From Model-Based to Data-Driven Simulation: Challenges and Trends in Autonomous Driving

Karlsruher Institut für Technologie

FZI

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Ferdinand Mütsch[1]*, Helen Gremmelmaier[2]*, Nicolas Becker[1]*, Daniel Bogdoll[12], Marc René Zofka[2], J. Marius Zöllner[12]
[1]KIT Karlsruhe Institute of Technology, Karlsruhe, Germany,
[2]FZI Research Center for Information Technology, Karlsruhe, Germany
*These authors contributed equally

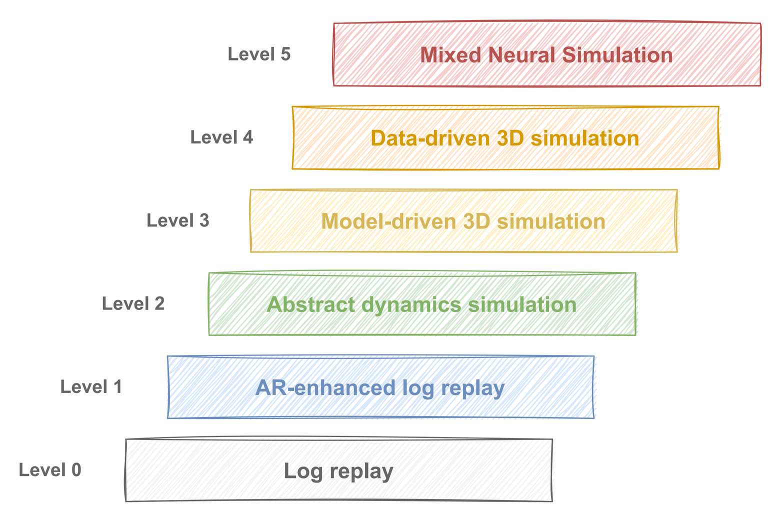
Where is AD Simulation Heading Towards?

- Simulation is crucial for development and testing of autonomous vehicles (AV). Where are we now? Where are we going?
- We provide an overview of current challenges and recent trends

Figure 1. Examples of our proposed simulation levels. Top-left to bottom-right: AR-enhanced (level 1), SUMO (level 2), CARLA (level 3), Block-NeRF (level 4)

- In addition, we present a classification scheme for simulation approaches

Levels of Simulation Approaches



- Categorization of simulations by comprehensives, realism, capabilities, etc.
- Higher level ≈ more "powerful" simulators
- Levels 0 to 3 well adopted and widely used
- Levels 4 and 5 of primary interest in research lately (e.g. [1, 2])

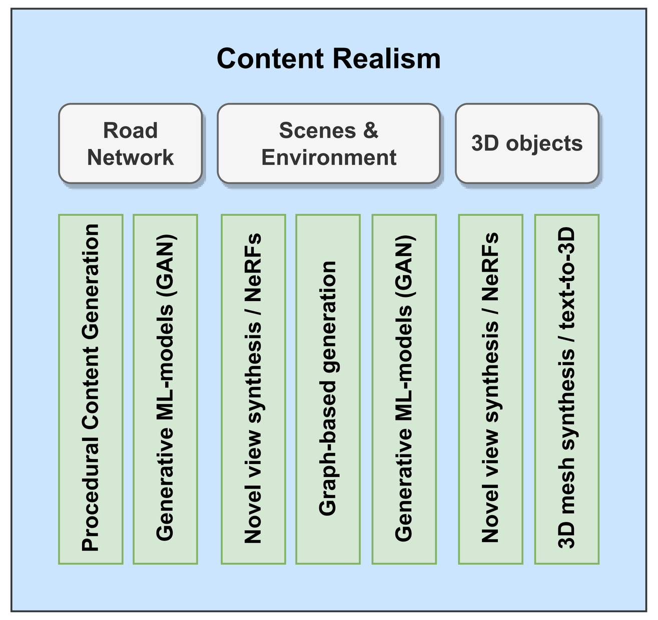
	Closed-loop / reactive	End-to-end development & testing	Visual fidelity	Diversity (content & behavior)	Object rep-resentation	Scalability	Control-lability	Key use cases	Examples
0. Log replay	✗	✗	high	low	implicit	very low	none	Perception	-
1. AR-enhanced log replay	partly	✗	mixed	medium	mixed	low	low	Perception	[29, 67, 69, 71]
2. Abstract dynamics simulation	✓	✗	-	medium	explicit	medium	high	Prediction, planning, control	SUMO [7], CityFlow [65]
3. Model-driven 3D simulation	✓	✓	low - high	medium	explicit	medium	high	E2E training / testing	CARLA [13], DriveSim [42]
4. Data-driven 3D simulation	✓	✓	medium - high	medium - high	implicit	high	low - medium	E2E training / testing	ViSTA 2.0 [3], Drive-GAN [27]
5. Mixed neural simulation	✓	✓	medium - high	high	mixed	high	medium - high	E2E training / testing	-

Table 1. Categories of Simulation Approaches for AD

[1] Alexander Amini et al. - Vista 2.0: An open, data-driven simulator for multimodal sensing and policy learning for autonomous vehicles.In IEEE ICRA, 2022.
[2] Seung Wook Kim et al. - DriveGAN: Towards a Controllable High-Quality Neural Simulation. In IEEE/CVF CVPR 2021.

