Context De-confounded Emotion Recognition

Dingkang Yang[1,2,3,4], Zhaoyu Chen[1], Yuzheng Wang[1], Shunli Wang[1], Mingcheng Li[1],Siao Liu[1], Xiao Zhao[1], Shuai Huang[1], Zhiyan Dong[1], Peng Zhai[1], Lihua Zhang[1,2,3,4]

[1]Academy for Engineering and Technology, Fudan University [2]Institute of Meta-Medical, IPASS

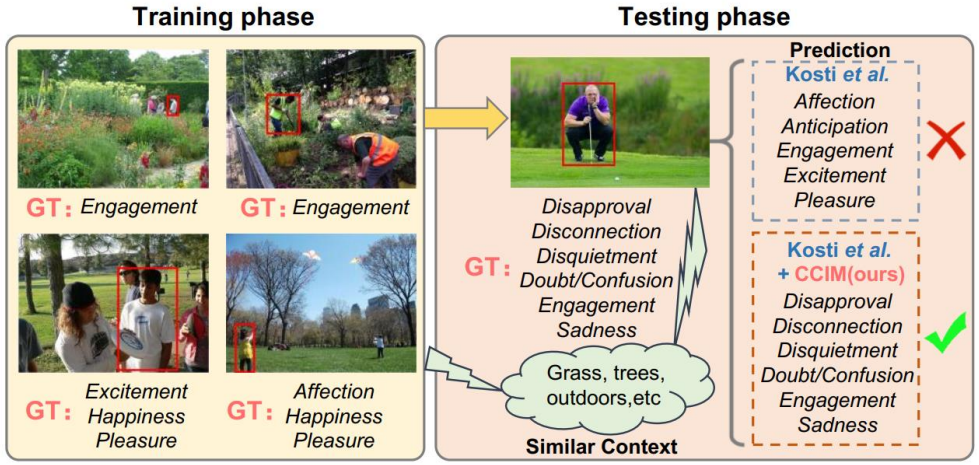
[3]Jilin Provincial Key Laboratory of Intelligence Science and Engineering

[4]Engineering Research Center of AI and Robotics, Ministry of Education

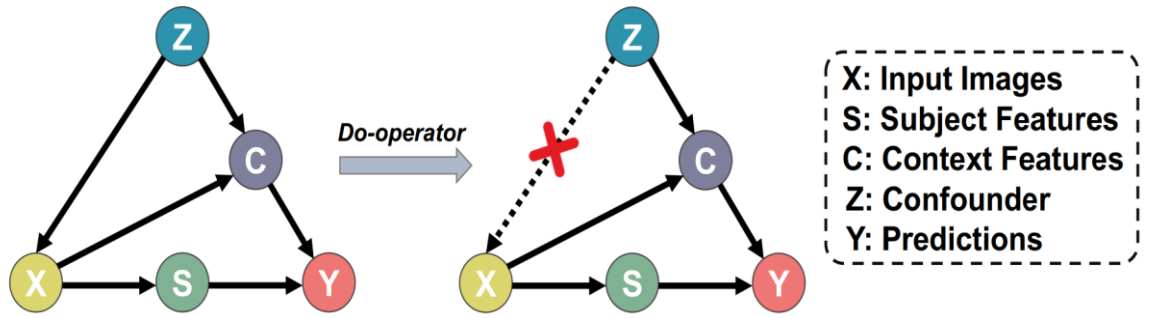
Poster Tag: **THU-AM-241**

Quick Preview

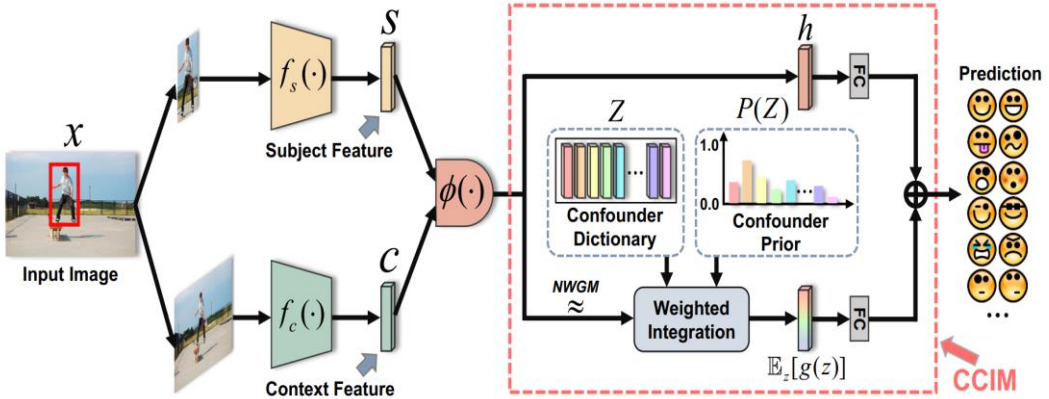
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



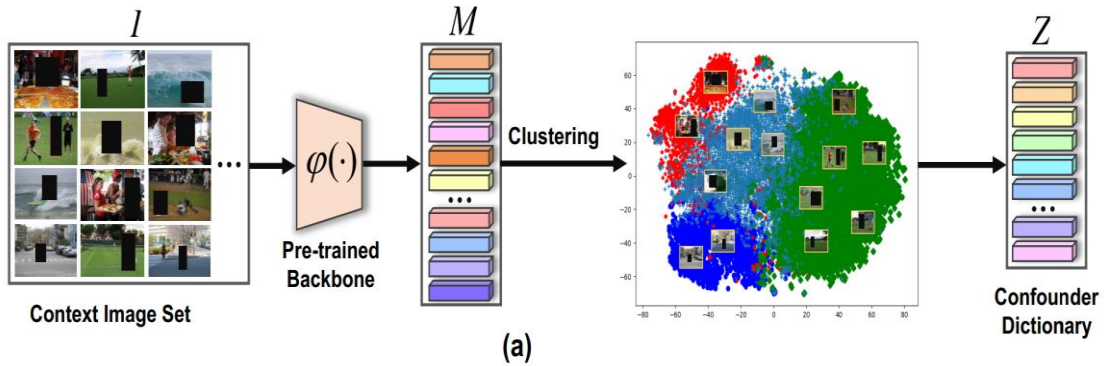
The context bias in the Context-Aware Emotion Recognition (CAER) task



The structured causal model for the CAER task



The context-deconfounded training by the proposed Contextual Causal Intervention (CCIM) Module



The confounder dictionary generation

Background & Motivation

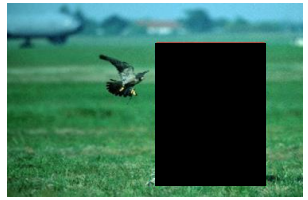
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Context-Aware Emotion Recognition (CAER)

Input image

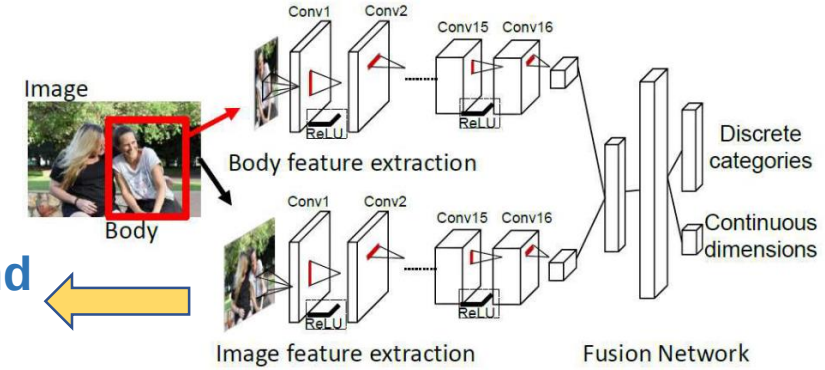
Person information

Context information



Emotion recognition

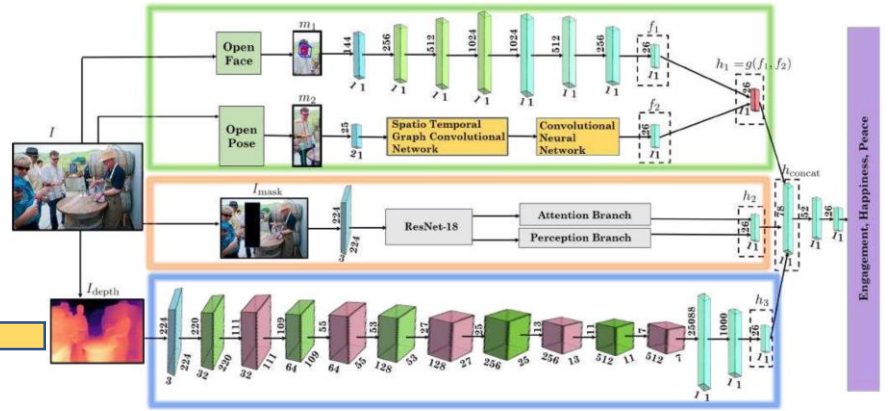
Background context



EMOT-Net

Kosti R, Alvarez J M, Recasens A, et al. Emotion recognition in context. In CVPR 2017: 1667-1675.

Scene and socio-dynamic contexts



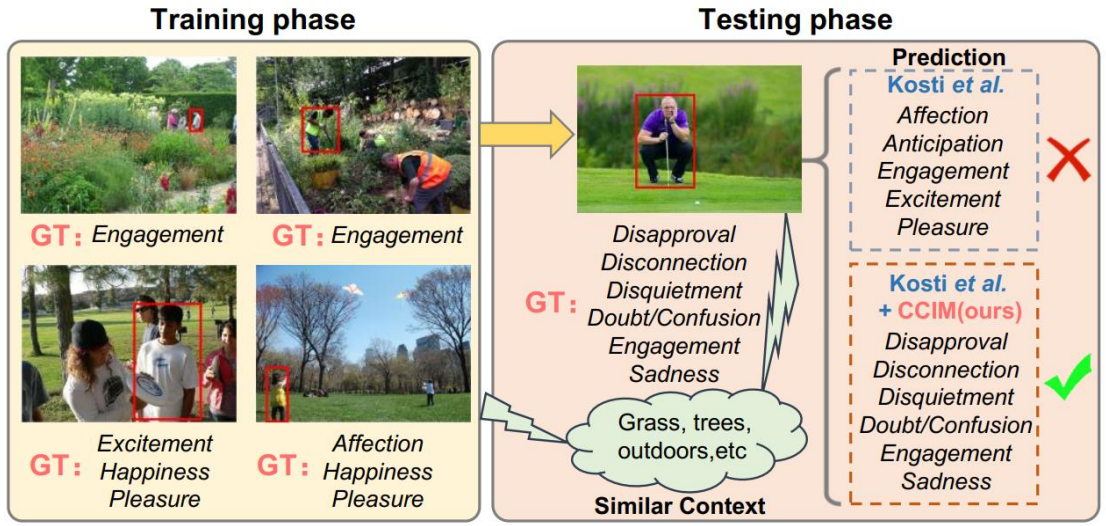
EmotiCon

Mittal T, Guhan P, Bhattacharya U, et al. Emoticon: Context-aware multimodal emotion recognition using frege's principle. In CVPR 2020: 14234-14243.

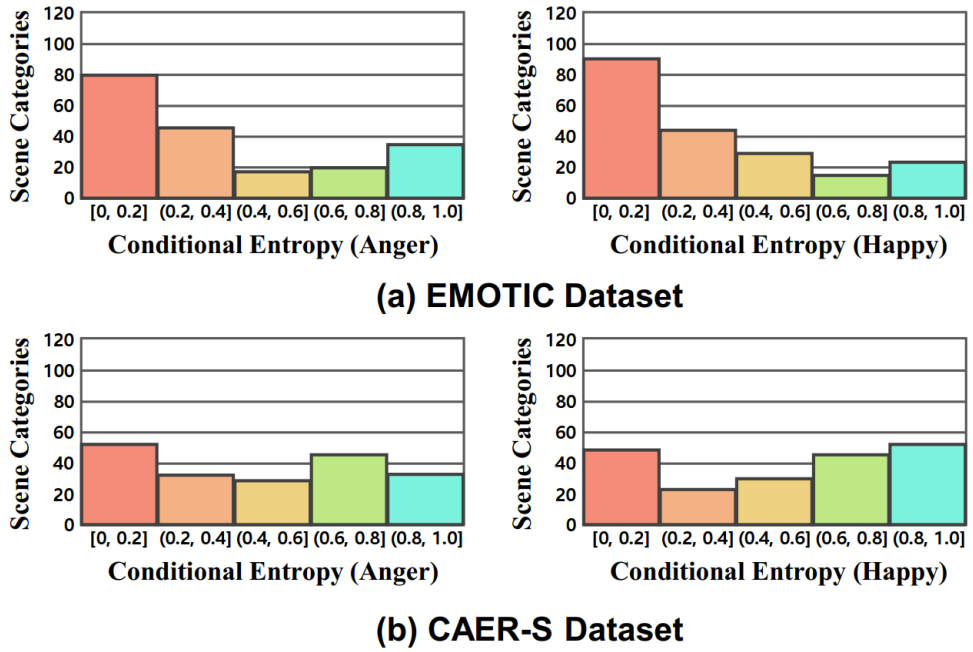
Background & Motivation

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

The Context Bias in the CAER Task



- Most images contain similar contexts in the training data with positive emotion categories.
- The baseline learns the spurious correlation between specific contexts and emotion categories and gives wrong results.

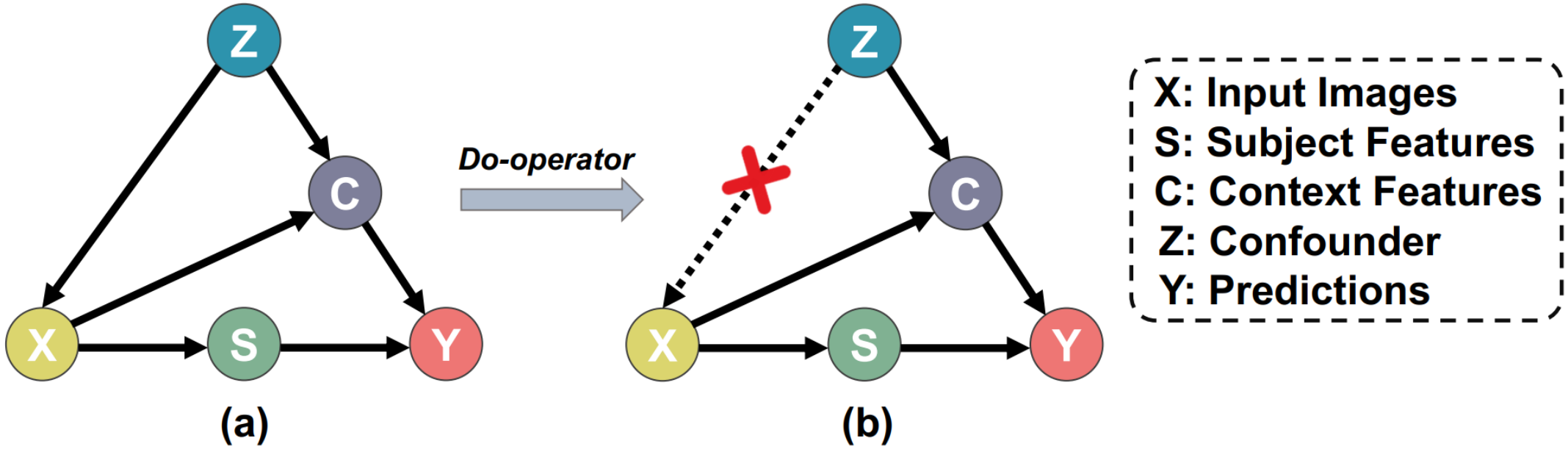


- For the EMOTIC dataset, about 40% of scene categories for anger have zero conditional entropy.
- About 45% of scene categories for happy (i.e., happiness) have zero conditional entropy.

Methodology

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Causal View at CAER Task



(a) The conventional likelihood

(b) The causal intervention

- The positive effects provided by contexts and subjects follow the path $X \rightarrow C/S \rightarrow Y$.
- The negative effects provided by the confounder Z follow the backdoor path $X \leftarrow Z \rightarrow C \rightarrow Y$.

Methodology

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

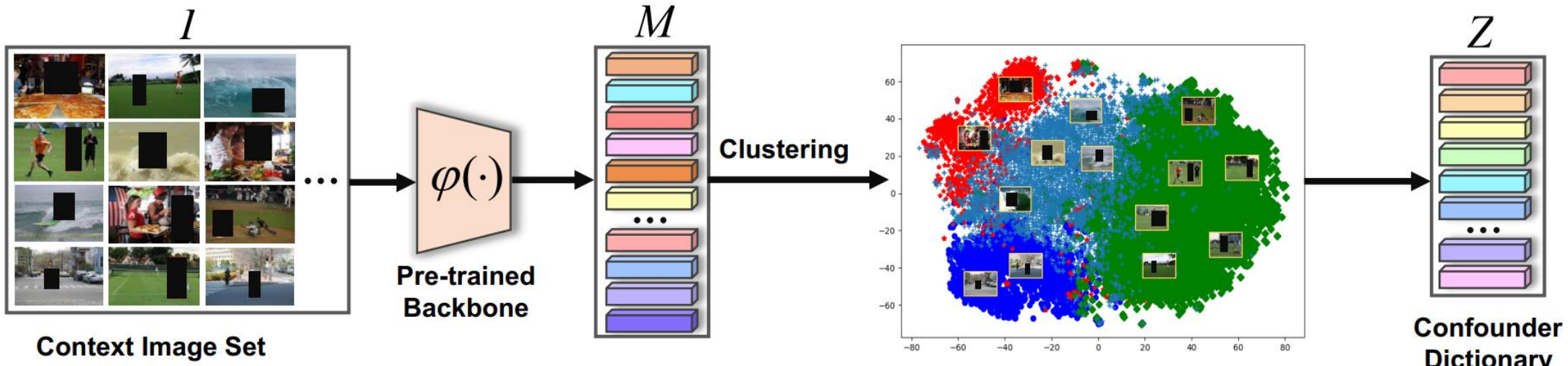
Causal Intervention via Backdoor Adjustment

Existing CAER methods rely on the traditional likelihood:

$$P(\boldsymbol{Y}|\boldsymbol{X}) = \sum_{\boldsymbol{z}} P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{S} = f_s(\boldsymbol{X}), \boldsymbol{C} = f_c(\boldsymbol{X}, \boldsymbol{z}))P(\boldsymbol{z}|\boldsymbol{X})$$

Causal intervention via backdoor adjustment:

$$P(\boldsymbol{Y}|do(\boldsymbol{X})) = \sum_{\boldsymbol{z}} P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{S} = f_s(\boldsymbol{X}), \boldsymbol{C} = f_c(\boldsymbol{X}, \boldsymbol{z}))P(\boldsymbol{z})$$



The confounder dictionary generation

Methodology

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

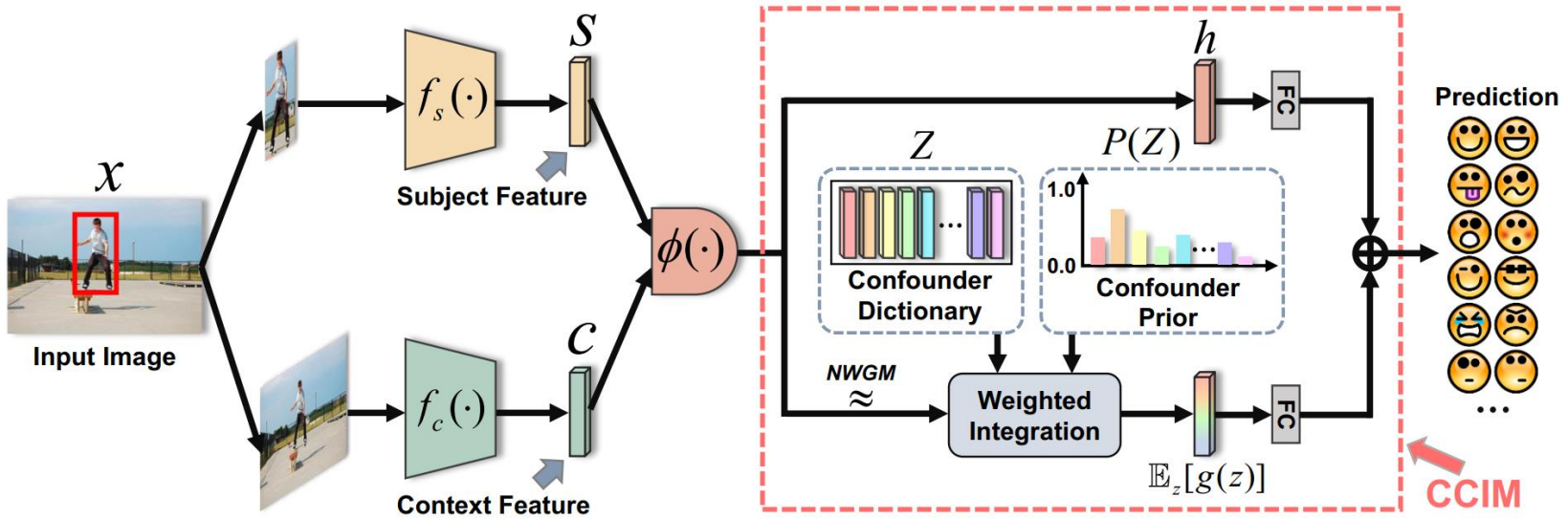
Context-Deconfounded Training with CCIM

Expectation approximation at feature level:

$$P(\boldsymbol{Y}|do(\boldsymbol{X})) \approx P(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{S} = f_s(\boldsymbol{X}), \boldsymbol{C} = \sum_{\boldsymbol{z}} f_c(\boldsymbol{X}, \boldsymbol{z}) P(\boldsymbol{z}))$$

A model is instantiated to approximate the conditional probability:

$$P(\boldsymbol{Y}|do(\boldsymbol{X})) = \boldsymbol{W}_h \boldsymbol{h} + \boldsymbol{W}_g \mathbb{E}_{\boldsymbol{z}}[g(\boldsymbol{z})] \quad \text{where} \quad \mathbb{E}_{\boldsymbol{z}}[g(\boldsymbol{z})] = \sum_{i=1}^{N} \lambda_i \boldsymbol{z}_i P(\boldsymbol{z}_i)$$



Experiments

Category	EMOT-Net [19]	EMOT-Net + CCIM	GCN-CNN [56]	GCN-CNN + CCIM	CAER-Net [22]	CAER-Net + CCIM	RRLA [23]	VRD [16]	EmotiCon [29]	EmotiCon + CCIM
Affection	26.47	34.87	47.52	36.18	22.36	23.08	37.93	44.48	38.55	40.77
Anger	11.24	13.05	11.27	12.53	12.88	12.99	13.73	30.71	14.69	15.48
Annoyance	15.26	18.04	12.33	13.73	14.42	15.28	20.87	26.47	24.68	24.47
Anticipation	57.31	94.19	63.2	92.32	52.85	90.03	61.08	59.89	60.73	95.15
Aversion	7.44	13.41	6.81	15.41	3.26	12.96	9.61	12.43	11.33	19.38
Confidence	80.33	74.9	74.83	75.01	72.68	73.24	80.08	79.24	68.12	75.81
Disapproval	16.14	19.87	12.64	14.45	15.37	16.38	21.54	24.54	18.55	23.65
Disconnection	20.64	27.72	23.17	30.52	22.01	23.39	28.32	34.24	28.73	31.93
Disquietment	19.57	19.12	17.66	20.85	10.84	18.1	22.57	24.23	22.14	26.84
Doubt/Confusion	31.88	19.35	19.67	20.43	26.07	17.66	33.5	25.42	38.43	34.28
Embarrassment	3.05	6.23	1.58	9.21	1.88	5.86	4.16	4.26	10.31	16.73
Engagement	86.69	88.93	87.31	96.88	73.71	70.04	88.12	88.71	86.23	97.41
Esteem	17.86	21.69	12.05	22.72	15.38	16.67	20.5	17.99	25.75	27.44
Excitement	78.05	73.81	72.68	73.21	70.42	71.08	80.11	74.21	80.75	81.59
Fatigue	8.87	9.96	12.93	12.66	6.29	9.73	17.51	22.62	19.35	15.53
Fear	15.7	9.04	6.15	10.31	7.47	6.61	15.56	13.92	16.99	15.37
Happiness	58.92	78.09	72.9	75.64	53.73	62.34	76.01	83.02	80.45	83.55
Pain	9.46	14.71	8.22	15.36	8.16	9.43	14.56	16.68	14.68	17.76
Peace	22.35	22.79	30.68	23.88	19.55	20.21	26.76	28.91	35.72	38.94
Pleasure	46.72	46.59	48.37	45.52	34.12	35.37	55.64	55.47	67.31	64.57
Sadness	18.69	17.47	23.9	22.08	17.75	13.24	30.8	42.87	40.26	45.63
Sensitivity	9.05	7.91	4.74	8.02	6.94	4.74	9.59	15.89	13.94	17.04
Suffering	17.67	15.35	23.71	18.45	14.85	11.89	30.7	46.23	48.05	21.52
Surprise	22.38	13.12	8.44	13.93	17.46	11.7	17.92	16.27	19.6	26.81
Sympathy	15.23	32.6	19.45	33.95	14.89	28.59	15.26	15.37	16.74	47.6
Yearning	9.22	10.08	9.86	11.58	4.84	8.61	10.11	10.04	15.08	12.25
mAP	27.93†	**30.88**† (↑ 2.95)	28.16†	**31.72**† (↑ 3.56)	23.85†	**26.51**† (↑ 2.66)	32.41*	35.16*	35.28†	**39.13**† (↑ 3.85)

Table 1. Average precision (%) of different methods for each emotion category on the EMOTIC dataset. ∗: results from the original reports. †: results from implementation. The footnotes ∗ and † of Tables 2 and 3 follow the same interpretation.

- CCIM helps raise the results of "Anticipation" and "Sympathy" in these CAER methods by 29%~37% and 14%~29%, respectively.
- Thanks to CCIM, the AP scores in "Aversion" and "Embarrassment" categories are achieved at about 12%~19% and 5%~16%.

Experiments

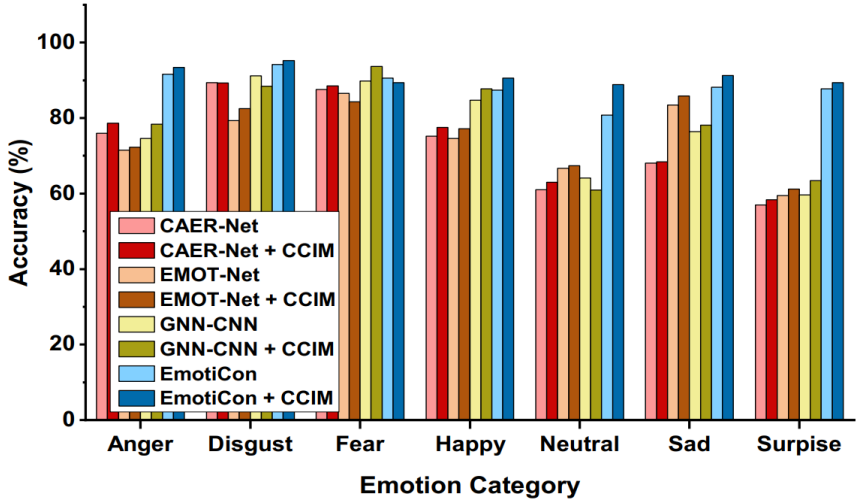
JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Category	EMOT-Net [19]	EMOT-Net + **CCIM**	GNN-CNN [56]	GNN-CNN + **CCIM**	CAER-Net [22]	CAER-Net + **CCIM**	EmotiCon [29]	EmotiCon + **CCIM**
Angry	57.65	62.41	51.92	54.07	45.18	50.43	68.85	75.93
Happy	71.32	75.68	63.37	70.25	56.59	60.71	72.31	79.15
Neutral	43.1	41.03	40.26	39.49	39.32	37.84	50.34	48.66
Sad	61.24	63.84	58.15	61.85	52.96	54.06	70.8	73.48
mAP	58.33†	**60.74**† (↑ 2.41)	53.43†	**56.42**† (↑ 2.99)	48.51†	**50.76** † (↑ 2.25)	65.58†	**69.31** † (↑ 3.73)

Table 2. Average precision (%) of different methods for each emotion category on the GroupWalk dataset.

Methods	**Accuracy** (%)	**Methods**	**Accuracy** (%)
CAER-Net [22]	73.47†	EmotiCon [29]	88.65†
CAER-Net + **CCIM**	**74.81**† (↑ 1.34)	EmotiCon + **CCIM**	**91.17**† (↑ 2.52)
EMOT-Net [19]	74.51†	SIB-Net [24]	74.56*
EMOT-Net + **CCIM**	**75.82**† (↑ 1.31)	GRERN [12]	81.31*
GNN-CNN [56]	77.21†	RRLA [23]	84.82*
GNN-CNN + **CCIM**	**78.66**† (↑ 1.45)	VRD [16]	90.49*

Table 3. Emotion classification accuracy (%) of different methods on the CAER-S dataset.



The performance of EMOT-Net, GCN-CNN, CAER-Net, and EmotiCon is consistently increased by CCIM, making each context prototype contribute fairly to the emotion classification results.

Experiments

ID	Setting	EMOTIC mAP (%)	CAER-S Accuracy (%)	GroupWalk mAP (%)
(1)	EMOT-Net + **CCIM**	**30.88**	**75.82**	**60.74**
(2)	EmotiCon + **CCIM**	**39.13**	**91.17**	**69.31**
(3)	(1) w/ Random $\boldsymbol{Z}$	26.56	73.36	57.45
(4)	(2) w/ Random $\boldsymbol{Z}$	35.12	87.34	65.62
(5)	(1) w/ ImageNet Pre-training	28.72	74.75	58.96
(6)	(2) w/ ImageNet Pre-training	37.48	90.46	68.28
(7)	(1) w/ ResNet-50	29.53	75.34	59.92
(8)	(2) w/ ResNet-50	38.86	90.41	68.85
(9)	(1) w/ VGG-16	28.78	74.95	59.47
(10)	(2) w/ VGG-16	37.93	89.82	68.11
(11)	(1) w/ Additive Attention	30.79	75.64	**60.85**
(12)	(2) w/ Additive Attention	**39.16**	91.08	69.26
(13)	(1) w/o λ_i	30.05	75.21	59.83
(14)	(2) w/o λ_i	38.53	89.67	68.75
(15)	(1) w/o $P(\boldsymbol{z}_i)$	30.63	75.59	59.94
(16)	(2) w/o $P(\boldsymbol{z}_i)$	39.05	90.06	69.15
(17)	(1) w/o Masking Strategy	29.86	74.84	59.22
(18)	(2) w/o Masking Strategy	38.06	90.57	67.79

Rationality of Confounder Dictionary

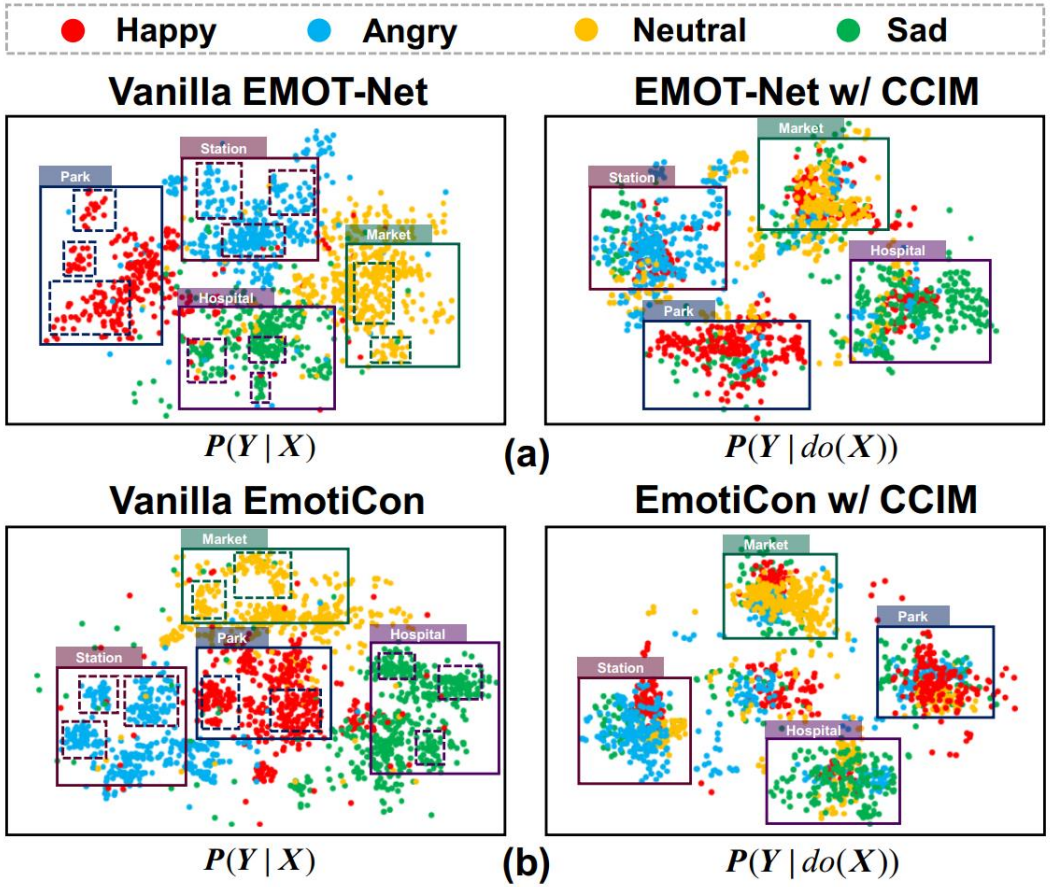
Robustness of Pre-trained Backbones

Effectiveness of Components

Necessity of Masking Strategy

Experiments

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA



Difference between likelihood and intervention

	Input Image	Ground Truth	Vanilla Method	w/ CCIM
EMOTIC		Engagement Happiness Peace Pleasure Surprise	Annoyance Doubt/Confusion Sadness Suffering	Happiness Engagement Peace Pleasure Surprise
		Anticipation Engagement Sadness	Aversion Engagement Peace Suffering	Anticipation Engagement Fatigue Sadness
CAER-S		Neutral	Happy	Neutral
		Disgust	Anger	Disgust
GroupWalk		Neutral	Sad	Neutral
		Angry	Happy Neutral	Angry

Case study of causal intervention

Conclusion

- ✓ Our causal debiasing strategy effectively mitigates the harmful bias of uneven distribution of emotional states across diverse contexts in the CAER task.

- ✓ We believe that the model-agnostic and lightweight CCIM provides better insights for the community than the complex module stacking in previous CAER methods.

Thanks!