# Semi-supervised Learning for Compound Facial Expression Recognition with Basic Expression Data

Rongkang Dong and Kin-Man Lam
*Department of Electrical and Electronic Engineering*
*The Hong Kong Polytechnic University*
Hong Kong, China
Email: rongkang97.dong@connect.polyu.hk, enkmlam@polyu.edu.hk

***Abstract*—Automatic facial expression recognition (FER) is a key technique in human-machine interaction. However, existing works on FER are mainly on basic expressions, paying far less attention to compound expressions. Even worse, labeled compound expression data is limited in quantity. Labeling this data requires prior psychological knowledge, and it is a time-consuming and labor-intensive task. Fortunately, much larger labeled basic expression datasets than compound expressions are available. Since basic expression images can potentially be compound expressions, they can be used to train models for compound FER (CFER). To achieve this goal, in this work, we propose to adopt a semi-supervised learning framework that utilizes basic expression data to generate pseudo-compound-emotion labels, thereby enhancing the model ability for CFER. In our method, we further exploit the basic labels and propose basic-label smoothing. Experimental results on the RAF-DB database and EmotioNet compound subset show that the proposed method achieves a great improvement over the baseline methods.**

***Index Terms*—Compound facial expression recognition, semi-supervised learning, pseudo-compound-emotion labels.**

## I. Introduction

We humans often convey more than one emotion through a facial expression. However, limited research has focused on compound facial expression recognition (CFER). Moreover, compound expression data for CFER research is still limited. The amount of basic facial expression data is larger than compound facial expression data. Basic expressions can potentially be compound expressions, although they are labeled with basic expressions for basic expression recognition. Relabeling them is difficult. This motivates us to utilize them for CFER. In this study, we propose to adopt a semi-supervised learning framework to generate pseudo-compound-emotion labels for basic expression data (SSL-BED-CFER). Moreover, we leverage the existing basic labels in CFER training by introducing a basic-label smoothing method. Fig. 1 shows the framework of the proposed method.

## II. Experimental results

Table I shows the experimental results. The semi-supervised learning methods (i.e., Pseudo-label, FixMatch, Ada-CM [3], and our SSL-BED-CFER) are based on the ResNet-18 model trained with sampled compound-expression data and basic-expression data. Two baseline methods, i.e., the backbone only and the EMA model, are trained with only the sampled compound data and are validated on the remaining compound data. We can see that our proposed method can achieve the best accuracy results over the other methods.

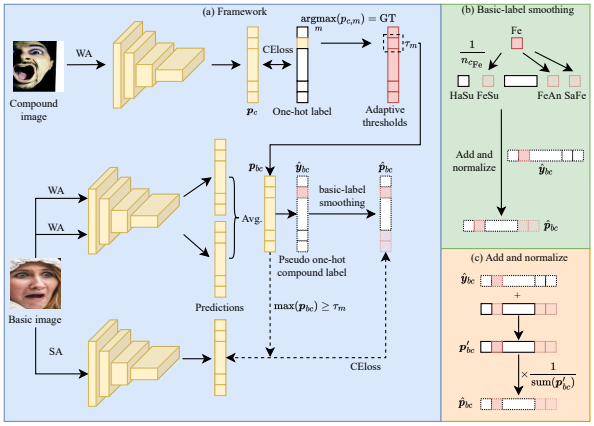


Fig. 1. Framework of the proposed method. The operation abbreviations "WA", "SA", "CEloss", and "Avg." stand for "weakly augmented", "strongly augmented", "cross-entropy loss", and "average", respectively. "GT" means ground truth. The basic emotion "fearful" is abbreviated as "Fe", so "FeSu" means "fearfully surprised", "HaSu" means "happily surprised", etc. $n_{c_{\mathrm{Fe}}}$ is the number of compound classes with the fearful emotion as the basic expression.

TABLE I
Experimental results evaluated on the RAF-DB compound subset [1], the EmotioNet compound subset [2], and the RAF-DB basic subset. "FT" stands for fixed threshold.

| Method | RAF-DB_C | EmotioNet_C |
|---|---|---|
| Backbone only | 58.09 | 63.40 |
| EMA model | 58.77 | 65.02 |
| Pseudo-label | 65.71 | 64.66 |
| FixMatch (FT=0.8) | 66.66 | 67.63 |
| Ada-CM w/o contrastive objective w/ threshold clip | 67.16 | 68.53 |
| SSL-BED-CFER w/ threshold clip (Ours) | 69.97 | 73.92 |

## References

[1] S. Li, W. Deng and J. Du, "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2017, pp. 2852–2861.

[2] C. Fabian Benitez-Quiroz, R. Srinivasan, and A. M. Martinez, "Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2016, pp. 5562-5570.

[3] H. Li, N. Wang, X. Yang, X. Wang, X. Gao, "Towards semi-supervised deep facial expression recognition with an adaptive confidence margin," in *Proc. IEEE Conf. Comput. Vision Pattern Recognit.*, 2022, pp. 4166-4175.