Prototype for Gender and Age Prediction using PyTorch

Sonal S R
Information Science and Engineering
R V College of Engineering
Bengaluru,India

Sushma G
Information Science and Engineering
R V College of Engineering
Bengaluru,India

Dr G S Mamatha
Information Science and Engineering
R V College of Engineering
Bengaluru, India

Prof Priya D
Information Science and Engineering
R V College of Engineering
Bengaluru, India

***Abstract* -- Gender prediction and Age estimation from face images has become useful in an extremely large number of applications. Since the advancements in social media, this topic has attracted various researchers and engineers. Even though there are many existing methods, the accuracy of these methods is still low on real world face images. This paper summarises the various methods of gender prediction and age estimation using deep learning. Various CNN(Convolutional Neural Network) architectures, trained and tested on various datasets are considered and their accuracies are compared.**

***Keywords*— CNN, PyTorch, LBP, FPLBP**

I. INTRODUCTION

Age and gender are part of the identity of a person. It is one of the key demographic characteristics to identify a person. Each person's face is different from another. Each individual has different facial features that are unique. Many languages in the world use separate grammar rules and vocabularies to address people based on their age and gender. Social interactions and communications become easier when we get to know the age and gender of the person we are communicating with. For example different types of reference such as he, she, him, her, salutations such as Mr, Mrs, Miss based on gender and similarly referring to a person as sir/madam based on age will all be relevant when we know the age and gender of the person we are referring to. There are many methods which are being used currently to determine age and gender based on facial images of people. This paper summarises some of those methods which are based on deep learning.

II. LITERATURE SURVEY

Adience dataset is used for the analysis. It gives the detailed architecture of the network two fully connected layers along with 3 convolutional layers and The model is trained from scratch by Initializing the weights from zero mean Gaussian and the standard deviation of 0.01. The risk of overfitting is avoided by using dropout layers where the output values of neurons is set to zero and by the process of data augmentation by cropping the images to be used in the forward -backward propagation process. It introduces two methods called center crop(cropping the centered face image) and Over sampling(extracting five cropped portions,4 from the corners of the image and one in the center) and the final prediction value is the average value of 5 predicted values. Two Important conclusions can be that CNN is able to have a better performance in prediction even with less number of unconstrained images which can definitely be improved with large dataset [1].

This paper projects a CNN model with 2 steps used for the estimation of age and gender of unconstrained facial images. It introduces pre-processing techniques like detection, alignment as frontal image in depth to handle the noise in unconstrained images such as difference in pose, resolution, background. Three datasets IMDB-Wiki, Morph-II and Adience are being used. The model proposed consists of 3 steps: (i) Preprocessing of images (ii) Extraction of features and learning/training of model (iii) Age and Gender Estimation. CNN model defined consists of 4 convolutional and 2 fully connected layers. The model is evaluated using exact and one off accuracy over the test dataset. The model is able to deal with images in the real world because of its robust preprocessing algorithm that handles variability [2].

This is a hybrid model which uses CNN and ELM(Extreme Learning Machine) and eases the process of age and gender prediction. CNN is used to pull out the features and ELM model to classify the gender and age. Adience benchmark dataset and Morph II dataset is used to analyze the model. The model is presented with the design of neural network layers, parameters and the importance of backpropagation over different iterations of training. It is made up of : Convolutional Layer, Max Pooling Layer, Contrast Normalization Layer,ELM Classification Layer. The paper also talks about the possible reasons of age or gender misclassification such as unconstrained images inkling blurred, heavy makeup and low resolution images. The risk of overfitting can be avoided by extracting a cropped face image from the given input, no tuning of weights of biases and other dropout methods. Hybrid models prove to have better accuracy results and also accelerate the process [3].

