



ICCV2025

Hierarchy UGP:

Hierarchy Unified Gaussian Primitive for Large-Scale Dynamic Scene Reconstruction

Hongyang Sun

22351179@zju.edu.cn
3445314103@qq.com



Table of Contents

- I. Overview of Street Scene Reconstruction**
- II. Street Scene Reconstruction Technologies**
- III. Introduction to Our Research Work**



Overview of Street Scene Reconstruction: Problem Description

Input: Vehicle-collected data (camera-captured videos, LiDAR data, etc.)

Output: Novel-view images (images from new viewpoints)



Overview of Street Scene Reconstruction: Application

Application 1: Rendering novel views can generate data for different vehicle models, which can be used to train end-to-end planner algorithms.

Challenge: Dynamic street scene reconstruction.

Application 2: The reconstructed 3D scenes can serve as realistic simulators for evaluating planner algorithms.

Challenge: Large-scale scene reconstruction.



Overview of Street Scene Reconstruction: Problem

Problem 1: Rendering novel views to generate clip data for different vehicle models — *Dynamic Street Scene Reconstruction*

It is necessary to achieve high-quality reconstruction of dynamic objects such as vehicles and pedestrians in order to generate high-quality clip data for autonomous driving planners.

Related Research: Neural Scene Graph, Street Gaussian, OmniRe, etc.

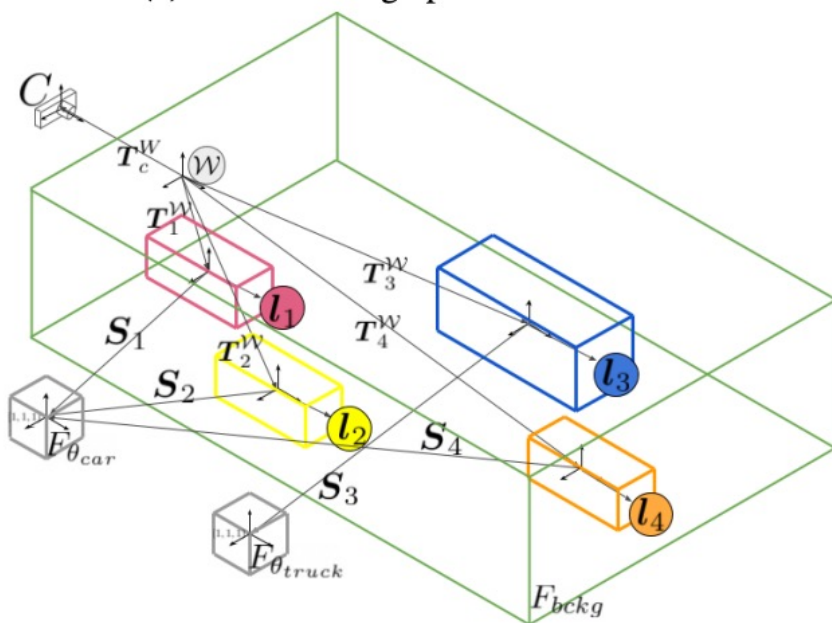
Problem 2: Reconstructing real-world scenes as realistic simulators — *Large-Scale Scene Reconstruction*

It is essential to realize high-quality large-scale scene reconstruction and real-time rendering in order to build a high-quality simulator for evaluating autonomous driving planner algorithms.

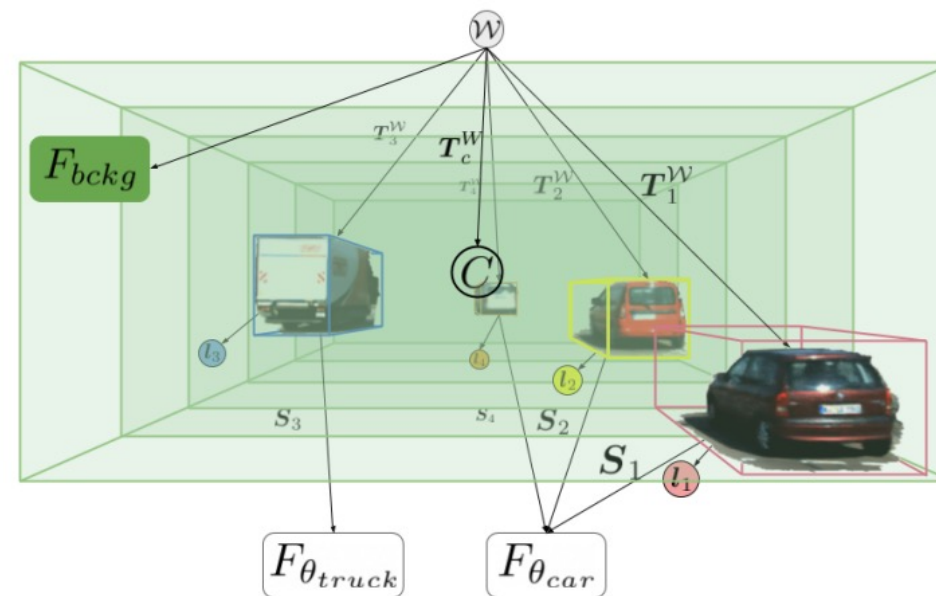
Related Research: Block NeRF, Vast Gaussian, Hierarchical Gaussian, etc.

The Neural Scene Graph is proposed to model dynamic street scenes.

(a) Neural scene graph in isometric view.

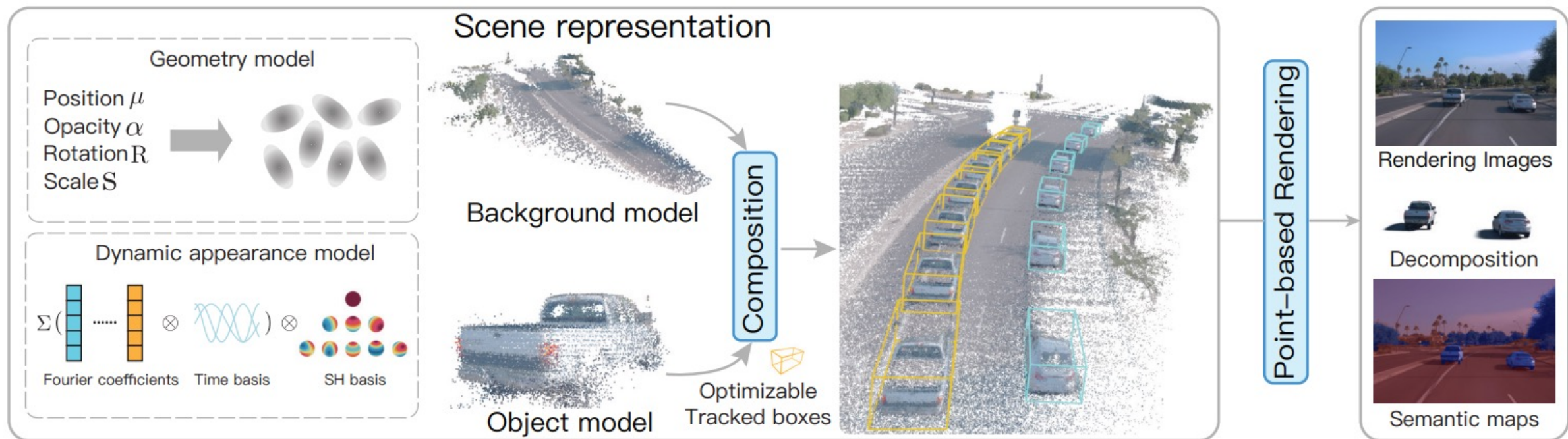


(b) Neural scene graph from the ego-vehicle view.



