Age, Gender and Ethnicity Recognition

Bowen Wang
Faculty of Science
University of Western Ontario
London, Canada
bwang489@uwo.ca

Zixuan Wang
Faculty of Science
University of Western Ontario
London, Canada
zwang863@uwo.ca

Didulani Salwathura Acharige
Faculty of Engineering
University of Western Ontario
London, Canada
dsalwath@uwo.ca

***Abstract*—Face Analysis has become the trend for different applications motivated by security and commercial uses nowadays. There are three features that are vital along with the image data which are age, gender and ethnicity. We propose three approaches, Convolutional Neural Network (CNN), Deep Neural Network (DNN), a combination of Principal component analysis (PCA) and DNN to process image data and aim to find the best performing network model and the architecture among age, gender and ethnicity while exploring and understanding the black box architecture of neural networks. Moreover, we build a generative network model to generate new facial images for specific ages of a person. Among different models the CNN model reaches 88% accuracy for gender prediction and 76% accuracy for ethnicity prediction. As for age regression, CNN also has a better performance among all the methods, it has the lowest mean squared error and closest prediction on the age of the facial image.**

***Index Terms*—CNN, DNN, PCA, conditional generative adversarial network**

I. Introduction

Significant progress has been made in recent years in the field of image processing, particularly in facial recognition [1]. With the development of machine learning technologies, most applications are inspired by solving problems related to security, safety and commercial considerations. For example, identification and tracking of criminals, search for lost persons, and analysis of customer behavior, etc [1]. Moreover, age, gender and ethnicity are three vital and useful features to be analyzed in image processing.

Different machine learning and deep learning algorithms have been used in applications depending on the type of the problem. Among them Convolutional Neural Network (CNN), Deep Neural Network (DNN), a Principal component analysis (PCA), and a conditional generative adversarial network (CGAN) are four machine learning and deep learning algorithms that are used to process image data in this project. In general CNNs have been promising to use for image data while the PCA approach is promising for dimensional reduction requirement by extracting most significant data. DNNs have been used for different types of applications and CGAN is a generative network having promising capability to generate image data.

Therefore, this project aims to 1) find out the best performing networks for three different parameters and compare the performances of networks, 2) explore through black box architecture by visualizing filters, 3) implement a generative network to create a facial image of a specific person corresponding to a specific age.

II. Related Work

A. Face Analysis

Automated face analysis is well-developed in face attribute classification and face recognition fields. Many companies including Google, IBM, Microsoft and Face++ have developed and released commercial software on automated face analysis applications [1].

Some previous work in the field of gender classification include extracting local binary patterns (LBP) histograms and performing a classification by support vector machines (SVMs) [2].

Tejas et al. [3] study the performance of discriminant functions including PCA, Linear Discriminant Analysis (LDA) and Subclass Discriminant Analysis (SDA) over a heterogeneous data-set of 8112 images that includes variations in illumination, expression, minor pose and ethnicity. The results conclude that PCA performs slightly better than PCA+LDA, PCA+SDA and PCA+SVM.

Samarasena Buchala et al. [4] experimented with the same target facial image properties (age, gender and ethnicity) as in this project. They used PCA to encode these properties and obtained a nice performance. They also implemented PCA and CCA methods on each section of the face. The PCA method results in an 87.5% accuracy compared to other methods.

B. Face Aging

Previous works related to face aging are mainly constructed by physical model approaches and prototype approaches [5]. Suo et al. [6] experimented physical model approaches by learning the parameters of the dynamic model from the facial image data set and explicitly modelling the stochasticity of face aging in the dynamics. Kemelmacher-Shlizerman et al. [7] propose the prototype approach by computing the average image sub spaces that are pixel-to-pixel aligned, modelling variable lighting, capturing the differences in shape and texture between each age prototype and lastly transferring the texture differences to the test set.

III. Data

The data-set is from a challenge hosted by Kaggle [8], and that includes 23,704 of facial images along with the age, gender and ethnicity information recorded. Fig. 1 demonstrates the ethnicity distribution in our data-set. There are five categories in total, the shape of the distribution implies imbalance as the first category has about ten thousand data entries which is more than one third of the whole data-set. Since the imbalance data usually refers to a challenge, the model could be more biased towards the highest frequent category reducing the generalization capability of the model.

