MRP-Net: A Light Multiple Region Perception Neural Network for Multi-label AU Detection

Speaker: Yang Tang

Yang Tang[1], Shuang Chen, Honggang Zhang*, Gang Wang, Rui Yang
Pattern Recognition and Intelligent System Lab. (P.R.I.S. Lab.),
Beijing University of Posts and Telecommunications (BUPT)

北京邮電大学
Beijing University of Posts and Telecommunications

p.r.i.s laboratory

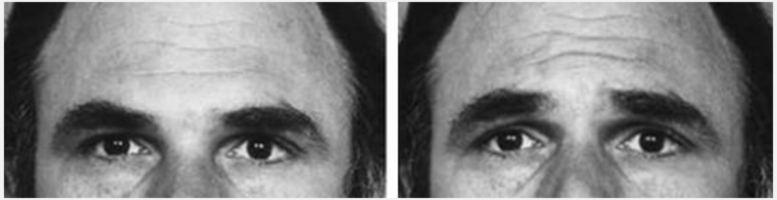
ICPR 2020

AUs and FACS

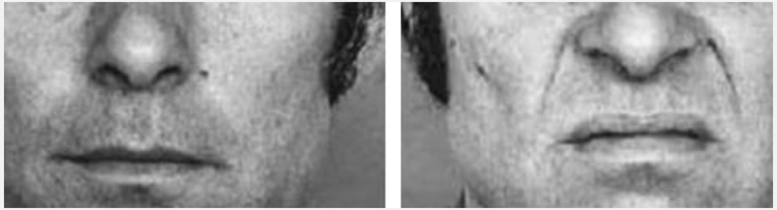
Facial expressions are intuitive responses to human emotions and natural ways of communication. Facial expressions detection has been an important topic in the field of Computer Vision.

AUs can be labeled with Facial Action Coding System(FACS). In FACS, each expression is considered to be composed of multiple AUs, which can effectively eliminate ambiguity in labeling.

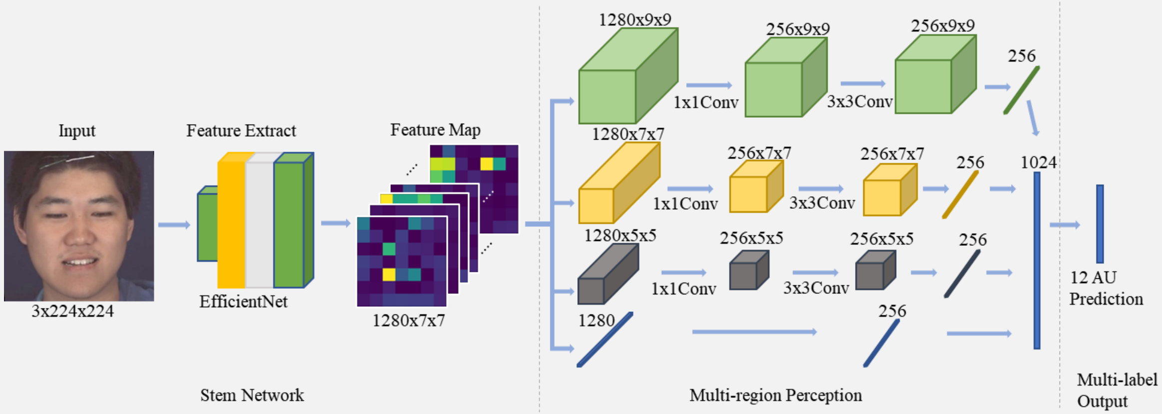
AU1 Inner Brow Raiser



AU10 Upper Lip Raiser



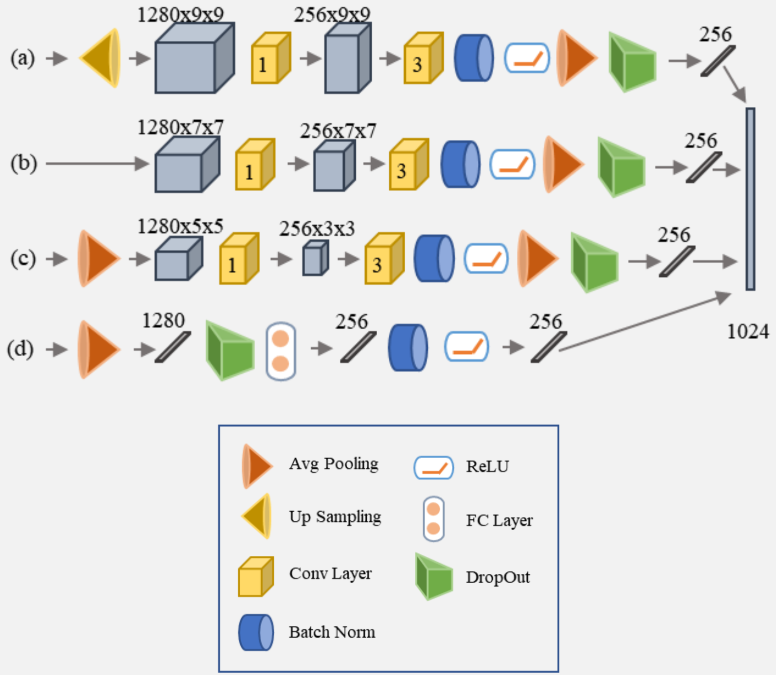
Overview of MRP-Net



We use Efficient-Net as stem network. Only use the features which Efficient-Net outputs and delete the pooling and fully connected layers.

