GenEx: Generative World Explorer

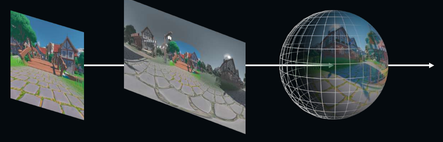
genex.world

Taiming Lu, Tianmin Shu, Junfei Xiao, Luoxin Ye, Jiahao Wang, Cheng Peng, Chen Wei,
Daniel Khashabi, Rama Chellappa, Alan Yuille, **Jieneng Chen**

Turn a single image into a world adventure.

World Initialization

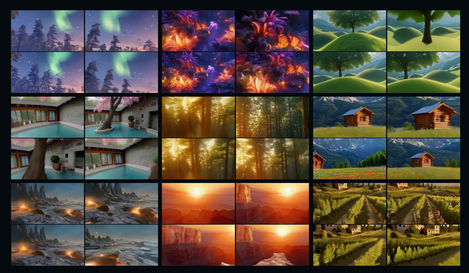
Single Image Input



World Exploration

Action Control

Diverse Generation

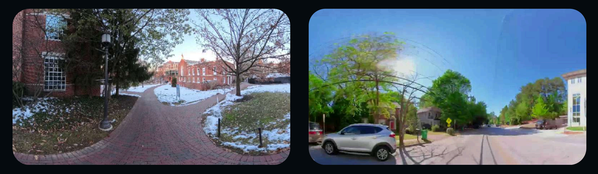


- Generative imagination guides exploration, forming priors of unseen environments.
- Builds 3D-consistent worlds from a single RGB image, generating panoramic video.
- Maintains loop consistency, preserving coherence over long trajectories.
- Enables active 3D mapping, refining beliefs and predicting unseen regions.
- Supports both goal-driven navigation and open-ended exploration for embodied AI.

Dataset Curation

UNREAL ENGINE Unity

Hand-held Collections Web Videos

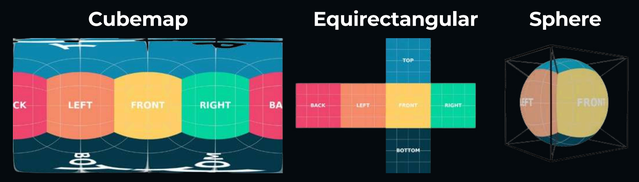


Our data curation leverages physical engines, utilizing realistic city assets from UE5 and animated world assets from Unity. We also collect real-world videos from hand-held cameras and mining from web.

ICLR International Conference On Learning Representations

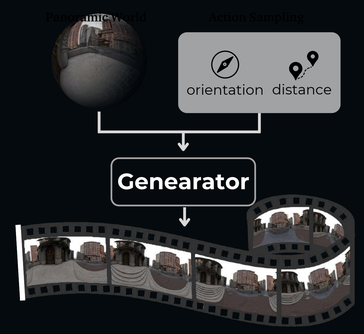
Scan here to follow on X

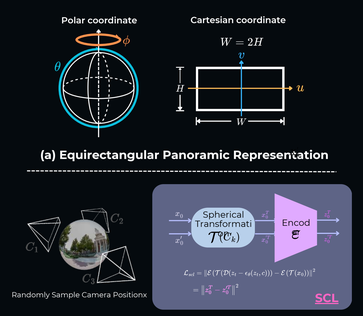
1 World Initialization



- We represent the 360° world using the panoramic view of the agent. Panoramic images capture a complete 360° × 180° view of a scene from a fixed viewpoint.
- Our world initialization model builds on a SOTA text-to-panorama model, tuned from the text-to-image model FLUX.1.
- We extend it to condition on both text and an image, enabling the generation of a coherent 360-degree environment.

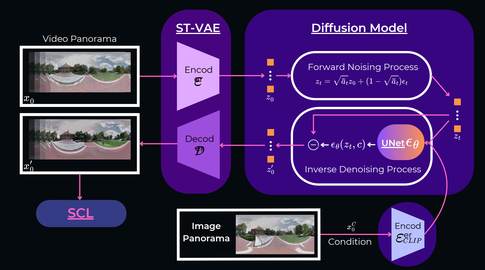
2 World Transition





(a) Equirectangular Panoramic Representation

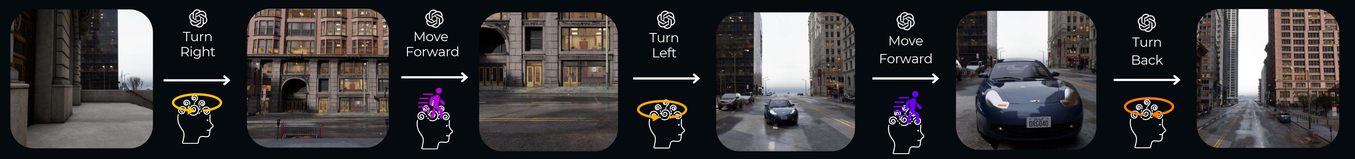
(b) Spherical-Consistent Learning (SCL)



(c) Spherical-Consistent Egocentric Panoramic Video Diffusion Model

3 World Exploration

Instruction: “Plan to move to the position of the blue car, then turn back.”



- The agent follows a goal-driven navigation instruction, where GPT plans high-level actions, and GenEx iteratively refines exploration, updating images step-by-step for controlled and targeted navigation.

Advancing Embodied AI

In our generative world, we explore unseen regions, gather comprehensive information, and refine beliefs for informed decision-making, framing this as an "**imagination-augmented policy**" that shapes the future of embodied AI.