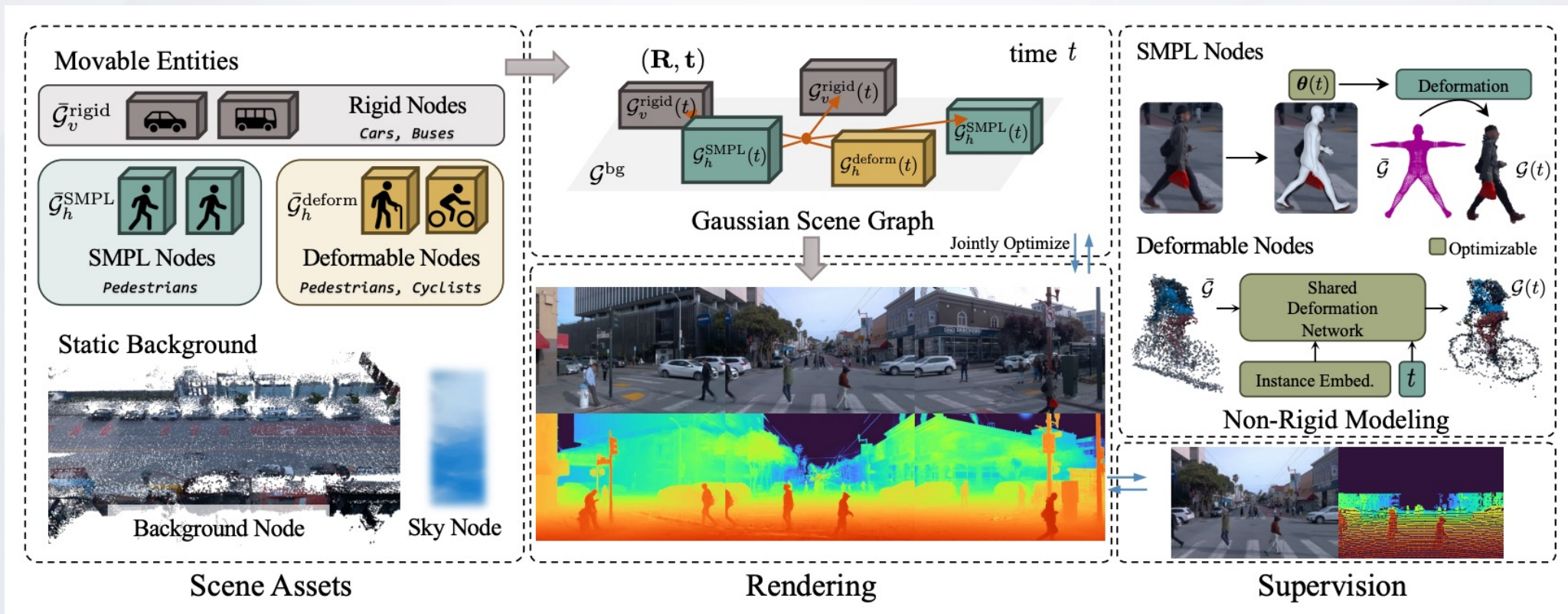
Neural scene graph, Julian et al, CVPR2021

The first to propose a Gaussian Scene Graph for modeling dynamic street scenes.



Street Scene Reconstruction Technology: Dynamic Street Scene Reconstruction

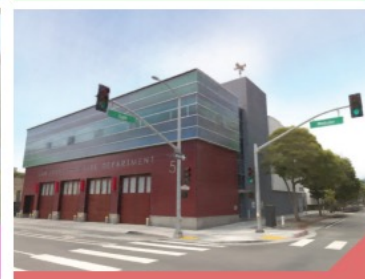
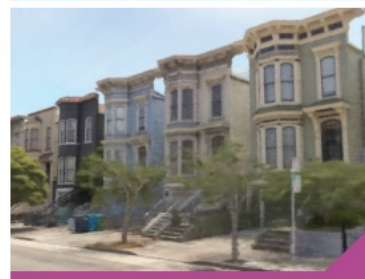
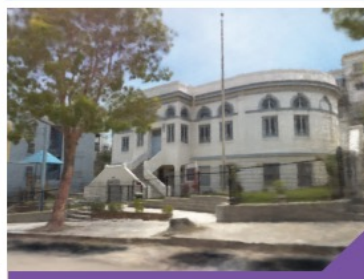
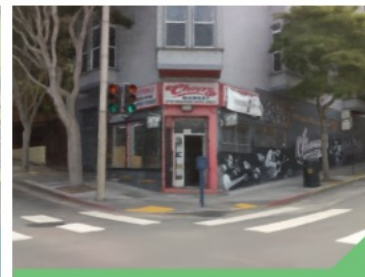
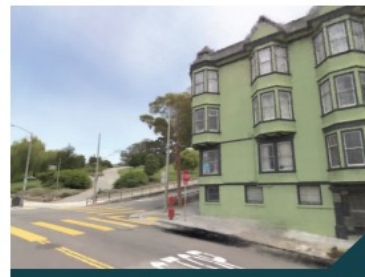
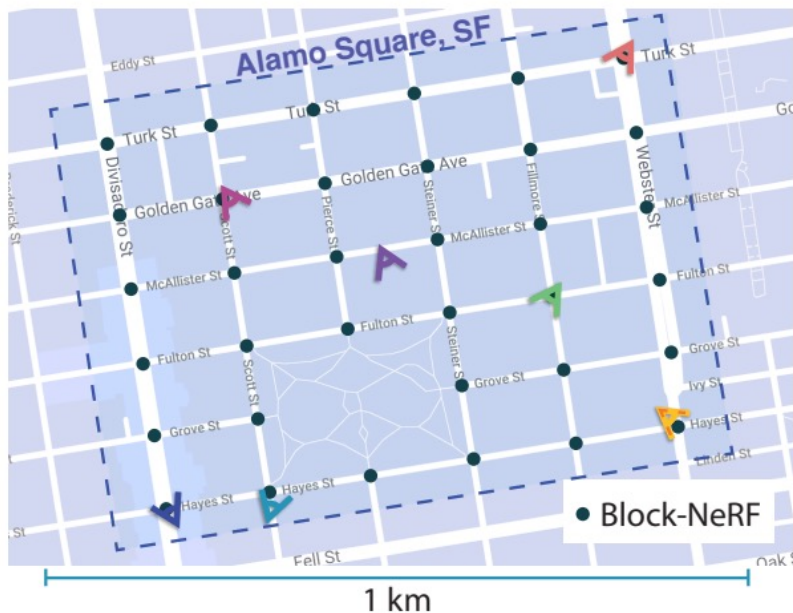
Building upon Street Gaussian, SMPL is utilized to achieve pedestrian reconstruction.



