

PanoDiT: Panoramic Videos Generation with Diffusion Transformer (Supplementary Material)

Anonymous submission

In this section, we present the qualitative experimental results for DTM-LoRA.

Qualitative Ablation for DTM-LoRA

In the main text, we quantitatively verify that the weights of DTM-LoRA are positively correlated with the motion scores of the output videos. Due to space constraints, we present the qualitative experiments in this section.

By fixing the prompt, we eliminated its potential impact on the motion patterns in the video. The model used is PanoDiT-B, with the resolution set to the same as in the main text, in a 512×1024 format. The total output consists of 144 frames at a rate of 24 FPS.

As shown in Figures 1 and 2, the results without DTM-LoRA exhibit chaotic motion outputs, demonstrating that our DTM-LoRA is highly effective in generating stable and directed ERP video motion.

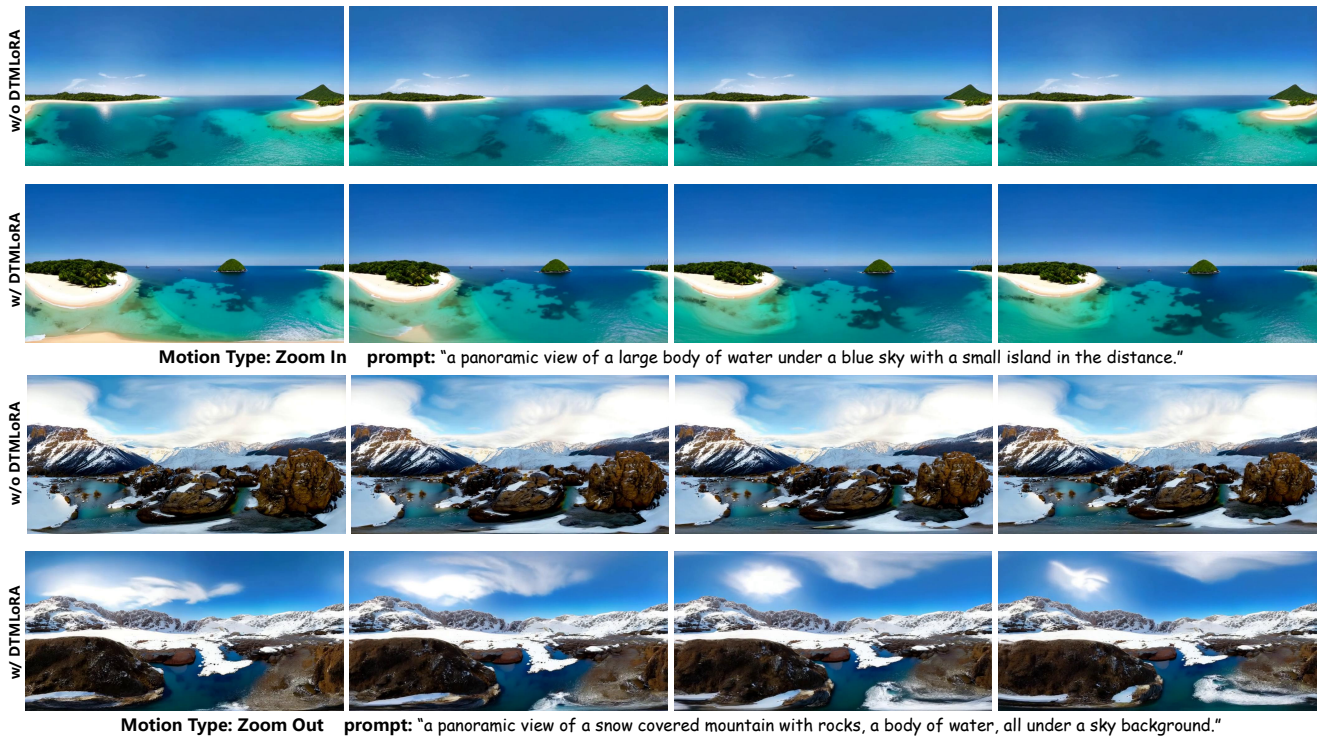


Figure 1: The figure shows the ablation experiment for DTM-LoRA, with the LoRA weight set to 0.8. The two images illustrate the zoom effect caused by forward and backward camera movements.

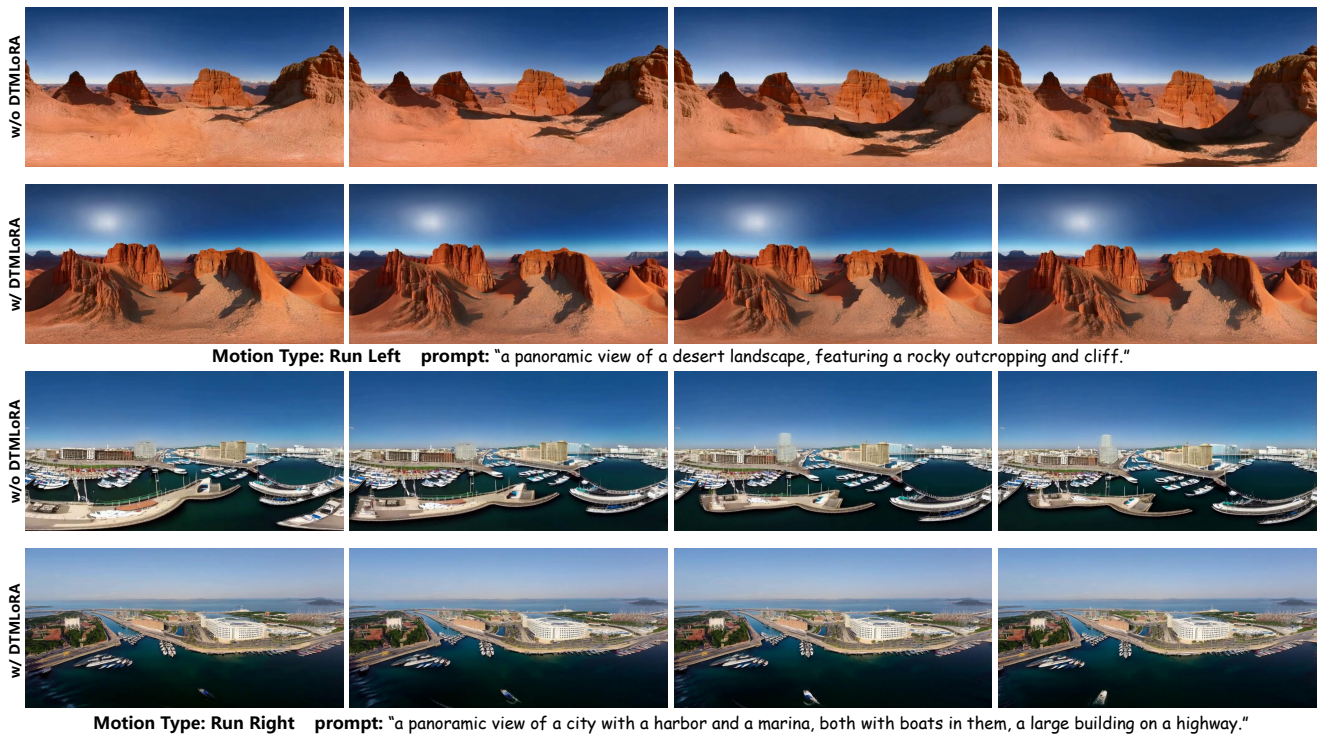


Figure 2: More result of ablation experiment for DTM-LoRA.