南开国际先进研究院（深圳福田）

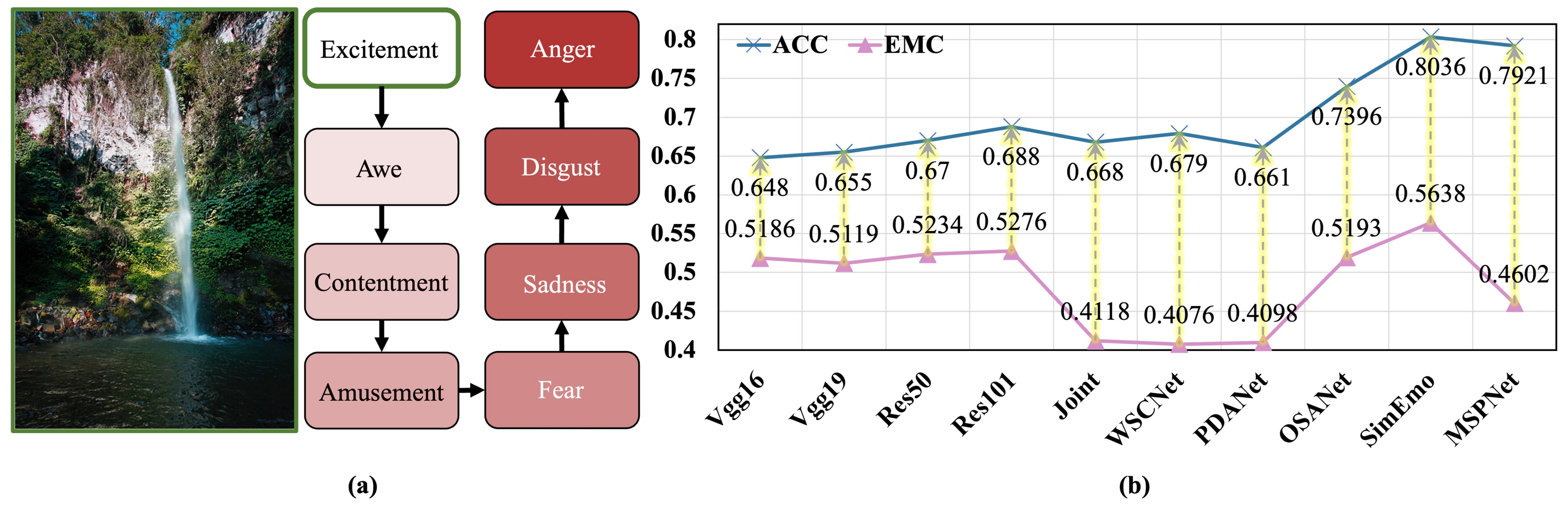
Nankai International Advanced Research Institute (SHENZHEN· FUTIAN)

To Err Like Human: Affective Bias-Inspired Measures for Visual Emotion Recognition Evaluation

Chenxi Zhao, Jinglei Shi*, Liqiang Nie, Jufeng Yang

NEURAL INFORMATION PROCESSING SYSTEMS

1.Introduction



(a) (b)

In the task of emotional classification, misclassifying a certain emotion into one class may be more severe than another, e.g., misclassifying 'excitement' as 'anger' apparently is more severe than as 'awe'.

2.Method

2.1 Definition of Emotional Distance

Previous scholars[1] provide a definition between emotional labels based on Mikel's wheel, However, this definition of emotion distance neglects the emotional polarity, e.g., the distance between 'fear' and 'excitement' being the same as the distance between 'fear' and 'sadness'. We thus define the emotional distance as follows:

$$W_{i,j} = \begin{cases} 1+\text{dist}\left(e_i,e_j\right) & e_i,e_j \in C_p \\ \mathbf{C}+\text{dist}\left(e_i,e_j\right) & e_i,e_j \notin C_p \end{cases} \quad (1)$$