OmniRe, Chen et al, ICLR2025

Scene Reconstruction Technology: Large-Scale Scene Reconstruction

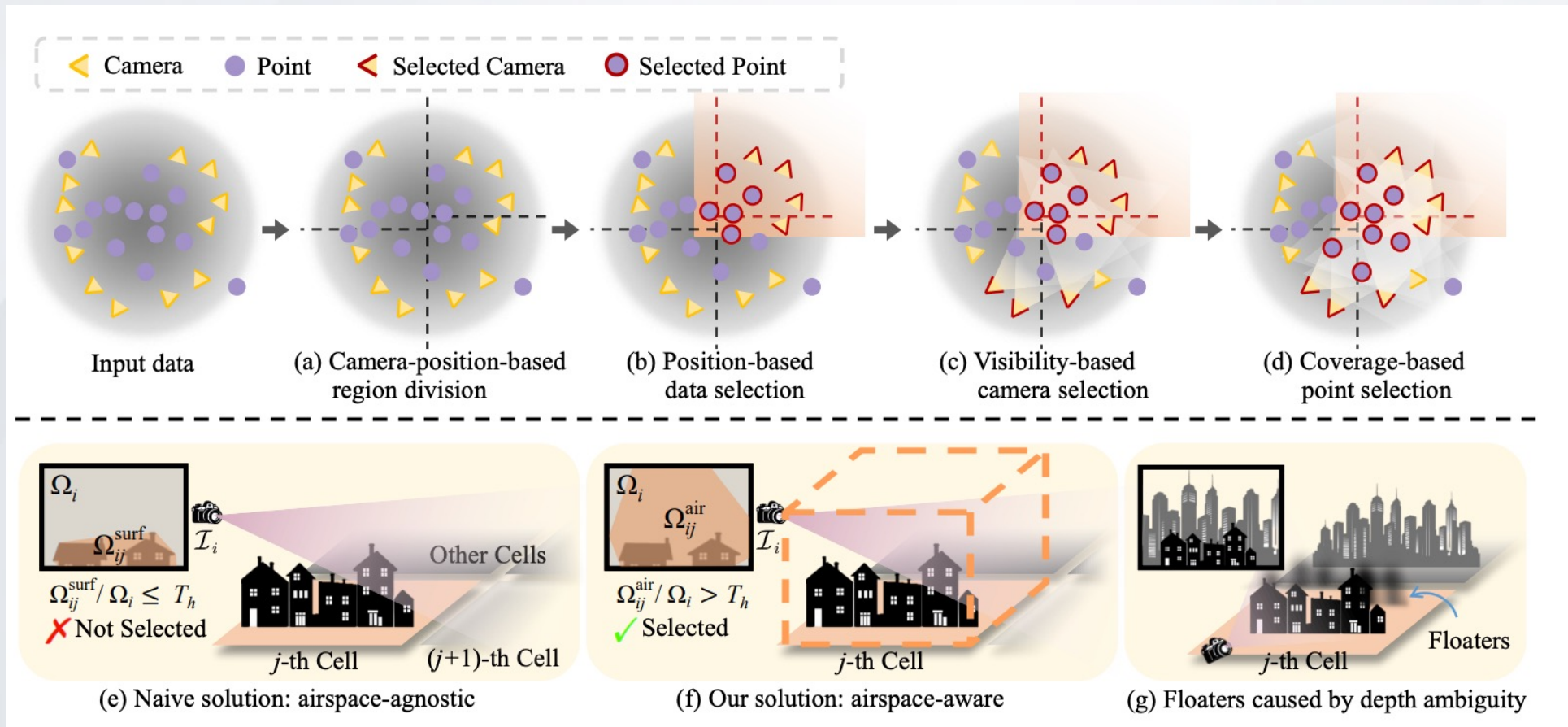
Divide the large-scale street scene into multiple blocks, with each block reconstructed using an independent NeRF.



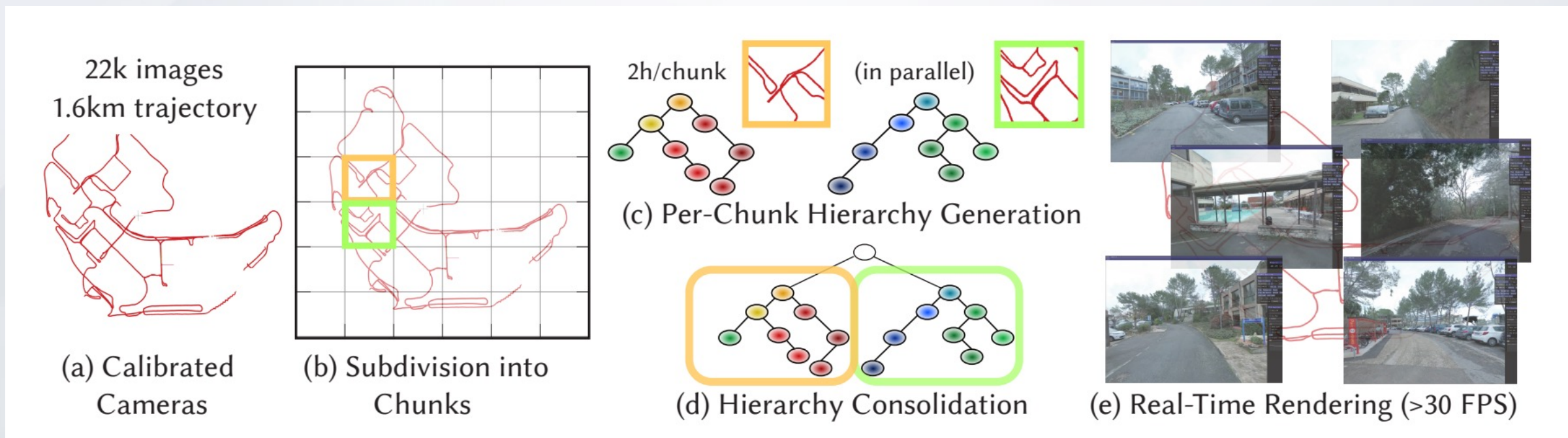
Block NeRF, Tancik et al, CVPR2022

Scene Reconstruction Technology: Large-Scale Scene Reconstruction

Divide the large-scale scene into multiple blocks, with each block reconstructed using 3D Gaussian



