switches on printing of running heads

Proceedings of Seminar and Project

TITLE

SEMESTER

Oliver Wasenmüller and Prof. Didier Stricker
Department Augmented Vision
University of Kaiserslautern and DFKI GmbH

Introduction

The seminar and project TITLE (INF-XX-XX-S-X, INF-XX-XX-L-X) are continuative courses based on and applying the knowledge taught in the lectures 3D Computer Vision (INF-73-51-V-7) and Computer Vision: Object and People Tracking (INF-73-52-V-7). The goal of the project is to research, design, implement and evaluate algorithms and methods for tackling computer vision problems. The seminar is more theoretical. Its educational objective is to train the ability to become acquainted with a specific research topic, review scientific articles and give a comprehensive presentation supported by media.

In the XXX semester XXX, XXX projects addressing XXX were developed. Moreover, XXX seminar works addressed XXX. The results are documented in these proceedings.

Organisers and supervisors

The courses are organised by the Department Augmented Vision (`http://ags.cs.uni-kl.de`), more specifically by:

Oliver Wasenmüller
Prof. Dr. Didier Stricker

In the XXX semester XXX, the projects were supervised by the following department members:

NAME

MONTH YEAR

Apparent/Real Age Estimation using Deep Learning

Basavaraj Hampiholi[1] and Mohamed Selim[2]

[1] basavaraj.hampiholi@dfki.uni-kl.de
[2] mohamed.selim@dfki.de

Abstract. This project describes the estimation of real/apparent age estimation in still face images using deep convolutional neural networks(CNNs). In this project, a special CNN architecture VGG-16 is used for training. Although age is regression problem, we considered it for classification due to the availability of larger datasets like IMDB-WIKI, MORPH-II, LAP, FG-NET provided more samples per each class. Also, DEX[8] shows that, classification with expected value yield better results than regression. A pre-trained VGG-16 on Imagenet is used as a classifier to train the IMDB-WIKI dataset. The resulting model is fine-tuned using LAP dataset to find the apparent age. In this,first face detection is applied on test images and then CNN estimates the age from an ensemble of 20 networks. Our results are comparable to the State of the art(SoA) approach DEX. We can further improve the results by having more images and pre-trained model on facial data.

Keywords: DEX, IMDB-WIKI, LAP, CNN, VGG-16

1 Introduction

Face analysis is one of the most important and rapidly growing area of research in Computer Vision and Pattern Recognition community. Automatic age estimation from facial images is also one of the most challenging topics because of following reasons: aging process is uncontrollable, aging patterns are personalized as age depends upon food, race, gender etc and variation among faces of the same age. Age estimation has many applications like customer profiling, search optimization in large databases, assistance of bio-metrics systems,video surveillance, Demographic statistics collection.

Majority of researches focus on real age estimation [6], but with less significant results compared to CNNs. This field regained interest with the availability of large databases FG-NET,MORPH-NET. With larger number of samples, the discretization error between each class is low and hence people started estimating age with classification rather than regression. In contrast, the estimation of apparent age, that is the age perceived by others is also progressing rapidly since the introduction of the ChaLearn's LAP dataset for apparent age estimation.

Main motive of this study is to estimate real and apparent age using deep convolutional neural networks(CNNs). Although there were many researches in this field, introduction of the larger public datasets like IMDB-WIKI(real age) by [8] and LAP(apparent age) has really promoted research in this area. IMDB-WIKI is the largest available dataset for real age estimation and LAP dataset is the first State of the art(SoA) database for apparent age.

The rest of the paper is organized as follows: In section 2, the datasets like IMDB-WIKI and LAP are described. Section 3, describes different data augmentation techniques, Section 4, discusses about approach followed for implementation for real/apparent age estimation, mainly DEX. In section 5, evaluation protocol for age estimation, results and also the variance of each predicted age. Finally section 6, summarizes the conclusions and future work.

