Supplementary Materials for "Harnessing Label Uncertainty to Improve Modeling: An Application to Student Engagement Recognition"

Arkar Min Aung (aaung@wpi.edu) and Jacob Whitehill (jrwhitehill@wpi.edu)

1 Deeper Architectures for Automatic Student Engagement Recognition

In order to enable a tight comparison with prior research on automatic student engagement recognition on the same dataset [1] used in our paper, we employed the Gabor + LogisticRegression architecture described in our paper, which can be implemented as a three-layer neural network. However, in an effort to achieve higher recognition accuracy, we actually conducted extensive computational experimentation using deeper convolutional neural networks. In particular, we tried (using Caffe) many variations (hundreds of hyperparameter choices, extra batch normalization layers, different non-linear activation functions, label-preserving data augmentation, etc.) on the following base architecture for engagement regression:

`Input(48x48) - Conv(5,5,8) - ReLu - MaxPool(2,2) - Conv(3,3,8) - ReLu - MaxPool(2,2) - FC(8) - ReLu - FC(1)`

On a "positive control" dataset (the GENKI4K dataset [2] containing 2000 training faces, 2000 testing faces, from approximately as many unique individuals) for smile/non-smile classification on 48×48 faces, this architecture achieved a high classification accuracy ($> 97\%$ AUC) and significantly outperformed the accuracy attained by a Gabor + LogisticRegression network. Yet for automatic engagement detection, we observed very high variance across the four cross-validation folds, and overall a mean accuracy that was always lower than using the Gabor-based approach.

Our hypothesis for this disappointing performance using a deeper neural network is that, since the number of subjects in HBCU is small (only 20 in total), the variance captured in the 10698 images is not high enough to prevent the network from overfitting, even when the total number of network weights was kept small.

References

[1] J. Whitehill, Z. Serpell, Y.-C. Lin, A. Foster, and J. R. Movellan, "The faces of engagement: Automatic recognition of student engagementfrom facial expressions," *IEEE Transactions on Affective Computing*, vol. 5, no. 1, pp. 86–98, 2014.

[2] `http://mplab.ucsd.edu`, "The MPLab GENKI Database, GENKI-4K Subset."