Weakly-Supervised Text-driven Contrastive Learning for Facial Behavior Understanding

Xiang Zhang[1], Taoyue Wang[1], Xiaotian Li[1], Huiyuan Yang[2], and Lijun Yin[1]

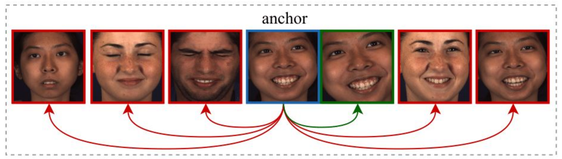
[1] Binghamton University [2] Rice University

ICCV23 PARIS

Motivation

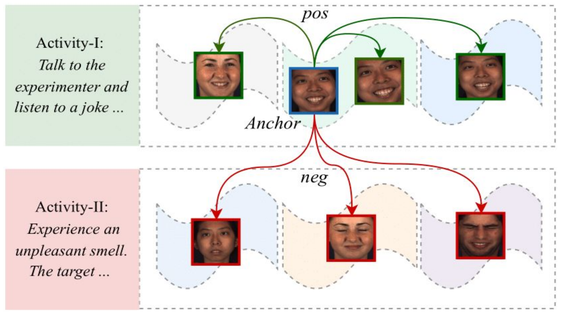
Problem

Self-supervised contrastive learning paring on face database will push away similar images



Idea

Paring based on activity category



Contribution

- We proposed a weakly-supervised contrastive learning method(a.k.a **CLEF**) that effectively leverages coarse-grained activity information.
- Text-driven contrastive learning is explored on FER and AUR tasks, where the performance is improved by incorporating text information.
- Experiments have been conducted on both in-the-lab and in-the-wild datasets.The SOTA performance is achieved.

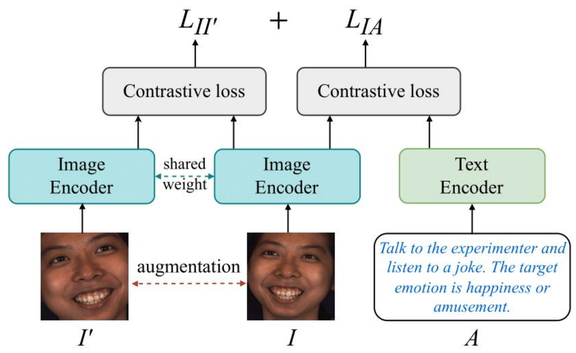
Acknowledgement

This work is supported in part by the NSF under grant CNS-1629898 and the Center of Imaging, Acoustics, and Perception Science (CIAPS) of the Research Foundation of Binghamton University.

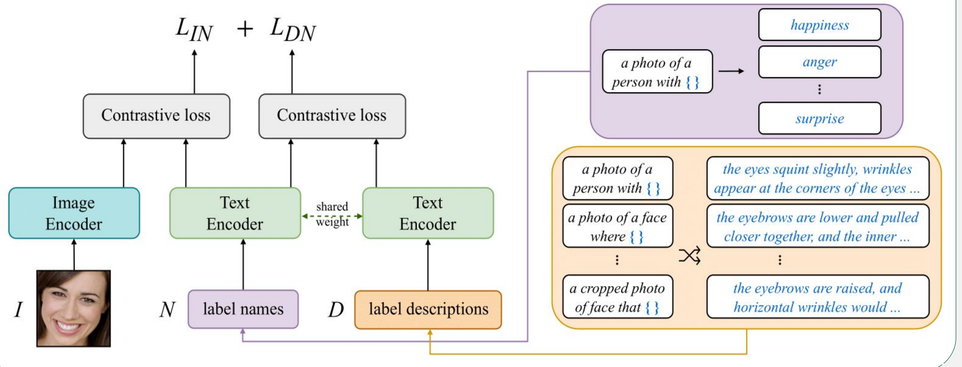
[1] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." ICML. 2021.
[2] Zheng, Yinglin, et al. "General facial representation learning in a visual-linguistic manner." CVPR. 2022.

