Recent Advances of NeRF in Autonomous Driving

Leheng Li 李乐恒

Ph.D. student at HKUST(GZ)

Contents

- Basic of NeRF
- NeRF in autonomous driving (NSG, Block NeRF, UniSim)
- AIGC helps downstream task (Lift3D)

Background of Leheng Li

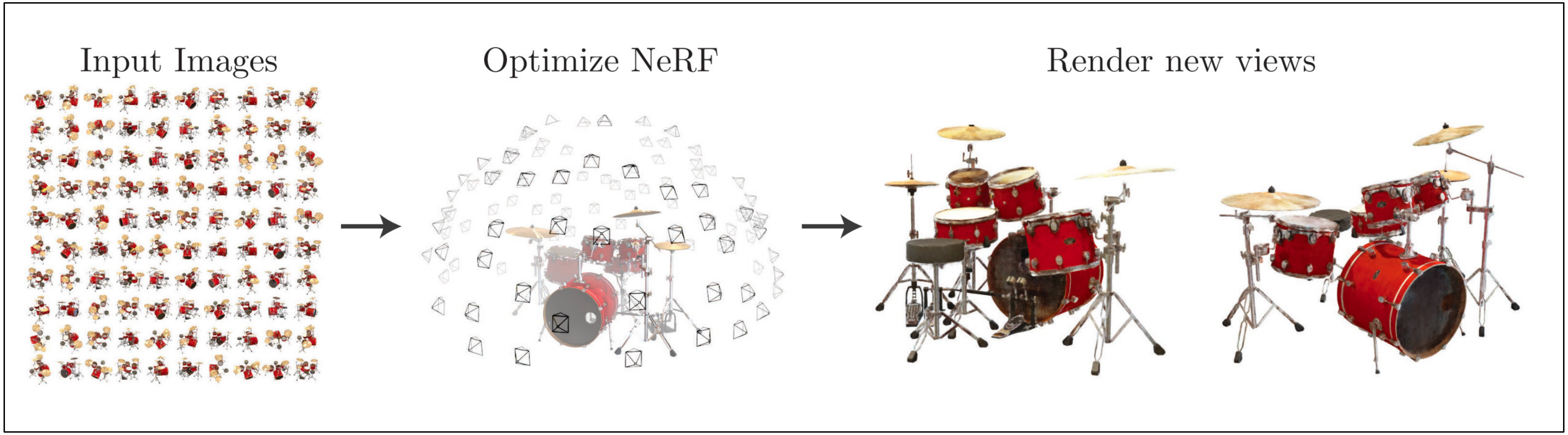
- The Hong Kong University of Science and Technology (Guangzhou)
- Ph.D. student in AI, advised by Prof. Ying-Cong Chen. 2022 - present

- Dalian University of Technology
- B.Sc. in Mathematics. 2018 – 2022

- I previously interned at NIO and MEGVII Technology.

NeRF: represent 3D scenes as neural nets

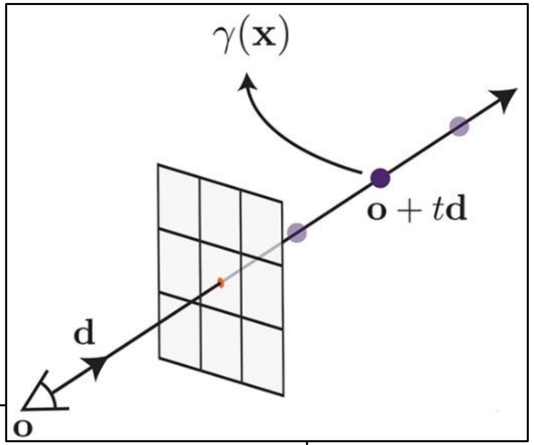
- Input: multi view images, intrinsic and extrinsic
- Training: optimize a MLP to fit the scene
- Inference: query the MLP to render novel view images
- Objective: PSNR, SSIM. Measure the image similarity

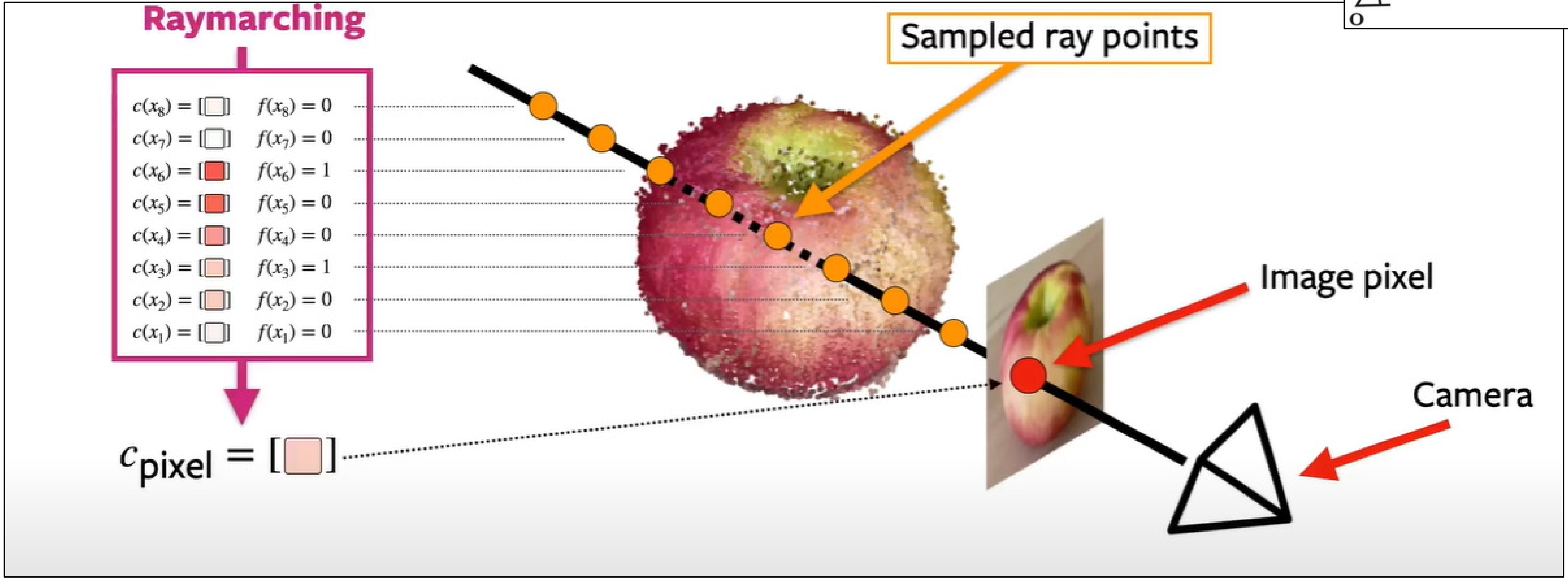


NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

NeRF: represent 3D scenes as neural nets

- Implicit neural representation: $(x, y, z, \theta, \phi) \rightarrow F_{\Omega} \rightarrow (r, g, b, \sigma)$





NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

NeRF: represent 3D scenes as neural nets

- Implicit neural representation: $(x, y, z, \theta, \phi) \xrightarrow{F_\Omega} (r, g, b, \sigma)$

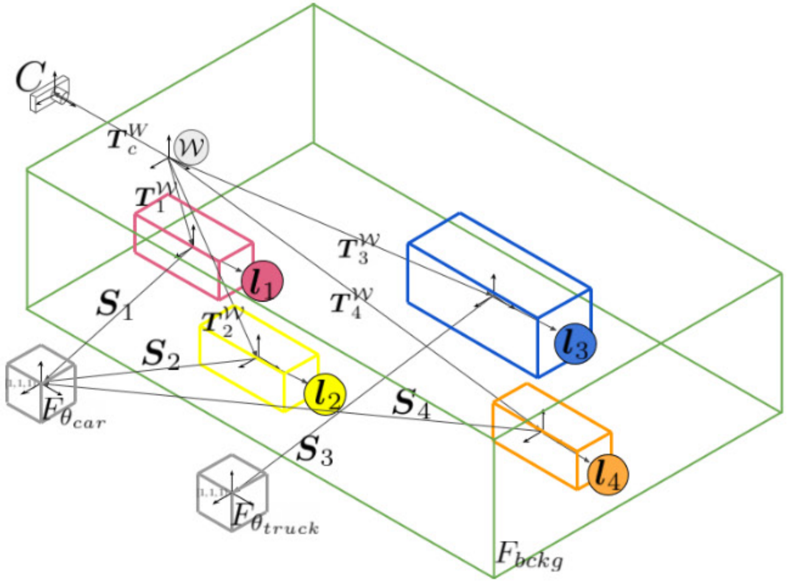
NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis, ECCV 2020

