OpenDriveLab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Tutorial on **Predictive World Model**

Zetong Yang

OpenDriveLab at Shanghai AI Lab

June 10, 2024

Outline

- 课程目标
    - 掌握世界模型的基础概念
    - 了解世界模型的典型做法和挑战
    - 了解世界模型的潜在问题和未来研究方向

Outline

- 世界模型概述 / Introduction
  - 背景与动机 / Motivation
  - 发展历程 / Roadmap
- 基础方法 / Method
  - 生成模型概述 / Generation Model
  - 世界模型涉及 / World Model
- 关键研究内容与挑战 / Frontiers and Challenges
- Q & A

OpenDriveLab

OpenDriveLab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

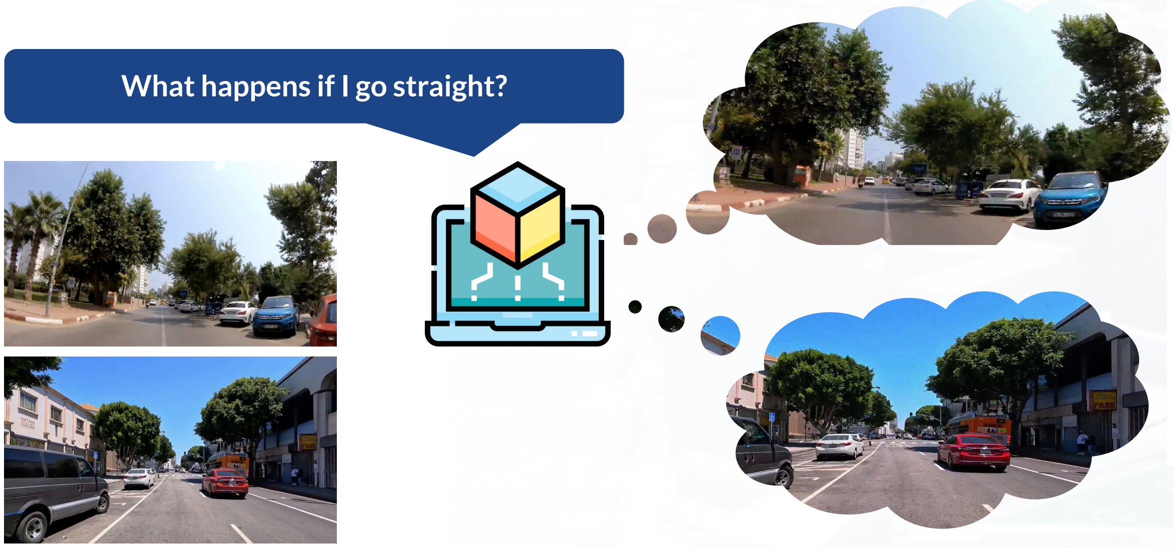
世界模型概述 / Introduction

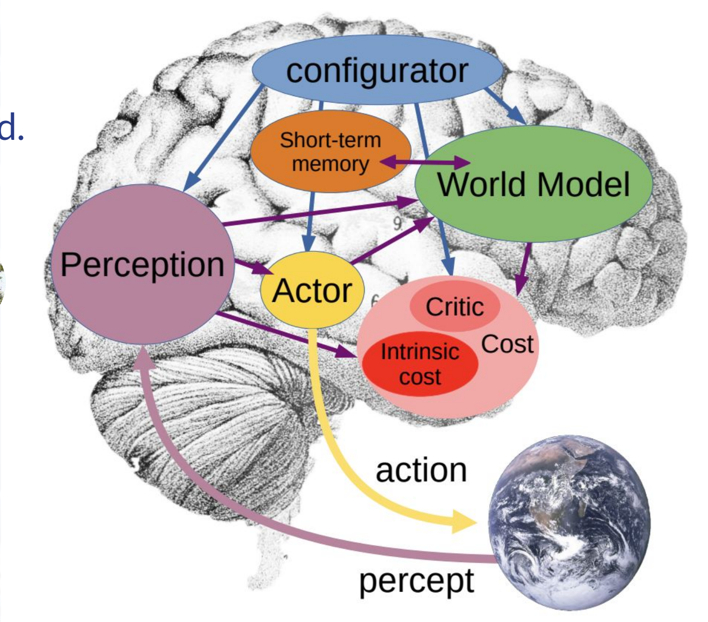
World Model

A Path Towards Autonomous Machine Intelligence, Yann Lecun

Task / Objective:

- **Represent the world & Learn to predict and re-act**
  - Simulate the world without **REAL** interaction with the world.



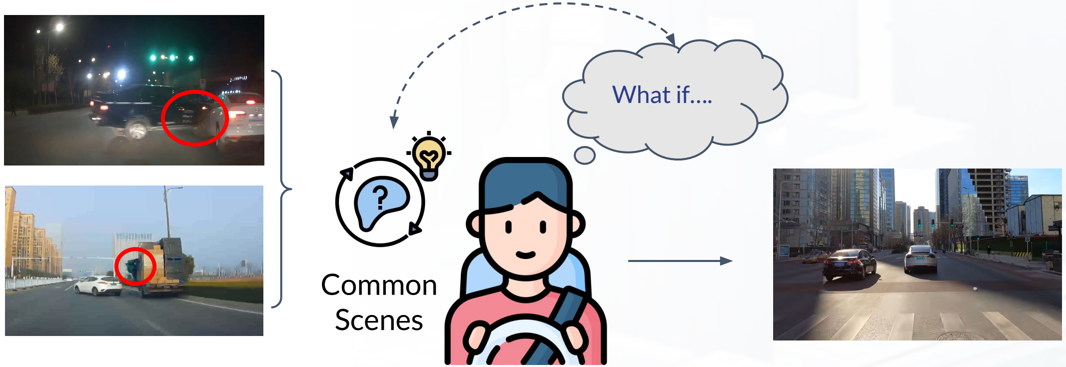


World Model

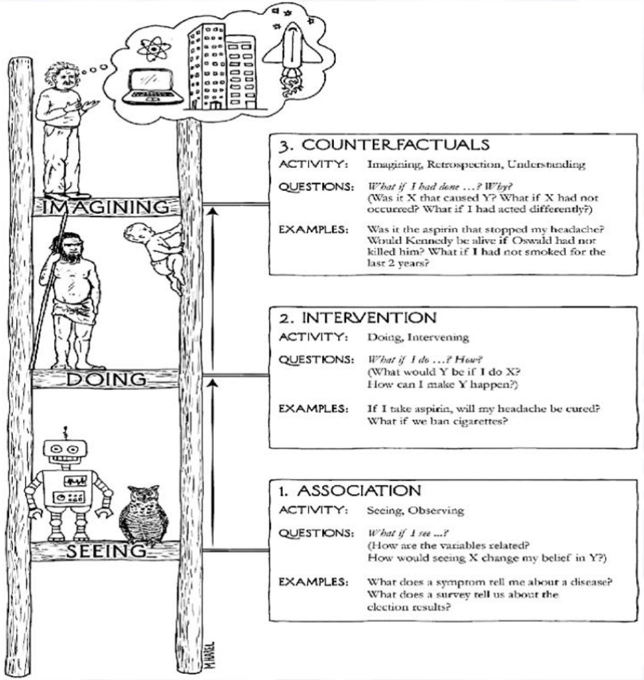
A Path Towards Autonomous Machine Intelligence, Yann Lecun

Motivation (Why study world model):

- **Simulate the world: learn new skills with very few trials**
  - Human and non-human animals model the world, infer and act in imagination, then make final decision.



Observation   Human Driver   Drive Safely

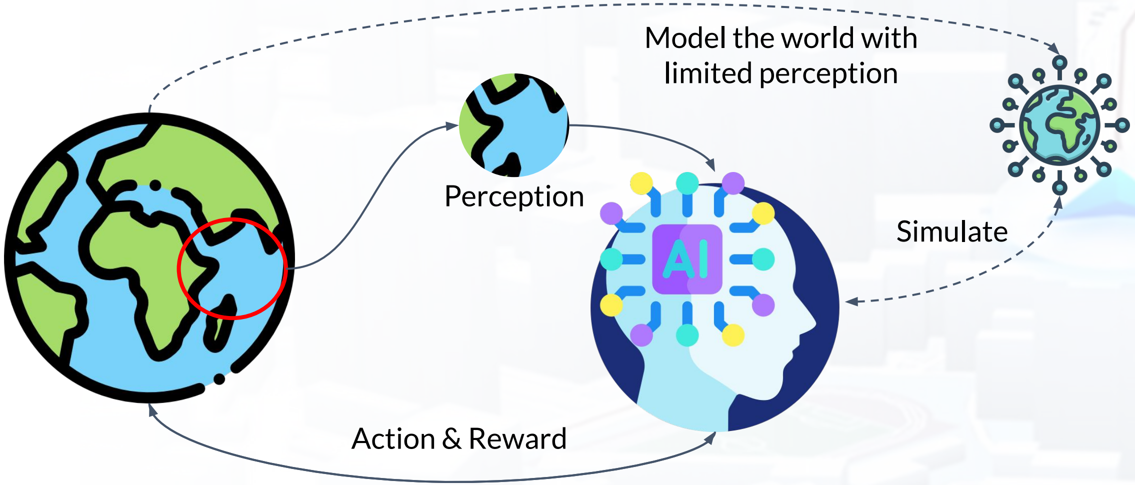


World Model

A Path Towards Autonomous Machine Intelligence, Yann Lecun

Motivation (Why study world model):

- **Enable agent: intelligent agents can perceive the world.**
  - The agent can predict what happens if taking some actions.

