Adrian Marinovich
Springboard Data Science Career Track
Capstone Project #1

Detection of smiles in images of faces

Final Report Slide Set
December 11, 2018

Mentor: Hobson Lane

The problem

Detect smiles in images of faces.

In other words, build a model that will classify images of faces as either smiling or not smiling.

Why is this interesting?

The smile detector may find eventual implementation in:

- A human-machine interface, in enabling emotional communication tools, which may allow:
    - control of such things as musical instruments via MIDI
    - more safe and reliable human-robot interactions

Potential clients:

- Music, robotics, and physical programming fields

- May be interested in smile detection as step towards emotional classification from facial expression
    - Market research, to better gauge interest in and reaction to products
        - Guide decisions on product design
    - Gaming
        -Support decisions to create more immersive and exciting player experiences

Description of data

LFWcrop dataset consists of 13,233 images (https://conradsanderson.id.au/lfwcrop/), from the Labeled Faces in the Wild (LFW) dataset

- Available as both 3-color and grayscale.
- Faces are centered on the image with the background largely omitted.
    - Resolution of 64x64 pixels

Small smiles dataset (https://data.mendeley.com/datasets/yz4v8tb3tp/5):

- A labelled subset of the cropped version of LFWcrop
- List of face images labelled as smiles: 600 images
- List of face images labelled as non-smiles: consists of 603 images

Large smiles dataset (https://github.com/hromi/SMILEsmileD/tree/master/SMILEs)

- Also a labelled subset of the cropped version of LFWcrop
    -- Additional image reclassification and cleaning performed for this study
- Smiles: 3719 images
- Non-smiles: 9199 images
- Appeared to contain wider variability of smile expressions with greater range of nuance than seen in small dataset, with some ambiguous smiles
- Attempts made to balance the need to exclude smile-like expressions such as smirking, grimacing, and ‘posed’ smiles with the need to include a range of subtle smiles

Smile defined as ‘a facial expression in which the eyes brighten and the corners of the mouth curve slightly upward and which expresses especially amusement, pleasure, [or] approval...’
(https://www.merriam-webster.com/dictionary/smile)

Features

Maximum dimensionality of each image is 12,288 with 3 colors (64x64x3).

Limiting to grayscale images yields reduced dimensionality, D, of approximately 4,096.

- Small dataset: D/N ratio of 6.8 (4,096/600)
- Large dataset: D/N ratio of 1.1 (4,096/3719)

Data wrangling

Data accessed via AWS S3 buckets
- Pipeline from the Jupyter notebook to the S3 buckets using Boto

Diversity of age, sex, ethnicity, head position, facial hair, and presence of eyeglasses in both the smile and non-smile sets

Young children are largely absent from both sets, and were removed where present

Overcropped images not having two eyes and a mouth visible in the image, were excluded

Additional wrangling and preparation for analysis

Small dataset was split into training and validation datasets:
- 1000 images in the training dataset
- 203 images in the validation dataset

Large dataset split into three sets:
- training (80%): 10334 images
- validation (15%): 1938 images
- hold-out (test) (5%): 646 images

Large dataset was randomly sampled for training and validation using a generator function

Data augmentation using image preprocessing with zoom, rotation, and shear

StandardScaler-type standardization using training set mean and standard deviation

Initial findings from exploratory analysis

Exploratory data analysis (EDA) conducted after an initial machine learning approach, random forest:
- Performed to classify the images into smile and non-smile categories
- Before the random forest was performed:
    - Gini vs. entropy impurity criteria compared using a lone decision tree classifier
- Entropy criterion selected for random forest classifier

EDA:
- Colormapping pixel feature importances and reviewing at a contingency table in conjunction with that

- Feature importances colormapped to a plot of pixel location show clustering of high importances in:
    - Mouth and cheek regions
    - Fainter clusters in the forehead and inferolateral cheek regions
    - Very faint signal is also seen in inferior nose/nostril and nasolabial folds regions

Figure 1a. Face image examples (labelled as smiling and non-smiling) from the large dataset.

