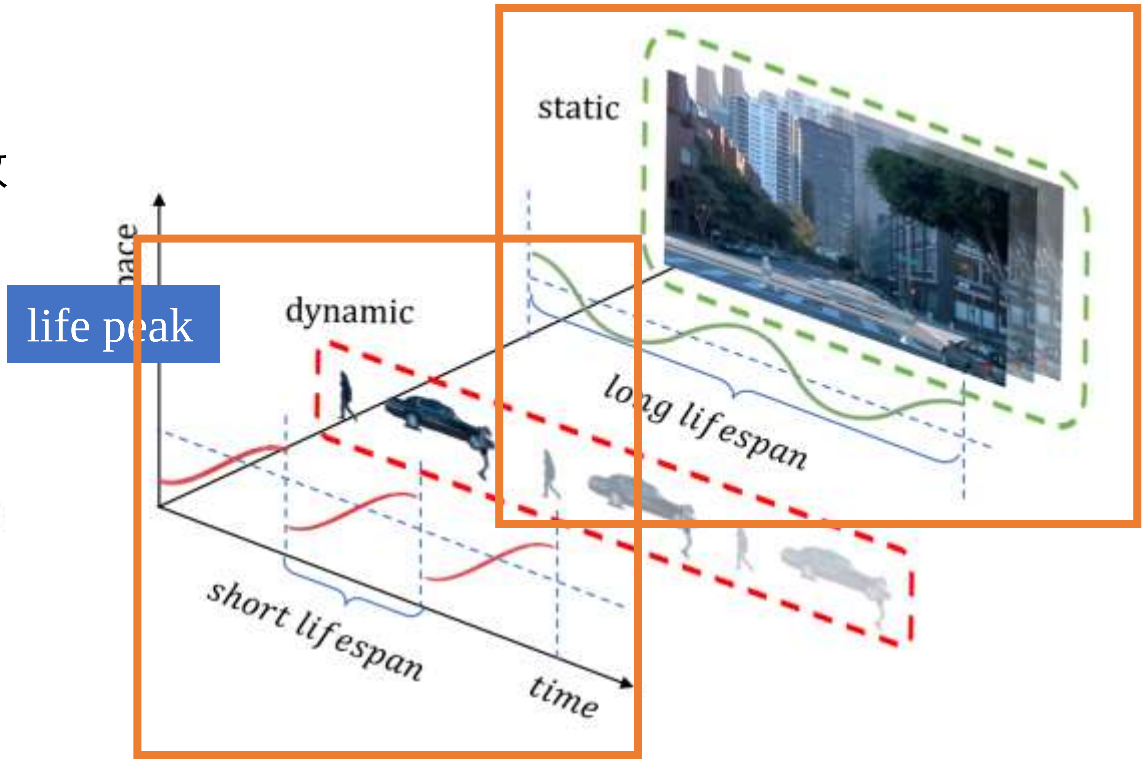
Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-time Rendering

Code (Soon) | arXiv 2311.18561 | Fudan University | University of Surrey

1 Periodic Vibration

- 把均值和不透明度改成依赖于时间的函数

$$
\begin{cases}
\mathcal{H}(t) = \{\widetilde{\boldsymbol{\mu}}(t), \boldsymbol{q}, \boldsymbol{s}, \widetilde{o}(t), \boldsymbol{c}\}, \\
\widetilde{\boldsymbol{\mu}}(t) = \boldsymbol{\mu} + \mathbf{A} \cdot \sin\left(\frac{2\pi(t-\tau)}{l}\right), \\
\widetilde{o}(t) = o \cdot e^{-\frac{1}{2}(t-\tau)^2\beta^{-2}},
\end{cases}
$$



Periodic Vibration Gaussian: Dynamic Urban Scene Reconstruction and Real-time Rendering

