COMP1804
Applied Machine Learning

Name: Arun Narayanaswamy
ID: 001122220
Module Leader: Dr. Dimitrios Kollias
Module instructor: Dr. Mahtab Hossain

School of Computing and Mathematical Sciences

UNIVERSITY of GREENWICH

Contents

List of Figures:

Introduction:

We encounter the applications of Facial Recognition in out day to day activities. Our cell phone uses Face ID to unlock, cameras give a small square highlighting the faces of human, used in social media etc. I like googles facial recognition where all the similar faces from our google photos will be collected and asks for their identity, after providing the identity the same will be updated in our contacts.

This is a fascinating technology where the identification of faces is done by facial attributes like colour, shape, wears glasses or not, hair type etc. "As a typical classification problem, face attribute prediction has been addressed using deep learning. Current state-of-the-art performance was achieved by using two cascaded Convolutional Neural Networks (CNNs)." (Chen, Jenkins 2017)

In this module I would be attempting to train a deep neural network based on facial attributes like wrinkles, freckles, glasses, hair colour and hair top. This would be a stepping stone in understanding the how the above network works in a real world. In this network I would be feeding the neurons with images which will be annotated and try to predict the output by passing random images.

Related Work:

Facial recognition plays an important role in security. Previously we had Keys, pins and ID etc. now all these has been replaced by using digital patterns, OTP and facial recognition. Facial recognition is considered to be much safer because it focus on detecting features like eyes, nose, ears and replicating that would not be possible. (Chen, Jenkins 2017)

For this scenario of facial recognition lets let's consider images of 40 individuals with each of them provides 10 different poses. These images would be different from one another considering the lightings, facial expressions, facial details. The image sizes are said to be 92*112 pixels. (Chen, Jenkins 2017)

In facial recognition PCA is used to find principal components of a given set of images and represent each face image as a smaller size in lower dimensional face.

The initial size of the image I has size m×n. The training of the set of N faces can be written as I= (I1, I2, IN).

The average of image generated is

A=1N∑Ni=1Ii

the vector Yi=Ii−A, is computed, which is the difference image of each face image.

C=1N∑Ni=1YiYT

After getting the result there are 3 machine learning classification executed. (Chen, Jenkins 2017)

1. linear discriminate analysis(LDA)
2. support vector machine (SVM)
3. k nearest neighbors (KNN)

LDA is an image optimiser it converts original data into much lower dimension. LDA separates the class and reduce dimension space. (Chen, Jenkins 2017)

"The optimal dimension reducing transformation GT∈Rl×m(l<m) for LDA is the one that maximizes the between-class covariance matrix S_B and minimizes the total within-class covariance matrix Sw" (Chen, Jenkins 2017)

Each image class I (1 ≤i≤r) has n_i elements, and the total number of data is n=∑ri=1ni. SB is the between-class covariance matrix given by.

SB=∑γi=1ni(ci−c)(ci−c)T

K-nearest neighbour is a classification method where each class is treated to be a cluster of points and a centroid needs to be defined. Centroid will treat the data points closest to it as one cluster. KNN is an unsupervised classification, KNN needs to find the centre of each cluster by itself. (Chen, Jenkins 2017)

Support vector machine (SVM) is a learning model to classify and analyse the data. The given training sample is treated as one class, it represents the examples as points in space and maps.

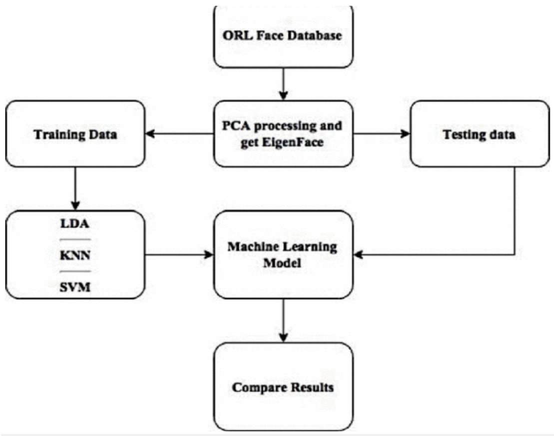

Figure 1 illustration of facial recognition two-stage processing.