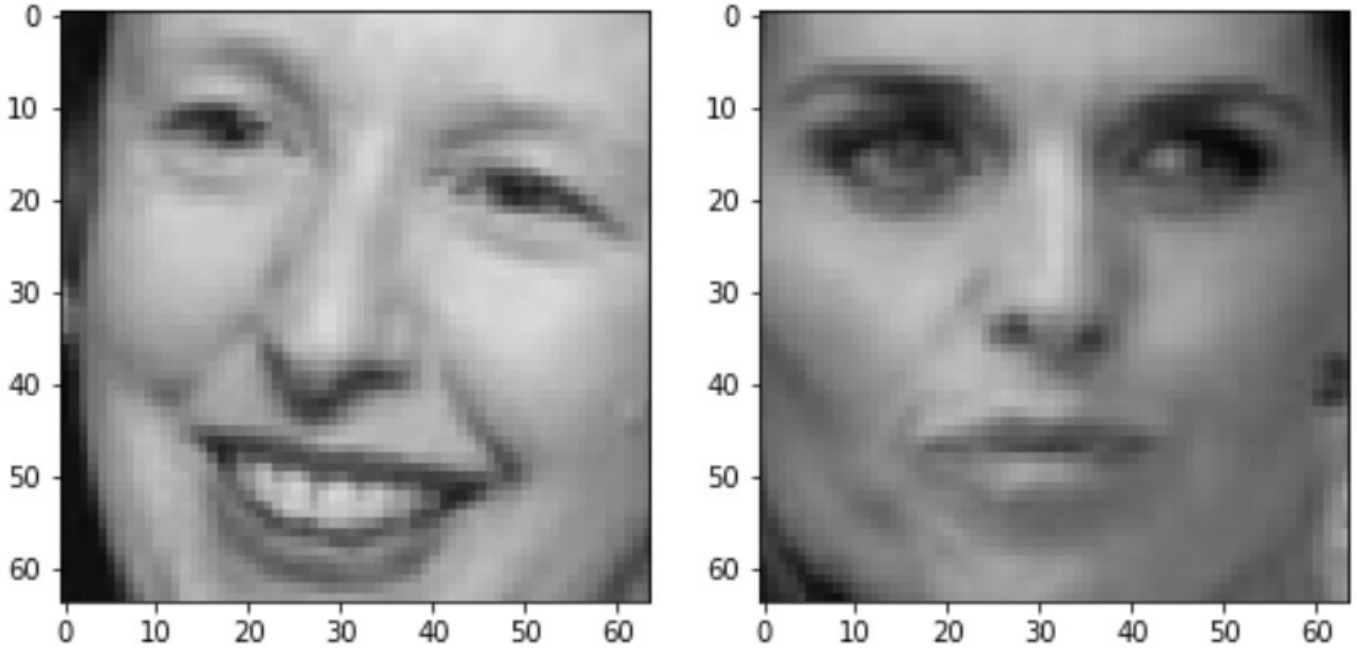
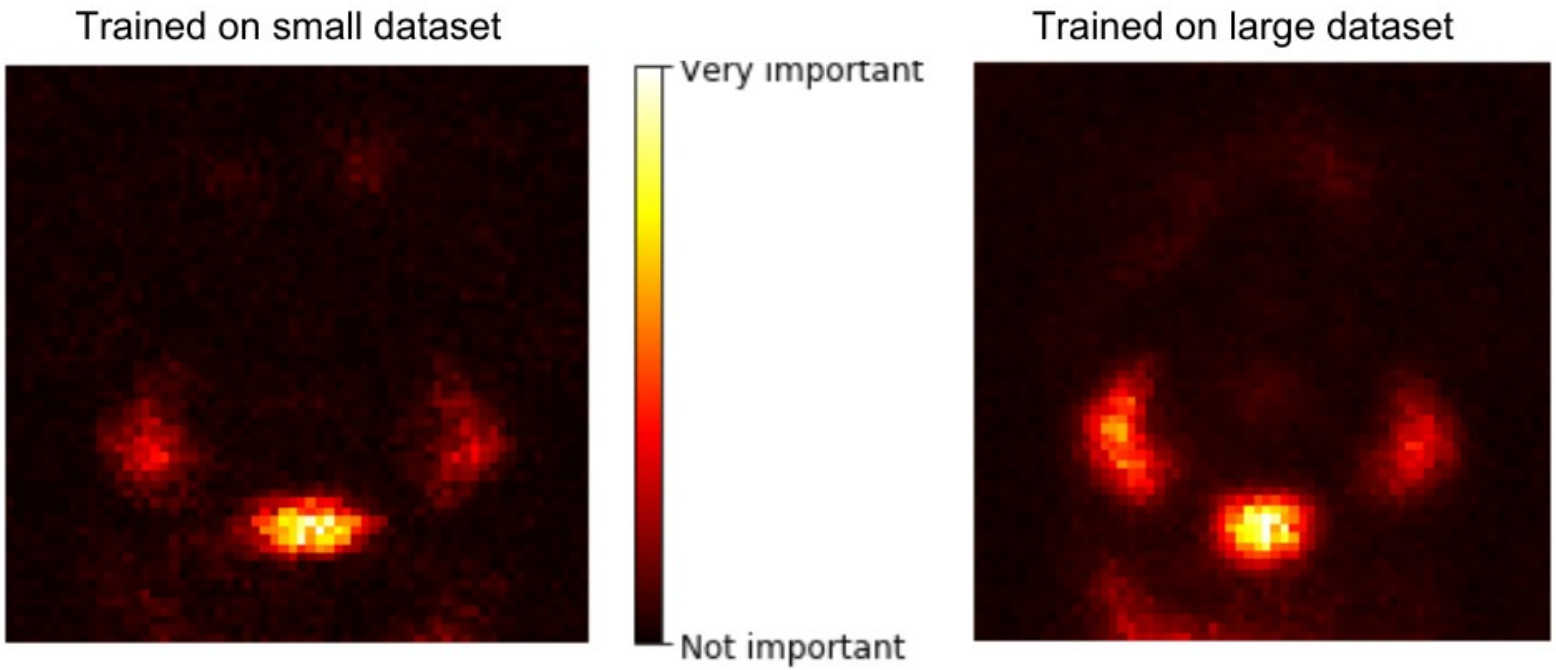


