

426 A Visualization of the generated videos

427 We provide visualization comparison between SAPAttn and Dense Attention on HunyuanVideo and
 428 Wan 2.1. Results in Figure 9 and Figure 10 demonstrate that SAPAttn can preserve high pixel-level
 429 fidelity, achieving similar generation quality compared with the dense attention. Real video samples
 430 are provided in the supplementary materials.

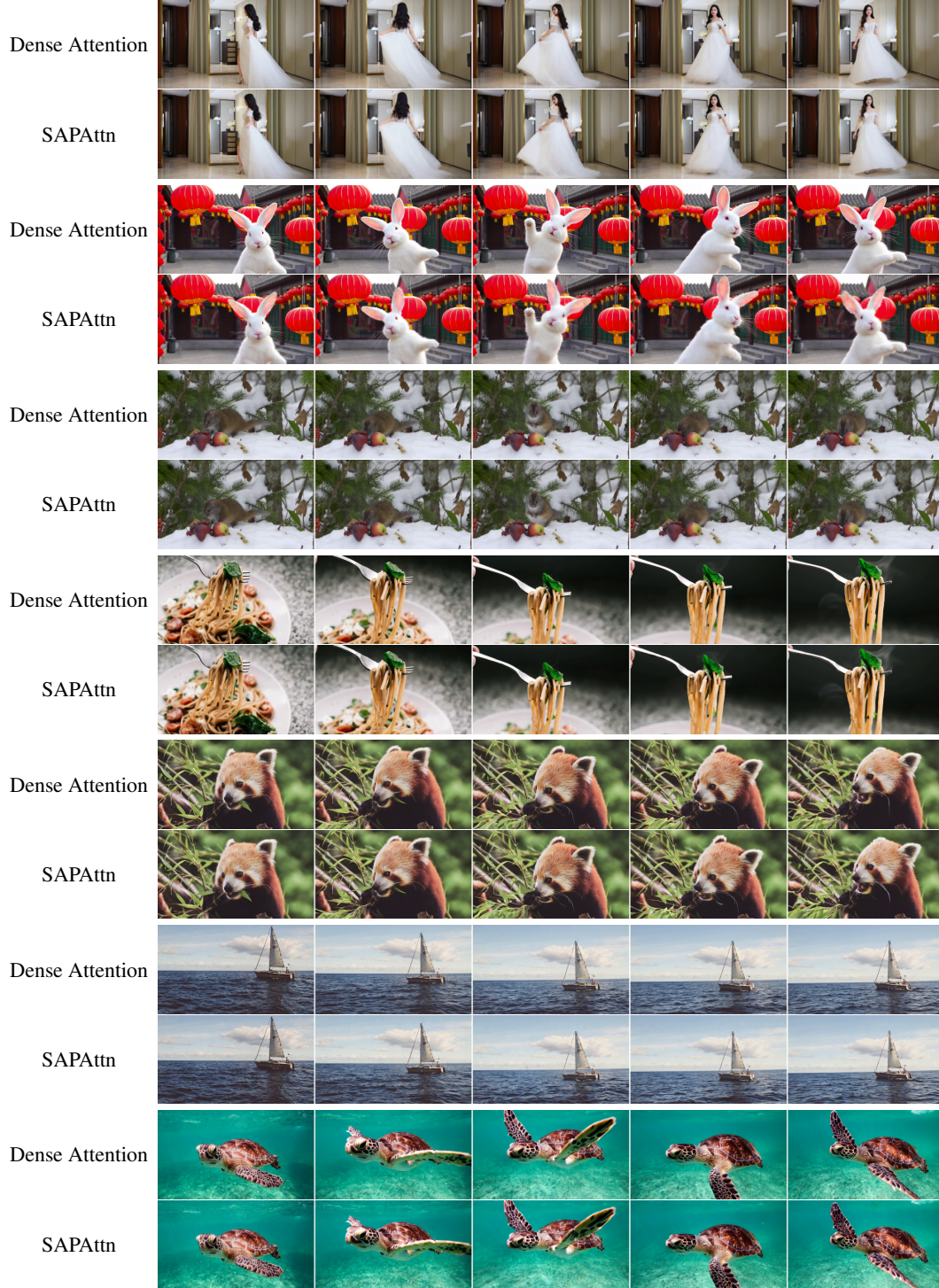


Figure 9: Comparison of Dense Attention and SAPAttn on HunyuanVideo and Wan 2.1 Text-to-Video generation.

