Journal of Electronic Research and Application, 2025, Volume 9, Issue 2
http://ojs.bbwpublisher.com/index.php/JERA
ISSN Online: 2208-3510
ISSN Print: 2208-3502

Face Expression Recognition on Uncertainty-Based Robust Sample Selection Strategy

Yuqi Wang, Wei Jiang*

School of Information Engineering, North China University of Water Resources and Electric Power, Zhengzhou 450046, China

Corresponding author: Wei Jiang, jiangwei@ncwu.edu.cn

Abstract: In the task of Facial Expression Recognition (FER), data uncertainty has been a critical factor affecting performance, typically arising from the ambiguity of facial expressions, low-quality images, and the subjectivity of annotators. Tracking the training history reveals that misclassified samples often exhibit high confidence and excessive uncertainty in the early stages of training. To address this issue, we propose an uncertainty-based robust sample selection strategy, which combines confidence error with RandAugment to improve image diversity, effectively reducing overfitting caused by uncertain samples during deep learning model training. To validate the effectiveness of the proposed method, extensive experiments were conducted on FER public benchmarks. The accuracy obtained were 89.08% on RAF-DB, 63.12% on AffectNet, and 88.73% on FERPlus.

Keywords: Facial expression recognition; Uncertainty; Sample selection strategy.

Online publication: April 3, 2025

1. Introduction

Emotions are expressed through various non-verbal means, with facial expressions being the most intuitive and widespread form of emotional communication. In recent years, numerous deep learning-based facial expression recognition (FER) methods have been proposed with promising performance. However, there are annotation discrepancies between different labelers for similar expressions, which leads to uncertainty [1-4]. Typically, uncertainty can lead to insufficient feature learning by the model, particularly when dealing with complex expressions where key facial features are not adequately captured. Studies have shown that in the final classification results of a facial expression recognition task trained with the ResNet-18 model, misclassified samples are often uncertain ambiguous expression images, accompanied by high confidence scores. The model gradually memorizes the facial features of noisy samples and overfits as the training epochs progress, leading to overly confident misclassifications.

To address the aforementioned challenges of uncertainty, this paper proposes a sample selection framework based on the ResNet-18 model. The framework utilizes an uncertainty determination mechanism to select samples that positively contribute to model training. Specifically, the training data is first divided into a clean set and an uncertain set. To enhance training stability, the sample exclusion strategy prevents uncertain samples from participating in the training process and temporarily reserves them for reassessment in the next epoch. Additionally, this framework incorporates RandAugment [5] during data preprocessing, utilizing a richer feature space to improve the model's ability to distinguish between different facial expression categories, thereby increasing intra-class compactness and inter-class separability. The main contributions of this paper are as follows:

We propose an efficient sample exclusion framework to mitigate the effect of uncertain samples on model training, allowing the model to learn clean facial expression features. We conduct comprehensive experiments on both real-world and synthetic noise Facial Expression Recognition (FER) datasets to validate the strong robustness of our method.

2. Related work

2.1. Facial expression recognition

The uncertainty stemming from inter-class similarity and annotation ambiguity in facial expressions makes it challenging to accurately recognize emotions. Recent methods [1-4] leverage learnable uncertainty to enhance model robustness. SCN [1] suggests quantifying the uncertainty of each sample and ranking them to suppress uncertain samples through re-labeling. DMUE [2] introduces a multi-branch learning framework to explore latent distributions and decay the weights of uncertain samples using confidence scores. RUL [3] quantifies the relative uncertainty of images and uses it for the weight of facial features mix. IPA2LT [4] is the first to address label inconsistency in FER datasets, proposing to assign pseudo-labels to each image to obtain a latent distribution.

2.2. Learning with noisy labels

Learning with noisy labels is currently mainly classified into two categories: adjusting the loss function [6,7] and clean sample selection methods [8,9]. Adjusting the loss function [6,7] focuses on estimating the noise transition matrix, inferring the probability of different class data points being corrupted using clean samples, and modifying the loss of each sample to minimize the adverse effects of label noise. Sample selection research [8,9] currently focuses on methods based on probability distribution. These methods typically utilize confidence to address noisy labels, where confidence reflects the model's certainty about each sample prediction.

