



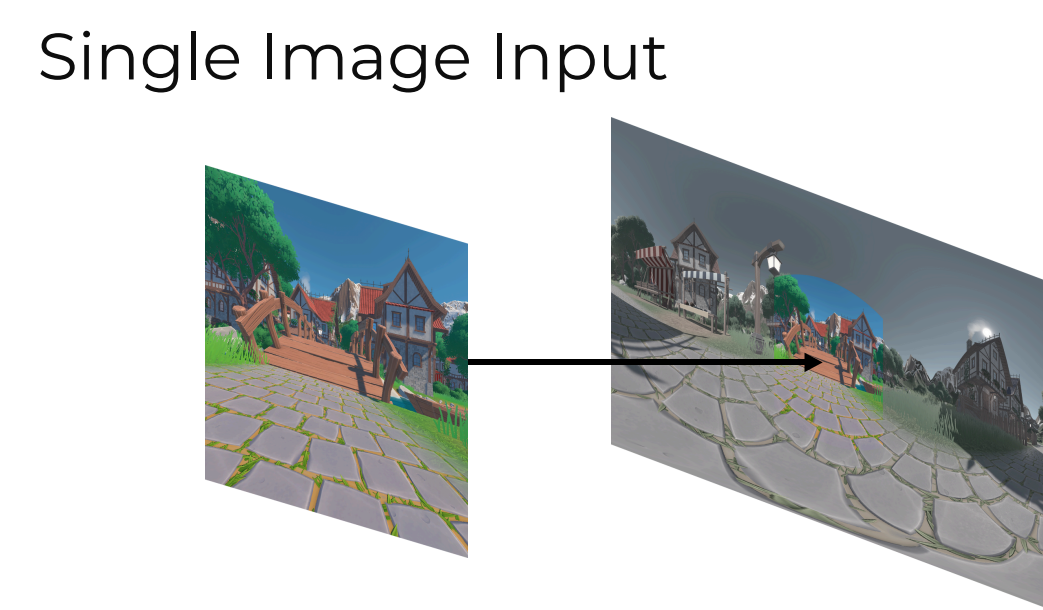
GenEx: Generative World Explorer

genex.world

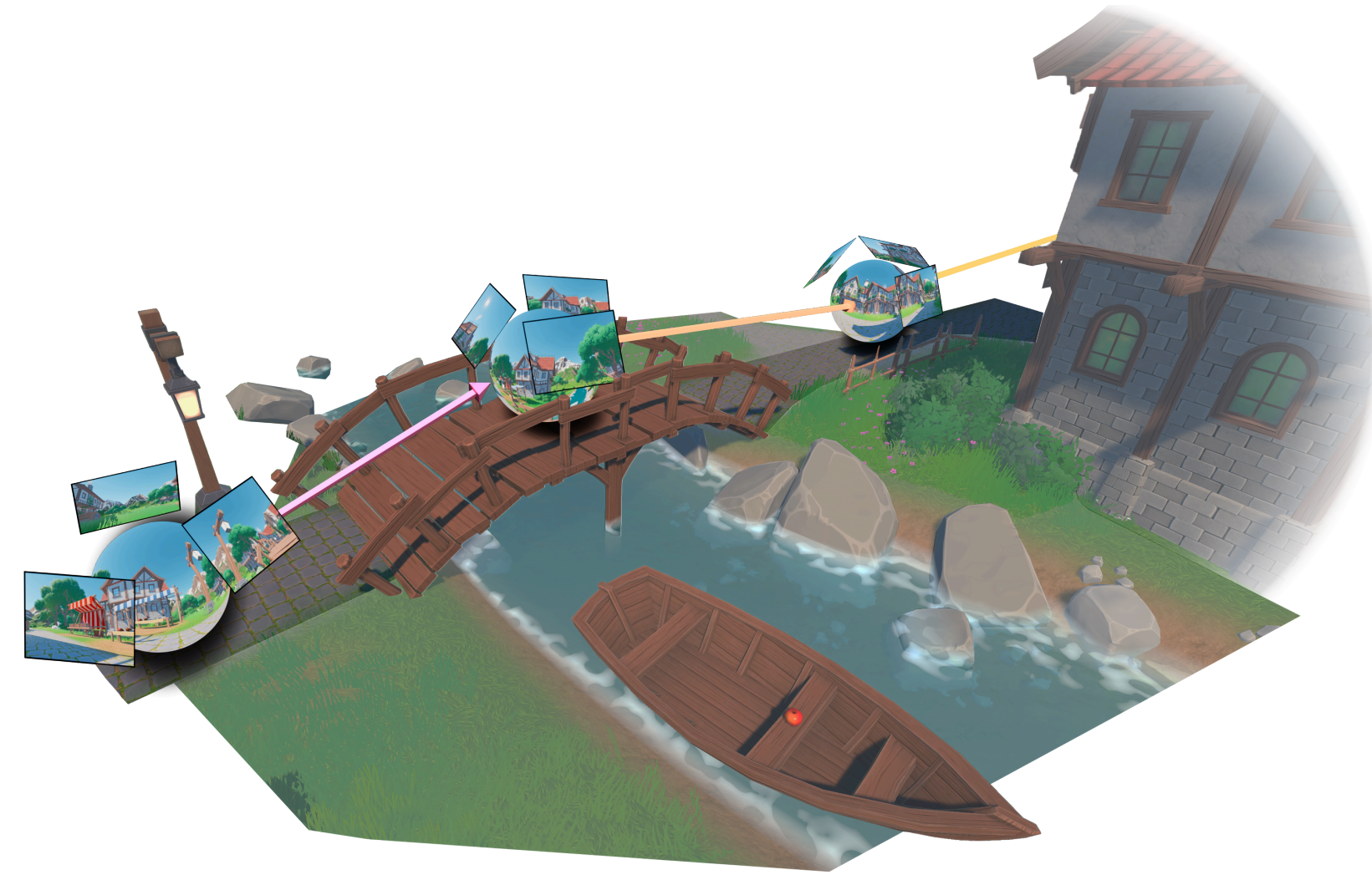
Taiming Lu, Tianmin Shu, Junfei Xiao, Luoxin Ye, Jiahao Wang, Cheng Peng, Chen Wei, Daniel Khashabi, Rama Chellappa, Alan Yuille, **Jieneng Chen**

Turn a single image into a world adventure.

