Real-time Action Unit Intensity Detection

Saurabh Hinduja and Shaun Canavan
Computer Science and Engineering, University of South Florida, Tampa, Florida

***Abstract*— We present a real-time system for action unit *intensity* detection. We train a convolutional neural network, with images from DISFA+, to detect the intensities of 12 action units that are commonly found in the literature. Along with real-time capabilities, the system is also able to detect intensities on static images, and in the wild videos. While the focus of this work is on the detection of action unit intensities, we are able to implicitly detect the occurrence of them as well. This is done by setting all action units with an intensity greater than 0 as active. In doing this, we calculate the F1-micro and F1-binary scores for the DISFA+ dataset.**

I. Introduction to Work

There has been an impressive body of the work in action unit (AU) detection in recent years. Li et al. [4] proposed a deep region-based enhancing and cropping method to detect action units. They found that when adding these layers to a deep network it can learn both of these features showing improvement over state of the art. Girard et al. [2] conducted experiments to learn how much training data is needed to detect action units. Their core findings reveal that more variation in subjects is needed, compared to more frames per subject, to increase performance. Motivated by these works, in this demo, we present a real-time, deep neural network-based, AU intensity detection system.

II. AU Intensity Detection

The overview of our system is shown in Fig 1. Broadly, our system consists of two main stages. First, face detection and alignment are performed on the input frame. Secondly, a deep convolutional neural network is used for AU intensity prediction. The system and the results are described below.

Data. We used DISFA+ [5] to train our intensity detection CNN, which is an extension of the popular spontaneous dataset, DISFA [6], that includes 27 subject ages 18-50 years. We have used all AUs in the dataset for the demo (Table I).

System Architecture. The face detection and alignment is done using the existing dlibs library [3], and the CNN used for intensity prediction is inspired from Ertugrul et al. [1]. The network is a multiple tail network with each output layer giving us a categorical output of the intensity for the AU. The network has 3 convolution layers with kernel size 128, 64 and 32 respectively. Each convolution layer is followed by batch normalization and max pooling layers, and finally a fully connected layer with 400 neurons. The feature vector from the fully connected layer is used separately by 12 fully connected output layers with 6 neurons (for 6 intensity predictions 0-5), the loss function for all 12 output layers is categorical cross-entropy (see Fig. 1).

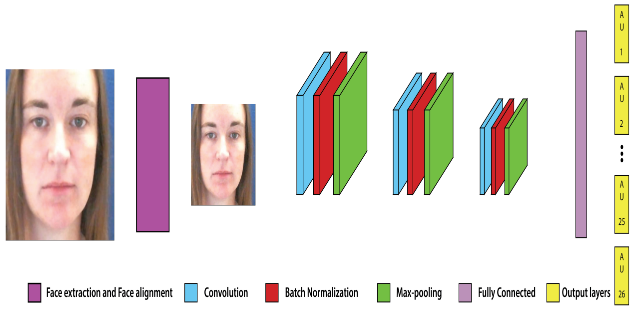

Fig. 1: Overview of the proposed real-time AU intensity detection system. *NOTE: best viewed in color.*

Evaluation. To evaluate the performance of our demo we performed one subject out cross-validation, each fold was trained for 50 epochs, and we report the F1-micro and F1-binary results (Table I).

TABLE I: F1-micro and binary scores from DISFA+ [5]

AU	F1 Micro	F1 Binary
1	0.7946	0.5435
2	0.7883	0.4659
4	0.6520	0.5259
5	0.7927	0.5935
6	0.7809	0.5461
9	0.9229	0.5312
12	0.7158	0.4072
15	0.8179	0.3736
17	0.7624	0.3858
20	0.8777	0.2915
25	0.7273	0.5566
26	0.8559	0.5338
Average	**0.7907**	**0.4795**

Acknowledgment

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

References

[1] I. Ertugrul et al. Cross-domain au detection: Domains, learning approaches, and measures. *FG*, 2019.
[2] J. Girard et al. How much training data for facial action unit detection? In *FG*, 2015.
[3] D. E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.
[4] W. Li et al. Action unit detection with region adaptation, multi-labeling learning and optimal temporal fusing. In *CVPR*, 2017.
[5] M. Mavadati et al. Extended disfa dataset: Investigating posed and spontaneous facial expressions. In *CVPRW*, 2016.
[6] S. M. Mavadati, M. H. Mahoor, K. Bartlett, P. Trinh, and J. F. Cohn. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2):151–160, 2013.