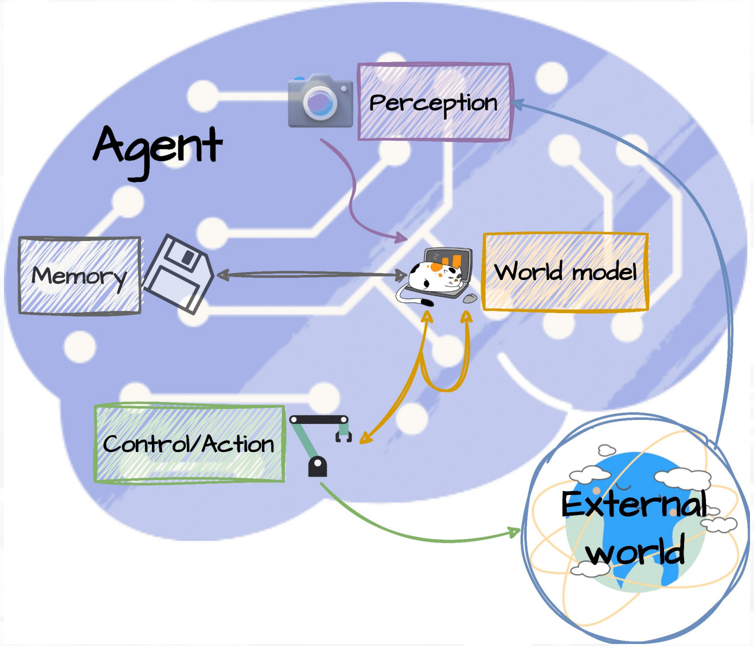


World Model

Big Picture of World Model

Intelligent Agent:

- **Perception Model:** estimate **state** from **observation**
- **Action Model:** propose **actions** given current state.
- **World Model:** predict **future states** given **actions and states.**
- **Reward:** compute **"penalty"** (GOAL: minimize penalty), from estimated **future states.**
- **Memory:** keep track of **states and rewards**.

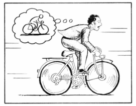


World Models for Autonomous Driving: An Initial Survey

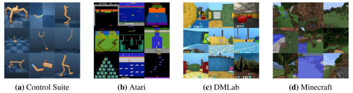
World Model

A Comprehensive Survey on General World Models and Beyond

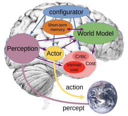
Roadmap to Autonomous Driving World Model:



World Models – *model RL environments.*



Dreamer Series– *towards general agents and scalable world models.*

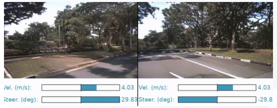


Position Paper (by LeCun) – *simulate the world, rehearse in the mind.*

UniPi/UniSim – *action/goal controlled universal video generation.*

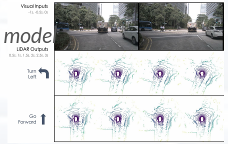
GAIA-1 – *action controlled realistic driving video generation.*

Tesla General World Model – *end to end world model for driving.*



DriveDreamer – *world model derived from real-world driving scenarios.*

Drive-WM – *the first driving world model compatible in E2E planning.*



ViDAR – *predicting future point clouds from historical visual input.*

GenAD – *2000 hours of driving videos and a generative driving model.*

Vista – *high-fidelity, versatile, and generalizable driving world model.*

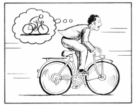
RL Agents | 18.3 | 20.3 | Vision | 22.6 | 23.2 | Driving | 23.6 | 23.9 | 23.11 | 24.3 | 24.5

From simulated agents to real world driving systems

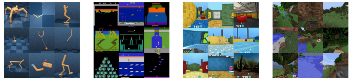
World Model

A Comprehensive Survey on General World Models and Beyond

Roadmap to Autonomous Driving World Model:



World Models – *model RL environments.*



(a) Control Suite (b) Atari (c) DMLab (d) Minecraft

Dreamer Series– *towards general agents and scalable world models.*

RL Agents

18.3

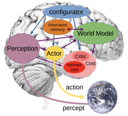
20.3

World models initially emerged in the field of reinforcement learning (RL) to model the environment, allowing an agent to evaluate actions without taking real actions, and thereby make the best decisions.

World Model

A Comprehensive Survey on General World Models and Beyond

Roadmap to Autonomous Driving World Model:



Position Paper (by LeCun) – *simulate the world, rehearse in the mind.*

UniPi/UniSim – *action/goal controlled universal video generation.*

Vision

22.6

23.2

In 2022, LeCun published a position paper proposing a pathway to achieving autonomous machine intelligence, where the world model is the most critical component. This paper presented an ideal vision of future artificial intelligence.

World Model

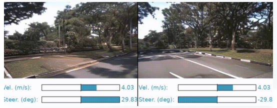
A Comprehensive Survey on General World Models and Beyond

Roadmap to Autonomous Driving World Model:

Subsequently, world models have flourished in areas such as video generation, autonomous driving, and autonomous agents.

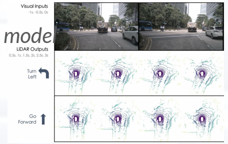
GAIA-1 – *action controlled realistic driving video generation.*

Tesla General World Model – *end to end world model for driving.*



DriveDreamer – *world model derived from real-world driving scenarios.*

Drive-WM – *the first driving world model compatible in E2E planning.*



ViDAR – *predicting future point clouds from historical visual input.*

GenAD – *2000 hours of driving videos and a generative driving model.*

Vista – *high-fidelity, versatile, and generalizable driving world model.*

Driving

23.6 | 23.9 | 23.11 | 24.3 | 24.5

OpenDriveLab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

基础方法 / Method

World Model

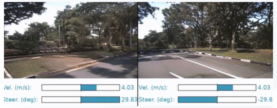
A Comprehensive Survey on General World Models and Beyond

Roadmap to Autonomous Driving World Model:

Subsequently, world models have flourished in areas such as video generation, autonomous driving, and autonomous agents.

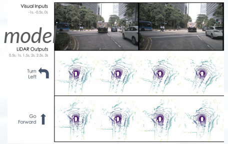
GAIA-1 – *action controlled realistic driving video generation.*

Tesla General World Model – *end to end world model for driving.*



DriveDreamer – *world model derived from real-world driving scenarios.*

Drive-WM – *the first driving world model compatible in E2E planning.*



ViDAR – *predicting future point clouds from historical visual input.*

GenAD – *2000 hours of driving videos and a generative driving model.*

Vista – *high-fidelity, versatile, and generalizable driving world model.*

Driving

23.6 | 23.9 | 23.11 | 24.3 | 24.5

Big Picture of World Model

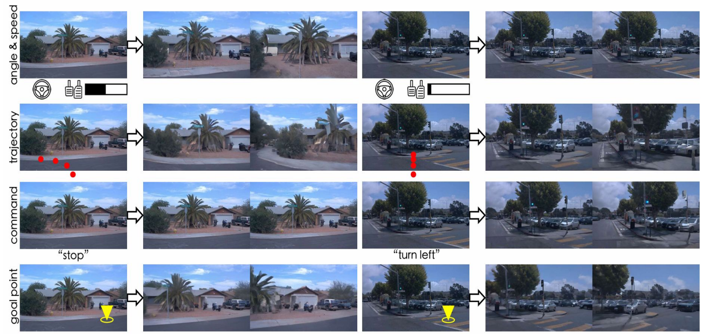
How to achieve world model:

- **From the most general perspective, World Model = $\boxed{\text{Generation}}$ + $\boxed{\text{Control}}$**

Common control types in autonomous driving

- GAN-based
- Diffusion-based
- Autoregressive modeling-based
- Masked modeling-based

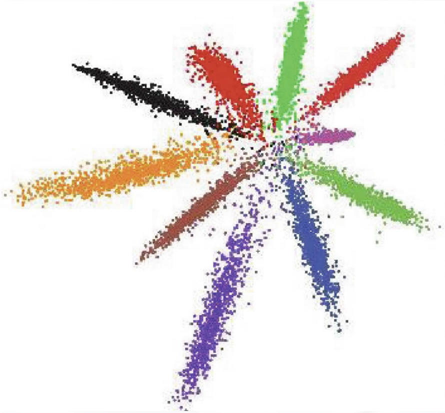
- Texts
- Destinations & Trajectories
- Ego-vehicle actions



Controllable Video Generation for World Model

Generation Models

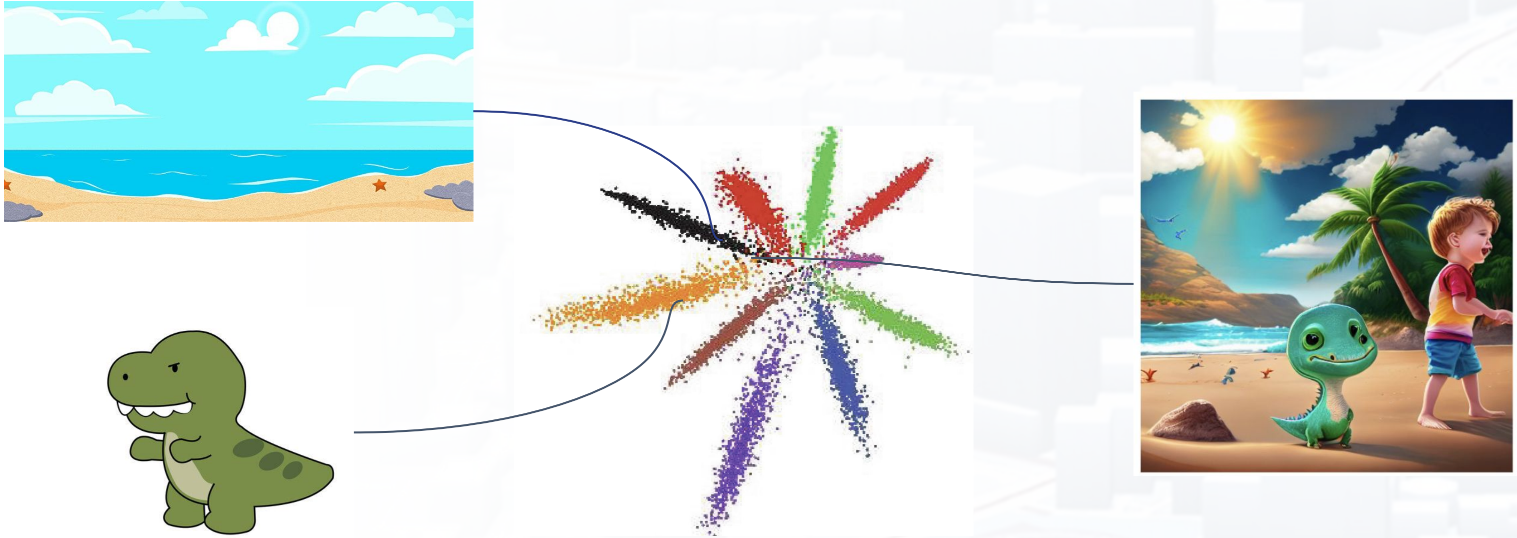
A function to map samples to a distribution.



