Gender and Age Classification

$M.Sumanth^1$, $G.RavitejaGowtham^2$ and $B.Vinay^3$
Indian Institute Of Information Technology Chittoor, Sri City, A.P., India
$sumanth.m15@iiits.in^1$, $raviteja.g15@iiits.in$, $vinay.b15@iiits.in$

***Abstract*— Now-a-days, Photos are being uploaded by everyone on social media platforms. Gender and Age Classifications, for the photos uploaded, are becoming popular in our day to day life. Ever since the evolution of social media and social platforms, Classification in regard with age and gender has widely enlarged with its content. Usage of deep-convolutional neural networks has been increased to obtain the higher accuracy in performance.Our Project is an implementation of CNN Model which classifies the person in the image into gender and age. Our CNN can work even better with limited amount of dataset. Adience Benchmark is the dataset being used in our project. In short the key idea is to introduce a simple convolutional net architecture that works well even when the amount of learning data is limited. Our model have been evaluated on recent Adience benchmark for age and gender estimation and show that it performs better than current state-of-the-art methods.**

I. INTRODUCTION

Gender and Age have been playing an important roles in social as well as mutual interactions. We have reserved salutations along with rules for grammar to be used for men and women seperately. We have very often usage of vocabularies while addressing the elders in comparison with adults. Inspite of the basic roles, these attributes comes along in our day to day lives. To estimate these attributes reliably and efficiently is some what far in reaching the expectations from commercial applications. In relation of face recognition, this has been very puzzling.

The recent approaches in estimating these attributes from images captured, which have centered face, have been concentrated via tailored face descriptors or differences in facial features dimensions. Many classification schemes have been raised in particularly estimating gender and age classification. The main challenge behind all these schemes is unconstrained imaging conditions. However, Machine Learning techniques which have been employed for these schemes have not exploited fully in regard of improvement in performance.

The CNN we have implemented consists of three convolution layers and two fully connected layers. Selu is the activation function that we have been using. Before feeding input the images are center cropped through preprocessing to make sure that the face region pixels are more concentrated. The implementation of the network is deep networks compared to the existing networks. So over fitting of the data can be prevented. The dataset used for training the network contains images which are not constrained i.e., there will be pose variation and motion blur in the image. Prediction of age will be done by distinguishing among eight classes and gender among two classes.

Figure 1. Audience benchmark dataset containing faces with pose and illumination variation. Faces from the Adience benchmark. These images represent some of the challenges of age and gender estimation from real-world, unconstrained images. Most notably, extreme blur (low-resolution), occlusions, out-of-plane pose variations, expressions and more

II. DATASET & PREPROCESSING

- All the images are captured in the wild environment. There are 8 labels for the age as follows (0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60-)
- The Audience benchmark dataset comprises of about 26,580 photos
- Gender labels are mentioned for the images along with the age labels.
- Images are center cropped for better feature learning so that only face pixels are used for processing.

We have already mentioned that our dataset is limited. So, we have gone through pre-processing steps like data augmentation which is strong combination of crop, flip, bright, contrast, scale, rotate. In order to get good results, we have been gone through data augmentation process. Many Images dont have gender and age labels. Some Images are not labeled as class, instead given as integer. Data-preprocessing is not an easy task instead long winded task. The data rotation is in range of -10x (2π / 360) to 10x (2π / 360). Scaling is in range of 0.8 to 1.2 along both directions of x and y. Brightness is under the constraints of max_delta with 0.5, upper bound and lower bound values of contrast are under 0.5 & 1.5 respectively.

Figure 2. Images after pre-processing

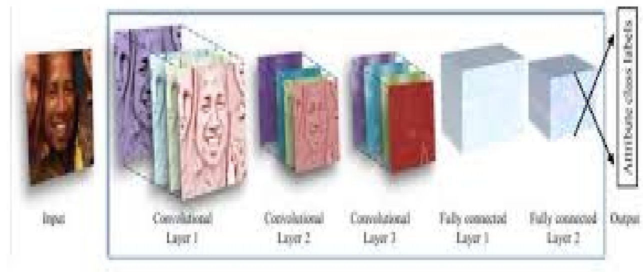

Figure 3. Illustration of our CNN architecture. The network contains three convolutional layers, each followed by a Scaled Exponential linear operation and pooling layer. The first two layers also follow normalization using local response normalization. The first Convolutional Layer contains 96 filters of 77 pixels, the second Convolutional Layer contains 256 filters of 55 pixels, The third and final Convolutional Layer contains 384 filters of 3 3 pixels. Finally, two fully-connected layers are added, each containing 512 neurons.

III. CNN for Gender & Age Classification

To gather such a large along with labeled training set containing images for gender and age classification from user personal information in social media platforms is an tedious task. It consumes alot of time for manual labeling. Datasets for gender and age classification from real-world social images are relatively limited in terms of size. Over fitting is the most common problem in usage of machine learning techniques on such small scale images. To overcome this, one has to collaborate with deep convolutional neural networks with model parameters. One has to be taken care of overfitting. Below is the detailed implementation of cnn network in regard of paper.