The data sets will be divided into training and testing samples form ORL face database. To minimize the calculation PCA would reduce the image size into 5×5,20×20 or 80. The influence of the size of the eigenfaces would images as 5*5 will be divided into training and testing data. After repeating the process 10 times the mean and standard deviation is obtained by LDA,KNN,SVM.

		LDA		KNN		SVM	
		mean	SD	mean	SD	mean	SD
Training data size	3	63.82%	1.54%	69.64%	4.51%	71.04%	3.04%
	5	66.50%	2.04%	77.60%	2.09%	80.25%	2.06%
	7	95.33%	2.19%	93.58%	2.15%	95.75%	1.90%

Figure 2 Recognition result from different size of training data

Form the table we can say that using larger size images for training provides better accuracy. LDA has lower recognition accuracy, KNN has a medium level of recognition accuracy and high standard deviation, SVM has the highest recognition accuracy and best standard deviation. From this we can prove that KNN is the best option to carry out image recognition. (Chen, Jenkins 2017)

The execution time is also an important factor in considering the face recognition algorithm. "Different running times of LDA, KNN and SVM are obtained from different sizes of eigenfaces and different sizes of training data". The process is repeated 10 times to process the average time. (Chen, Jenkins 2017)

Average running time(s)		LDA			KNN			SVM		
Training Group Size		3	5	7	3	5	7	3	5	7
Eigen-Face Size	5x5	1.14	0.97	1.07	0.92	0.89	0.84	3.59	3.51	3.42
	20x20	1.08	1.01	1.10	0.76	0.85	0.86	4.07	4.07	4.02
	80x80	1.20	1.22	1.17	0.83	0.91	0.79	4.59	4.97	5.03

Figure 3 Average running time

Form the above table it is clear that when the eigenfaces in 20*20 the execution time taken is more however the \average running time does not follow the proportion of the size of training data.

"Among these classifications, KNN has the fastest running speed and SVM has the lowest running speed. LDA is a little slower than KNN but much faster than SVM". (Chen, Jenkins 2017) category of thick hair. If the images had bald spots, only then it was under the category 0.

Dataset Preparation

It is challenging to know how to prepare image data when training a convolutional neural network. The initial images received was 1999 from professor and an additional 2000 images i requested to provide so that more images would be available for training the model. Better the image recognition takes place.

There are approximately 4000 images that needs to be annotated. The following images were converted into videos .mp4 using the given code in Lab 8. The converted the images had few repetitive inserted during the conversion. The mp4 video was later converted into .mov to upload the file into the given annotation tool.

The .mov file was loaded into annotation which had 7996 images which needed to be annotated based on the facial attributes like the image has Wrinkles or not, Freckles or not, wears normal glasses or sunglasses, what is the hair colour of the image, hair top like bald or less hair or thick hair.

Annotations was quit challenging and there were few assumptions made, suppose the image was unclear the hair colour black or brown considered it to be black for most of the images, if grey and white considered to be white. Continuing the assumption, if the images had little hair and medium quantity considered it as a category as few hair, if the hair was up to eye level then those images fell on the

After the annotations was completed a .txt file was generated from the annotations tool. The .txt file was uploaded to convert the file into .csv format. The operations involved in removing the line with N/A and all the binary values were organised linking to the respective annotations.

This file is called as InputFile.csv

The .csv files has 4000 images that needs to extracted to start with pre-processing the data.

ML Method using ResNet50

Deep learning has been successfully applied in diverse application domains including computer vision, image classification, speech recognition, natural language processing, etc.
(Hussain Mujtaba 2020)

Deep neural network is divided into 3 types for different applications

1. Feed-forward neural networks (FCNs)
2. Convolutional neural networks (CNNs)
3. Recurrent neural networks (RNNs)

For this course work I will be using convolution neural network(CNN). CNN is the most popular neural network model being used for image classification problem.

There are different models in CNN like LeNet, AlexNet, VGG, GoogLeNet, ResNet etc.

I would be using ResNet50 for (Freckles, Hair colour) and VGG16 for (Wrinkles, Glasses, Hair top, Not human) for facial recognition.

ResNet50: (Hussain Mujtaba 2020)

ResNet is abbreviated as residual network model. ResNet follows a unique procedure called connection called "Skip connect" which means that it allow gradients to flow easily from layer to layer. This helps to train really deep networks. The skip connect solves the vanishing gradient in deep neural network.