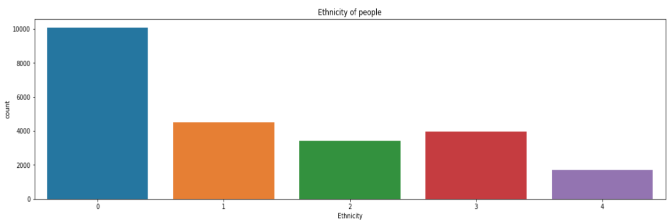

Fig. 1. Ethnicity Distribution

However, the Fig. 2 shows the data collected are fairly distributed between male and female.

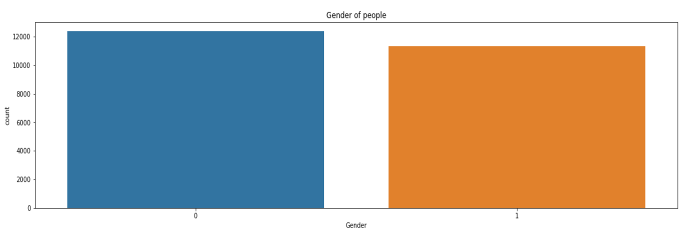

Fig. 2. Gender Distribution

Fig. 3 is the distribution of age attribute of all facial images. The shape indicates another imbalanced distribution of people's age. A peak occurs around the age of 20 to 30, and moreover, it has a long right tail.

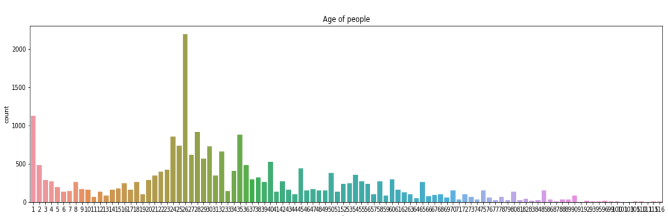

Fig. 3. Age Distribution

IV. Methods

A. Research Objectives

The problem to solve in this project is classified under three main objectives. The first research objective is to recognize people's ethnicity, age, and gender from a facial image. With the help of CNN and DNN along with the PCA approach, we were expected to train a classifier that can classify images into five ethnicity categories: a classifier that classifies images to two gender categories; a regressor that estimates the age of the person in the image. Three separate deep networks were expected to be implemented to find out the most effective feature extracting network model and architecture for the image data.

The black box architecture of deep learning models is a common challenge among the AI community. Therefore, as the second objective we were interested in exploring that by visualizing kernels of CNN model and visualizing most effective principal components of PCA model to understand the networks more.

The third research objective was to generate a facial image for a specific age of a person, as it has interesting applications and, implementing CGANs for regression labels is lacking in the literature to the best of our knowledge at this point. Therefore, we were interested in implementing a generative model to generate real like facial images.

B. Research Methodology

1) Preprocessing: The image data is saved as a list of gray scale pixels in the data-set and before building the networks, the data needs to be preprocessed for better results. First, the pixels were obtained from the data-set and reshape them into a 2D array so that it would be easier for the following process. Then for the standardization the gray scale matrix was divided by 255 so that the figure can remain in range [0, 1].

As the next step the data-set was split into training, testing and validation data-set with proportions of 80, 10 and 10. The training data-set is used to do model training while the testing set and validation set are used in model testing and evaluation respectively.

2) Age, gender, ethnicity classifier: Recognizing age, gender and ethnicity out of a facial image has been an essential area of investigation for different applications. Therefore, under the first objective we implemented and compared the performances of three separate models for three parameters: age, gender and ethnicity. A convolutional neural network, deep neural network and the combination of principal component analysis along with the deep neural network models were implemented for three separate tasks. The idea of using convolutional neural networks was, from the previous literature we found that convolutional neural networks perform better and promising with image data as the convolutional approach can extract the important features out from an image.

Principal component analysis is an approach which is capable of reducing the dimensionality by extracting most important features from the data. Therefore, we were interested in using the extracted features from the principal component analysis approach as the input data to the deep network and compare the performances with convolutional neural network performance. In order to identify the true capability behind the principal component analysis we further implemented a deep neural network as the third network by feeding the row image data as the input without changing the architecture we used in the combined network (PCA and deep network) except the number of input nodes.

