Emotion Recognition for Multiple Context Awareness

Dingkang Yang*, Shuai Huang*, Shunli Wang, Yang Liu, Peng Zhai,
Liuzhen Su, Mingcheng Li, Lihua Zhang
Academy for Engineering and Technology, Fudan University, Shanghai, China

ECCV TEL AVIV 2022

Introduction

- **Background**: Context-Aware Emotion Recognition (CAER) is a crucial yet challenging task that aims to perceive the emotional states of the target person with the assistance of context information.
- **Motivation**: Existing CAER methods lack reliable context semantics to mitigate uncertainty in expressing emotions and fail to model multiple context representations complementarily.
- **Contributions**: We present a CAER framework from a psychological and sociological perspective, which incorporates four context information. We propose a fusion module that focuses on the interactions among diverse contexts and adaptively assigns higher weights to beneficial contexts.We release HECO, a new dataset for emotion recognition in context.

Performance

Discrete Classification Results

Category	Kosti *et al.*[30]	Zhang *et al.*[72]	Lee *et al.*[31]	Mittal *et al.*[41]	Ours L_{disc}	Ours L_{comb}	Category	Kosti *et al.*[30]	Zhang *et al.*[72]	Lee *et al.*[31]	Mittal *et al.*[41]	Ours L_{disc}	Ours L_{comb}
Peace	22.35	30.68	19.55	**35.72**	25.5	26.24	Affection	26.47	**47.52**	22.36	38.55	41.61	37.66
Esteem	17.86	12.05	15.38	**25.75**	21.98	20.29	Anticipation	57.31	63.2	52.85	60.73	62.75	**63.31**
Engagement	86.69	**87.31**	73.71	86.23	74.69	75.23	Confidence	**80.33**	74.83	72.68	68.12	72.22	74.42
Happiness	58.92	72.9	53.73	80.45	83.58	**85.25**	Pleasure	46.72	48.37	34.12	67.31	67.26	**67.68**
Excitement	78.05	72.68	70.42	80.75	85.64	**86.56**	Surprise	22.38	8.44	17.46	19.6	25.31	**27.03**
Sympathy	15.23	19.45	14.89	16.74	24.7	**25.87**	Doubt/Confusion	31.88	19.67	26.07	**38.43**	23.44	24.96
Disconnection	20.64	23.17	22.01	28.73	27.64	**28.95**	Fatigue	8.87	12.93	6.29	19.35	32.35	**33.58**
Embarrassment	3.05	1.58	1.88	10.31	9.63	**10.57**	Yearning	9.22	9.86	4.84	**15.08**	10.88	11.12
Disapproval	16.14	12.64	15.37	18.55	23.41	**23.52**	Aversion	7.44	6.81	3.26	11.33	13.19	**15.28**
Annoyance	15.26	12.33	14.42	24.68	28.98	**29.02**	Anger	11.24	11.27	12.88	14.69	15.47	**17.84**
Sensitivity	9.05	4.74	6.94	13.94	22.53	**24.89**	Sadness	18.69	23.9	17.75	40.26	46.75	**47.8**
Disquietment	19.57	17.66	10.84	**22.14**	19.36	21.17	Fear	15.7	6.15	7.47	16.99	36.06	**36.68**
Pain	9.46	8.22	8.16	14.68	18.26	**19.27**	Suffering	17.67	23.71	14.85	**48.05**	45.37	46.74
							mAP	27.93	28.16	23.85	35.28	36.87	**37.73**

Performance on EMOTIC

Category	Kosti *et al.*[30]	Zhang *et al.*[72]	Lee *et al.*[31]	Mittal *et al.*[41]	Ours L_{disc}	Ours L_{comb}
Surprise	28.45	34.87	24.27	**38.37**	38.04	38.12
Excitement	42.16	45.74	37.97	48.59	53.2	**55.04**
Happiness	62.82	63.26	55.81	66.53	67.26	**69.16**
Peace	51.64	54.17	47.57	55.97	57.23	**57.31**
Disgust	45.37	49.43	41.74	50.48	52.28	**54.95**
Anger	40.76	45.22	38.39	51.29	53.04	**53.43**
Fear	32.74	35.67	30.51	**40.81**	40.08	40.27
Sadness	22.53	27.28	20.92	32.65	34.17	**36.94**
mAP	40.81	44.46	37.15	48.09	49.41	**50.65**

Performance on HECO

Category	Kosti *et al.*[30]	Zhang *et al.*[72]	Lee *et al.*[31]	Mittal *et al.*[41]	Ours
Anger	57.65	51.92	45.18	68.85	**70.54**
Happy	71.32	63.37	56.59	72.31	**72.38**
Neutral	43.1	40.26	39.32	50.34	**52.54**
Sad	61.24	58.15	52.96	70.8	**71.42**
mAP	58.33	53.43	48.51	65.58	**66.72**

Performance on GroupWalk

Method	Accuracy
Fine-tuned VGGNet[9]	64.85
Fine-tuned ResNet[4]	68.46
CAER-Net-S[6]	73.51
Ours	**79.57**

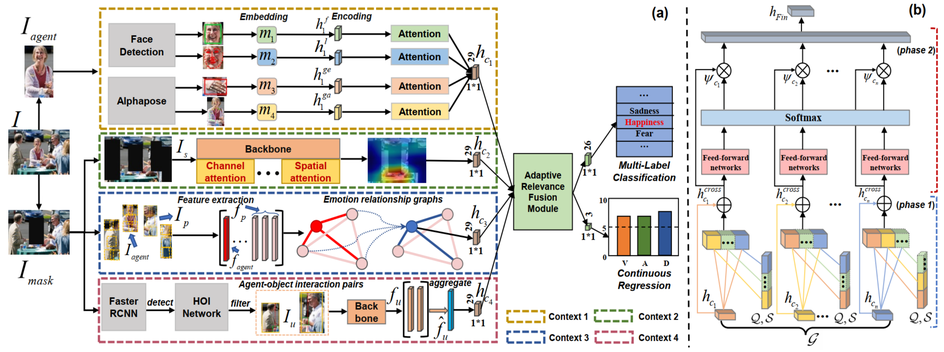
Performance on CAER-S

Continuous Regression Results

Method	Dataset	Valence	Arousal	Dominance	mER	Dataset	Valence	Arousal	Dominance	mER
Kosti *et al.*[30](L_{cont})	EMOTIC	1.0	1.5	0.8	1.1	HECO	0.9	1.3	0.8	1.0
Kosti *et al.*[30](L_{comb})		0.9	1.2	0.9	1.0		0.9	1.2	**0.6**	0.9
Zhang *et al.*[72](L_{cont})		0.8	1.6	1.2	1.2		0.9	1.1	1.0	1.0
Zhang *et al.*[72](L_{comb})		0.7	1.0	1.0	0.9		**0.6**	1.1	0.7	0.8
Ours(L_{cont})		**0.6**	1.3	0.8	0.9		0.8	1.0	**0.6**	0.8
Ours(L_{comb})		0.8	**0.9**	**0.7**	**0.8**		0.7	**0.8**	**0.6**	**0.7**

Performance on EMOTIC and HECO

Method

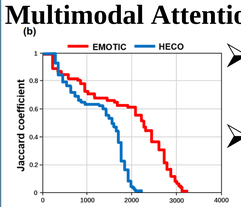


Overview of the Proposed Method

- **Context 1** is the agent-centric multimodal emotion recognition.
- **Context 2** adopts the channel and spatial attention modules to obtain the emotion semantics of the scene context.
- **Context 3** explores the emotion transmission among agents.
- **Context 4** aggregates the emotion cues from the interactions between surrounding agents and objects.

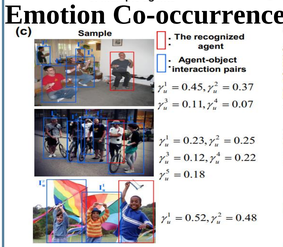
Visualization

Multimodal Attention **Emotion Semantic Capture of Scene**

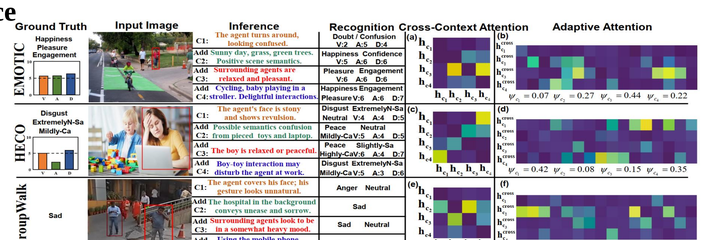


Emotion Co-occurrence

- Interesting visual examples of each context branch reveal different aspects of the emotional phenomenon.
- Cross-context and adaptive attention demonstrates the complementary recognition ability of the proposed contexts.

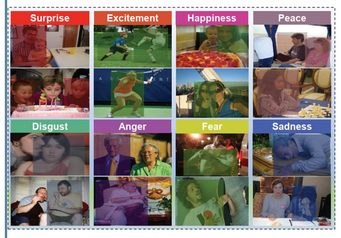


Aggregation Strategy



Complementary Recognition Analysis of Four Contexts

Dataset



The dataset contains 9,385 images and 19,781 annotated agents. Eight categories: Surprise, Excitement, Peace, Happiness, Disgust, Anger, Fear, and Sadness.

Discrete Emotion Categories

Different Scores of Valence, Arousal, and Dominance

Ablation Study

Model Design	Dataset EMOTIC mAP	EMOTIC mER	GroupWalk mAP	GroupWalk mER	HECO mAP	HECO mER
Full (ours)	**37.73**	**0.8**	**66.72**	-	**50.65**	**0.7**
C_1	22.51	1.6	44.76	-	37.27	1.4
$C_1 + C_2$	29.23	1.1	54.42	-	41.93	1.0
$C_1 + C_3$	27.56	1.3	57.36	-	39.62	1.1
$C_1 + C_4$	26.29	1.2	52.09	-	37.93	1.2
$C_1 + C_2 + C_3$	36.18	0.8	64.34	-	48.07	0.8
$C_1 + C_2 + C_4$	34.93	0.9	60.27	-	47.25	0.8
$C_1 + C_3 + C_4$	33.45	1.0	62.61	-	44.23	0.9
C_2 (Mask Face)	35.57	1.0	64.34	-	47.71	1.0
C_2 (Mask Body)	37.02	0.9	65.12	-	49.46	0.8

Model Design	Dataset EMOTIC mAP	EMOTIC mER	GroupWalk mAP	GroupWalk mER	HECO mAP	HECO mER
C_3 (GCNs) [28]	35.25	1.2	63.94	-	48.84	1.1
C_3 (Depth) [32]	36.39	1.0	65.19	-	49.02	0.9
Concatenation [52]	30.47	1.2	59.87	-	43.6	1.1
Multiplication [41]	36.52	0.9	65.24	-	50.22	0.7
ARF (phase 1)	36.27	0.9	65.33	-	49.73	0.8
ARF (phase 2)	34.65	1.0	64.13	-	47.51	0.9
C_1 (OpenFace [1])+OpenPose [5])	37.45	0.8	66.39	-	50.28	0.8
C_4 (R-FCN [13]+HOI-Net [33])	37.52	0.9	66.45	-	50.34	0.7
VGG19 [56] (C_1,C_2)+Res101 [22](C_3, C_4)	37.76	0.9	66.68	-	50.87	0.7
Res34[22] (C_1,C_2)+Res152 [22](C_3,C_4)	36.83	0.8	65.85	-	50.24	0.8

We analyse the effect of context branches, masking strategies, fusion strategies, detectors, and the CNN backbones on model performance through several ablation studies.

Conclusions

- We explore emotion-rich representations from contexts at the visual level to advance the development of effective visual-only driven emotion recognition applications.
- Numerous qualitative and quantitative analyses clearly demonstrate the superiority of our method.