Seeks to improve the existing approaches for age and gender prediction which are based on CNN. Mainly the four important factors of CNN training are focussed. They are (i) the depth of the CNN model, (ii)the loss function and age encoding, (iii) the necessity to pretrain the network, and (iv) the strategy used for training which could be mono-task or multi-task. The IMDB-Wiki_cleaned dataset is used for training and private dataset for gender and age is used for testing which is a balanced dataset of both men and women and of all age groups. Some of their important conclusions include (i) For Age estimation CNN training, Label distribution age encoding (LDAE) is extra powerful. (ii) Gender recognition is more simpler than Age estimation (iii) Pre Training is crucial for face recognition for age and gender CNNs. (iv) When a CNN is trained only from the beginning, age estimation and gender recognition multi task training is useful [4].

It proposes the multistage process with two stages, first stage being the marking of each pixel of the image as person or non person, and using only the pixel named person for further processing. The second stage makes use of the model to study the global, local information and their interaction for the final prediction. The model is tested on three datasets FG-Net, Adience and CACD. Also proposes seven contrasting tasks that might have achieved good results in age estimation such as Deep Convolutional Neural Network, Deep Expectation, Hierarchical Age Estimation, Multiagent, Metric Regression with CNN, Ordinal Regression with CNN, Ranking CNN. But the major drawback is that the images in the dataset contain one and only one person which is not ideal in the real world situation where there may be no person at all or many people in the same image. And it talks about the interclass and intraclass differences which poses difficulty in the gender and age estimation process[5].

It majorly focuses on developing a robust system for age and gender detection with the constrained(i.e images in real time conditions) computing power in mobile applications. A variant of CNN called Lightweight multi task CNN is proposed. The Adience Dataset is made use of for the process which also has blurred, low resolution, occlusions, out of pose images and thus sets a wild benchmark for joint gender and age estimation when compared to MORPH II where the pictures are taken in a controlled environment. Hard Parameter sharing and Soft parameter sharing are often used as learning techniques in multi tasking. Soft Parameter sharing is when the tasks defined have their own deep neural network and Hard Parameter sharing when there is a single deep neural network model with several task specific output layers. It proposes the use of MobileNet whose architecture is built around the conversion of general convolution networks into depthwise and point convolutions. Hence the model LMT CNN(Lightweight Multitasking CNN) tries to perform multiple tasks at the same time while also maintaining the accuracy of the model[6].

It is majorly based on previous work done in the paper " Age and gender classification using convolutional neural networks".It uses Adience face dataset to train and test the proposed project, which contains 26,580 photos totally ,out of which 2,284 are unique faces. Each of those images has details of the person in that image(age and gender). It uses different classifiers for different genders. A face image of size 256 x 256 pixels is provided as the input to the algorithm, and it is cropped to 227 x 227 pixels, and is then fed into one of the classifiers. The age of the face is obtained in a range. There are 8 ranges for age in the dataset and hence the classifier gives a number from 0 to 7 representing the age range. The gender classifier gives either 0(indicating male or 1(indicating female) . A chained gender-age network is developed that would first classify an image as male or female(gender classification), and then based on the gender classification, the image is fed into an age classifier trained only on men or only on women. With this they have obtained highest accuracy of 92% for age range 15-20 years and lowest accuracy of 27% for age range 0-2 years [7].

The method proposed in this paper has three deep neural networks where residual learning methods is adopted. The training of the model is performed using the IMDB-WIKI dataset. Images from the internet are also collected training as the dataset has less number of images for ages less than 20. There is one gender prediction network and two networks designed for age estimation of male and female respectively in the proposed model.These models have residual connections between them for continuous learning to improve performance. The result of the gender prediction network is given as input to the age prediction network. The final age result is defined as the weighted average of two age prediction networks. This also shows that the regression model is better than the classification model for real age estimation[8].

