NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

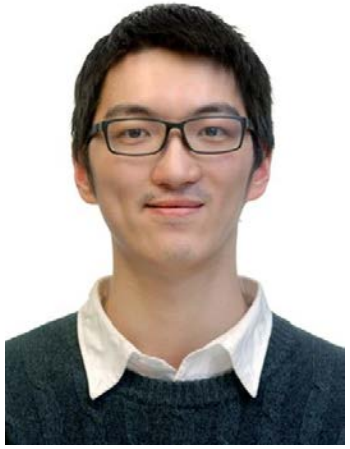
Toward World Models: From 3D to 4D City Generation

Haozhe Xie (谢浩哲)

Research Fellow at MMLab@NTU

haozhe.xie@ntu.edu.sg

Haozhe Xie (谢浩哲)



- Research Fellow at MMLab@NTU
- #3D Vision #AIGC #Robotics

haozhexie.com

@hzxie

@hzxie

@zjhzxhz

Work Experience

MMLab @NTU

Tencent AI Lab

商汤 sensetime

Representative Works

GaussianCity: Generative Gaussian Splatting for Unbounded 3D City Generation

CVPR 2025 Stars 161

CityDreamer: Compositional Generative Model of Unbounded 3D Cities

CVPR 2024 Stars 637

GRNet: Gridding Residual Network for Dense Point Cloud Completion

ECCV 2020 Stars 316

Pix2Vox: Context-aware 3D Reconstruction from Single and Multi-view Images

ICCV 2019 Stars 503

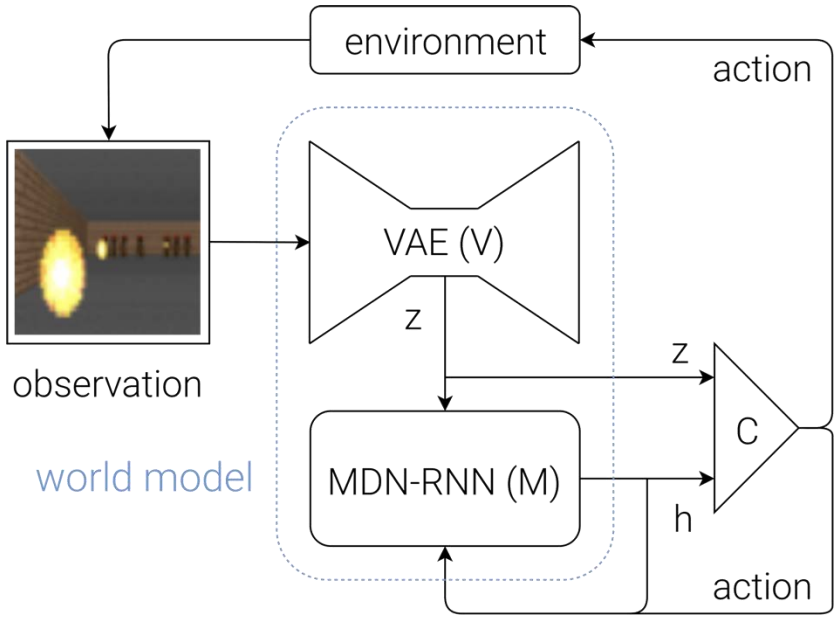
Outline

- **What is a World Model?**
  The Ultimate Goal of 3D/4D City Generation
- **CityDreamer**
  Learning 3D Unbounded Cities from Google Earth
- **GaussianCity**
  60x Faster 3D City Generation!
- **DynamicCity**
  4D Occupancy Generation for Self-Driving
- **CityDreamer4D**
  The First True 4D Unbouded City!

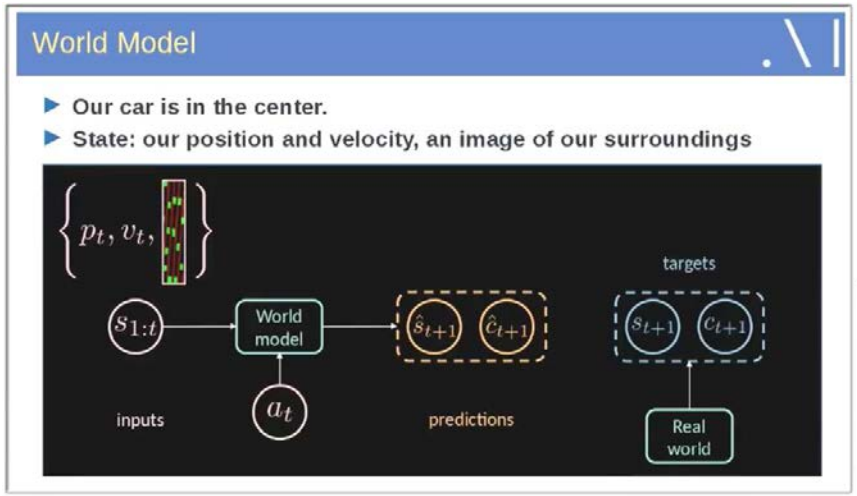
What is a World Model?

The Ultimate Goal of 3D/4D City Generation

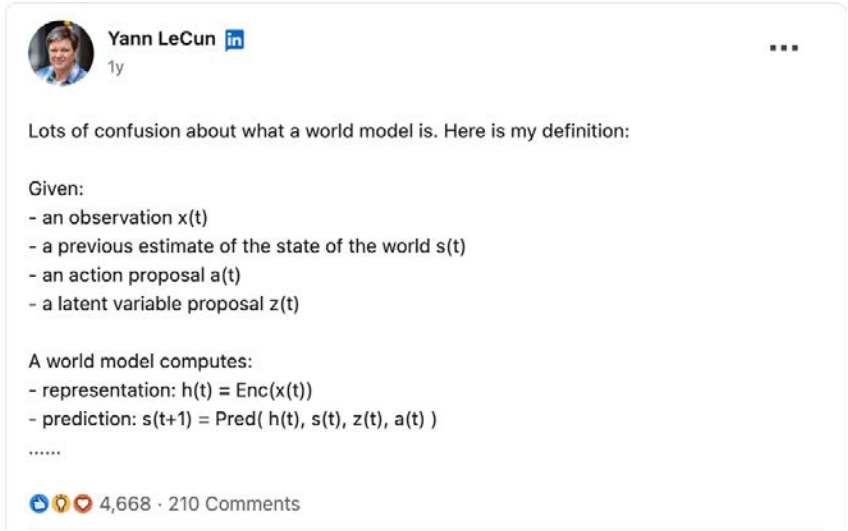
World Model Definition



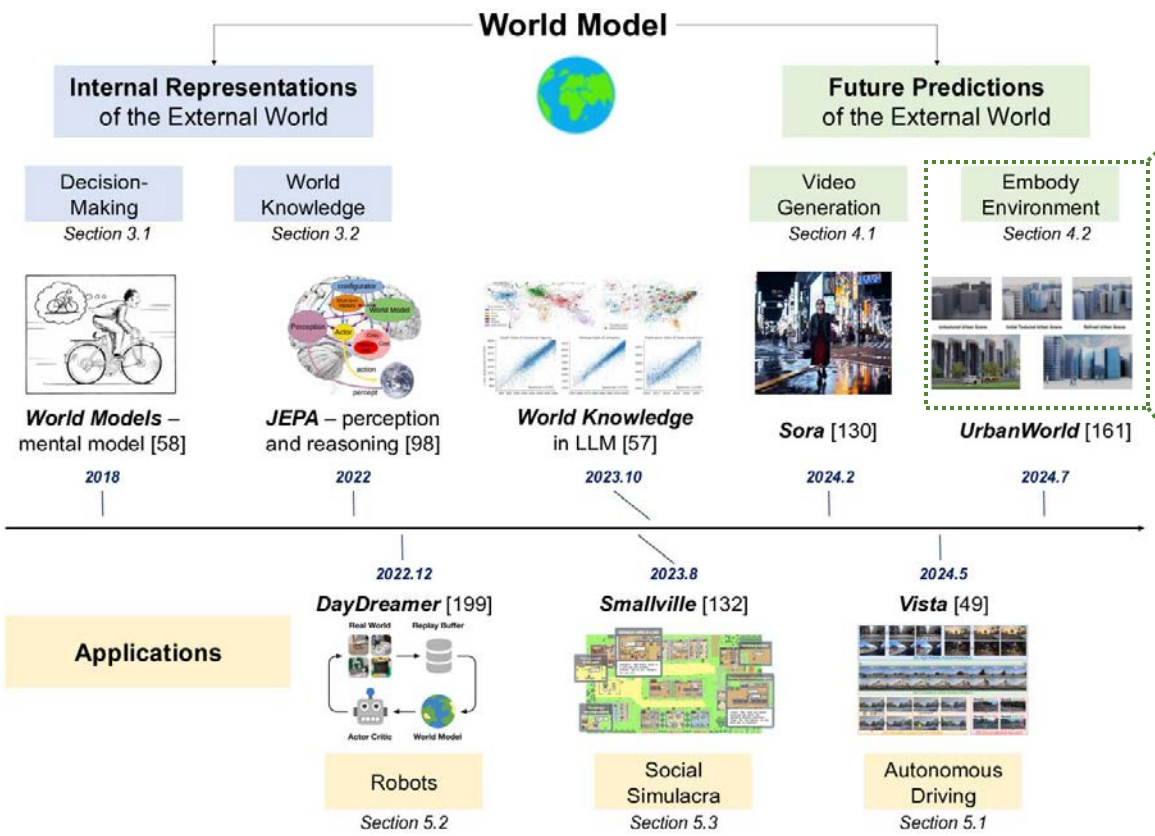
David Ha and Jürgen Schmidhuber. World Models. NeurIPS 2018.

