SREEKANTH PAI
S7 CSB 30

Learning Affective Video Features for Facial Expression Recognition

One key challenging issues of facial expression recognition (FER) in video sequences is to extract discriminative spatiotemporal video features from facial expression images in video sequences. Here we propose a new method of FER in video sequences via a hybrid deep learning model. The proposed method first employs two individual deep convolutional neural networks (CNNs), including a spatial CNN processing static facial images and a temporal CN network processing optical flow images, to separately learn high-level spatial and temporal features on the divided video segments. These two CNNs are manipulated on target video facial expression datasets from a pre-trained CNN model. Then, the obtained segment-level spatial and temporal features are linked into a deep fusion network built with a deep belief network (DBN) model. This deep fusion network is used to learn spatiotemporal features. Finally, an average pooling is performed on the learned DBN segment-level features in a video sequence, to produce a fixed-length global video feature representation. Based on the global video feature representations, a linear support vector machine (SVM) is employed for facial expression classification tasks. The extensive experiments on three public video-based facial expression datasets, i.e., BAUM-1s, RML, and MMI, show the effectiveness of our proposed method, outperforming the state-of-the-arts.