After that, we get 1280 feature maps with a size of 7x7. Then, the features are processed by MRP module to a vector of 1024-dimension to predict AU probabilities.

Multi Region Perception



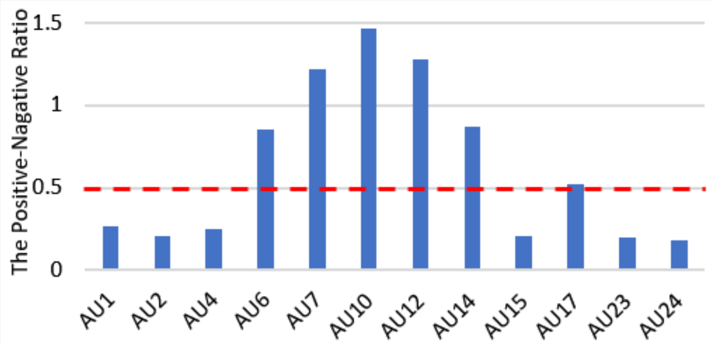
This picture gives the detailed structure of MRP module for multiple region perception.

For the feature maps of stem network, we want to capture the information of different locations and sizes rather than simply reducing the dimension of features.

Therefore, we divide the input feature maps into a, b, c, and d groups for processing. MRP module outputs a vector of 1024-dimension for predicting the probability of AUs.

Batch Balanced Learning

The samples of different AUs in the dataset are unbalanced.



N: Batch size

P_c: the positive-negative ratio for AU_C

$$PB_c = \frac{positive + N}{negative + N} \qquad A_c = \frac{PB_c}{P_c}$$

$$Loss(Y, \hat{Y}_{nc}) = -\frac{1}{N} \sum_{N}^{n=1} \sum_{C}^{c=1} \frac{1}{A_c} \cdot Y \cdot \log \hat{Y}_{nc} + A_c \cdot (1 - Y) \cdot \log(1 - \hat{Y}_{nc}).$$

We proposed Batch Balanced Learning to solve this problem.

Comparison with Other Methods

F1 SCORE ON BP4D DATASET

Method	AU1	AU2	AU4	AU6	AU7	AU10	AU12	AU14	AU15	AU17	AU23	AU24	Avg.
LSVM [32]	23.2	22.8	23.1	27.2	47.1	77.2	63.7	64.3	18.4	33.0	19.4	20.7	35.3
JPML [7]	32.6	25.6	37.4	42.3	50.5	72.2	74.1	65.7	38.1	40.0	30.4	42.3	45.9
DRML [10]	36.4	41.8	43.0	55.0	67.0	66.3	65.8	54.1	33.2	48.0	31.7	30.0	48.3
CPM [19]	43.4	40.7	43.3	59.2	61.3	62.1	68.5	52.5	36.7	54.3	39.5	37.8	50.0
EAC-Net [9]	39.0	35.2	48.6	76.1	72.9	81.9	86.2	58.8	37.5	59.1	35.9	35.8	55.9
ROI [25]	36.2	31.6	43.4	77.1	73.7	**85.0**	87.0	62.6	**[45.7]**	58.0	38.3	37.4	56.4
JAA-Net [33]	47.2	**44.0**	54.9	77.5	74.6	84.0	86.9	61.9	43.6	60.3	42.7	41.9	60.0
DSIN [34]	**51.7**	40.4	**56.0**	76.1	73.5	79.9	85.4	**62.7**	37.3	**[62.9]**	38.8	41.6	58.9
LP-Net [26]	43.4	38.0	54.2	**77.1**	**76.7**	83.8	**87.2**	**[63.3]**	**45.3**	60.5	**48.1**	**[54.2]**	**61.0**
MRP-Net	**[54.0]**	**[47.5]**	**[56.3]**	**[78.5]**	**[76.9]**	**[85.5]**	**[89.3]**	60.9	44.6	**61.5**	**[50.4]**	**48.7**	**[62.8]**

Bracketed and bold numbers indicate best performance; bold numbers indicated second best.

F1 SCORE ON DISFA DATASET

Method	AU1	AU2	AU4	AU6	AU9	AU12	AU25	AU26	Avg.
LSVM [32]	10.8	10.0	21.8	15.7	11.5	70.4	12.0	22.1	21.8
DRML [10]	17.3	17.7	37.4	29.0	10.7	37.7	38.5	20.1	26.7
APL [8]	11.4	12.0	30.1	12.4	10.1	65.9	21.4	26.9	23.8
ROI [25]	41.5	26.4	66.4	**[50.7]**	8.5	89.3	88.9	15.6	48.5
JAA-Net [33]	**43.7**	**[46.2]**	56.0	41.4	44.7	69.6	88.3	**58.4**	56.0
DSIN [34]	42.4	**39.0**	68.4	28.6	46.8	70.8	90.4	42.2	53.6
LP-Net [26]	29.9	24.7	**[72.7]**	46.8	**49.6**	**72.9**	**93.8**	**[65.0]**	**56.9**
MRP-Net	**[49.68]**	29.1	**70.9**	**50.1**	**[54.33]**	**[79.74]**	**[94.15]**	52.0	**[60.0]**

Bracketed and bold numbers indicate best performance; bold numbers indicated second best.