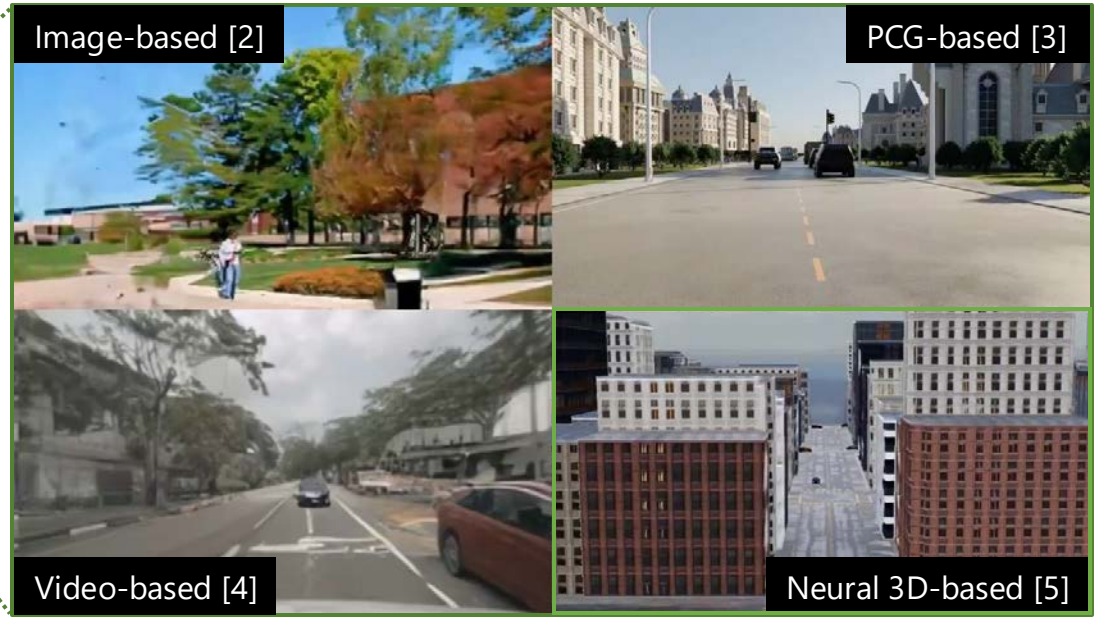


The Next Step Towards Artificial Intelligence
Yann LeCun



Existing World Models





[1] Understanding World or Predicting Future? A Comprehensive Survey of World Models. arXiv 2411.14499.
[2] Wonderjourney: Going from Anywhere to Everywhere. CVPR 2024.
[3] CityX: Controllable Procedural Content Generation for Unbounded 3D Cities. arXiv 2407.17572.
[4] MagicDrive: Street View Generation with Diverse 3D Geometry Control. ICLR 2024.
[5] CityDreamer4D: Compositional Generative Model of Unbounded 4D Cities. arXiv 2501.08983.

CityDreamer

Learning 3D Unbounded Cities from Google Earth

hzxie/CityDreamer

hzxie/city-dreamer

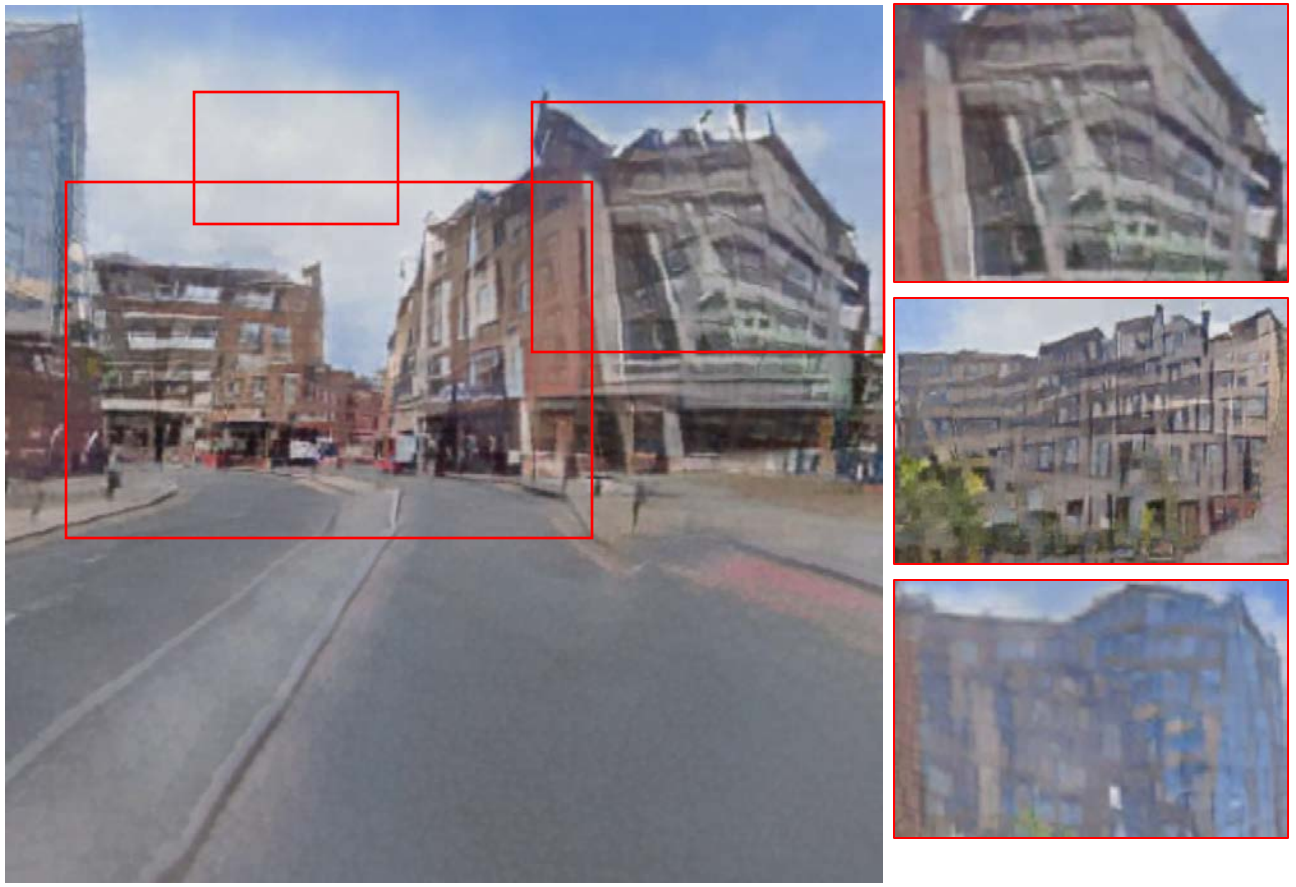
CityDreamer: Compositional Generative Model of Unbounded 3D Cities. CVPR 2024.

Challenges

- Creating cities is **more complex** compared to generating natural scenes

GANCraft [CVPR'21]

SceneDreamer [TPAMI'23]

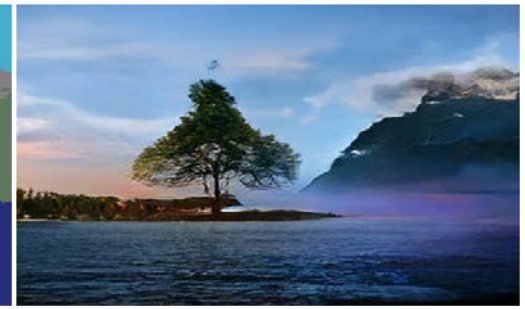


InfiniCity [ICCV'23]