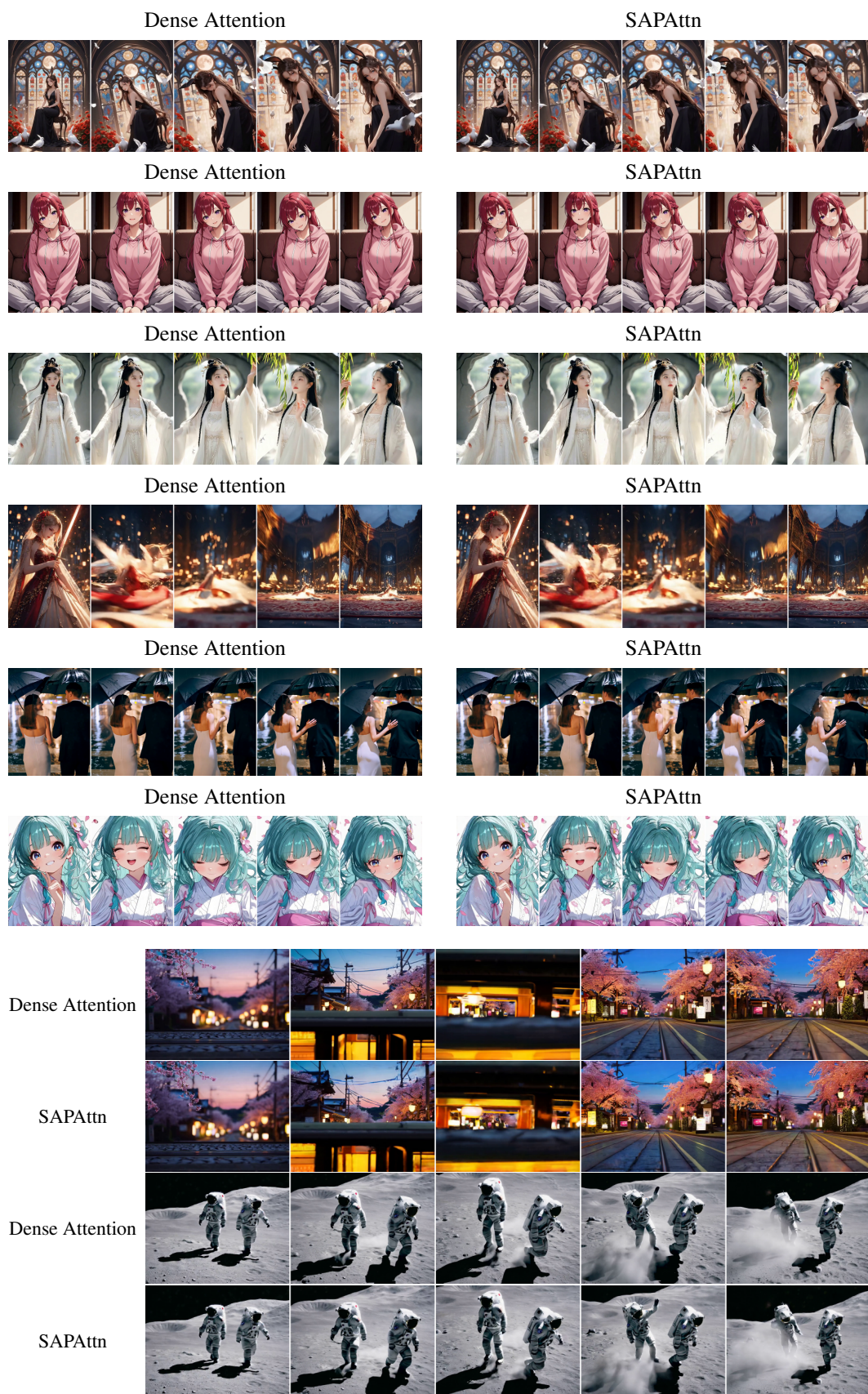


Figure 10: Comparison of Dense Attention and SAPAttn on Wan 2.1 Image-to-Video generation.