MRP-Net have an average F1 score improvement of 2.95% on BP4D and 5.43% on DISFA.

Ablation Study

F1 SCORE OF ABLATION EXPERIMENTS ON BP4D DATASET

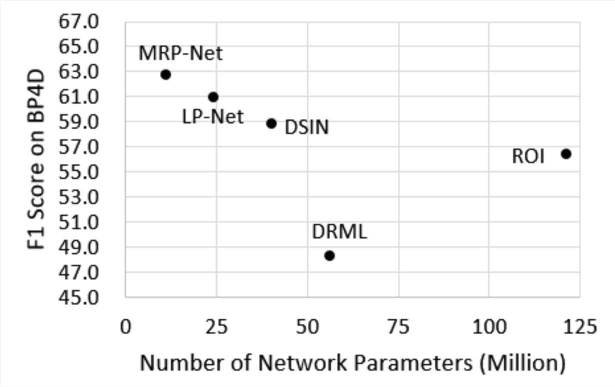
Method	AU1	AU2	AU4	AU6	AU7	AU10	AU12	AU14	AU15	AU17	AU23	AU24	Avg.
EF+CE	44.3	42.8	51.9	71.5	68.6	80.7	80.8	57.9	**[49.4]**	54.9	43.5	45.8	57.7
EF+MRP+CE	43.5	**44.2**	**52.9**	76.8	75.6	84.7	**88.6**	58.9	**46.7**	59.8	44.3	**[50.2]**	60.5
EF+MRP+DCE	**[57.24]**	38.7	52.0	75.9	**[78.2]**	83.3	76.9	**[61.6]**	40.4	56.9	42.7	46.7	59.2
EF+MRP+FL	49.8	41.6	46.8	**[83.4]**	76.1	**85.1**	87.0	59.7	42.8	**[64.6]**	**49.1**	47.9	**61.2**
EF+MRP+BBL	**54.0**	**[47.5]**	**[56.3]**	**78.5**	**76.9**	**[85.5]**	**[89.3]**	**60.9**	44.6	**61.5**	**[50.4]**	**48.7**	**[62.8]**

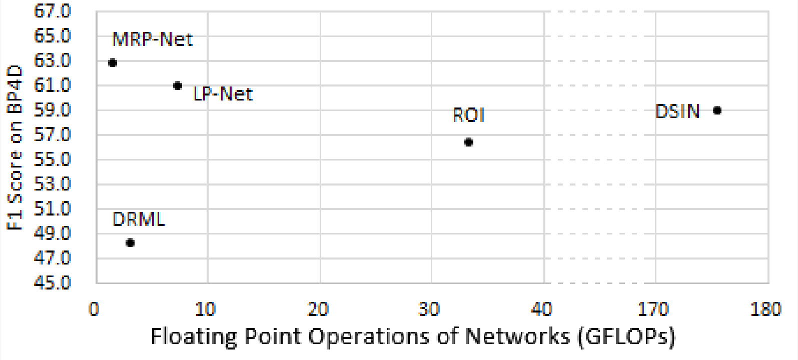
Bracketed and bold numbers indicate best performance; bold numbers indicated second best.
EF:Efficient Net B1; **CE**:Cross Entropy; **MRP**:MRP Module; **FL**:Focal Loss; **BBL**:Batch Balanced Learning.

MRP module have improved F1-57.7 to 60.5.

Batch Balanced Learning have improved F1-61.2 to 62.8 to Focal Loss.

Comparisons of Parameters and FLOPs





Compared with LP-Net, MRP decreases the number of network parameters by 54.62% and the number of network FLOPs by 19.6%.

Conclusion

In this paper, we proposed an end-to-end trainable network MRP-Net, which consists of feature extraction network and multiple region perception module for automatic AU detection without depending on facial landmark information.

We proposed a batch balanced learning method to solve the samples unbalanced problem in multi-label learning.

Our work outperforms the existing methods on two widely used AU datasets, with a significant reduction in the number of network parameters that can be efficiently migrated to portable devices.

MRP-Net: A Light Multiple Region Perception Neural Network for Multi-label AU Detection

Speaker: Yang Tang

Yang Tang[1], Shuang Chen, Honggang Zhang*, Gang Wang, Rui Yang
Pattern Recognition and Intelligent System Lab. (P.R.I.S. Lab.),
Beijing University of Posts and Telecommunications (BUPT)

北京郵電大學
Beijing University of Posts and Telecommunications

p.r.i.s laboratory

ICPR 2020