Detecting Age and Gender through Convolutional Neural Networks

Kuber Shahi, Satpreet Makhija, Dhruv Khandelwal

December 6, 2020

ABSTRACT:

The human face conveys much information, which we have a remarkable ability to extract, identify, and interpret. Age and gender are two such information that we can reliably infer. Recently, there has been an increase in the development machine based systems that can mimic these abilities of our visual system. Among many other approaches, convolutional neural networks (CNNs) have been remarkably successful in imitating our visual system. In this paper, we build a CNN to predict the age and gender of a person with the aid of an image provided as input to the CNN and analyse our results with the industry benchmark.

1 INTRODUCTION

Collection of data in this age has become an utmost important task. Age and gender are two important attributes of such data. It plays an essential role in a wide range of real-world applications such as targeted advertisement, forensic science, visual surveillance, content-based searching, human-computer interaction systems, etc. However, gender classification is still an arduous task due to various changes in visual angles, face expressions, pose, age, background, and face image appearance. It is more challenging in the unconstrained imaging conditions. Facial recognition is a field that has been growing at an exponential rate, and age and gender identification is a problem at the heart of this field. It is therefore important to be able to correctly estimate a person's gender and age from facial images.

1.1 DESCRIPTION OF DATA SET

The database used for this project is the Adience Data set[5] which contains a total of 26,580 photos of 2,284 unique subjects that are collected from Flickr. Each image is annotated with the person's gender(assuming binary genders) and age-range (out of 8 possible ranges). The images are subject to various levels of occlusion, lighting, and blur, which reflects real-world circumstances. Our model is trained on 12,111 examples due to absence of proper labelling in the rest of the examples. The details of the **training dataset** is mentioned below.