Code (Soon) arXiv 2311.18561

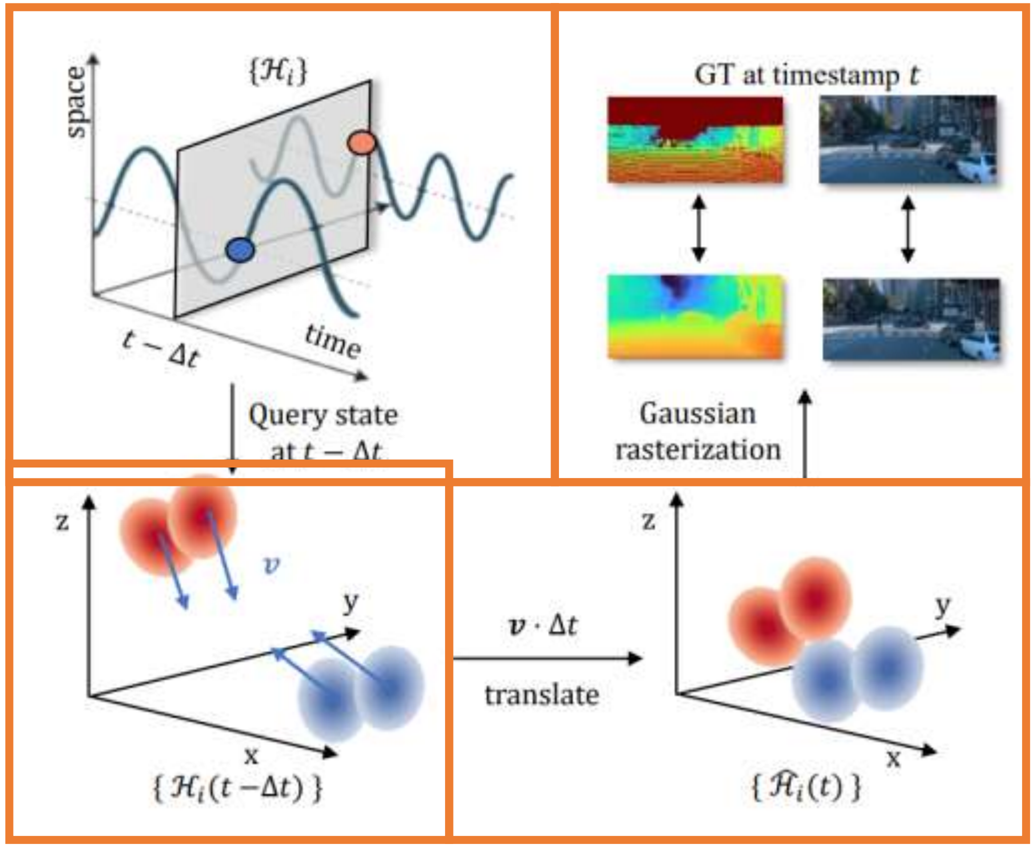
Fudan University | University of Surrey

2 Temporal smoothing by scene flow

- 定义 scene velocity， 表达 PVG 连续状态之间的线性关系： $\boxed{\rho = \frac{\beta}{l}}$

$$\boldsymbol{v} = \frac{\mathrm{d}\widetilde{\boldsymbol{\mu}}(t)}{\mathrm{d}t}\Big|_{t=\tau} \cdot \exp(-\frac{\rho^2}{2}) = \frac{2\pi\mathbf{A}}{l} \cdot \exp(-\frac{\rho^2}{2}).$$

