

ABSTRACT

Provided is a video generation method and related devices, which relates to the field of artificial intelligence technology, and in particular to the fields of computer vision, deep learning, large models, and augmented reality, and can be applied to scenarios such as digital humans. The method includes: extracting (S101) a reference feature of a target object from a reference image; determining (S102) a mixed pose sequence; wherein a plurality of target parts in the mixed pose sequence adopt a corresponding description mode to enhance a pose characteristic of a corresponding part; introducing (S103) initial noise to a denoising network, and performing a reverse denoising operation on the reference feature and the mixed pose sequence, to generate a video of the target object.

[Figure 1]