Cascade Attention Networks For Group Emotion Recognition with Face, Body and Image Cues

Kai Wang, Xiaoxing Zeng, Jianfei Yang, Debin Meng, Kaipeng Zhang, Xiaojiang Peng* and Yu Qiao*

Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
kai.wang@siat.ac.cn

Contributions

- We propose three types of CNNs, namely face based emotion CNNs, global image based CNNs and body based CNNs.
- Particularly, we introduce an effective attention networks for face based emotion CNNs.

Introduction of Our Approaches

Global Image Based CNNs

- A global image can provide an important clue for group-level emotion prediction. Therefore, we train global image based CNNs namely VGG19, ResNet101, and SE-net154.

Table 1: Results of global image based CNN models on the EmotiW validation set.

	VGG19		ResNet101		SE-net154	
	Softmax	L-Softmax	Softmax	L-Softmax	Softmax	L-Softmax
Accuracy	67.2	**73.0**	72.8		**74.9**	

Body Based CNNs

- Body information can be helpful from previous work. We use OPENPOSE to extract all the human bodies and we finetune ResNet101 and SE-net154 for human bodies.

Table 2: Results of body based CNN models on the EmotiW validation set.

	ResNet101	SE-net154
	Softmax	Softmax
Accuracy	69.05	**70.5**

Face based Emotion CNNs

- Facial emotion for each face is one useful cue for prediction of group emotion. We utilize two kinds of networks, namely the aligned facial emotion CNN and cascade attention networks.

Table 3: Results of individual facial emotion CNN models and Cascade Attention Networks on the EmotiW validation set.

	Aligned Faces		Cascade Attention Networks
	Softmax	L-Softmax	Softmax
ResNet64	-	70.5	-
VGG-FACE	72.5	-	-
ResNet34	72.3	-	-
ResNet18	69.5	-	71.9
SE-net154	70.6	-	-

Our Pipeline

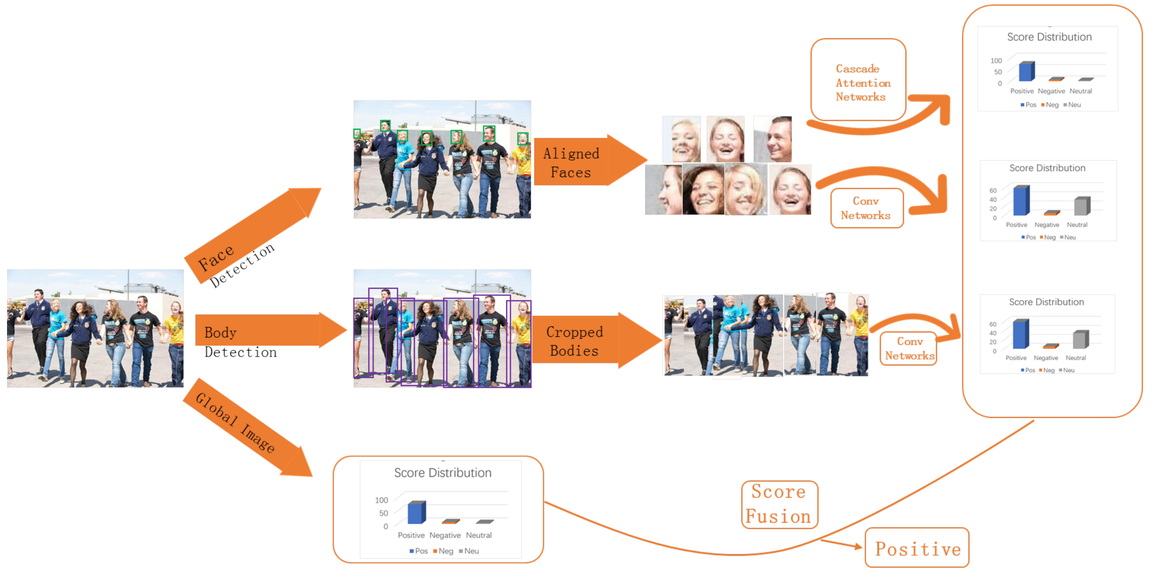

Figure 1: The system pipeline of our approach. It contains three kinds of CNNs, namely the face based emotion CNNs, the global image based CNNs and the body based CNNs. Particularly, we train two types of face based emotion networks, namely the norm individual face CNNs and cascade attention networks. The final prediction is made by averaging all the scores of CNNs from all faces, the global image and all bodies.

