Age and Gender estimation from faces using deep learning: A comparative analysis

North South University

Department of Electrical and Computer Engineering

1. Monjur Hasan, 1821591642
monjur.hasan@northsouth.edu

2. Mostak Ahamed ,1821737042
mostak.ahamed@northsouth.edu

3.Muhtasin Mashrur Adit ,1821831042
muhatasin.adit@northsouth.edu

Abstract—Age and gender, 2 of the key facial attributes, play a really foundational role in social interactions, creating age and gender estimation from a single face image a crucial task in intelligent applications, like access management, human-computer interaction etc. Advancement in computer vision makes this prediction even more practical and open to all, thus enabling the world to come up with datasets, one of which, used in this paper, is UTK-Face that has 23705 pictures of male and female ageing from 0 to 116. In this paper, we propose a Convolution Neural Network (CNN) to predict age and gender. CNN is a Neural Network (NN) algorithm that extracts the deep features from the image and specifies the desired output at the final layers. Age prediction is approximately near to the real values with a five difference in both ways. Gender prediction is accurate in all the test data presented to the model. Validating with arguments shows no change in training and validation. Our model successfully executed approximately 80% in gender prediction and 60% in age prediction that can be furtherly advanced with pipelining with other classification models and much larger real-world dataset.

Keywords—*UTK-face, Classification, CNN, optimizer, cross-entropy-Loss*

Introduction:

Face recognition has consistently been one of the most difficult and significant tasks in Computer Vision over the years. Face recognition has sparked a lot of attention. Initially, it was linked to security concerns, as automatic face analysis might aid security services in detecting threats Fraudulent passports, the identification of criminals or missing children, the prevention of identity theft, and so on. The initial algorithm was developed by Face recognition technology and was first introduced in the mid-1960s ***(Chan., 1965)***. And Since then, several methods to the problem have been offered in literature. Prior to the introduction of deep learning in 2006 ***(Hinton, 2007.)***, Face analysis algorithms were created by constructing algorithms. The age estimation plays a prominent role in the applications.

like biometric evaluation, 3d face marking, virtual makeup, and virtual try-on applications for jewelry and eye-ware by mapping the face according to the age found. Lens kart is such an application that gives the try-on option for their customers. Age estimation is a subfield of face recognition and face tracking which in combination can predict the health of the individual. Many health care applications use this mechanism to keep track of health by monitoring their daily activities. China uses this face detection technique in service driver identification and jaywalker identification. Some other countries use it for worshipper identification and advertising etc. ***(Anda, 2018)***. Finding the correct dataset for training the model is a crucial task. Since the real-time data is massive, the

computation and the time to prepare the model are high. It's been a tough task after implementing several methods from machine learning, and the accuracy increases drastically. By eliminating the barrier of expressions, we have the possibility to find the best features leading to an accurate measure of gender and age. To predict the age and gender, we use a vital range of machine learning and deep learning algorithms. CNN (convolution neural network) is one of the most used techniques for age and gender detection. In this paper, we use open cv and CNN to

predict the age and gender of any given person's image.

2. Related works:

Md. Hafizur Rahman, Md. Abul Bashar has identified the age and gender by converting the RBG (red, blue, green) image to YCbCr (luminance, chrominance blue, and red) format as it is used in video compression. The primary usage of this format is to find the skin-tone. The lighting compensation (LC) algorithm is used in the pre-processing phase to enhance the image and restore natural colors. Gabor filter is used for feature extraction from the pre-processed image and classified using the logistic regression. ***(M. H. Rahman, 2013)***

The age and gender are detected using a deep residual learning network with connections that includes a gender estimation network and a gender-specific age network, with the outputs from the gender network being used as weights to estimate the outputs of the two age networks. For age estimation, the suggested model employed regression. ***(S. H. Lee, 2018)***

Recently, to better handle the non-stationary property of the aging process, ordinal regression is employed for age estimation. Yang et al. employ the Rank Boost algorithm, which is a single hyperplane ranker in the feature space, for age estimation. Chang et al. [***(Lin., 2006.)***] employ the parallel hyperplanes model OR-SVM [***(K. Chang, 2011.)***]

for age estimation, and further extended it to a more flexible scenario, i.e., several possibly non-parallel hyperplanes [***(B.-C. Chen, 2014.)***].