Divide the large-scale scene into multiple blocks, with each block reconstructed using 3D Gaussian



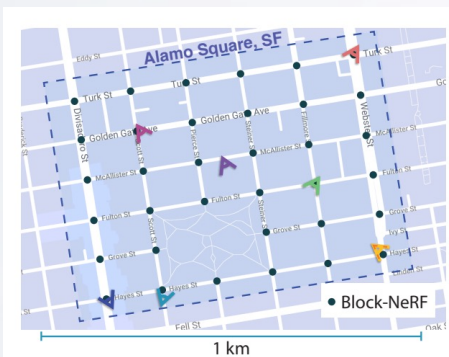
Hierarchical 3DGS, KERBL et al, Siggraph2024 TOG

Introduction to Our Research Work

Research Objective:

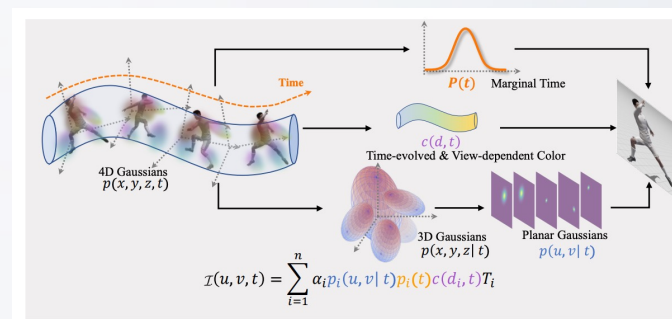
To address both dynamic street scene reconstruction and large-scale scene reconstruction within a unified framework.

Problem 1: *Large-Scale Scene Reconstruction*



Inspired by Block NeRF, we divide large-scale dynamic scenes into multiple smaller sub-scenes.

Problem 2: *Dynamic Object Reconstruction*

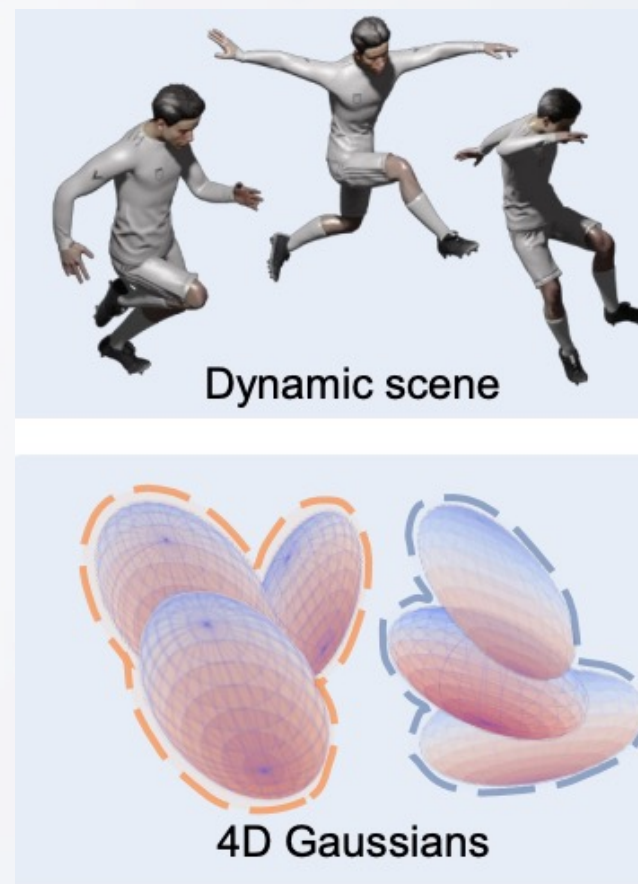
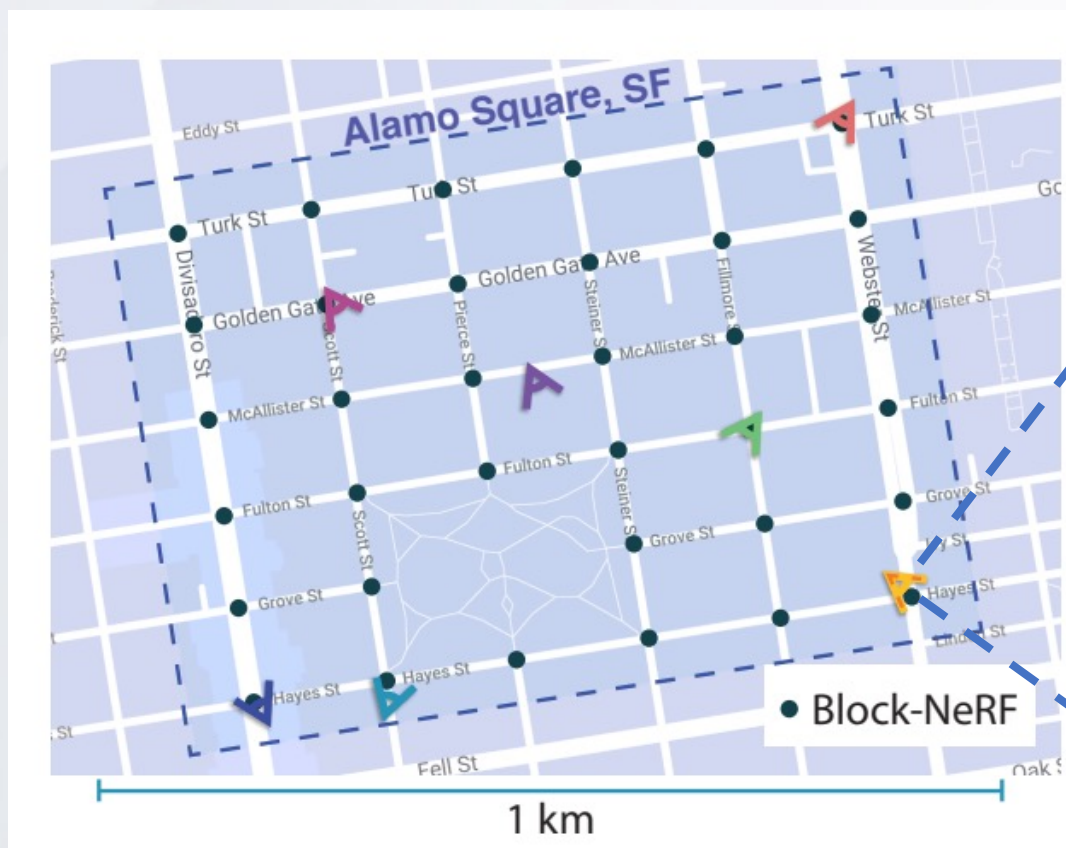


We use 4D Gaussian Splatting (4DGS) as the reconstruction primitive to achieve high-quality reconstruction of dynamic objects.