Challenges

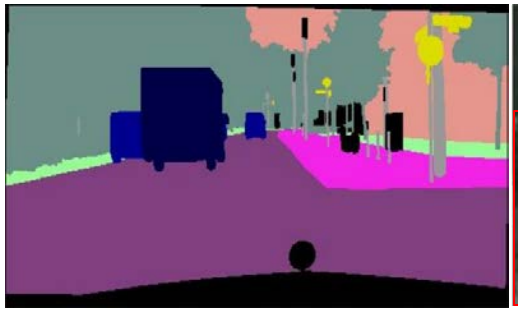
- Objects are similar in natural scenes


- Synthetic nature images are realistic

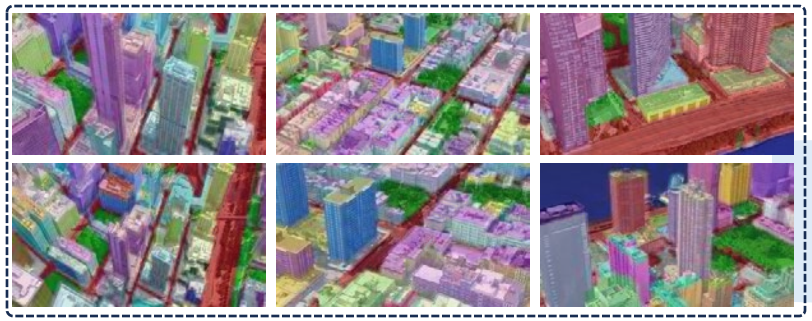



- Buildings are diverse in cities

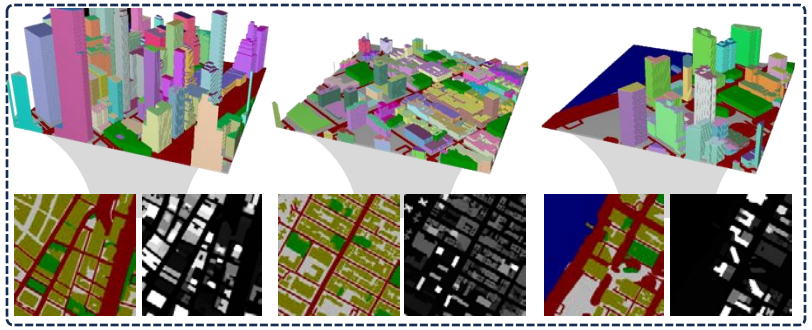

- Synthetic city images are not realistic



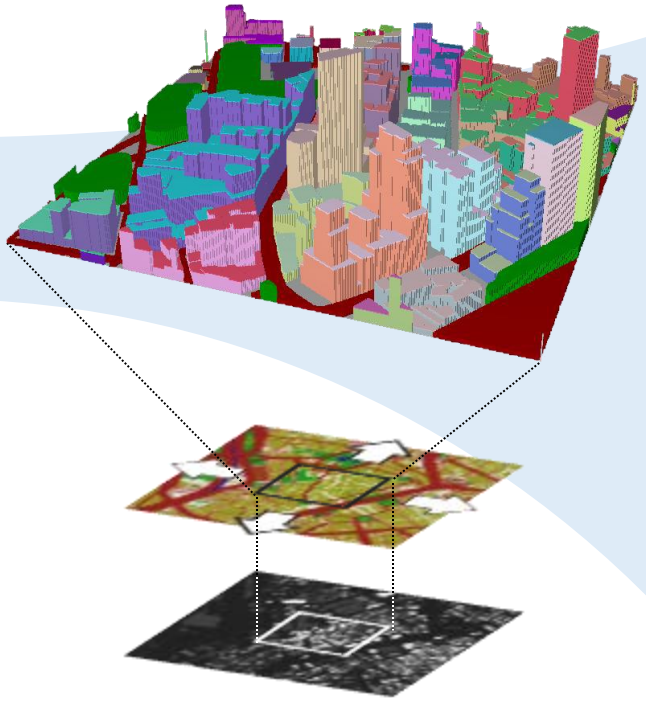

CityDreamer Overview



(a) Google Earth Dataset: Real-world City Appearance



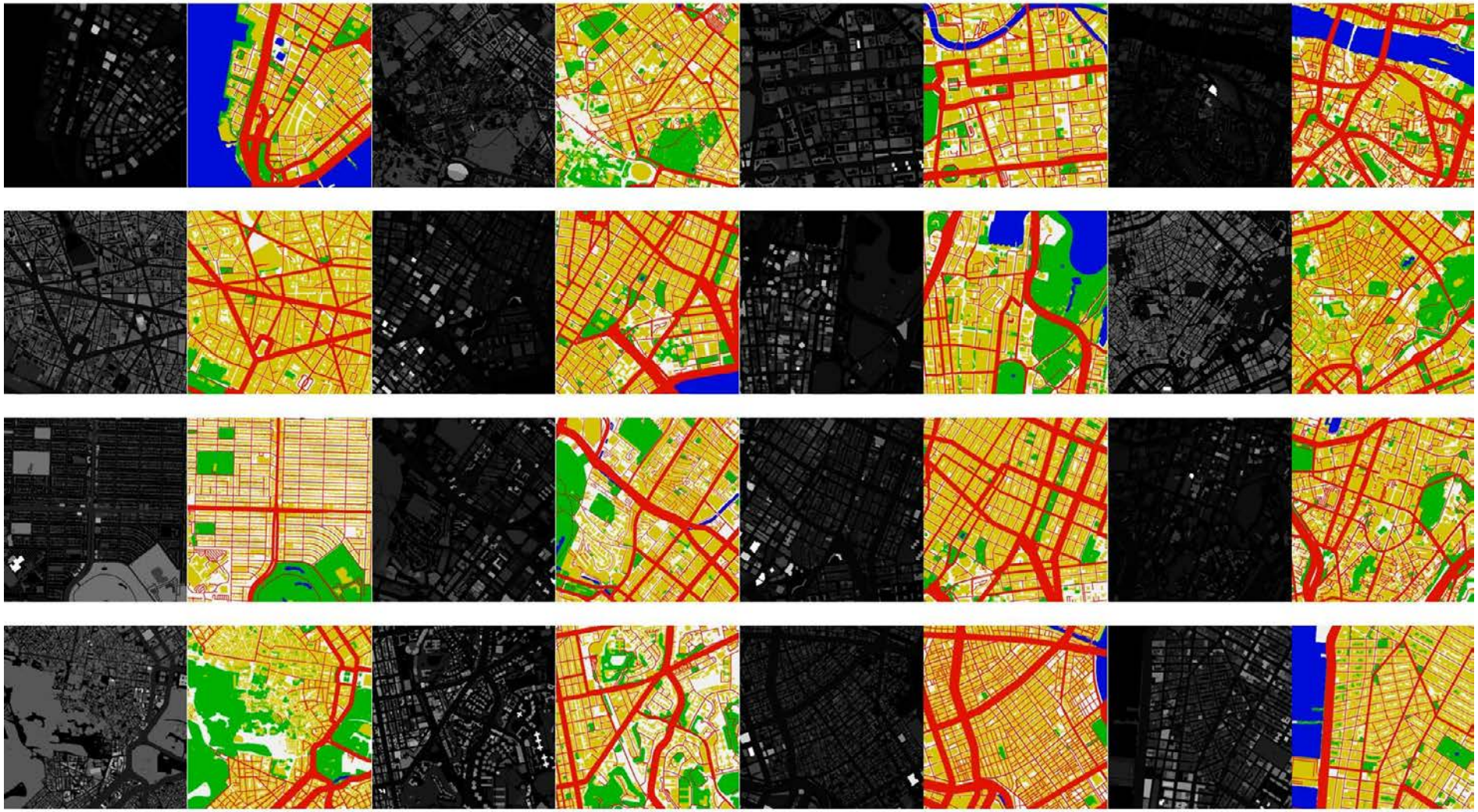
(b) OSM Dataset: Real-world City Layout



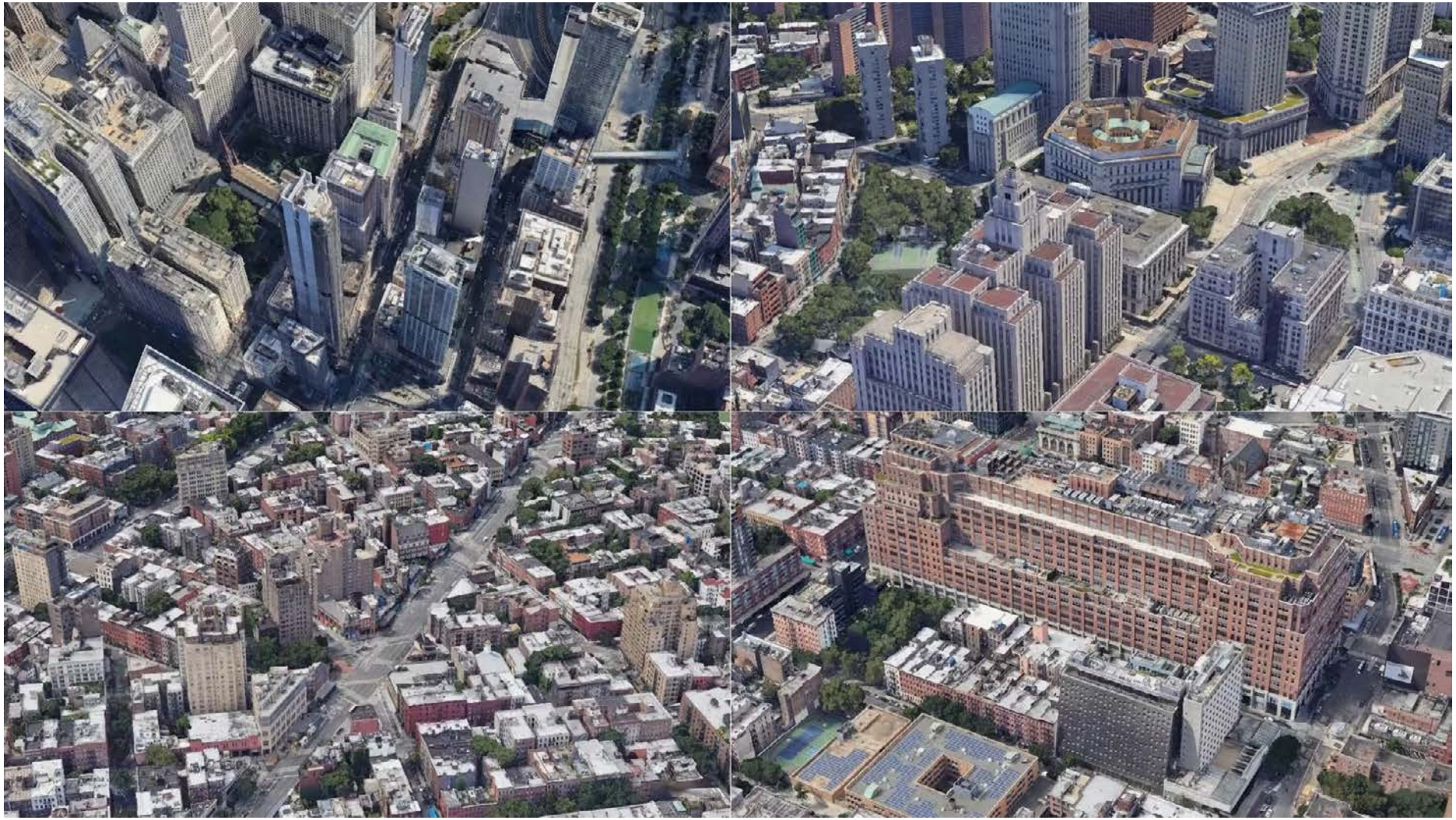
(c) Unbounded City Layout Generation

(d) CityDreamer Generated 3D Cities

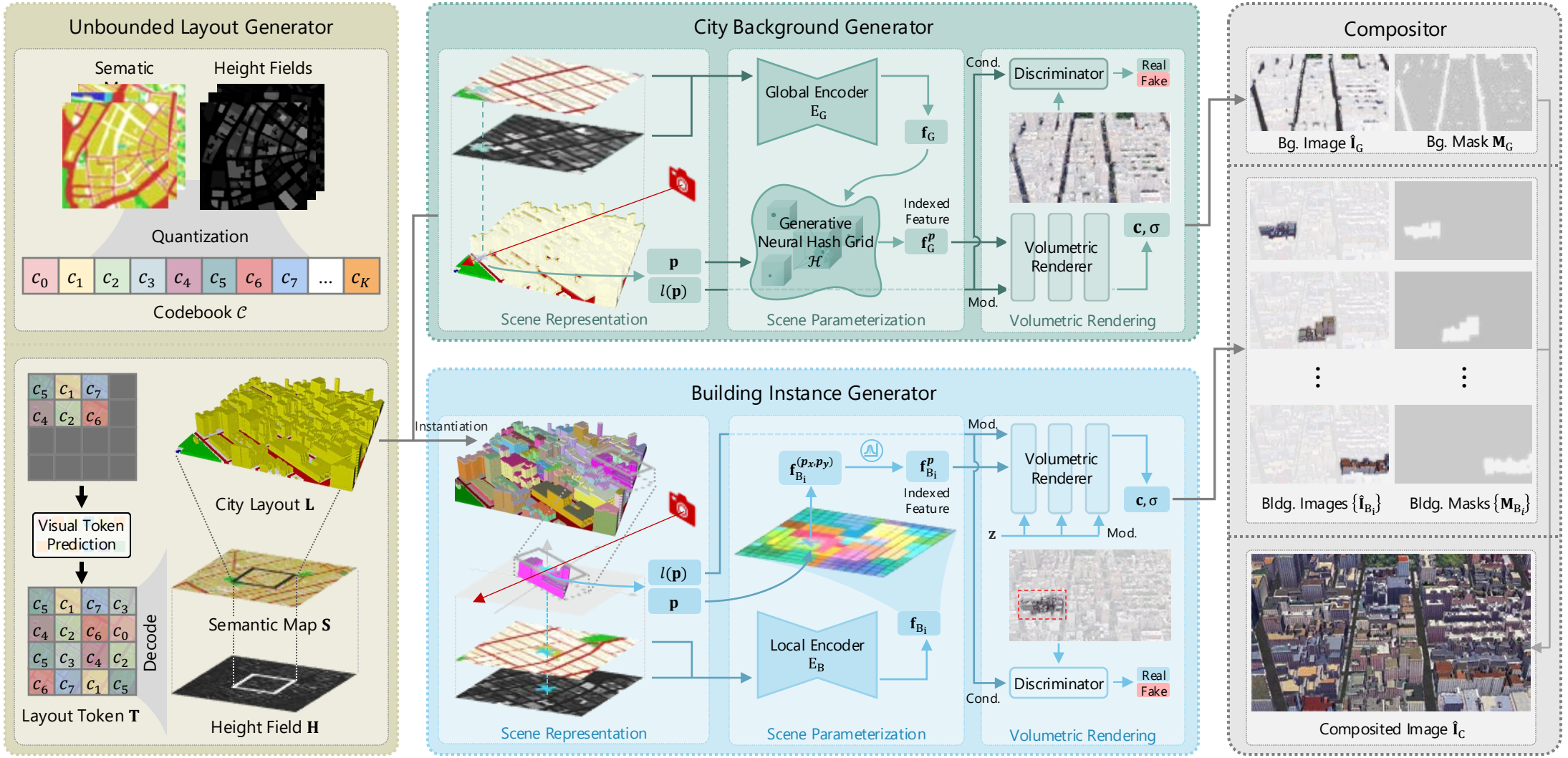
