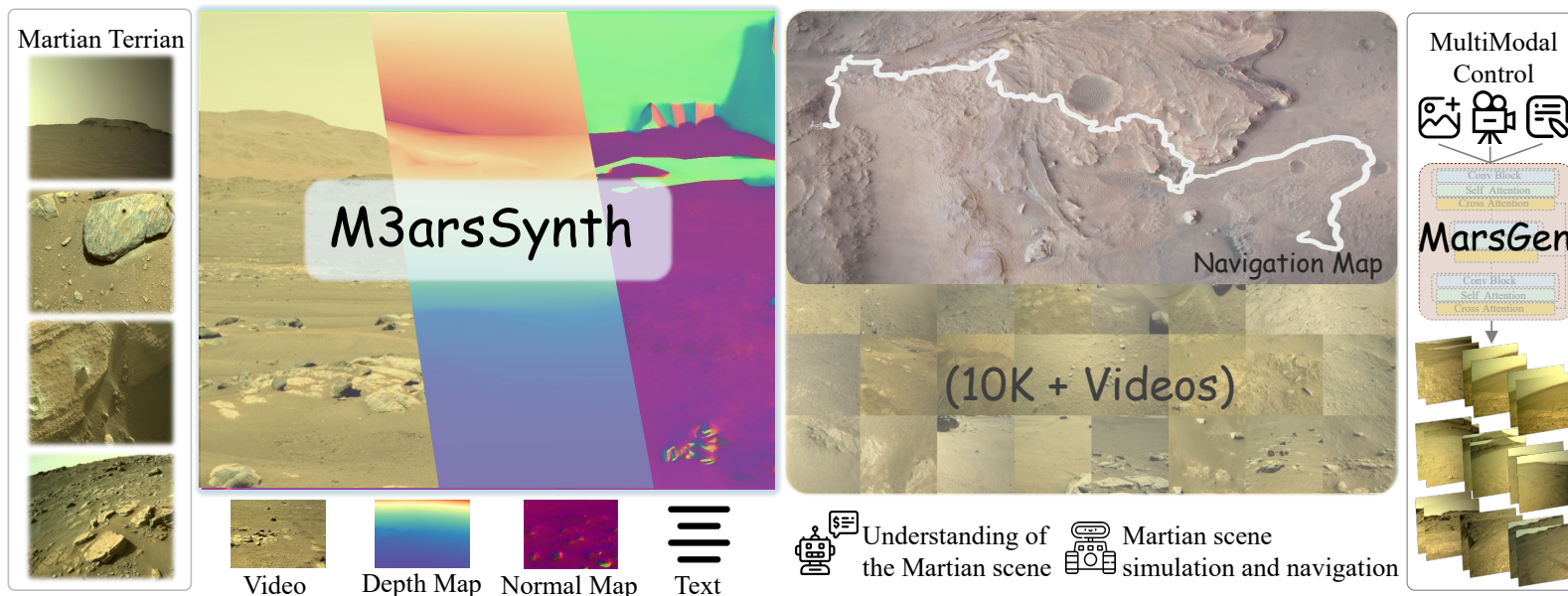


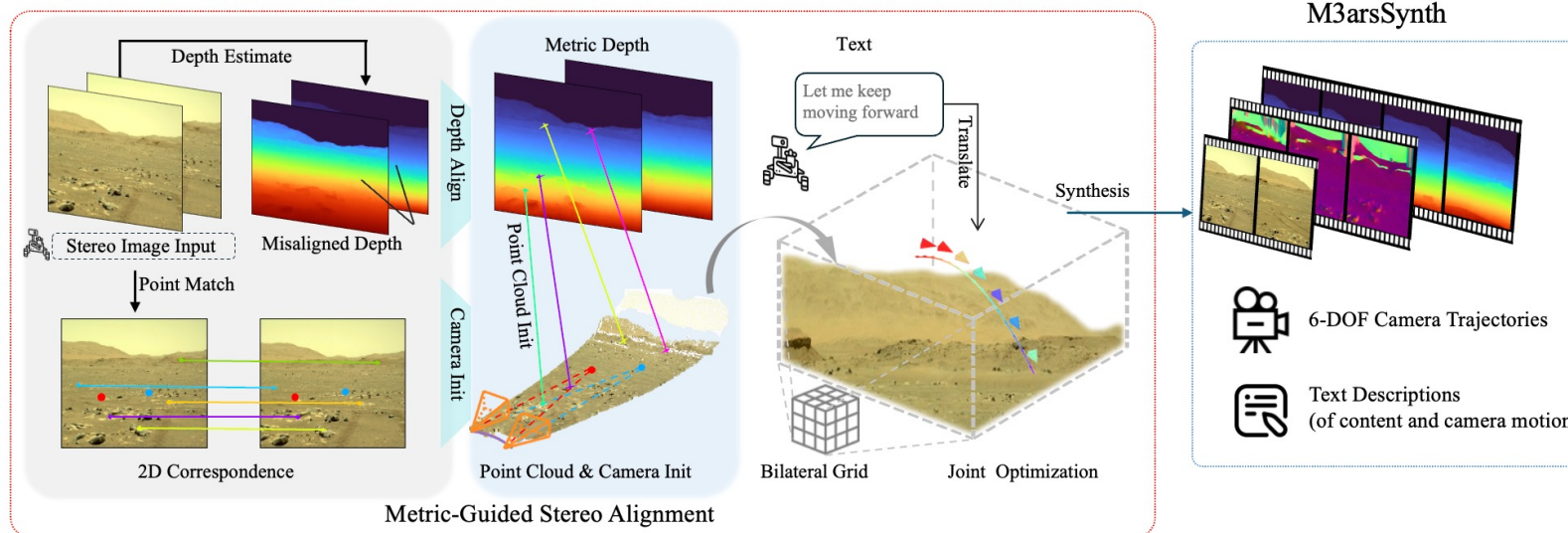


# Martian World Model: Controllable Video Synthesis with Physically Accurate 3D Reconstructions

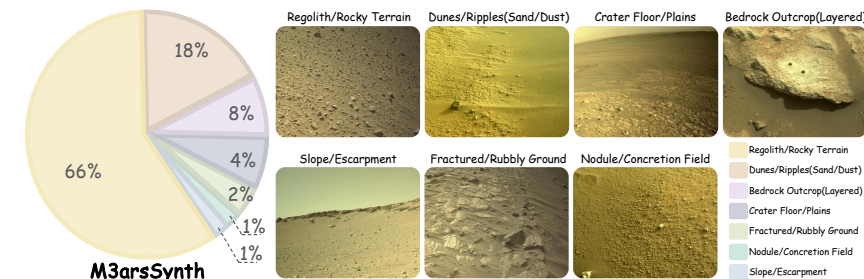
Longfei Li, Zhiwen Fan, Wenyan Cong, Xinhang Liu, Yuyang Yin, Matt Foutter, Panwang Pan, Chenyu You, Yue Wang, Zhangyang Wang, Yao Zhao, Marco Pavone, Yunchao Wei



## ➤ M3arsSynth (Data Engine)



## ➤ Dataset Distribution



Distribution of primary terrain types within the dataset, showcasing the diversity of Martian environments covered.

## ➤ Performance

Model	Visual Fidelity		3D Consistency	Novel View Synthesis		
	FID ↓	FVD ↓	Warp Err ↓	PSNR ↑	SSIM ↑	LPIPS ↓
<i>Image-to-Video</i>						
Pyramidal-Flow <a href="#">Jin et al. (2024)</a>	78.495	637.952	17.930	—	—	—
CogVideoX <a href="#">Yang et al. (2024)</a>	48.912	411.808	6.866	—	—	—
Kling <a href="#">Brooks et al. (2024)</a>	74.632	727.130	24.793	—	—	—
Sora <a href="#">Brooks et al. (2024)</a>	142.954	823.418	10707.713	—	—	—
<i>Camera Control Image-to-Video</i>						
CameraCtrl <a href="#">He et al. (2024)</a>	123.386	772.476	17.410	20.014	0.288	0.408
ViewCrafter <a href="#">Yu et al. (2024)</a>	169.942	2297.899	501.734	13.143	0.262	0.586
Ours	<b>38.779</b>	<b>364.822</b>	<b>6.071</b>	<b>21.239</b>	<b>0.336</b>	<b>0.351</b>

- Models trained on Earth data produce thematically inconsistent or uncontrolled videos.
- MarsGen generates coherent Martian landscapes that follow camera controls.

## Overview of the M3arsSynth dataset construction.

We process stereo image pairs using a metric-aware foundation model and solve the Perspective-n-Point (PnP) problem to reconstruct metric-scale 3D Martian scenes.