Solution Approach:

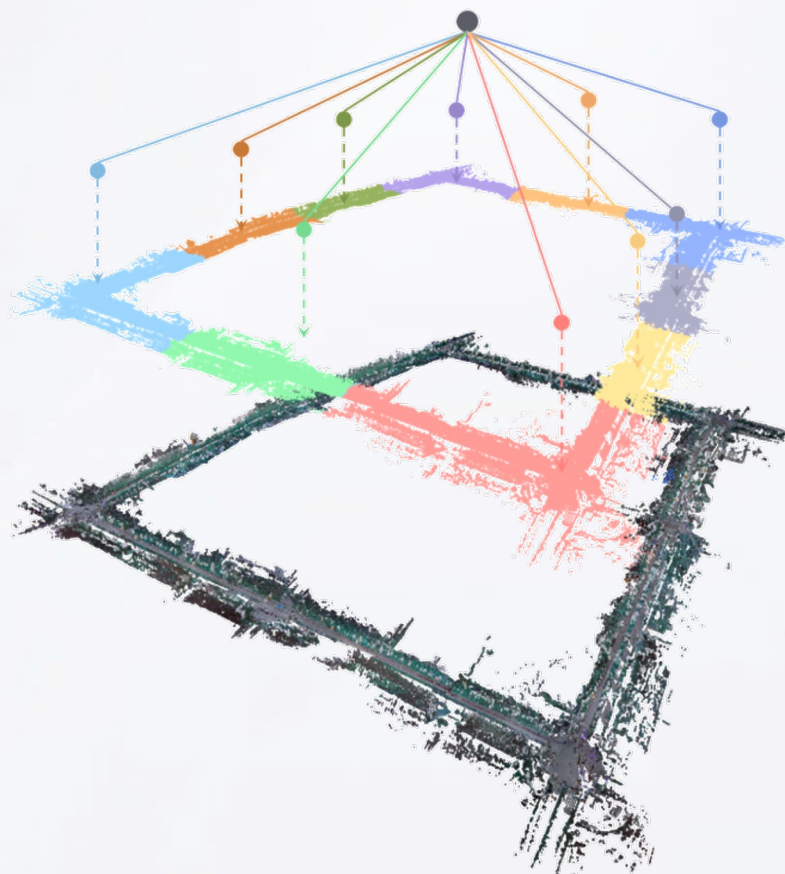
Model large-scale dynamic scenes using *Hierarchical Unified Gaussian Primitives*.

Divide the large-scale dynamic scene into multiple sub-scenes, and use UGP (4DGS) to model each sub-scene.

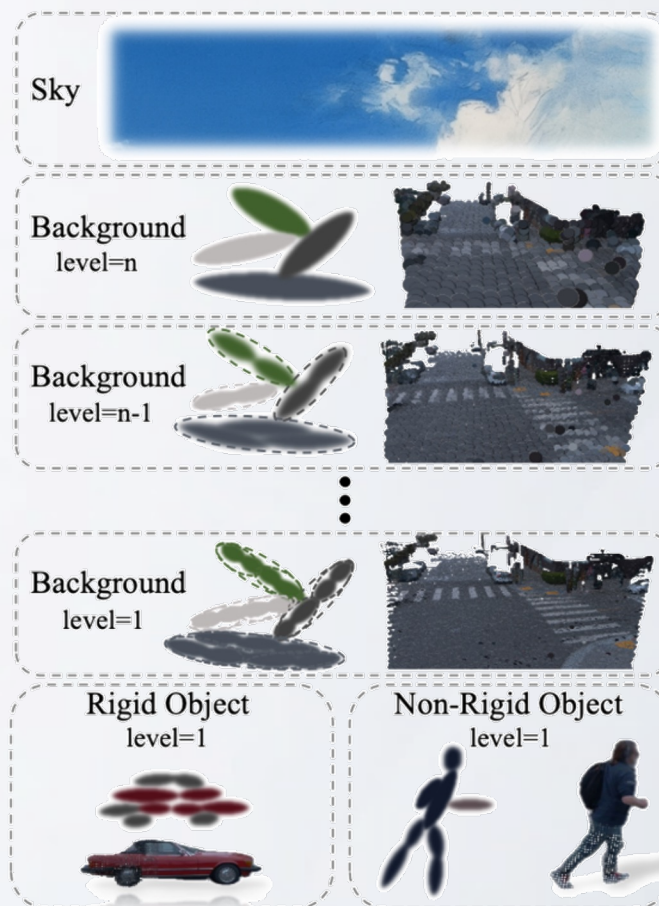


The **Hierarchy UGP** framework is proposed, achieving for the first time **high-quality reconstruction and real-time rendering** of large-scale dynamic street scenes.