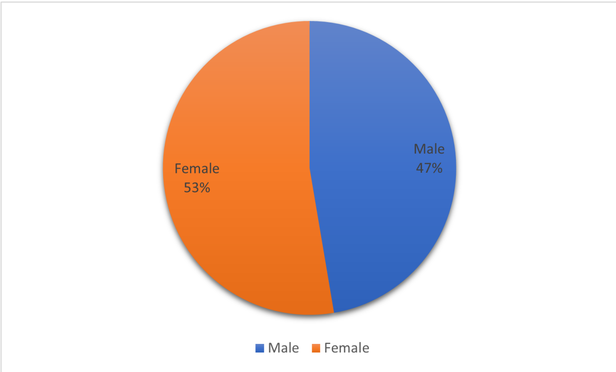

Figure 1.1: Gender Split of Data

Figure 1.2: Some example Images

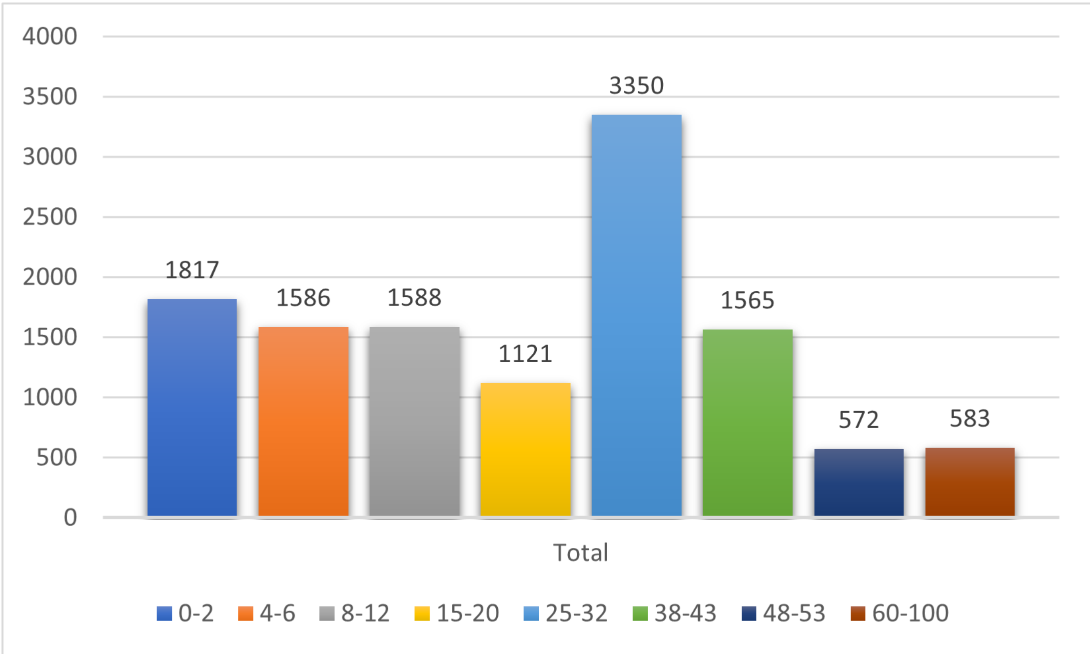

Figure 1.3: Age Split of Data

2 Existing Approaches

The earliest methods which addressed the gender and age classification problem are known as appearance based methods. In appearance based methods, features are extracted from face images considering the face as a one-dimensional feature vector and then a classification tool is used. The features were extracted using Principle Component Analysis which was not performing optimally[6].Some earlier researchers extracted pixel intensity values as well and then fed these values to the classifiers. Typically, the classification is performed through a binary classification strategy. The mostly used classifier for the automatic gender classification is a support vector machine; other classifiers applied included decision trees and neural networks. Another method used for gender classification is known as the **geometric method**.These models extract facial landmark information from face images and build a model based on landmarks information. The geometric models maintain a certain geometric relationship between different face parts. These models are affected by pose variation and discards valuable pixel information.

However, Deep Convolutional Neural Networks has shown outstanding performance for various image recognition problems. The CNN based methods are applied to both feature extraction as well as classification algorithm for the automatic gender classification. A hybrid system for gender and age classification was presented in [4] . Features were extracted through CNNs, and an extreme learning machine (ELM) was used for classification. This hybrid model is known as ELM-CNNs. The ELM-CNNs were evaluated on two public databases, MORPH-II and Adience.The ELM-CNN is the best algorithm performing on gender classification thus far. Recently, gender classification through face segmentation was done, and had performance very close to deep CNNs. This model used CNNs to extract segments of faces, and then used classification models on the same. They achieved high accuracy.

To the best of our knowledge, the current state of the art results for age and gender predictions are reported in [7] with 64% and 91% accuracy respectively. The model from [7] used the VGG-16 layer architecture, which has been pre-trained on the IMDB-WIKI face data set. The network architecture of the VGG-16 model is incredibly complex with about 16 layers.

3 Our Implementation

The multiplicity of dimensions constricts us to not use a simple neural network. A one megapixel image which has an RGB channel gives us $1000 \times 1000 \times 3 = 3$ million input features. Consider the case where the first layer of our simple neural network with 1000 nodes. The weight matrix will be of dimension 3×1000 million . It is not feasible to train a model with such a large weight matrix dimension due to computation power limitations whereas using a CNN reduces the number of parameters exponentially which provides us the capability to create a CNN architecture with multiple layers. Hence, we used CNNs for our age and gender detection model and trained two different CNNs, one for gender identification and another for age identification.

3.1 CNN Architecture

Gender Model: The gender CNN is a sequential model of three convolutional layers followed by three fully connected layers. It takes an input image of size $(150 \times 150 \times 3)$. The images are taken in RGB mode: hence the 3 channels in the input image size. We decided to take the input images in RGB mode as they represent the real-world images without incurring the loss in information which is not the case with grayscale images. Each convolutional layer has 32/48 filters with ReLU as the activation function which is followed by a 2D max pooling of size (2×2). This reduces the input image size to $(75 \times 75 \times 3)$. Finally, a dropout is applied at the end of each convolutional layer. The first two fully connected layers have 128 and 64 neurons respectively with ReLU as the activation function which is followed by a dropout at the end. The final layer of fully connected layers has only one neuron with sigmoid activation function as the gender classification is a binary classification.