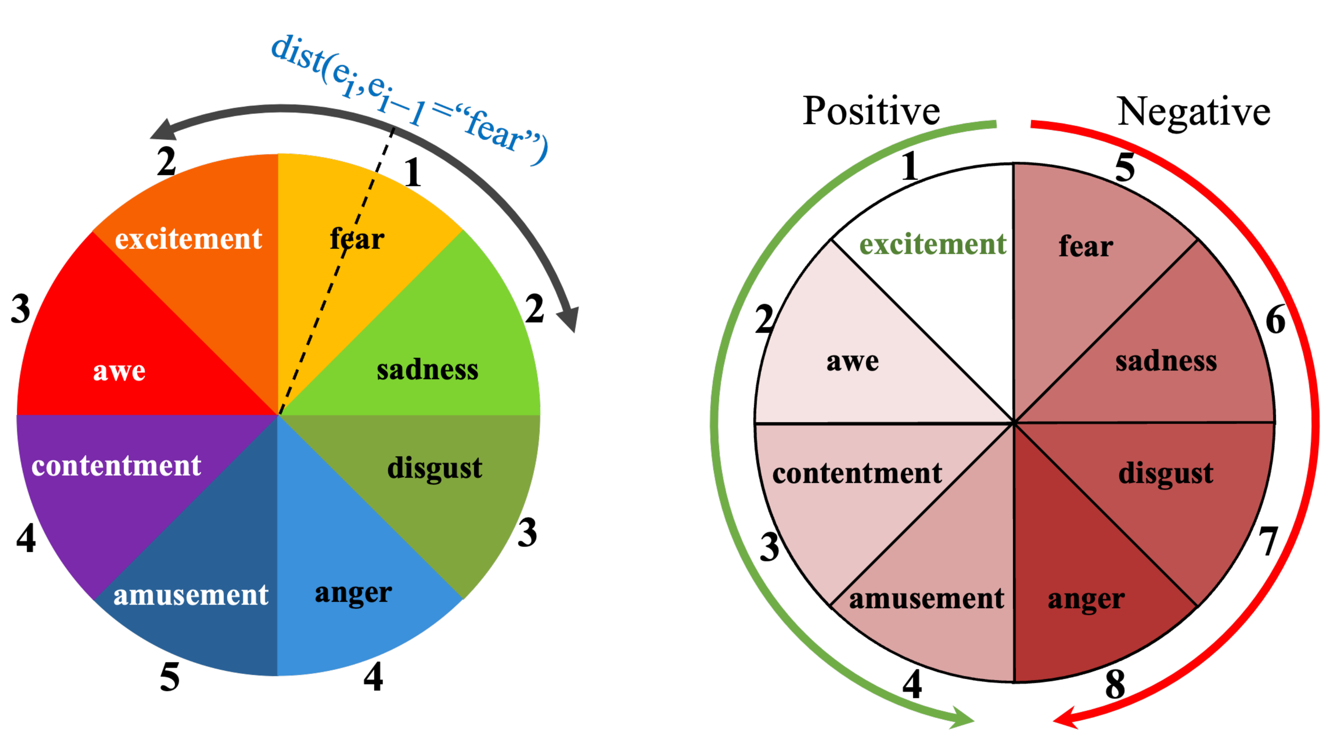
2.2 Emotion Confusion Confidence (ECC)

ACC can be re-formulated as the product of the confusion matrix and a modulation factor $M_{i,j}$ as:

$$\text{ACC} = \frac{\text{N}_\text{c}}{\text{N}} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{c} S_{i,j} \times M_{i,j}}{\text{N}}, M_{i,j} = \begin{cases} 1 & i=j \\ 0 & i \neq j \end{cases} \quad (2)$$

we rely on the reciprocal of the emotional distance $W_{i,j}$ defined in Eq. 1 to replace the modulation factor $M_{i,j}$ in Eq. 2 for the misclassified samples to obtain ECC in Eq. 3.

$$\text{ECC} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{c} S_{i,j} \times \frac{1}{W_{i,j}}}{\text{N}} = \text{ACC} + \frac{\sum_{j=1}^{c}\sum_{i=1,i\neq j}^{c} S_{i,j} \times \frac{1}{W_{i,j}}}{\text{N}} \quad (3)$$