To implement the principal component analysis a 2D array images (48 $\times$ 48 pixels) were converted to a 2304 of 1D array pixels. After implementing PCA the significance of data was plotted with respect to the principal components which is shown in Fig. 4. The plot describes that 500 principal components include more than 90% of data in the images and to visualize it, the images were re-constructed with 500 principal components (See Fig. 5). In Fig. 5 The left image

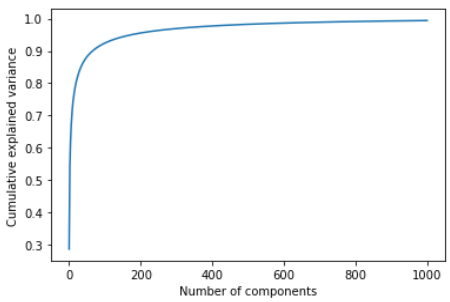

Fig. 4. Cumulative explained variance vs. Principal components

represents the original image while the right image represents the re-constructed image for facial data. The resolutions of the reconstructed images with 500 principal components are lower compared to the original images but do not show a significant difference in other factors.

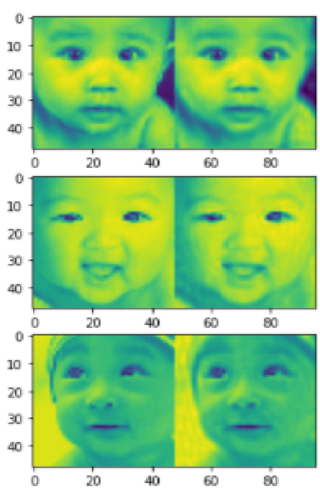

Fig. 5. original image vs. re-constructed image

a) Age model: The first approach was to implement a convolutional neural network as a regression network for the age parameter and the final architecture after tuning the hyper parameters is given in the table . The deep network of the age model has 2304 input nodes which is the size of the 1D array of images and the combined network (PCA and DNN) has 500 input nodes. The final architecture of DNN and combined approach of PCA and DNN are also given in Table I below. For a fair comparison all the hyper-parameters and both network architectures kept similar except the number of input nodes of the deep network and the combined approach network. The mean squared error (equation (1)) between the real and the predicted age values from the network was used as the loss function for all three networks.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (1)$$

TABLE I
NETWORK ARCHITECTURES FOR AGE MODEL

Network	CNN	DNN	PCA+DNN
Input size	48 $\times$ 48	2304	500
# of convolutional layers	3	-	-
# of nodes in the hidden dense layers	1024,1024, 512,256, 128,1	1024,1024, 512,256, 128,1	1024,1024, 512,256, 128,1
# of channels in the convolutional layers	32,64,64	-	-
Type of the pooling layers	Maxpooling	-	-
Pool size in each pooling layer	(2,2)	-	-
Sizes of strides in each convolutional layer	(3,3)	-	-
Activation function	ReLU Linear	Elu Linear	Elu Linear
# of dropout layers and dropout percentage	6 (0.3,0.3,0.3,0.5, 0.5,0.4,0.4)	4 (0.25,0.25,0.25, 0.25,0.25)	4 (0.25,0.25,0.25, 0.25,0.25)
Loss function	Mean squared error	Mean squared error	Mean squared error
Optimizer	Adam	Adam	Adam

b) Gender model: Gender model is a classification model which has two categories to classify images. Therefore, all the models have two output neurons with softmax activation in the last layer. The fine tuned model architectures are given in Table II. Same as the age model, three separate networks were implemented as classification networks. The binary cross entropy (equation (2)) was used as the loss function for all three networks.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i log(p(y_i)) + (1 - y_i) log(1 - p(y_i)) \quad (2)$$

TABLE II
NETWORK ARCHITECTURES FOR GENDER MODEL

Network	CNN	DNN	PCA+DNN
Input size	48 × 48	2304	500
# of convolutional layers	4	-	-
# of nodes in the hidden dense layers	128,2	1024,1024, 512,256, 128,2	1024,1024, 512,256, 128,2
# of channels in the convolutional layers	16,16, 32,64	-	-
Type of the pooling layers	Maxpooling	-	-
Pool size in each pooling layer	(2,2)	-	-
Sizes of strides in each convolutional layer	(3,3)	-	-
Activation function	ReLU Linear	Elu Linear	Elu Linear
# of dropout layers and dropout percentage	4 (0.25,0.25, 0.25,0.25)	4 (0.5,0.25, 0.4,0.5)	4 (0.5,0.25, 0.4,0.5)
Loss function	Binary Crossentropy	Binary Crossentropy	Binary Crossentropy
Optimizer	Adam	Adam	Adam