(a) Hierarchy UGP



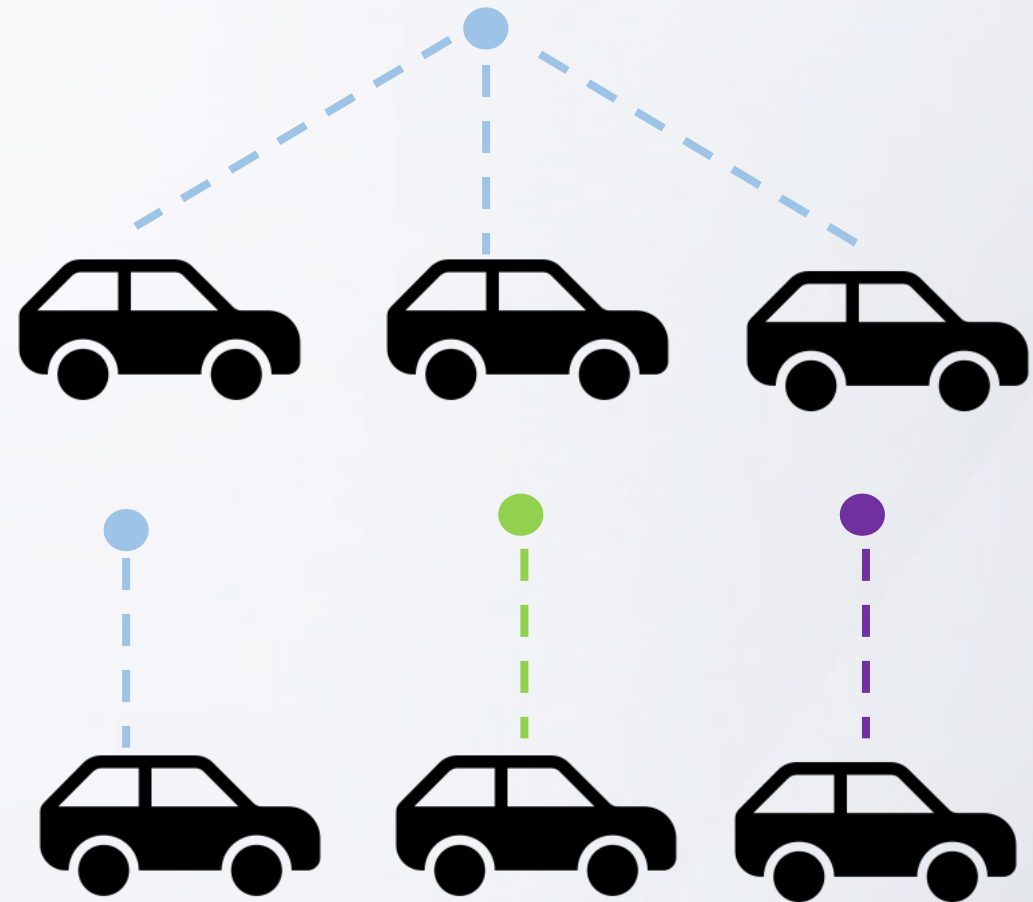
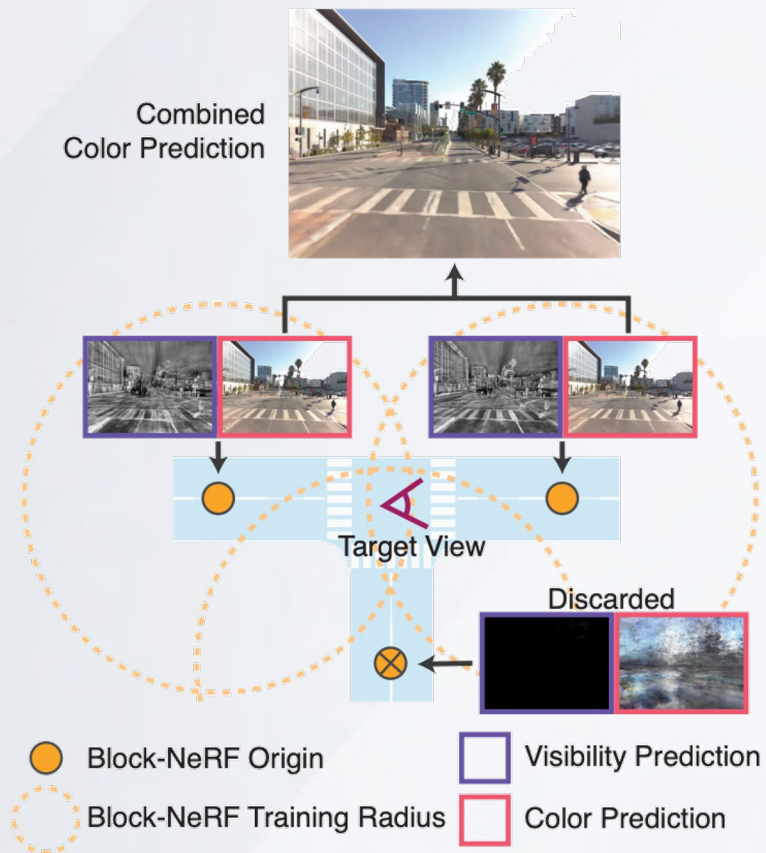
(b) Primitive Level





Training Strategy

Block-wise Objects: For vehicles that span across multiple sub-scenes.