Applications of NeRF in autonomous driving

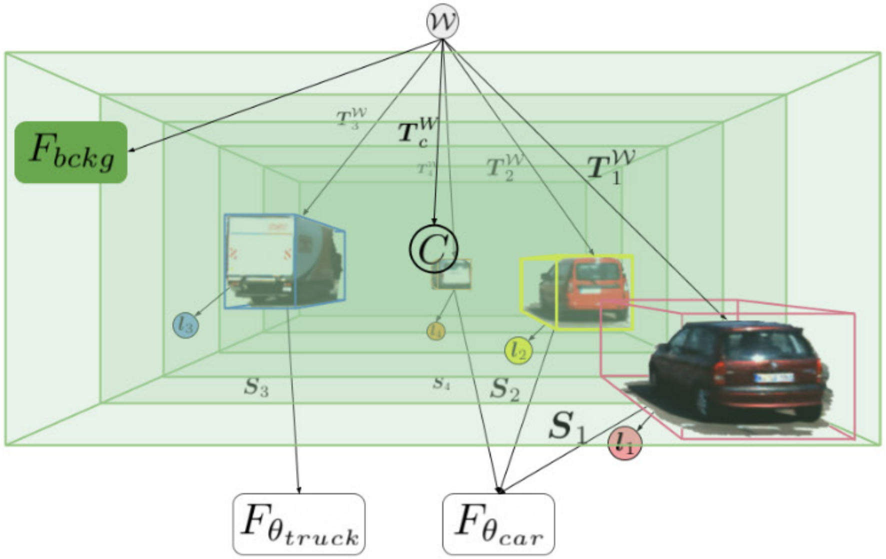
- Motivation:
- Generate free training data by AIGC (GAN, NeRF, diffusion…)
- Provide realistic evaluation and simulation

- Advantage:
- 1. No need for human annotation
- 2. Controllable (6D pose, lighting), easy to create long-tail scenes / corner cases
- 3. Nearly the same distribution with real world data, thus no need for domain adaptation
- 4. Photorealistic appearance compared with graphic engine (Unreal …)

Neural Scene Graphs for Dynamic Scenes



(a) Neural scene graph in isometric view.

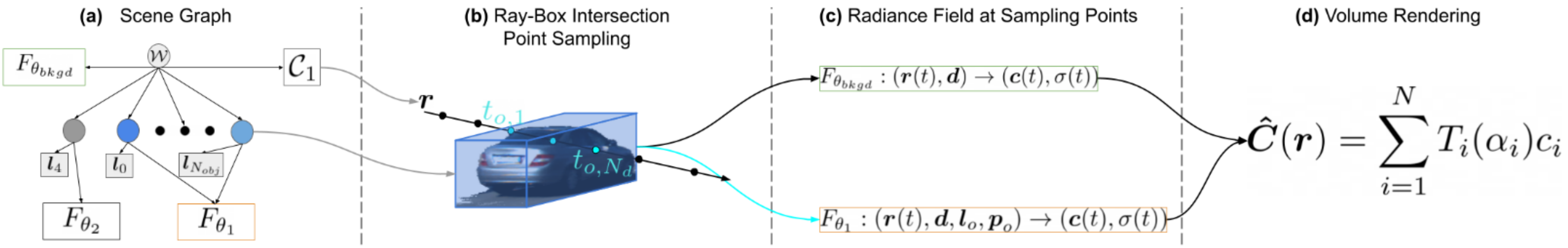


(b) Neural scene graph from the ego-vehicle view.

- NSG provides the first exploration of NeRF in driving scenes.
- NSG disentangle dynamic objects and static background by explicit 3D boxes.
- The sequential 3D boxes are obtained from GT or detection+tracking

Neural Scene Graphs for Dynamic Scenes, CVPR 2021

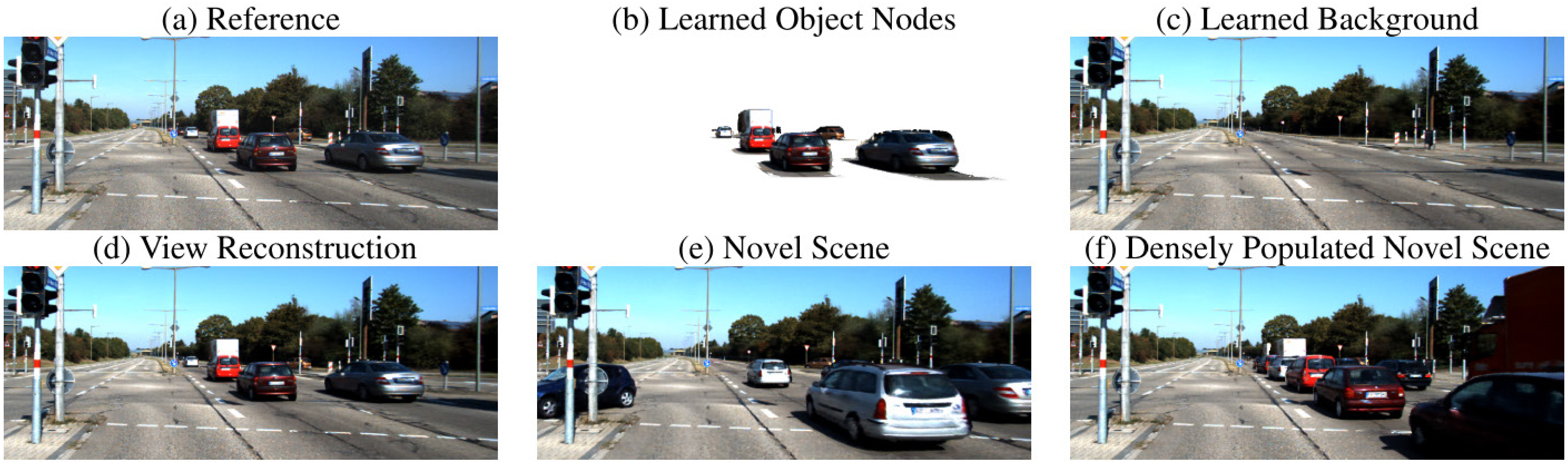
Neural Scene Graphs for Dynamic Scenes



- Learning paradigm:
- Each ray is assigned to a specific object or background by ray-box intersection.
- The sampling points are restricted to the 3D box
- Volume rendering and compute loss

Neural Scene Graphs for Dynamic Scenes, CVPR 2021

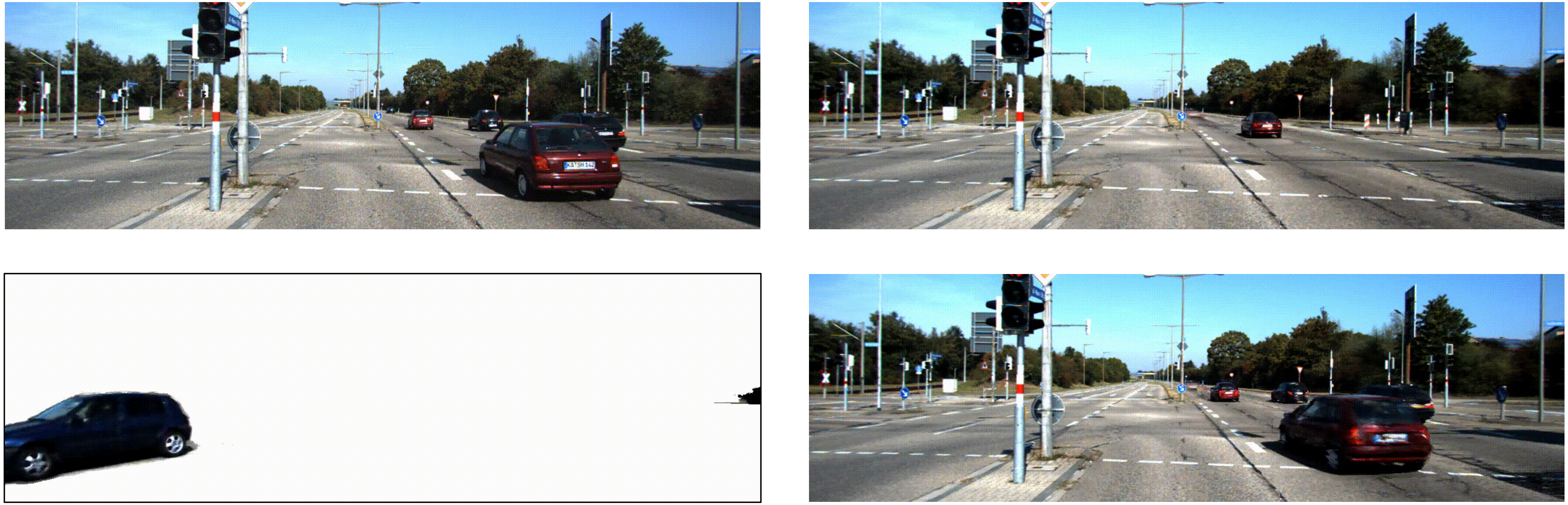
Neural Scene Graphs for Dynamic Scenes



- Application:
- 1. foreground and background disentanglement
- 2. object pose and camera pose manipulation

