Abstract

A VIDEO ACTION RECOGNITION METHOD AND APPARATUS

The present application relates to the field of computer vision, and discloses a video action recognition method, including: obtaining a video; inputting the video into a hidden layer of a neural network model for processing to obtain a recognition object and a motion corresponding to the recognition object of the input video, where the hidden layer of the neural network model includes a plurality of processing units; and outputting an action recognition result of the input video based on the recognition object in the video and the motion corresponding to the recognition object; where in at least one processing unit, sequentially identifying spatial features and temporal features of the input video as input of the at least one processing unit, concatenating the spatial features and the temporal features of the input video in parallel, performing a pointwise convolution operation, and outputting spatial semantics and temporal semantics of the input video. According to the technical scheme provided by the present application, the video processing process has a greater ability to encode spatio-temporal relationships, allowing more meaningful features to be extracted with a smaller number of parameters, thus allowing more useful information to be learned from the dataset using a more compact structure. In the implementation of video processing, only a single processor is required to process the amount of video that previously required multi-processor processing, and the hardware configuration is more optimized, resulting in a more streamlined internal structure of the computer system.

FIG. 4