

Image Emotion Recognition using Cognitive  
Contextual Summarization Framework  
**Supplementary Material**

**Contents**

<b>1</b>	<b>Experiment and Results</b>	<b>2</b>
1.1	Learning Capability of the Models . . . . .	2
1.2	Ablation Study . . . . .	3

# 1 Experiment and Results

We present supplementary plots and diagrams in this section.

## 1.1 Learning Capability of the Models

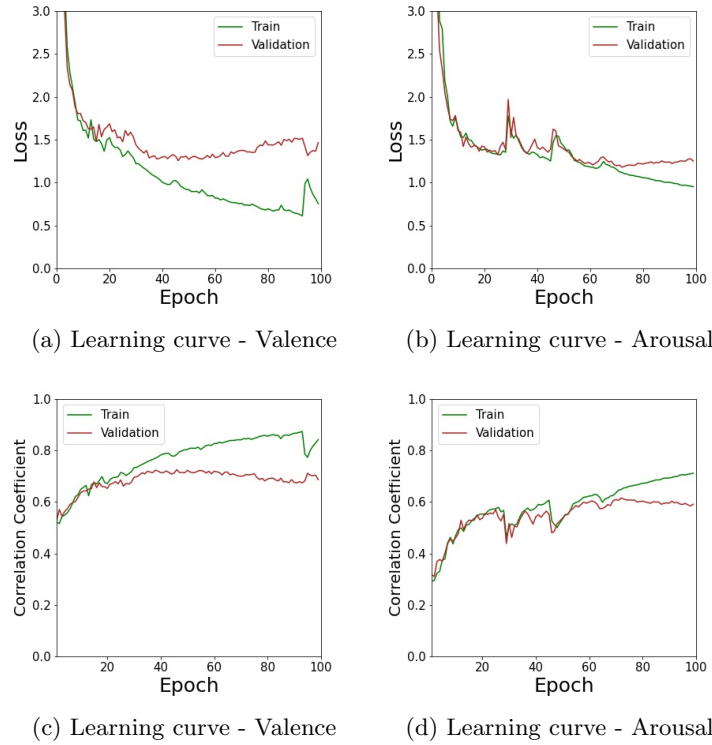
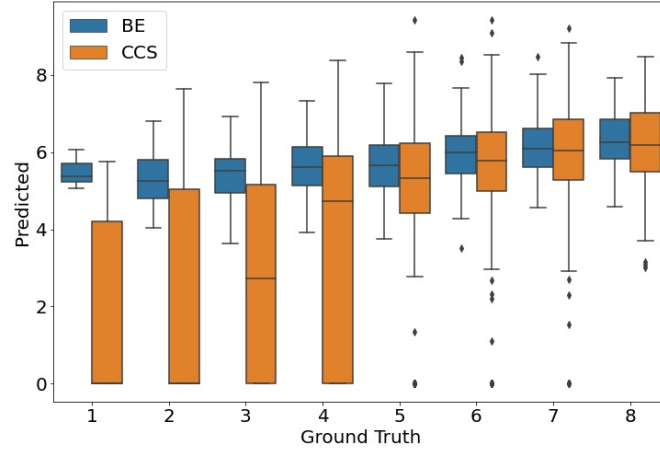


Figure 1: Learning curves for valence and arousal for the CCS model trained on the BE dataset up to 100 epochs. Top row: MSE loss. Bottom row: Correlation coefficient between predicted values and ground truth

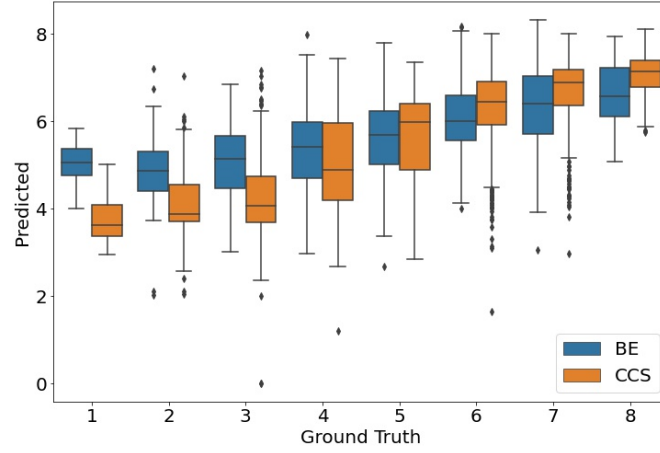
We can see from the Figure 1 that the CCS model which has more number of parameters continues to learn across epochs, correlation between the ground truth and predicted VA increases across epochs.

We can see from the Figure 2 that the CCS model predicts better than the BE model. For instance, in Figure 2b) mean of BE prediction is close to 5 for the ground truth bins 1 - 4, whereas CCS learns the trend well as can be seen in the mean of CCS prediction.

As regards choice of 6 fully connected layers in the CCS model, Table 1



(a) Epoch 1



(b) Epoch 40

Figure 2: Validation (box) plots of Ground Truth and Predicted valence values for CCS and BE models trained on the BE dataset.

## 1.2 Ablation Study

The Table 1 shows the ablation study of the fully connected layers showcasing that 6 FC layers contribute better to emotion prediction.

<b>Model</b>	<b>Valence</b>	<b>Arousal</b>
CCS (3 FC layers)	1.63	1.45
CCS (4 FC layers)	1.49	1.30
CCS (5 FC layers)	1.65	1.44
CCS (6 FC layers)	1.44	1.28
CCS (7 FC layers)	1.84	1.65
CCS (8 FC layers)	1.64	1.22

Table 1: Ablation Study of the CCS model in terms of MSE on the BE dataset.