Block-wise Objects

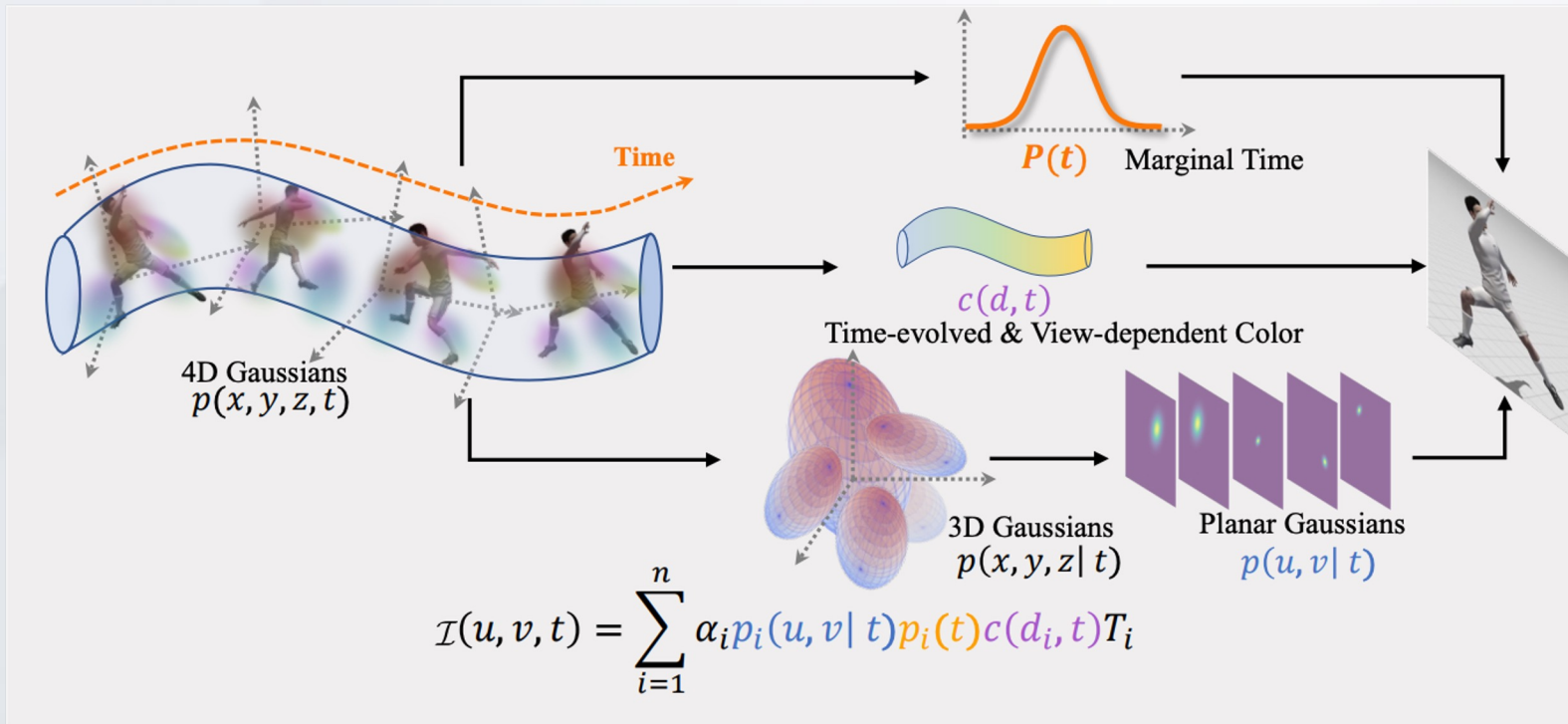


w/ Block-wise Objects



w/o Block-wise Objects

Temporal Scale Initialization



$$S = \text{diag}(s_x, s_y, s_z, s_t)$$



Training Strategy

Temporal Scale Initialization

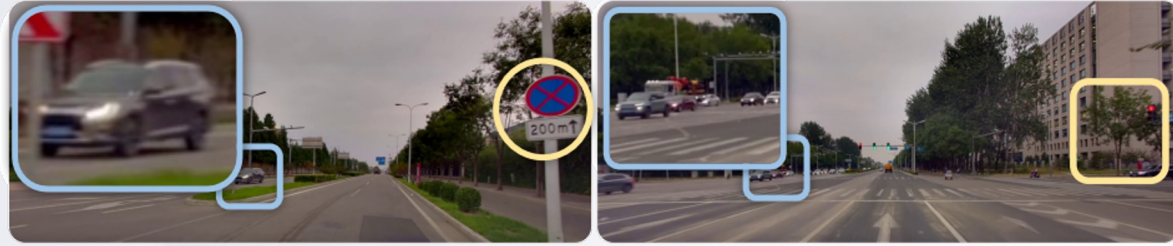


w/ Temporal Scale Initialization



w/o Temporal Scale Initialization

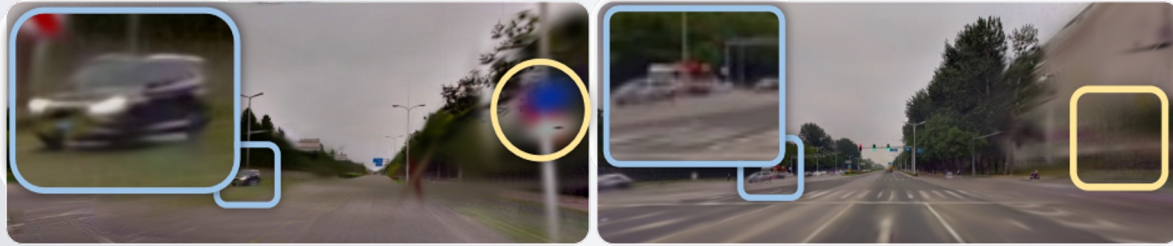
Experimental Results (Qualitative)



(a) GT



(b) Ours



(c) OmniRe



(d) Street GS



(e) 4D-GS



(f) Hierarchical GS

Experimental Results (Quantitative)

	Large 001			Large 002			Large 003			Large 004		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
PVG [3]	25.03	0.825	0.321	23.27	0.792	0.373	23.79	0.788	0.341	21.82	0.743	0.366
OmniRe [4]	24.48	0.776	0.336	22.17	0.734	0.386	23.00	0.729	0.323	23.07	0.737	0.222
Hierarchical GS [9]	27.00	0.870	0.171	25.33	0.852	0.185	24.73	0.840	0.191	25.68	0.857	0.186
Ours	27.29	0.871	0.164	25.46	0.853	0.182	25.24	0.845	0.185	26.5	0.860	0.181

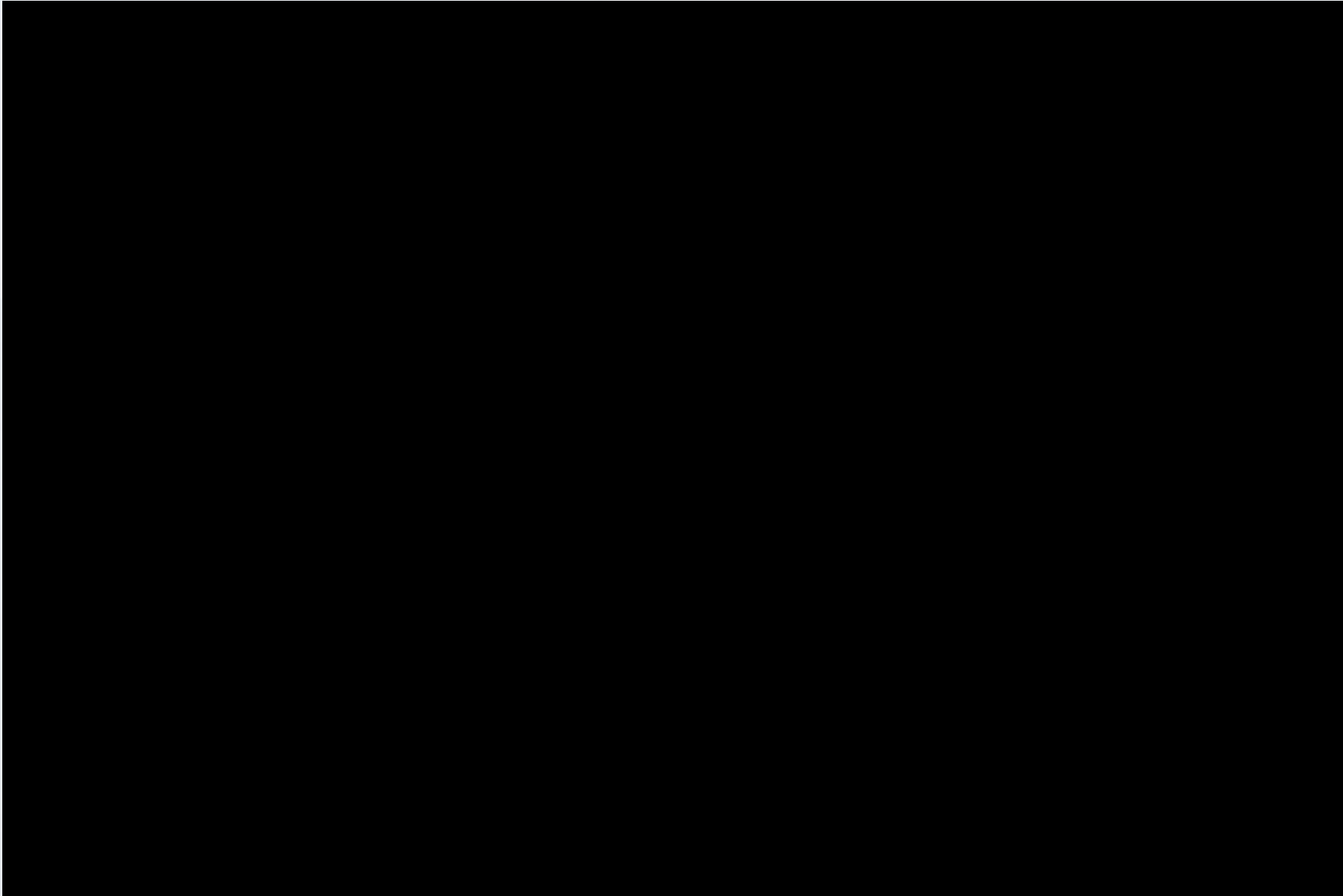
Table 2. **More Quantitative Comparison on Dynamic City.** We selected every 10th frame as a test frame and computed visual quality metrics in four large scenes. **used LOD level $\tau = 1$.** Each cell is colored to indicate the **best**.