TABLE III
NETWORK ARCHITECTURES FOR GENDER MODEL

Network	CNN	DNN	PCA+DNN
Input size	48 × 48	2304	500
# of convolutional layers	4	-	-
# of nodes in the hidden dense layers	128,5	1024,1024, 512,256, 128,5	1024,1024, 512,256, 128,5
# of channels in the convolutional layers	16,16, 32,64	-	-
Type of the pooling layers	Maxpooling	-	-
Pool size in each pooling layer	(2,2)	-	-
Sizes of strides in each convolutional layer	(3,3)	-	-
Activation function	Elu Softmax	ReLU Softmax	ReLU Softmax
# of dropout layers and dropout percentage	4 (0.4,0.4, 0.25,0.25)	4 (0.5,0.25, 0.4,0.5)	4 (0.5,0.25, 0.4,0.5)
Loss function	Categorical Cross Entropy	Categorical Cross Entropy	Categorical Cross Entropy
Optimizer	Adam	Adam	Adam

c) Ethnicity model: Same as the gender model the ethnicity model is a classification model which has five categories. Therefore, all three networks were implemented with five output neurons with softmax activation for the last layer. The network architectures of three networks are shown in the Table III. The categorical cross entropy function has been used as the loss function in ethnicity models.

3) Filter Visualization: To address the second objective the idea was to explore through the black box architecture of neural network models. For that purpose the filters of convolutional neural network models were visualized in each convolutional layer to understand how each layer learns the important features of the images to predict the output.

To understand the principal component analysis, the most important principal components were visualized (see Fig. 24) as it helps to understand the most significant features of the images. Since we feed the extracted features to the deep network, visualizing principal components help to understand the input data of the networks from the combined approach more.

4) Generative model: Generative Adversarial Networks (GANs), are trained in an adversarial manner and they are capable of generating new images out of noise vectors. But there is no way to control the types of images generated by the GAN network. Therefore, a new generative network called Conditional Generative Adversarial Network (CGAN) was introduced and it was a type of GAN that involved the conditional generation of images by a generator model. It can be conditional on a class label allowing the targeted generation of images of a given type.

Therefore, the third goal of the project was addressed by implementing a conditional generative adversarial network as a generative model. CGAN network has two separate networks: a discriminator and the generator which train in an adversarial manner. The generator generates real looking images out of a noise vector with the intention of fooling the discriminator which tries to discriminate between the real and generated images.

Fig. 6 below illustrates the conditional generative adversarial network model. In the CGAN the fine tuned generative network architecture consists of 2 upsampling layers, 3 convolutional layers with 128, 64, 1 channels in each layer, and two inputs: the label value along with 100×1 noise vector. The input size of the discriminator was $48\times48\times2$, which has two channels: the input image and the label value. Further, it has 4 convolutional layers having 32, 64, 128, 256 channels in each layer with (2, 2) strides in each. The final layer of the discriminator network is a dense layer with 1 neuron. Binary cross entropy (equation (2)) is the loss function for both discriminator and generator networks while using adam optimizer. The activation function for all layers for both networks were LeakyReLu except the last layer. The activation functions of the last layer for the generator and discriminator are relu and sigmoid activations respectively.

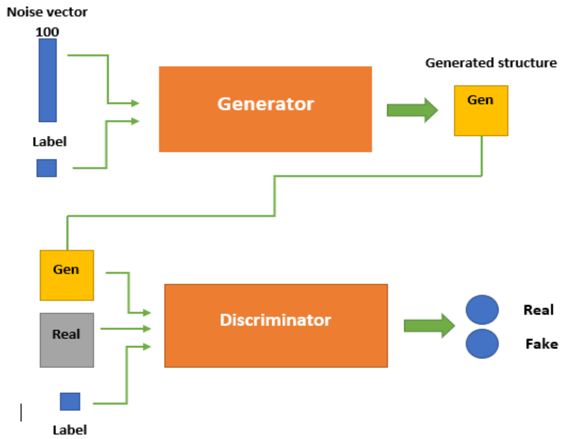

Fig. 6. Conditional Generative Adversarial Network (CGAN) architecture

V. Results

The generalization capability of the trained models were evaluated based on the test data. The test samples were pre-processed before feeding to models with the similar procedure used for training and validation data.

