[[1]](#footnote-2)

Age Prediction using Facial Images

**Abhi Shah\*, Rachit Shah\*\*, Prof. Pooja Shah\*\*\***

\*Computer Engineering Department, Institute of Technology Nirma University,

Ahmedabad 380009, India. Email: 14bce107@nirmauni.ac.in

\*\*Computer Engineering Department, Institute of Technology Nirma University,

Ahmedabad 380009, India. Email: 14bce110@nirmauni.ac.in

\*\*\*Computer Engineering Department, Institute of Technology Nirma University,

Ahmedabad 380009, India. Email: pooja.shah@nirmauni.ac.in

*Abstract*—In this paper, we tried to solve apparent age prediction problem from facial images using deep learning. We have used various deep learning architectures with transfer learning. We have tried to implement this problem on LAP (Looking At People) dataset and imdb-wiki dataset. We have taken this problem as classification problem of 101 classes for age 0 to 100. Our method first extracts faces from the given images and aligns them. Then we extracted the features of images using different architectures. We have tried various architectures such as VGG16, Xception, ResNet for feature extraction and then finetuned them on our dataset. We achieved 87% accuracy using ResNet architecture on LAP dataset.

*Index Terms*—Convolution Neural Network, Deep Learning

# INTRODUCTION

Age as perceived from looks is called apparent age. Unlike the biological age which is based on multiple factors of human body, apparent age is the age which we can estimate by looking at just the face. As such it is difficult for the apparent age results from our models to be the exact same as the biological age of the person. But we can train our model to extract features relevant and try to bring the error between the real biological age and the apparent age as low as possible.

With the current quick development of the intelligent applications there is a developing interest for automatic extraction of biometric data from confront pictures or, then again videos. Applications where age estimation can play

an imperative part include: (I) access control, e.g., confining

the entrance of minors to sensible items like liquor from vending machines or to occasions with grown-up content; (ii) human-PC connection (HCI), e.g., by a smart operator evaluating the age of an adjacent individual or, on the other hand a commercial board adjusting its offer for youthful, grown-up, or elderly individuals, as needs be; (iii) law authorization, e.g., automated examining of video records for suspects with an age estimation can help amid examinations; (iv) observation, e.g., automated identification of unattended kids at surprising hours and spots; (v) perceived age e.g., there is a vast enthusiasm of the general public in the apparent age, moreover pertinent while surveying plastic surgery, theater and motion picture role casting, or, on the other hand human resources help for open age part business.

One should take note of that the intelligent applications need to handle age estimation under unconstrained settings, that is, the face isn't adjusted and under known, unaltered, light and foundation conditions. Thusly, in the wild, a face needs initially to be identified, at that point adjusted, and, at long last, utilized as contribution for an age estimator. It is especially this setup we focus in our paper with our framework. Regardless of the current advance the treatment of faces in the wild and the exact prediction of age remains a challenging issue.

In the next section, we have mentioned state of the arts results achieved in this problem and their details. In Section III, proposed method is elaborated with diagrams. In Section IV, all the details about datasets which we have used for our models is mentioned. In Section V, various pretrained architectures are described which we have used. In the last section, results are mentioned with some figures reflecting it.

# State of the Art

DEX(Deep Expectation) method(6) has achieved state of the art results in this problem. They have used VGG16 (1) architecture which is pretrained on the ImageNet. They have collected their dataset from the imdb and Wikipedia and trained their model on that large dataset. They achieved 90% accuracy on LAP dataset.

# Proposed Method

We have utilized 2 datasets for training and testing our models. The first one is the ChaLearn LAP’s Apparent Age dataset (CVPR’16). It is an expanded version of a previous dataset (ICCV2015) and contains 8000 image seach with one face of a person. It is worth noting that the image labels for this dataset are all apparent ages and not the real ages. The apparent ages are determined by multiple individuals using a Facebook collaborative survey and Amazon Mechanical Turk.

The other dataset which we used is the IMDB-Wiki dataset from the DEX Paper by Rasmus Rothe et al.The dataset consists of approximately 500 thousand images (460723 from IMDband 62328 from Wikipedia). The labels of this dataset includes real ages of people along with their gender.

We utilize convolutional neural networks (CNNs) for our training. But before training on our dataset we actually use the pre-trained weights of models from architectures like VGG16, Xception and ResNet. This architectures are able to classify images but not suitable for age estmation. Thus we can utilize pretrained weights to detect faces well and then fine-tune our CNN model for age prediction.