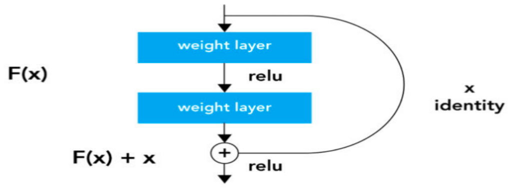

Figure 4 Skip connect in ResNet

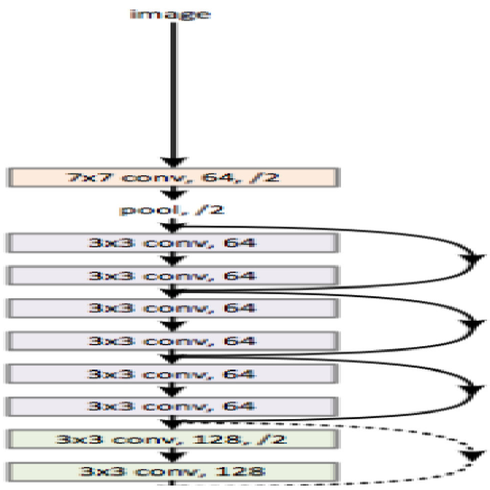

Figure 5 ResNet Architecture

Let's consider the first block where the input image is passed to a convolution layer from there to batch normalisation and then to activation function relu. Again the image is passed to next convolution layer and batch normalisation, before sending the output to activation function the output and the previous input will be added. Every dotted line indicates the input needs to be changed. The network chooses the layers that needs to be processed and carry out the iterations. By doing so the network will not forget the previously learnt features. The deeper the layer better the accuracy obtained. (Hussain Mujtaba 2020)

The models trained using ResNet has provide a better prediction on the image recognition, before using ResNet I had tried with VGG for (freckle and hair colour) but the network could not identify most of the images.

ResNet has provided better prediction than compared to VGG.

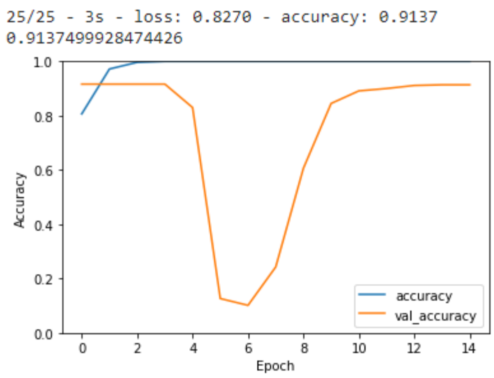

Figure 7ResNet graph for freckle

From the above graph we can say that ResNet has better accuracy when compared with VGG and it has provided good prediction with minimum loss. The model was able to predict random images having freckles from random testing samples.

When comparing both the graphs we can say that if there are variations in accuracy and validation accuracy the model would be trained and can predict specific features.

Evaluation of ResNet model

After normalising the images. The images are split into training and validation datasets. training data consisted of 80% of images and validation of 20%.

<u>Freckles:</u>

After training the VGG model on Freckles the error rate was less and accuracy was good, however the model could not recognise the freckle feature and showed the output of no freckles for all the images. I tried implementing data augmentation like zooming, cropping, re-splitting datasets. There were no changes in the model, still could not predict the images.

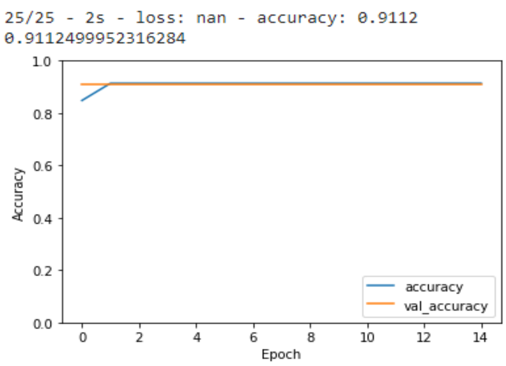

Figure 6 VGG16 graph for freckles

The above graph clearly indicates over fitting where the accuracy and validation accuracy are very high. When the same model is applied in validating the images the model was not able to recognise the features and did not predict proper output.