A. Age, Gender and Ethnicity model

1) Age model:

a) Convolutional Neural Network: Fig. 7 shows the training and validation loss variation with respect to epochs of CNN for age model.

- The loss decreases gradually as the epoch increases on both training and validation sets.
- There is no overfitting present in the training process of the CNN network.

Fig. 8 is a representation of the variation between real and the predicted age values for the testing set.

- The variation between real and predicted values shows a positive correlation.
- But the prediction capability of age values greater than 80 is considerably low.

From that it can be concluded that less number of data samples in the dataset for people who are greater than 80 years old might be the reason for this.

Fig. 7. Loss variation of the Age model (CNN)

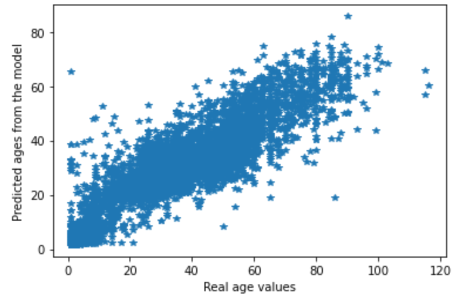

Fig. 8. Predicted age value vs real age values for the test set (CNN)

b) Deep Neural network: Fig. 9 represents the mean squared error variation of the training and validation set of the deep neural network over each epoch for 100 epochs. Fig. 10 is a representation of the predicted age values with real age values for the test sample.

- Fig. 10 implies that deep neural network is not doing a good job at predicting the age values for the facial images in the test set.
- Comparatively very high loss values are showing for some epochs while training this network

This model is mainly implemented to represent the potential behind PCA in extracting most important features from image data which is represented in the next section.

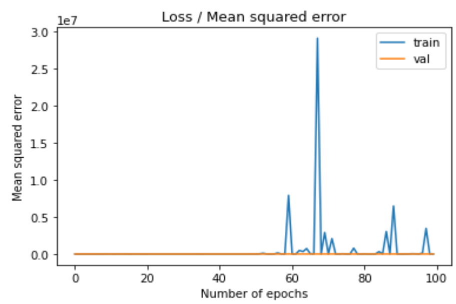

Fig. 9. Loss variation of the Age model (DNN)

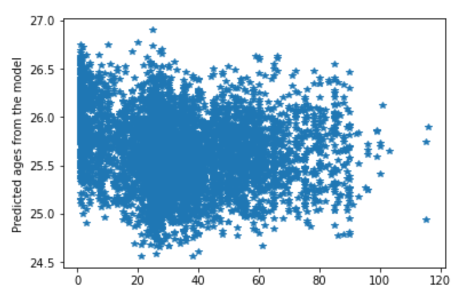

Fig. 10. Predicted age value vs real age values for the test set (DNN)

c) Deep Neural Network along with PCA approach: To test this network testing dataset was obtained after performing PCA to the original test set.

- The validation loss of the trained network (see Fig. 11) is lower than the validation loss obtained from the DNN model and higher than the validation loss obtained from the CNN model.
- Comparatively very high loss values are showing for some epochs while training this network.

Fig. 12 is a representation of the predicted age values compared to the real age values. When compared with the plot obtained from DNN (Fig. 10), it can conclude that extracting most significant features from PCA and feeding only important data to the DNN works better than feeding all features of the images to the DNN. The reason for that is that the PCA approach has the potential to increase the interpretability of data.

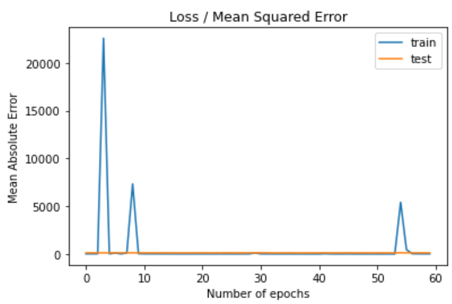

Fig. 11. Loss variation of the Age model (PCA+DNN)

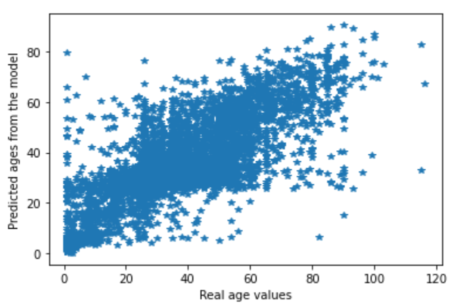

Fig. 12. Predicted age value vs real age values for the test set (PCA+DNN)

2) Gender model:

a) Convolutional Neural Network: Fig. 13 shows the training and validation accuracy of the gender model implemented with convolutional neural networks. Fig. 14 shows the training and validation loss variation with respect to epochs of CNN for the gender model.