Convolutional Neural Networks Inspired by the animal visual cortex, convolutional neural networks have been impressive in solving computer vision and pattern recognition problems. Alex Krizhevsky et al.[5], trained a large deep convolutional neural network(CNN) to classify the 1.2

million high-resolution images in the ILSVRC-2012 and achieved a winning top error rate of 15.4% compared to next best result of 26.2%. This shows CNNs work better compared to any other SoA with image data. Since our problem is estimating the age by looking at face images, the CNNs can be used as solution.Followed by AlexNet, many CNN architectures were introduced by several researchers and major ofthem are GoogleNet[10], VGG Net[9] and Microsoft ResNet[4]. Among them, VGG-16 is a simple and deep architecture with significant less number of parameters.

2 Related Works

Automatic age estimation is an important and challenging problem in facial analysis for computer vision and pattern recognition community. Real age estimation has made significant progress with impressive accuracies after decades of research due to the availability of large public face databases. The pioneer research in this field is based on cranio-facial development theory and skin wrinkle analysis[6]. This is implemented to classify input images into three age-groups: babies, young adults, and senior adults. The implementation involves two phases. The primary phase to extract the features of the face like mouth,chin,nose. In the secondary phase a wrinkle geography map is used to guide the detection and measurement of wrinkles. The wrinkle index computed is sufficient to distinguish seniors from young adults and babies.

Another effort in this area is AGES[2]. The basic idea is to model the aging pattern, which is defined as the sequence of a particular individuals face images sorted in time order, by constructing a representative subspace. The proper aging pattern for a previously unseen face image is determined by the projection in the subspace that can reconstruct the face image with minimum reconstruction error, while the position of the face image in that aging pattern will then indicate its age.

One more unique SoA in image processing is Bio-inspired features to perform feature extraction and classification. Guo et. al[3] investigated the biologically inspired features (BIF) for human age estimation from faces. In this, a pyramid of Gabor filters are used at all positions of the input image for the S1 units to extract face features and form gabor jets. Then PCA is applied for dimensionality reduciton of the data. Finally the data is classified using support vector machine(SVM).

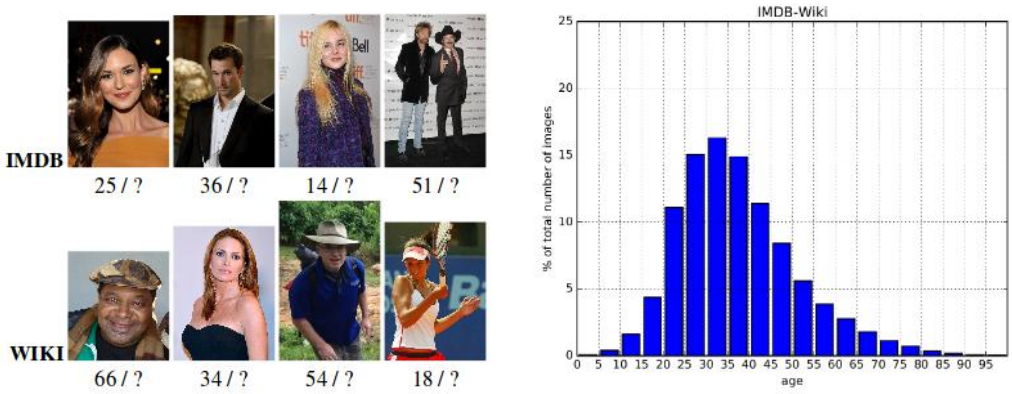
The foremost SoA for age estimation using CNN is by Dong Yi et.al,[11]. Convolutional neural network (CNN) is used to the age estimation problem, which leads to a fully learned end-to-end system can estimate age from image pixels directly.In this, 23 image scale patches are generated and 23 sub-networks are created to process them respectively and fuse their responses in the final full connected layer to estimate the age.