Controllable Video Generation for World Model

Generation Models

Sample unseen images from the distribution.

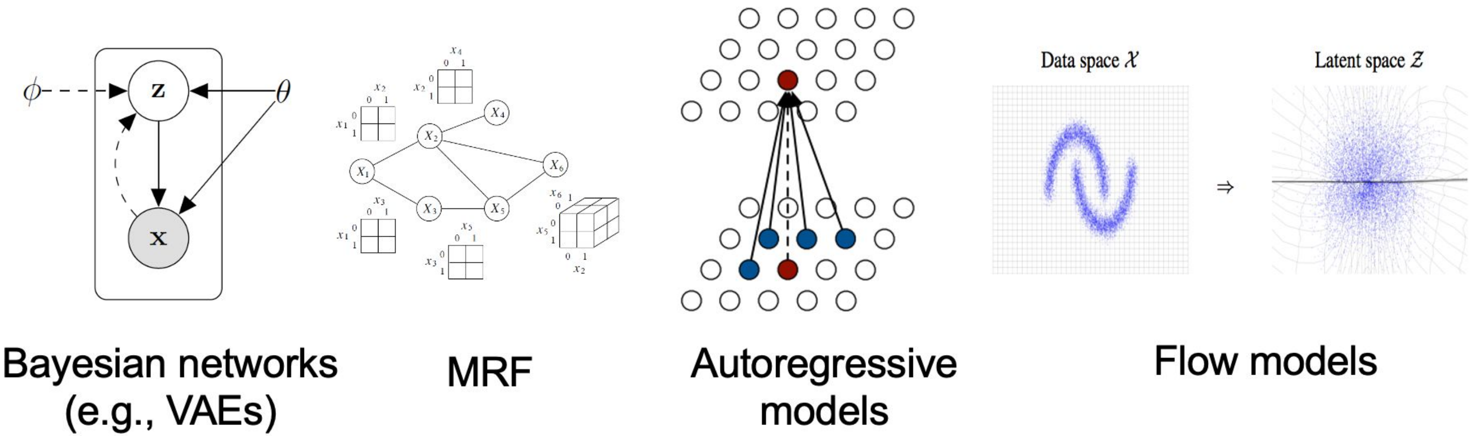


Controllable Video Generation for World Model

How to represent probability distribution of natural images?

Generative Models can be grouped into:

- Likelihood-based models
  - Directly learn the distribution function via maximum likelihood.

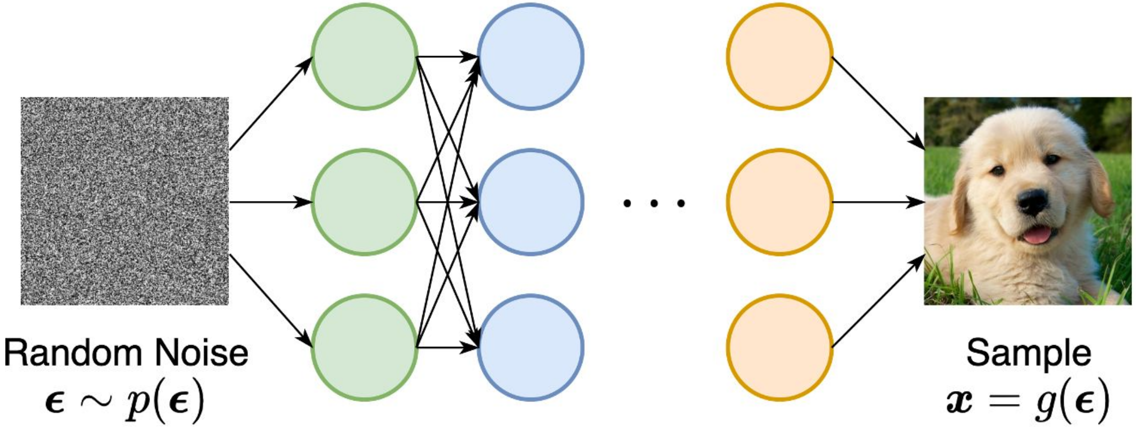


Controllable Video Generation for World Model

How to represent probability distribution of natural images?

Generative Models can be grouped into:

- Implicit generation model
  - the distribution is implicitly represented by a model. (GAN)

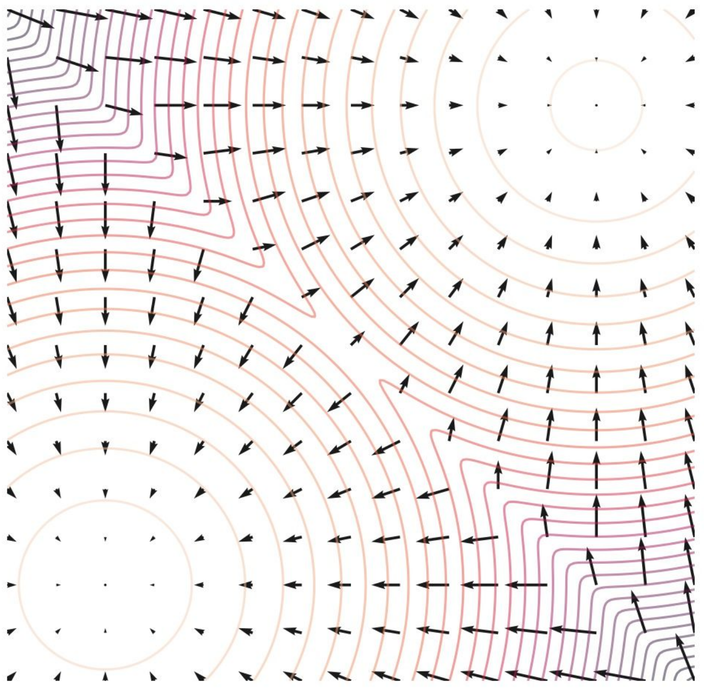


Controllable Video Generation for World Model

How to represent probability distribution of natural images?

Generative Models can be grouped into:

- Diffusion model (Score-based model)
  - Model the gradient of the log probability density function, instead of distribution itself.

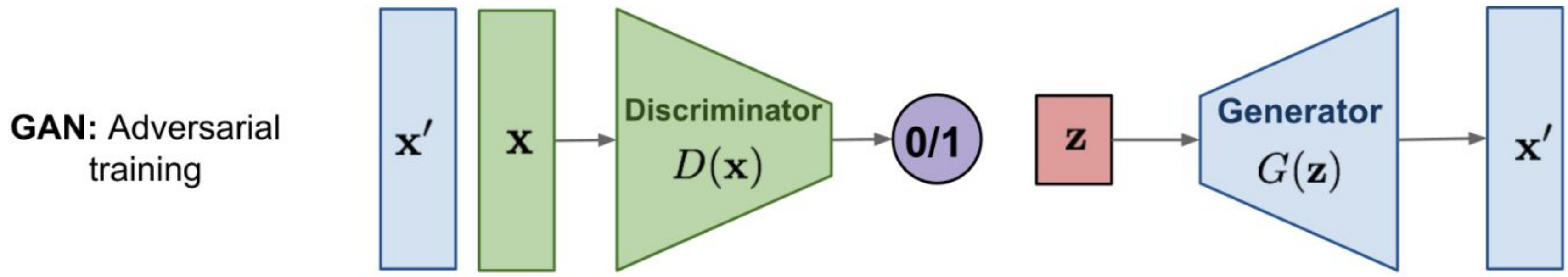


Generative Adversarial Network (GAN)

Architecture

Two models, competing with each other and making other stronger.

- Generator
    - Outputs synthetic samples given a noise variable input
- Discriminator
    - A critic to tell the fake samples from the real ones



Generative Adversarial Network (GAN)

Architecture

Discriminator (D):

- Real samples: maximize the probability $\mathbb{E}_{x \sim p_r(x)}[\log D(x)]$
- Fake samples: output a probability close to zero, by maximizing $\mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$

Generator (G):

- Increase the chances of producing a high probability for a fake example, thus to minimize $\mathbb{E}_{z \sim p_z(z)}[\log(1 - D(G(z)))]$

Generative Adversarial Network (GAN)

Convergence Issue

Hard to achieve Nash Equilibrium because generator and discriminator update themselves independently.

- Nash Equilibrium: a situation where no player could gain by changing their own strategy.

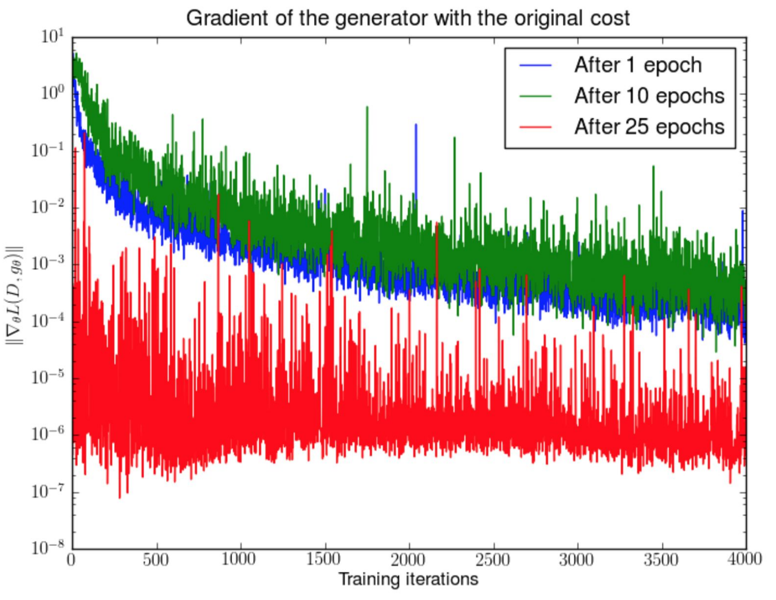
囚犯的博弈矩阵		囚犯乙	
		招供	不招供
囚犯甲	招供	各判刑2年	甲立即释放，乙判刑10年
	不招供	甲判刑10年，乙立即释放	各判刑半年

Generative Adversarial Network (GAN)

Convergence Issue

Gradient vanishing when two distributions have no overlap:

- If the discriminator behaves badly, the generator does not have accurate feedback and the loss function cannot represent the reality.
- If the discriminator does a great job, the gradient of the loss function drops down to close to zero and the learning becomes super slow or even jammed.