Takes distinctive face images taken from smart phones and mobiles which are uploaded online to image repositories which have labels for age and gender. Dropout-SVM approach is made use of, to avoid overfitting for face attribute estimation. C++ and Python are combined for the system pipeline. For LBP and FPLBP descriptors, their own Python implementations are produced. OpenCV routines are made use of, for face detection for the system. The Gallagher benchmark and the Adience dataset is used for testing and greater accuracy for both age and gender is obtained with Gallaghar benchmark.[9]

Describes a system for emotion recognition in addition to age and gender recognition. Provides a large number of images data for age, gender and emotion recognition which is used to train the proposed system as well. Pre-processing is done on each image before they are fed to Deep neural networks. First step is face detection, followed by facial landmark detection and alignment. As the proposed network is for a single face only, the most centered face is chosen if there are more than one detected faces in the input image. It is shown that their new proposed deep architecture including their large number of internally collected dataset is better than other algorithms which are present with several benchmarks.[10]

Speaks about Automatic Age Estimation by facial features using different convolutional neural network models and gives their performance comparison with respect to traditional models with hand crafted features. The dataset used are MORPH, FACES, FG-NET,PubFig and Casia-Web. The best age predictor is definitely assessed under blurred,noisy ,tilted,low contrast conditions.This also discusses the way of optimizing the number of neural network layers required to fine tune the task of age prediction. Among the models investigated, XCeption network proved to be the best followed by InceptionV3, ResNet, VCG19, VCG16 respectively .The models were also tested for robustness by adding the gaussian noise.The effect of different gender, ethnicity, expression is studied on the task of age prediction. It was found that the process of age estimation is not affected much by gender , whereas it is highly affected with the change in ethnicity and expression[11].

Focuses on optimizing the deep learning neural networks such that they can work under low computational resources and made portable for widespread use. Five datasets used are VCGFACE, Adience, MIVIA-FACE, LFW, IMDB-WIKI dataset. The model is investigated to check the effect of decrease in the number of layers on performance under different configuration settings of input size and width multiplier.
XCeption architecture is majorly explored in the paper which is built using depthwise convolution over Inception Architecture.Later SqueezeNet and ShuffleNet networks are checked for their speed optimization along with their performance. Problems generated by detection errors are mentioned which can be a major source of study. Various other predictions of age, emotion, ethnicity etc can be considered in the future work[12].

Builds a CNN model using unconstrained facial images. The face detection task can be used to develop a model for age estimation with improved performance. Also various steps are taken to avoid overfitting. The model is trained using 2D images. The face recognition VGG model is trained using "Labelled Faces in the Wild(LFW) "[16] and "Face Recognition in Unconstrained videos with matched background similarity" [17]. Each CNN layer also has Rectification layers, normalisation layer and max pooling layer.
The input image is cropped to 224 x 224 pixel image. Learning rate is set to 0.1 and regular changes are made to it depending on the results obtained during the iterations of training. Distribution of date in different age and gender labels is shown for uniformity. Evaluation metrics are given in the form confusion matrix and accuracy[13].

Focuses on applying a number of post processing methods to the deep learning pretrained models to increase the performance. Only some features are selected using dimensionality reduction method, which is then fed to the Feed Forward Neural Network. for age estimation. WIKI and AmI-Face Dataset is used for training. Adience public dataset is chosen as test data. Mean absolute error(MAE) is chosen as the evaluation metric. Experiments are performed to analyse the model in less constrained conditions and comparison is made to various other related works[14]