3 Dataset Preparation

There are many datasets available for real age estimation, like IMDB-WIKI, MORPH-II, FG-NET, Adience. Some of them contains the samples for age group instead of single age value, like Adience. The datasets for apparent age estimation are very less and the only available public dataset is LAP. The SoAs presented in this paper used IMDB-WIKI and LAP dataset for real and apparent age estimation respectively. Hence we discuss the process of dataset preparation for IMDB-WIKI and LAP in below sections.

3.1 Dataset for Real Age Estimation

This is largest dataset available for real age estimation problem. Most popular 100,000 actors were listed as per the IMDB ranking and automatically crawled from their profiles birth dates, images, and annotations. Difference between the date of birth and photo taken year is labeled as real age. Images without photo taken year and with multiple high scored face detections are removed, but the accuracy of the dataset is not guaranteed as many images are stills from movies and have wrong time stamps. In total, 461,871 face images of celebrities were obtained.

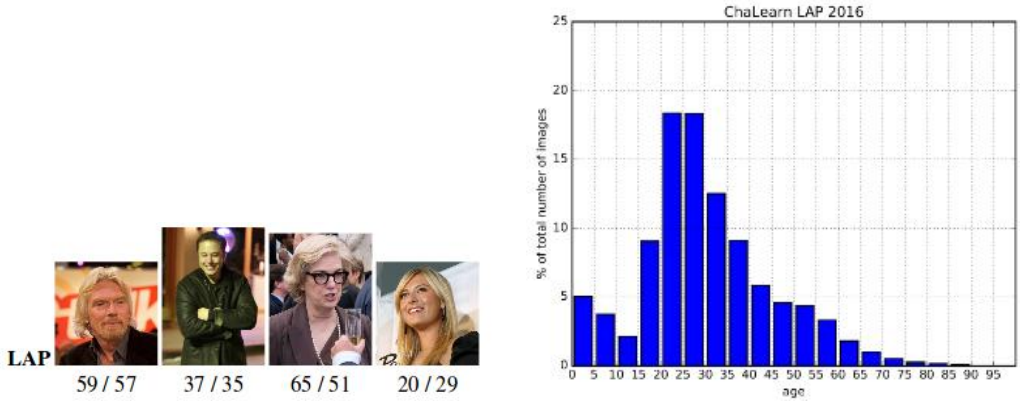


(a) (b)

Fig. 1. IMDB-WIKI dataset:(a) Sample images show real age labels and what would be the predicted age(?) [8]. (b)The distribution of samples per category(0-100)[1]

Wikipedia profile pictures were crawled and filtered as per the same criteria applied to IMBD images and collected 62,359 images. Finally, total of 524,230 images were taken from both the sources with age information. In case of images with several faces, images with second strongest face detection score below a threshold value were taken into consideration. Age distribution was equalized by randomly dropping some of the images of the most frequent ages.

We used IMDB-WIKI, Adience datasets for real age estimation. Fig.1.b shows the distribution of IMDB-WIKI dataset and it's clear that there are less number of samples in the category of children and senior citizens. So we extracted the images of children(age:0-20) and senior citizens(age:60-100) from Adience dataset and prepared combined dataset. After performing all the filtering, the total images for IMDB-WIKI were around 223K and Adience were around 8K. This combined dataset has around 231K images for with age information. Finally, dataset is split is done as follows: training- 167K images, validation- 41K images, test- 23K images. We used only training and validation images during training and test images were untouched.