Generative Adversarial Network (GAN)

Improve GAN training

Stablize the training stage:

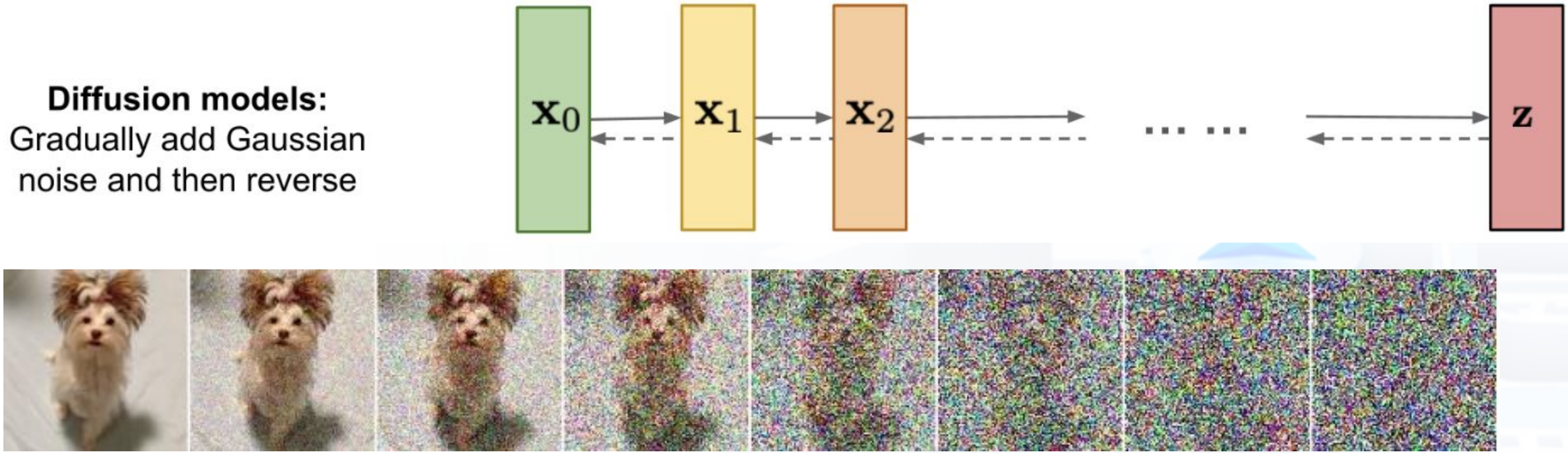
- **Historical Averaging:** penalizes the training speed when is changing too dramatically in time.
- **Adding Noises:** create higher chances for two probability distributions to have overlaps.
- **Virtual Batch Normalization:** each data sample is normalized based on a fixed batch (“reference batch”) of data rather than within its minibatch
- **Minibatch Discrimination:** add more data points into GAN loss.
- **Wasserstein distance:** to solve the situation where two distributions (real / fake) have no overlap.

Diffusion Model (Score-based Model)

Explicit distribution modeling

A markov process to slowly add noise to data and then learn the inverse.

- Explicitly model the data distribution via probability density function.



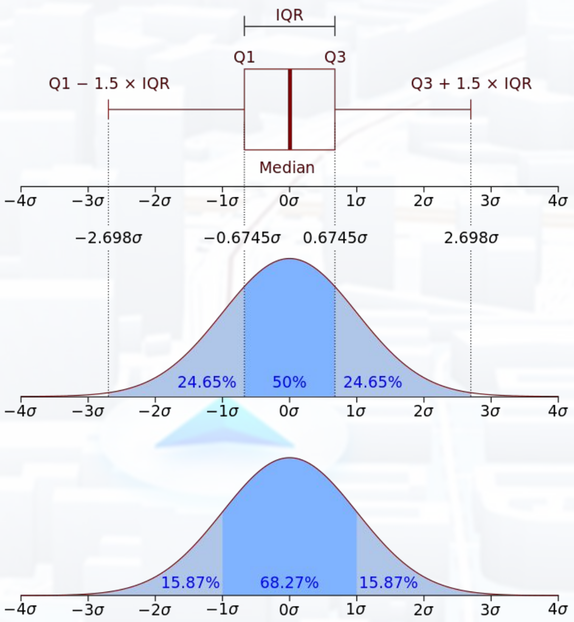
Diffusion Model (Score-based Model)

Explicit distribution modeling

Probability Density Function (PDF)

- A relative likelihood that the value of the random variable.

$$\Pr[a \le X \le b] = \int_a^b f_X(x)\,dx.$$



Diffusion Model (Score-based Model)

How to model PDF in score-based model

Suppose X represents data samples; $p(\mathrm{x})$ represents the underlying data distribution:

- First design the PDF as

$$p_\theta(\mathbf{x}) = \frac{e^{-f_\theta(\mathbf{x})}}{Z_\theta}$$

A similar form to gaussian distribution, where Z_θ is a normalization form, to ensure the integration to be 1. θ is the learnable parameter.

Diffusion Model (Score-based Model)

How to train a generation model

A simple solution is to **Maximize the PDF for each sample:**

$$\max_{\theta} \sum_{i=1}^{N} \log p_{\theta}(\mathbf{x}_i)$$

Diffusion Model (Score-based Model)

How to train a generation model

A simple solution is to **Maximize the PDF for each sample:**

$$\max_{\theta} \sum_{i=1}^{N} \log p_\theta(\mathbf{x}_i)$$

Undesired

$$p_\theta(\mathbf{x}) = \frac{e^{-f_\theta(\mathbf{x})}}{Z_\theta}$$

Need to deal with Z_θ, which typically is **intractable** given we don't know f_θ

Diffusion Model (Score-based Model)

How to train a generation model

Alternative approach: optimize the **gradients** of p_θ

- We define a function s_θ, as the gradient of p_θ :

$$\mathbf{s}_\theta(\mathbf{x}) = \nabla_\mathbf{x} \log p_\theta(\mathbf{x}) = -\nabla_\mathbf{x} f_\theta(\mathbf{x}) - \underbrace{\nabla_\mathbf{x} \log Z_\theta}_{=0} = -\nabla_\mathbf{x} f_\theta(\mathbf{x})$$

- By doing so, we are still training the f_θ, but avoid the intractable Z_θ !

The function s_θ is called the **score function**, and a model for the score function is called **score-based model**.

Diffusion Model (Score-based Model)

How to train a generation model

Then, we can train the score-based model by minimizing the discrepancy between real data distribution and the model:

$$\mathbb{E}_{p(\mathbf{x})}[\|\nabla_{\mathbf{x}} \log p(\mathbf{x}) - \mathbf{s}_\theta(\mathbf{x})\|_2^2]$$

Diffusion Model (Score-based Model)

How to train a generation model

Then, we can train the score-based model by minimizing the discrepancy between real data distribution and the model:

$$\mathbb{E}_{p(\mathbf{x})}\left[\|\boxed{\nabla_{\mathbf{x}} \log p(\mathbf{x})} - \mathbf{s}_\theta(\mathbf{x})\|_2^2\right]$$

How to track the ground-truth data score? Score matching

Diffusion Model (Score-based Model)

Score Matching

Where the add-noise / de-noise procedure stands out.

- Since hard to estimate $p(\mathrm{x})$, how about using conditional probability for estimation?

$$q_\sigma(\tilde{x}) \triangleq \int q_\sigma(\tilde{x}|x)p_{\mathrm{data}}(x)dx$$

Add small noise to the original data distribution, and ensure $q_\sigma(\tilde{x})$ to be similar to $p_{\mathrm{data}}(x)$

Diffusion Model (Score-based Model)

Score Matching

Then we can transfer the original loss function from:

$$\mathbb{E}_{p(\mathrm{x})}\left[||\nabla_{\mathrm{x}} \log p\left(\mathrm{x}\right) - s_{\theta}\left(x\right)||_2^2\right]$$

$$q_{\sigma}(\tilde{x}) \triangleq \int q_{\sigma}(\tilde{x}|x)p_{\mathrm{data}}(x)dx$$

to:

$$\mathbb{E}_{q_{\sigma}(\tilde{\mathrm{x}})}\left[||\nabla_{\tilde{\mathrm{x}}} \log q_{\sigma}\left(\tilde{\mathrm{x}}\right) - s_{\theta}\left(\mathrm{x}\right)||_2^2\right]$$

P. Vincent. A connection between score matching and denoising autoencoders. Neural computation

$$\mathbb{E}_{q_{\sigma}(\tilde{\mathrm{x}}\,|\,\mathrm{x})p_{\mathrm{data}}(\mathrm{x})}\left[||\nabla_{\tilde{x}} \log q_{\sigma}\left(\tilde{\mathrm{x}}\,|\,\mathrm{x}\right) - s_{\theta}\left(\mathrm{x}\right)||_2^2\right]$$

Diffusion Model (Score-based Model)

Score Matching

Then we can transfer the original loss function from:

$$\mathbb{E}_{p(\mathrm{x})}\left[||\nabla_{\mathrm{x}} \log p(\mathrm{x}) - s_\theta(x)||_2^2\right]$$

$$q_\sigma(\tilde{x}) \triangleq \int q_\sigma(\tilde{x}|x) p_{\mathrm{data}}(x) dx$$

to:

$$\mathbb{E}_{q_\sigma(\tilde{\mathrm{x}})}\left[||\nabla_{\tilde{\mathrm{x}}} \log q_\sigma(\tilde{\mathrm{x}}) - s_\theta(\mathrm{x})||_2^2\right]$$

P. Vincent. A connection between score matching and denoising autoencoders. Neural computation

$$\mathbb{E}_{q_\sigma(\tilde{\mathrm{x}} \mid \mathrm{x}) p_{\mathrm{data}}(\mathrm{x})}\left[\boxed{||\nabla_{\tilde{x}} \log q_\sigma(\tilde{\mathrm{x}} \mid \mathrm{x}) - s_\theta(\mathrm{x})}||_2^2\right]$$