The OSM Dataset



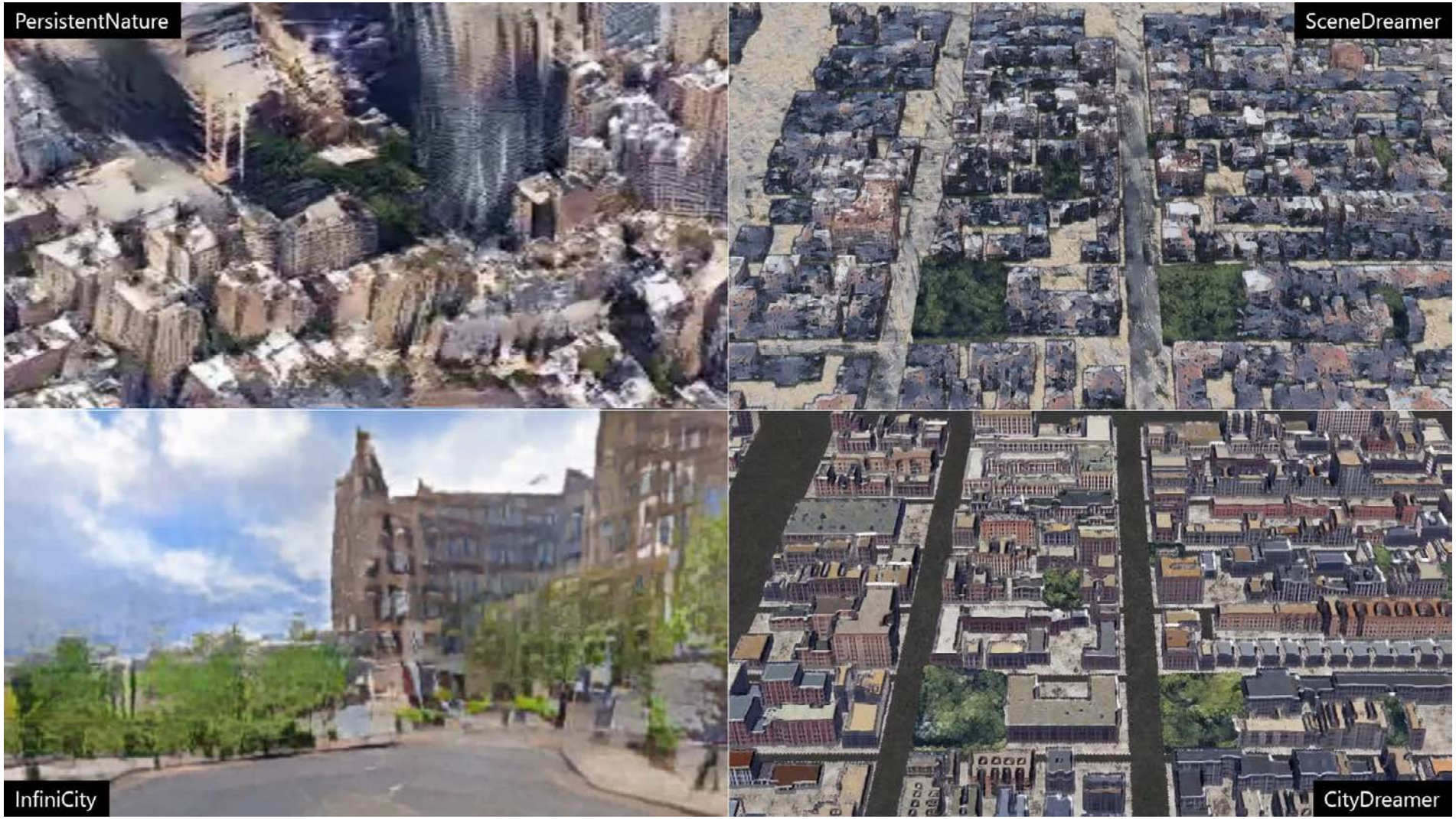
The GoogleEarth Dataset



CityDreamer Framework



Comparison to SOTA Methods



Arbitrary View Rendering

GaussianCity

60x Faster 3D City Generation!

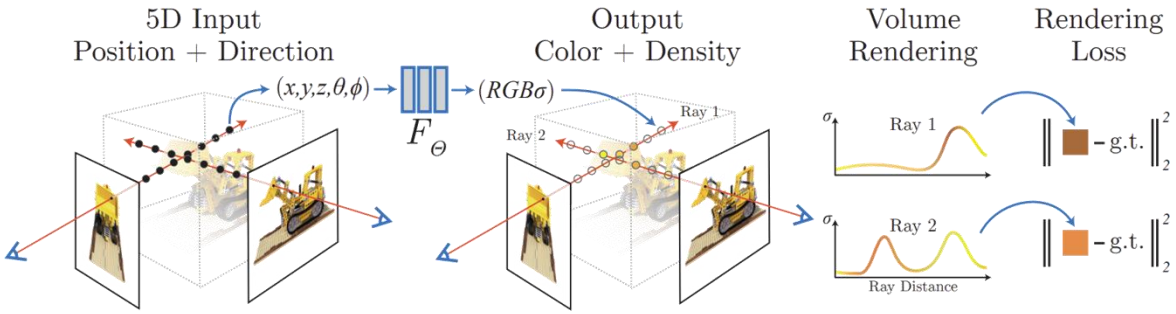
hzxie/GaussianCity

hzxie/gaussian-city

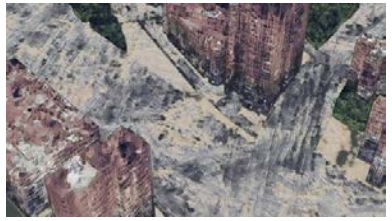
Generative Gaussian Splatting for Unbounded 3D City Generation. CVPR 2025.

Challenges

NeRF is Time-inefficient



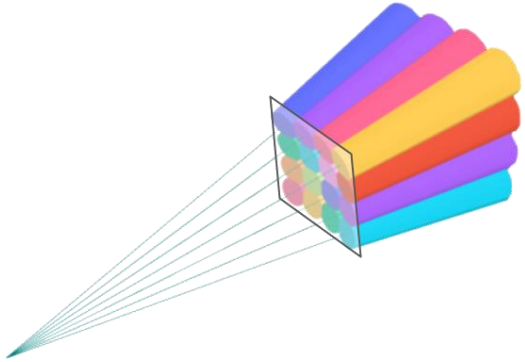
Pers.Nature (5.99 FPS)



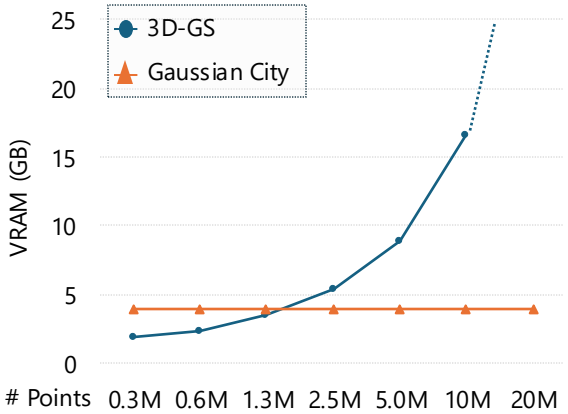
SceneDreamer (1.61 FPS)

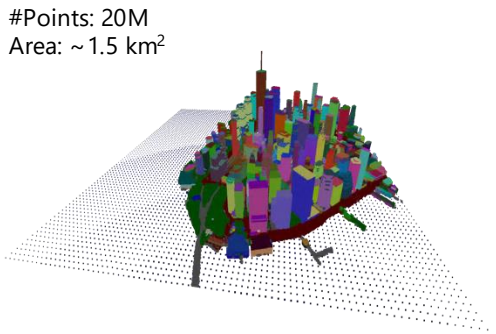
CityDreamer (0.18 FPS)