Data-driven approach has also been used. Chen et al. propose a coding framework called Cross-Age Reference Coding (CARC). The method is based on the assumption that if people look alike when they are young, they might also look similar when they both grow older. By leveraging a large-scale image dataset freely available on the Internet as a reference set, CARC is able to encode the low-level feature of a face image with an age-invariant reference space. Two images of the same person have similar representations using CARC for the reason that they both look similar to certain reference people with different ages ***(K. Chang, ranking approach for human age estimation based on face images, 2010)***.

Real age estimation, as in the case of apparent age estimation, involves target biases: those that are intrinsic to the target's visual face appearance; those that depend on gender, ethnicity, face expression, and so on. The work of [***(J. Xing, 2017)***] deals with the biases introduced by gender and ethnicity by posing the age estimation problem as a multi-task classification problem. In [***(Z. Lou, 2018)***], expression-invariant age is estimated using structured learning.

3. Methodology:

In this paper the UTK Face dataset was used to create a CNN model that distinguishes age and gender. We use 0 to 116 ageing people to estimate their age and their gender. The data is trained, and the testing is done on splitted images. We trained our model in the pre-trained resnet architecture and also built our own model and trained. Since the data is broad and computation and prediction with this kind of data will be hard and time-consuming. So, there is a requirement of the optimizer that is Adam optimizer. All the models use the TensorFlow platform in the background to generate the age and gender of the particular image.

3.1 Convolution Neural Network:

CNN is the most popular deep learning algorithm, used tremendously in computer vision. It is computationally efficient and instinctively spots prominent features without any supervision. CNN has the same layers compared with the traditional neural network, but the hidden layer consists of different kinds of internal layers, specifically the convolutional layer, pooling layer, fully connected layer, and normalization layer. ***(P. Yang, 2010.)***CNN has a prominent role in computer vision and image processing, to process an image which is an array of pixels taken in the form of a matrix. An image can be in color or black and white form that is RBG, grayscale, respectively. The image matrix that enters the convolution layer merges with the kernel or convolution filter that leads away from the feature map. In our model we have five convolutional layer five (2*2) Max pooling layer, fully connected layer and two output layers. And in the end, we make 2 output layers in one output.

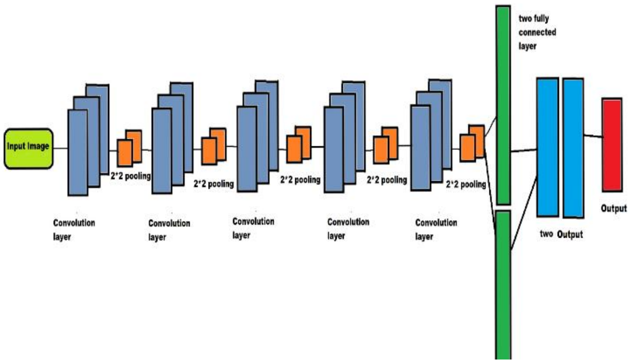

Figure 1: Model Architecture

We give (48*48*3) Image in the input and filtering over layer by layer. By the change of the filter edge detection, sharpening, identity, and box blur operations can be performed on the input. The sliding of the filter in this layer is determined by the stride padding. After convolution, we use max-pooling for every time and the data is reduced dimensionally. In our model we use five (2*2) max pooling. The output of the pooling layer is flattered to generate a vector as a fully connected layer takes vectors as input. The process of flattering includes the conversion of a 3-dimensional matrix to a one-dimensional vector. For calculating loss, we are using Binary cross entropy for gender estimation and Mean absolute error for Age estimation. We also use sigmoid as the last activation function for gender and for age we use 'Relu' activation function.

MAE= Mean Absolute Error

MAE= $\frac{\sum_{i=1}^{n}|y_i-x_i|}{n}$

Yi= prediction

Xi= true value

N= total number of data point

Binary-cross-entropy loss

Loss=$-\frac{1}{n}\sum_{1}^{n}\sum_{1}^{m} y\mathbf{i}j\, log\left(p_{ij}\right)$

N= number of rows

M= number of classes

3.2 Resnet-50

We also use pretrained Resnet-50 models to train our model. its a 50 layer deep neural network and it has a large training parameter. This model contains five stages with residual blocks containing three layers, each with the convolutions. As the traditional NN, this network with remaining blocks provides the output of one layer to the next layer and also provides by hoping 2-3 layers away. This process is named as identity connections.