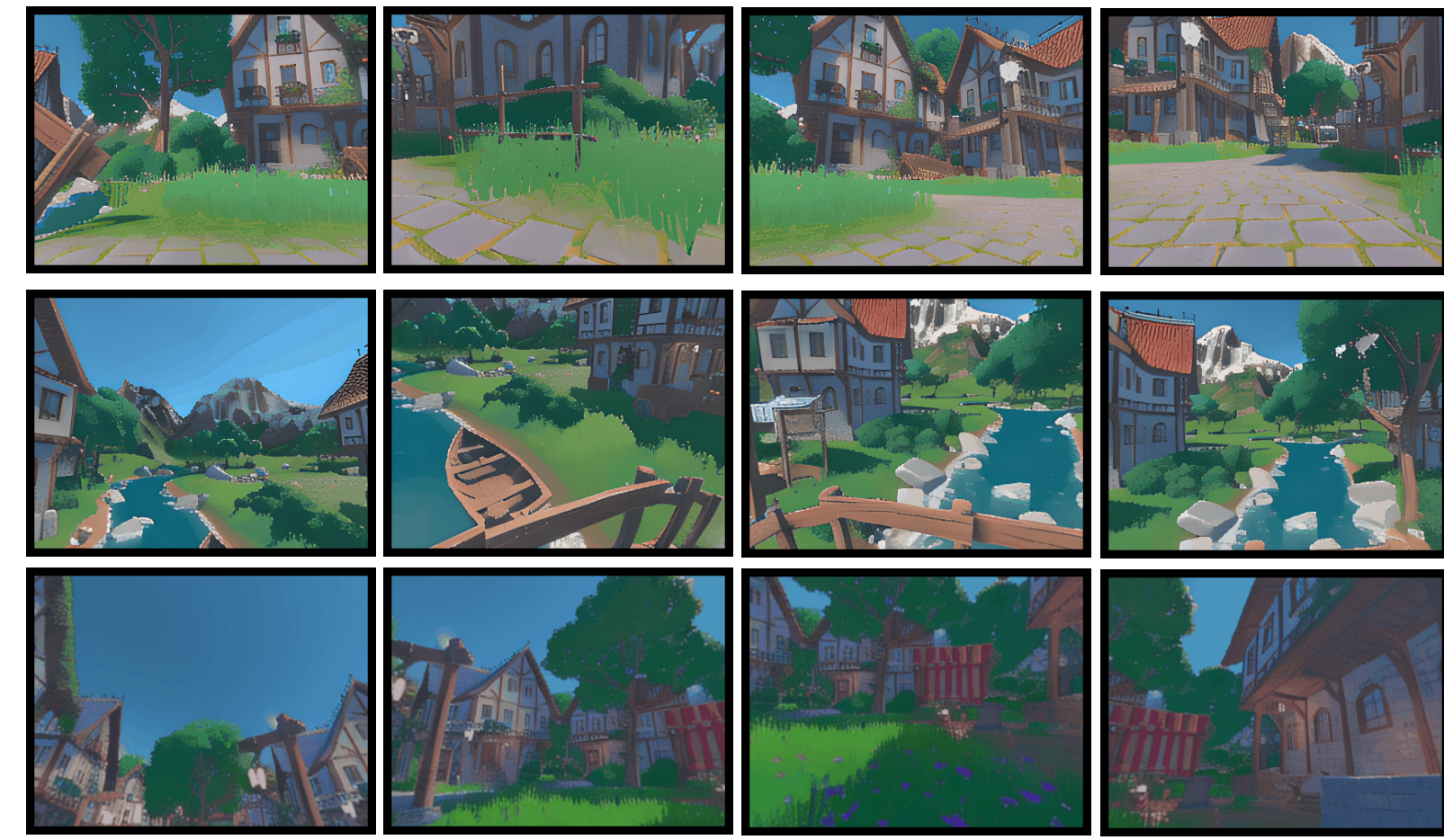
World Initialization



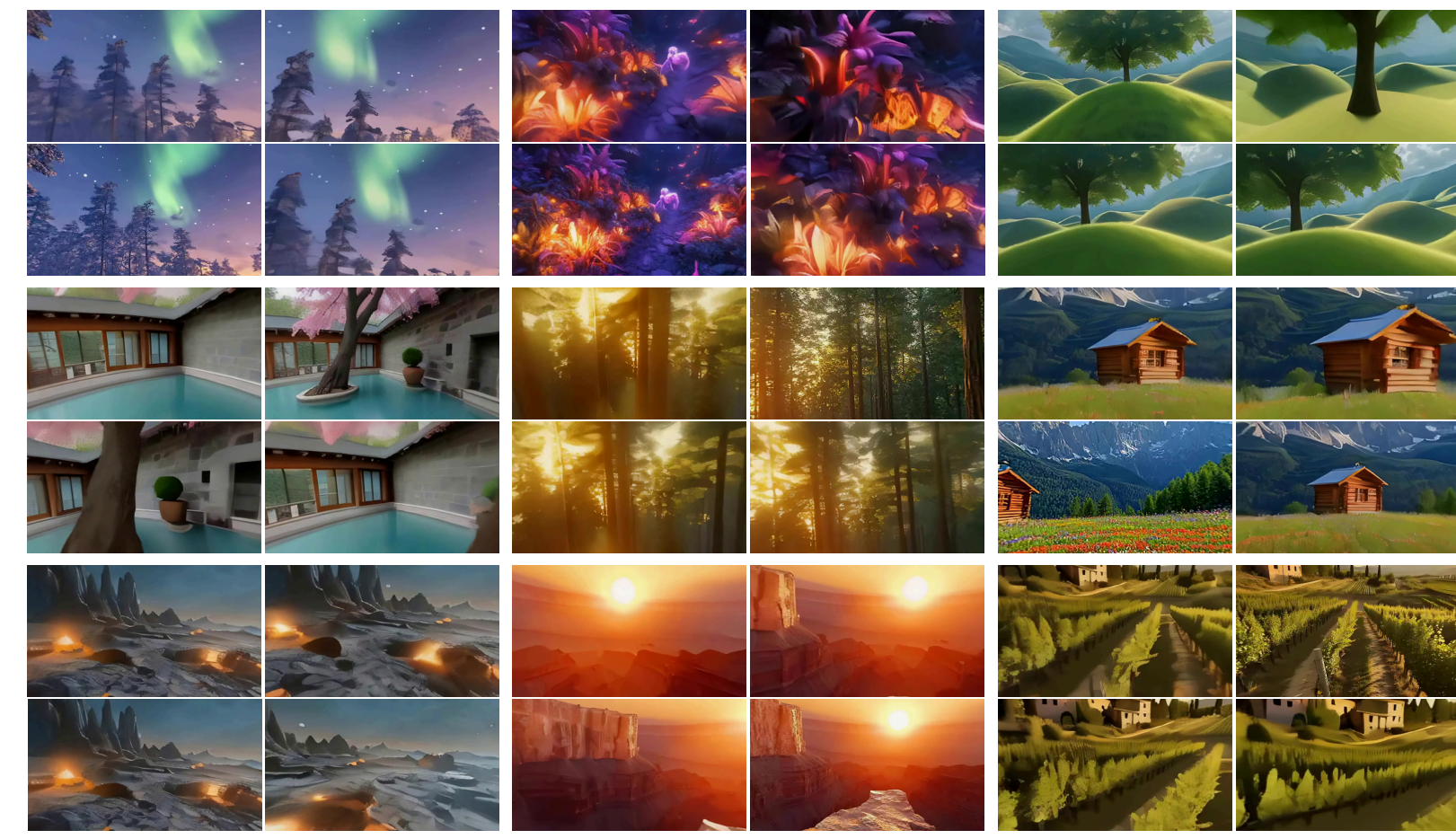
World Exploration



Action Control

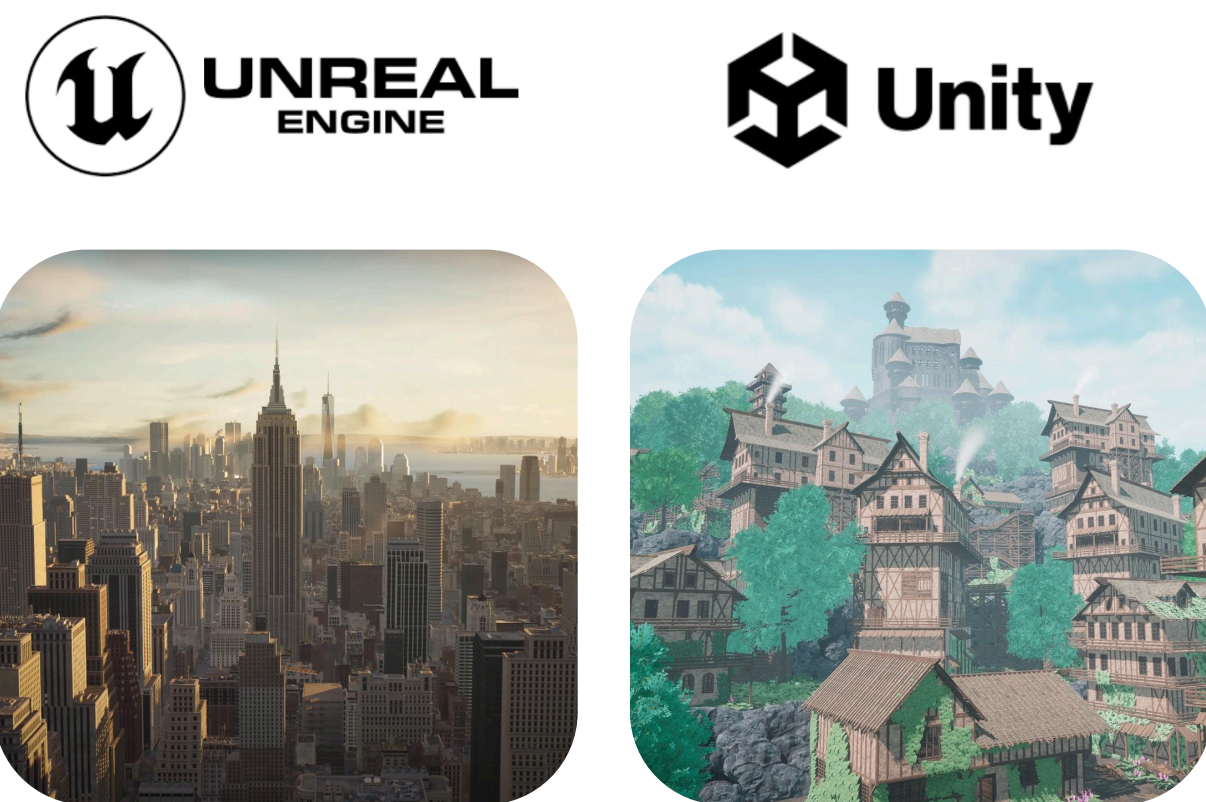


Diverse Generation