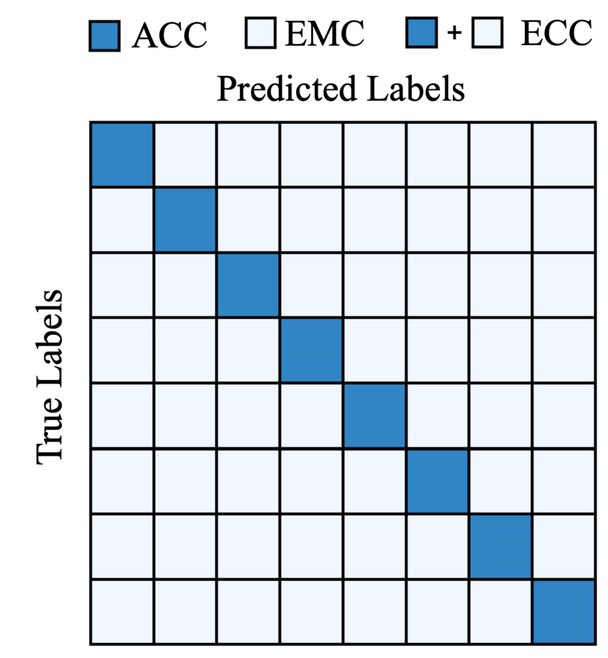


(a) Mikel's emotion distance (b) Our emotion distance/rank

2.3 Emotional Misclassification Confidence (EMC)

In order to consider the cases of misclassification, one can extract the second term that excludes ACC from Eq. 3 as an indicator for evaluating misclassification.

$$\text{EMC} = \frac{\sum_{j=1}^{c}\sum_{i=1,i\neq j}^{c} S_{i,j} \times \frac{1}{W_{i,j}-1}}{\text{N}-\text{N}_\text{c}} \quad (4)$$



(c) Confusion matrix

3.Experiment

3.1 Adjustment of Confidence Threshold

we treat the pseudo-labels generated from weakly augmented unlabeled samples as the ground truth, and use the predicted labels from strongly augmented samples as the model's predictions to calculate EMC. If EMC is high, we can lower the confidence threshold, allowing more pseudo-labeled samples to participate in the training process. We can realize the above confidence adjustment mechanism by simply setting:

$$\tau_t' = \tau \cdot \frac{e}{\textbf{EMC}_t} \quad (5)$$

3.2 Adjustment of Confidence Threshold

The ListMLE[2] loss function is designed to constrain the final prediction probability of the sample to conform to a given rank arrangement. To prove that model with higher ECC or EMC is better for pesudo labeling, we train the same network with different loss functions: cross-entropy loss and the combined loss:

$$\mathcal{L}_\text{c} = \mathcal{L}_\text{CE} + \alpha\mathcal{L}_\text{ListMLE} \quad (6)$$

The Model using combined loss have lower ACC, but the pseudo-labels become closer to the ground truth. In such cases, these pseudo labels can still have a positive impact on the training process and thereby improve the model's accuracy.

		FI			EmoSet		
	label num	80	800	1600	400	4000	8000
TA	Fixmatch [32]	28.2±0.78	37.4±0.51	42.2±0.29	31.1±0.41	42.3±0.65	45.8±1.25
TA	Flexmatch [52]	29.7±0.90	38.2±0.49	40.6±0.55	30.4±0.78	42.8±0.34	44.9±1.24
TA	Ours(4.2.1)	**31.2±0.12**	**40.8±0.34**	**42.7±0.21**	**31.6±0.56**	**43.7±0.69**	**47.6±0.61**
State-of-the-art	Comatch [20]	36.7±0.87	43.5±0.39	47.9±0.26	30.3±0.97	44.2±0.41	46.8±0.49
State-of-the-art	Simmatch [58]	31.4±1.26	41.9±0.57	43.7±0.62	36.3±0.22	44.7±0.34	50.2±0.71
State-of-the-art	Freematch [39]	26.0±1.66	37.3±0.43	39.9±0.87	31.2±1.63	41.5±0.59	46.3±0.61
State-of-the-art	Softmatch [5]	30.7±1.31	37.9±0.78	40.7±0.19	30.8±0.35	44.0±1.28	45.8±0.25
State-of-the-art	S^2-VER [16]	39.1±0.66	46.9±0.46	51.8±0.21	44.9±0.35	57.5±0.51	60.2±0.34
State-of-the-art	Ours(4.2.2)	**40.2±1.08**	**48.9±0.91**	**52.1±0.33**	**47.0±0.18**	**59.0±0.33**	**61.5±0.12**