3D-GS is Time-efficient

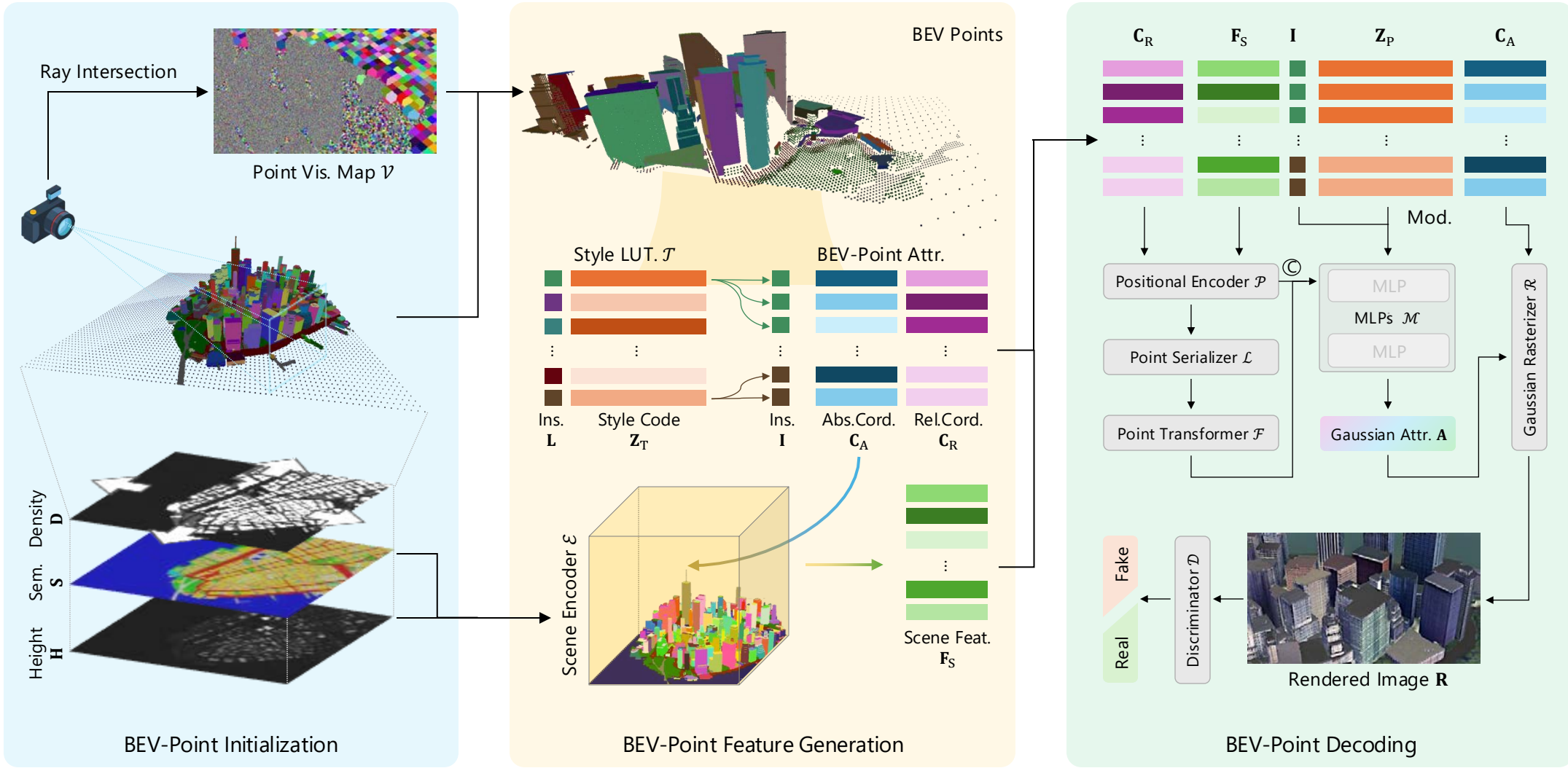


BUT! 3D-GS is Storage-inefficient

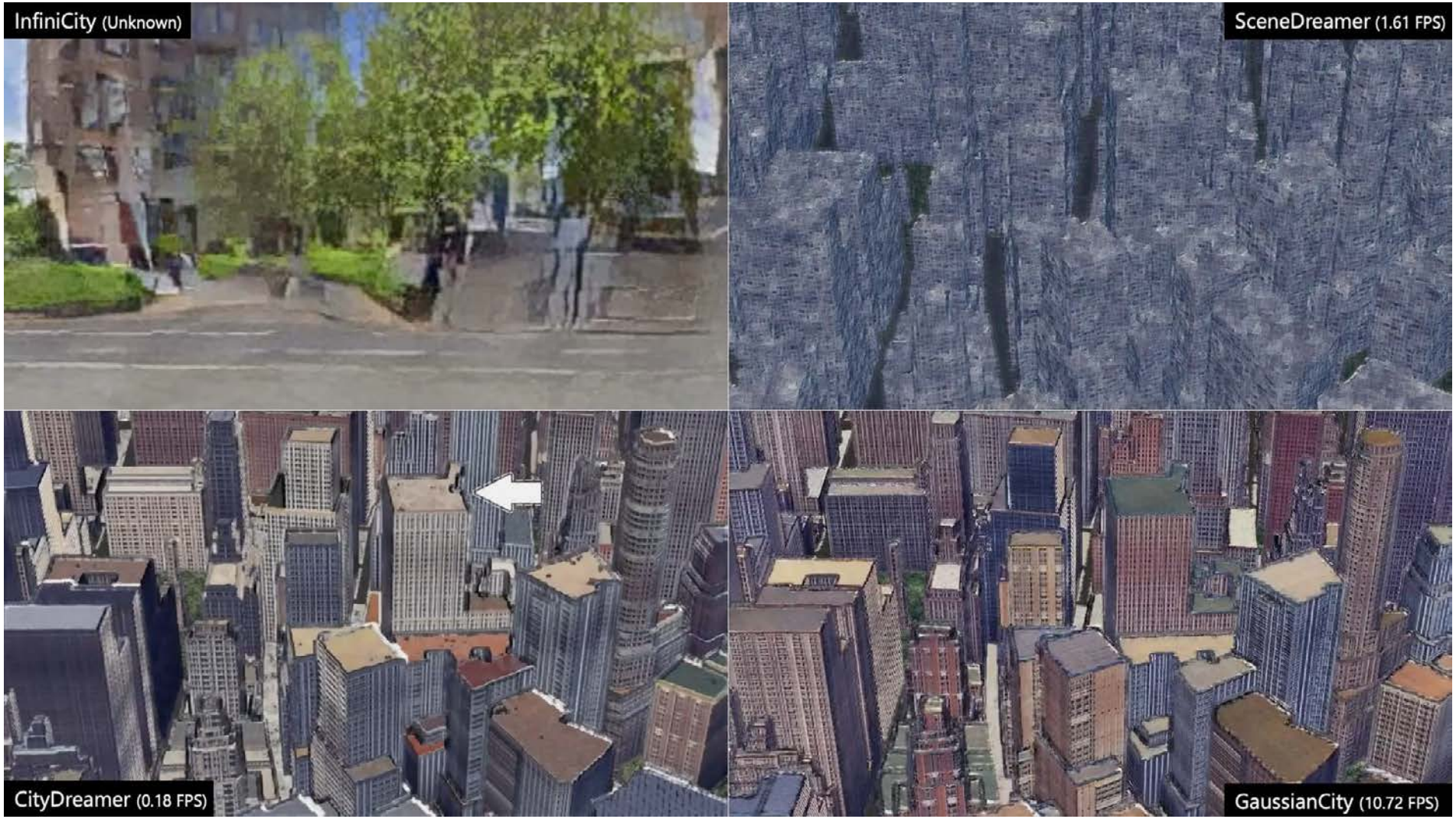




GaussianCity Framework

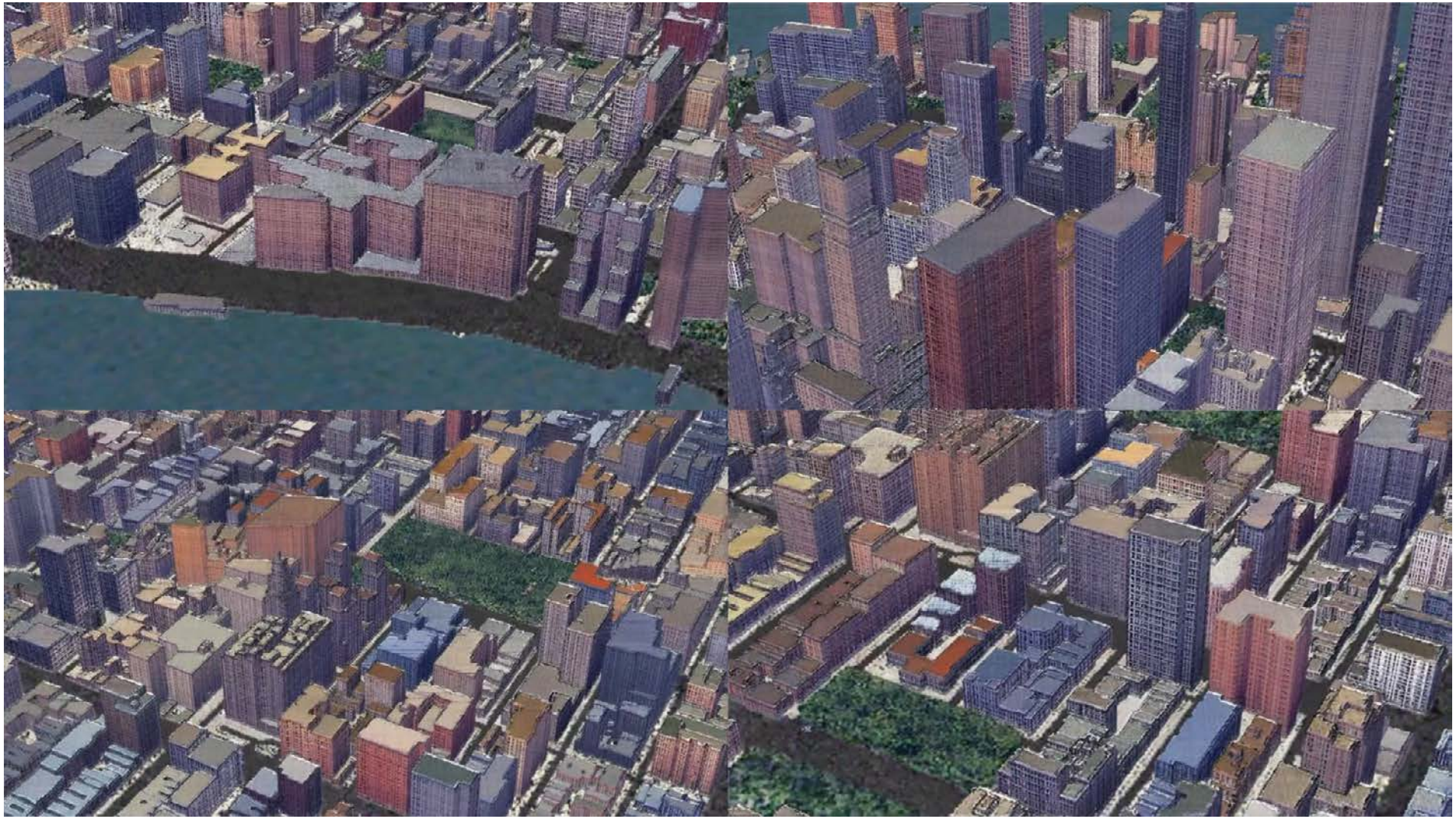


Comparison to SOTA Methods



Better Consistency

Arbitrary View Rendering



DynamicCity

4D Occupancy Generation for Self-Driving

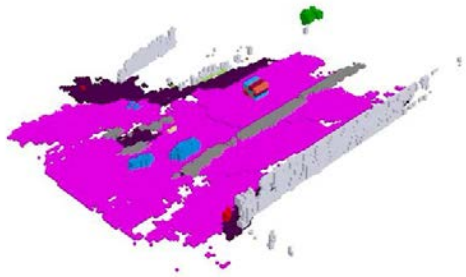
3DTopia/DynamicCity

DynamicCity: Large-Scale 4D Occupancy Generation from Dynamic Scenes. ICLR 2025.