Method

Pretraining



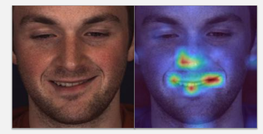
Finetuning



Task

Facial expression recognition

Facial action unit detection



Results

Quantitative Results

Results indicate F1-score on BP4D and accuracy on RAF-DB. PA: pre-trained with activity texts. PI: pre-trained with image; I: image encoder; N: label names; D: label descriptions.

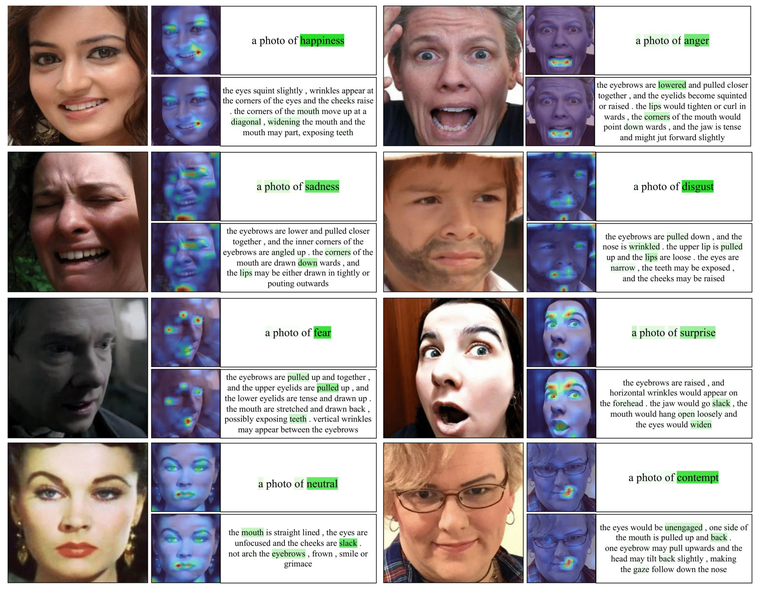
Methods	PA	PI	I	N	D	BP4D	RAF-DB
CLIP [1]			✓			63.4	87.88
CLIP			✓	✓		64.0	88.72
CLIP			✓	✓	✓	64.4	89.70
FaRL [2]			✓			63.7	88.31
FaRL			✓	✓		64.1	88.69
FaRL			✓	✓	✓	64.6	88.78
CLEF		✓	✓			65.0	89.67
CLEF	✓	✓	✓			64.2	89.34
CLEF		✓	✓	✓		64.7	88.57
CLEF	✓	✓	✓	✓		64.9	89.57
CLEF		✓	✓	✓	✓	64.8	89.44
CLEF	✓	✓	✓		✓	65.7	89.73
CLEF	✓	✓	✓	✓	✓	**65.9**	**90.09**

Zero-shot

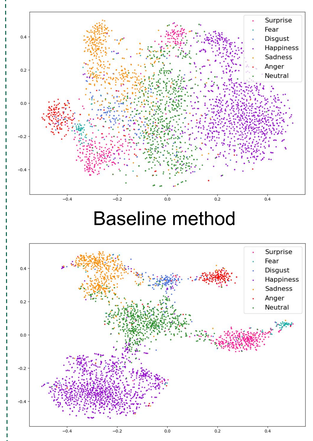
Left: trained on some expressions and test on other expressions
right: trained on AU detection and test on FER

Methods	RAF-DB	FER+	RAF-DB	FER+
FaRL	16.21	25.73	13.10	21.20
CLEF	**29.14**	**34.40**	**29.47**	**24.90**

Visualization



T-sne plot of latent features on RAF-DB



Baseline method

Our method

Discussion

- Pairing method: Each activity is deliberately designed to elicit a specific expression, more likely grouping similar expressions images together than self-supervised pairing.
- Extension: Using texts as label names facilitates easy extension with other information. For example, intensity details can be integrated into label names by including phrases, “with low intensity”, or “with high intensity”.
- Future work: Extend the work to the databases without activity descriptions; Investigate the prompting strategy.