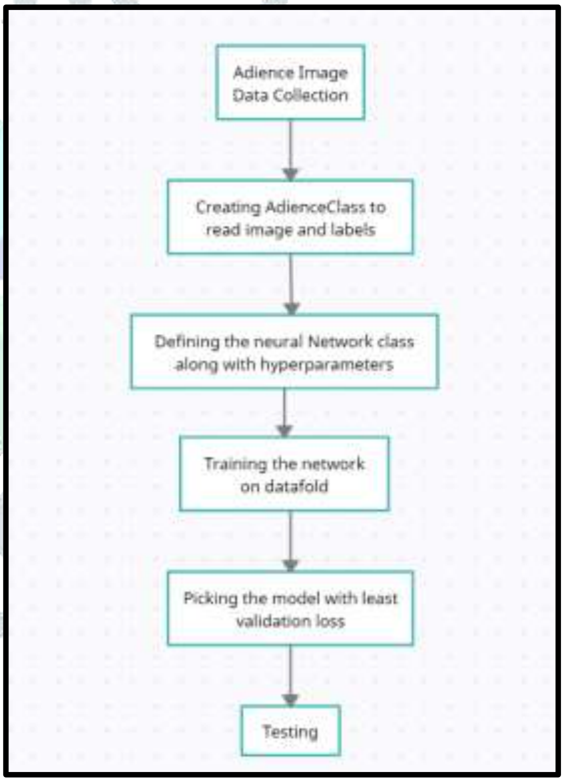
Propose a gender recognition system using VGGNet architecture based on Deep Convolutional Neural Network to get better accuracy and which improves the method that was formerly used. The training on the proposed VGGNet - face network is done such that it can perceive more than 10000 identities of celebrities. The images are changed to 256 x 256 size and are cropped to 227 x 227 size before feeding to the network. Pytorch, OpenCV, tensorflow with python language are used for implementing the code along with Graphical Processing unit(GPU) which has given great accuracy. This is the first research for gender prediction to have used VGGNet with Celebrity faces dataset to obtain high accuracy as high as 95%. [15]

Comparison Table of Various different Deep Learning Methods

Paper	**Method**	**Dataset**	**Accuracy for Gender(%)**	**Accuracy for Age-Exact (%)**	**Accuracy for Age-One-off (%)**
[1]	Deep-convolutional neural networks	Adience	86.8 ± 1.4	50.7 ± 5.1	84.7 ± 2.2
[2]	CNN	IMDB-Wiki Morph II Adience	96.2	83.1	93.8
[3]	CNN-ELM	Adience, Morph-II	88.2 ± 1.7	52.3± 5.7	-
[4]	CNN	IMDB-Wiki_cleaned for training, private balanced dataset for testing	98.7	**Age MAE=4.26	-
[5]	Deep neural networks	FG-Net, Adience, CACD	On FG-Net=98.80% On Adience=93.52% On CACD=95.01	**Age relative difference On FG-Net=8.49 On Adience =6.61 On CACD=16.23	-
[6]	Light-weight Multi-task CNN	Adience	85.16	44.26	-
[7]	CNN	Adience	80.8	54.5	84.1
[8]	Deep residual learning network	IMDB-WIKI + Images from internet	88.5± 1.4	52.2 ± 6.1	92.0± 2.4

[9]	dropout-SVM	Gallagher benchmark and Adience dataset	On Gallagher benchmark=88.6 ± 0.3 On Adience dataset=76.1 ± 0.9	On Gallagher benchmark=66.6 ± 0.7 On Adience dataset=45.1 ± 2.6	On Gallagher benchmark=95.3 ± 0.2 On Adience dataset=80.7 ± 1.1
[10]	CNN	Face and gender recognition: Own four million images, Age estimation: own dataset of ~ 600, 000 images	On Adience dataset: 91.00 %	On Group dataset: 70.5% On Adience dataset: 61.3 ± 3.7%	On Group dataset: 96.2 %
[11]	-	-	-	-	-
[12]	Deep Convolutional Neural Networks	VGG Face, LFW Dataset, MIVIA-Gender dataset, IMDB-WIKI, Adience dataset	-	Best obtained: 98.73%	-
[13]	Deep Convolutional Neural Network	Adience dataset	-	59.9%	90.57 %
[14]	Pre-trained Convolutional Neural Networks	WIKI AmI-Face for training, Adience for testing	-	58.49%	-
[15]	VGG Net architecture of Deep Convolutional Neural Network	Celebrity face data-set, LFW face dataset	95%	-	-

III. METHODOLOGY



The dataset chosen is Adience, which consists of 26,580 images of 2284 people. Each image in the dataset is labelled with 2 gender classes(0 and 1 for male and female) and 8 age classes(0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53, 60 and above)

PyTorch library is chosen to build the model as it provides strong GPU support to speed up the process of training on the large dataset, along with large number of builtins helpful to train a neural network such as the modification of dynamic graphs to reach the local minimum following the procedure of global minimum.