3. Proposed method

3.1. Overview of proposed method

This paper designs a training framework based on confidence error to prevent deep networks from overfitting uncertain facial images. We observe that the model tends to memorize noisy facial features during training, which leads to a gradual increase in the classification accuracy of uncertain labels after an early fluctuation. Inspired by [9], we compute the confidence error of training samples. To prevent the model from memorizing uncertain facial images, samples with confidence errors above the fixed threshold are excluded from training and retained for re-evaluation in the next epoch. In addition, we employ RandAugment during the data preprocessing stage to expand the feature space of facial expressions.

3.2. Data augmentation module

To mitigate the deterioration of training caused by uncertainty, we introduce RandAugment [5]. RandAugment expands the training dataset by randomly selecting transformations for each sample in a mini-batch and increases the diversity of images. In FER tasks, facial expression images may be blurry or occluded and the model may fail to learn useful knowledge for distinguishing uncertain samples. Therefore, we choose to use RandAugment to introduce perturbations to simulate the real-world data distribution, making the model sensitive and adaptive to uncertainty.

3.3. Confidence error

Confidence error [9] is defined as the difference between the predicted label and the original label of a sample, serving as a sieving strategy to distinguish clean samples from noisy samples. During the training of CNN for Facial Expression Recognition (FER), the neural network model $F(x_i\,, \theta) \in R^{m\times k}$ is a k-class classifier with trainable parameters θ. The probability computed by the softmax activation function for each class represents the confidence of the sample in that class. Assuming a classification task on a training dataset $D = \{(x_i, y_i)|x_i \in X, y_i \in Y\}_{i=1}^{n}$, where *n* is the number of samples, *X* and *Y* denote the training sample and label spaces respectively.

Given a set of facial expression images $D_1 = (x_i, \widetilde{y_i})$, we apply RandAugment to the input images and then feed them into the neural network model to obtain the predicted probability for each class, denoted as $P = F(x_i\,, \theta)$. The model confidence $P^{(l)} = F(x_i\,, \theta)^{(l)}$ is generated through the original labels $l \in \{1\,, \ldots, k\}$. Subsequently, the predicted confidence is considered as the current maximum probability: $p^{arg} = \text{argmax}\big(F(x_i\,, \theta)\big)$. From these two confidences, we can extract the confidence error of facial expression images:

$$E_P(D_1) = P^{arg} - P^{(l)} \qquad (1)$$

3.4. Uncertainty judgment module

To select uncertain data samples, this paper employs a proven effective sample exclusion method that uses cross-entropy to exclude samples exceeding a fixed threshold from training. Furthermore, these excluded samples can be re-evaluated instead of being deleted in the next epoch. For the multi-class facial expression task, we denote our loss as Sample Exclusion Loss (L_{SEL}), formulated as follows:

$$L_{\text{SEL}} = \sum_{b=1}^{m} \mathbf{1}\,(E_P(D_1) \le \delta) H(P, l) \qquad (2)$$

Where *H*(*P*,*l*)is defined as the cross-entropy of the probability distribution , with as the fixed threshold. We calculate the loss for clean samples based on the formula. Importantly, samples exceeding the threshold do not participate in the current model training round, but their confidence error will be recalculated for judgment in the next epoch.

4. Experiments

4.1. Implementation details

This experiment utilizes the RAF-DB [10], FERPlus [11] and AffectNet [12] datasets, implemented with the PyTorch framework and executed on three GTX 1080 Ti GPUs. By default, we employed a ResNet-18 [13] pre-trained on MS-Celeb-1M [14] and trained it end-to-end as the backbone network. Facial images were resized to 224×224 pixels for fair comparison. To improve the effectiveness of the sieving strategy, we applied horizontal flipping with a probability of 0.5, Random Erasing [15], and RandAugment [5] to the images. The batch size was set to 1024 during training. The initial learning rate was 0.001 and training ended at epoch 60. Additionally, we used the Adam optimizer with a weight decay of 0.0001 to expedite convergence. To decrease the learning rate after each epoch, the ExponentialLR learning rate scheduler was set with a gamma of 0.9.

4.2. Evaluation on noise for datasets

To quantitatively analyze noisy labels, we explore the robustness of our method across three parameters as shown in **Table 1**.

