Ambiguous Facial Expression Recognition Using Local Features

Moeka Miyata
School of FSE,
Waseda University
Tokyo, Japan
moeka28miyata@moegi.waseda.jp

Hiroshi Watanabe
Graduate School of FSE,
Waseda University
Tokyo, Japan
hiroshi.watanabe@waseda.jp

***Abstract*—Facial expression recognition plays an important role in estimating emotional states in real-world environments. However, many facial images contain ambiguous expressions in which emotional cues are weak or intentionally suppressed. Such ambiguity often leads to unstable annotations and noisy labels, which degrade recognition performance. In this paper, we propose a two-stage learning framework that evaluates ambiguity from both local and global facial features. Specifically, we first estimate emotion using the periocular region, where emotional changes tend to appear prominently, and then incorporate this information into whole-face emotion recognition. Experimental results demonstrate that the proposed method improves recognition accuracy for ambiguous expressions compared with the conventional Navigating Label Ambiguity method.**

***Index Terms*—Facial expression recognition, label ambiguity, local feature, noise-aware weighting, periocular region.**

I. INTRODUCTION

In interpersonal sports situations and remote communication environments, emotions are often intentionally suppressed due to tactical intentions or social considerations. Therefore, the observed facial expressions often become ambiguous, which makes it difficult to accurately infer emotional states from facial appearance alone.

Recent studies have investigated label ambiguity in facial expression datasets, where subjective annotations introduce label noise. Navigating Label Ambiguity (NLA)[1] has addressed this issue by adaptively adjusting the contribution of each training sample according to its ambiguity level.

However, NLA evaluates ambiguity only based on global facial features. As a result, local facial characteristics that strongly reflect emotional changes are not sufficiently considered. Therefore, ambiguity evaluation should incorporate both local and global perspectives. In this paper, we propose a two-stage learning framework that evaluates ambiguity using both periocular features and whole-face features. The periocular region is particularly informative for emotion recognition. By integrating local ambiguity estimation into global learning, the proposed method improves performance for ambiguous expressions.

II. RELATED WORK

Navigating Label Ambiguity (NLA) introduces Noise-aware Adaptive Weighting (NAW)[1] to evaluate the ambiguity of each sample. NAW uses two probabilities: the predicted probability of the ground-truth label, denoted as P_{GT}, and the highest predicted probability among non-ground-truth classes, denoted as P_{NN}.

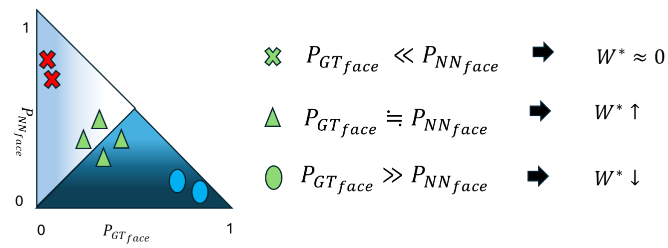

Fig. 1. Overview of the NAW.

The overall mechanism of NAW is illustrated in Fig. 1. If $P_{GT} > P_{NN}$, the expression is considered clear. When $P_{GT} \approx P_{NN}$, the sample is regarded as ambiguous but informative. Otherwise, $P_{GT} < P_{NN}$, the sample is treated as noisy. Based on these relationships, NAW adjusts the loss weight of each sample.

Although this strategy effectively suppresses harmful noisy labels, it relies only on global facial representations.

III. PROPOSED METHOD

In this study, we propose a two-stage learning framework that evaluates facial expression ambiguity using both local facial features and whole-face features. The overall framework of the proposed method is illustrated in Fig. 2.

Among various facial regions, the periocular region plays an important role in facial expression recognition[2][3]. Therefore, we first focus on the periocular region as a local feature source.

In the first stage, we extract the periocular region from each face image using MediaPipe[4]. We then train an emotion recognition model based on these cropped regions. During training, we apply Noise-aware Adaptive Weighting and Jensen–Shannon Divergence based consistency regularization to estimate the ambiguity level of each sample.

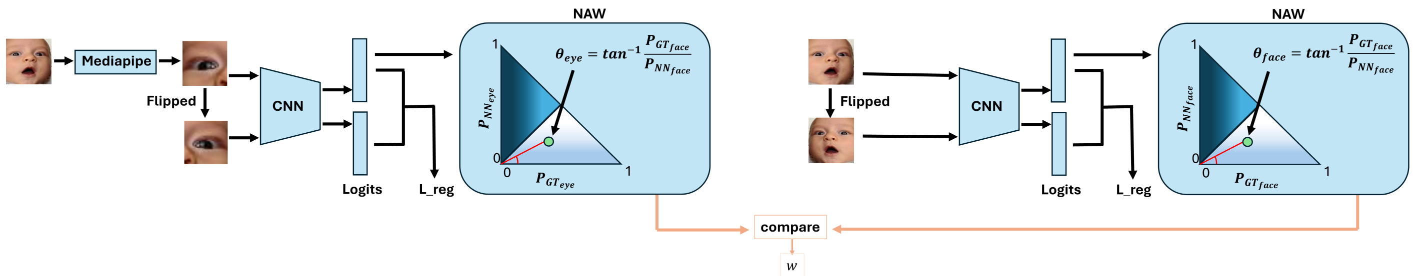

Fig. 2. Overview of the proposed method.