Challenges

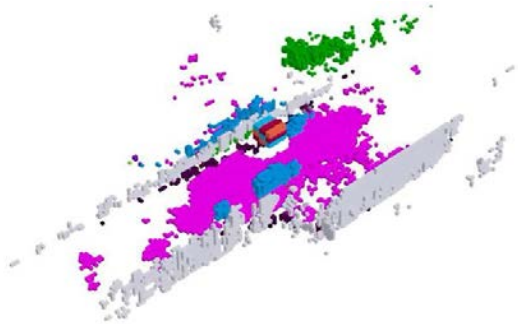
- Inefficient VAEs for 4D data (low compression, poor reconstruction)
- Suboptimal generation quality
- Limited control over the generation process

OccSora



Go Straight

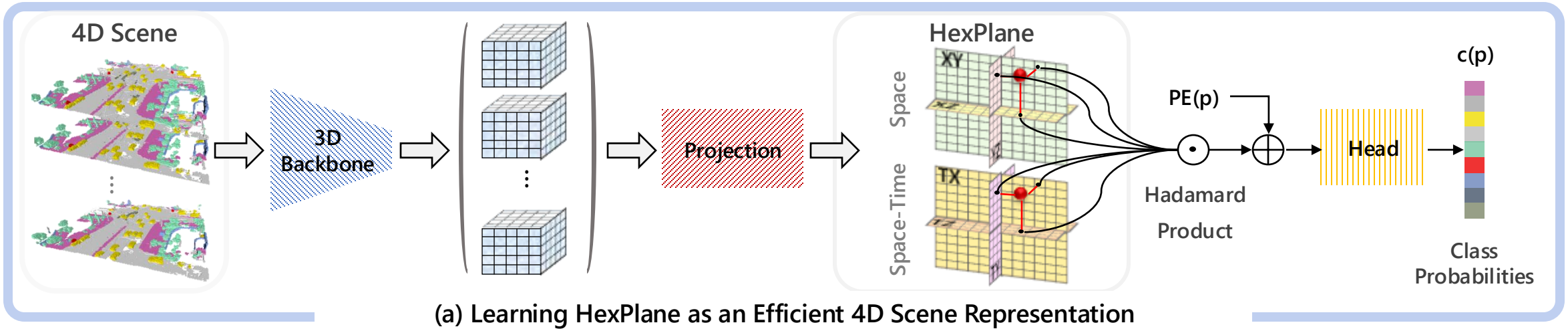
Turning



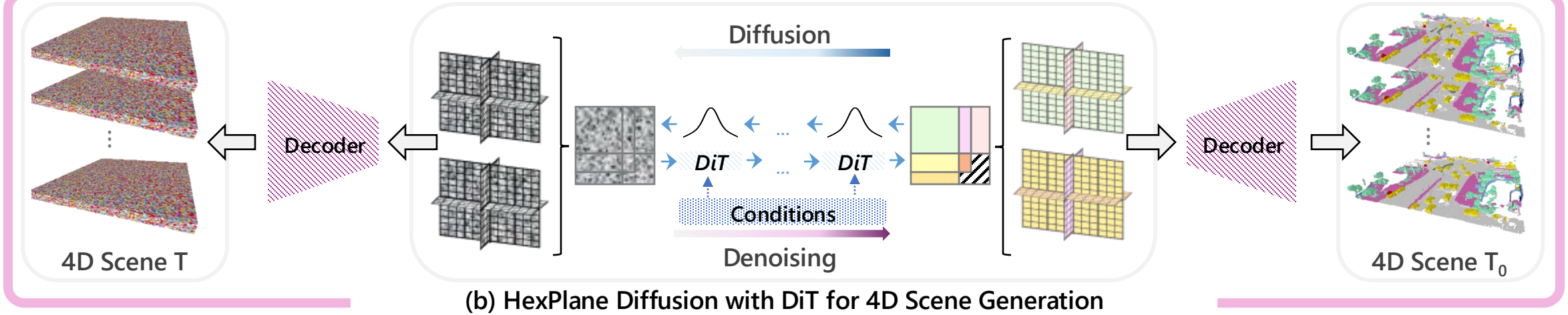
Motionless

OccSora: 4D Occupancy Generation Models as World Simulators for Autonomous Driving. arXiv 2405.20337.