因此 PVG 点的两个相邻时刻的状态通过 scene flow translation 线性连接在一起。



DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

Code (Soon) | arXiv 2312.07920

Peking University | Google Research | University of California, Mercedz

EmerNeRF VS Ours

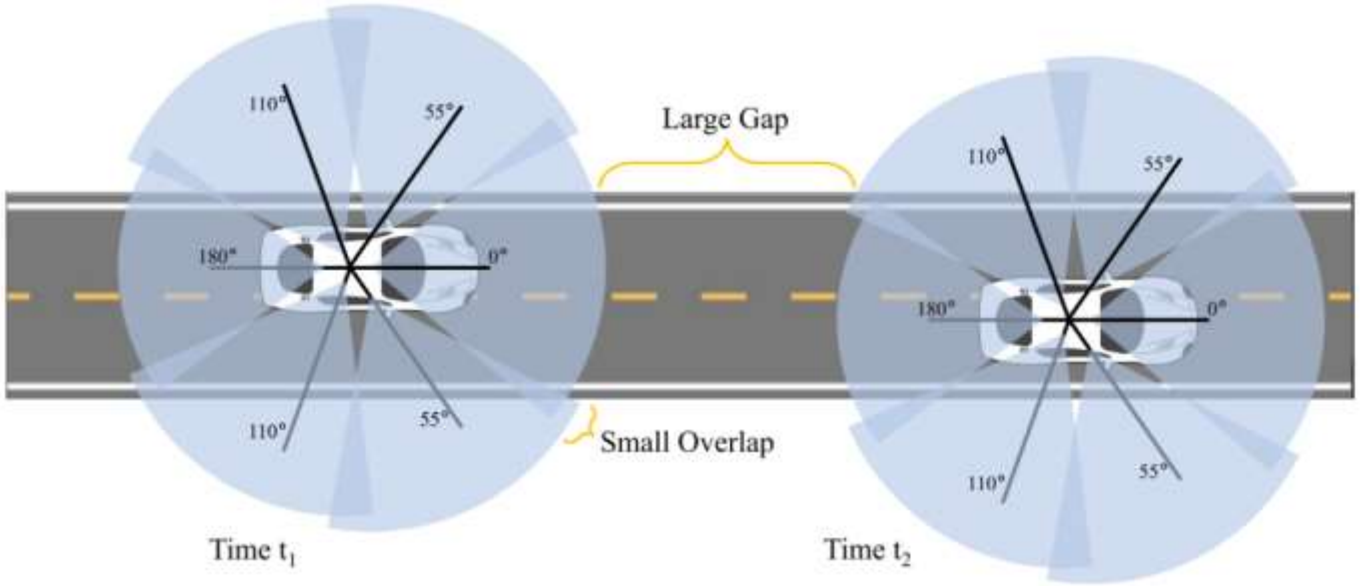
DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

Code (Soon) | arXiv 2312.07920

Peking University | Google Research | University of California, Mercedz

360°大规模自动驾驶场景建模面对的挑战:

1. 自身车辆和动态物体都在以相对高的速度运动，且视角受限，只能捕捉到快速变化;
2. 多相机采集时，朝外的视角重叠较少;
3. 不同方向的光照变化，复杂的几何，时空不连续性……



Autonomous Driving

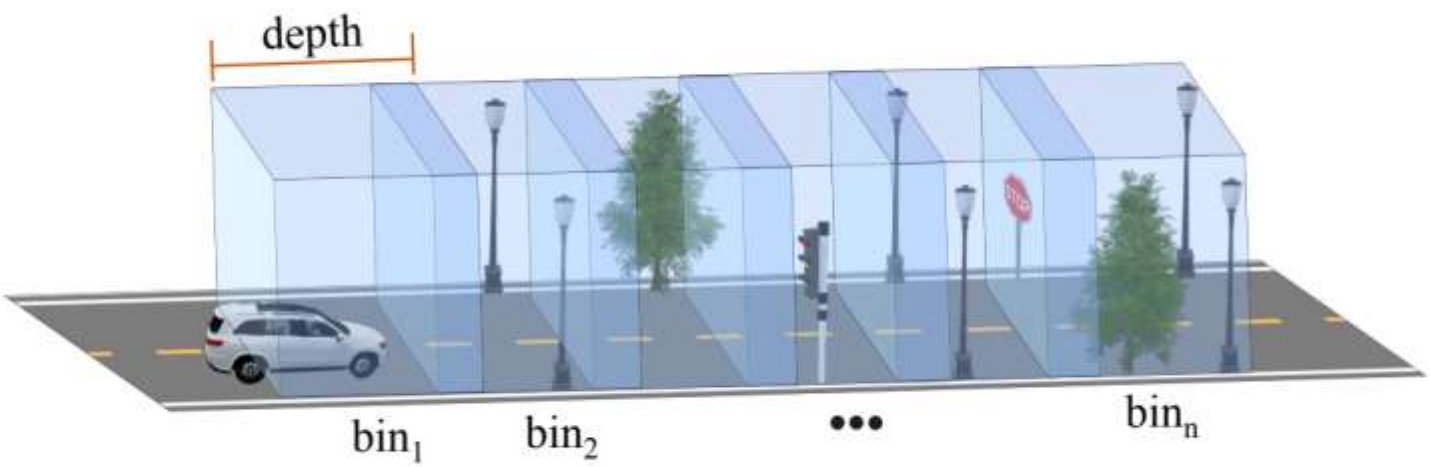
DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

