Unified multi-stage fusion network for affective video content analysis

Yun Yi[1], Hanli Wang[2], and Pengjie Tang[2]

[1]Gannan Normal University
[2]Tongji University

June 18, 2022

Abstract

Affective video content analysis is an active topic in the field of affective computing. In general, affective video content can be depicted by feature vectors of multiple modalities, so it is important to effectively fuse information. In this work, a novel framework is designed to fuse information from multiple stages in a unified manner. In particular, a unified fusion layer is devised to combine output tensors from multiple stages of the proposed neural network. With the unified fusion layer, a bidirectional residual recurrent fusion block is devised to model the information of each modality. Moreover, the proposed method achieves state-of-the-art performances on two challenging datasets, i.e., the accuracy value on the VideoEmotion dataset is 55.8%, and the MSE values on the two domains of EIMT16 are 0.464 and 0.176 respectively. The code of UMFN is available at: https://github.com/yunyi9/UMFN.

Hosted file

latex.pdf available at https://authorea.com/users/489738/articles/573344-unified-multi-stage-fusion-network-for-affective-video-content-analysis

1