Rayeesa Mehmood, Rumaan Bashir, and Kaiser J. Giri

https://doi.org/10.1142/S0219649223500417 | **Cited by:** 0 **(Source: Crossref)**

Previous Next

Abstract

Generating videos is a novel area of computer vision research that is still far from being addressed. The reason for the same being that videos are very complex in nature where both spatial and temporal coherence needs to be taken care of. Compared to the unconditional video generation, an automated video generation from the text description is an even more difficult task, in which maintaining semantic consistency and visual quality are very crucial. The video generation from the text description seems to be non-trivial owing to the intrinsic complexity that occurs in the frames and video framework. The conditional generative models are required to be implemented for this challenging task of text-to-video generation. "Generative adversarial networks (GANs)" have had a lot of success in producing images conditioned over the natural language description. But, it is yet to be employed for producing realistic videos from text that are temporally and spatially coherent and semantically consistent with the text descriptions. Thus, a new Optimised Dual Discriminator Video Generative Adversarial Network (ODD-VGAN) for text-to-video generation is suggested in this paper. The hyper-parameters of ODD-VGAN are optimised using the improved reptile search algorithm (IRSA). The efficiency of the proposed approach is demonstrated by both qualitative and quantitative experimental results.

Keywords: Text-to-video generation ▪ Optimised Dual Discriminator Video Generative Adversarial Network ▪ improved reptile search algorithm ▪ optimised deep learning model ▪ conditional GAN