3.3 Adam Optimizer:

This optimization is used in place of stochastic gradient descent. The key features of the algorithm are computationally useful, less memory required, suitable for extensive data, easy to implement, used for noisy data.

3.4. Experimentation:

We use the UTK-face dataset for the identification of age and gender. We use this dataset to train our model, testing, modeling and evaluate our model.

3.4.1 Dataset: In the UTK-Face dataset there are different male female ageing people (0-116 years) and there are 23705 images there. Each image has close-up and clear face characteristics, as well as expressions and diverse stances, and the images are of varying resolutions. Each image is labeled by age and gender. formatted like [age]_[gender]_[race]_[date&time].jpg. There are a total 105 classes. Some classes are missing on this dataset. Missing classes are [97,98,102,104,106,107,107,109,112,113,114].

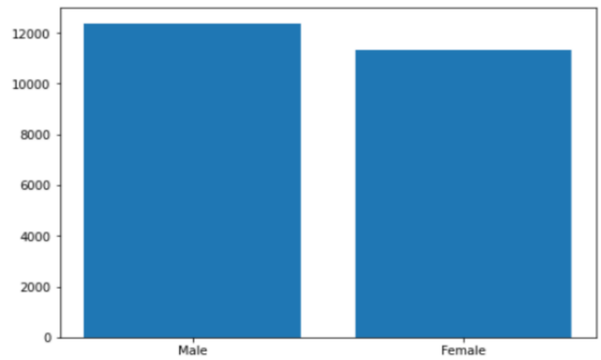

Figure 2: number of male female in the dataset

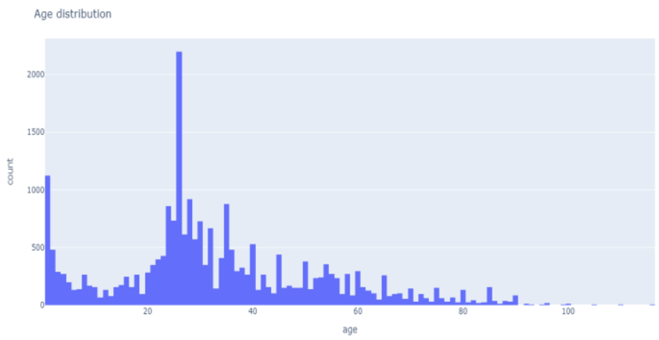

Figure: Age distribution

3.4.2 Pre-Processing:

Before applying images to our model, we pre-process our images. At this stage, all the images in the dataset are inspected to find out if there is an image file without the label in the format of its representation ***(Avuthu Sai Meghana, 2020)***. If any such image file is found, then it is either removed/any estimated values are assigned. Since we have the cropped images along with the data set else, this process of cropping can be categorized under the pre-processing phase. And we also store every picture in their particular classes before applying to our model. And this is also done under the pre-processing section.

3.4.3 Training, Validation, and Testing Data:

We split our image into training, validation and testing parts. At first, we separate 20% images from whole images for training. Then also split another 10% for validation. In the UTK-Face dataset, the cropped images don't have any borders or margins around the face, which is called tightly cropped.

3.4.4 Modeling:

For estimating the gender and age through Images, we used our own neural network. In our model we assume age and gender simultaneously ,there is no need for two different models for different attributes.. By the various stages of CNN, the input images are first convoluted according to the filter, and later the required features are pooled out, and finally, by fully connected layer, the features and weights are mapped such that we could estimate the age and gender of the person. We use 64 batch sizes and run 50 Epochs to our model.

3.4.5 Metrics:

We use different metrics to evaluate our model.

Metrics	Formula	Evaluation Focus
Accuracy (acc)	$\frac{TP+Tn}{TP+FP+Tn+Fn}$	Accuracy (acc) In general, the accuracy metric measures the ratio of correct predictions over the total number of instances evaluated.
Average d Accuracy	$\sum_{L=1}^{L}\frac{F_P\mathbf{i}+Tn\mathbf{i}}{TP\mathbf{i}+Fn_i+FP\mathbf{i}+Tn\mathbf{i}}$	The average effectivenes s of all class

3.4.6: Strategy:

Our main strategy is we didn't build two models for two attributes. We build one model that could estimate age and gender at once. and give us the output of two things. In the last layer we extracted two features and summing these two outputs we made one output and ultimately gave us one output of two attributes

4. Evaluation:

At this phase, we cross-validate between the original and our extracted age and gender of that particular person in the image and find the error rate. We use different loss functions in our model. Best accuracy was given by Binary-cross-entropy for gender and Mean absolute error for age estimation.