(a) (b)

Fig. 2. LAP dataset:(a)Sample images show apparent age labels(mean) and the predicted age [8]. (b)The distribution of samples per category(0-100)[1]

3.2 Dataset for Apparent Age

The first state of the art database for apparent age estimation rather than real age estimation. To the date, ChaLearn LAP dataset V2 contains around 8000 images, which are labeled by multiple individuals(at least 10) using a collaborative Facebook implementation and Amazon Mechanical Turk. Hence the dataset is labeled with mean and standard deviation. The votes variance is used as a measure of the error for the predictions. The dataset is split into 4113 images for training, 1500 for validation and rest of the images for testing. The age distribution is the same in all the three sets of the LAP dataset.In Fig.2.(a) shows some sample images from LAP annotated with apparent age and Fig.2.(b) shows the dataset distribution. From the dataset distribution it can be seen that there are less number of samples for children and senior citizens compared to adults. We applied Mathias face detector on this dataset, extracted the face locations and cropped the images accordingly.

3.3 Data Augmentation

The dataset distribution for IMDB-WIKI shows less number of samples children and senior citizen category. Hence we performed 4 different data augmentation in these categories. First, rotation of images by -10 to 10, and scaling by 1.1 to 1.6 of the original size. Also performed random distortion on images with grid size 4x4. Finally we skew the images. Then the dataset is downsampled to maximum of 4000 images. The distribution of finally augmented train dataset for real age estimation is shown in fig.3.

For apparent age estimation, LAP train dataset is augmented 10 times. We applied rotation between -10 to +10, Scaling 1.1 to 1.6, Distortion with grid size 4x4, skew corner, shearing between 20 and -20 and Mirroring(flip left right).

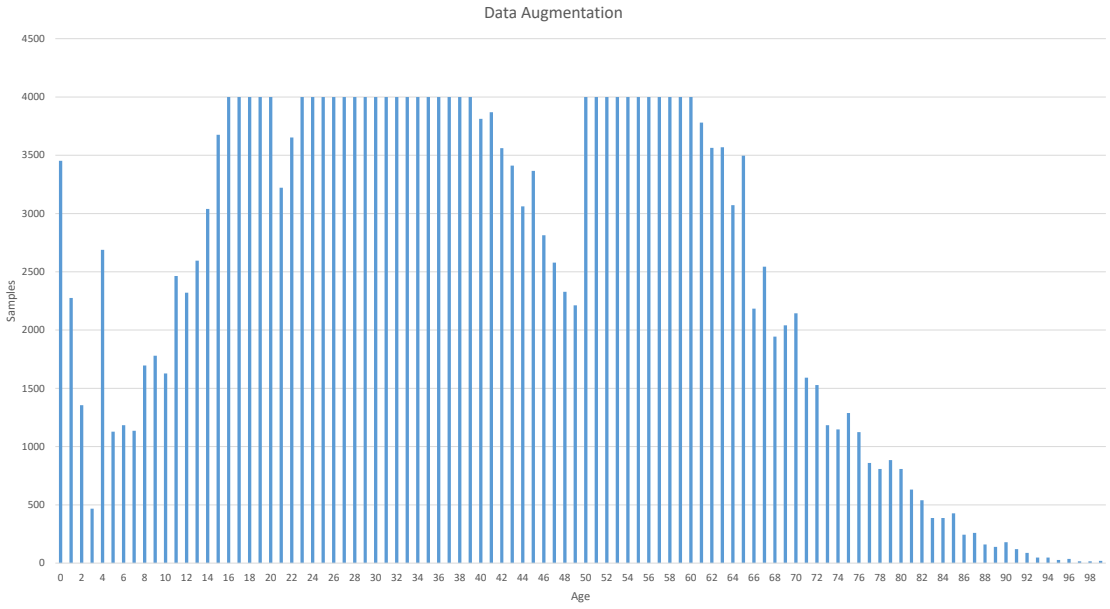


Fig. 3. IMDB-WIKI-ADIENCE dataset: The total dataset is split into train, validation and test data. We have augmented training data by rotation,scale,mirror,skew operations. Finally undersampled to 4000 images per class, but still it is imbalanced dataset

4 Approach

There are many SoA approaches for age estimation are available using CNNs itself. One among them is DEX: Deep expectation of apparent age from single still faces[8] which attained excellent results and won LAP-2015 challenge for apparent age estimation. The recent approach that has achieved better results is Children Specialized DEX by Gregory et.al[1]. This is based upon DEX itself but achieved better results than DEX. It is because of improvements in (a) the model used in children specialized DEX is pretrained on face dataset[7] and in DEX the model is pretrained on ImageNet (b) preparation of private dataset of children(5.7K) (c)Face alignment using facial landmark detection. Although children speacialized DEX has better results than the DEX, we chose to implement DEX itself as children dataset is publicly unavailable.