- Since Images are in 256 x 256 size, we have cropped into 227 x 227 size.
- In the first convolutional layer, the image is multiplied 96 times with 3 x 7 x 7 size filter.
- Later, it is passed through SeLU Activation Function, followed by a max pooling layer of 3 x 3 size with stride value as 2 and a local response normalization layer.
- In the second convolutional layer, the previous layers output is 96 x 28 x 28 is multiplied 256 times with 96 x 5 x 5 size filter. Later, it is again passed through SeLU and a max pooling layer with same parameters as above along with local response normalization.
- In the final convolutional layer, the output of previous layer is multiplied 384 times with 256 x 3 x 3 size filter. Later on, it is passed through SeLU Function and a max pooling layer.
- The first and foremost fully-connected layer which has 512 neurons gets input from third convolutional layer and later it is passed through SeLU and followed by a dropout layer.
- The second fully-connected layer undergoes changes same as first fully-connected layer.
- The third fully-connected layer classifies into the output.
- The output from third FC Layer is passed through Soft-Max Layer to perform non-maxima suppression.
- Brief description about architecture is illustrated above:

IV. Training & Testing

A. TF Records

To save the large amount of data (images) into a single TF Records format file and load it batch-wise to train your network in tensorflow. To feed the input for tensorflow, we use an input pipeline which takes a list of filenames (any supported format), shuffle them (optional), create a file queue, read, and decode the data. However, TF Records is the recommended file format for Tensorflow. We were provided with a list of files which have details such as labels of gender and age for each image in the entire data set. We have created 8 tf records for gender, age and age&gender seperately. We have used 6 for training and 2 for testing.

B. Reading TF Record Data

We have used inbuilt functions to read and decode the data. We can retrieve the label and image from those inbuilt functions which are nothing but training images and training labels respectively. For testing, we do same as above for testing image and testing labels. Later we can use them for image batch as well as label batch.

C. Initialization

- Classes of age: (0,2) is 0, (4,6) is 1, (8,12) is 2, (15,20) is 3, (25,32) is 4, (38,43) is 5, (48,53) is 6 and (60,100) is 7.
- We have made some assumptions and rounded the classes for age.
- Classes for gender are labeled as shown here m is 0 and f is 1.

	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	*Total*
Male	745	928	934	734	2308	1294	392	442	8192
Female	682	1234	1360	919	2589	1056	433	427	9411
Both	1427	2162	2294	1653	4897	2350	825	869	19487

Table I
THE ADIENCEFACES BENCHMARK.

Architecture	Framework	Accuracy Observed
Proposed architecture	Tensorflow	86.1 +/- 4.7

Table II
GENDER ESTIMATION RESULTS ON ADIENCE BENCHMARK

Architecture	Framework	Accuracy Observed
Proposed architecture	Tensorflow	50.2 +/- 1.2

Table III
AGE ESTIMATION RESULTS ON ADIENCE BENCHMARK

- The weights are initialized using pre-trained VGG Face network weight. VGG face network is a network for face recogniton, We use it as a face feature extractor. It is a good base model.
- We performed pre-processing steps in order to get multiple variations and pose variations.

D. *Network and Training*

In view of limiting the risk of the main challenge called overfitting, we used dropout learning with a ratio of 0.5 (50% chance of setting a neurons output value to zero) and Data Augmentation mentioned above along with cropping the image of 256 x 256 into 227 x 227. So that we can achieve multiple crop variations and mirror images in different variations. Training itself is performed using Adam Optimizer with image batch size of 64 images. The initial learning rate is 0.005.

E. *Usage of Activation Functions*

We have tried with different activation functions to see the efficiency which is higher with. The activation functions we have used are ReLU(Rectified Linear Unit) and SeLU(Scaled Exponential Linear Unit). RELU seems to be doing a much better job than SELU. We discovered ReLU is more converging than SeLU. SeLU takes more computation time than ReLU. To be genuine, it is possible that SELU is good in some configurations. Replacement of ReLU with SeLu does not show higher accuracy in models. But to our model, we have gained 2% higher accuracy than ReLU. So, to propaganda our higher accuracy, we used SeLU.