Code (Soon) | arXiv 2312.07920

Peking University | Google Research | University of California, Mercedz

1 Incremental Static 3D Gaussians

- 根据雷达先验的深度范围，将静态场景划分为 N 个 bins，
- 用来自前一个 bin 的高斯点作为位置先验，并基于重叠区域对齐相邻 bins。



DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

Code (Soon) | arXiv 2312.07920 | Peking University | Google Research | University of California, Mercedz

2 Composite Dynamic Gaussian Graph

- 使用数据集提供的 bounding box 从静态背景分解动态前景物体，
- 建立动态高斯图，每个动态物体分配一个ID，以及时间戳对应的外观；

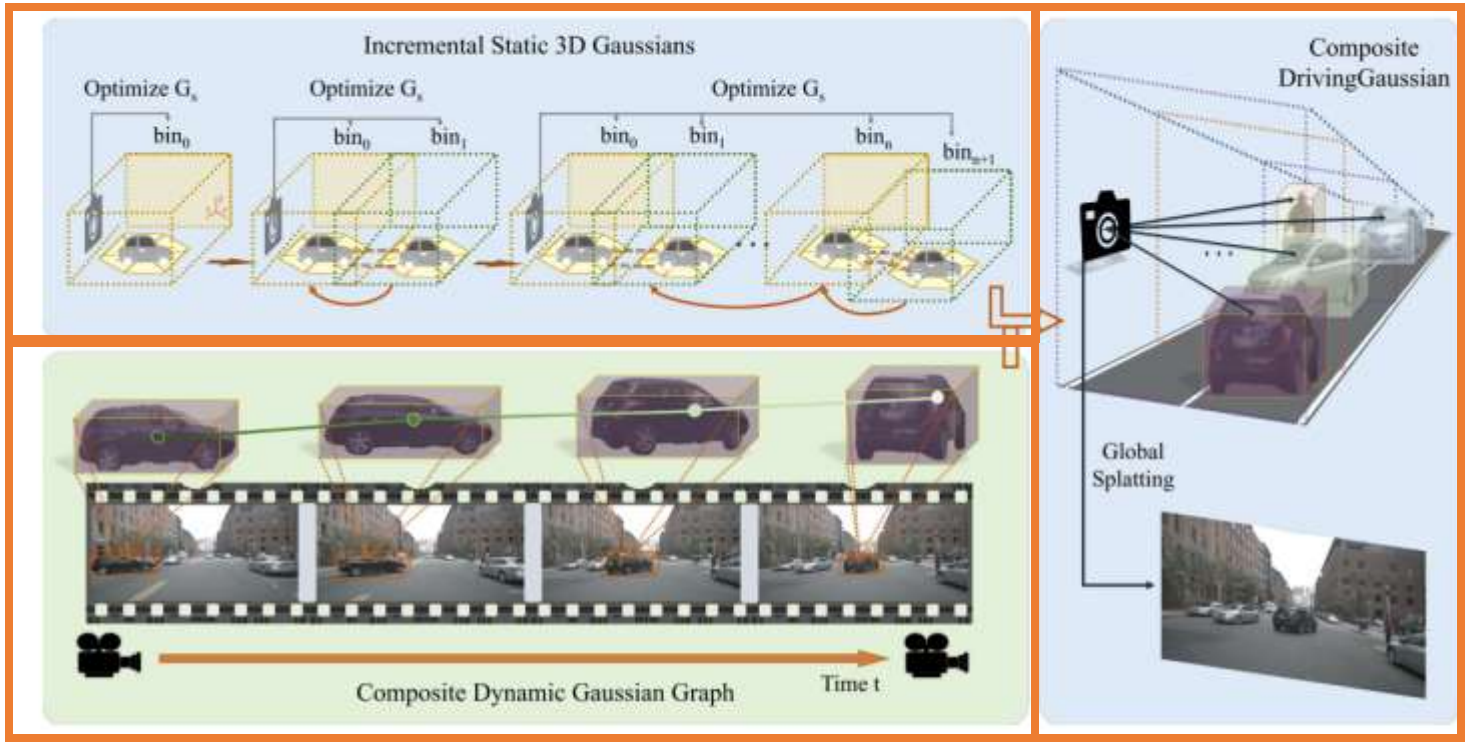
$$H =< O, G_d, M, P, A, T >,$$

O： instance object
G： 动态高斯点
M： 物体的变换矩阵
P： bbox 的坐标中心
A： t 时刻的 bbox 的方向
T： 时间

DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes

Code (Soon) | arXiv 2312.07920 | Peking University | Google Research | University of California, Mercedz