- 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



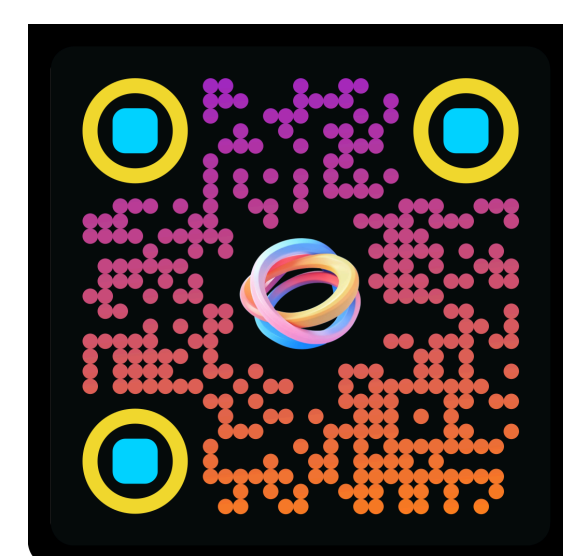
Hand-held Collections



Web Videos

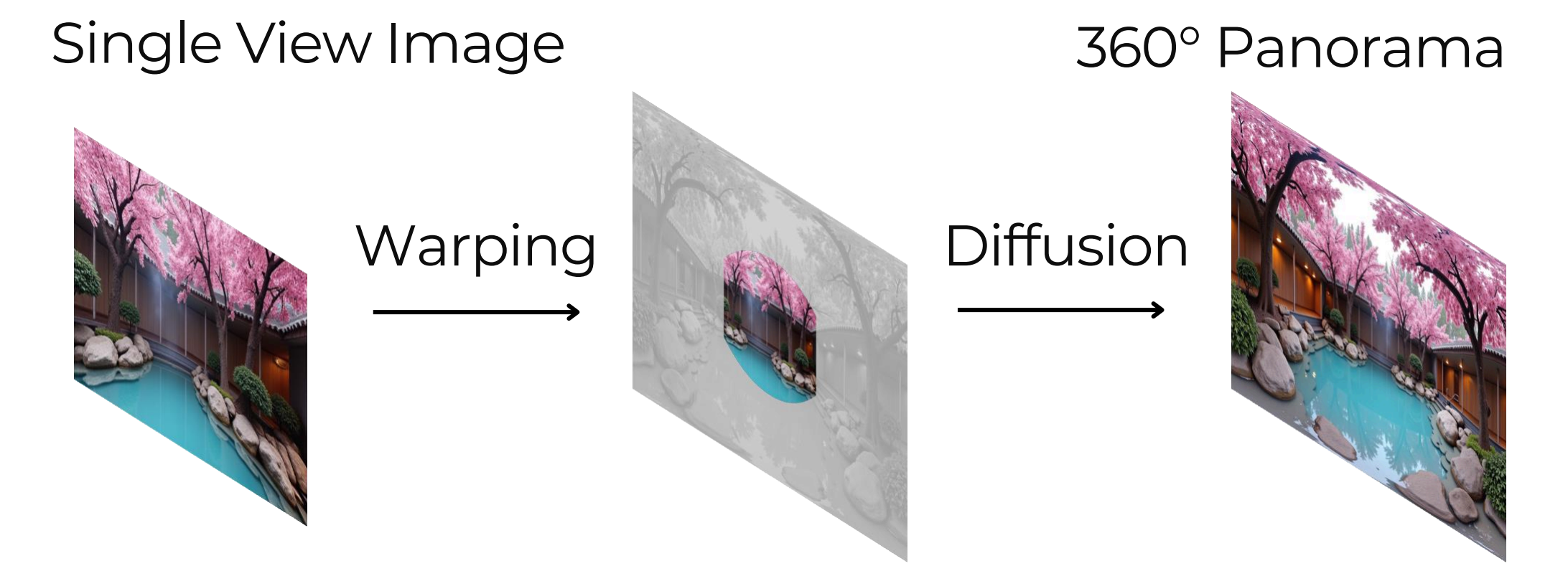
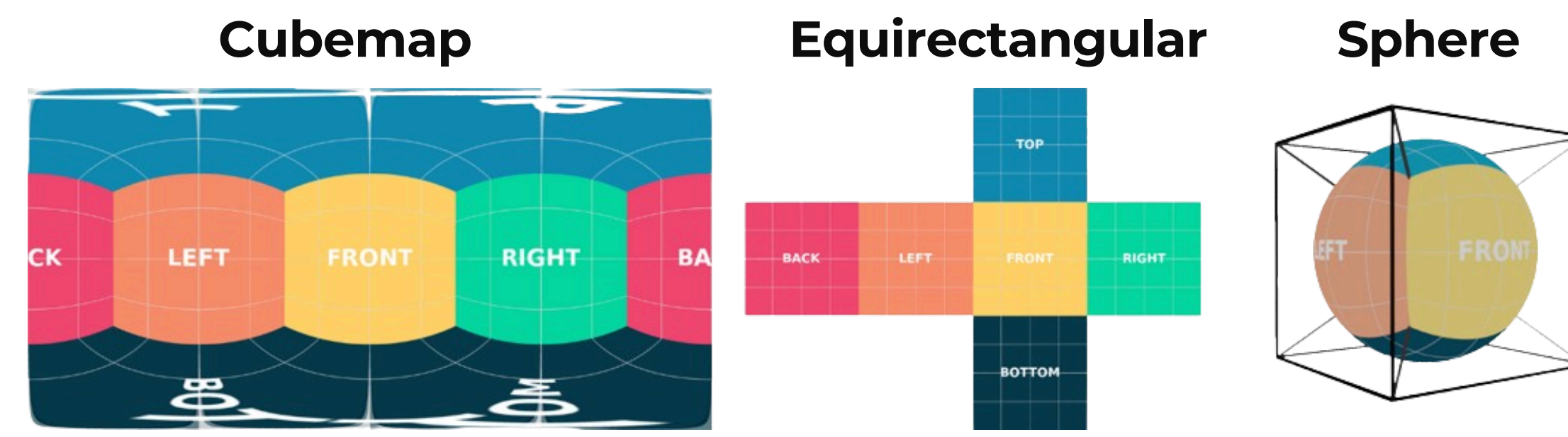


Our data curation leverages physical engines, utilizing realistic city assets from UE5 and animated world assets from Unity. We also collect real-world videos from hand-held cameras and mining from web.



