GeoSim: Realistic Video Simulation via Geometry-Aware Composition for Self-Driving

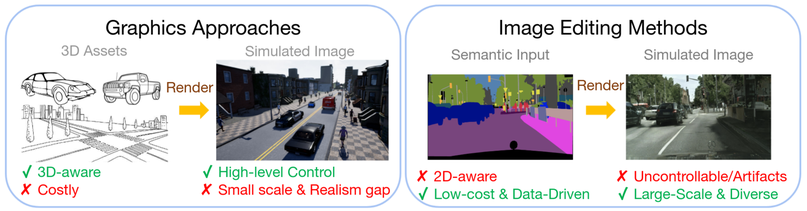
Yun Chen*, Frieda Rong*, Shivam Duggal*, Shenlong Wang, Xinchen Yan,
Sivabalan Manivasagam, Shangjie Xue, Ersin Yumer, Raquel Urtasun

Uber ATG · University of Toronto · MIT · Stanford University

Introduction

Scalable camera simulation is critical for validating self-driving systems because it is costly and risky to test it in the real-world.

Existing methods can be divided into 2 paradigms. But they either lack scalability, realism or 3D-awareness.

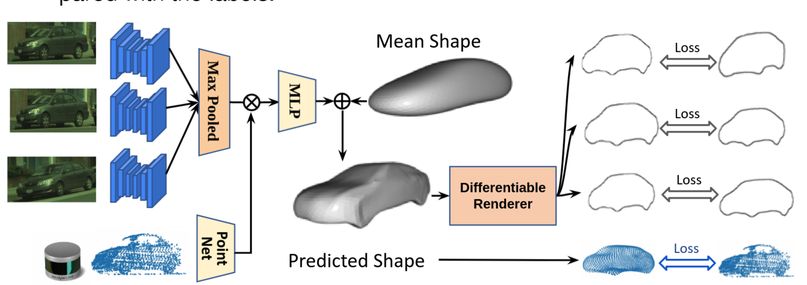


GeoSim combines graphics and image-editing methods to get the best of both worlds. GeoSim reconstructs 3D assets and then geometrically simulating the new objects in video.

Asset Creation

GeoSim uses real world data to automatically build vehicle asset bank from the wild without 3D groundtruth.

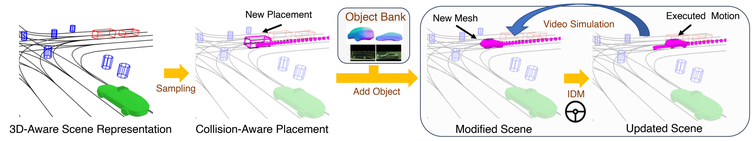
1. Multi-sensor feature is extracted from camera and LiDAR.
2. Per-vertex deformation is applied on a learnable mean-shape to predict shape.
3. Silhouette is differentiably rendered from the predicted mesh and compared with the labels.



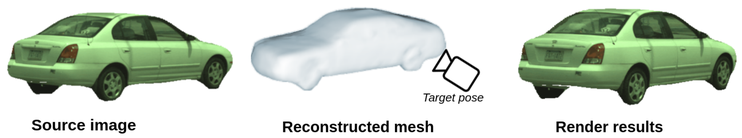
Simulation Pipeline

With the reconstructed asset bank, GeoSim can automatically simulate new objects in videos. The approach is scalable and the results are realistic and geometry-consistent.

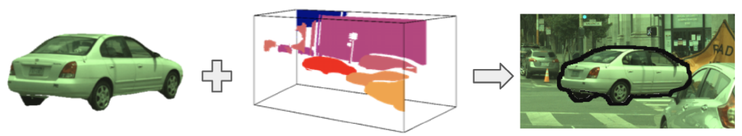
1. **Scenario Generation:** (1) 3D-aware placement using HD Map, (2) asset selection for rendering, (3) trajectory simulation for video generation.



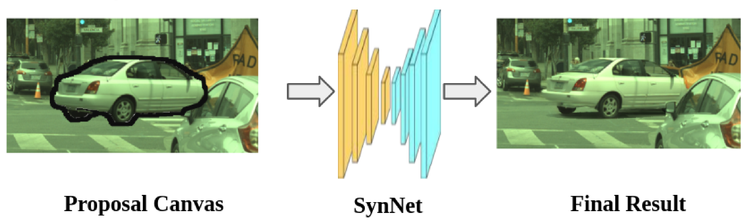
2. Using the reconstructed mesh as correspondence, the selected asset can be warped to the target pose.



3. The dense depth is applied for **occlusion reasoning** between the rendered object and the background.



4. **Post-composition refinement** is applied to naturally blend new objects into the target scene.



Experiments

- Qualitative comparison of image simulation approaches on UrbanData.

- Quantitative comparison (left) and ablation on rendering approaches (right). *HS*: human score (% of participants who prefer our GeoSim results over baseline)

Method	HS(%)	FID ↓
Guided Editing	94.3	20.3
Cut-Paste	98.5	22.1
CAD	94.3	17.3
GeoSim	-	**14.3**

Approach	Shadow	HS (%)	FID ↓
Physics	Yes	94.2	17.3
2D Synthesis	-	75.7	**13.7**
Geo Synthesis	No	71.9	**13.7**
Geo Synthesis	Yes	-	14.3

- Qualitative on Argoverse and video results on UrbanData.

Video results in 4K
tinyurl.com/cvpr2021

- **Sim2Real:** Augmenting labeled real data with GeoSim results leads to consistent improvement in segmentation.