The dataset images are unzipped into a folder, where data folds consisting of a text file with image path and age and gender labels are maintained. A dataset class is defined to perform read operation on the image and the text data file.This class inherits from the Dataset class of PyTorch.

Various Hyperparameters required while building a neural network model are defined accordingly.
Batch size: The number of samples of data given to the network at one go while training, No of Epochs: The number of iterations on the original dataset required to fully train the model, Learning Rate: It decides the rate at which the weights in the model should be updated while training.

A simple Neural network class is defined by inheriting from the nn.module class provided by PyTorch. The model is developed by following the paper G Levi and Hassner, hence made up of three convolutional layers followed by two fully connected layers.

The model is trained on different data folds, and in each iteration the training and validation loss is calculated . Various models trained on different batches of training data are stored along with its loss metrics. The age and gender model with least validation loss is picked as the best model obtained from the process.

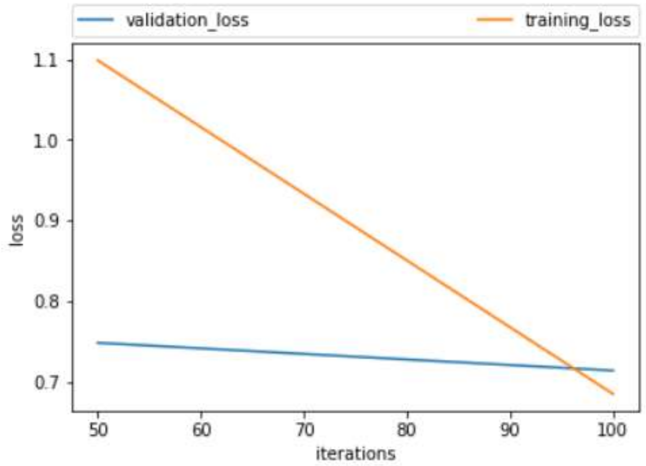
IV. RESULTS AND CONCLUSION

Performance metric is given in the form of accuracy which is the ratio of number of exact matches to the total number of samples tested. As age is difficult to predict compared to gender because of the large number of variations and classes, 'one off accuracy' is defined.

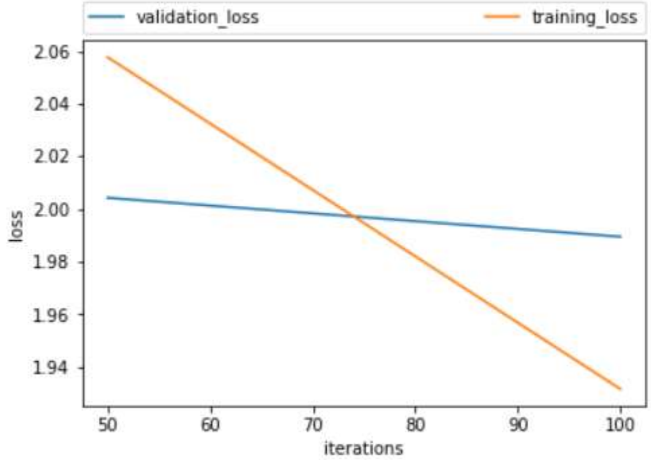
Results:

	Accuracy
Gender	Exact:82.37%
Age	Exact : 50.72% One-off : 81.99%

Validation loss and training loss for gender prediction.



Validation loss and training loss for age prediction.



It can be seen that over iterations, the model has lowered training and validation loss, hence the improved performance.

CONCLUSION

Various frameworks based on CNN were studied in this paper. The accuracy for age estimation is very low in almost all approaches compared to gender prediction, reasons being that there are only 2 classes for gender and more number of classes for age, gender specific facial features are more distinctive than age specific features, some approaches making use of predicted gender to predict age based on gender, some datasets lacking images for some age ranges, lacking unconstrained images as in real world situations.

ACKNOWLEDGEMENT

We would like to express our special thanks to our guides(Dr. G S Mamatha and Prof.Priya D) who gave us a lot of guidance while working on this paper, which also helped us to explore and learn a lot of new things.

