

Peer Review

Review of: "AVID: Adapting Video Diffusion Models to World Models"

Jiacheng He¹

1. University of Electronic Science and Technology of China, China

This article is well-organized, but there are still some issues that need to be addressed.

1. In Section 3.3, the phrase "it's parameters" should be corrected to "its parameters."
2. The explanation of biased sampling (Equation 4) is dense and may confuse readers unfamiliar with stochastic differential equations. Consider adding a brief intuition or example to clarify the practical implications.
3. The manuscript does not adequately justify why some baselines, such as Classifier-Free Guidance, perform poorly. Consider a deeper discussion of the limitations of these methods in the context of action-conditioned video generation.
4. The computational budget for each method is detailed, but it is unclear how these budgets compare in terms of cost-effectiveness. For example, does AVID consistently outperform others in scenarios with stricter computational constraints?
5. AVID should be defined in full the first time it appears.
6. The summary of related studies is insufficient and should be supplemented.

Declarations

Potential competing interests: No potential competing interests to declare.