Video Generation

Prompt: "a bear climbing a tree"









Hierarchical Prompting Generation

Split video generation prompts into 3 stages:

- 1. Static scene description. Describe all the objects or characters in the input prompt.
- 2. Action/motion directions. Describe the motion or action of all the objects or characters in the input prompt.
- Expected ending state. Describe the ending scene of the video.

Now process this input: a bear climbing a tree



["A bear is hugging the base of a tall tree in a forest, with leaves rustling gently in the breeze", "The bear is climbing the tree with strong movements, using its legs and arms, branches swaying slightly with each step", "The bear has reached the top of the tree, sitting comfortably with a view of the surrounding forest, the tree swaying gently in the wind"]

Stage 1: Text-to-Image Verification

Does the first frame not completely fail to meet the requirements stated in the text "A bear is hugging near the base of a tall tree in a forest, with leaves rustling gently in the breeze"?





Yes. 🍳



Stage 2: Medium Frames Verification

Does the medium frames not completely fail to meet the requirements stated in the text "The bear is climbing the tree with strong movements, using its legs and arms, branches swaying slightly with each step"?





Yes. 😂



Stage 3: Final Verification

Does the video not completely fail to meet the requirements stated in the text "The bear has reached the top of the tree, sitting comfortably with a view of the surrounding forest, the tree swaying gently in the wind"?





Yes. 😂



Does the whole video meet most of the requirements stated in the text "a bear climbing a tree"?



Yes. 💝