Methods	StreetGS				OmniRe				Ours			
	PSNR↑	LPIPS↓	PSNR*↑	SSIM*↑	PSNR↑	LPIPS↓	PSNR*↑	SSIM*↑	PSNR↑	LPIPS↓	PSNR*↑	SSIM*↑
327	32.46	0.120	26.66	0.823	36.24	0.090	28.56	0.839	36.43	0.085	34.52	0.947
614	30.95	0.161	30.74	0.897	30.97	0.169	26.36	0.795	30.76	0.158	35.41	0.964
621	32.01	0.102	24.98	0.810	35.05	0.074	27.23	0.825	35.04	0.071	32.72	0.936
703	31.16	0.142	25.98	0.820	33.37	0.094	27.09	0.806	33.51	0.098	32.65	0.933
788	30.71	0.161	25.92	0.824	31.24	0.150	25.43	0.789	31.73	0.139	32.96	0.949

Table 3. **More Quantitative Comparison on Waymo.** We computed visual quality metrics, where * denotes the metrics for the pedestrian regions. Each cell is colored to indicate the **best**.



Demo1



bilibili 直播
1799624863

下次一定YAY bilibili

从仿真单个场景到仿真完整区域，进一步加快VLA模型强化训练速度。

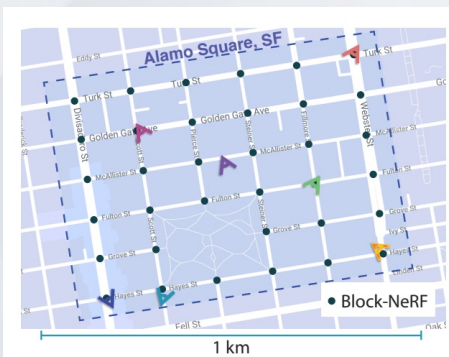


Baidu, Google, Alibaba, Tencent, Microsoft, Intel, NVIDIA, and others are contributing to the development of the VLA model. The VLA model is a large-scale model that can be trained on a wide range of data, including text, images, and video. The VLA model is designed to be a general-purpose model that can be used for a wide range of tasks, including natural language processing, computer vision, and robotics. The VLA model is a key component of the Baidu AI ecosystem, and it is expected to play a major role in the development of future AI applications.

Research Objective: To address both dynamic street scene reconstruction and large-scale scene reconstruction within a unified framework.

Problem 1:

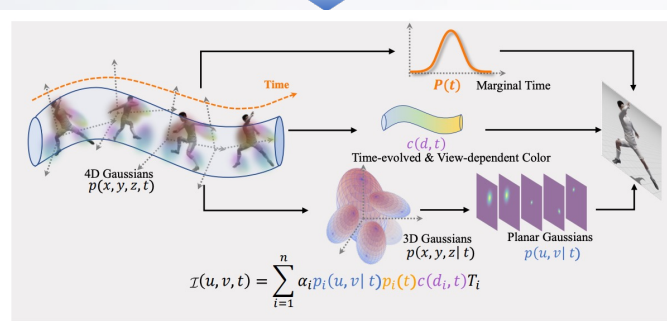
Large-Scale Scene Reconstruction



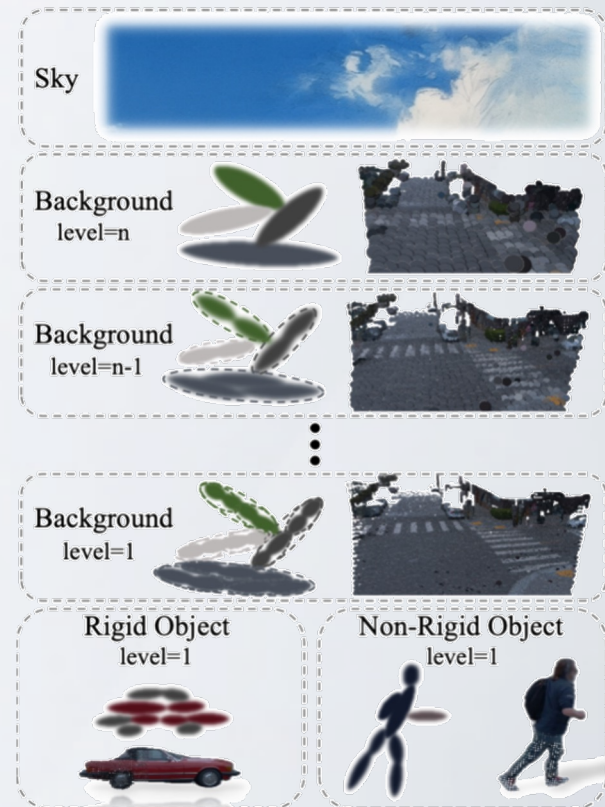
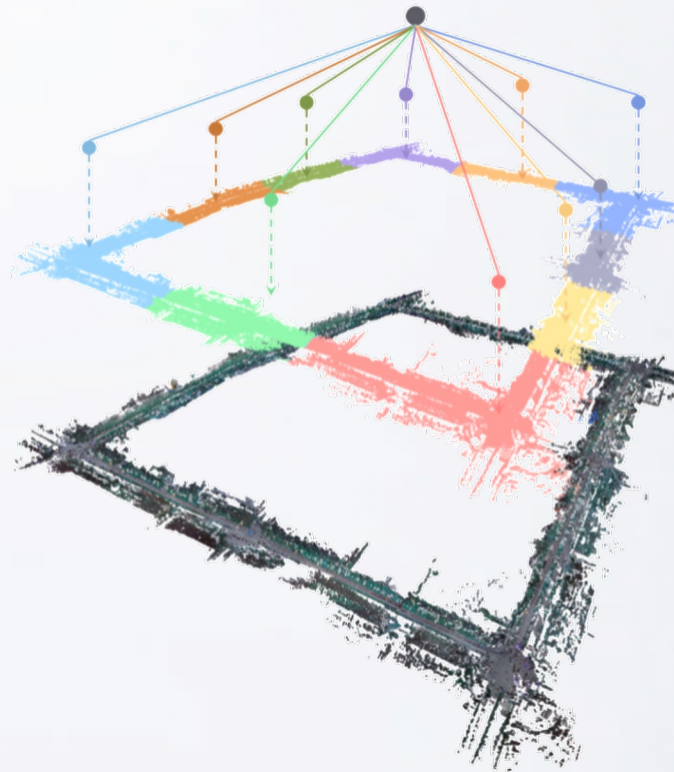
Inspired by Block NeRF, we divide large-scale dynamic scenes into multiple smaller sub-scenes.

Problem 2:

Dynamic Object Reconstruction



We use 4D Gaussian Splatting (4DGS) as the reconstruction primitive to achieve high-quality reconstruction of dynamic objects.





Thank You!

Hongyang Sun
22351179@zju.edu.cn