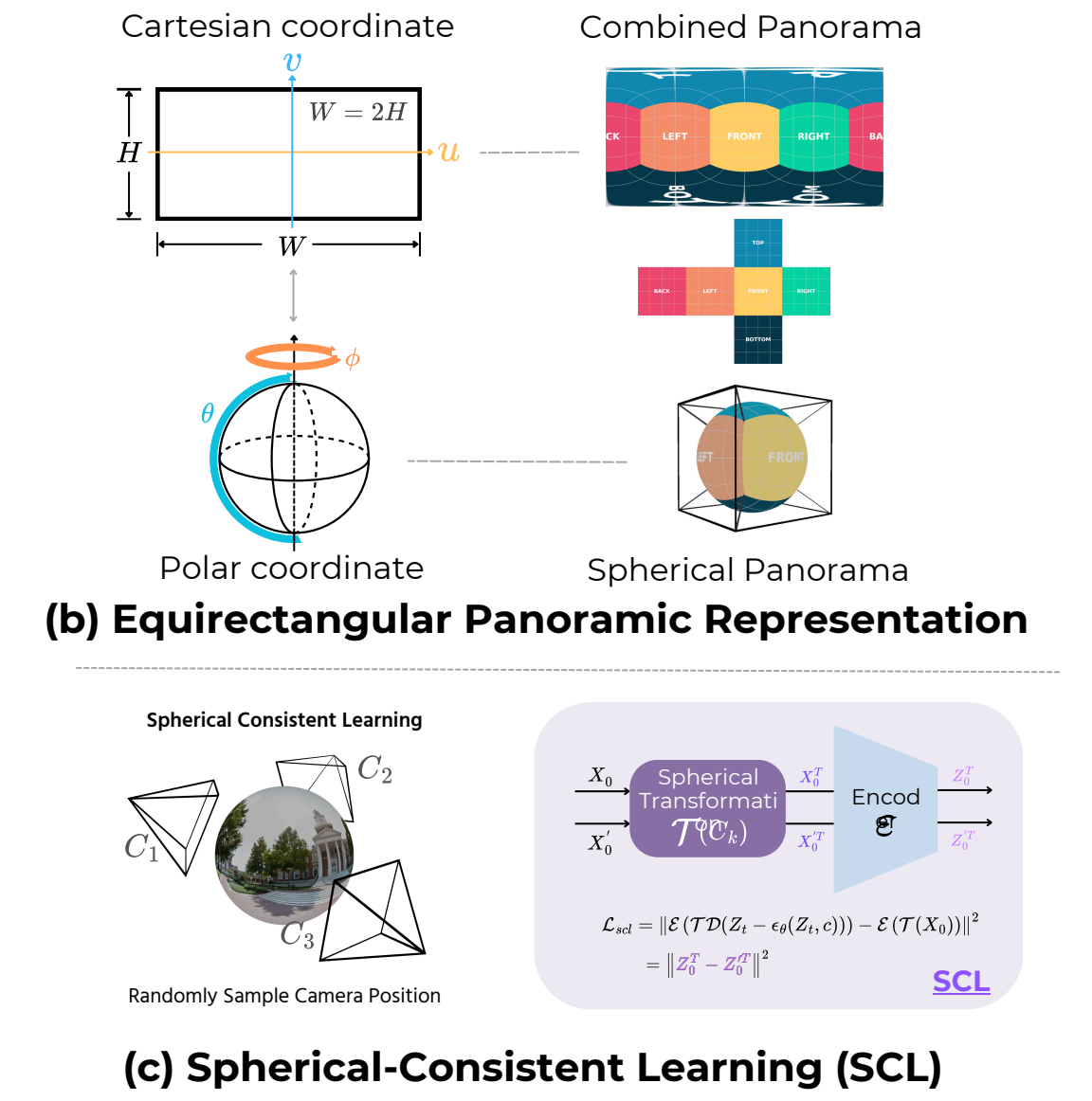
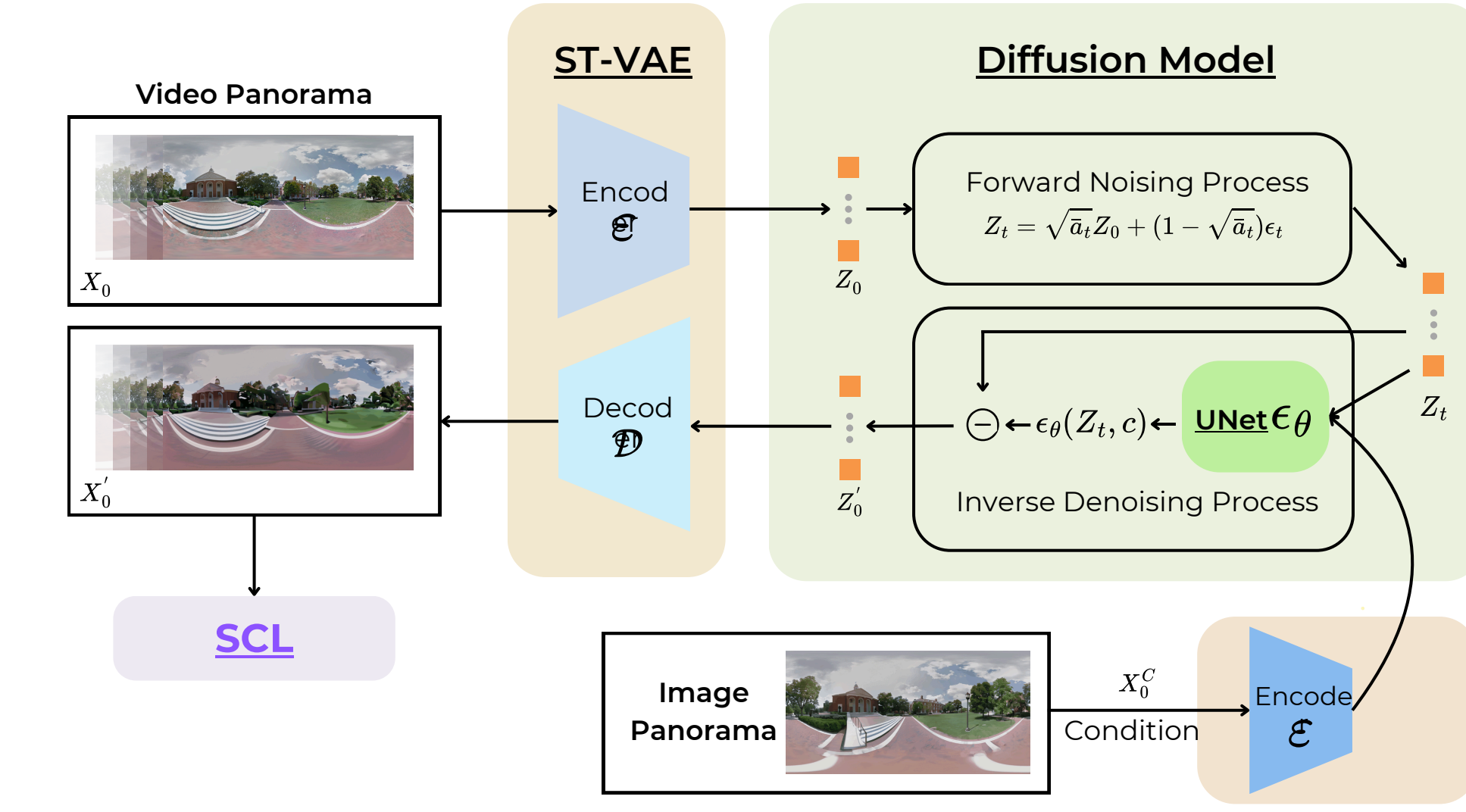