Age Model: The age CNN too is a sequential model but has four convolutional layer followed by four fully connected layers. It also takes an input image of size $(150 \times 150 \times 3)$. Each convolutional layer has 48/64 filters with ReLU as the activation function which is followed by a 2D max pooling of size (2×2). This reduces the input image size to $(75 \times 75 \times 3)$. Finally, a dropout is applied at the end of each convolutional layer. The first three fully connected layers have 128, 128 and 64 neurons respectively with ReLU as the activation function which is followed by a dropout at the end. The final layer of fully connected layers has only 8 neurons to represent 8 output classes of age with softmax activation function applied on them.

A visualization of both CNNs has been shown below.

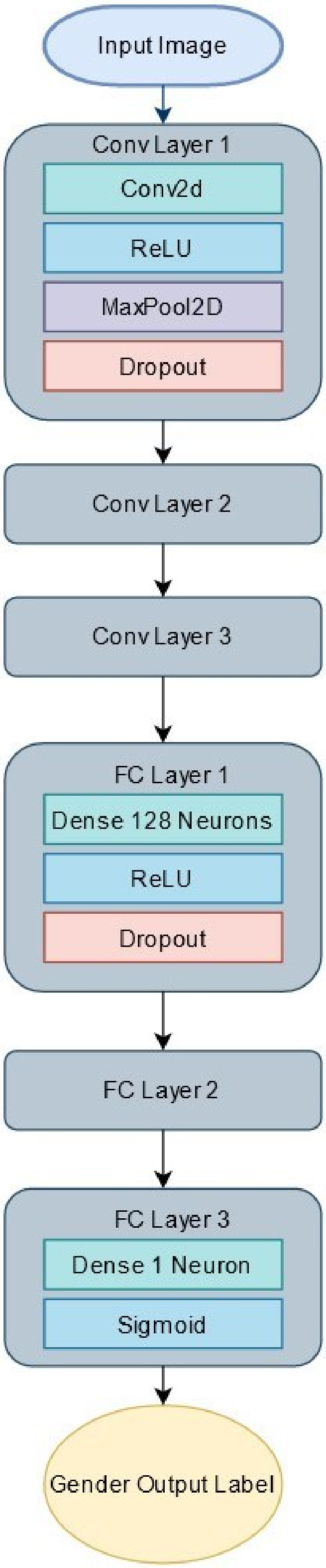

Figure 3.1: Gender Model

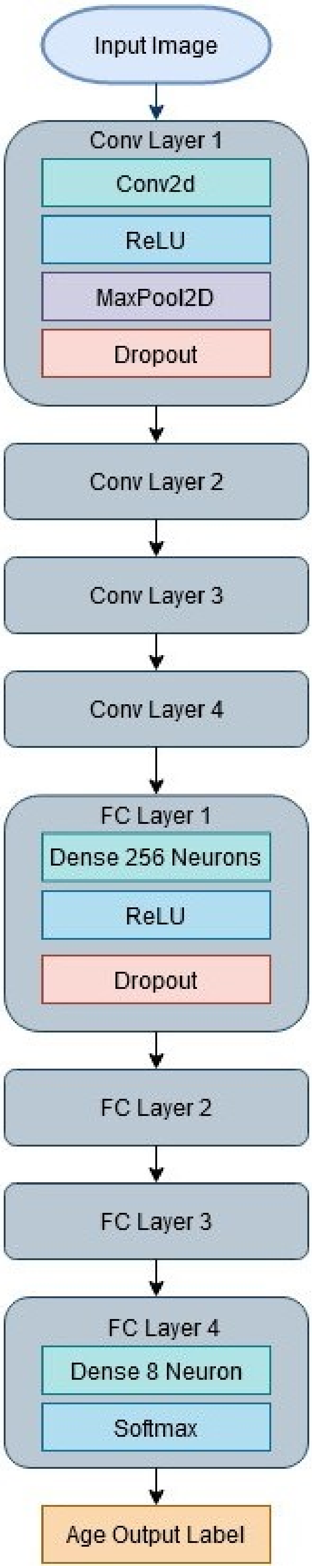

Figure 3.2: Age Model

3.2 Training & Testing

All three color channels are processed directly by the CNN. The input images are rescaled to 150 x 150 pixels. We used data-augmentation by rotating the image at multiple examples, followed by mirroring of the image. Multiple zoom percentages are also applied. This helped in increasing the diversity of data available for training. All the weights are initialised using Gaussian parameters with a standard deviation of 0.05 and mean value 0. We included multiple dropout layers with mean dropout ratio 0.3 to reduce overfitting. We have distributed the dataset as **80 % training, 10 % validation and 10 % testing**. We have used **Binary Cross Entropy as loss function for Gender Model** and **Categorical cross entropy for Age Model** respectively.

