
Image as a World: Generating Interactive World from Single Image via Panoramic Video Generation

Supplementary Materials

Notes: More videos are in the supplementary material and best viewed with a 360 video player.

A Parameters Setup

We present the hyperparameters used in our experiment in Table A. The same hyperparameters are applied to both IaaW-IM and IaaW-CM, and our pipeline is built upon the CogVideoX codebase

Parameters	Value
Video Height	1024
Video Width	2048
Video Frames	49
Batch Size	4
Mixed Precision	bf16
Optimizer	AdamW
Optimizer Betas	(0.9, 0.95)
Optimizer Weight Decay	1e-4
Learning Rate	2e-5
Warmup Steps	100
Inference Steps	25
Frame Per Second	16
Masks Number	6

Table A: Hyperparameters setting of our experiments.

B Masks Setup

We present our masks setup in Fig. A, which contains six perspectives of one panoramic image/video and these can together seamlessly reconstruct the full panoramic scene.

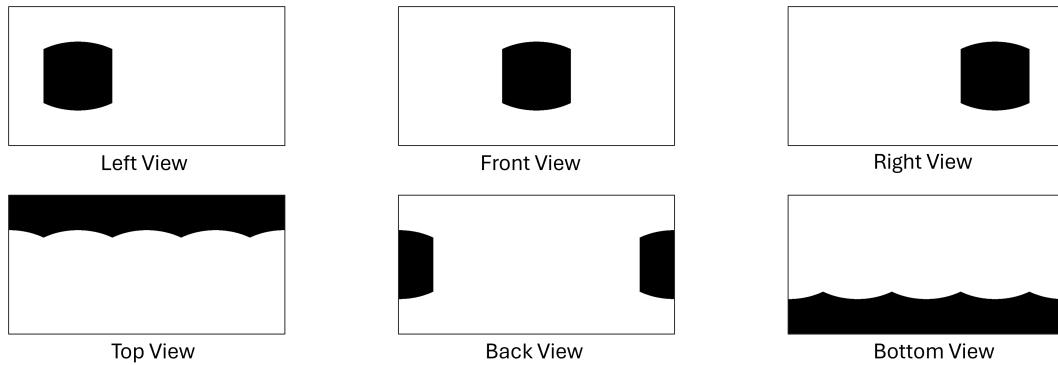


Figure A: Different view masks setup in our method.

C More Visualization Results

We present additional visualization results of our world initialization in Fig. B, which demonstrate that our method exhibits significant diversity across various scenarios.

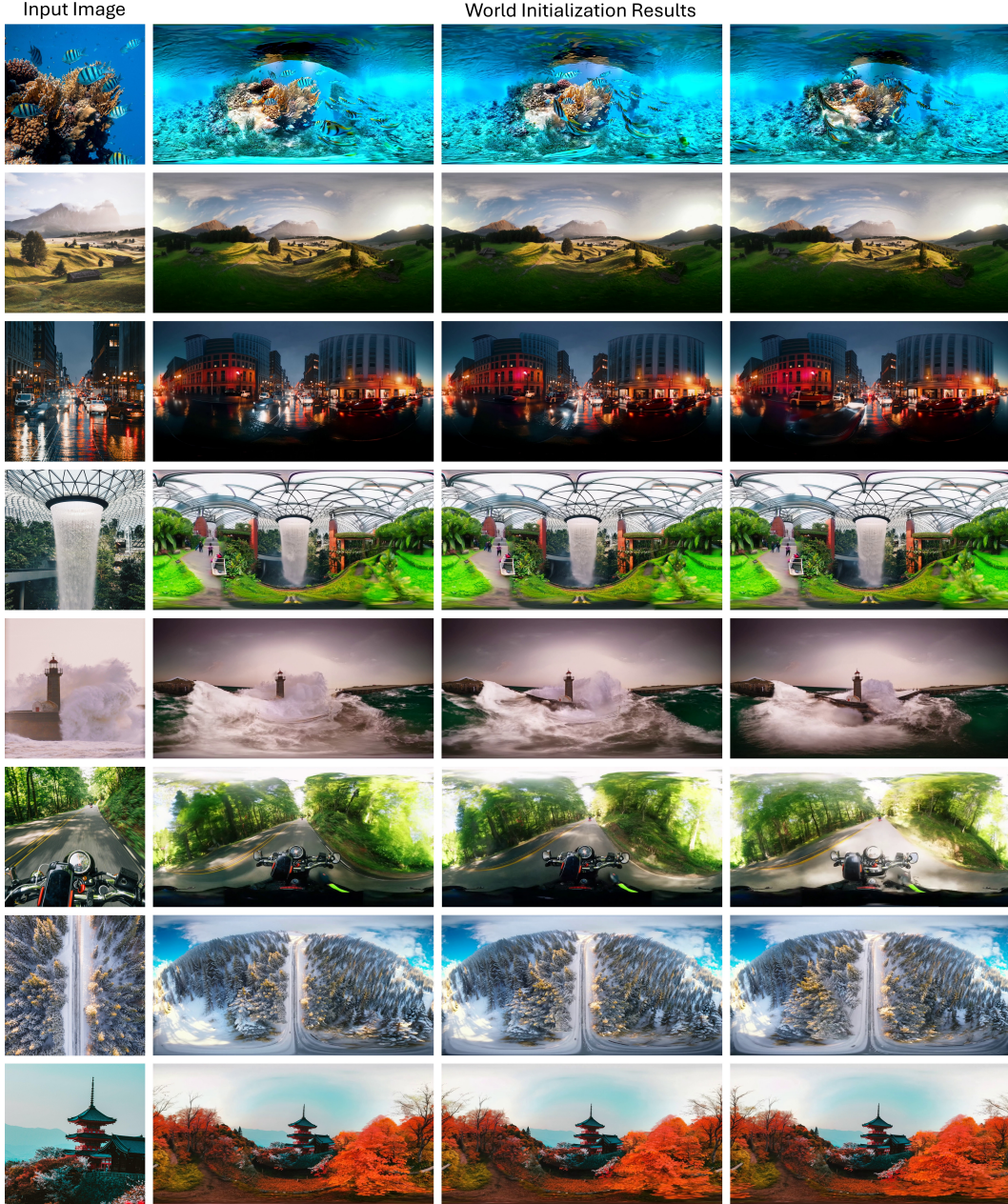


Figure B: More visualization results of world initialization.