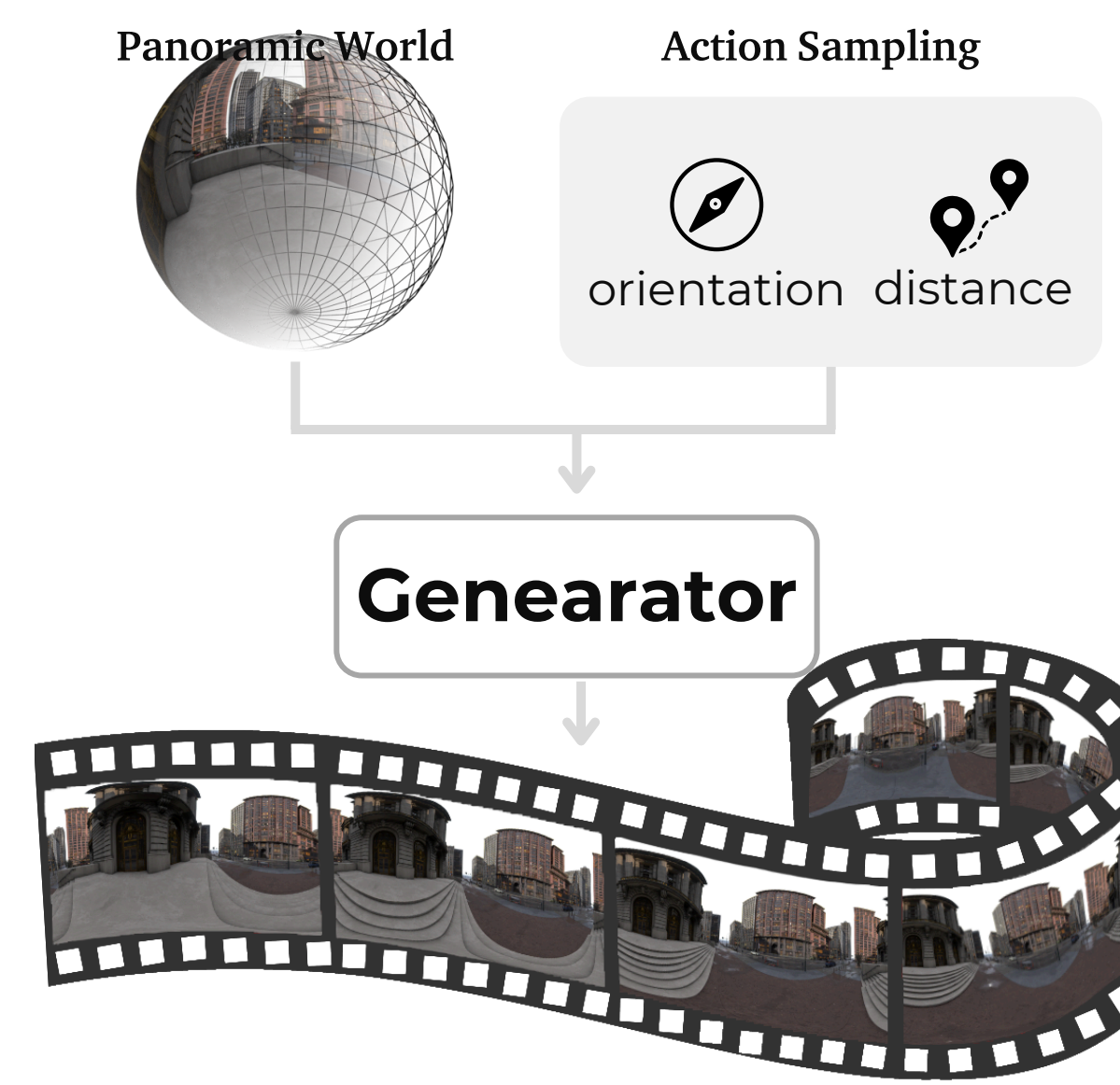
Scan here to follow on X

