Empirical Evaluation of Residual CNN in Emotion Recognition

Rongxin Liu

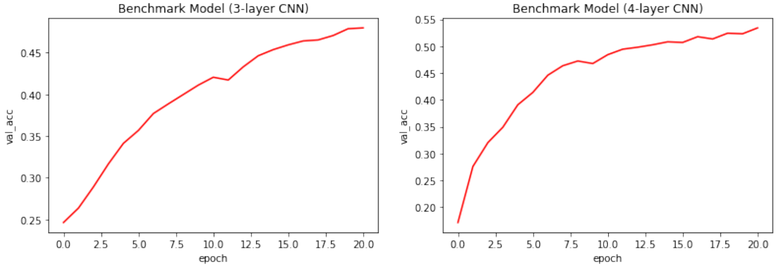
CS231n: Convolutional Neural Networks for Visual Recognition

Overview

Empirical experiment on emotion recognition using residual CNN model. Final tuned model achieved an accuracy of 85%, surpassing the human accuracy on the FER-2013 dataset (65.5%).

Goal: Detect and classify human emotion in real time. Through experiments, we want to find out the optimal CNN model architecture and its corresponding activation functions, optimization functions, and other applicable regularization methods.

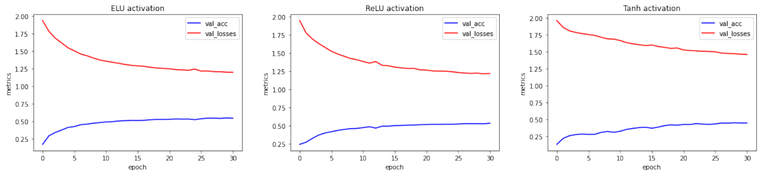
Baseline Model Architecture



A deeper CNN could improve accuracy, and this is illustrated in the above figures that the 4-layer baseline CNN model is superior to the 3-layer baseline CNN model.

Activation Functions

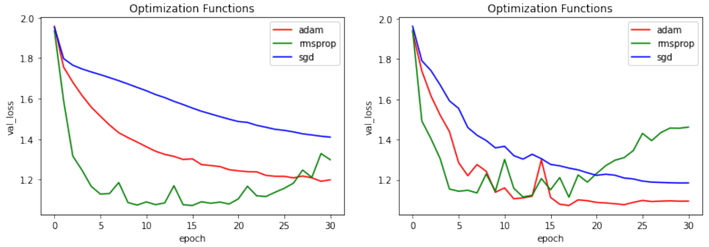
We experimented with ReLU, ELU, and Tanh activation functions for the 4-layer CNN model.



Both ReLU and ELU achieved similar validation accuracy, and the corresponding validation loss also decreased to lower values steadily compared to the model that uses the Tanh activation function. We chose ReLU as our activation function for the CNN model.

Optimization Functions

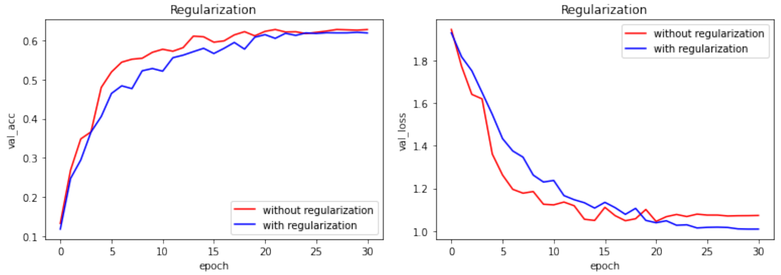
The loss function for our CNN model would be the cross-entropy loss. To find out the optimal optimization function for our CNN model, we experimented with 3 popular optimization functions (Adam, RMSProp, and SGD) and plotted the loss curve for each of them, respectively.



- With higher learning rates, we might be moving too much in the direction opposite to the gradient and could increase the loss. Learning rate scheduling and gradient clipping could help with this issue (right).
- We can observe that RMSProp helps the CNN model converge quickly, but then the loss rate increases gradually even with gradient clipping and learning rate scheduling. Adam optimization seemed to be the best fit for our CNN model from both experiments on optimization functions.

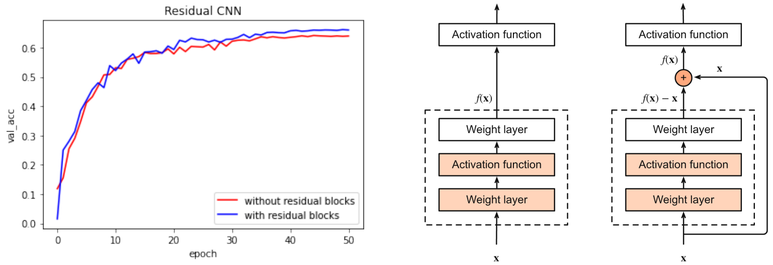
Regularization

After applying dropout and weight-decay regularizations, our CNN model achieved a lower loss while maintaining similar accuracy to the CNN model trained without applying any regularization.



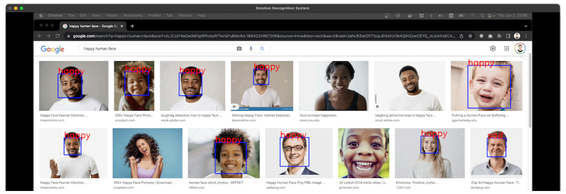
Residual CNN

Lastly, we added residual blocks to our CNN model, and we ran experiments on training a CNN model with or without adding residual blocks and observing their obtained validation accuracy.



We can see that the CNN model with residual blocks achieves a higher validation accuracy than the CNN model without residual blocks.

Realtime Emotion Recognition



We utilized OpenCV to capture video feeds from a video device (OBS virtual display) and convert them to frames. Haar cascades were used to identify and draw a bounding box around faces as the region of interest (ROI). We then fed the ROI to our prediction model and obtained the classified emotion.

Results & Conclusion

- Compared to our baseline CNN model with an accuracy of 47%, the final tuned residual CNN achieved an accuracy of around 85%.
- It also surpassed the human-level accuracy of 65.5% for the FER-2013 dataset.
- The experimented results showed that the CNN approach was appropriate for solving the facial expression recognition problem, and the trained model could be applied in our proposed emotion recognition system.