TABLE I
COMPARISON OF RECOGNITION ACCURACY IN FACIAL EXPRESSION RECOGNITION TASKS

Method	Surprise	Fear	Disgust	Happiness	Sadness	Anger	Neutral	Average
NLA	0.8844	0.6486	0.6125	0.9578	**0.8828**	**0.8333**	0.8779	0.8139
Ours	**0.8875**	**0.6756**	**0.6312**	**0.9603**	0.8744	0.8271	**0.8897**	**0.8208**

We define the ambiguity indicator for the periocular region as

$$\theta_{\text{eye}} = \tan^{-1}\left(\frac{P_{GT}^{\text{eye}}}{P_{NN}^{\text{eye}}}\right), \tag{1}$$

where P_{GT}^{eye} denotes the predicted probability of the ground-truth label and P_{NN}^{eye} denotes the highest predicted probability among the non-ground-truth classes. A larger θ_{eye} indicates clearer recognition in the periocular region.

In the second stage, we train a model using the whole-face image. We compute the ambiguity indicator for the whole-face prediction in the same manner as

$$\theta_{\text{face}} = \tan^{-1}\left(\frac{P_{GT}^{\text{face}}}{P_{NN}^{\text{face}}}\right), \tag{2}$$

Based on both ambiguity indicators, we adjust the training weight of each sample. We determine the final weight as

$$w = \begin{cases} \gamma & (\theta_{\text{face}} > \frac{\pi}{4} \ and \ \theta_{\text{eye}} < \theta_{\text{face}}), \\ w_0 & (\text{otherwise}), \end{cases} \tag{3}$$

where γ controls the contribution of informative ambiguous samples and w_0 denotes the original weight defined in Noise-aware Adaptive Weighting. We set γ to 0.5 based on empirical observation.

This two-stage design enables the model to reconsider samples that are treated as noisy when only global features are used. By incorporating local facial cues into ambiguity estimation, we utilize informative but difficult samples more effectively during training.

The final loss for whole-face training follows the NAW formulation:

$$L_{NAW-CE} = (1 + w) \cdot L_{CE}, \tag{4}$$

where w is determined by (3).

The total loss is computed as

$$L_{total} = \lambda L_{NAW-CE} + (1 - \lambda) L_{reg}, \tag{5}$$

where L_{reg} denotes the Jensen–Shannon divergence based regularization term.

IV. EXPERIMENTS

We evaluate the proposed method on the RAF-DB[1] dataset using a seven-class emotion classification task. We compare our method with the conventional NLA framework under the same experimental conditions.

The classification accuracy for each emotion category and the macro average are shown in Table I. The proposed method achieves higher average accuracy than NLA. In particular, improvements are observed in five out of seven emotion classes.

These results indicate effective utilization of informative ambiguous samples.Consequently, the model achieves more robust learning under label ambiguity.

V. CONCLUSION

In this paper, we proposed a two-stage learning framework that evaluates facial expression ambiguity based on both periocular and whole-face features. Experimental results on RAF-DB demonstrated that incorporating local facial features improves recognition performance for ambiguous expressions.

Although this study focuses only on the periocular region as a local feature, other regions, such as the mouth and eyebrow areas, may also influence ambiguity evaluation. Future work will investigate multiple facial regions to further improve recognition accuracy.

REFERENCES

[1] J. Lee et al., “Navigating Label Ambiguity for Facial Expression Recognition in the Wild,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 39, no. 4, pp.4517–4525, Apr. 2025.

[2] F. Ma, Y. He, B. Sun, and S. Li, “Multimodal Prompt Alignment for Facial Expression Recognition,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 12581-12591, Oct. 2025.

[3] X. He, C. Luo, X. Xian, B. Li, M. H. Khan, Z. Ge, W. Xie, S. Song, L. Shen, B. Ghanem, and X. Yue, “SynFER: Towards Boosting Facial Expression Recognition with Synthetic Data,” in Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 10184-10195, Oct. 2025.

[4] T. E. Köksal, “Deep Learning-Based Real-Time Sequential Facial Expression Analysis Using Geometric Features,” arXiv preprint arXiv:2512.05669, Dec. 2025.