Table 1. Evaluation of the sample selection framework on synthetic uncertainty FER datasets

Noise (%)	Method	RAF-DB (%)	AffectNet-7 (%)
10	SCN [1]	82.14	58.56
	DMUE [2]	83.19	61.21
	RUL [3]	86.22	60.54
	Ours	88.17	61.24
20	SCN [1]	79.79	57.21
	DMUE [2]	80.31	58.66
	RUL [3]	84.34	58.36
	Ours	87.09	60.84
30	SCN [1]	77.46	54.84
	DMUE [2]	79.41	56.88
	RUL [3]	82.06	56.65
	Ours	85.10	59.63

Levels on the RAF-DB and AffectNet datasets. Specifically, we select 10%, 20%, and 30% of the training data in each category and randomly change their labels to assign them labels from other categories. For a fair comparison, we choose ResNet-18 as the backbone network and compare its performance with other state-of-the-art FER uncertainty quantification methods based on ResNet-18. As shown in **Table 1**, our method outperforms superior performance compared to SCN and other state-of-the-art FER uncertainty methods.

5. Conclusions

This paper proposes a simple framework based on confidence error to prevent deep networks from overfitting uncertain facial images. The sample selection framework employs an uncertainty Judgment module to filter out samples above the fixed threshold and retain them for training, which prevents the model from overfitting uncertain facial expression images. Additionally, this paper introduces RandAugment to simulate real-world data distribution, making the model sensitive and adaptive to uncertainty. Experimental results on multi-dimensional synthetic and real-world FER datasets demonstrate the robustness of this framework. Furthermore, compared to other uncertainty training methods, the proposed sample selection framework achieves state-of-the-art performance.

Disclosure statement

The authors declare no conflict of interest.

References

[1] Wang K, Peng X, Yang J, et al., 2020, Suppressing Uncertainties for Large-scale Facial Expression Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6897–6906.

[2] She J, Hu Y, Shi H, et al., 2021, Dive Into Ambiguity: Latent Distribution Mining and Pairwise Uncertainty Estimation for Facial Expression Recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 6248–6257.

[3] Zhang Y, Wang C, Deng W, 2021, Relative Uncertainty Learning for Facial Expression Recognition. Advances in Neural Information Processing Systems, 34: 17616–17627.

[4] Zeng J, Shan S, Chen X, 2018, Facial Expression Recognition with Inconsistently Annotated Datasets. Proceedings of the Proceedings of the European Conference on Computer Vision (ECCV), 222–237.

[5] Cubuk ED, Zoph B, Shlens J, et al., 2020, Randaugment: Practical Automated Data Augmentation with a Reduced Search Space. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 702–703.

[6] Yao Y, Liu T, Gong M, et al., 2021, Instance-dependent Label-noise Learning Under a Structural Causal Model. Advances in Neural Information Processing Systems, 34: 4409–4420.

[7] Yao Y, Liu T, Han B, et al., 2020, Dual t: Reducing Estimation Error for Transition Matrix in Label-noise Learning. Advances in Neural Information Processing Systems, 33: 7260–7271.

[8] Nguyen D, Mummadi C, Ngo T, et al., 2019, Self: Learning to Filter Noisy Labels with Self-ensembling. arXiv: 1910.01842. https://doi.org/10.48550/arXiv.1910.01842.

[9] Torkzadehmahani R, Nasirigerdeh R, Rueckert D, et al., 2022, Label Noise-robust Learning using a Confidence-based Sieving Strategy. arXiv: 2210.05330. https://doi.org/10.48550/arXiv.2210.05330.

[10] Li S, Deng W, Du J, 2017, Reliable Crowdsourcing and Deep Locality-preserving Learning for Expression Recognition in the Wild. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2852–2861.

[11] Barsoum E, Zhang C, Ferrer C, et al., 2016, Training Deep Networks for Facial Expression Recognition with Crowd-sourced Label Distribution. Proceedings of the Proceedings of the 18th ACM International Conference on Multimodal Interaction, 279–283.

[12] Mollahosseini A, Hasani B, Mahoor M, 2017, Affectnet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild. IEEE Transactions on Affective Computing, 10(1): 18–31.

[13] He K, Zhang X, Ren S, et al., 2016, Deep Residual Learning for Image Recognition, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 770–778.

[14] Guo Y, Zhang L, Hu Y, et al., 2016, Ms-celeb-1m: A dataset and Benchmark for Large-scale Face Recognition. Proceedings of the Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part III 14, 2016, Springer International Publishing, 87–102.

[15] Zhong Z, Zheng L, Kang G, et al., 2020, Random Erasing Data Augmentation. Proceedings of the AAAI Conference on Artificial Intelligence, 34(07): 13001–13008.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.