Neural Scene Graphs for Dynamic Scenes, CVPR 2021

Neural Scene Graphs for Dynamic Scenes



- NSG can control 6D pose of each object by changing the 3D box layout
- The 3D box layout is descripted by rotation and location of object in each frame

Neural Scene Graphs for Dynamic Scenes, CVPR 2021

Block NeRF

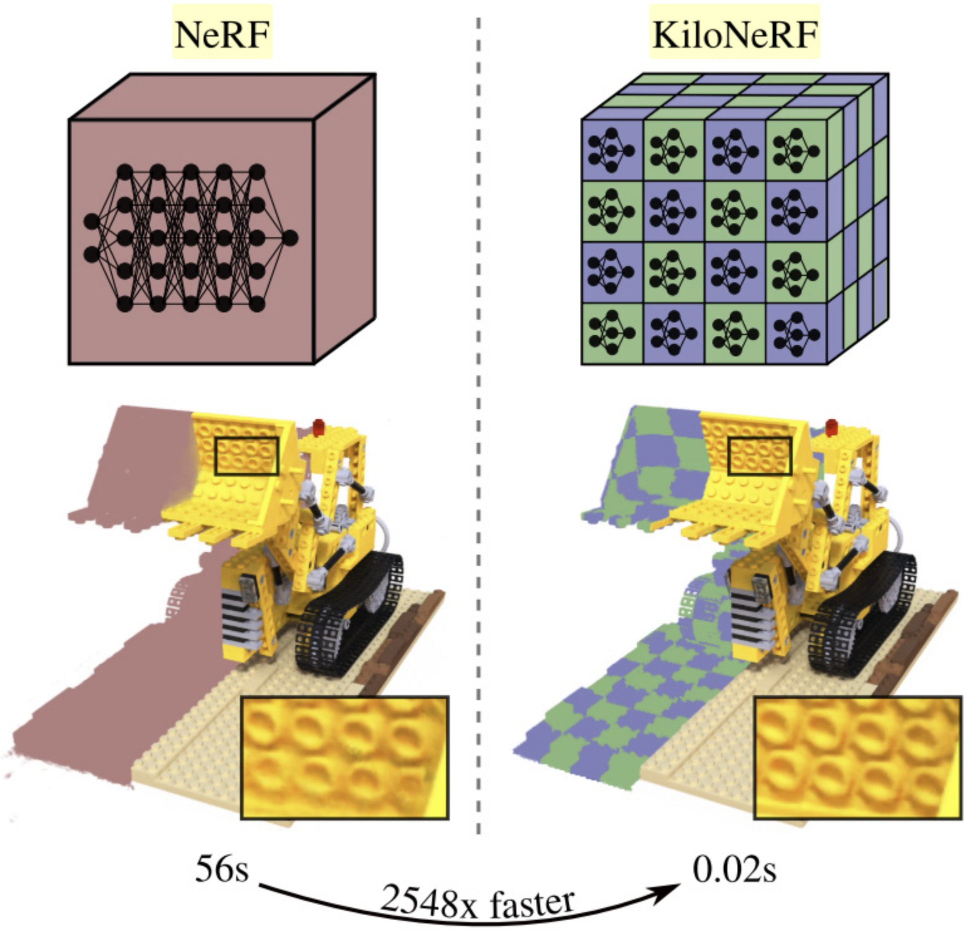
- Scale NeRF to city level.
- Divided the whole dataset into multiple blocks, then use multiple NeRF to reconstruct the whole scene.
- Limits: Block NeRF can only reconstruct static scenes. Dynamic objects are filtered by segmentation mask.

Block-NeRF: Scalable Large Scene Neural View Synthesis, CVPR 2022

Block NeRF

- The scaling issue:
- Single MLP does not have the capacity to reconstruct a large scene.

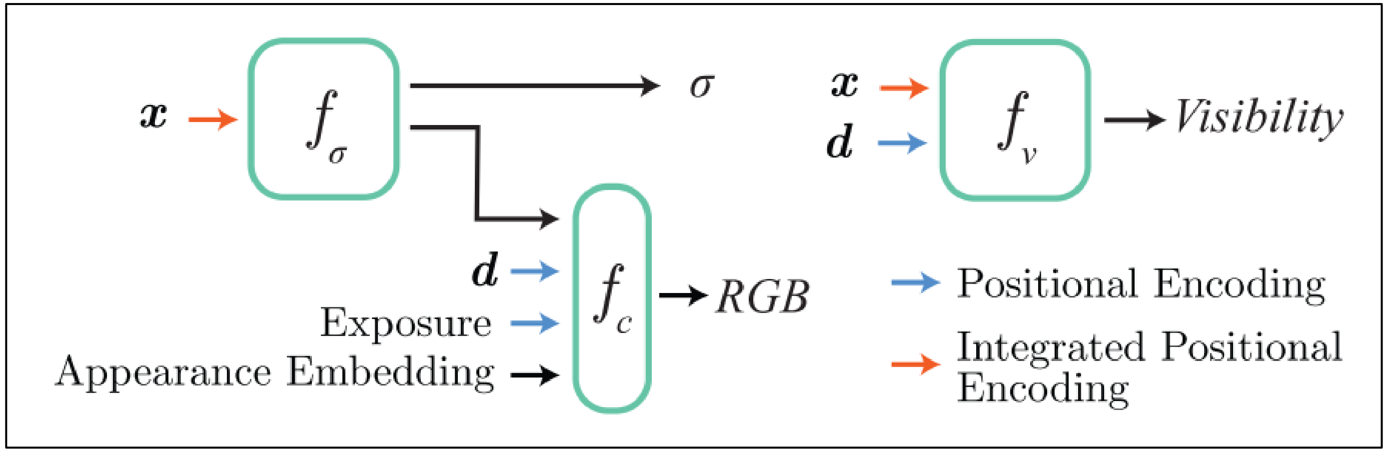
- Solution:
- Split the whole scene into regular grids in 3D space. Each grid is modeled by a specific MLP.



KiloNeRF: Speeding up Neural Radiance Fields with Thousands of Tiny MLPs, ICCV 2021

Block NeRF

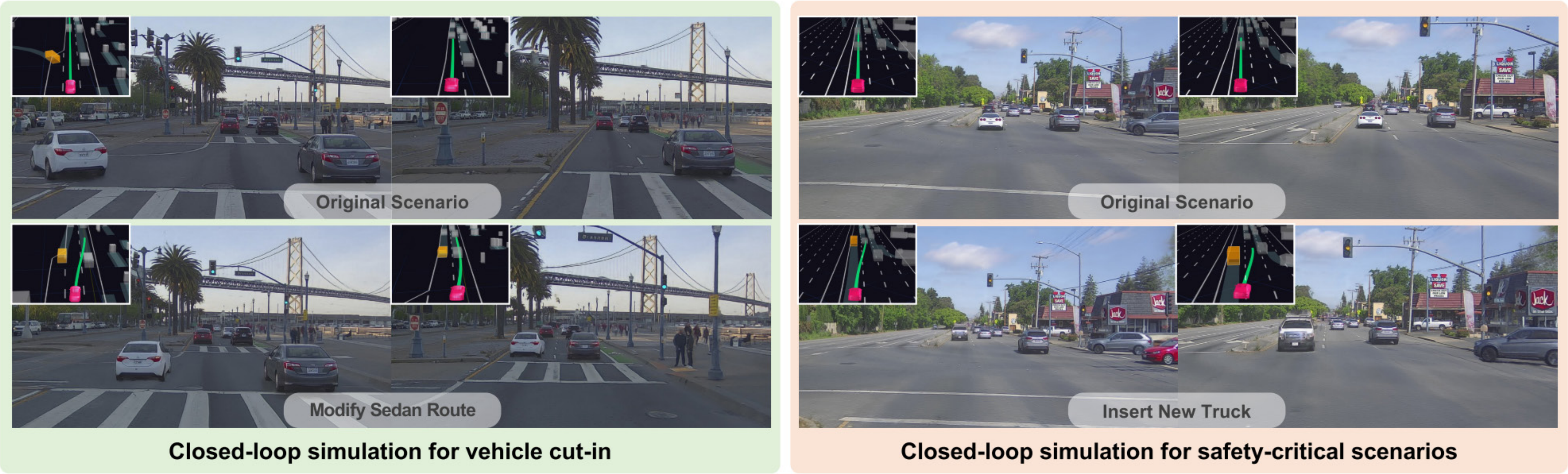
- Challenge: lighting variation and time variation
- Solution: using conditional learnable embedding to learn final RGB