Gender	Male	Female
Train	4525	5174
Validate	607	602
Test	601	603

Age	0-2	4-6	8-12	15-20	25-32	38-43	48-53	60-100
Train	1514	1288	1299	916	2895	1259	482	483
Validate	154	149	150	107	228	157	43	50
Test	149	149	139	98	227	149	47	50

Figure 3.3: Data Split Between Training, Validation and Test

We trained and tested our network using RMSprop, which is a gradient descent method that uses adaptive learning rates for faster convergence. This normalization balances the step size, decreasing the step for large gradients to avoid exploding, and increasing the step for small gradients to avoid vanishing.

The training of the Gender Model was relatively easier as it did not suffer from overfitting. This was also because the distribution of male and female data points was of the ratio 1.2: 1, i.e., the dataset for gender detection was not skewed. Although, **the model did not predict gender for babies with higher accuracy**. This is because during the infancy stage, both the genders share many facial patterns and hence it because difficult to separate the two on the basis of gender. With increase in age, the two genders start differing more in their facial patters.

We encountered the following problems while training the Age Model:

1. **Data Skewness and class weights:** Our dataset is skewed. We have 5 times more number of data points for age label (25-32) than for age labels (48-53) and (60-100). Similarly, there is disparity in the ratio of examples of different labels with the above mentioned being the most extreme. This reduced the accuracy drastically as the model was trained on one type of label more than the others. Hence, to correct the skewness of our data, **We introduced class weights to create an asymmetrical loss function**. We used the formula $\frac{\text{total number of samples}}{8\times \text{ number of samples in a class}}$ to assign higher weights to the classes with relatively fewer examples. This helped improve our accuracy by a substantial amount.

2. **Overfitting:** While training our model, we noticed that after epoch 4, the training accuracy and the validation accuracy were going in opposite direction with the former increasing and latter decreasing. This provided evidence for overfitting of the data. To overcome this

problem, we performed the following strategies:

a) **Early Stopping:** While training our model we noticed that the last epoch was changing the loss by a substantial amount and hence overfitting the model to the data. Therefore, we reduced the epochs for the age model from 8 to 6 to 5, which gave best accuracy. Epoch 4 was still underfitting and therefore epoch 5 gave us the sweet spot.

b) **Dropout:** We tried many different dropout rates in the Age Model. Since dropout functions reduces overfitting, we set up a higher dropout rate for the age model which was effective to some extent in reducing overfitting.