- Both accuracy and loss curve become stable after 10 epoches.
- Accuracy curve has an increasing trend that finally remains between 0.85 and 0.90.
- Loss curve has a decreasing trend that finally remains between 0.25 and 0.3.

Fig. 15 is the confusion matrix that aggregates the prediction result compared with the true label.

- Most of the predictions are correct, located on the diagonal from left top to right bottom of the matrix.

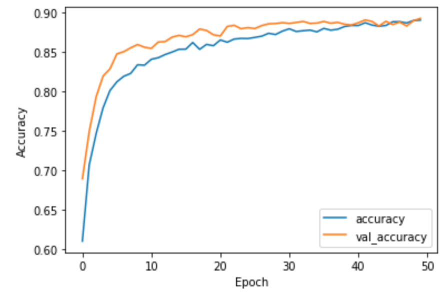

Fig. 13. Training and validation accuracy of the gender model (CNN)

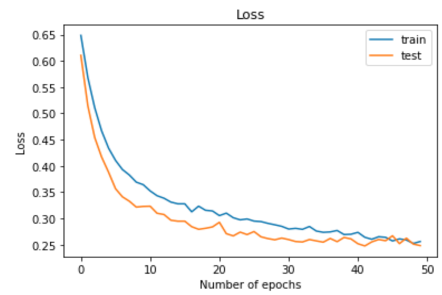

Fig. 14. training and validation loss variation of the gender model (CNN)

- The accuracy from the confusion matrix can be calculated as 0.8856.

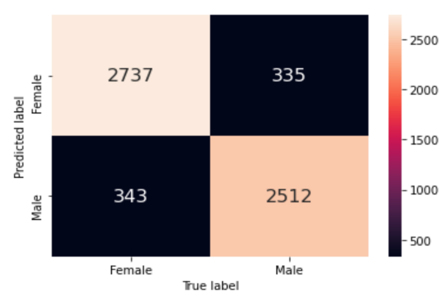

Fig. 15. confusion matrix of gender classifier (CNN)

Fig. 16 is the ROC curve that shows the corresponding true positive rate against false positive rate along the moving threshold

- The area under the ROC curve is 0.96 very close to 1, which indicates that it has a very good performance on prediction accuracy.

b) Deep Neural Network model and DNN with PCA approach: The training and validation accuracy and loss variation of both networks does not change over epochs. And the accuracy of both networks are constant around 50% and, therefore, the classification was not done properly.

- In classification networks the output tries to predict the probability of being included in a specific class. Therefore, the output of the DNN trained from reduced

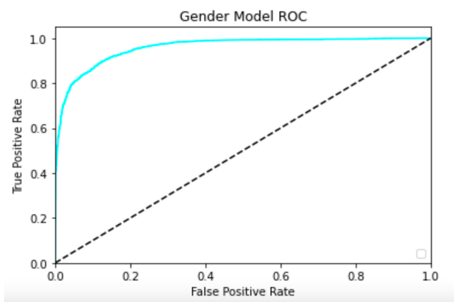

Fig. 16. ROC curve of the gender classifier (CNN)

dimensionality data from the PCA approach is not a good choice for age classification tasks.

3) Ethnicity model:

a) Convolutional Neural Network: Fig. 17 shows the training and validation accuracy of the ethnicity model implemented with convolutional neural networks. Fig. 18 shows the training and validation loss variation with respect to epochs of CNN for the ethnicity model.

- Both accuracy and loss curve become stable after 5 epoches.
- Accuracy curve has an increasing trend that finally remains around 0.75.
- Loss curve has a decreasing trend that finally remains between 0.7 and 0.8.