From UTK-face dataset that we use for training the model. we accept our prediction as:

Figure 3: UTK-face dataset sample

And we try to predict the image accurately as shown in the picture with their age leveling.

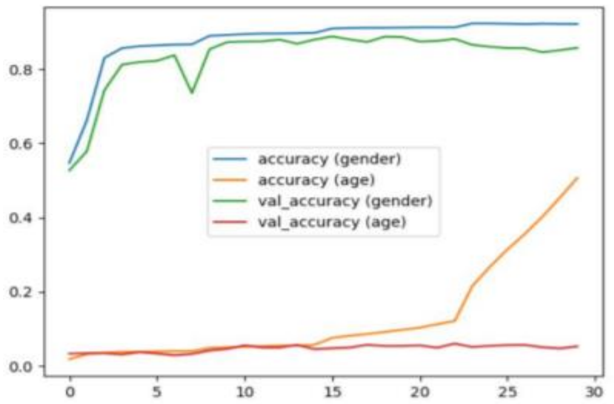

Figure 4: accuracy of Resnet 50 (Epochs vs percentage)

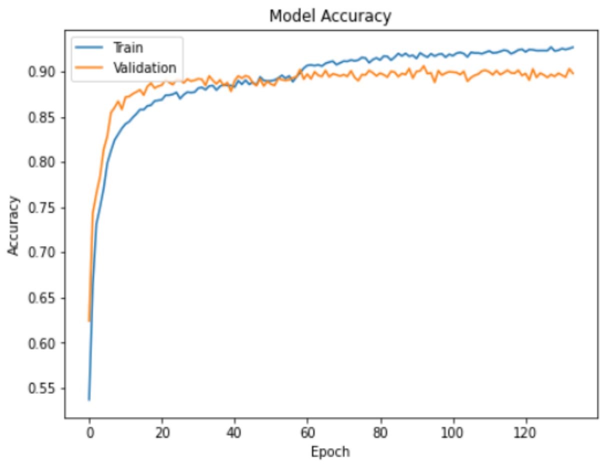

Figure 5: accuracy of our model (Epochs vs percentage)

Here you can see that our model is much more accurate in terms of accuracy we try with resnet-50. In this graph accuracy was average accuracy of age and gender.

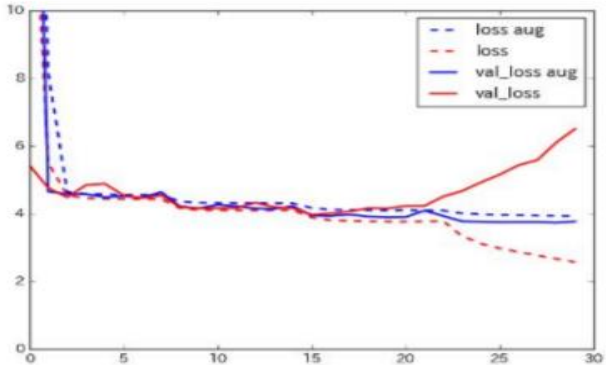

Figure 6: loss of Resnet-50 (Epochs vs loss)

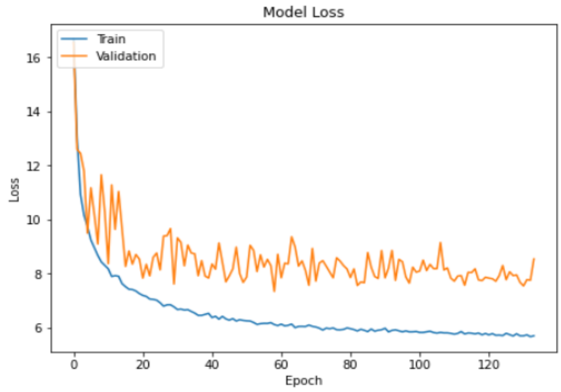

Figure 7: model loss ((Epochs vs loss)