Block-NeRF: Scalable Large Scene Neural View Synthesis, CVPR 2022

UniSim: Closed-Loop Sensor Simulator

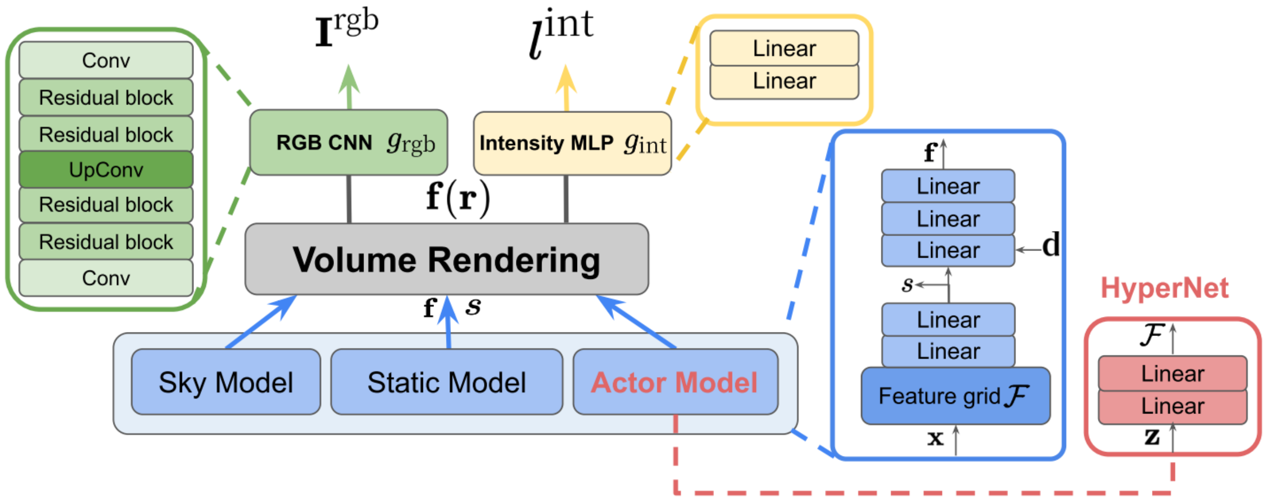
- An extension to NSG
- Sensor simulation: camera images and lidar point cloud
- UniSim provide a test bed for autonomous driving algorithm

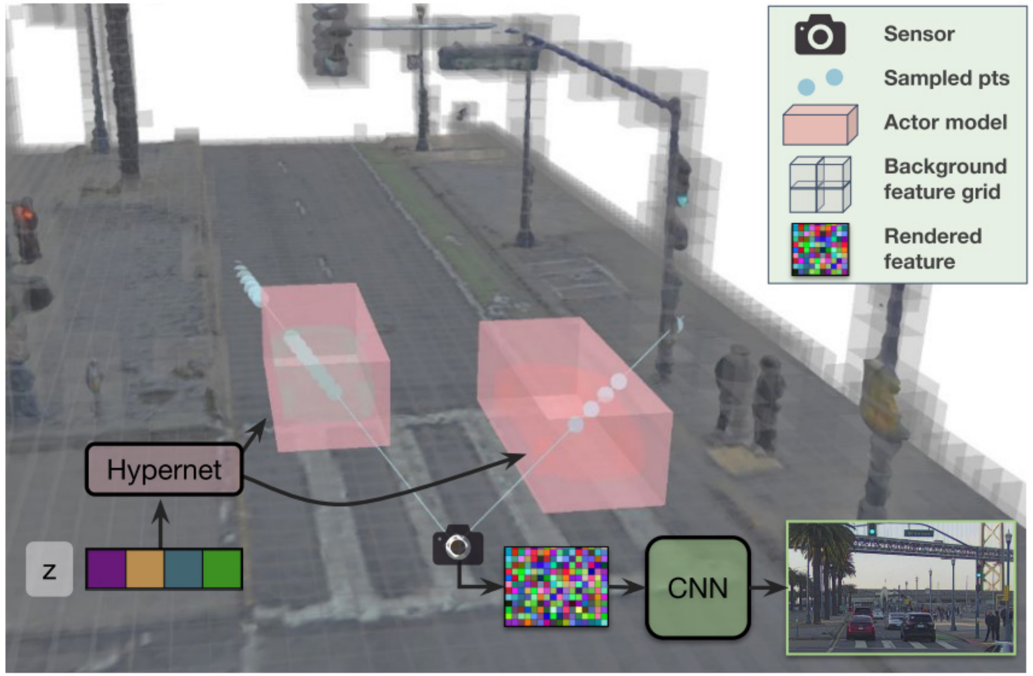


UniSim: A Neural Closed-Loop Sensor Simulator, CVPR 2023

UniSim: Closed-Loop Sensor Simulator

- Build upon advances in NeRF:
- 1. grid-based feature *vs* pure MLP
- 2. occupancy grid sampling *vs* two stage sampling





UniSim: A Neural Closed-Loop Sensor Simulator, CVPR 2023

UniSim: Closed-Loop Sensor Simulator

- Build upon advances in NeRF:
- 1. grid-based feature *vs* pure MLP
- 2. occupancy grid sampling *vs* two stage sampling

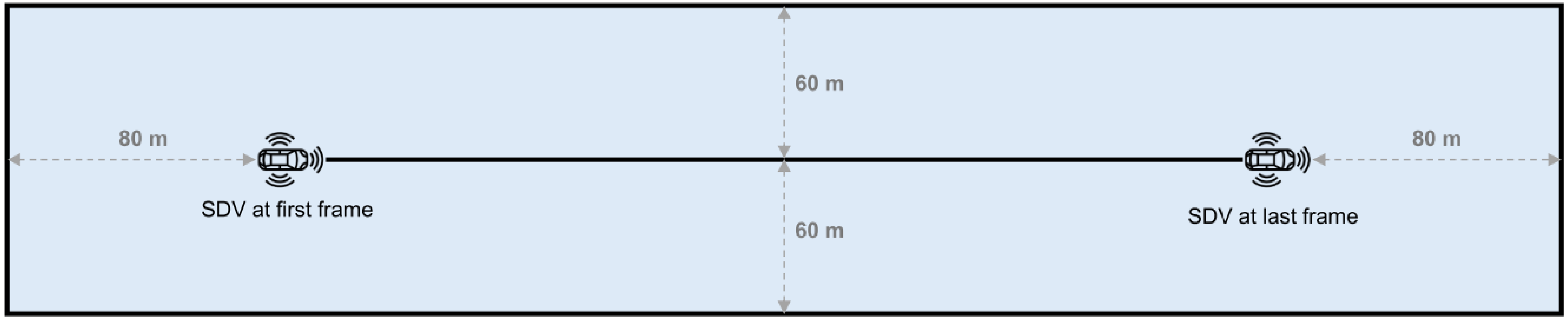


Figure 2. **Region of interest of our scene representation.**

UniSim: A Neural Closed-Loop Sensor Simulator, CVPR 2023

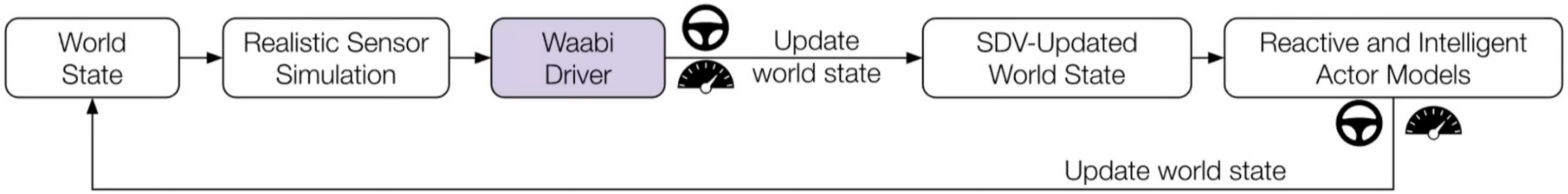
UniSim: Closed-Loop Sensor Simulator

Waabi World

Waabi World Engine

Close loop simulator

- Like playing a video game: every action has a reaction
- Truly experience how the scenario would play out if it were in the real world
    1. **Immersive** – need for sensor simulation (e.g.. camera, lidar)
    2. **Reactive** – the SDV reacts to the actors and the actors to the SDV
    3. **Diversity** of the real world in both behavior and appearance
    4. **Scale:** need to be efficient
- Evaluator that can automatically assess the driving skills



waabi

17

https://www.youtube.com/watch?v=0RjF9xbkiAY&t=928s, CVPR 2023 workshop

Our work: use NeRF to synthesize training data

Lift3D: Synthesize 3D Training Data by Lifting 2D GAN to 3D Generative Radiance Field