DynamicCity Framework

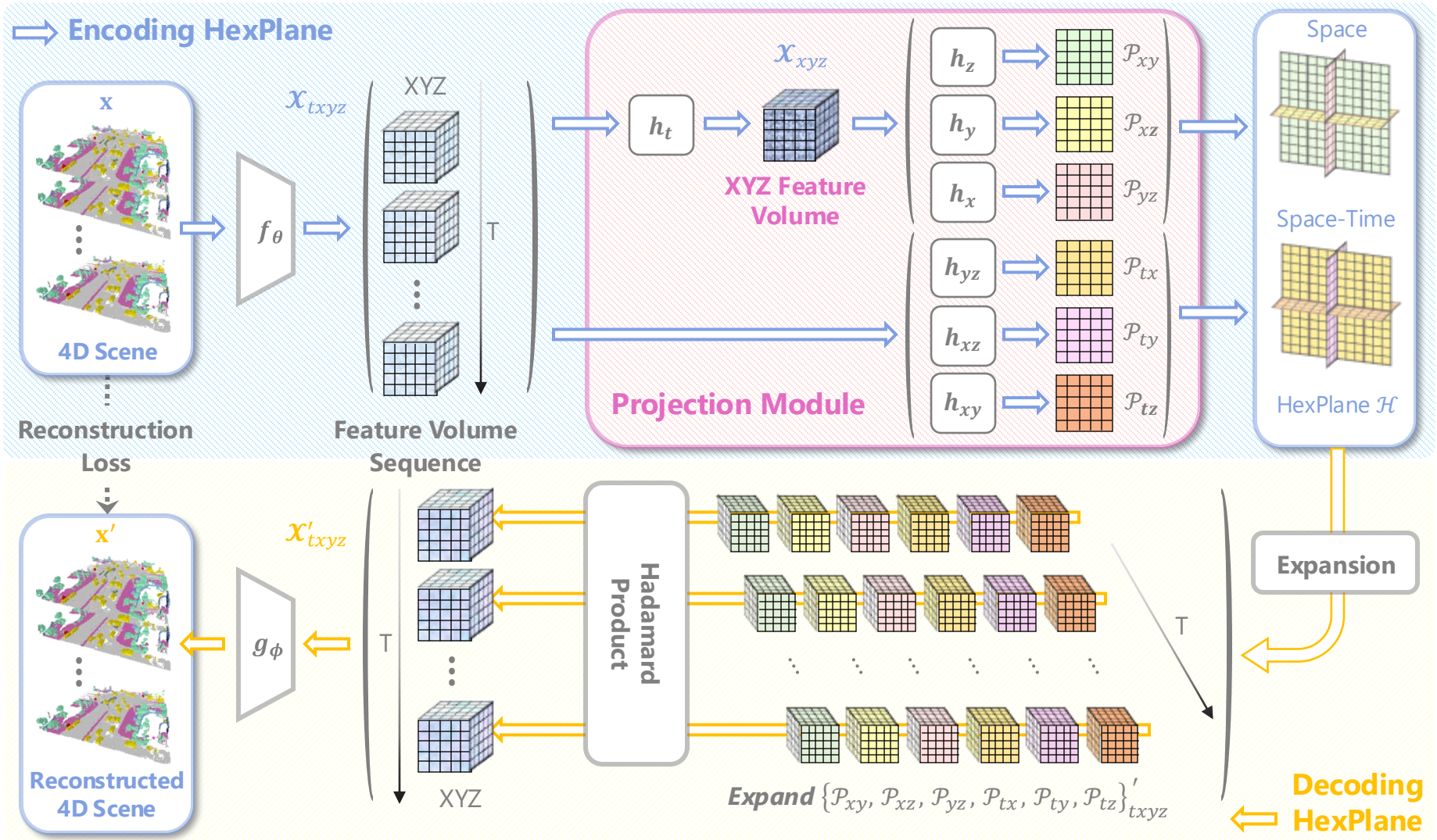


(a) Learning HexPlane as an Efficient 4D Scene Representation



(b) HexPlane Diffusion with DiT for 4D Scene Generation

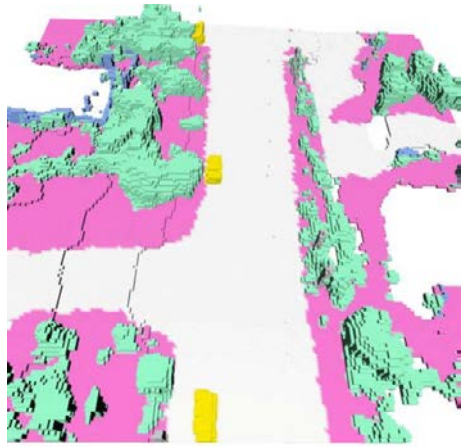
VAE for 4D Scenes

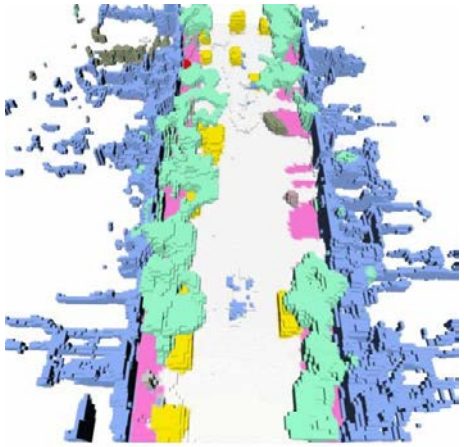


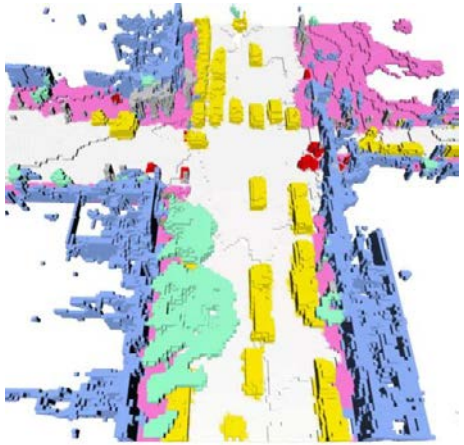
Unconditional 4D Generation

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

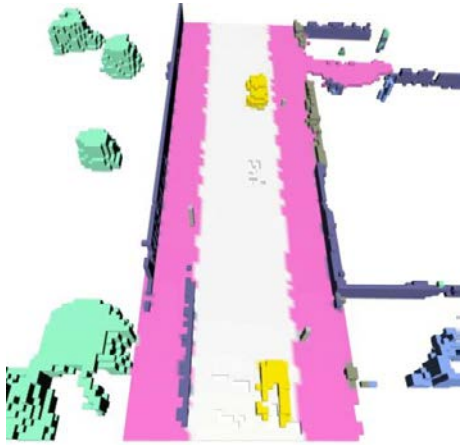
Occ3D-Waymo

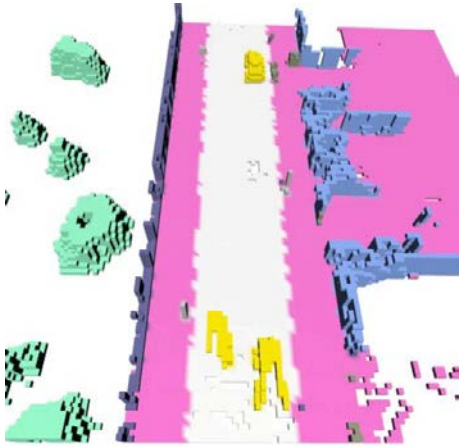


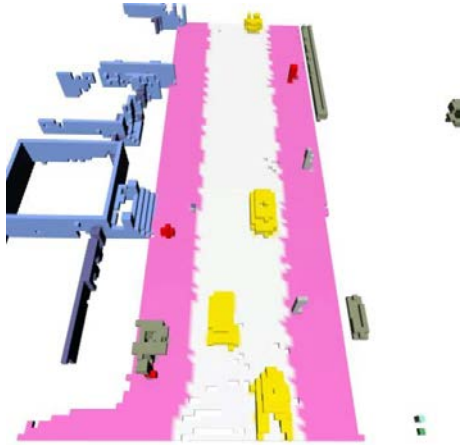




CarlaSC

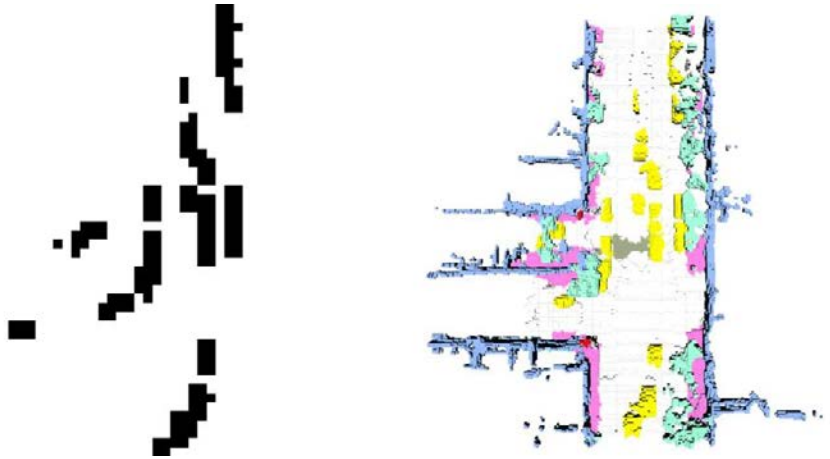




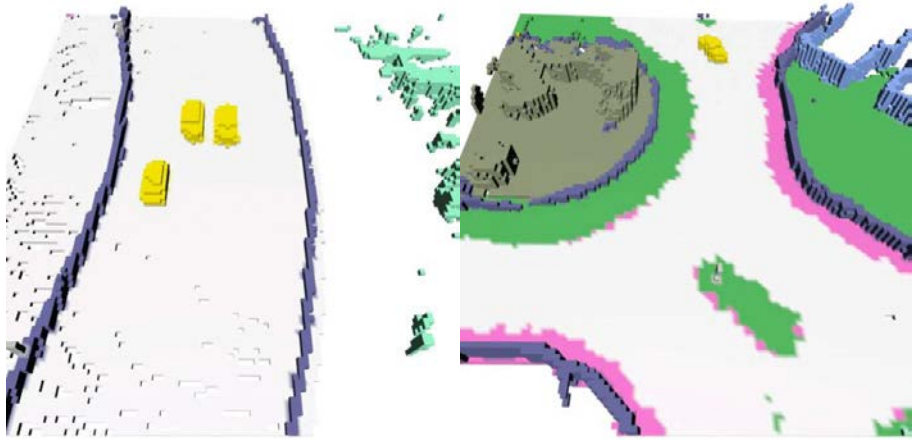


Conditional 4D Generation

Layout-conditioned



Trajectory-conditioned



Inpainting

Outpainting

CityDreamer4D

The First True 4D Unbounded City!

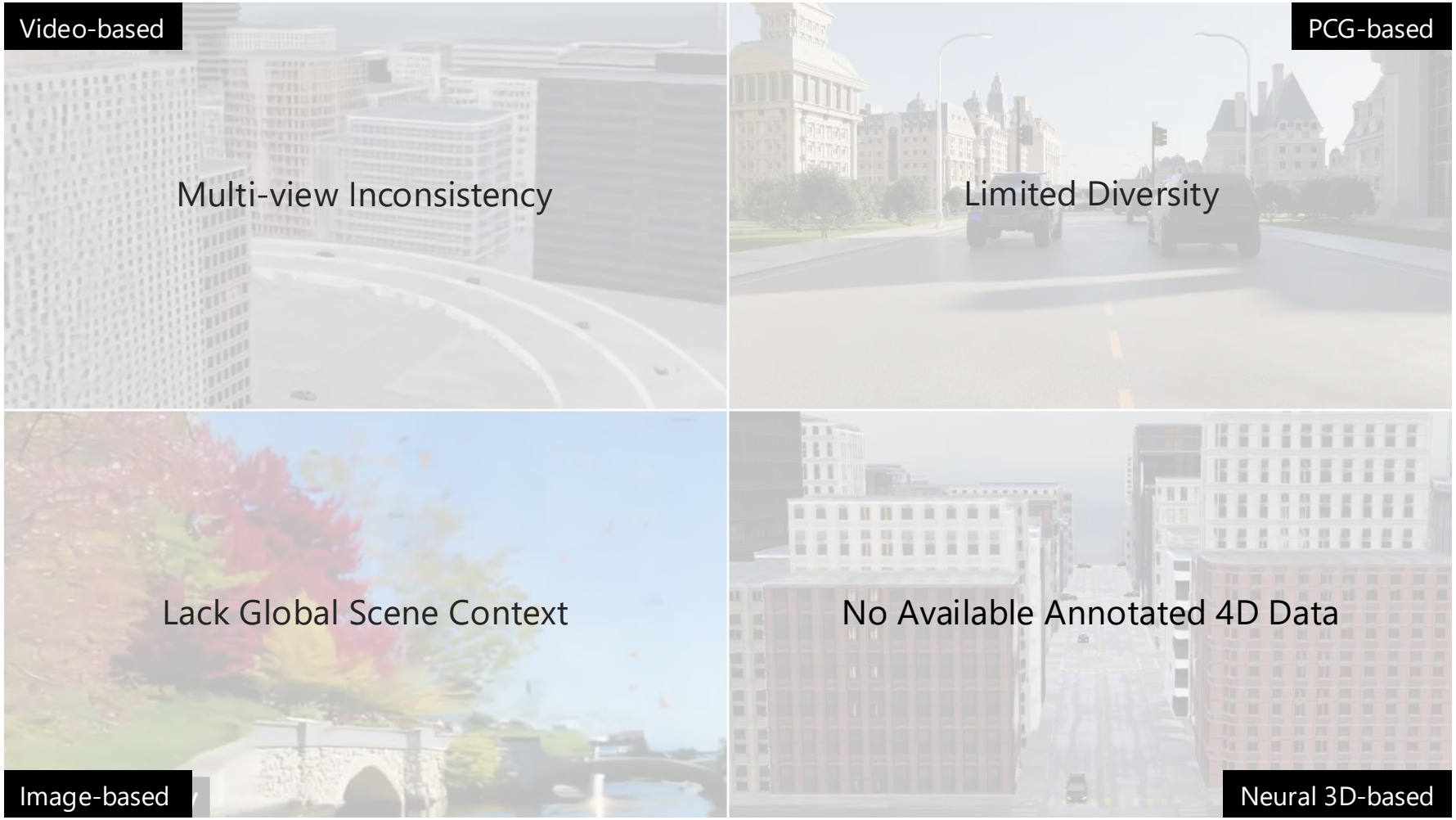
hzxie/CityDreamer4D

CityDreamer4D: Compositional Generative Model of Unbounded 4D Cities. arXiv 2501.08983.