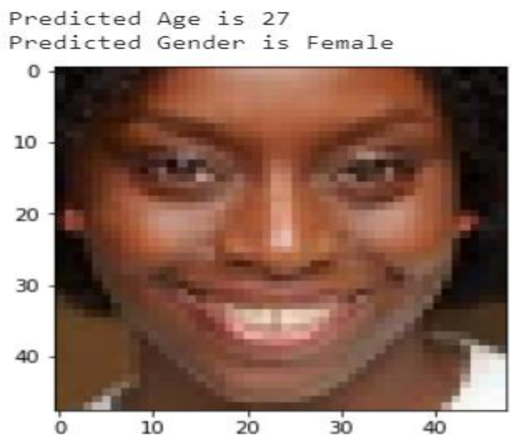

Figure 8& 9: Testing in our model

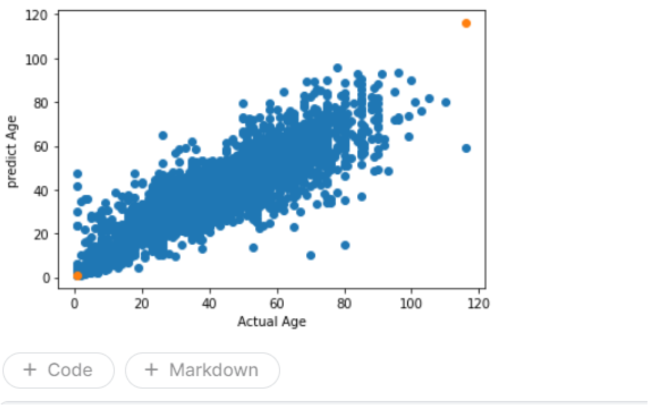

Figure 10: Predicted age vs Actual Age

Conclusion:

We have implemented an efficient method in the detection of gender and age. In comparison to other deep learning approaches, the identification procedure was simple, took less time to compute, and used less memory with our CNN. The resulting results are precise and straightforward, with accurate predictions; the person's gender is appropriately predicted as male or female. Due to the application of bias, the algorithm produces precise results when considering age. We found that after epochs 50 our validation loss is not improving. So our model is accurate on 50 iterations. And our model is precise in estimating age and gender.

I. References

1.Anda, F. (2018). Evaluating Automated Facial Age Estimation Techniques for Digital Forensics. *IEEE Security and Privacy Workshops*. Retrieved from https://doi.org/10.1109/SPW.2018.00028

2. Avuthu Sai Meghana, S. s. (2020). Age and Gender prediction using Convolution, ResNet50 and. *International Journal of Advanced Trends in Computer Science and Engineering*, (p. volume 9 no 2). Tamil nadu.

3. B.-C. Chen, C.-S. C. (2014.). Cross-Age Reference Coding for Age-Invariant Face Recognition and Retrieval. *pringer International Publishing, Cham*, 768–783.

4. Chan., W. W. (1965). A man-machine facial recognition system:. *some preliminary results. Panoramic Research, Inc,* . Palo Alto, California.: echnical Report PRI A.

5. Hinton, G. E. (2007.). Learning multiple layers of representation. *rends in cognitive sciences, 11(10*, (pp. 428–434,).

6. J. Xing, K. L. (2017). Diagnosing deep learning models for high accuracy age estimation from a single image., (pp. 66:106–116,).

7. K. Chang, C. C. (2011.). Ordinal hyperplanes ranker with cost sensitivities for age estimation. . 585– 592.

8. K. Chang, C. C. (2010). ranking approach for human age estimation based on face images. *ICPR,* .

9. Lin., L. L. (2006.). Ordinal regression by extended binary classification. *NIPS,*, (pp. 865–872).

10. M. H. Rahman, M. A. (2013). An automatic face detection and gender identification from color images using logistic regression,. *2013 International Conference on Informatics, Electronics and Vision (ICIEV)*, (pp. pp. 1-6). Dhaka.

11.P. Yang, L. Z. (2010.). Ranking model for facial age estimation. *ICPR.*

12. S. H. Lee, S. H. (2018). Age and gender estimation using deep residual learning network," 2018 International Workshop on Advanced Image Technology., (pp. pp. 1-3). Chiang Mai.

13. Z. Lou, F. A. (2018). Expression-invariant age estimation using structured learning. *IEEE transactions on pattern analysis and machine intelligence*, (pp. 40(2):365–375).