**Applied machine learning system ELEC0132 19/20 report**

*SN: 19132626*

#### Abstract

This section provides a brief overview of the methodology/results presented in the report.[[1]](#footnote-1)

**Index Terms—** One, two, three, four, five

**1. Introduction**

*This section introduces the problem, a brief bird's-eye view of the methodologies you adopted and the organization of this report.*

Face recognition is one of the hottest topics of today’s machine learning deployments and research. Google scholar yielded 500,000 results to the search for “face recognition”. Computer vision and face recognition software is being actively deployed for commercial applications, such as advertising, video surveillance, and user authentication [1] One of many state-of-art commercial examples is Apple’s FaceID, which uses infrared and light scans to identify a user’s face and unlock Phone X [1].

However, the accuracy of the models is still inconsistent in real-life situations, and biases are common in face analysis to determine sex of female and darker-skinned faces [2]; higher rates of false positives have been found as more common for Asian and African American faces relative to images of European / Caucasian faces [3]. These biases are one of several hot discussion topics in today’s public debate and research community.

Recognition of emotions also holds high promise for future applications in the areas of advertising, retail and healthcare, among others. Smiling, for example, is an important factor of social interaction and indicator of mental health.

Until early 2000s, the methodologies for face recognition were based on supervised learning techniques, using image intensity and texture analysis with such classification algorithms as SVM (support vector machines) and AdaBoost. The article “Face recognition using eigenfaces”, published in 1991, was cited 7528 times.

Use of deep learning / convolutional neural networks, while introduced in the end of 80s/early 90s, started being used for image, face and emotion classification in mid-2000s/2010a.

In this paper, several of the above-mentioned models are used for gender and emotion classification, including a pre-trained CNN model to determine the gender of the celeba image set, and a combination of image classifier and well-established classification SVM algorithm.

*A citation example is given here [1].*

**2. Literature survey**

*This section should focus on an overview of potential approaches to solve the tasks. You can introduce some classical and state-of-the-art machine learning algorithms.*

The approaches to detecting or classifying gender attributes have used differences in facial feature dimensions (29 in Hassner’s paper) or “tailored” face descriptors (10). Hassner notes that a detailed survey of gender classification methods can be found in 34 and 42. A brief summary is as follows:

“One of the early methods for gender classification [17]

used a neural network trained on a small set of near-frontal

face images. In [37] the combined 3D structure of the

head (obtained using a laser scanner) and image intensities were used for classifying gender. SVM classifiers

were used by [35], applied directly to image intensities.

Rather than using SVM, [2] used AdaBoost for the same

purpose, here again, applied to image intensities. Finally,

viewpoint-invariant age and gender classification was presented by [49].

More recently, [51] used the Webers Local texture Descriptor [6] for gender recognition, demonstrating nearperfect performance on the FERET benchmark [39].

In [38], intensity, shape and texture features were used with

mutual information, again obtaining near-perfect results on

the FERET benchmark.

… [46] experimented on the popular Labeled Faces in the Wild (LFW) [25] benchmark, primarily used for face recognition. Their method is a combination of LBP features with an AdaBoost classiﬁer.”

CNN was initially used for optical character recognition by Y.LeCun et al in 1989. As the availability of data and compute power improved, CNN began to be used for image classification.

The most common approaches to facial expression recognition include static image texture analysis, feature point-based expression classifiers, 3D face modelling, and dynamic analysis of video sequences.

The classical machine learning approaches to face recognition include the following popular image classifiers:

* Eigenfaces (1991, which uses PCA)
* Fisherfaces (1997, which uses LDA)
* Local Binary Pattern Histograms (LBP), 1996, which uses texture based descriptors. LBP is currently widely used due to its computational simplicity and good performance. The concept of LBP was first introduced as early as 1993 and was made popular following the paper published by Ojala et al [ADD]. LBP computes a local representation of texture, comparing each pixel with its nighbouring pixels and calculating and storing the LBP value in output 2D array.
* Scale Invariant Feature Transform (SIFT) (1999)
* Speed Up Robust Features (SURF) (2006)

Local representation of facial features

State-of-the-art algorithms use deep learning and focus on improving the accuracy of face and object recognition in difficult light conditions, in real-time situations:

* Deep learning facial embeddings
* Whitehill et al focussed on smile detection and optimising performance in real-time real-world imaging conditions. They collected the GENKI database of 36,000 images of real-life faces.

**3. Description of models**

In this section, you should briefly describe the model you are using for each task, along with the rationale. You may opt to use a single learning algorithm to solve the problem or multiple ones, but bear in mind there are page limitations and that you should explain your rationale behind your choices. That is, the algorithmic description must detail your reasons for selecting a particular model.

You can clarify them with flow charts, figures or equations. An example of how to draw an image is demonstrated in Fig. 1.

**3.1. Task A1: Gender detection: male or female**

CelebA dataset was used for gender detection task.

For gender detection task, gender representations were learned through the use of deep-convolutional neural network (CNN). The rational for choosing this model was the interest in testing how existing CNN models would perform when applied to a new dataset, the availability of the model itself in open access, as well as is its compact size (44.5MB) and computational efficiency.

