Human Emotion Classification & Prediction of the OASIS Image Dataset

Protip Roy
protip@stanford.edu | https://youtu.be/ky0vdhsZtCk
Dept. of Computer Science, Stanford University: CS230 Fall 2019

Abstract

Images evoke human emotions, such as fear, anger, joy, sympathy, disgust etc. It might be assumed that such emotions are highly subjective, however studies such as IAPS[3] and OASIS[1] have shown that such responses are largely uniform across a large group of subjects. Project OASIS has published an open-access image dataset labeled by standardized normative emotional ratings. This projects goal is to build a CNN model and train it over the OASIS dataset in order to be able to predict such emotional classifications.

Dataset

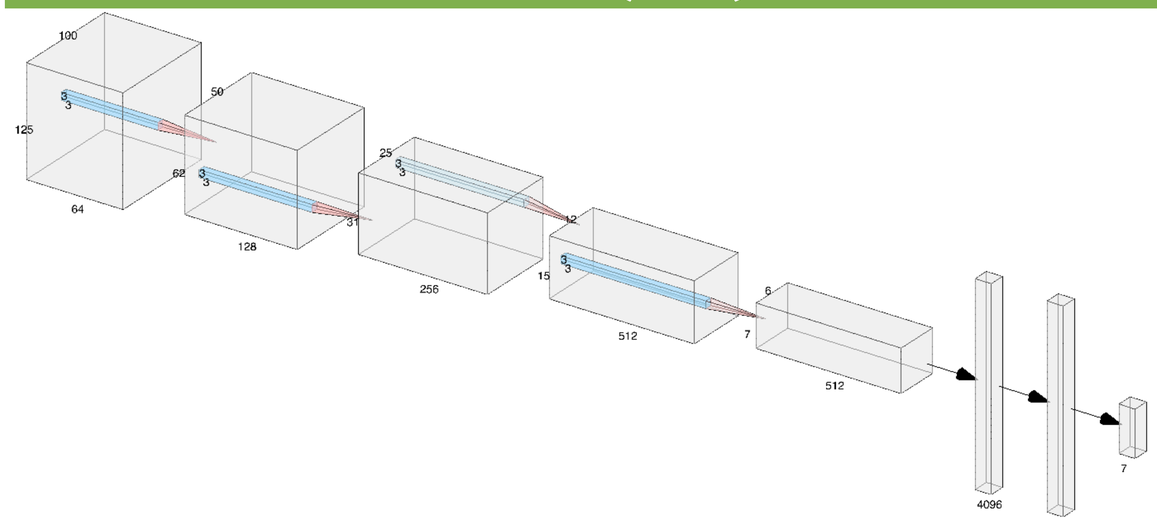
The original dataset comprises of 900 color images of resolution 400Hx500Wx3C and a spreadsheet (CSV)containing the associated *ratings: Valence & Arousal*. Expressed in a 7-point Likert scale - 1 through 7.

Valence High(6.2) to Low(1.3): L to R

Valence - the degree of positive or negative affective response

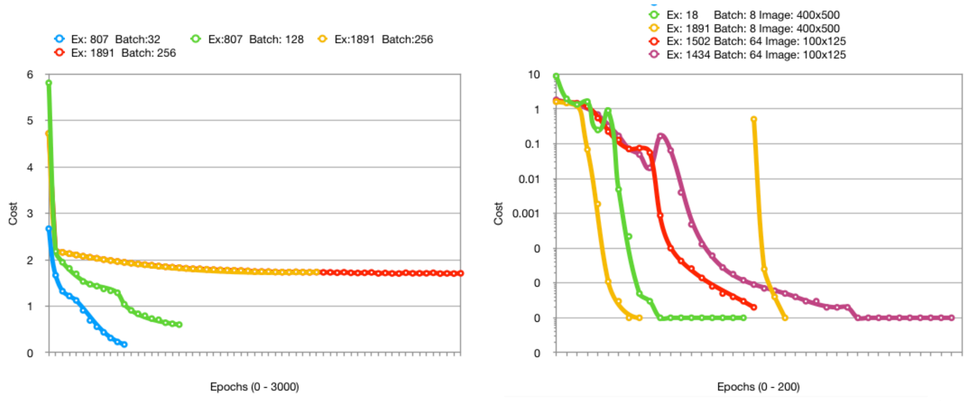
Arousal - the intensity of the affective response.

Model (CNN)



I have built two NN Models:

- Model V1: Started with a simple Tensorflow.v1 model comprising of 3 layers with a 7-class Softmax activation at the output.
- Model V2: A CNN identical to the structure of VGG16 [2]. This model implements the VGG16 structure, i.e. stacked Convolution2D and MaxPool layers followed by 3 FC/Dense layers as shown above (Conv2D sub-layers are not shown). The last FC layer comprises of only 7 units in order to match the degree of classification of our ground truth.



Training Results: Model V1(Std.) vs Model V2 (CNN)

Accuracy

	Image Resolution	Train Set Size	Batch Size	Epochs	Training Accuracy	Dev/Test Accuracy
Model V1	400x500	807	32	600	0.94423795	0.22222222
Model V1	400x500	807	128	1000	0.87608427	0.3
Model V1	400x500	1891	256	2000	0.3395029	0.3270142
Model V1	400x500	1891	256	3000	0.3395029	0.3270142
Model V2	400x500	18	8	100	1.0	0.5
Model V2	400x500	1891	8	110	crash	crash
Model V2	100x125	1502	64	100	1.0	0.4011976
Model V2	100x125	1434	64	200	1.0	0.3625

Discussion

Model V1 flatlines Cost at 1.7 even with > 3000 epochs taking 12+ hours.

Model V2 trains within an hour, however, runs into GPU memory resource limits. Requiring down scaling of images, which in turn leads to compromise of accuracy.

Table above shows that even though Training Accuracy reaches 1.0 with the Convolution model (V2), test accuracy gets saturated at ~0.5. Clearly this model is over-fitting to the training set. Also, dev/test accuracy was compromised when original image resolution was down scaled.

Future Work

- Regularization - L2 and Dropout
- Migrate to PyToch or TF V2

References

[1] Introducing the Open Affective Standardized Image Set (OASIS)
Benedek Kurdi1 & Shayn Lozano1 & Mahzarin R. Banaji1
Published online: 23 February 2016 © Psychonomic Society, Inc. 2016

[2] VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION
Karen Simonyan+ & Andrew Zisserman+
Visual Geometry Group, Department of Engineering Science, University of Oxford {karen,az}@robots.ox.ac.uk

[3] International Affective Pictures System (IAPS)