<u>Hair colour:</u>

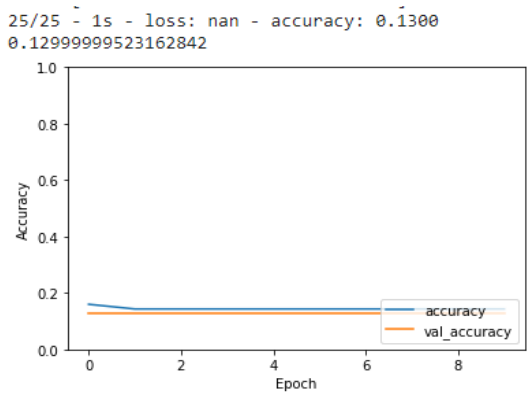

Figure 8 VGG16 graph for Hair colour

The above graph clearly indicates under fitting where the accuracy and validation accuracy are very low. After the model was trained I tried implementing the mode on validating data. The model could not recognise the colour of the hair and gave wrong prediction. This is the reason to drop VGG model and opt ResNet as VGG under fits the data. In ResNet, the model gave a decent prediction on the images and was able to recognise hair colour, few predicated output was incorrect but most of it was right.

Hence considered using ResNet for Freckles and hair colour.

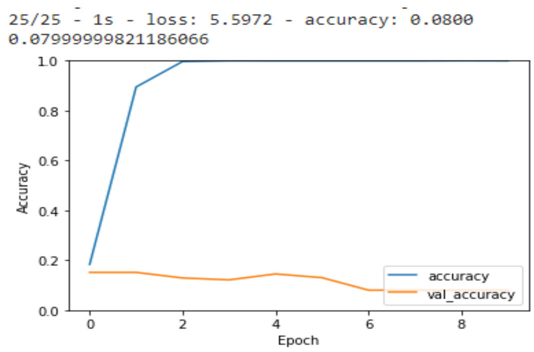

Figure 9 ResNet graph for hair colour

From the above graph we can conclude that ResNet has better accuracy and has provided good prediction with minimum loss. The model was able to predict (black, brown, blond) images from validation and could predict the hair colour when random images was passed.

ML Method using VGG16

"Face recognition is generally used for identification and verification. There are several features that can be used for recognising a person VGG16 is widely used for this task."
(Perdana, Prahara 2019)

VGG16 is one of the most prefered CNN architecture in the recent years. This has 16 convolution layers and has an uniform architecture. (Perdana, Prahara 2019)

The architecture consists of two time (2 convolution layers and pooling layers) followed by three times (3 convolution layers and 3 pooling layers) finally connected to fully connected dense layers.

Each time when an image passes through a convolution layer the size reduces from 224-112-56-28-14-7.
(Perdana, Prahara 2019)

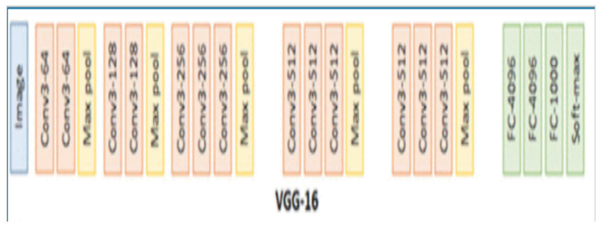

Figure 10 VGG layers

After when the data is normalised. The images are split into training and testing data. In training the VGG model the input image will be resized and the model will be trained. A sample data set would be kept aside for testing and checking the prediction that how well the model can read the features from the image and predict the output.

I am using VGG16 to train my model for facial attributes like Wrinkles, Glasses and hair top.

Evaluation of VGG model

Wrinkle:

The below graph shows overfitting where the accuracy and validation are very high. After the model was trained on wrinkles, the model was validated on test images the result was not good and the model was unable to predict wrinkles in the images. I tried with different model but comparatively VGG16 was slightly better.

The total images with wrinkles was 250 where I felt it was not sufficient to train the model which the result was shown in validation and testing random images.

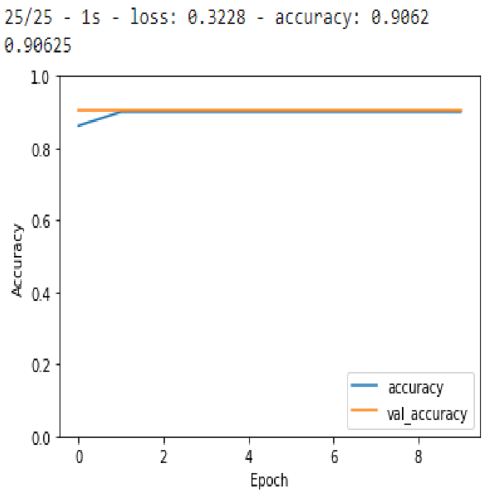

Figure 11 VGG16 graph for Wrinkles

Glasses

For glasses recognition I have used VGG16 where the training data contains 80% of images and validation with 20%.