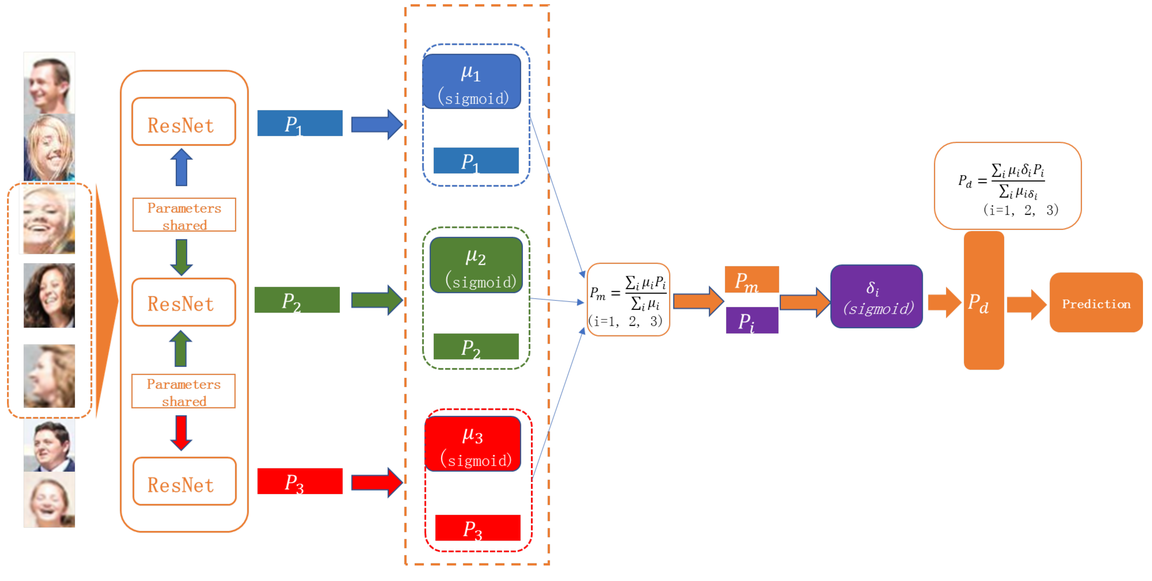

Figure 2: The structure of our cascade attention networks. It contains two stages,one is calculate each face contribution for the whole image, the other is use these contributions to predict the result of the whole image.

Our Submissions

1. face: VGG-FACE, ResNet34, SE-net154, ResNet18(CAN); global: ResNet101, VGG19, SE-NET154; body: SE-net154 ***(average)***
2. face: VGG-FACE, ResNet34, ResNet64, SE-net154, ResNet18(CAN); global: ResNet101, SE-net154, VGG19; body: SE-net154 ***(average)***
3. face: VGG-FACE, ResNet34, ResNet64; global: ResNet101, SE-net154, VGG19; body: SE-NET154
4. face: VGG-FACE, ResNet34, ResNet64, ResNet18(CAN); global: ResNet101, SE-net154, VGG19; body: SE-NET154
5. face: VGG-FACE, ResNet34, ResNet64, ResNet18(CAN); global: ResNet101, VGG19, SE-net; body:SE-net154; ***(average)***
6. face: VGG-FACE, ResNet34, SE-net154, ResNet64, ResNet18(CAN); global: ResNet101, VGG19, SE-NET154; ***(average)***
7. face: VGG-FACE, ResNet34, ResNet64, SE-net154, ResNet18(CAN); global: ResNet101, SE-net154, VGG19; body: SE-NET154 ***(only train on the training set)***

Table 4: Results of our final submissions.

Runs	Validation	Test			
	Overall	Positive	Neutral	Negative	Overall
1	86.7	77.09	55.89	65.62	**67.48**
2	86.53	77.01	56.33	64.41	67.25
3	86.26	75.27	56.00	63.93	66.29
4	86.53	76.46	56.76	63.20	66.82
5	86.9	76.93	57.20	63.44	67.22
6	86.28	77.25	55.89	65.37	**67.48**
7	86.9	75.90	53.71	64.17	65.92

Conclusions

- We present our approach for the group-level emotion recognition in the Emotion Recognition in the Wild Challenge 2018.
- We propose three types of Convolutional Neutral Networks(CNNs), namely face based emotion CNNs, global image based CNNs and body based CNNs.
- Particularly, we introduce an effective attention networks for face based emotion CNNs.
- We utilize a large-margin softmax loss for discriminative learning, and explore different fusion strategies. Experimental results show the effectiveness of our approach, and we win second place of the group-level emotion recognition task.

中国科学院深圳先进技术研究院
SHENZHEN INSTITUTES OF ADVANCED TECHNOLOGY
CHINESE ACADEMY OF SCIENCES