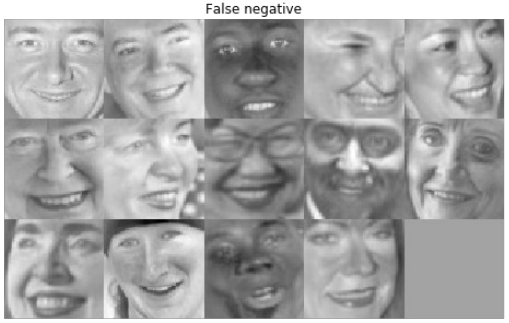
Figure 1b. The mapping of pixel (feature) importance using the random forest smile classifier, entropy criterion, trained on small and large datasets.



Contingency table showing images from the small dataset validation set comparing random forest smile classification against the target labels:

True positive

False negative



False positive

True negative

Initial findings from exploratory analysis

Mapping of pixel importances, coupled with contingency table, suggests some higher-level features might have a role in random forest misclassification:

- False negatives may in part result from face rotation, scaling and centering/cropping differences.

- False positives may in part result from prominent nasolabial folds (as might be seen in smirking or grimacing), presence of facial hair, as well as face rotation, scaling and centering/cropping differences.

Model development

Decision tree

Random forest

Support vector machine model with a radial basis function kernel (SVM-RBF)

Fully-connected deep neural network (DNN)

Convolutional neural network (CNN)

Validation accuracies improved in a progression of models as ordered above

Models produced higher validation accuracies for the small dataset compared to the large dataset, except for the final and best-performing VGG-like CNN model, which performed marginally better on the large dataset at 20 epochs.

This VGG-like CNN model predicted smile in the hold-out (test) set with an accuracy of 0.915

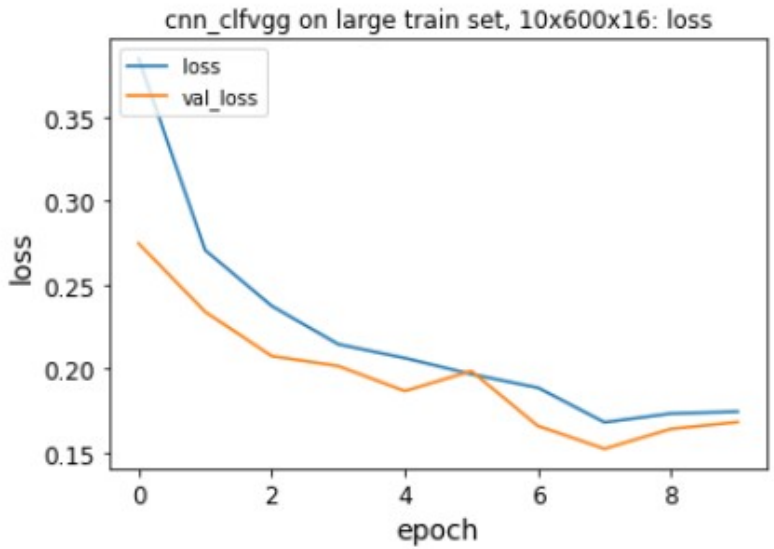
Table 1. Validation set accuracy by model and dataset.