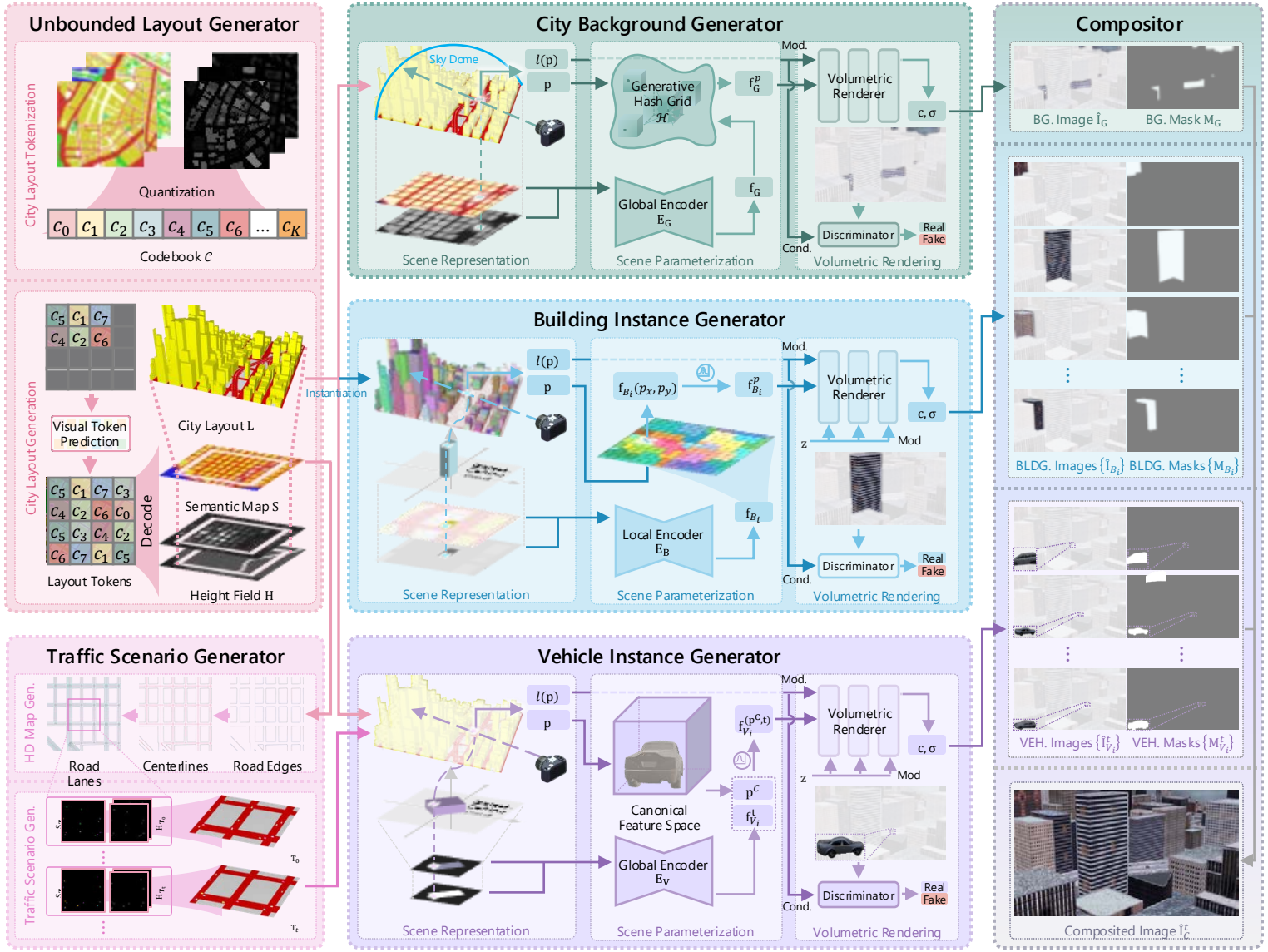
Challenges

[1] Wonderjourney: Going from Anywhere to Everywhere. CVPR 2024.
[2] CityX: Controllable Procedural Content Generation for Unbounded 3D Cities. arXiv 2407.17572.
[3] DimensionX: DimensionX: Create Any 3D and 4D Scenes from a Single Image with Controllable Video Diffusion. arXiv 2411.04928.

Challenges


Video-based
Multi-view Inconsistency
PCG-based
Limited Diversity
Lack Global Scene Context
Image-based
No Available Annotated 4D Data
Neural 3D-based

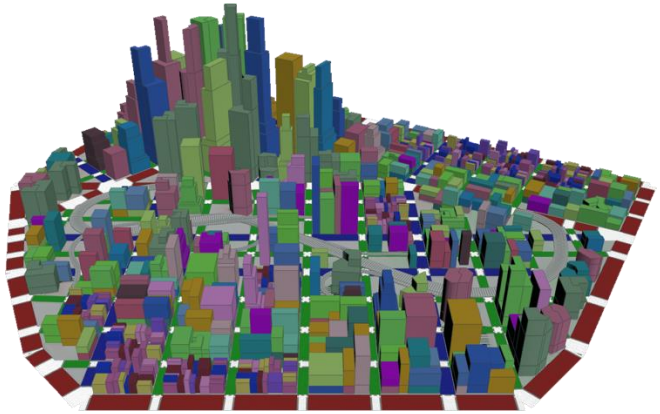
CityDreamer4D Framework



The CityTopia Dataset

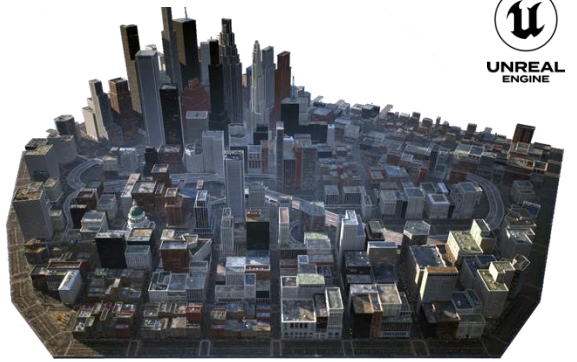
3D Assets
(Small set for Visualization)

UNREAL ENGINE

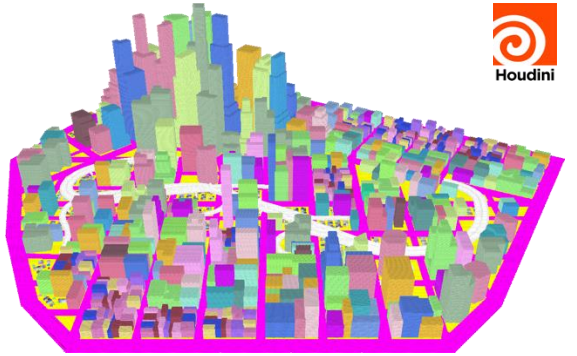


City Prototype

Houdini



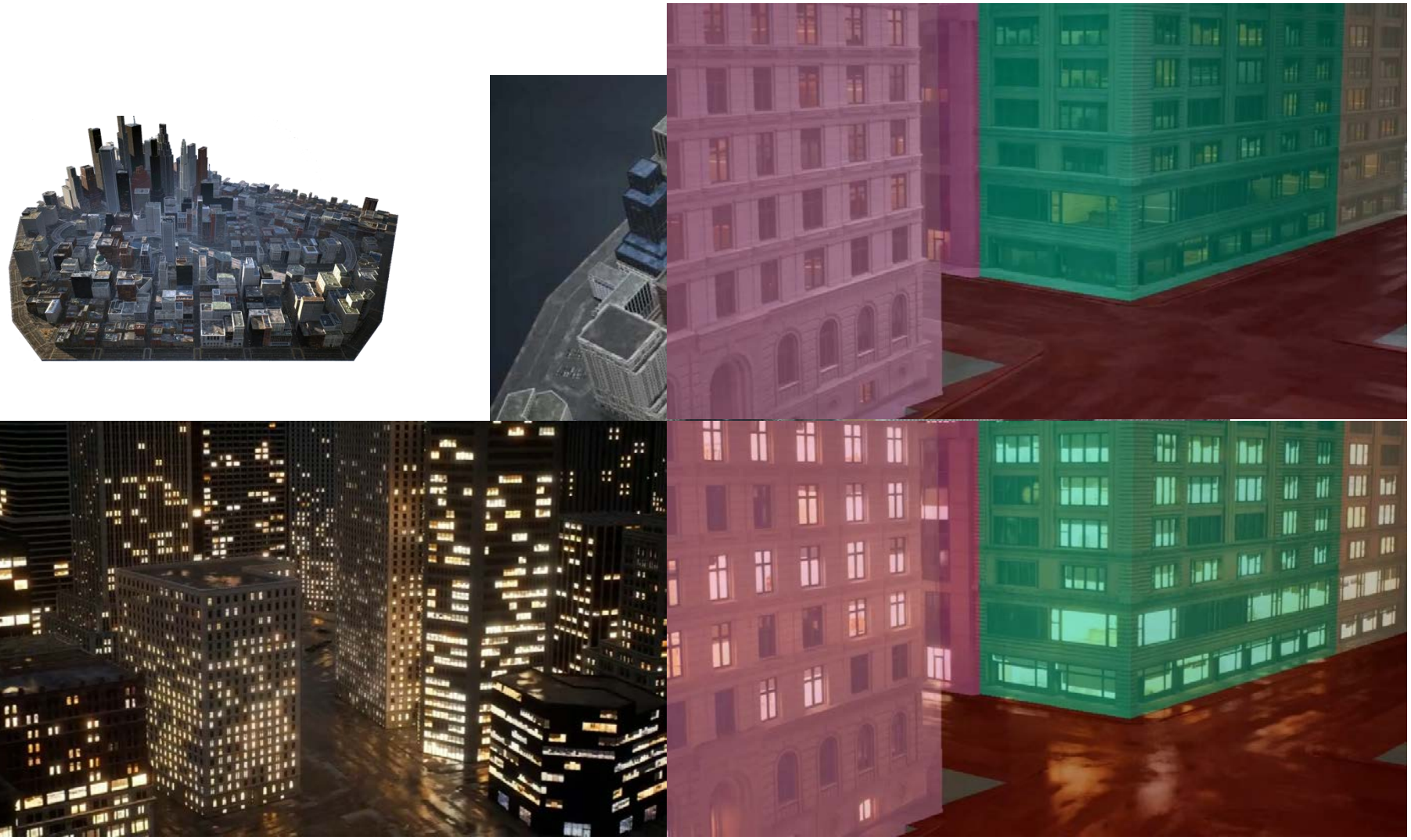
Generated 3D City



Houdini

3D Instance Annotation

The CityTopia Dataset



Comparison to SOTA Methods

Arbitrary View Rendering

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE | S-LAB FOR ADVANCED INTELLIGENCE

Thank You

Haozhe Xie

Research Fellow at MMLab@NTU