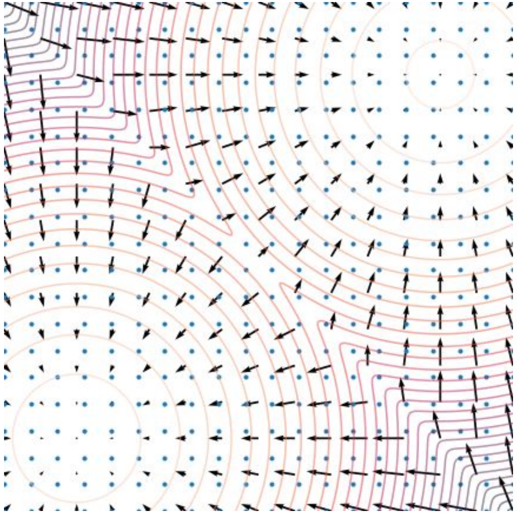
That's why diffusion model use **Gaussian Noise** as supervision

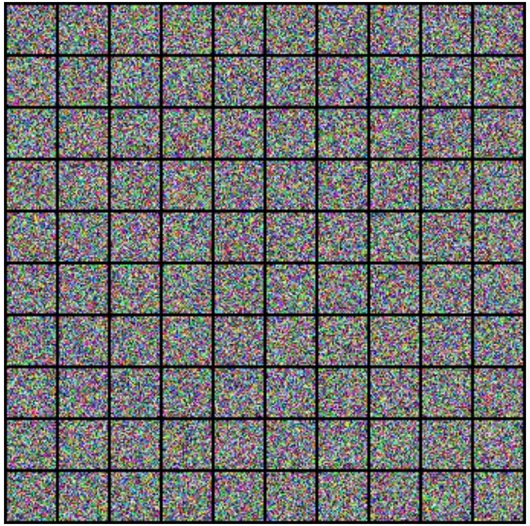
Diffusion Model (Score-based Model)

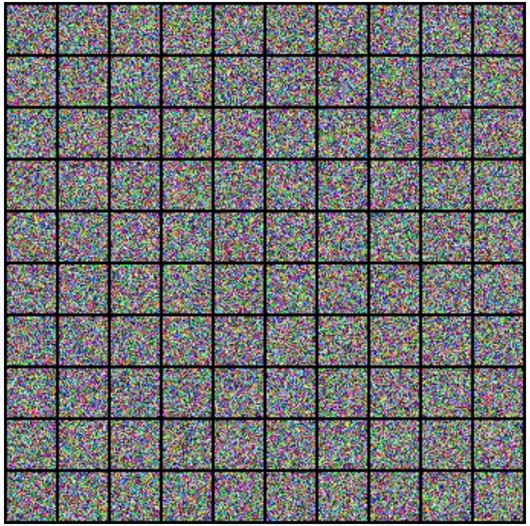
Langevin dynamics

Draw samples from score-based models:

$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i + \epsilon \nabla_{\mathbf{x}} \log p(\mathbf{x}) + \sqrt{2\epsilon}\, \mathbf{z}_i, \quad i = 0, 1, \cdots, K,$$





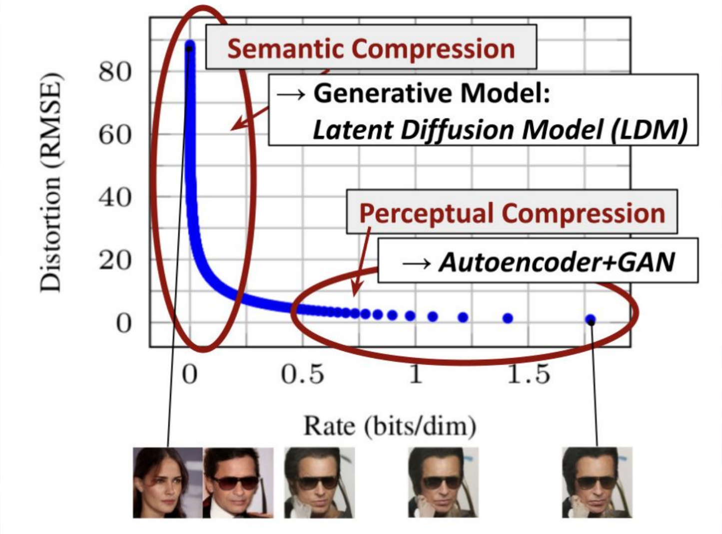


Diffusion Model (Score-based Model)

How to scale up diffusion model?

Latent Diffusion Model (CVPR 2021):

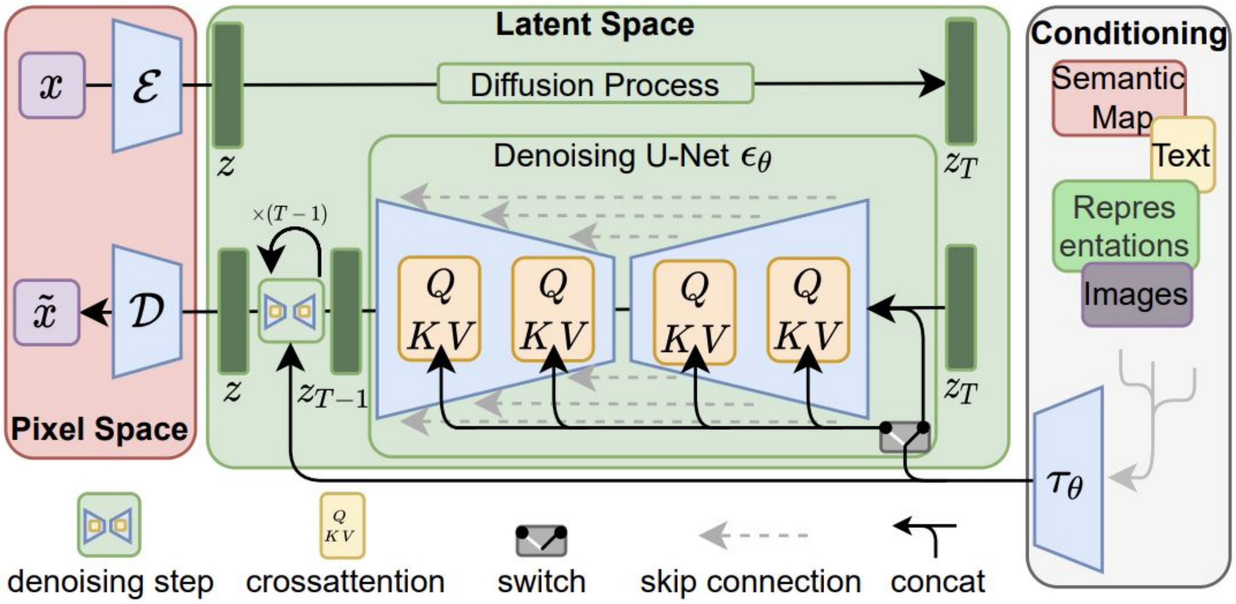
- **Key Insight:** Diffusion learning can be roughly divided into two stages:
  - **Perceptual compression stage** which removes high-frequency details but still learns little semantic variation.
  - **Semantic compression stage** learns the semantic and conceptual composition of the data.
- **Key Idea:** find a perceptually equivalent, but computationally more suitable space



Diffusion Model (Score-based Model)

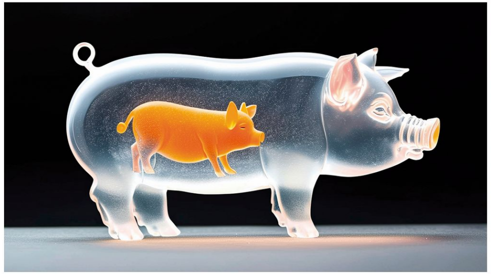
How to scale up diffusion model?

Latent Diffusion Model (CVPR 2021):



Diffusion Model (Score-based Model)

How to scale up diffusion model?



Prompt: Translucent pig, inside is a smaller pig.

Prompt: A massive alien space ship that is shaped like a pretzel.

Prompt: A kangaroo holding a beer, wearing ski goggles and passionately singing silly songs.

Prompt: An entire universe inside a bottle sitting on the shelf at walmart on sale.

Prompt: A cheeseburger with juicy beef patties and melted cheese sits on top of a toilet that looks like a throne and stands in the middle of the royal chamber.

Prompt: This dreamlike digital art captures a vibrant, kaleidoscopic bird in a lush rainforest

Prompt: A car made out of vegetables.

Prompt: Heat death of the universe line art

World Model

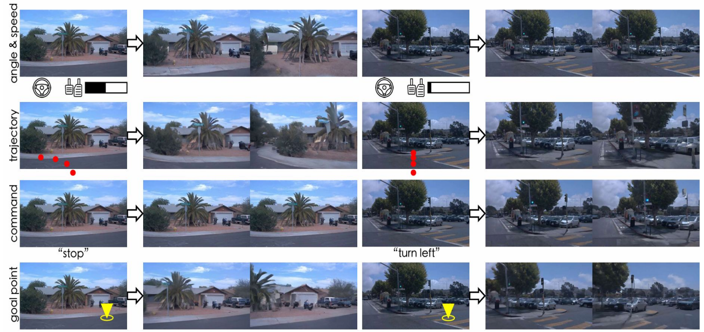
Big Picture of World Model

How to achieve world model:

- From the most general perspective, World Model = $\boxed{\text{Generation}}$ + $\boxed{\text{Control}}$

Common control types in autonomous driving

- Texts
- Destinations & Trajectories
- Ego-vehicle actions



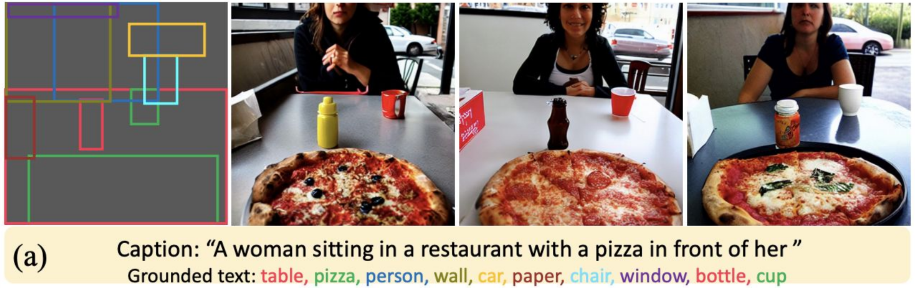
Controllable Generation

Concept

Generate images following control signals: $I = f(Z \mid C)$

- Z : a random variable
- C : a random variable.

a teddy bear on a skateboard in times square

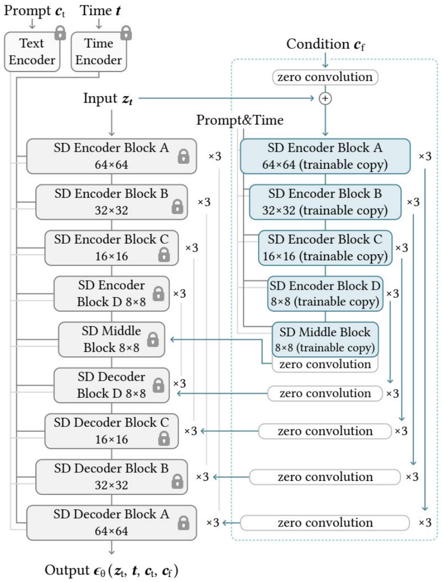


Hierarchical Text-Conditional Image Generation with CLIP Latents
GLIGEN: Open-Set Grounded Text-to-Image Generation
AnyControl: Create Your Artwork with Versatile Control on Text-to-Image Generation.

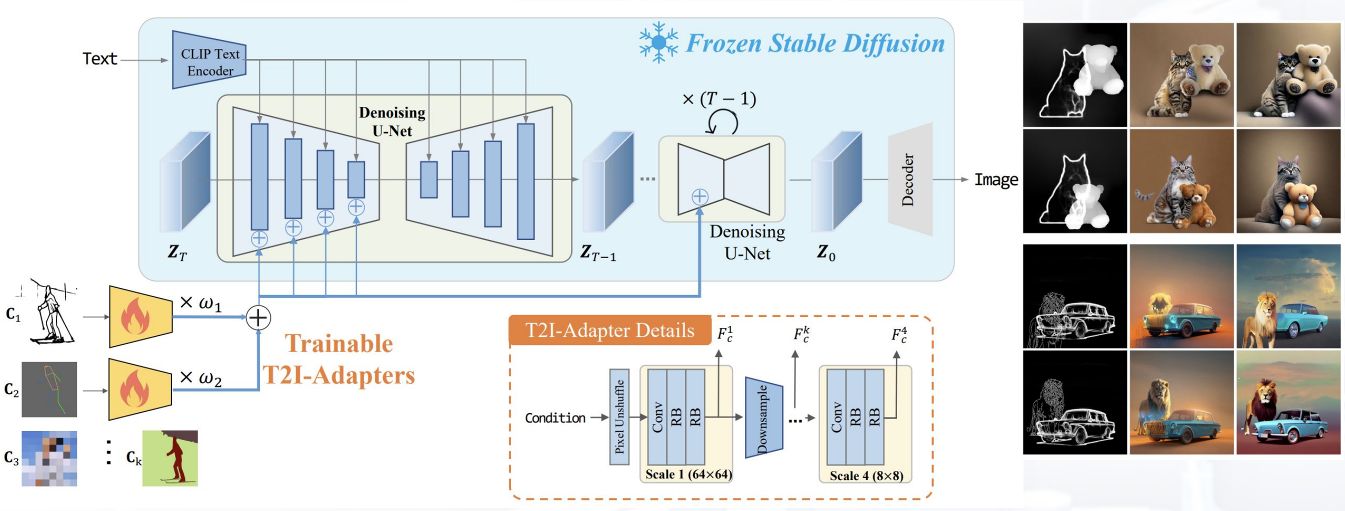
Controllable Generation

Typical methodology

Fine-tuning with control-image pairs:



(a) Stable Diffusion (b) ControlNet



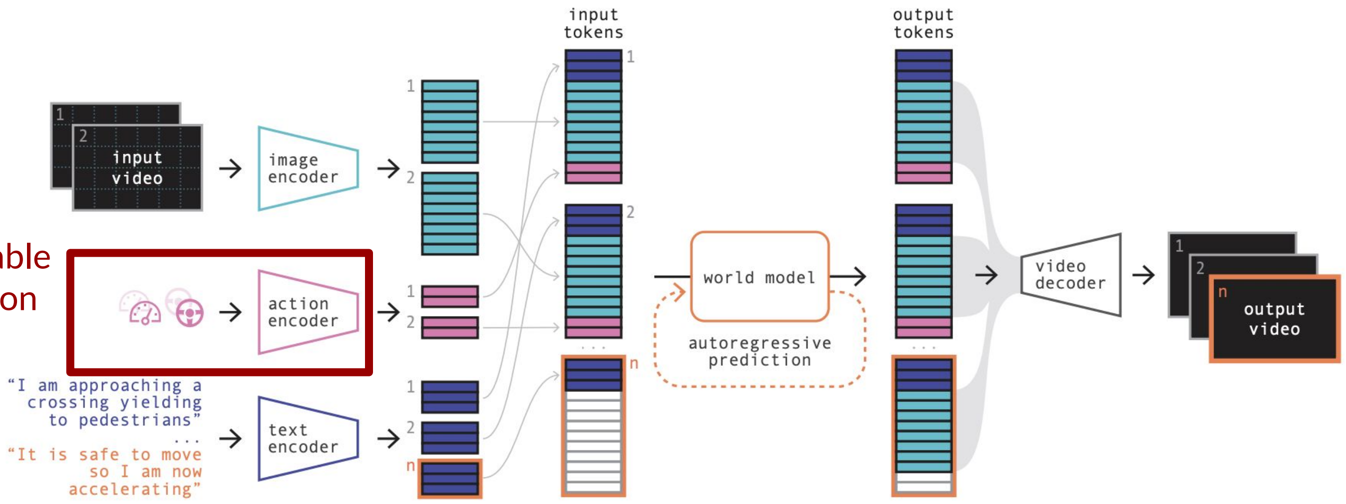
Adding Conditional Control to Text-to-Image Diffusion Models
T2I-Adapter: Learning Adapters to Dig out More Controllable Ability for Text-to-Image Diffusion Models
AnyControl: Create Your Artwork with Versatile Control on Text-to-Image Generation.

Predictive World Model

World model via controllable future generation

Predict future following the action signal: $s_{i+1} = f\left(s_i \mid a_i\right)$

Action-controllable Future Generation



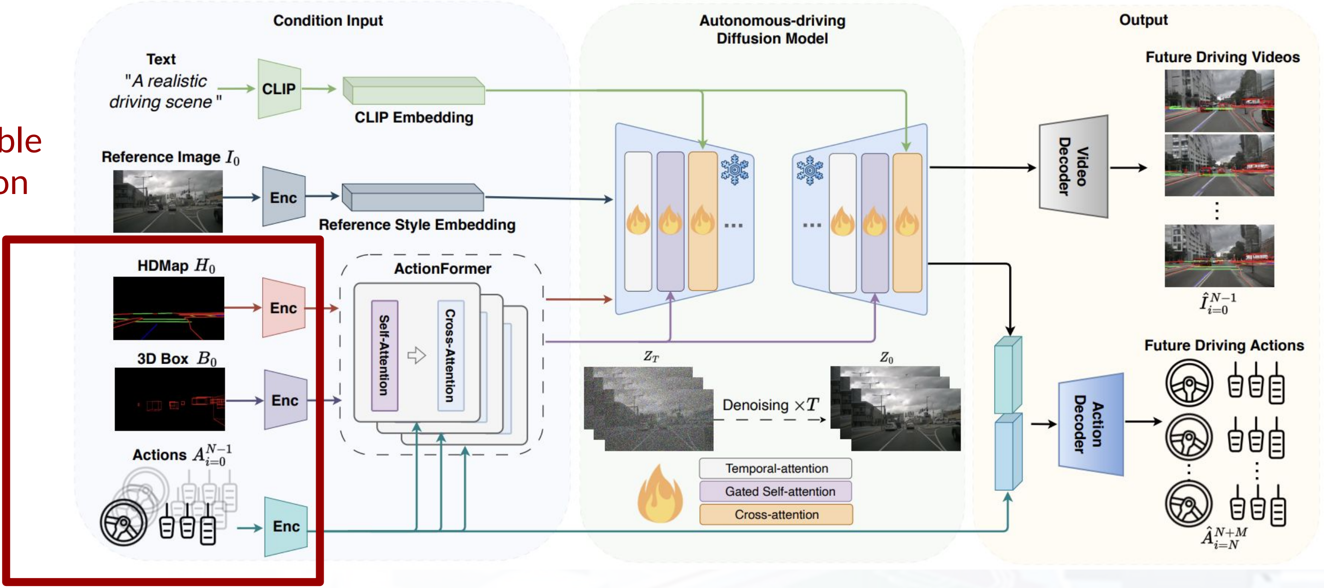
GAIA-1: A Generative World Model for Autonomous Driving

Predictive World Model

World model via controllable future generation

Predict future following the action signal: $s_{i+1} = f\left(s_i \mid a_i\right)$

Action-controllable Future Generation

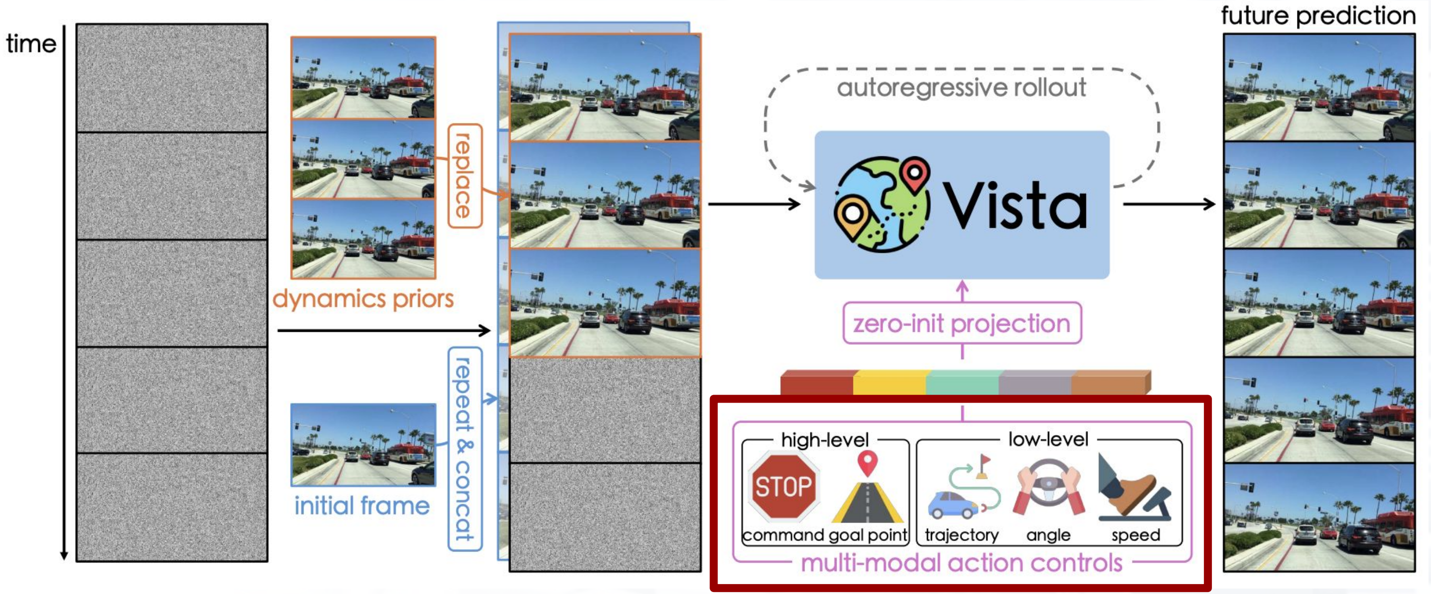


DriveDreamer: Towards Real-world-driven World Models for Autonomous Driving

Predictive World Model

World model via controllable future generation

Predict future following the action signal: $s_{i+1} = f\left(s_i \mid a_i\right)$



Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

Predictive World Model

World model via controllable future generation

Action control modes:

Method	Model Setups Data Scale	Frame Rate	Resolution	Action Control Modes Angle&Speed	Trajectory	Command	Goal Point
DriveSim [96]	7h	5 Hz	80×160	✓			
DriveGAN [63]	160h	8 Hz	256×256	✓			
DriveDreamer [118]	5h	12 Hz	128×192	✓			
Drive-WM [120]	5h	2 Hz	192×384		✓		
WoVoGen [84]	5h	2 Hz	256×448	✓			
ADriver-I [57]	300h	2 Hz	256×512			✓	
GenAD [130]	2000h	2 Hz	256×448		✓	✓	
GAIA-1 [50]	4700h	25 Hz	288×512	✓			
Vista (Ours)	1740h	10 Hz	576×1024	✓	✓	✓	✓

Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

Predictive World Model

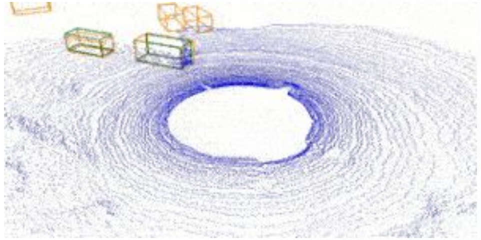
World model via controllable future generation

Vista: A Generalizable Driving World Model with High Fidelity and Versatile Controllability

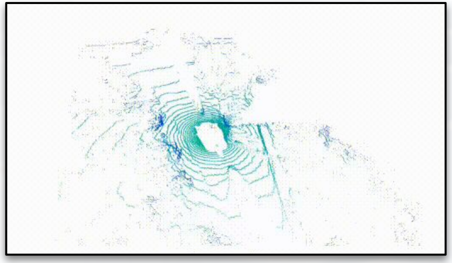
Predictive World Model

World model in other modalities.

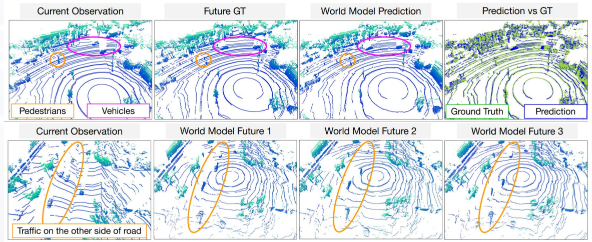
Point Cloud



S2Net — Point cloud future prediction for planning



4D-Occ — Ego Future Tracjectory



Copilot4D — LiDAR world model with diffusion techniques

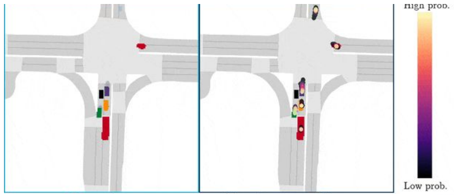
2022

2023

2024

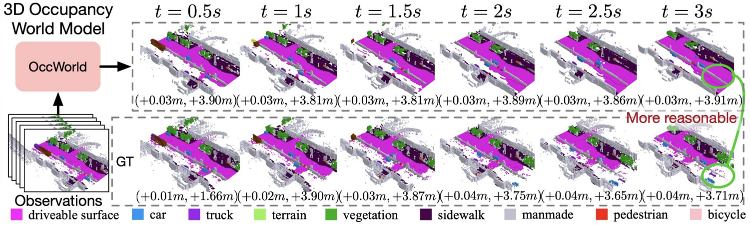
BEV

Fiery



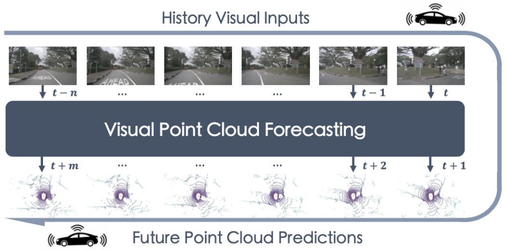
Occupancy

Occworld — occupancy world model



Multi-modality

ViDAR
multi-modal world model



OpenDriveLab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

关键研究内容与挑战 / Frontiers and Challenges

Frontiers and Challenges

How to use predictive world model?

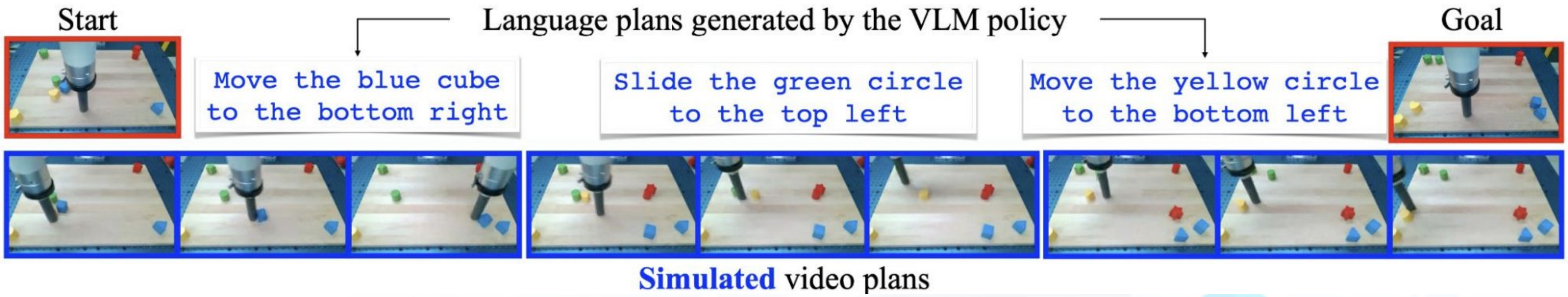
Why do we study world model?

- As powerful simulator to train agent in an unsupervised manner.
- Help decision by imagining the future.
- As a foundation model.

Frontiers and Challenges

As powerful simulator.

Prevalence in Embodied AI, but underexplored in Autonomous Driving.



Why this situation? **We think it is due to the much complex scenarios for autonomous driving compared to robotics.**

UniSim: Learning Interactive Real-World Simulators

Frontiers and Challenges

Autonomous Driving Scenarios v.s. Robotics

V.S.

Frontiers and Challenges

Help the decision process by imagining the future.

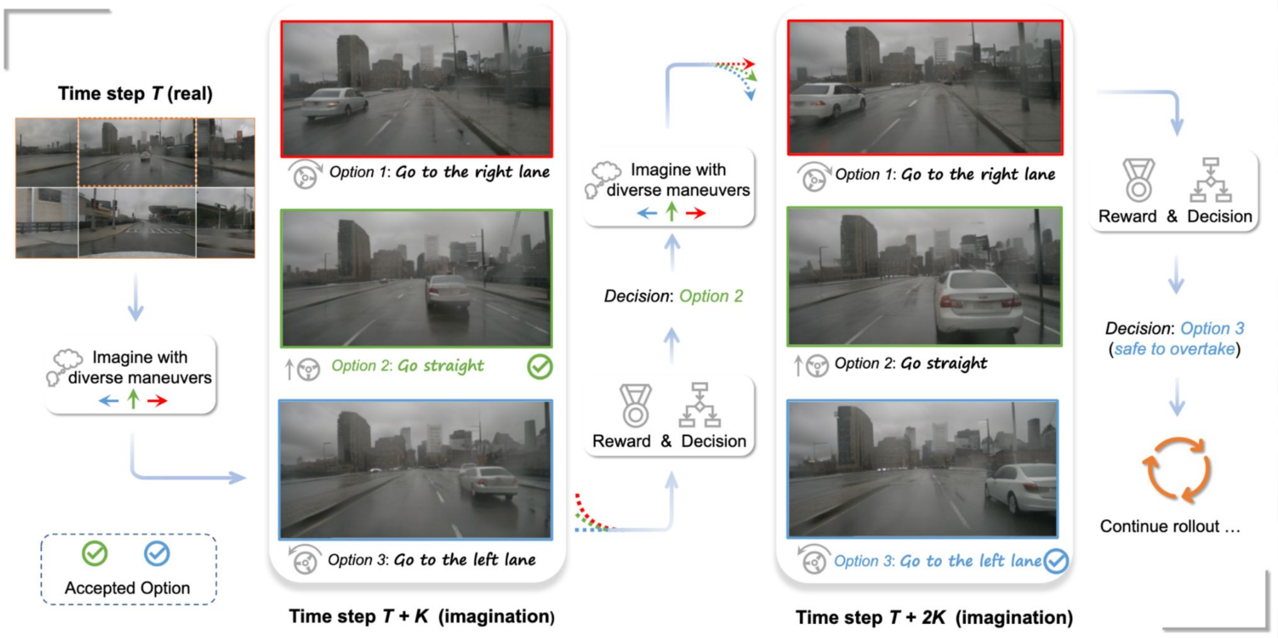
Drive into the Future: Multiview Visual Forecasting and Planning with World Model (CVPR 2024)

- **Predicting future in advance** and evaluating the foreseeable risks to empower autonomous vehicles for better planning their actions and enhancing safety and efficiency on the road.