Challenges & Trends

Content Realism



- Is about accurately representing real-world objects (static and dynamic) and environments and their diversity (*what's "in" a scene*)

Road Network

- 3D mapping or modeling in CAD is costly and lacks scalability
- Navigation maps are not suitable for autonomous driving

➞ Traditional procedural content generation (PCG) from gaming, GANs for road graph generation, adversarial RL for PCG (e.g. [3])

Scenes & Environment

- Lack in (visual) fidelity / photo-realism of objects or entire 3D-rendered scenes
- Hand-crafting scenarios for level 2 / 3 simulations is time-consuming and lacks scalability
- Lack in diversity and / or realism of scene topologies
- Limited controllability over generated content (with levels 4 / 5)

➞ GANs for novel view-point synthesis (NVP), large-scale NeRFs for NVP, ML-facilitated 2D to 3D to 2D projection for end-to-end simulation, RL or sequence models for scene graph generation (e.g. [1, 4])

3D objects

- Lack in diversity / level of detail of scene objects (e.g. vehicles)
- Hand-crafting object models is time-consuming and lacks diversity

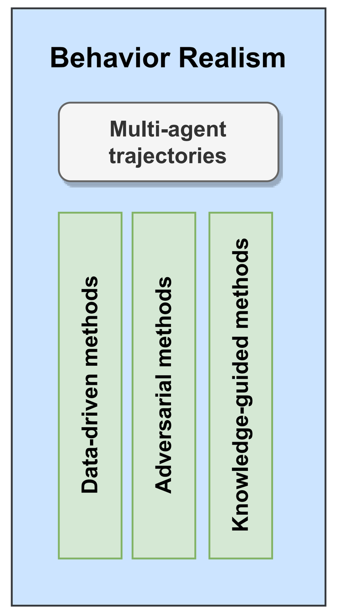
➞ NeRFs for NVP ("scanning" objects), end-to-end 3D mesh generation, text-to-3D models (e.g. [5, 6])

Cross-Cutting Aspects

Standardization | Data & Compute Power | Validity & Transferability

[3] Linus Gisslen et al. - Adversarial Reinforcement Learning for Procedural Content Generation. In 2021 IEEE COG.
[4] Yuanbo Xiangli et al. - BungeeNeRF: Progressive Neural Radiance Field for Extreme Multi-scale Scene Rendering. In The European Conference on Computer Vision (ECCV), 2022.
[5] Jun Gao et al. - GET3D: A Generative Model of High-Quality 3D Textured Shapes Learned from Images. In Advances In Neural Information Processing Systems, 2022.
[6] Chen-Hsuan Lin et al. - Magic3D: High-Resolution Text-to-3D Content Creation. arXiv preprint, 2022.

Behavior Realism



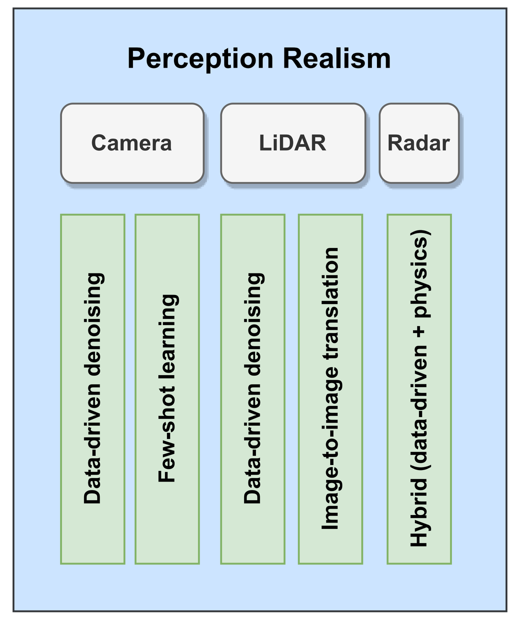
- Covers dynamic aspects of scenes, such as non-static actors' motion characteristics (*what's "happening" in a scene*)
- It is hard to design challenging, yet plausible maneuvers to cover the "long tail of events"

➞ Data-driven methods

➞ Adversarial methods

➞ Knowledge-guided methods

Perception Realism



- Is about replicating the appearance of the real world from the perspective of different sensors (*how a scene "looks like"*)
- Lack in fidelity of camera-, lidar- and radar noise models
- Lack in generalization from sensor-specific characteristics

➞ Deep learning for data-driven noise models, image-to-image translation, "hybrid" approaches (e.g. [7, 8])

In Summary

We observe simulations being pushed heavily towards levels 4 and 5, particularly facilitated by generative ML methods, NeRFs, adversarial RL and others. Heading in this direction, important future research questions include:

1. How to better understand and control data-driven simulations?
2. How to quantify quality & validity of data-driven simulations?
3. How to extract / generate training data at scale for different OODs?
4. How to leverage gestures & facial expressions in AD simulations?

[7] Sivabalan Manivasagam et al. - LiDARsim: Realistic LIDAR Simulation by Leveraging the Real World. In IEEE/CVF CVPR, 2020.
[8] Microsoft Research - Data-driven Sensor Simulation for Realistic LIDARs, 2022.