2 Global Rendering via Gaussian Splatting



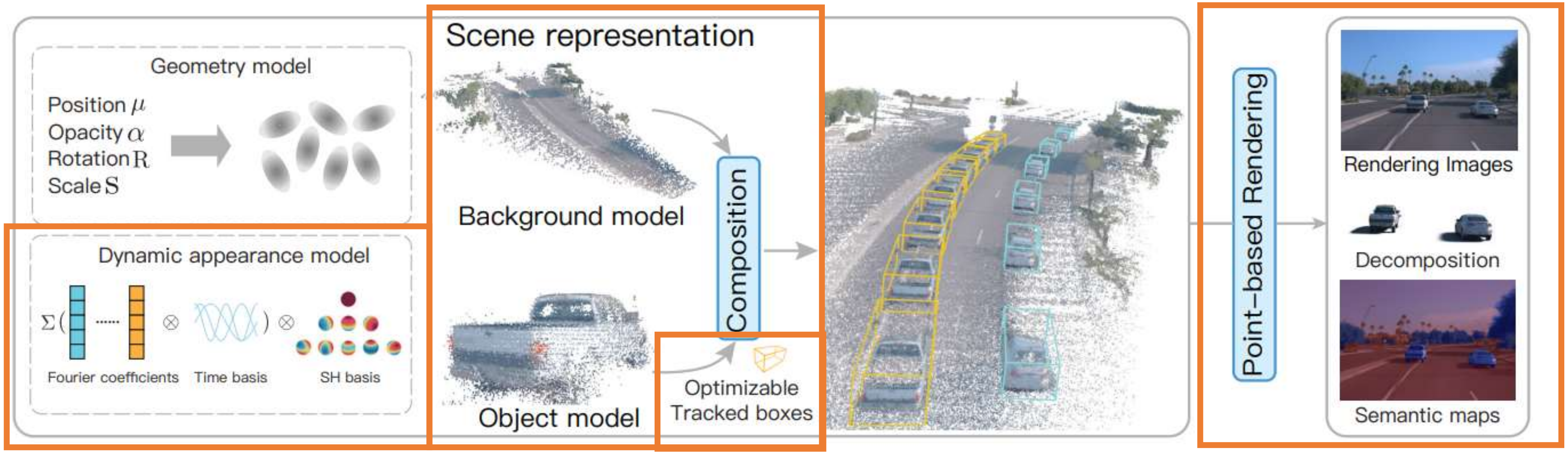
Street Gaussians for Modeling Dynamic Urban Scenes

Code (Soon) | arXiv 2401.01339 | Zhejiang University | Li Auto

Street Gaussians for Modeling Dynamic Urban Scenes

Code (Soon) | arXiv 2401.01339 | Zhejiang University | Li Auto

Overview



Autonomous Driving

Street Gaussians for Modeling Dynamic Urban Scenes

Code (Soon) | arXiv 2401.01339 | Zhejiang University | Li Auto

1 Object model

- 使用4D球谐模型建模，每个 t 时刻球谐参数从离散的傅里叶逆变换获得，这样就能够把时间信息编码到外观中。

$$z_{m,l} = \sum_{i=0}^{k-1} \boldsymbol{f}_i \cos\left(\frac{i\pi}{N_t} t\right).$$

傅里叶变换与球谐函数展开有何相似之处？ - 知乎

两者都是分解到空间中的一组正交基，可以看作从实空间到频率空间之间的变换，它们有没有更进一步的联系？...