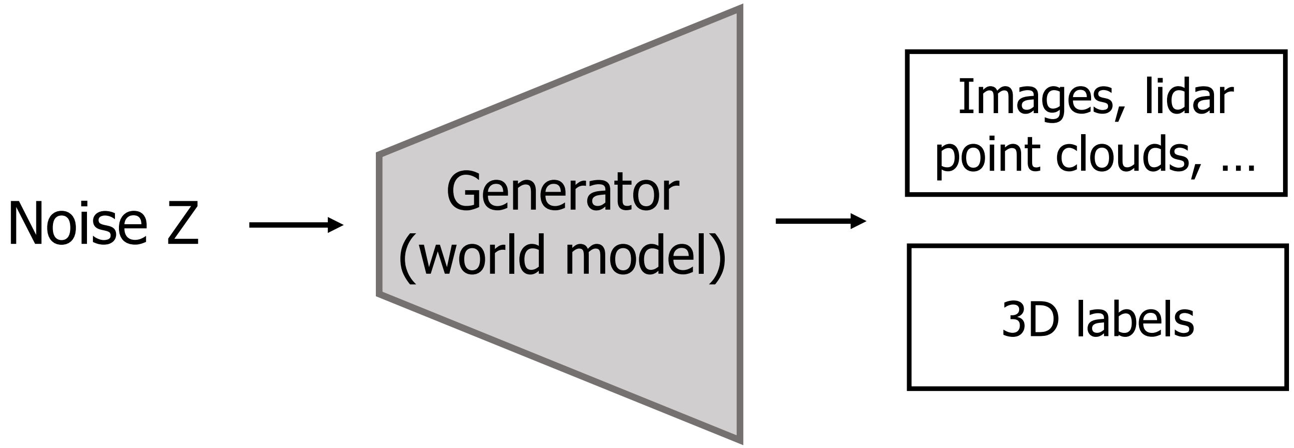
Leheng Li[1], Qing Lian[2], Luozhou Wang[1], Ningning Ma[3], Ying—Cong Chen[1,2]

[1]HKUST(GZ), [2]HKUST

[3]NIO

Project page: https://len-li.github.io/lift3d-web/

Imagine there is an AIGC algorithm that generate training data for free



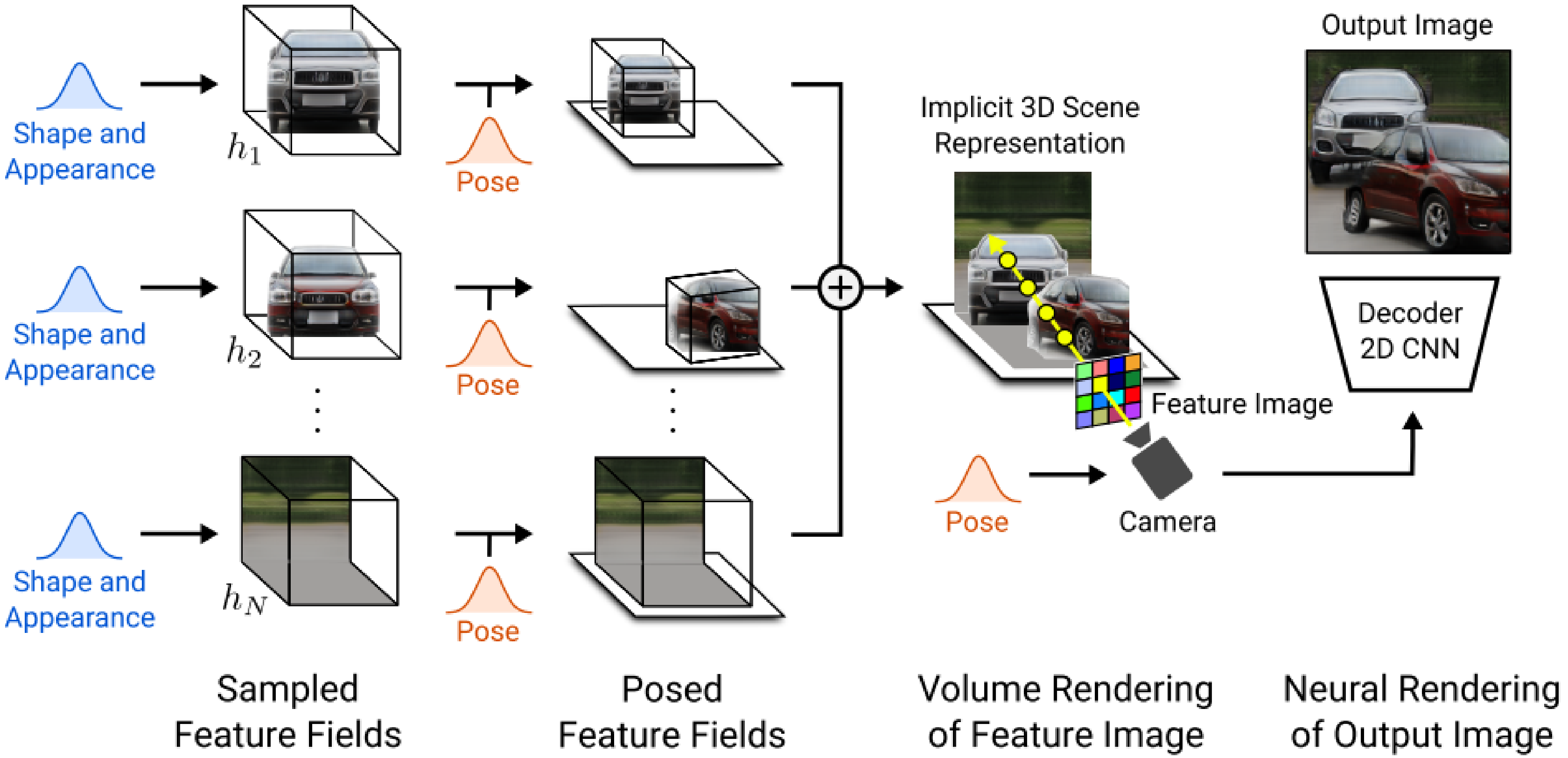
Evaluation setting: data augmentation

- A pure generative model is hard to guarantee the data distribution with real world data
- We instead evaluate the generated data by its benefit of data augmentation.

Lift3D: Synthesize 3D Training Data by Lifting 2D GAN to 3D Generative Radiance Field, CVPR 2023

Baseline: GIRAFFE (CVPR 2021 best paper)

