

Frame-Transformer Emotion Classification Network

Jiarui Gao¹

¹School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing,
Fudan University, China
jrgao14@fudan.edu.cn

ABSTRACT

Emotional content is a key ingredient in user-generated videos. However, due to the emotions sparsely expressed in the user-generated video, it is very difficult to analyze emotions in videos. In this paper, we propose a new architecture—Frame-Transformer Emotion Classification Network (FT-EC-net) to solve three highly correlated emotion analysis tasks: emotion recognition, emotion attribution and emotion-oriented summarization. We also contribute a new dataset for emotion attribution task by annotating the ground-truth labels of attribution segments. A comprehensive set of experiments on two datasets demonstrate the effectiveness of our framework.

KEYWORDS

Video emotion recognition; video emotion attribution; video emotion-oriented summarization and spatial-transformer network

ACM Reference format:

Jiarui Gao¹. 2017. Frame-Transformer Emotion Classification Network. In *Proceedings of*, , , 1 pages.
DOI: <http://dx.doi.org/xxx>

REFERENCES

- [1] Jie Zhang, Shiguang Shan, Meina Kan, and Xilin Chen. 2014. Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment. In *Computer Vision—ECCV 2014*. Springer, 1–16.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2017 ACM. ACM ISBN xxx...\$15.00
DOI: <http://dx.doi.org/xxx>