知乎 https://www.zhihu.com/question/52859906

Autonomous Driving

Street Gaussians for Modeling Dynamic Urban Scenes

Code (Soon) arXiv 2401.01339 Zhejiang University | Li Auto

2 Tracking pose optimazion

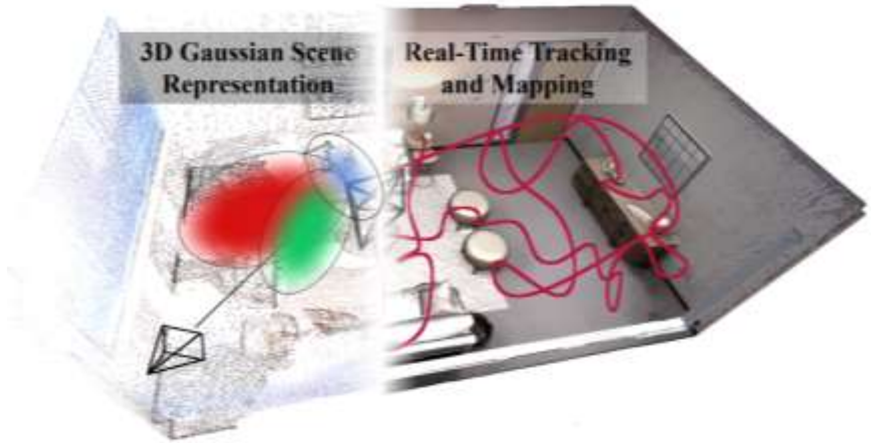
- 渲染使用的物体旋转被表示为原始旋转乘以修正量，平移表示为原始位置加上偏移量，这样一来在更新的时候这部分梯度不需要额外的计算量。

$$
\begin{aligned}
\mathbf{R}_t' &= \mathbf{R}_t \Delta \mathbf{R}_t, \\
\mathbf{T}_t' &= \mathbf{T}_t + \Delta \mathbf{T}_t,
\end{aligned}
$$

其他工作 Other interesting works

无位姿先验&SLAM

Reconstruction with No Pose Prior,
Simultaneous Localization and Mapping



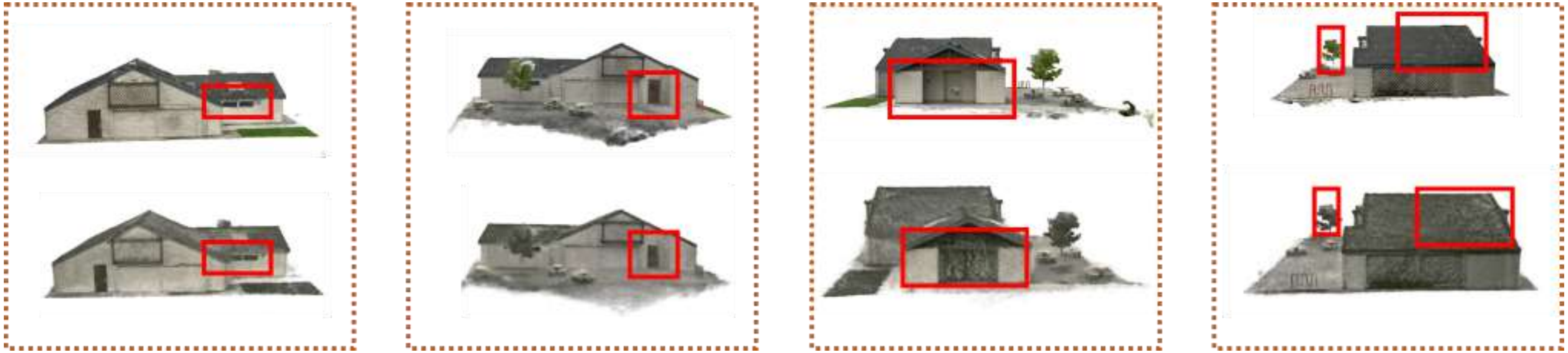
- SGS-SLAM: Semantic Gaussian Splatting For Neural Dense SLAM
- GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting
- SplaTAM: Splat, Track & Map 3D Gaussians for Dense RGB-D SLAM
- Gaussian Splatting SLAM
- Gaussian-SLAM: Photo-realistic Dense SLAM with Gaussian Splatting
- Photo-SLAM: Real-time Simultaneous Localization and Photorealistic Mapping for Monocular, Stereo, and RGB-D Cameras
- COLMAP-Free 3D Gaussian Splatting

