ABSTRACT

Action recognition is one form of implementation of the deep learning method, which is currently used in a wider field related to information technology, sports, and the arts. Erhu is a stringed instrument originating from China. In playing this instrument, there are rules on how to position the player's body and hold the instrument correctly. We need a system to detect every Erhu player's movement to meet these needs. So that in this study will discuss action recognition on video using three methods such as 3D-CNN, YOLOv3, and GCN. The 3D-CNN method is a method that has a CNN base. CNN is a method commonly used to perform image processing. 3DCNN has been proved effective in capturing motion information from continuous video frames. To improve the ability to capture every information stored in every movement, combining an LSTM layer in the 3D-CNN model is necessary. LSTM is an advanced RNN, a sequential network. It is capable of handling the vanishing gradient problem faced by RNN. Another method used in this study that has the ability in image processing is YOLOv3. YOLOv3 is an object detector with a relatively good accuracy level and can detect objects in real-time. Then to maximize the performance of YOLOv3, this study will combine YOLOv3 with GCN so that we can use the body key points to help YOLOv3 methods be easier for classification. GCN performs spatial convolution by merging several features of nodes around local neighbors on the graph. This research uses RGB video as a dataset, and there are three main parts in preprocessing and feature extraction. The three main parts are the body, erhu pole, and bow. To perform preprocessing and feature extraction, this study proposes two approaches. The first approach uses a segmentation process on the input video by utilizing the MaskRCNN method. The second approach uses a body landmark to perform preprocessing and feature extraction on the body segment. In contrast, the erhu and bow segments use the Hough Lines algorithm. The three main sections will then be divided into several sections according to the class that has been defined. Furthermore, for the classification process, this study proposes two algorithms to be used, namely, deep learning. This study will combine all deep learning methods with traditional image processing algorithm methods. These combination algorithm processes will produce an error message output from every movement displayed by the erhu player.

Keywords: Action Recognition, CNN, 3D-CNN, LSTM, YOLOv3, GCN