Univ.AI

Final Project: Real-time facial expression prediction

Arnab Chattopadhyay, Kunal Vartia, Siddharth D P, Neeraj K Udupa

Motivation and About the Project

- The field of Computer Vision is one of the most popular fields in AI today
- One of it's key highlights is that it allows us to automatically extract knowledge from complex raw data structures such as images.
- The goal of this project is building a face expression recognition system with Deep Learning
- We built a custom Deep CNN, as well as experimented with transfer learning.
- We trained the model from scratch using our facial expression dataset
- Finally, we fed real time images to the model to see if it could classify the expressions successfully.

Data and Labels

- We used a dataset from a Kaggle challenge. (Link below)
- The data consists of 48x48 pixel grayscale images of faces.
- Each image corresponds to a facial expression in one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).
- The dataset contains approximately 36,000 images.

References

- https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge
- https://arxiv.org/pdf/1409.1556
- https://docs.opencv.org/master/
- https://keras.io/

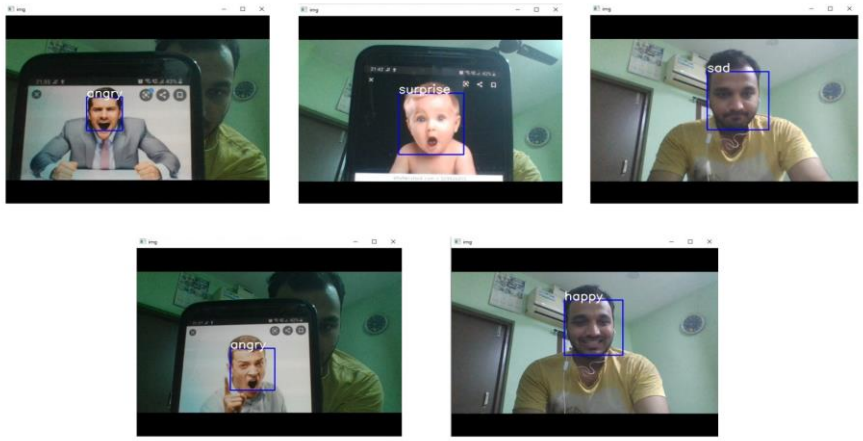
Model

Model: "functional_1"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 48, 48, 3)]	0
block1_conv1 (Conv2D)	(None, 48, 48, 64)	1792
block1_conv2 (Conv2D)	(None, 48, 48, 64)	36928
block1_pool (MaxPooling2D)	(None, 24, 24, 64)	0
block2_conv1 (Conv2D)	(None, 24, 24, 128)	73856
block2_conv2 (Conv2D)	(None, 24, 24, 128)	147584
block2_pool (MaxPooling2D)	(None, 12, 12, 128)	0
block3_conv1 (Conv2D)	(None, 12, 12, 256)	295168
block3_conv2 (Conv2D)	(None, 12, 12, 256)	590080
block3_conv3 (Conv2D)	(None, 12, 12, 256)	590080
block3_pool (MaxPooling2D)	(None, 6, 6, 256)	0
block4_conv1 (Conv2D)	(None, 6, 6, 512)	1180160
block4_conv2 (Conv2D)	(None, 6, 6, 512)	2359808
block4_conv3 (Conv2D)	(None, 6, 6, 512)	2359808
block4_pool (MaxPooling2D)	(None, 3, 3, 512)	0
block5_conv1 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv2 (Conv2D)	(None, 3, 3, 512)	2359808
block5_conv3 (Conv2D)	(None, 3, 3, 512)	2359808
block5_pool (MaxPooling2D)	(None, 1, 1, 512)	0
global_average_pooling2d (Gl	(None, 512)	0
dense (Dense)	(None, 256)	131328
batch_normalization (BatchNo	(None, 256)	1024
dropout (Dropout)	(None, 256)	0
dense_1 (Dense)	(None, 512)	131584
batch_normalization_1 (Batch	(None, 512)	2048
dropout_1 (Dropout)	(None, 512)	0
dense_2 (Dense)	(None, 7)	3591

Total params: 14,984,263
Trainable params: 13,247,239
Non-trainable params: 1,737,024

Results



- The model was successfully able to detect the facial expression of the live-fed images.

Conclusion and Future Work

In conclusion:

- We used transfer learning, using the VGG16 model, and added some of our own custom layers on top to suit our problem.
- There were 467 angry instances in test set. We could classify 214 angry items correctly. On the other hand, the model classified 9 items as disgust but these items are actual angry ones.
- Clearly, accuracy should not express right impression for multi class classification problems.
- To help in interpreting the results better, we have included the confusion matrix of the model in our presentation.

Future Work:

- For this project, we trained all the models from scratch using CNN packages in Keras.
- In the future, we would like to implement a Facial Action Coding System (FACS) for determining facial expressions.
- This posits that expressions result from the change of facial parts and each muscular movement is tracked.
- For this, we could investigate the efficacy of pretrained SOTA such as the GoogLeNet