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

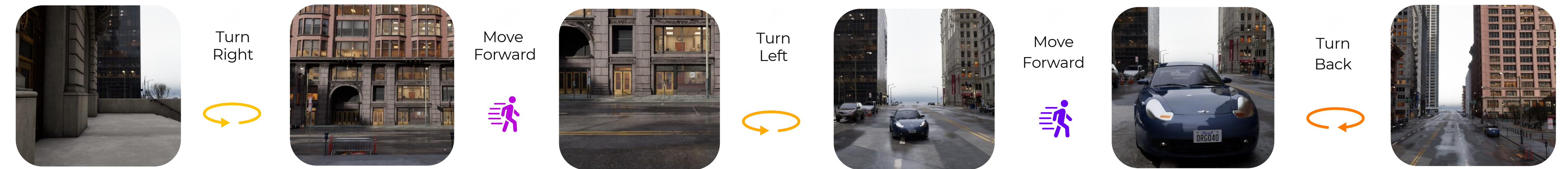


(a) Spherical-Consistent Egocentric Panoramic Video Diffusion Model

(c) Spherical-Consistent Learning (SCL)

3 World Exploration

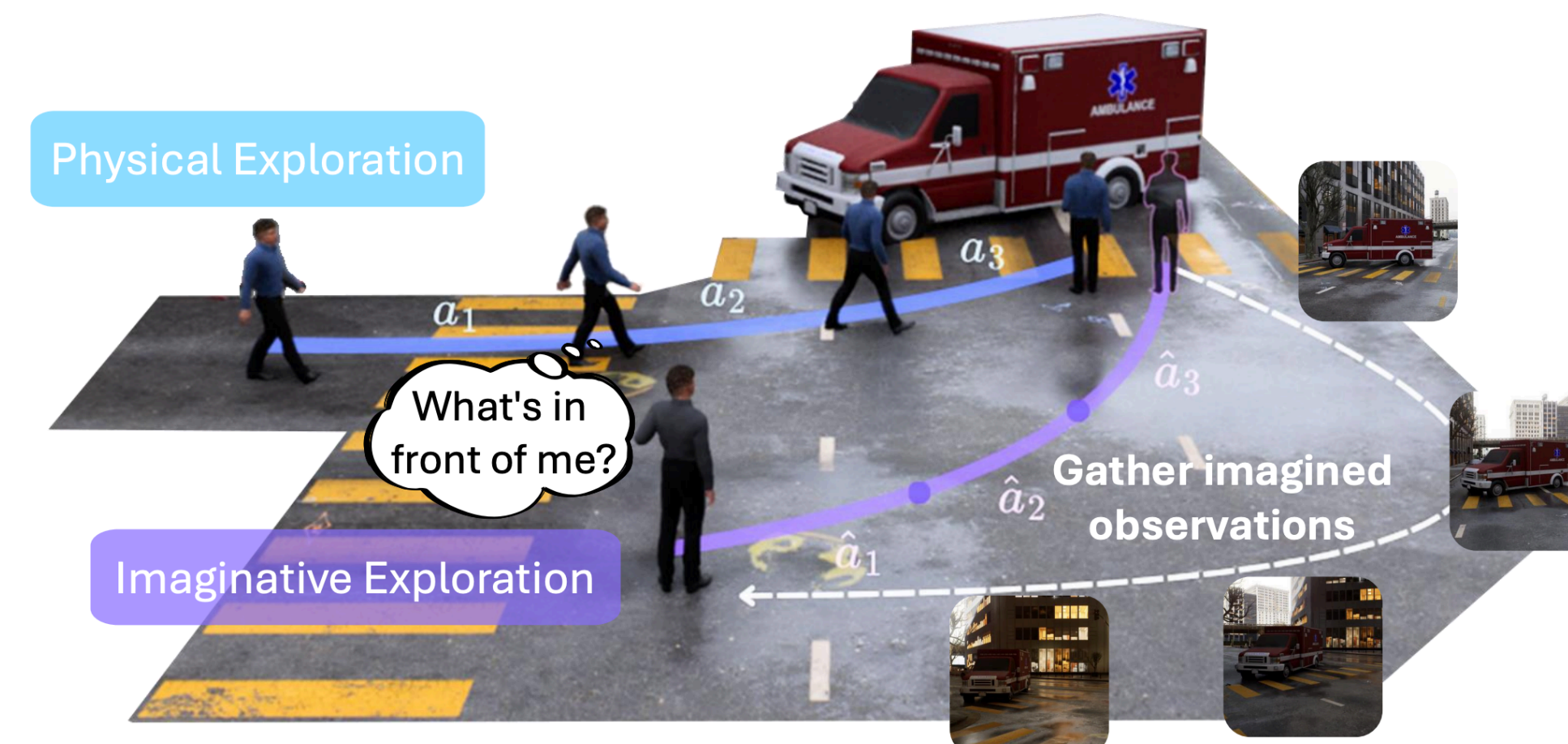
Instruction: "Plan to move to the position of the blue car, then turn back."