Model	Validation accuracy	
	Small set*	Large set**
Convolutional neural network - VGG-like model Keras layers: cnn_clfvgg = Sequential() cnn_clfvgg.add(Conv2D(32, (3, 3), activation='relu', input_shape=(64, 64, 1))) cnn_clfvgg.add(Conv2D(32, (3, 3), activation='relu')) cnn_clfvgg.add(MaxPooling2D(pool_size=(2, 2))) cnn_clfvgg.add(Dropout(0.25)) cnn_clfvgg.add(Conv2D(64, (3, 3), activation='relu')) cnn_clfvgg.add(Conv2D(64, (3, 3), activation='relu')) cnn_clfvgg.add(MaxPooling2D(pool_size=(2, 2))) cnn_clfvgg.add(Dropout(0.25)) cnn_clfvgg.add(Flatten()) cnn_clfvgg.add(Dense(256, activation='relu')) cnn_clfvgg.add(Dropout(0.5)) cnn_clfvgg.add(Dense(1, activation='sigmoid')) Loss: binary cross-entropy Optimizer: 'adam'	0.921 Epochs: 10 Batch size: 10 Train total: 10,000 Valid: 203	0.926 Epochs: 10 x 600 steps Batch size: 16 Train total: 96,000 Valid: 2,000 0.943 Epochs: 20 x 600 steps Batch size: 16 Train total: 192,000 Valid: 2,000

*Small, or first set: 1000 training and 203 validation images

**Large, or second set: see generator settings by model for training and validation sampling sizes

Figure 2a. Loss and accuracy plots for VGG-like CNN model, first 10 epochs, large dataset.



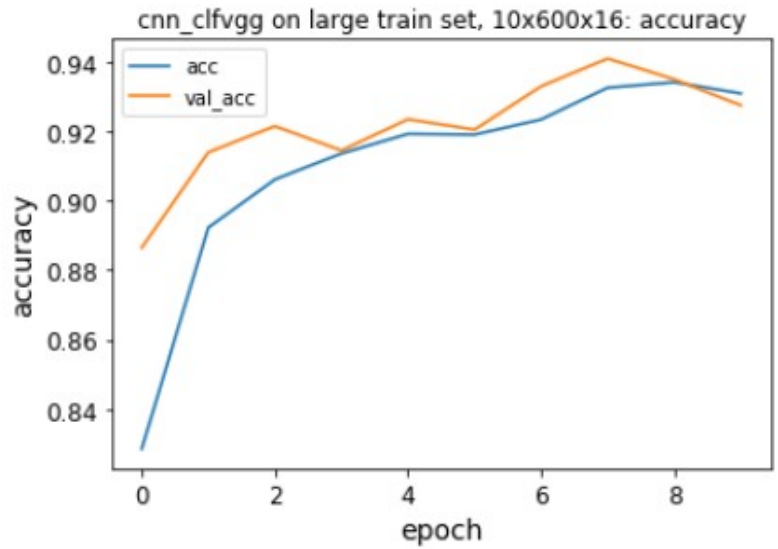
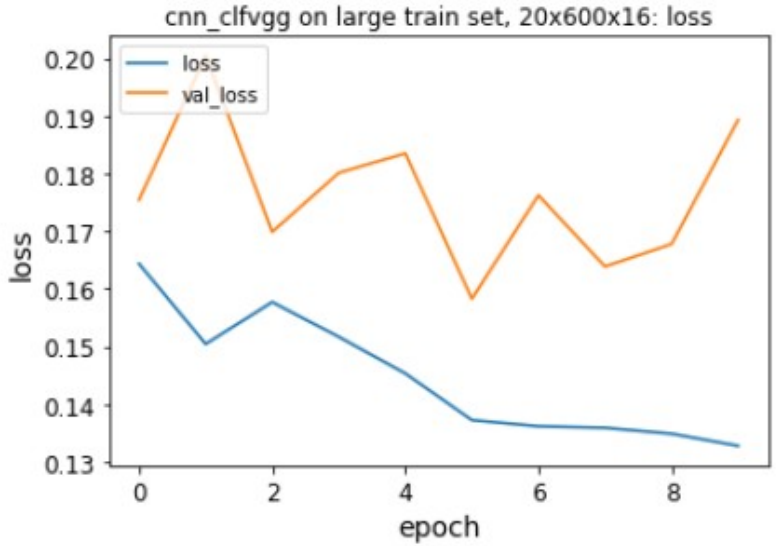