上周总结 Last week review

科研进展

大尺度高保真场景重建论文及实验

多视角融合大尺度场景融合调研

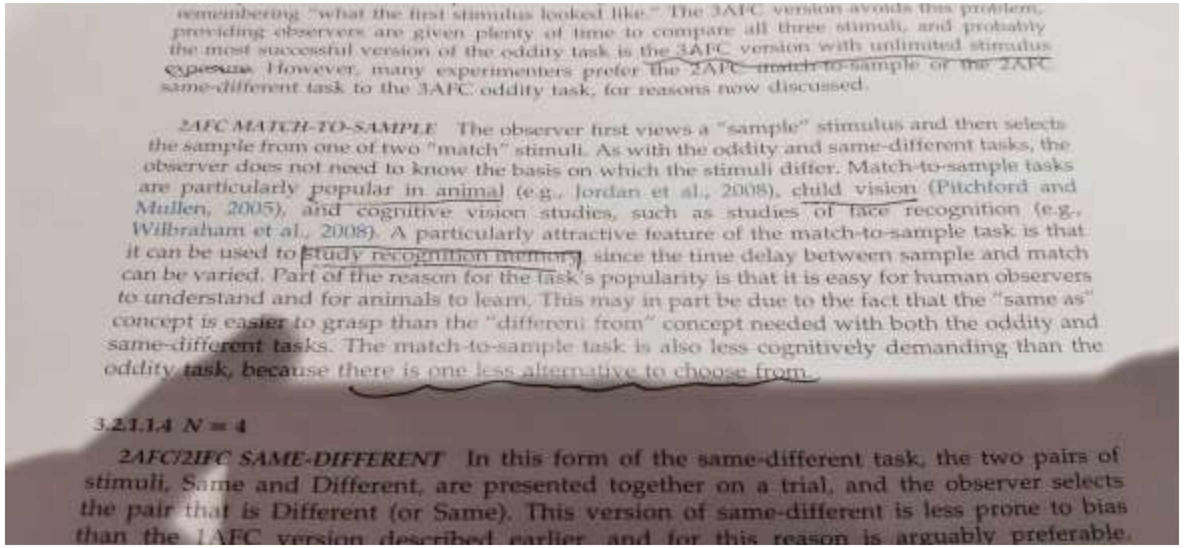


上周总结 Last week review

工程进展

大尺度高保真场景重建无人机部署（ROS+Cuda）

移动目标的OCR识别


remembering "what the first stimulus looked like." The 3AFC version avoids this problem, providing observers are given plenty of time to compare all three stimuli, and probably the most successful version of the oddity task is the 3AFC version with unlimited stimulus exposure. However, many experimenters prefer the 2AFC match-to-sample or the 2AFC same-different task to the 3AFC oddity task, for reasons now discussed.

2AFC MATCH-TO-SAMPLE The observer first views a "sample" stimulus and then selects the sample from one of two "match" stimuli. As with the oddity and same-different tasks, the observer does not need to know the basis on which the stimuli differ. Match-to-sample tasks are particularly popular in animal (e.g., Jordan et al., 2008), child vision (Pitchford and Mullen, 2005), and cognitive vision studies, such as studies of face recognition (e.g., Wilbraham et al., 2008). A particularly attractive feature of the match-to-sample task is that it can be used to study recognition memory, since the time delay between sample and match can be varied. Part of the reason for the task's popularity is that it is easy for human observers to understand and for animals to learn. This may in part be due to the fact that the "same as" concept is easier to grasp than the "different from" concept needed with both the oddity and same-different tasks. The match-to-sample task is also less cognitively demanding than the oddity task, because there is one less alternative to choose from.

3.2.1.1.4 N = 4

2AFC/2IFC SAME-DIFFERENT In this form of the same-different task, the two pairs of stimuli, Same and Different, are presented together on a trial, and the observer selects the pair that is Different (or Same). This version of same-different is less prone to bias than the 1AFC version described earlier, and for this reason is arguably preferable.

2AFC MATCH-TO-SAMPLE. The observer first views a "sample" stimulus and then selectsthe sample from one of two "match" stimuli. As with the oddity and same-different tasks, theobserver does

准确率：92%

Thanks

大规模&自动驾驶场景重建

汇报人：杨泽鹏

2024/10/18