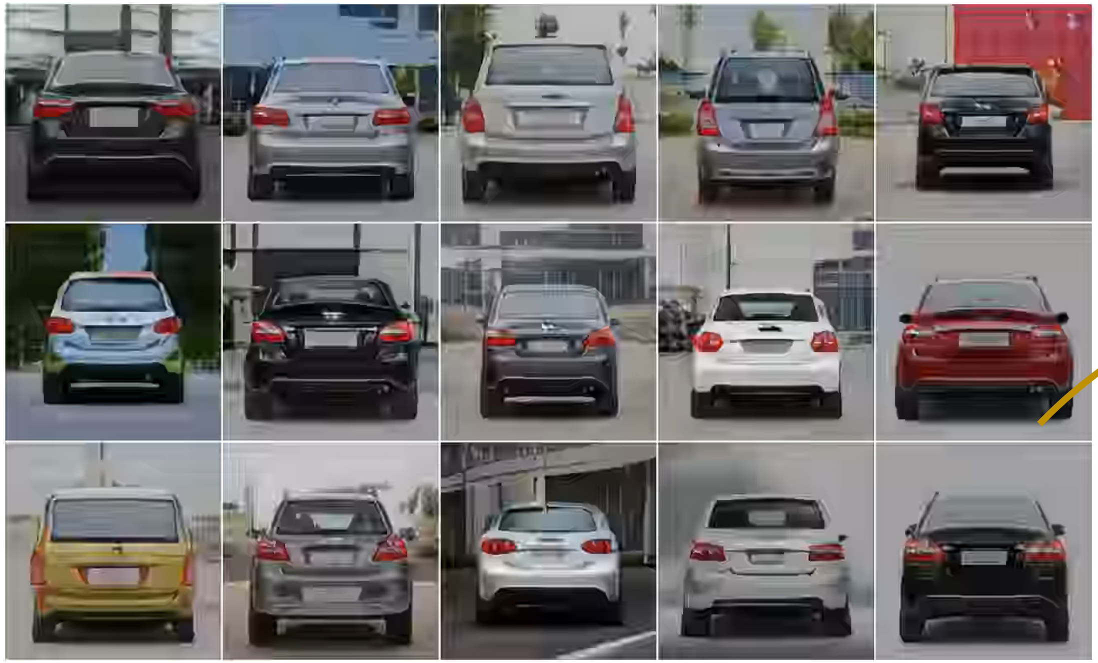
- Method: NeRF + GAN



GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields

Use GIRAFFE to augment existing dataset

- Generate new objects and add them to existing scenes



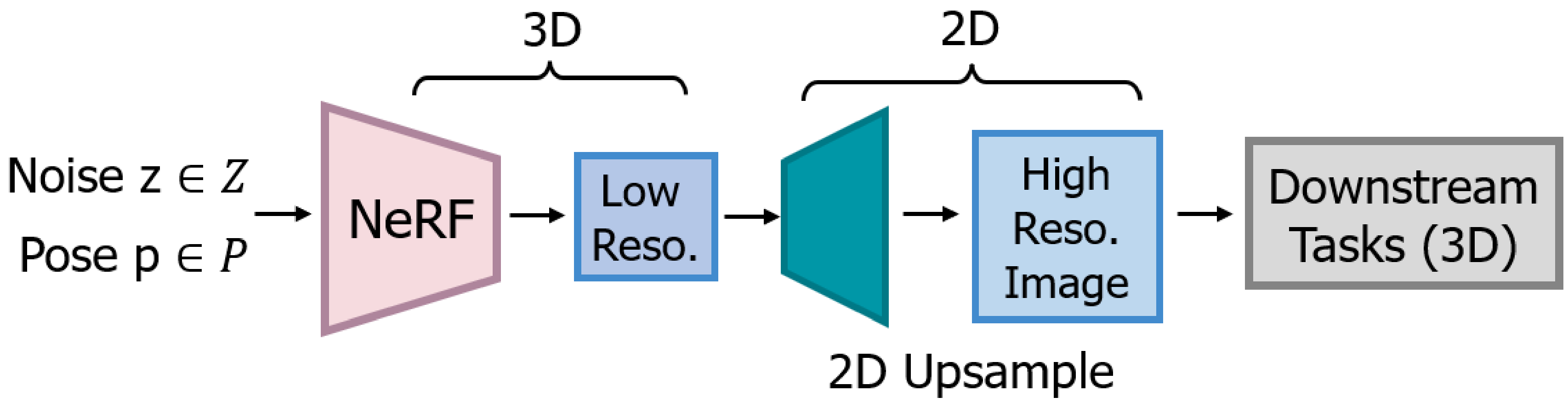
Generated objects

Add objects

nuScenes dataset

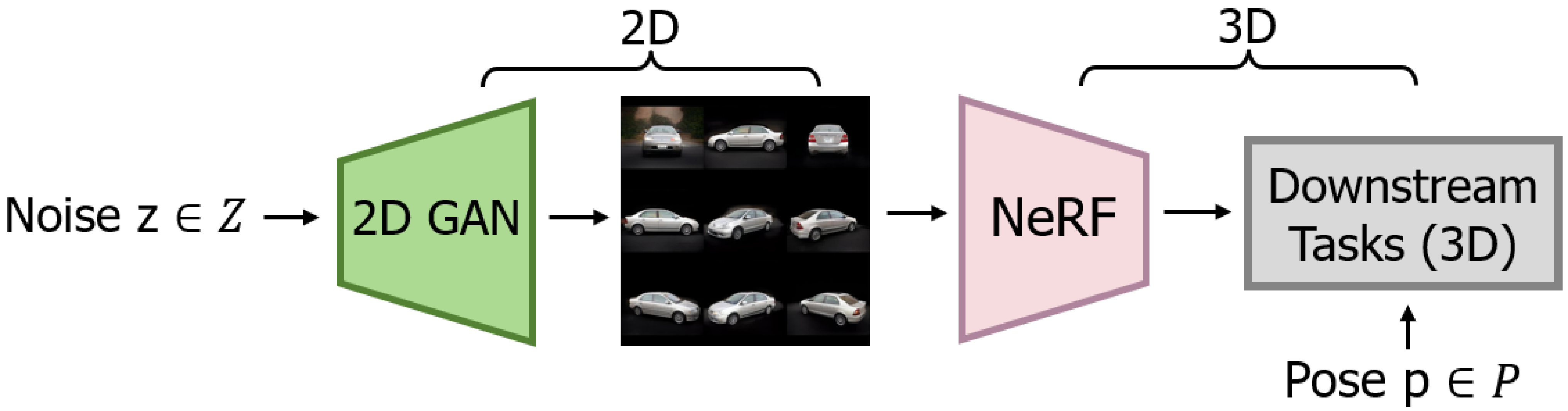
Why previous work fall short of 3D consistent generation?

- Due to sample efficiency, NeRF-based GAN typically adopt a two stage pipeline:
- 1. use volume render to generate the low resolution feature.
- 2. upsample the feature to the final image by 2D upsampler.
- Empirical results show that this pipeline does not strictly preserve 3D consistent synthesis due to 2D upsampler.



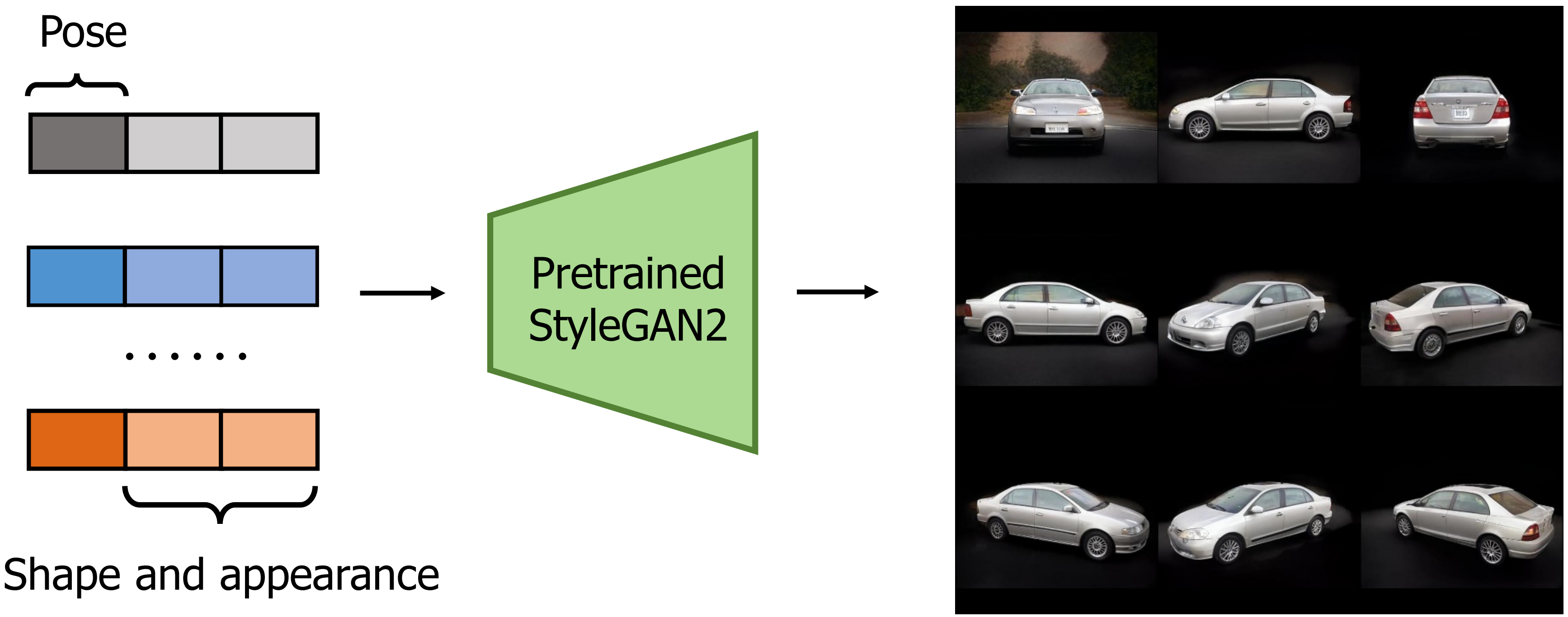
How to escape the computational bottleneck?

- Our method: Disentangle the 2D-3D generation.
- 2D GAN: provide photorealistic image synthesis, NeRF: provide 3D synthesis
- Without relying on fixed-resolution 2D upsampler, Lift3D perform strict 3D consistent synthesis that generalize to any camera parameters.