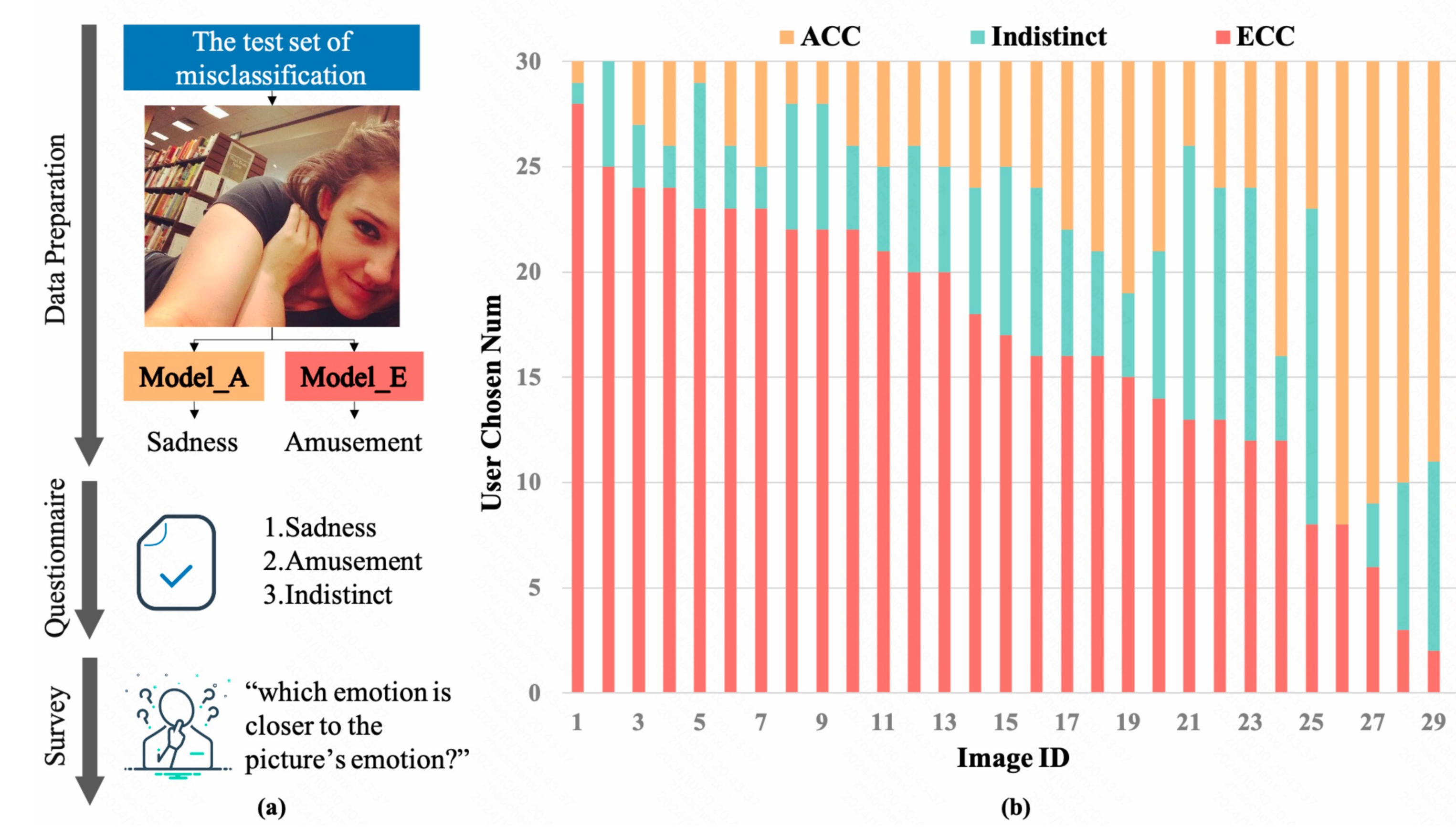
3.3 Compare with ACC_2

Dataset	FI						EmoSet					
Backbone	Resnet18		Resnet50		Resnet101		Resnet18		Resnet50		Resnet101	
Loss function	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$	$\mathcal{L}_\text{CE}$	$\mathcal{L}_\text{c}$
ACC	**65.8**	64.4	**67.6**	66.2	**68.1**	65.6	**73.9**	72.4	**76.2**	74.3	**76.7**	74.5
ACC_2	79.0	**86.2**	83.7	**86.0**	84.7	**86.0**	85.0	**85.6**	85.3	**85.8**	85.7	**85.8**
ECC	**76.1**	75.8	**77.2**	76.8	**77.9**	76.7	**82.6**	81.9	**84.2**	83.2	**84.5**	83.3
EMC	50.1	**54.8**	49.0	**53.2**	51.9	**54.8**	57.0	**59.3**	57.0	**59.9**	56.9	**60.5**

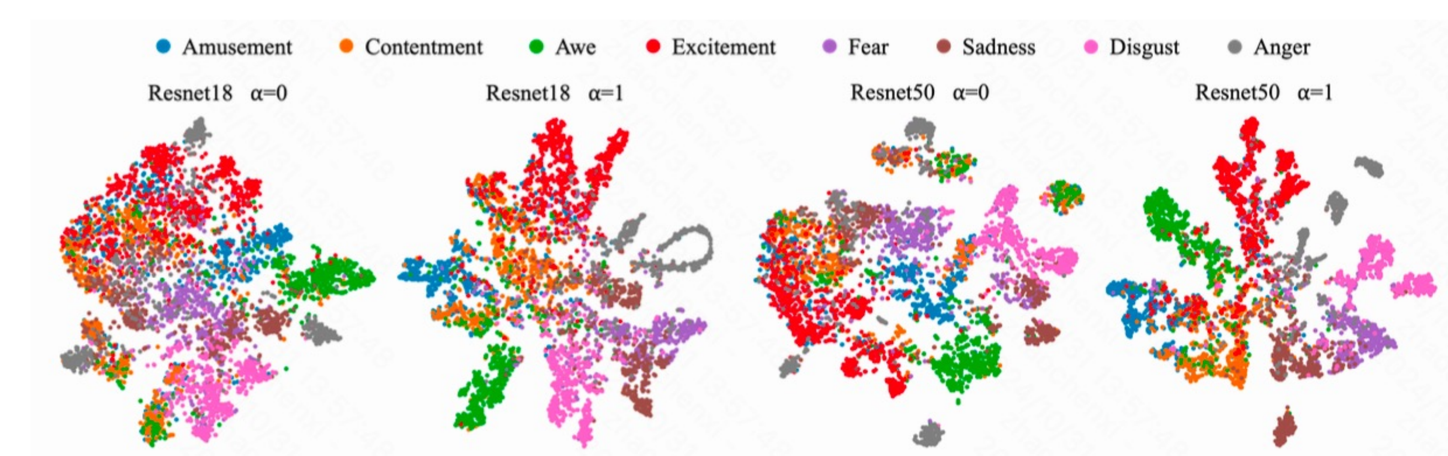
3.4 Validity of Emotional Distance Definitions

	Resnet18			Resnet50			Resnet101		
Label Rank	ACC	ECC	EMC	ACC	ECC	EMC	ACC	ECC	EMC
RE w/o R1	60.6	70.2	40.4	63.4	73.1	43.3	63.8	73.6	44.4
RA w/o R1	61.2	71.5	48.6	63.8	74.1	48.6	64.4	74.6	49.5
Our Rank	**63.9**	**75.1**	**58.4**	**65.7**	**77.0**	**51.5**	**67.9**	**77.6**	**52.2**

3.5 User study



3.6 Combined Loss analysis



Github NKU CVLab WeChat

References

[1].Zhao, Sicheng, et al. "Predicting personalized emotion perceptions of social images." Proceedings of the 24th ACM international conference on Multimedia. 2016.

[2]. Xia, Fen, et al. "Listwise approach to learning to rank: theory and algorithm." Proceedings of the 25th international conference on Machine learning. 2008.