Fig. 1 Gender Detection Model Flow

The paper noted that it tested its model on the Adience benchmark for gender classification of unfiltered face images (10 in Hassner model) and “outperformed existing state of the art by substantial margins.” However, the authors also noted that there was further room for improvement, and the trained model was provided for public access.

**3.2. Task A2: Emotion detection: smiling or not smiling**

CelebA dataset of 5,000 images was used for emotion detection task.

To detect smiles, we trained the model on 4,000 random images to recognize smiling, and predicted the labels on the remaining 1,000 images. Prior to training the model, we used OpenCV to detect faces using Haar cascade image representations; we recognized faces using Fisherface and BLP face recognizers for the prediction.

Whitehill et al [1] published

**3.3. Task B1: Face Shape recognition: 5 types of face shapes**

Hello world!

###### 3.4. Task B2: Eye color recognition: 5 types of eye colors

Hello world!



Fig. 1 A nice view of Roberts Building

**4. Implementation**

This section must provide the detailed implementation of your models. In particular, you must provide the name and use of external libraries, explain hyper-parameter selection, training pipeline (if any) and key modules/classes/functions/algorithms.

You also must provide a detailed description of the dataset (content, size, format, etc.), any data pre-processing that was applied and how you separate your dataset into training, validation and test sets.

The execution of your models also should be reported here. In particular, this section should include a thorough discussion on the training convergence and stopping criterion (it is recommended that learning curves graphs be used to this effect).

**4.1. Task A1: Gender detection: male or female**

*4.1.1. Description of the model*

The open access Caffe model for gender classification was provided by Tal Hassner in his git depository [ADD]. The method was implemented using the Caffee open-source framework.

The code used for the model was adapted from the code provided in N.Singh Chauhan post for TowardsDataScience web site [ADD].

The images were processed using Haar cascade frontalface classifier in OpenCV library. OpenCV dnn package was used to deploy the model. The dnn package uses a class called Net, used to populate a neural network.

The model uses 3 convolutional and 3 fully connected layers. “The output of the last fully connected layer is fed to a soft-max layer that assigns a probability for each class. The prediction itself is made by taking the class with the maximal probability for the given test image.” Data augmentation is used by image crop to 227x227 pixels, and the training is performed using stochastic gradient descent with image batch size of 50 images.

**4.1. Task A2: Emotion detection: smiling or not smiling**

The emotion detection implementation used Box Filter image representations, namely Haar features that use the Viola-Jones algorithm, such as Haar face, eyes and smile cascades to detect a face and a smile. Two different face recognizers – FisherFace and LBP – were used to recognize faces. In Whitehill paper [4], the use of LBP was shown to improve the performance, compared with the use of BF alone.

A model was then trained on 4,000 images using existing labels, and “smiling” label predicted on 1,000 images.

The code for the detection was adapted from the Python Smile Detection Using OpenCV [1] [ADD] and How to Build a Smile Detector [ [2]ADD] articles.

Code for LBP – Model training was adapted from Paul van Gent’s article [ [6]]

**4.1. Task B1: Face shape recognition: 5 types of face shapes**

Hello world!

**4.1. Task B2: Eye color recognition: 5 types of eye colors**

Hello world!

**5. Experimental Results and Analysis**

This section describes and discusses your results. Additionally, this section should include accuracy prediction scores on a separate test dataset, provided by the module organizers, but not used during your training and validation process.

We recommend you use a table to list the tasks, models and results before analysis.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task | Model | Face recognition Acc | Train Acc | Val Acc | Test Acc |
| A1 | Caffe model |  |  |  | 52.62% |
| A2 | Trained model | 95.3% |  |  | 52% |
| B1 |  |  |  |  |  |
| B2 |  |  |  |  |  |

**5.1. Gender Detection Analysis**

The first run of the gender detection pre-trained CNN model yielded very poor accuracy of 52.62% - barely over a random guess, and well below the results achieved in the source paper (85.9% +/-1.4 using single image crop).

A selection of misclassified images is provided below.

Fig. 1 Gender misclassifications

**5.1. Emotion Detection Analysis**

The accuracy of the emotion recognition model (52%) was calculated based on an average of 10 runs. Of these, 4 runs were performed using FisherFace recognizer, with an average accuracy of x%; and 6 runs using LBP recognizer with an average accuracy of 55.63%.

The accuracy of the predictions was affected by the accuracy of face recognition – this was 95.3% on average.

These results were lower than results reported in Whitehill et al [4], where a combination of image representation – classifier reported where 96.3% +\_0.27 was reported for smile detection.

**6. Conclusion**

This last section summarizes the findings and suggests directions for future improvements.

**12. References**

List and number all bibliographical references at the end of the paper. The references can be numbered in alphabetic order or in order of appearance in the document. When referring to them in the text, type the corresponding reference number in square brackets as shown at the end of this sentence [2]. An additional final page (the fifth page, in most cases) is allowed, but must contain only references to the prior literature.

[1] A.B. Smith, C.D. Jones, and E.F. Roberts, “Article Title,” *Journal*, Publisher, Location, pp. 1-10, Date.

1. The code is provided link-to-download-your-project.com and GitHub project: link-to-your-github-project [↑](#footnote-ref-1)