The validation data was successful in identifying the images with glasses but when a random images was passed to the model it failed to recognise the feature.

I tried with data augmentation and tried even then the prediction did not increase. I tried different models as well but even then very few images got recognised.

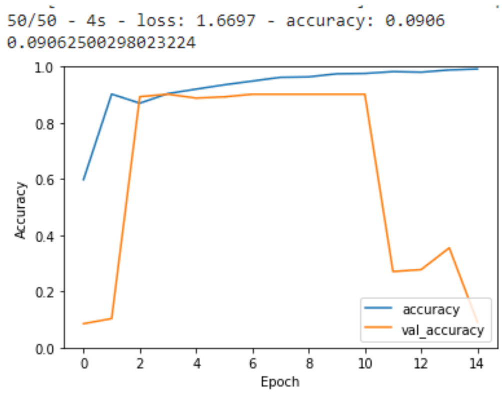

Figure 12 VGG16 graph for Glasses

From the graph the model accuracy and validation accuracy has a significant correlation, however the same prediction was not shown when random images were passed to the model

Hair Top

For Hair top recognition I have used VGG16 where the training data contains 80% of images and validation with 20%.

The validation data was successful in identifying the images with few hair, thick hair and bald. The validation images was a successfully identified.

When the random images were passed the model was partially successful in identifying thick hair and bald features.

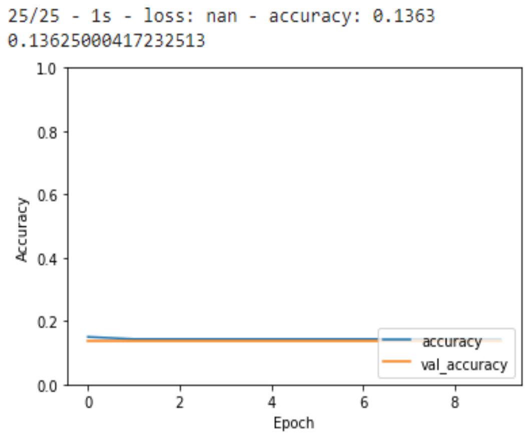

Figure 13 VGG 16 graph for hair top

From the above graph it is clear that the model is under fitting the image points. However, the predictions made by the model is good and can be accepted.

Future work

The obtained models in all the facial attributes was satisfactory. But I feel that there needs to be one model to pass all the images at once to check the facial attributes and improve upon the model in finding out the colour of eye ball, skin tone, happy/sad/angry.

Before creating separate models for all the attributes I tried creating one model to do prediction but it was not that successful as the implantation was lot harder understanding the error was a little tricky.

For my future work I would like to design a facial recognition which could read the live streaming data from HDFS and generate predictions about the user. This would be a game changer in solving anti money laundry where fraud users can be stopped instantly.

I have selected facial recognition as my academic project as well. I will continue to use these models and generate a better model for the project requirement.

References:

1. Chen,J., Jenkins,WK.,(2017)Title: "*Facial recognition with PCA and machine learning methods*" Available URL: https://ieeexplore.ieee.org/abstract/document/8053088, Accessed on: 14/04/2021

2. Mujtaba,H (2020) Title: "*Introduction to ResNet or Residual Network*" Available url: https://www.mygreatlearning.com/blog/resnet/#sh11, accessed on : 12/04/2020

3. Perdana,AB., Prahara,A., Title: "*Face Recognition Using Light-Convolutional Neural Networks Based On Modified Vgg16 Model*", Available URL: https://ieeexplore.ieee.org/abstract/document/9111481, Accessed on: 13/04/2021

4. All the lab codes were referred from the Labs 7,8.

5. The prediction and testing codes was worked in group including myself and vp1311k@greenwich.ac.uk (0011409014)

Thank you Note:

I sincerely thank prof. Dr. Dimitrios Kollias and Dr. Mahtab Hossain for teaching me this course on Applied Machine Learning and for provide feedbacks to improve my work.

I enjoyed the interaction provided by lectures especially Mentimeter.