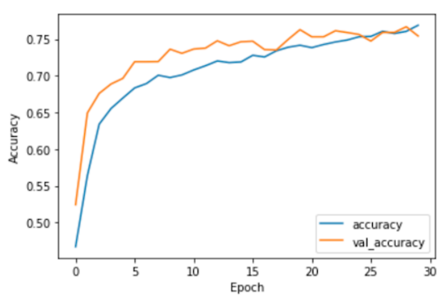

Fig. 17. Training and validation accuracy of the ethnicity model (CNN)

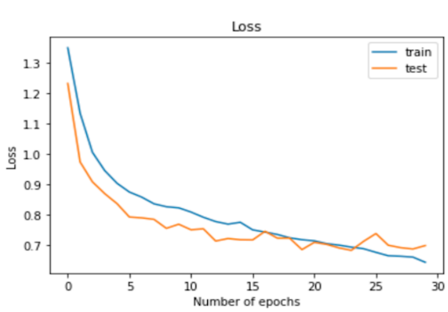

Fig. 18. Training and validation loss variation for the ethnicity model (CNN)

Fig. 19 is the confusion matrix that aggregates the prediction result compared with the true label.

- Most of the predictions are correct, located on the diagonal from left top to right bottom of the matrix.
- The accuracy from the confusion matrix can be calculated as 0.7682.

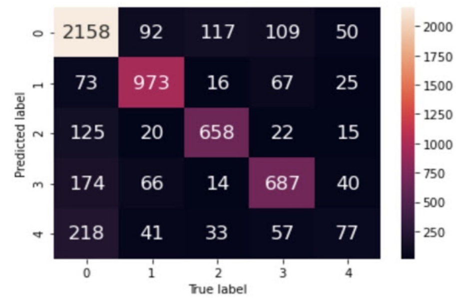

Fig. 19. Confusion matrix of the ethnicity classifier (CNN))

Fig. 20 is the ROC curve that shows the corresponding true positive rate against false positive rate along the moving threshold differentiated by each class.

- The area under the ROC curve of class 0, 1, 2, 3 are all over 0.9, which indicates great prediction accuracy of those classes.
- The area under the ROC curve of class 4 is 0.8 which is significantly smaller than the other classes. This is due to the less amount of data for the class 4 in the dataset.

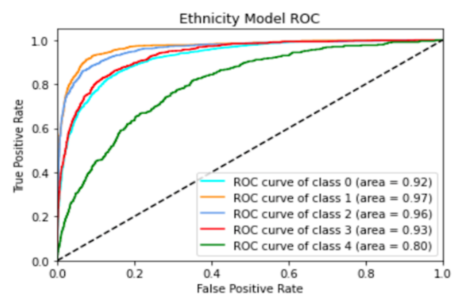

Fig. 20. ROC curve of the ethnicity classifier (CNN)

b) Deep Neural Network model and DNN with PCA approach: Same as the gender classifier, the training and validation accuracy and loss variation of both networks does not change over epochs. And the accuracy of both networks are constant around 50% and, therefore, the classification was not done properly. Therefore, the output of the DNN trained from reduced dimensionality data from the PCA approach is not a good choice for ethnicity classification tasks.

B. *filter visualization*

1) Filters of convolutional neural network: Due to the complexity of deep learning models, the inner working of neural networks is often a mystery. That is one of the main challenges the AI community has been facing over the decades. With the idea of explaining neural networks exploring through the black box architecture helps to understand why the model might be failing to classify or predict the output properly, hence fine-tuning the model for better accuracy and precision.

The objective of the convolution process of convolutional layers is to extract the high-level features from images. Therefore, visualizing filters of convolutional layers is a way to understand what type of features are extracted from each channel of the convolutional layers.

Fig. 21, Fig. 22, Fig. 23, are a representation of extracted features from the first, second and third convolutional layers respectively. From the visualized filters of CNN it can conclude that the eyes, nose, beard, mouth are some prominent features in the dataset.