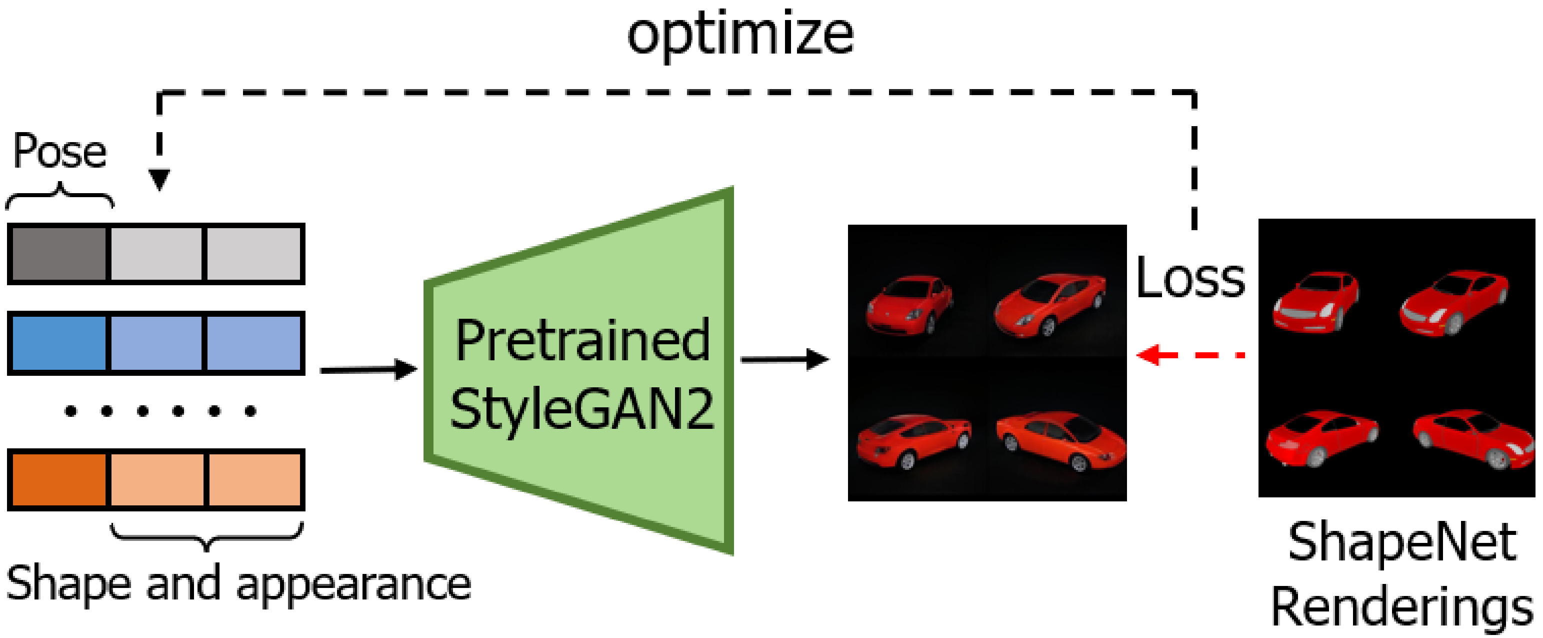
Two stage pipeline

- First stage: use StyleGAN2 to generate multi-view images
- StyleGAN2 provides photorealistic synthesis with rough 3D controllability
- Disentangled 2D GANs allow us to generate images with <u>3D pose label</u>



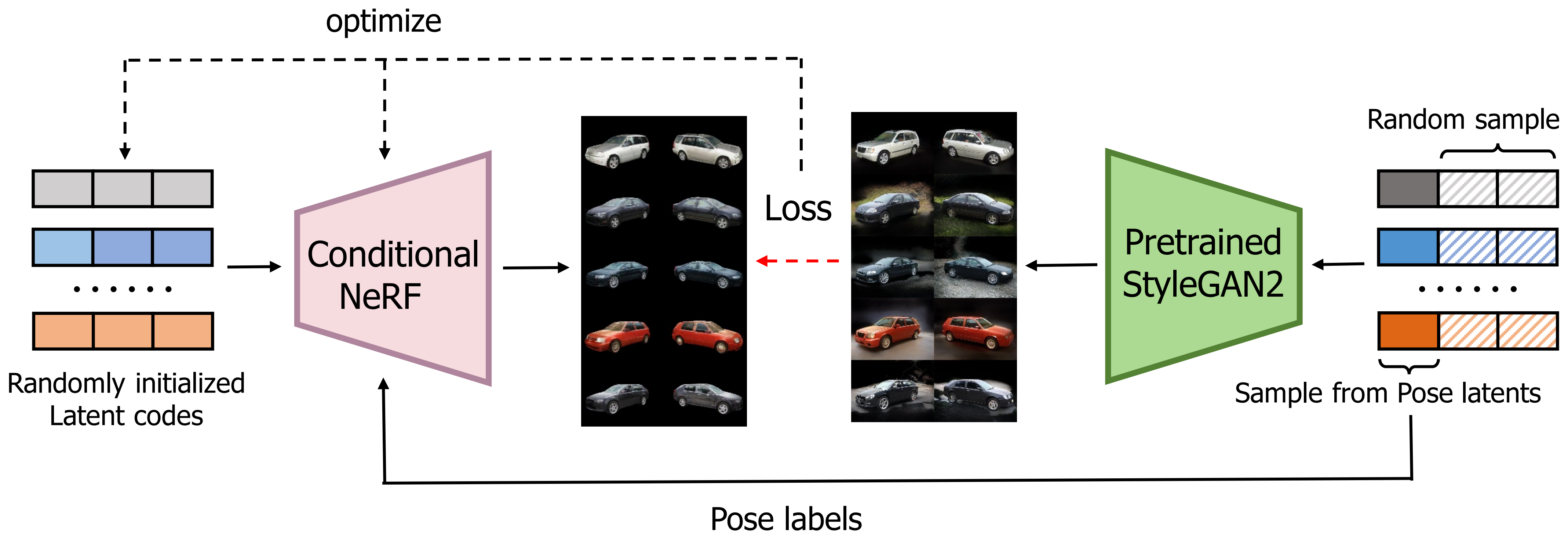
Two stage pipeline

- First stage: use StyleGAN2 to generate multi-view images
- Use synthetic data to automatically find pose label



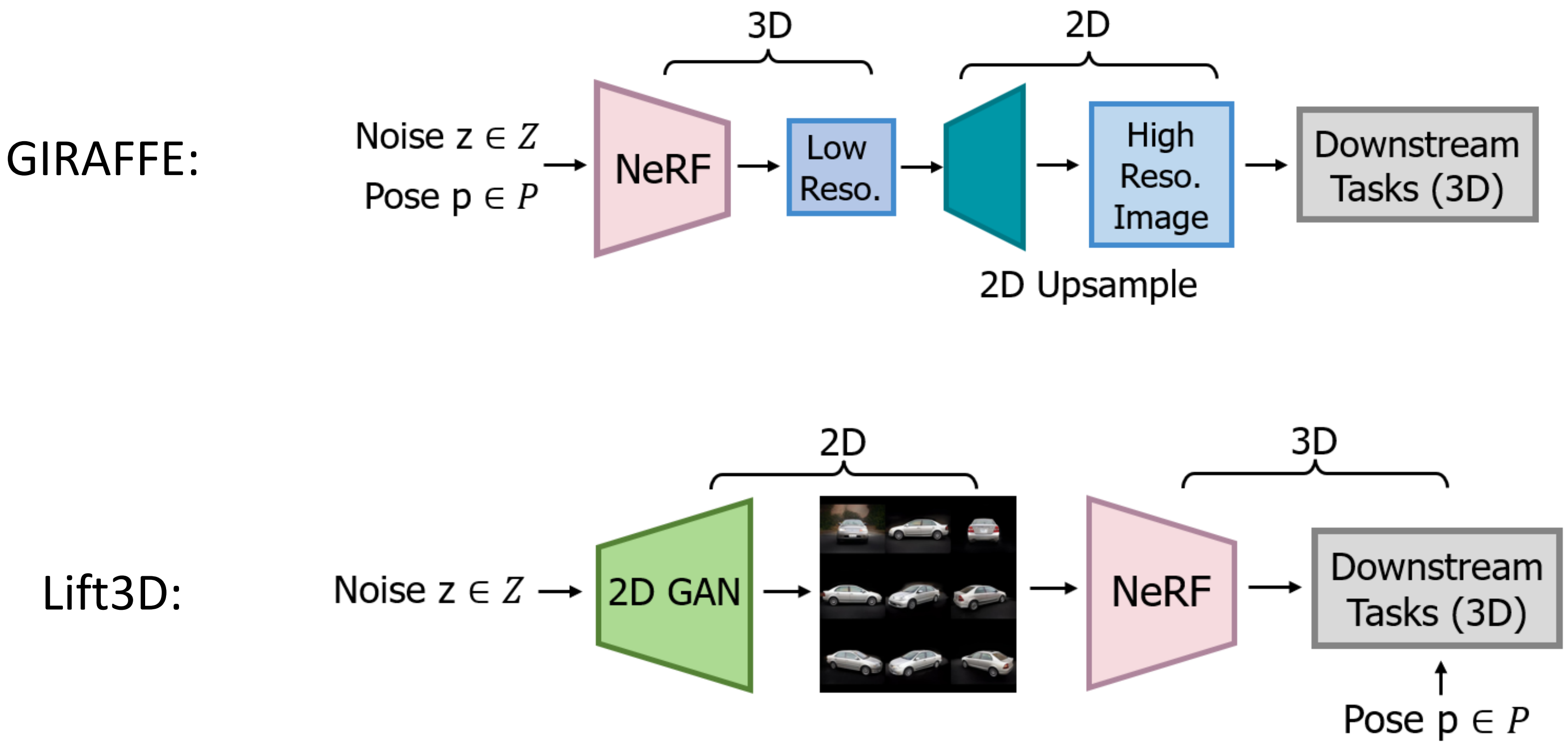
Two stage pipeline

- Second stage: lift multi-view images to 3D NeRF.
- All instances share the same NeRF network to encode prior.