Figure 2b. Loss and accuracy plots for VGG-like CNN model, additional 10 epochs, large dataset. (Note change in scales, compared to Figure 1 above.)



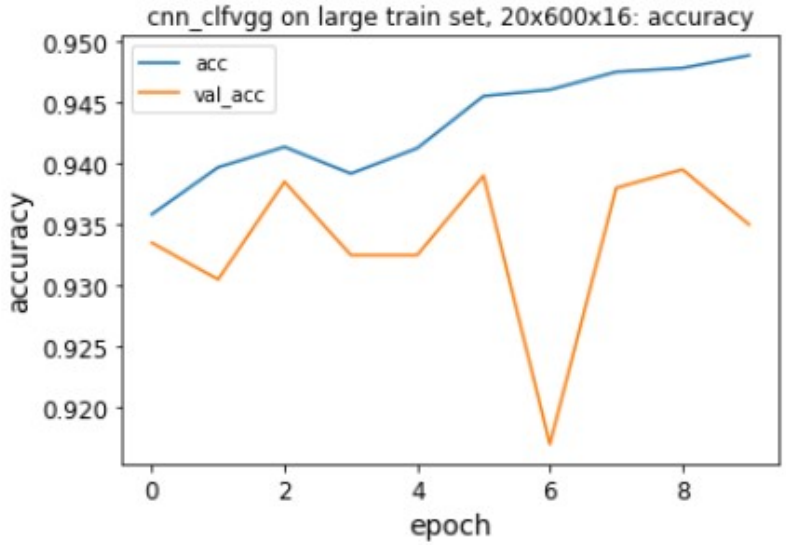


Figure 3. Contingency table showing images from the hold-out set by prediction outcome using the 20-epoch VGG-like CNN model.

Discussion

Role of ambiguity

- Prediction contingency table shows images in the ‘false negative’ and ‘false positive’ boxes that may justifiably fall into an ‘ambiguous’ category by another individual’s labelling, or even be reclassified

Future work may focus on:

- Broader validation and further pruning of the dataset to reduce missclassification or ambiguity...
  ... at cost of generalizability?

- Consider a three-class detector to allow for detection of ambiguous smile-like expressions...
  ... or an ordinal or continuous smile classification structure graded by intensity and clarity of perceived intent?

Conclusion

Parsimonious model shows potential for the development of similar models in the rapid detection of human facial expressions

- Could be implemented in human-machine interfaces.

Superior performance of the VGG-like CNN model indicates the importance of doubled convolutional layers and drop-out layers in CNN models to allow greater generalizability

Project code

Small set code:
https://github.com/adriatic13/springboard/blob/master/dsct_capstone1/Adrian_Marinovich_Cap1_smiles_data_wrangling.ipynb

https://github.com/adriatic13/springboard/blob/master/dsct_capstone1/Adrian_Marinovich_Cap1_smiles_eda.ipynb

Large set code:
https://github.com/adriatic13/springboard/blob/master/dsct_capstone1/
Adrian_Marinovich_Cap1_smiles_data_wrangling_large_set.ipynb

https://github.com/adriatic13/springboard/blob/master/dsct_capstone1/Adrian_Marinovich_Cap1_smiles_indepth.ipynb

References

Arigbabu, Olasimbo Ayodeji, et al. "Smile detection using hybrid face representation." Journal of Ambient Intelligence and Humanized Computing (2016): 1-12.

Huang GB, Mattar M, Berg T, Learned-Miller E (2007) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report.

https://github.com/hromi/SMILEsmileD/tree/master/SMILEs

https://keras.io/getting-started/sequential-model-guide/

https://www.robots.ox.ac.uk/~vgg/

https://www.merriam-webster.com/dictionary/smile