# Supplementary Material for ASM: Adaptive Sample Mining for In-The-Wild Facial Expression Recognition

Ziyang Zhang[1,2], Xiao Sun[1,2,3](✉), Liuwei An[1,2], and Meng Wang[1,2,3]

[1] School of Computer Science and Information Engineering,
Hefei University of Technology, Heifei, China
sunx@hfut.edu.cn

[2] Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machines, Hefei University of Technology, Heifei, China

[3] Institute of Artificial Intelligence, Hefei Comprehensive National Science Center, Heifei, China

## 1 Ablation Study

Table 1: Ablation study of the fixed threshold and adaptive threshold in ASM on RAF-DB and FERPlus. Note that FT denotes fixed threshold and AT denotes adaptive threshold.

| $[T_n, T_c]$ | FT/AT | RAF-DB(%) | FERPlus(%) |
|---|---|---|---|
| [0.3, 0.7] | FT | 89.75 | 89.13 |
| [0.2, 0.8] | FT | 89.86 | 89.35 |
| [0.1, 0.9] | FT | 90.03 | 89.67 |
| [0.05, 0.95] | FT | 90.22 | 89.79 |
| [0.3, 0.7] | AT | 90.35 | 89.92 |
| [0.2, 0.8] | AT | 90.58 | 90.21 |
| [0.1, 0.9] | AT | 90.28 | 89.92 |
| [0.05, 0.95] | AT | 90.12 | 89.89 |

**Fixed vs. Adaptive.** We carry out the ablation study to investigate the different threshold-generated strategies. As shown in Tab.1, several observations can be summarized as follows. First, our adaptive threshold is shown to achieve larger performance improvement than fixed threshold. This also confirms our viewpoint that dynamic threshold, as compared to fixed threshold, can more comprehensively reflect the recognition difficulty of different facial expression categories and facilitate accurate differentiation between ambiguous expressions

This work was supported by the National Key R&D Programme of China (2022YFC3803202), Major Project of Anhui Province under Grant 202203a05020011. This work was donein Anhui Province Key Laboratory of Affective Computing and Advanced Intelligent Machine.

and noise expressions. Additionally, with the incorporation of our carefully designed tri-regularization module, the robustness of the model can be further enhanced. Second, as for the fixed threshold, when $T_c$ is small, some ambiguous expressions are mixed into the clean set, which affects the efficient learning of the model. When $T_n$ is large, some ambiguous expressions are mixed into the noisy set, causing the model to not fully learn the label information of these samples. As $T_c$ increases and $T_n$ decreases, the model can accurately distinguish between clean and noisy samples, resulting in more ambiguous samples. This aligns with the distribution of ambiguous expressions in the datasets. As for the adaptive threshold, the model is capable of adaptively adjusting the confidence of each class based on the recognition difficulty of the samples and tolerance towards noise during the learning process. As a result, it can handle different categories more fairly.

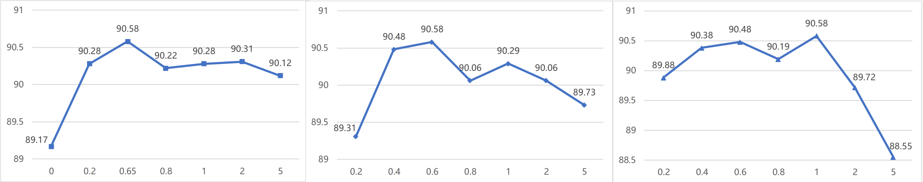

Fig. 1: Evaluation of the hyper-parameters $\beta$, $\omega$ and $\gamma$ on the RAF-DB dataset.

**Evaluation of $\beta$.** $\beta$ controls the shape of the sigmoid ramp-up function in the mutuality learning strategy. The smaller $\beta$, the faster the transition from supervised loss to contrastive loss and vice versa. We study different values from 0 to 5 in both clean and synthetic noisy RAF-DB datasets. As can be seen in

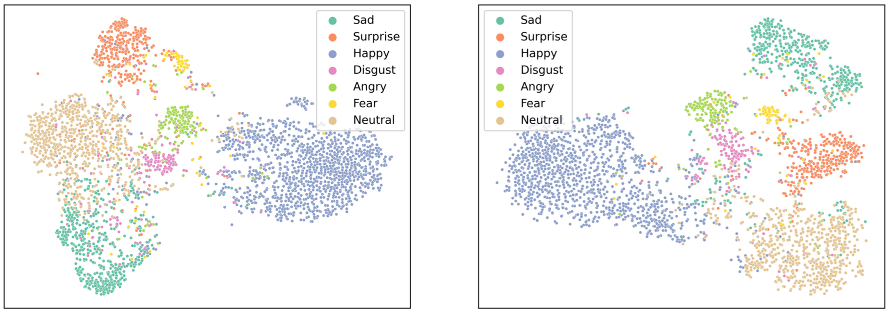

Fig. 2: t-SNE visualizations of facial expression features obtained by (a) Baseline and (b) ASM on RAF-DB test set.

Fig. 1 (left), $\beta = 0.65$ achieves better performance under all noise levels, which is also effective on other datasets.

**Evaluation of $\omega$.** $\omega$ is the weighting factor of mutuality loss. We evaluate different $\omega$ on all three dataset, and the results on RAF-DB are shown in Fig. 1 (middle). We can find that too small $\omega$ suppress the model to discriminate clean samples from ambiguous samples, and too large $\omega$ causes the model to focus on ambiguous expressions and not learn well enough from clean expressions.

**Evaluation of $\gamma$.** $\gamma$ is the weighting factor of unsupervised consistency loss. We evaluate different $\gamma$ on all three dataset, and the results on RAF-DB are shown in Fig. 1 (right). When $\gamma$ is too small, the model cannot take full advantage of the noisy label expressions. When $\gamma$ is too large, it is difficult for the model to accurately divide the dataset.

## 2 Feature Visualization

To evaluate the effectiveness of the proposed ASM method, we utilize t-SNE [1] to visualize the distribution of facial features extracted in a 2-D space by the Baseline model in Fig.2 (a), and by ASM in Fig.2 (b) for RAF-DB. It is evident that the facial expression features captured by the Baseline model exhibit poor discriminability due to significant intra-class and inter-class variations. In contrast, the ASM-based expression features demonstrate high intra-class similarity and distinct inter-class differences, thus yielding superior performance.

## References

1. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)