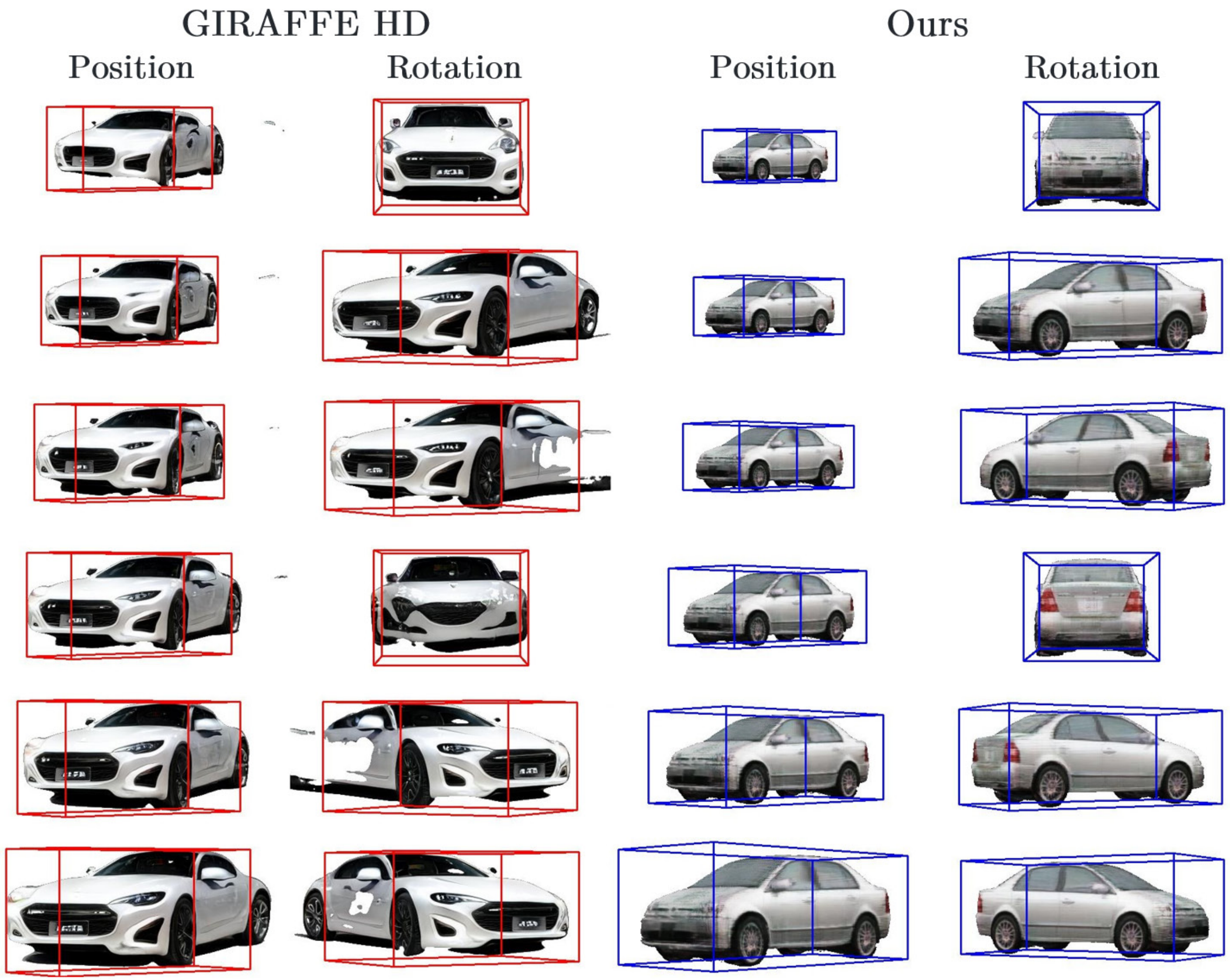
Mechanism

- Lift3D disentangles 3D generation from image synthesis
- Output image rendered by NeRF thus is strictly 3D consistent



Results

- Visualization of multi-view synthesis with plotted 3D box



Results

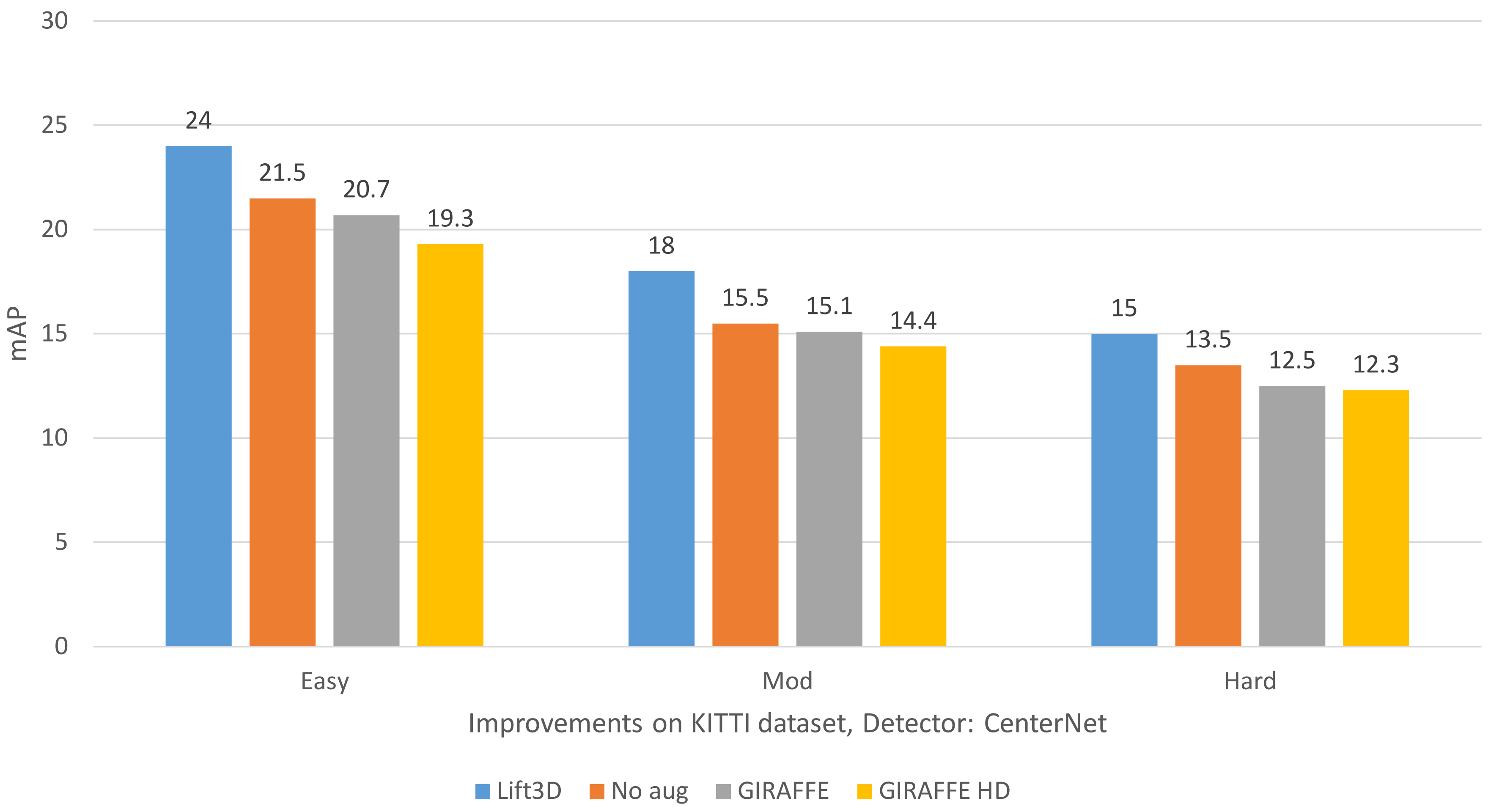
- Visualization result of augmentation

Original Dataset

Augmented Dataset

Results

- We display improvement of 3D detection accuracy on KITTI dataset



Summary

- Disentangled 3D generation provides tight 3D annotation
- Lift3D can synthesize images in any resolution by accumulating single-ray evaluation
- Without any domain adaptation, the generated data improves downstream task performance
- Achieve good qualitative and quantitative results

Thanks for listening!