Our model is a classification of ages into 101 labels (0,..,100). Though age estimated can be argued to be a regression problem, by classifying ages into these 101 ages and ignoring the rest, we can achieve better results. Though the limitation is that we can only predict the age of a person upto the age of 100.

Our method follows the methodology mentioned in figure. In following section whole procedure is explained.

## Face Detection and Alignment:

We have used opencv and dlib library to detect the faces and align them for the whole dataset. We have used haar cascade of opencv to detect the faces from the given image. If there are multiple faces in the image then we have extracted all faces from that image. If it fails to detect faces in the image then we have taken whole image as input.

To align the face, we used dlib library. First it predicts the shape using shape predictor which is then used for facealigner to align the faces.

We have resized these extracted faces to 224 \* 224 pixels to use these images as input in our convolutional neural network.

## Deep learning with CNNs

We have used different architectures to extract features from VGG16, Xception and ResNet. We have used the weights which are pretrained on ImageNet Dataset. Then this network is finetuned on imdb-wiki dataset. Different dense layers are added on top of that architecture and classification is performed in 101 classes corresponding to age 0 to 100.

## Expected Value

Figure 2 Pipeline of Model

Age estimation can be seen as a piece-wise regression or, then again, as a discrete classiﬁcation with various discrete names of significant worth. The greater the amount of classes is, the lesser the discretization error gets. For our problem, it is a one dimensional regression issue with the age being inspected from a value which is constant ([0,100]). We can upgrade the classiﬁcation definition for relapsing the age by vivaciously growing the amount of classes and thusly better approximating the mark and by uniting the neuron yields to recover the flag. Extending the amount of classes demands sufﬁcient preparing samples per each class and fabricates the shot of overﬁtting the preparation age scattering and of having classes not arranged truly as a result of a nonattendance of tests or unbalance. After various preparatory tests, we chose to work with 101 age classes. The softmax expected value, E, as takes after:

Where O = {0,1,2,3,…,100\_} = 101 output layer,

= softmax output probabilities belonging to O

= years discretely corresponding to each class

# Datasets

## ChaLearn LAP Dataset

The ChaLearn LAP dataset comprises of 8000 face images by and large age marked utilizing two applications - Facebook Collaborative and Amazon Mechanical Turk [5]. Each name is the averaged supposition of no less than 10 independent users. In this way, a standard deviation σ is additionally accommodated each age name. The LAP dataset is part into 2476 pictures for preparing, 1136 pictures for approval and 1087 pictures for testing. The age dissemination is the same in all the three arrangements of the LAP dataset. LAP covers the 20-40 years between time best, while for the [0,15] and [65,100] intervals it encounters unobtrusive number of tests each year.

### Evaluation Metrics

The outcomes are assessed either by utilizing the standard MAE measure or the Ɛ-error as characterized for the ChaLearn LAP challenge. The standard mean absolute error (MAE) is registered as the average of total errors between the evaluated age and the ground truth age. Note that the blunder does not get the defenselessness in the ground truth named age. The Ɛ - mistake covers such point of view. Ɛ-blunder. LAP dataset pictures are clarified with the normal and the standard deviation σ of the age votes tossed by various clients. The LAP challenge appraisal uses fitting an ordinary circulation with the mean µ and standard deviation σ of the votes for each picture:

Each prediction is evaluated using the following formula,

getting an error value between 0 (correct) and 1 (far from age). Not predicted images are evaluated with 1

## IMDB- Wiki Dataset

For good execution, ordinarily the expansive CNN designs require huge training datasets. Since the openly accessible face image datasets are regularly of little to medium size, once in a while surpassing a huge number of images, and frequently without age data, an extensive dataset of celebrities is gathered. For this reason, a rundown of the most well known 100,000 performers as recorded on the IMDB site and wikipedia are consequently crept from their proﬁles birth dates, images, and comments. The images without timestamp are removed (the date when the photograph was taken), likewise the images with numerous high scored confront discoveries. By expecting that the images with single countenances are probably going to demonstrate the on-screen character and that the time stamp and birth date are right, each such image is allocated the biological (genuine) age. Other than wrong time stamps, many images are stills from motion pictures, motion pictures that can have broadened creation times. Altogether there are 461,871 face images for big names from IMDB.