Framework	Hardware	Time for training
Tensorflow	GPU	10 hours

Table IV
TIME TAKEN FOR TRAINING

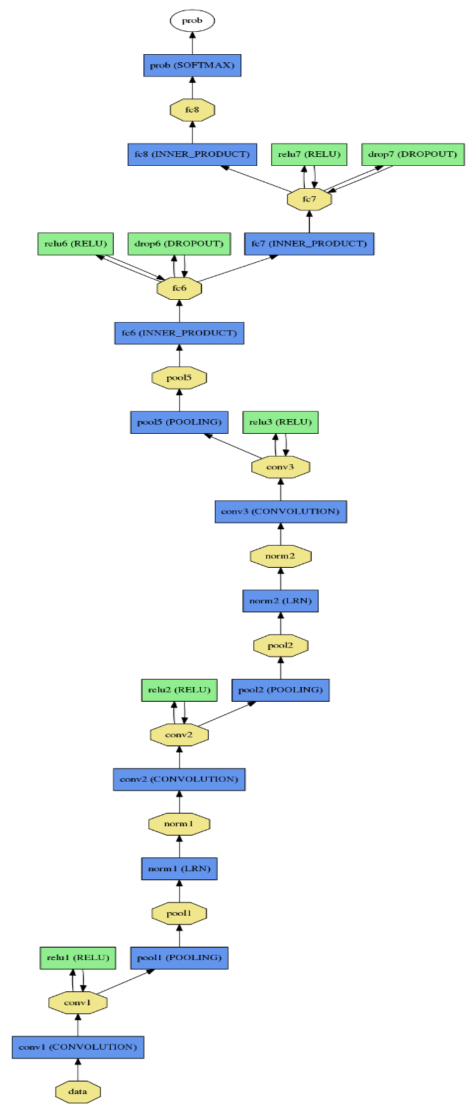

Figure 4. Detailed View of CNN Architecture

F. *Experimentation*

Our method is implemented using the Tensorflow framework [26]. Training was performed on an GPU machine. We have splitted our problem into three instances like gender data, age data and age & gender data. Training each network required about 10 hrs. Prediction running times can conceivably be substantially improved by running the network on image batches.

G. Prediction

We experimented with two methods of working with the neural network in order to produce gender and age classifications.

- Center Crop: Since Images are in 256 x 256 size, we have cropped into 227 x 227 size and fed it as input.
- Over-sampling: We have extracted four from the corners of the 256 X 256 face image, five 227 X 227 pixel crop regions and an additional crop region from the center of the face. The network is presented with all five images, along with their horizontal reflections. Its utmost final prediction is going to be the mean prediction value across all those extracted variations.

V. CONCLUSION AND RESULTS

We have mentioned the total number of images in the adience benchmark data set. We have also mentioned the table results regarding the accuracy of gender and age in table 2 and table 3 respectively. We have registered the highest accuracy of 91% in gender and 55% in age using VGG pretrained weights and SeLU Activation function.

VI. ACKNOWLEDGEMENTS

We implemented the paper Age and Gender Classification using Convolutional Neural Networks by Gil Levi and Tal Hassner. We acquired higher efficiency than state-of-the-art model in gender and small variation in age.
We have learned the difference in using ReLU and SeLU by working on it. Using pre-trained weights has really helped alot in reaching to the efficient output.

VII. MILESTONES

- Building a CNN model for classification of gender and age with higher accuracy than state-of-the-art models.
- Usage of SeLU and ReLU to understand the working of activation functions.
- Usage of VGG pre-trained weights instead of random initialization.

VIII. CHALLENEGES FACED

- Data-Preprocessing is long winded task.
- Computational Time is higher for Training.
- Many Images dont have gender and age labels.
- Some Images are not labeled as class, instead given as integer.
- Age Detection is not easy considering many factors. So We have been classifying the image.

REFERENCES

[1] Age and Gender Classification using Convolutional Neural Networks, Gil Levi and Tal Hassner, Department of Mathematics and Computer Science,The Open University of Israel,gil.levi100@gmail.com, hassner@openu.ac.il

[2] Understanding and Comparing Deep Neural Networks for Age and Gender Classification Sebastian Lapuschkin Fraunhofer Heinrich Hertz Institute Alexander Binder Singapore University of Technology and Design Berlin Institute of Technology Wojciech Samek Fraunhofer Heinrich Hertz Institute

[3] W.-L. Chao, J.-Z. Liu, and J.-J. Ding. Facial age estimation based on label-sensitive learning and age-oriented regression. Pattern Recognition, 46(3):628641, 2013.

[4] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. A study on automatic age estimation using a large database. In Proc. Int. Conf. Comput. Vision, pages 19861991. IEEE, 2009.

[5] A. Dehghan, E. G. Ortiz, G. Shu, and S. Z. Masood. Dager: Deep age, gender and emotion recognition using convolutional neural network. arXiv:1702.04280, 2017.

[6] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. IEEE Transactions on Information Forensics and Security, 9(12):21702179, 2014.

[7] F. Arbabzadeh, G. Montavon, K.-R. Muller, and W. Samek. Identifying individual facial expressions by deconstructing a neural network. In Pattern Recognition - 38th German Conference, GCPR 2016, volume 9796 of LNCS, pages 344354. Springer, 2016.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Neural Inform. Process. Syst., pages 10971105, 2012.