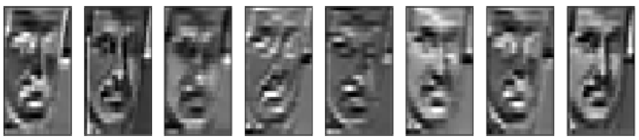

Fig. 21. 16 filters of first convolutional layer

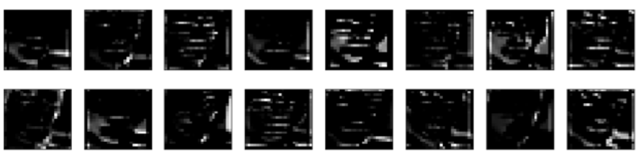

Fig. 22. 16 filters of second convolutional layer

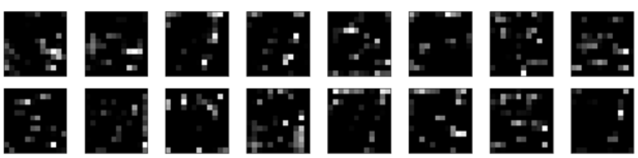

Fig. 23. first 16 filters of the third convolutional layer

2) Principal component Analysis: Fig. 24, is a representation of the first 60 principal components of the dataset. Principal components represent the directions of the data that explain a maximum amount of variance. That helps to understand the most significant features of facial data and what features have the most variance in the dataset. Therefore, it describes what features of the facial data make the highest impact to predict the age value corresponding to each facial image. According to the figure 24, it can be concluded that eyes, nose, hair, mouth are some of the prominent features of the dataset.

C. Generative model

The results from the generative network is shown in Fig. 25,. The representation from left top corner to right bottom corner represents the person from age 21 to 30 respectively. Even though there is a difference in specific pixels, there is no visible significance in generated images due to the lower resolution. There are few reasons for that,

Fig. 24. First 60 principal components of the dataset

- Generative networks are computationally more expensive than typical CNN or DNN networks. Therefore, training on CPU takes longer time and needs to run for a higher number of epochs to produce better results.
- Less popularity of data for some age values is also a reason for poor generation of facial images.

Fig. 25. The age from 21 - 30 of a person

VI. Conclusion & Future Work

Comparing the performances of three models, CNN does the best job on gender prediction as the accuracy becomes stable over 80% the AUROC reaches 0.889. CNN also has a great performance when doing ethnicity prediction, the accuracy becomes stable around 0.75, and the values on the diagonal of the confusion matrix is significantly larger than the others, which means the CNN model has good accuracy and sensitivity to predict. As for age regression, CNN still does the best prediction. From the predicted value vs true value plot, CNN has the strongest positive correlation among the three models. Thus, CNN is the best performing network.

PCA is promising in extracting most significant features from images and it reduces the dimensionality minimizing information loss while increasing the interpretability. Visualizing filters of CNN models and principal components of PCA approach help to understand the insight of the model and why the model might be failing to predict the output hence fine-tuning the model for better precision and accuracy. Generative models have interesting applications in many areas and it is comparatively computationally expensive.

As a future work, for the purpose of general applications the dataset can be improved to avoid class imbalance and for the generative model GPU training will be a solution to avoid computationally expensive problems. Further the generative model can be improved by improving the loss function as a customized loss to consider both the adversarial loss and design loss.

REFERENCES

[1] Castelluccia, C.& Inria, D. L. (2020). Impact Analysis of Facial Recognition: Towards a Rigorous Methodology. 2020. hal-02480647f. Retrieved from https://hal.inria.fr/hal-02480647/document.

[2] Lian, H. & Lu, B. (2007). Multi-view gender classification using multi-resolution local binary patterns and support vector machines. International Journal of Neural Systems, 17(06), 479-487. Retrieved from http://dx.doi.org/10.1142/s0129065707001317.

[3] Dhamecha,T. I., Sankaran, A., Singh, R. & Vatsa, M. (2011). Is gender classification across ethnicity feasible using discriminant functions? 2011. International Joint Conference on Biometrics (IJCB).

[4] Buchala, S., Gale, T. M., Davey, N., Frank, R. J. & Foley, K. (2005). Global and feature based gender classification of faces: A comparison of human performance and computational models. Modeling Language, Cognition and Action.

[5] Liu, S., Sun, Y., Zhu, D., Bao, R., Wang, W., Shu, X. & Yan, S. (2018). Face Aging with Contextual Generative Adversarial Nets. Retrieved from https://arxiv.org/pdf/1802.00237.pdf.

[6] Suo, J., Zhu, S. C., Shan, S., & Chen, X. (2010). A compositional and dynamic model for face aging. IEEE transactions on pattern analysis and machine intelligence, 32(3), 385–401. Retrieved from https://doi.org/10.1109/TPAMI.2009.39.

[7] Kemelmacher-Shlizerman, I., Suwajanakorn, S. & Seitz, S. M. (2014). Illumination-Aware Age Progression. 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014, pp. 3334-3341, doi: 10.1109/CVPR.2014.426.

[8] Age, Gender and Ethnicity (Face data) — Kaggle. (2020). Retrieved September 2020, from https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv.