World Models: Background, Applications and Opportunities

Presented by Sakura

October 10, 2025

bili_sakura@zju.edu.cn

Teaser: Plato's allegory of the cave. Image Credit: Wikipedia

Note: This keynote is still in progress.

Background and Related Works on World Modeling

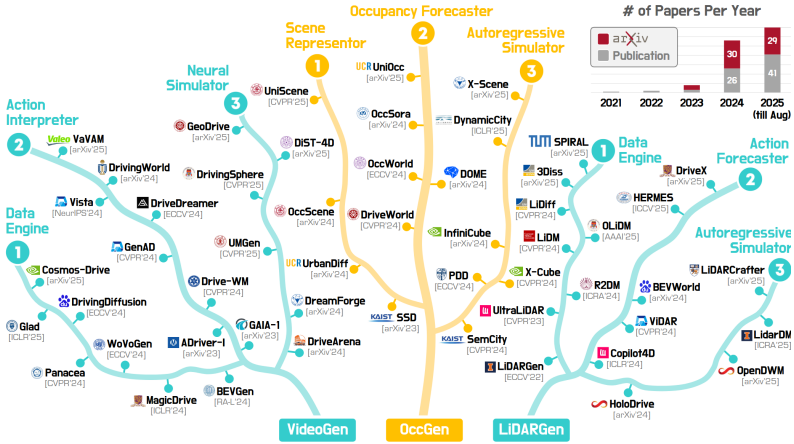


Figure: Summary of representative video-based generation (VideoGen), occupancy-based generation (OccGen), and LiDARbased generation (LiDARGen) models from existing literature. Image Credit: (Kong et al., 2025)

VideoGen: CogVideoX

Figure: CogVideoX (Yang et al., 2025) for Text-to-Video Generation.

Yang, et al. CogVideoX: Text-to-Video Diffusion Models with An Expert Transformer, ICLR, 2025.

Using CogVideoX for Image Editing

Intuition

Intuition of using large video generation model is that the sequence of frames in video clip reflect the real-world physics with strong **image consistency**.

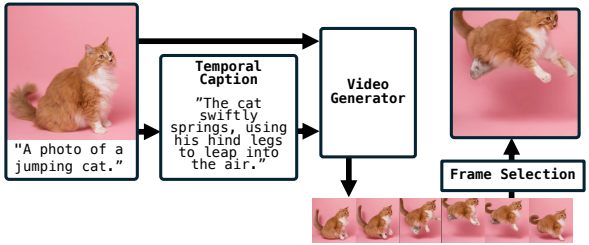
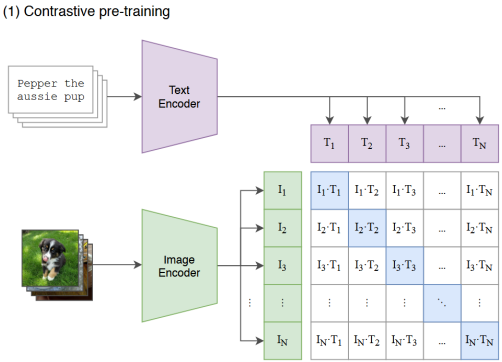


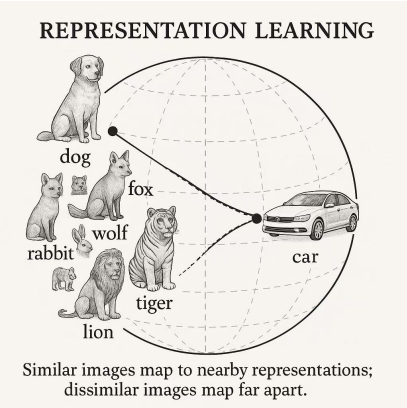
Figure: Frame2Frame (Rotstein et al., 2025) is a training-free method that uses a pretrained image-to-video diffusion model to synthesize a sequence of intermediate frames, and then selects the frame that best satisfies the edit.

Rotstein, et al. Pathways on the Image Manifold: Image Editing via Video Generation, CVPR, 2025.

Representation Learning



CLIP: Visual & Semantic Representation



Spherical latent space (Shared embedding)

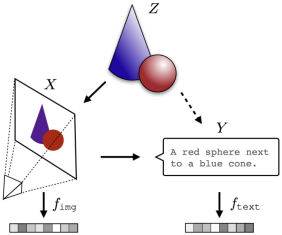
Figure: Overview of CLIP (Radford et al., 2021). A novel example of joint visual and semantic representation learning.

Radford, et al. Learning Transferable Visual Models From Natural Language Supervision, ICML, 2021.

Platonic Hypothesis

The Platonic Representation Hypothesis

Neural networks, trained with different objectives on different data and modalities, are converging to a shared statistical model of reality in their representation spaces.



Platonic solids (Abstraction of forms)

Plato's Allegory of the Cave

The Platonic Representation Hypothesis (Huh et al., 2024): Images (X) and text (Y) are projections of a common underlying reality (Z).

Huh, et al. Position: The Platonic Representation Hypothesis, ICML, 2024.