Multimodal Convolutional Neural Network for Musical Instrument Recognition

Authors : Yagya Raj Pandeya, Joonwhoan Lee

Abstract : The dynamic behavior of music and video makes it difficult to evaluate musical instrument playing in a video by computer system. Any television or film video clip with music information are rich sources for analyzing musical instruments using modern machine learning technologies. In this research, we integrate the audio and video information sources using convolutional neural network (CNN) and pass network learned features through recurrent neural network (RNN) to preserve the dynamic behaviors of audio and video. We use different pre-trained CNN for music and video feature extraction and then fine tune each model. The music network use 2D convolutional network and video network use 3D convolution (C3D). Finally, we concatenate each music and video feature by preserving the time varying features. The long short term memory (LSTM) network is used for long-term dynamic feature characterization and then use late fusion with generalized mean. The proposed network performs better performance to recognize the musical instrument using audio-video multimodal neural network.