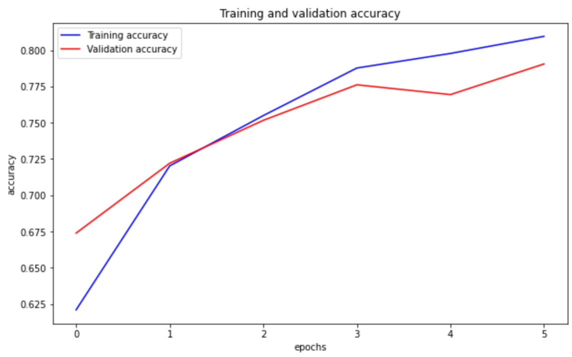

Figure 3.4: Accuracy of gender model

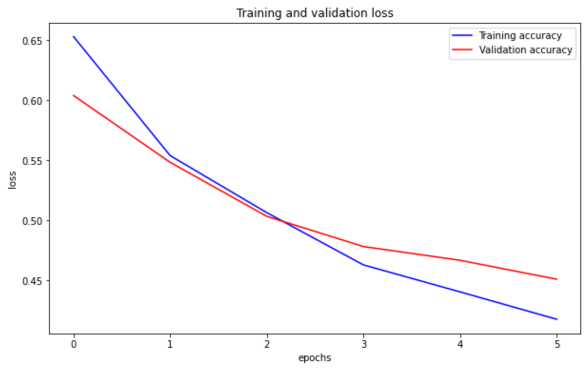

Figure 3.5: Loss of gender model

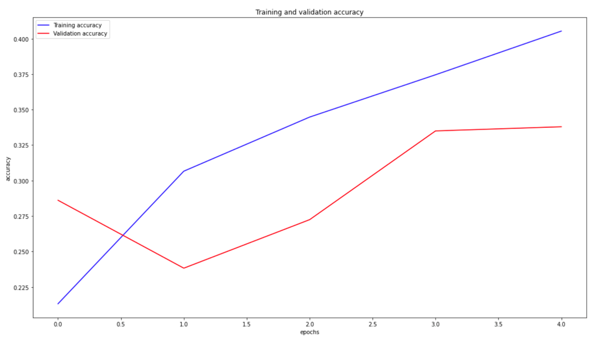

Figure 3.6: Accuracy of age model

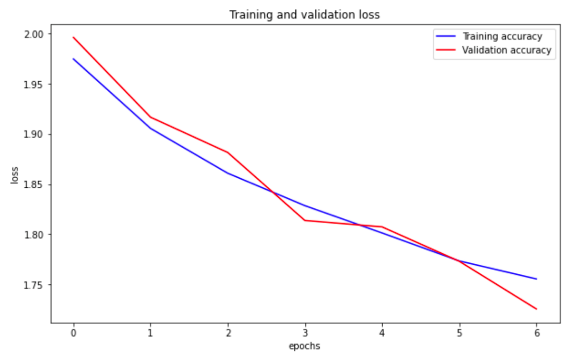

Figure 3.7: Loss of age model

4 Results and Analysis:

To determine our accuracy, we checked the percentage of correct labels against the test data. Our gender network performed with an accuracy of 80.5% , while our age network gave us an accuracy of 43%. The current industry benchmark is given by [1]. The table below summarises our results against the benchmark.

Parameters	Age	Gender
Benchmark	50.7	84.7
Our Model	43	80.5

Figure 4.1: Accuracy of our models

The benchmark performs better than our model for three reasons:

1. We used a dataset of around 12,000 images while the benchmark model is trained on around 26,000 images. Rest of the images had incomplete labeling and therefore were unusable from our point of implementation.

2. The benchmark model also used over-sampling, which gives them a boost of approximately 3-4%. Again this was not a viable option for us, due to lack of computing power.

3. The benchmark used CNNs with more features, bigger kernel sizes, and more layers which allowed them to train models with higher accuracy. However, this was not possible for us due to computational and technical limitations.

Nonetheless, to get a better understanding of what our model was doing, in other words, what it was doing right and where it was going wrong, we acquired the confusion matrix for both the models.