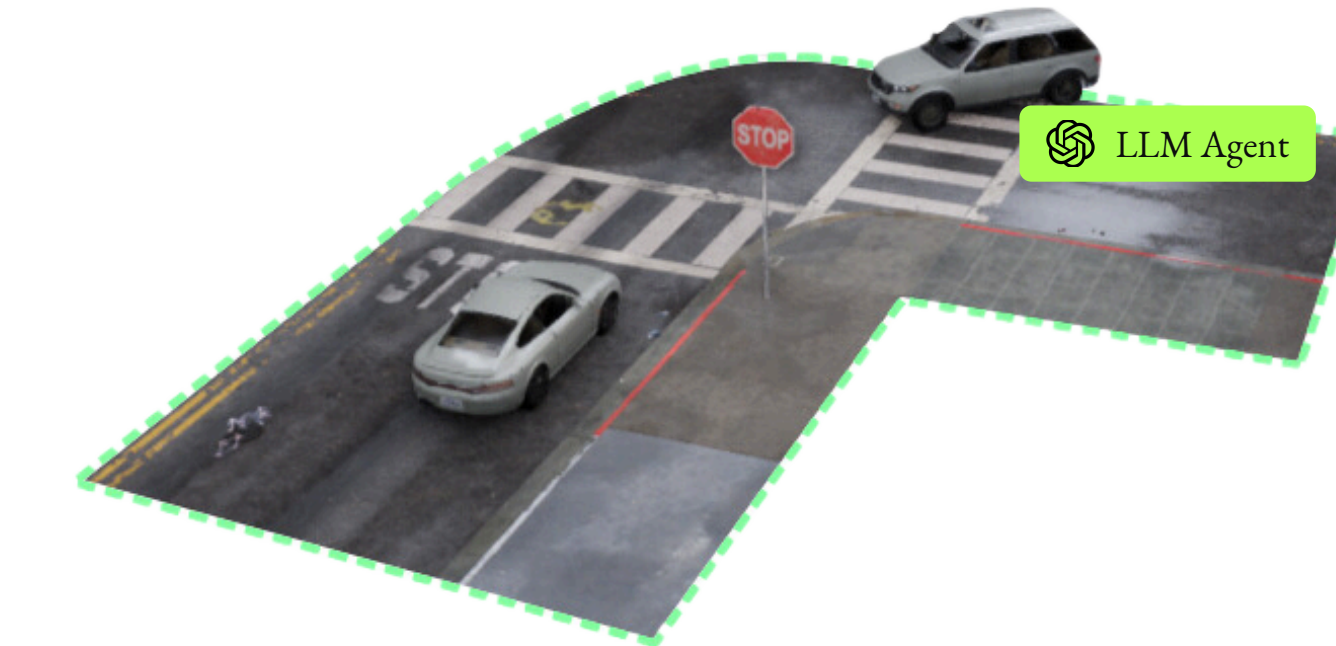
- The agent follows a goal-driven navigation instruction, where GPT plans high-level actions, and GenEx iteratively refines exploration, updating images step-by-step for controlled and targeted navigation.

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.



(a) Single-Agent



Observation

I'm turning left at an intersection with no traffic lights. A silver car is slowly moving ahead, and I'm unsure if it will stop. Should I wait?

Generative World Exploration Imagined Observation

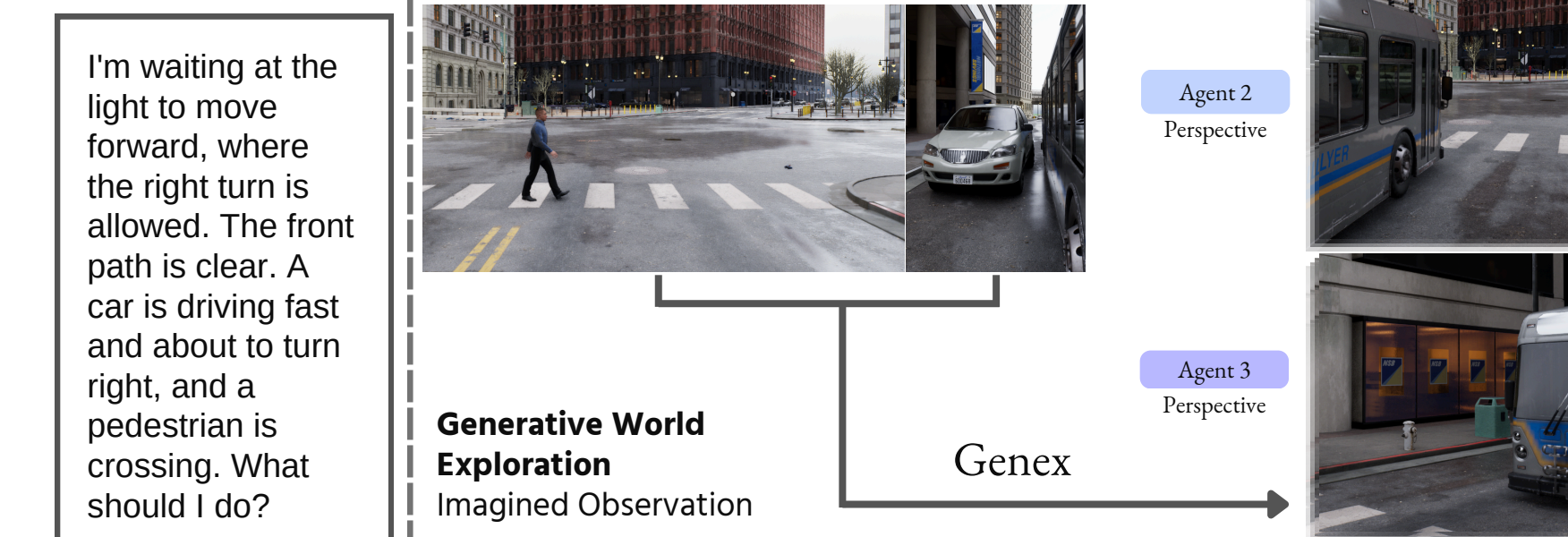
I should stop to avoid a potential collision, as the car might not stop.

The car sees a stop sign and will stop, so I should move to avoid blockage.

Egocentric Single-View Decision: Stop in place (X)

Decision with Imagination: Continue driving (✓)

Observation



I'm waiting at the light to move forward, where the right turn is allowed. The front path is clear. A silver car is driving fast and about to turn right, and a pedestrian is crossing. What should I do?

Generative World Exploration Imagined Observation

I want to drive forward, but the light is red, so I should wait in place.

I'm blocking the view between the car and pedestrian, and they might collide.

Egocentric Single-View Decision: Stop in place (X)

Decision with Imagination: Warn both parties (✓)

(b) Multi-Agent

