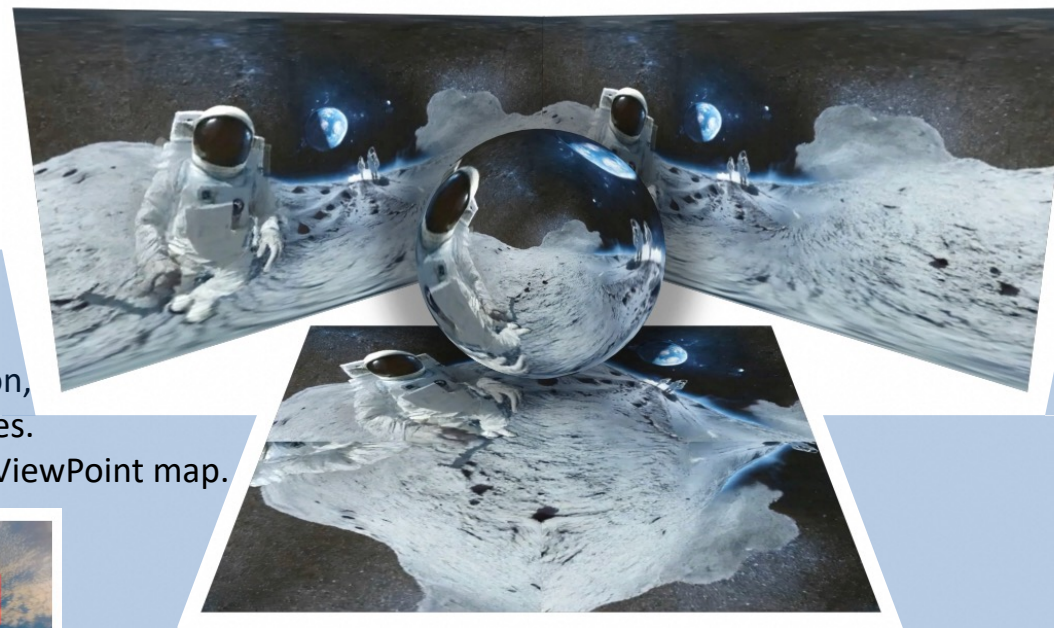


Video/Text to Panoramic Video Generation

Our model extends a perspective video into a panorama, or generates a panorama directly from text.

denotes the input video.



Motivation

Due to the scarcity of panoramic data and the modality gap introduced by the equirectangular projection (ERP) representation, existing methods often struggle with severe distortion at the poles. To alleviate this issue, we propose a novel representation called ViewPoint map.

Pipeline

We first remap the ERP or CP representation into a ViewPoint map and then propose a Pano-Pers attention mechanism to enhance the generation quality.

Evaluation

Method	subject consistency↑	imaging quality↑	motion smoothness↑	dynamic degree↑
Follow-Your-Canvas [5]	0.8284	0.4464	0.9655	0.8500
360DVD [39]	0.8633	0.5394	0.9703	0.5083
Imagine360 [33]	0.8547	0.5859	0.9720	0.8148
ViewPoint(AnimateDiff [9])	0.8681	0.5914	0.9786	0.8633
ViewPoint(Wan2.1 [35])	0.8793	0.5927	0.9800	0.9083

We evaluate our method on VBench across 4 dimensions.

zxfang1130@gmail.com

