

METHOD, DEVICE, AND MEDIUM FOR GENERATING TRANSITION VIDEOS WITH DIFFUSION MODEL

ABSTRACT

Implementations of the present disclosure provide a method, device, and medium for generating transition videos with a diffusion model. The method comprises obtaining a start frame and an end frame for the video, a first caption of the start frame, and a second caption of the end frame. The method further comprises generating a first latent noise in a latent space based on the start frame and a second latent noise in the latent space based on the end frame. The method further comprises generating third latent noises in the latent space corresponding to transition frames based on the first latent noise and the second latent noise. In addition, the method further comprises generating, by utilizing a pre-trained image-to-video diffusion model, the transition frames based on the third latent noises, the start frame, the end frame, the first caption, and the second caption.

FIG. 2