Frontiers and Challenges

Help the decision process by imagining the future.

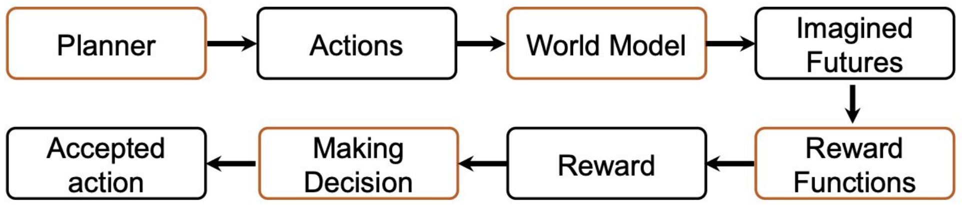
Drive into the Future: Multiview Visual Forecasting and Planning with World Model (CVPR 2024)



Frontiers and Challenges

Help the decision process by imagining the future.

Drive into the Future: Multiview Visual Forecasting and Planning with World Model (CVPR 2024)



Method	L2 (m) ↓ 1s	2s	3s	Avg.	Collision (%) ↓ 1s	2s	3s	Avg.
VAD (GT cmd)	0.41	0.70	1.05	0.72	0.07	0.17	0.41	0.22
VAD (rand cmd)	0.51	0.97	1.57	1.02	0.34	0.74	1.72	0.93
Ours	**0.43**	**0.77**	**1.20**	**0.80**	**0.10**	**0.21**	**0.48**	**0.26**

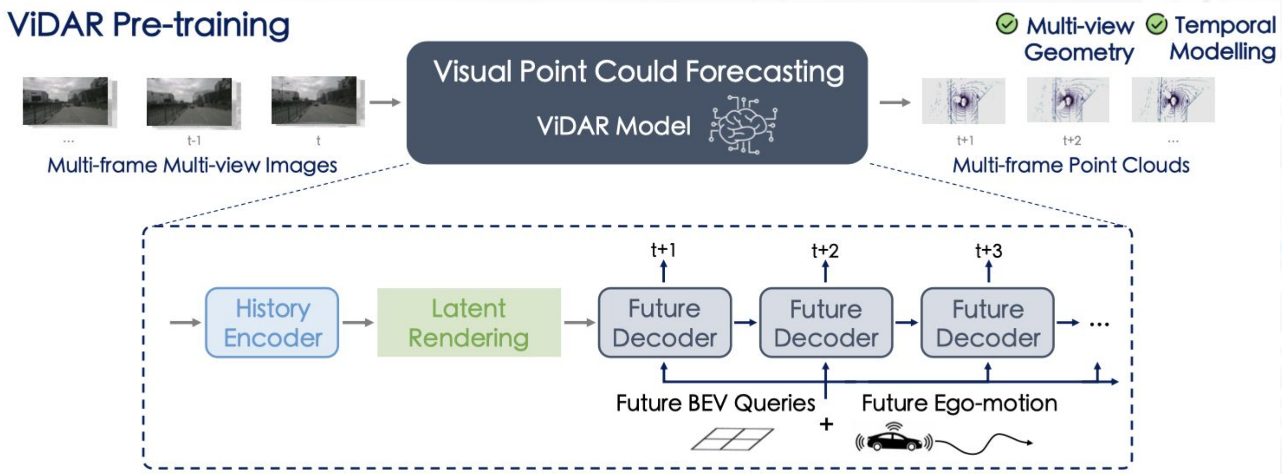
Imagination-then-Decision pipeline enhances the overall soundness of planning.

Frontiers and Challenges

Serve as powerful foundation model

Visual Point Cloud Forecasting enables Scalable Autonomous Driving (CVPR 2024, Highlight)

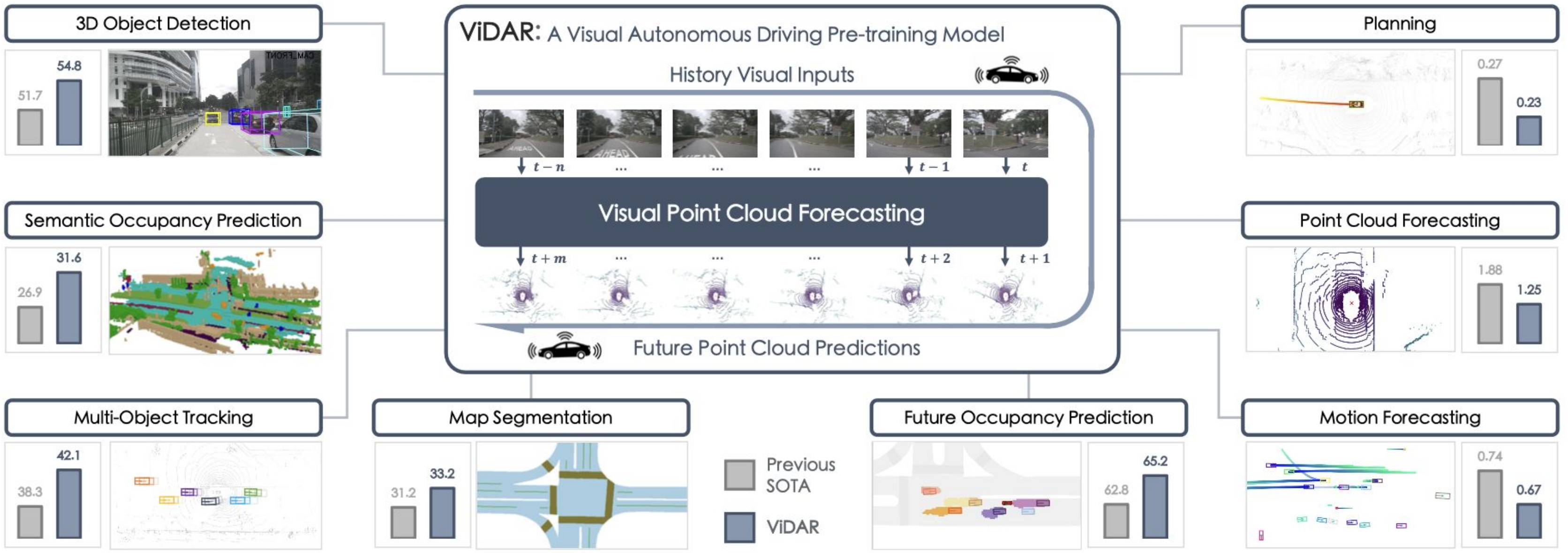
- **Visual point cloud forecasting** captures the synergic learning of semantics, 3D structures, and temporal dynamics. Hence it shows superiority in various downstream tasks.



Frontiers and Challenges

Serve as powerful foundation model

Visual Point Cloud Forecasting enables Scalable Autonomous Driving (CVPR 2024, Highlight)



Frontiers and Challenges

Summary

How to use world model for autonomous driving is still a big problem!

- **As simulator:** too complicated driving scenarios, hard to simulate.
- **Decision maker:** so slow inference, hard to make it real-time.
- **Foundation model:** Performance bottleneck?

One-page Takeaway

- Roadmap of predictive world model for autonomous driving
- Introduction of generation model & controllable future prediction
- Frontiers and challenges in utilizing predictive world models

Introduction to Generative Models

Reference

[1] Goodfellow, Ian, et al. "Generative adversarial nets." NIPS, 2014.
[2] Tim Salimans, et al. "Improved techniques for training gans." NIPS 2016.
[3] Martin Arjovsky and Léon Bottou. "Towards principled methods for training generative adversarial networks." arXiv preprint arXiv:1701.04862 (2017).
[4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein GAN." arXiv preprint arXiv:1701.07875 (2017).
[5] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028 (2017).
[6] Computing the Earth Mover's Distance under Transformations
[7] Wasserstein GAN and the Kantorovich-Rubinstein Duality
[8] Ferenc Huszár. "How (not) to Train your Generative Model: Scheduled Sampling, Likelihood, Adversary?." arXiv preprint arXiv:1511.05101 (2015).
[9] Jascha Sohl-Dickstein et al. "Deep Unsupervised Learning using Nonequilibrium Thermodynamics." ICML 2015.
[10] Max Welling & Yee Whye Teh. "Bayesian learning via stochastic gradient langevin dynamics." ICML 2011.
[11] Yang Song & Stefano Ermon. "Generative modeling by estimating gradients of the data distribution." NeurIPS 2019.
[12] Yang Song & Stefano Ermon. "Improved techniques for training score-based generative models." NeuriPS 2020.
[13] Jonathan Ho et al. "Denoising diffusion probabilistic models." arxiv Preprint arxiv:2006.11239 (2020).
[14] Jiaming Song et al. "Denoising diffusion implicit models." arxiv Preprint arxiv:2010.02502 (2020).
[15] Alex Nichol & Prafulla Dhariwal. "Improved denoising diffusion probabilistic models" arxiv Preprint arxiv:2102.09672 (2021).
[16] Yang Song, et al. "Score-Based Generative Modeling through Stochastic Differential Equations." ICLR 2021.
[17] Alex Nichol, Prafulla Dhariwal & Aditya Ramesh, et al. "GLIDE: Towards Photorealistic Image Generation and Editing with Text-Guided Diffusion Models." ICML 2022.
[18] Aditya Ramesh et al. "Hierarchical Text-Conditional Image Generation with CLIP Latents." arxiv Preprint arxiv:2204.06125 (2022).
[19] Salimans & Ho. "Progressive Distillation for Fast Sampling of Diffusion Models" ICLR 2022.
[20] Zhang et al. "Adding Conditional Control to Text-to-Image Diffusion Models." arxiv Preprint arxiv:2302.05543 (2023).
[21]. Yang Song's blog: https://yang-song.net/blog/
[22]. Lilian Wang's blog: https://lilianweng.github.io/

OpenDriveLab

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Q & A

End