REFERENCES

[1] G Levi and T Hassner, Age and Gender classification using convolutional neural networks , IEEE Conference on computer vision and pattern recognition workshops, 2015, pp

[2] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2015, pp. 34-42

[3] Agbo-Ajala, Olatunbosun & Viriri, Serestina. (2020). Deeply Learned Classifiers for Age and Gender Predictions of Unfiltered Faces. The Scientific World Journal. 2020. 1-12.

[4] Mingxing Duan, Kenli Li, Canqun Yang, Keqin Li, "A hybrid deep learning CNN–ELM for age and gender classification", Neurocomputing,Volume 275,2018, Pages 448-461,ISSN 0925-2312..

[5] Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, Jean-Luc Dugelay, “Effective training of convolutional neural networks for face-based gender and age prediction”, Pattern Recognition,Volume 72,2017,Pages 15-26,ISSN 0031-3203

[6] Jie Fang, Yuan Yuan, Xiaoqiang Lu, Yachuang Feng, "Multi-stage learning for gender and age prediction", Neurocomputing,2018

[7] Jia-Hong Lee, Yi-Ming Chan, Ting-Yen Chen, and Chu-Song Chen, "Joint Estimation of Age and Gender from Unconstrained Face Images using Lightweight Multi-task CNN for Mobile Applications",arXiv,2018

[8] Ari Ekmekji,"Convolutional Neural Networks for Age and Gender Classification",2016

[9] S. H. Lee, S. Hosseini, H. J. Kwon, J. Moon, H. I. Koo and N. I. Cho, "Age and gender estimation using deep residual learning network," 2018 International Workshop on Advanced Image Technology (IWAIT),2018

[10] E. Eidinger, R. Enbar and T. Hassner, "Age and Gender Estimation of Unfiltered Faces," in IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170-2179, Dec. 2014

[11] Dehghan, Afshin & Ortiz, Enrique & Shu, Guang & Masood, Zain. “DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network”.2017

[12] Alice Othmani, Abdul Rahman Taleb, Hazem Abdelkawy, Abdenour Hadid,"Age estimation from faces using deep learning: A comparative analysis",Computer Vision and Image Understanding,Volume 196,2020

[13] Greco, Antonio & Saggese, Alessia & Vento, Mario & Vigilante, Vincenzo,"A Convolutional Neural Network for Gender Recognition Optimizing the Accuracy/Speed Tradeoff ", IEEE ,2020

[14] Qawaqneh, Zakariya & Abumallouh, Arafat & Barkana, Buket, "Deep Convolutional Neural Network for Age Estimation based on VGG-Face Models", 2017

[15] A. Anand, R. D. Labati, A. Genovese, E. Muñoz, V. Piuri and F. Scotti, "Age estimation based on face images and pre-trained convolutional neural networks," 2017 IEEE Symposium Series on Computational Intelligence (SSCI), 2017, pp. 1-7.

[16] Amit Dhomne, Ranjit Kumar, Vijay Bhan, ”Gender Recognition Through Face Using Deep Learning”,Procedia Computer Science,Volume 132,2018,Pages 2-10

[17] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts,Amherst2007.

[18] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 529-534.

[19] Wei-Lun Chao, Jun-Zuo Liu, and Jian-Jiun Ding. “Facial age estimation based on label-sensitive learning and age-oriented regression”. Pattern Recognition. 2013.

[20] Zighem, Mohammed En-nadhir & Ouafi, Abdelkrim & Bekhouche, Salah Eddine & Benlamoudi, Azeddine & taleb-ahmed, Abdelmalik. . “Age Estimation Based on Color Facial Texture”. 2017

[21] Sung Eun Choi, Youn Joo Lee, Sung Joo Lee, Kang Ryoung Park, and Jaihie Kim. 2011. Age estimation using a hierarchical classifier based on global and local facial features.*Pattern Recognition* 44, 6 (June, 2011), 1262–1281.