Orbis: Overcoming Challenges of Long-Horizon Prediction in Driving World Models

Arian Mousakhan*

Sudhanshu Mittal*

Silvio Galesso*

Karim Farid*

Thomas Brox

Main Contributors*

University of Freiburg, Germany
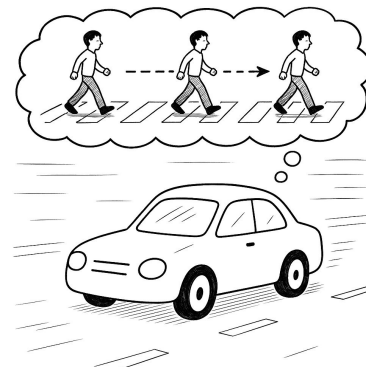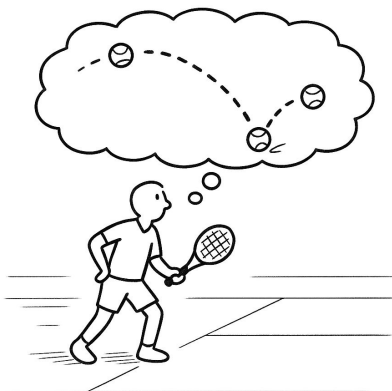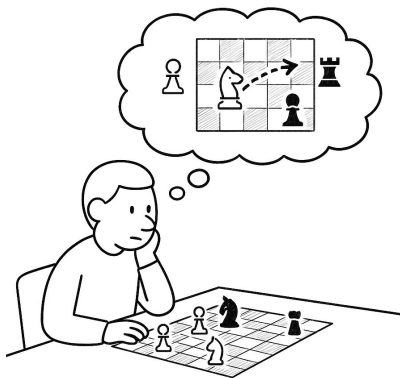
World Model

Learn dynamic representation $\rightarrow$ Predict $\rightarrow$ Simulate actions $\rightarrow$ Plan

universität freiburg

Related Work: Driving World Models

Non-testable:

- GAIA-1 *(Hu et al. 2023)*

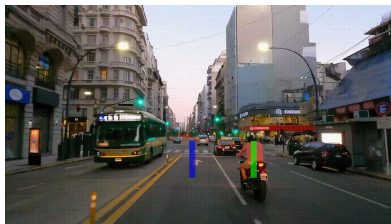- GAIA-2 *(Russell et al. 2025)*

Testable:

- Vista *(Gao et al. 2024)*

- GEM *(Hassan el al. 2024)*

- DrivingWorld *(Hu et al. 2024)*

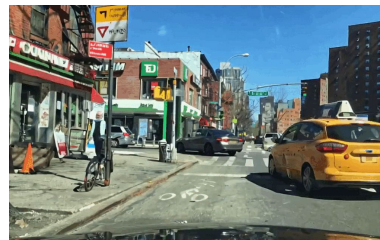- Cosmos *(Nvidia 2025)*

GAIA-1

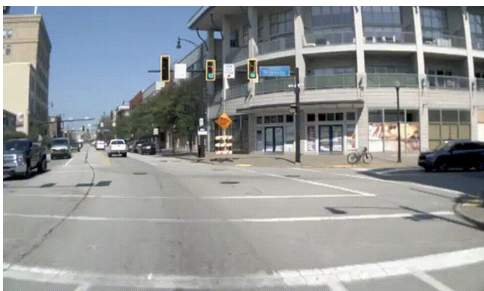Vista

GEM

DrivingWorld

Cosmos

State-of-the-art & Challenges
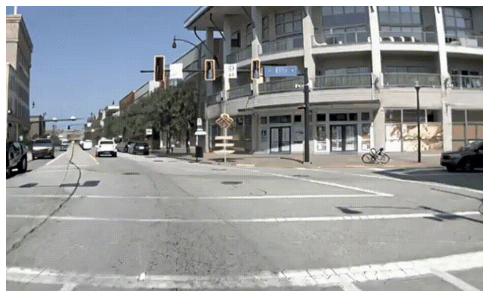
Generating new scenes and content is hard, models struggle with:

- Long-horizon generation
- Turns

GEM

Cosmos

DrivingWorld

universität freiburg
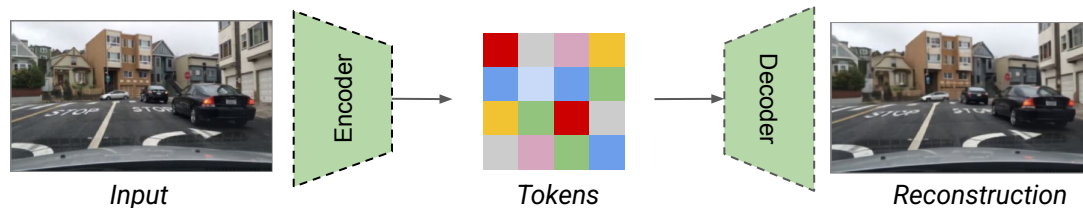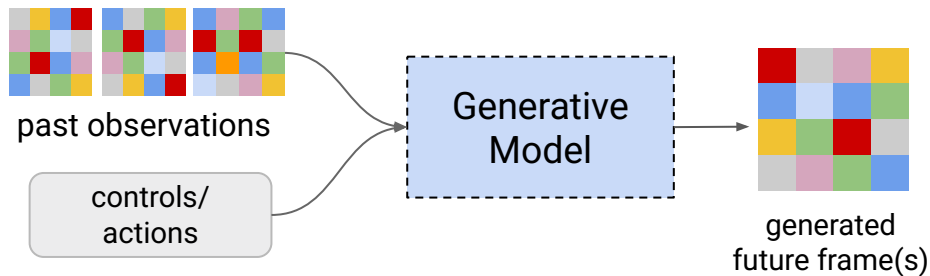
Two stage training

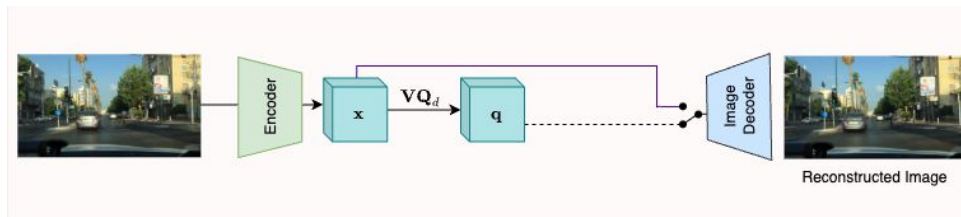- Stage 1: Tokenization

- Stage 2: World modeling

Discrete or continuous latent world model?

Stage-I
Hybrid Tokenizer

Stage-II
World Model
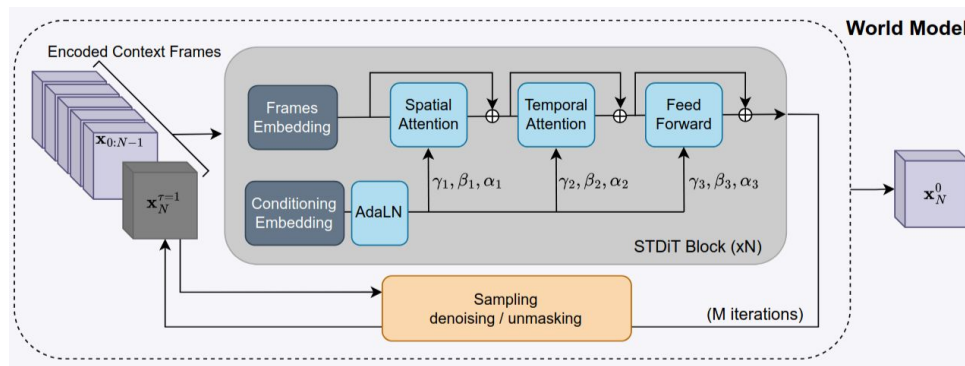
Iterative prediction:
- Continuous: denoising (Flow Matching)
- Discrete: unmasking (MaskGIT)

Continuous representation outperform Discrete one

	Discrete	Continuous
rFID ↓	FVD ↓ Orbis-MG	FVD ↓ Orbis
9.10	533.28	246.11

Does token representation play a role?