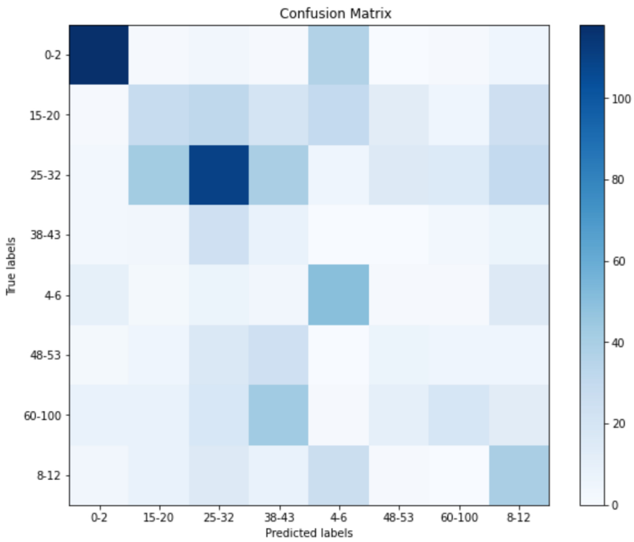

Figure 4.2: Age Confusion Matrix

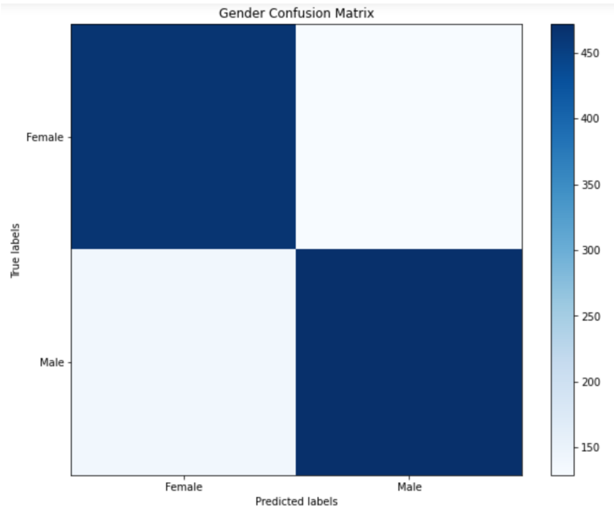

Figure 4.3: Gender Confusion Matrix

The facial features of individuals of age (0-2) is very different from that of every other age group. This is clearly evident to the naked eye. We induced the same from the results we obtained in the confusion matrix for Age (see Figure 4.2). The age labels (15-20) and (25-32) is relatively difficult for the model to predict accurately. This is because of two factors. First, these two age groups share a lot of facial features and second, we have more examples for age group (25-32), although we reduced this effect to a substantial amount by introducing class weights. It is because of these reasons that the model performed really well for age group (0-2) and (25-32) as compared to the other age groups. The model however faced difficult in predicting age labels (48-53) and (60-100) due to fewer training examples. Therefore, we felt that with a more balanced database, our model would perform much better.

4.1 Another Benchmark:

Though [1] is considered the industry benchmark, [2] and [3] performed very well too. [2] used a chained neural network and is based on [1]. They chain the gender network to the age network, and therefore train three networks in total. A gender network, a male age network and a female age network. This network outperformed the benchmark in the age network by achieving an accuracy of 54.5. The later[3] used [1] again for the construction of the CNN, but then fine tuned the network to get the best results yet! They used compacting, picking and unforgettable continual learning to get results of 89% in gender recognition and 57% on age recognition.

5 Conclusion

Before training the model, we had hypothesised the following for age and gender classification:

1. The Gender Model's accuracy on babies will be relatively low because of greater indifference in facial features of the two genders during infancy stage.

2. The Gender Model's overall accuracy will be in the high 80s because first, the dataset's distribution for gender was not skewed and second, it was a binary classification problem.

3. The Age Model's prediction will cross 50% with the introduction of class weights.

4. The probability of predicting the age for label (25-32) will be high due to skewness of dataset and also for label (0-2) due to clear distinction of facial features between label (0-2) and others.

Out of the above mentioned hypotheses, 1, 2 and 4 turned out to be true whereas 3 was false. Even with the introduction of class weights and tuning the model with variation in kernel size, number of filters, image augmentation, and optimizers and regularizers, we were not able to cross the 50% mark in accuracy for age prediction. In conclusion, gender prediction is relatively easier as compared to age prediction. But, we believe with a more balanced dataset, age prediction accuracy should increase. Every dataset has its own caveats and must be dealt with empirical evidence, and in our case too, we did the same to optimise our model for better results.

6 REFERENCES

1. Gil Levi & Tal Hassner Alexander, *Age and Gender Classification using Convolutional Neural Networks.*2015

2. Ari Ekmekji.*Convolutional Neural Networks for Age and Gender Classification* ,Stanford University.2016

3. Steven C. Y. Hung, & Cheng-Hao Tu. (n.d.). *Compacting, Picking and Growing for Unforgetting Continual Learning.* Institute of Information Science, Academia Sinica, Taipei, Taiwa.

4. Ali Maina Bukar, Hassan Ugail, David Connah, *Automatic age and gender classification using supervised appearance model*, J. Electron. Imaging25(6), 061605 (2016)

5. Eran Eidinger, Roee Enbar, and Tal Hassner, *Age and Gender Estimation of Unfiltered Faces*

6. Khalil Khan1, Muhammad Attique, Ikram Syed Asma Gul, *Automatic Gender Classification through Face Segmentation.*

7. R. Rothe, R. Timofte, and L. Van Gool. Deep expectation of real and apparent age from a single image without facial landmarks.International Journal of Computer Vision, pages 1–14, 2016.