Action Unit Selective Feature Maps in Deep Networks for Facial Expression Recognition

***Yuqian Zhou[1], Bertram Shi[2]**

[1,2]Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong, China

***E-mail: yzhouas@ust.hk**

Facial expression recognition based on hand-crafted features have achieved satisfactory performance on most databases. However, deep neural network, especially deep convolutional neural network (CNN) with high-level features has been proved to boost performance on vision tasks. In this project, we conduct a transfer learning to establish the existence and utility of feature maps selective to action units [1] in a deep CNN trained by generic images, starting from a network pre-trained on the Image-Net dataset and then transferred to the facial expression recognition task using the Karolinska Directed Emotional Faces (KDEF) [2], Radboud Faces Database(RaFD) [3] and extended Cohn-Kanade (CK+) [4] database. We propose two feature selection schemes. The first one is based upon identifying feature maps that respond to areas within the face by identifying the receptive fields of the maps using deconvolution [5]. And the second one is based upon identifying feature maps that are putative AU detectors by training multiple binary AU detectors on every single feature map via logistic regression, and ranking the maps as AU detectors according to their F1 score. We demonstrate that feature selection is critical in achieving robustness in this task, and that higher convolutional layers of the deep CNN trained on generic images are selective to action units in the Facial Action Coding System (FACS) [1]. We further demonstrate that these action unit selective feature maps are more critical in the facial expression recognition task. This provides further support to the hypothesis that both human and deeply learned CNNs use similar mechanism for recognizing facial expressions.

References

[1] P. Ekman and W. V. Friesen, "Facial action coding system," 1977.

[2] D. Lundqvist, A. Flykt and A. hman, "The Karolinska directed emotional faces (KDEF)," CD ROM from Department of Clinical Neuroscience, Psychology Section, Karolinska Institutet, pp. 91-630, 1998.

[3] O. Langner, R. Dotsch, G. Bijlstra, D. H. Wigboldus, S. T. Hawk and A. van Knippenberg, "Presentation and validation of the Radboud Faces Database," Cognition and Emotion, vol. 24, pp. 1377-1388, 2010.

[4] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The extended cohn-kanade dataset (ck): A complete dataset for action unit and emotion-specified expression," in IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, 2010, pp. 94-101.

[5] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in European Conference on Computer Vision (ECCV), 2014, pp. 818-833.