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.

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.

### 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.

![Image of emotion classification](image-url)

**Abstract**

**Model (CNN)**

**Accuracy**

<table>
<thead>
<tr>
<th>Image Resolution</th>
<th>Train Sel Size</th>
<th>Batch Size</th>
<th>Epochs</th>
<th>Training Accuracy</th>
<th>Dev/Test Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>400x500</td>
<td>867</td>
<td>32</td>
<td>600</td>
<td>0.84433795</td>
<td>0.32352323</td>
</tr>
<tr>
<td>400x500</td>
<td>867</td>
<td>128</td>
<td>1000</td>
<td>0.876947</td>
<td>0.3270142</td>
</tr>
<tr>
<td>400x500</td>
<td>1891</td>
<td>256</td>
<td>2000</td>
<td>0.3395029</td>
<td>0.3270142</td>
</tr>
<tr>
<td>400x500</td>
<td>1891</td>
<td>256</td>
<td>3000</td>
<td>0.3395029</td>
<td>0.3270142</td>
</tr>
<tr>
<td>400x500</td>
<td>18</td>
<td>8</td>
<td>100</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>400x500</td>
<td>1891</td>
<td>8</td>
<td>110</td>
<td>crash</td>
<td>crash</td>
</tr>
<tr>
<td>100x125</td>
<td>1502</td>
<td>64</td>
<td>100</td>
<td>1.0</td>
<td>0.4011976</td>
</tr>
<tr>
<td>100x125</td>
<td>1434</td>
<td>64</td>
<td>200</td>
<td>1.0</td>
<td>0.3625</td>
</tr>
</tbody>
</table>

**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)

2] VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

3] International Affective Pictures System (IAPS)