From Wikipedia, slithered all proﬁle images from pages of individuals and subsequent to ﬁltering them as indicated by similar criteria connected for the IMDB images, there were 62,359 images. Altogether there are 524,230 face images with slithered age data. As a portion of the images (particularly from IMDB) contain a few people, just utilize the photographs where the second most grounded confront location is beneath a limit. For the system to be similarly discriminative for all ages, even out the age distribution, i.e. arbitrarily disregard a portion of the images of the most continuous ages. This leaves 260,282 training images.

# Pre-training Architectures

## VGG16

This system is portrayed by its effortlessness, utilizing just 3×3 convolutional layers stacked over each other in expanding profundity[1]. Lessening volume estimate is dealt with by max pooling. Two completely associated layers, each with 4,096 nodes are then trailed by a softmax classifier.

A screenshot of a cell phone

Description generated with high confidence

## Xception

Xception was proposed by none other than François Chollet himself, the maker and maintainer of the Keras library.

Xception is an expansion of the Inception which replaces the standard Inception modules with depthwise distinct convolutions.

A screenshot of a cell phone

Description generated with very high confidence

Xception stands for "extreme inception". Rather like our past two architectures, it reframes the way we look at neural nets — conv nets specifically. Furthermore, as the name proposes, it takes the standards of Inception to an extreme.

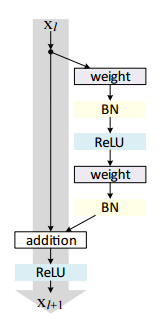
Here's the theory: "cross-channel correlations and spatial correlations are adequately decoupled that it is ideal not to delineate together."

## ResNet

Not at all like customary successive system designs, for example, AlexNet [2], OverFeat [3], and VGG [1], ResNet[4] is rather a type of "intriguing engineering" that depends on miniaturized scale design modules (likewise called "organize in-arrange structures").

The term smaller scale design alludes to the arrangement of "building squares" used to develop the system. A gathering of miniaturized scale engineering building hinders (alongside your standard CONV, POOL, and so forth layers) prompts the large scale design (i.e,. the end arrange itself).

In the first place presented by He et al. in their 2015 paper, Deep Residual Learning for Image Recognition, the ResNet engineering has turned into an original work, exhibiting that amazingly profound systems can be prepared utilizing standard SGD (and a sensible instatement work) using lingering modules:



# Results

We compared the results of all 3 models and found that ResNet gave the best results with a accuracy of 87%The following are some results taken as screenshots from the prediction python notebook.

V. CONCLUSION

Although many previous methods have tackled the problem of age classification of images, in this paper we establish a benchmark for the task based on state-of-the-art network architectures and show that chaining the prediction of age with can improve overall accuracy. If there had been more time, we would have dedicated more effort towards fine-tuning the parameters and the modified architectures we experimented with. Specifically, we would have liked to get the Adam learning algorithm in place with equal or improved performance to SGD, and I would have liked to replace the multiple fully connected layers at the end of the architecture with only one and instead shifted those parameters over to additional convolutional layers. By far the most difficult portion of this project was setting up the training infrastructure to properly divide the data into folds, train each classifier, cross-validate, and combine the resulting classifiers into a test-ready classifier. I foresee future directions building off of this work to include using age classification to aid face recognition, improve experiences with photos on social media, and much more. Finally I hope that additional training data will become available with time for the task of age and gender classification which will allow successful techniques from other types of classification with huge datasets to be applied to this area as well.

A person smiling for the camera

Description generated with very high confidenceA screen shot of a person

Description generated with very high confidenceA person smiling at the camera

Description generated with high confidenceA person wearing a suit and tie

Description generated with very high confidenceA person posing for the camera

Description generated with very high confidenceA group of people posing for the camera

Description generated with very high confidence

# References

1. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
2. Iandola, Forrest N., et al. "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size." *arXiv preprint arXiv:1602.07360* (2016).
3. Sermanet, Pierre, et al. "Overfeat: Integrated recognition, localization and detection using convolutional networks." arXiv preprint arXiv:1312.6229 (2013).
4. Szegedy, Christian, et al. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." AAAI. 2017.
5. Turk, Amazon Mechanical. "Amazon mechanical turk." Retrieved August 17 (2012): 2012.
6. Rothe, Rasmus, Radu Timofte, and Luc Van Gool. "Dex: Deep expectation of apparent age from a single image." Proceedings of the IEEE International Conference on Computer Vision Workshops. 2015.

1. [↑](#footnote-ref-2)