DINO	TF	Vocab Size	rFID ↓	FVD ↓ Orbis-MG	FVD ↓ Orbis
✗	✗	4096	9.33	1331.28	240.34
✓	✗	4096	12.17	1214.34	248.79
✓	✓	2×4096	9.10	533.28	246.11

universität freiburg

Orbis outperforms existing Driving World Models

Model	nuPlan	FVD↓ Waymo	nuPlan turns
Cosmos [1]	291.80	278.19	248.39
Vista [17]	323.37	422.58	413.61
GEM [25]	431.69	291.84	357.25
DW* [31]	298.97	N/A	334.89
Orbis (ours)	**134.06**	**167.57**	**239.20**

Model Comparison: FVD vs Training Data

universität freiburg
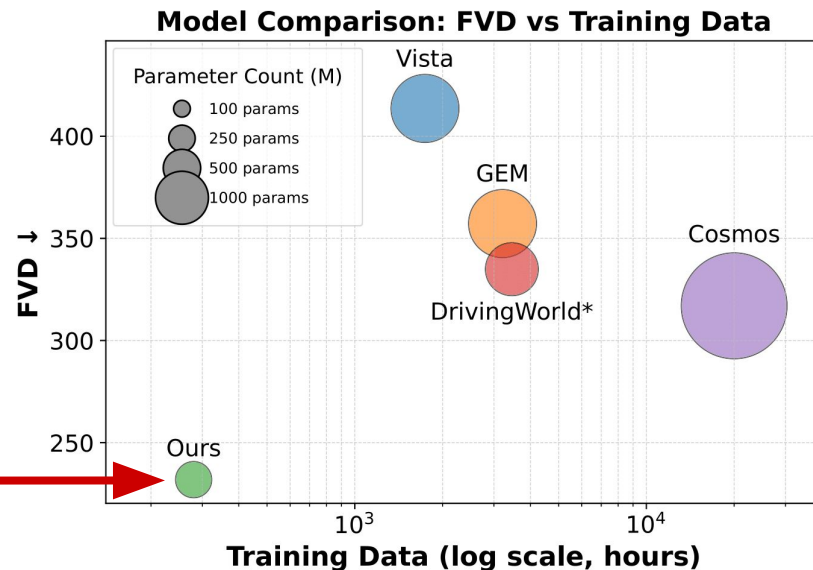
Orbis has more accurate trajectory

From video to trajectory: VGGT (Wang et al. 2025), as an inverse dynamics model.
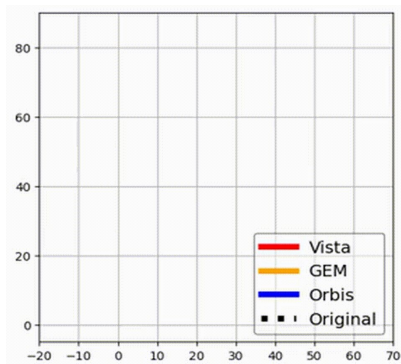
Comparing trajectories of real and generated videos.

Model	Frechet		ADE	
	Prec.	Rec.	Prec.	Rec.
Vista	0.39	0.45	0.25	0.48
GEM	0.33	0.54	0.27	0.47
Ours	**0.47**	**0.56**	**0.41**	**0.51**

Unconditional generation:
trajectory distribution similarity

Model	ADE ↓
Unconditional	5.20
+ ego-motion	2.40

Conditional generation of Orbis:
trajectory per-sample similarity

Conclusion

- Efficient Driving World model
- Better in challenging scenarios
- Continuous model outperforms the discrete model
- Proposed metric for trajectory evaluation

Project page

Poster Session 4
Neurips 2025
San Diego
Thu 4 Dec
4:30-7:30 p.m. PST

EurIPS 2025
Copenhagen

universität freiburg