4.1 DEX: Deep Expectation of Apparent Age

Age estimation using CNNs follows the process of face detection, face alignment and training. Fig.4. shows the pipeline in detail and the same is explained below. Mathias et al., face detector is used to obtain the location of the face. Face detector is run on the original image as well as on all rotated versions between -60 to 60 degrees in 5 degree steps and also on -90,90 and 180 degrees for upside down photos. The face with strongest detection score is taken and rotated it accordingly to up-frontal position. In case face detector failed to find the faces, entire image is taken. Then face size is extended by adding 40 margin to left, right, above and below.

DEX uses VGG-16 CNN architecture for training. One of the most simple and deep CNN architecture VGG-16 is used for training. A pre-trained model of VGG-16 on ImageNet, is fine-tuned on IMDB-WIKI dataset. While training the classifier the output layer is changed to 101(0-100) output neurons and for regression only one neuron at the output layer. For real age estimation, the IMDB-WIKI dataset is divided into training, testing splits and pre-trained VGG-16 is trained on it. For apparent age estimation, the LAP dataset is divided into 20 splits with equal age distribution in each split. In each split, 90 of the data is used for training and 10 for testing. The data augmentation is performed on LAP dataset before training. Finally all the split data is trained on ensemble of 20 networks and the prediction is average of all.

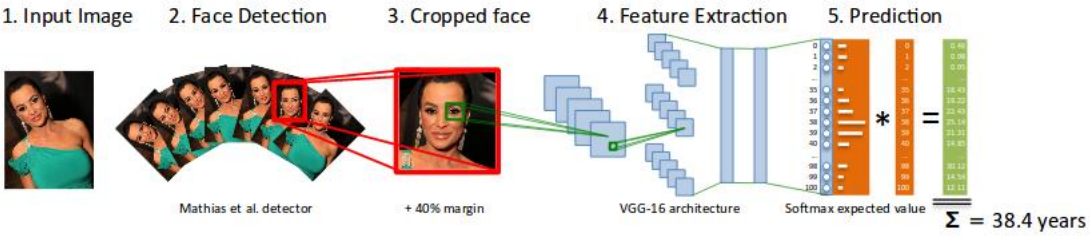


Fig. 4. Training pipeline of DEX method for age estimation[8]

Training Pipeline Pytorch provides the pre-trained model of VGG-16 on ImageNet. We used this model to fine-tune the dataset for real age estimation. While training the classifier we changed output layer to 100(0-99) output neurons. The training details are presented in Table.1. The learning rate is dropped by 0.1 for every 20 epochs. The model is trained with batch size of 32, stochastic gradient descent(SGD) optimizer and cross entropy loss.

The pipeline for apparent age estimation is shown in fig.5. For apparent age estimation, the LAP dataset augmented as discussed in section. Then we divided the dataset into 8 splits with equal age

CNN	Learning Rate	Epochs	Training Time
Real Age	0.0001	40	2days
Apparent Age	0.0001	35	12 hours

Table 1. Training Details: values for different hyper-parameters set during training

distribution in each split. We loaded the pre-trained model of real age dataset and trained all the 8 splits on ensemble of 8 models. We calculated the expected value for each model and the prediction is average of all. The model softmax expected value E is computed as: $E(O)=\sum_{i=0}^{100} y_i * o_i$, where O=0-99 is 100 dimensional output layer representing softmax probabilities o and y is the discrete years corresponding to each class i.

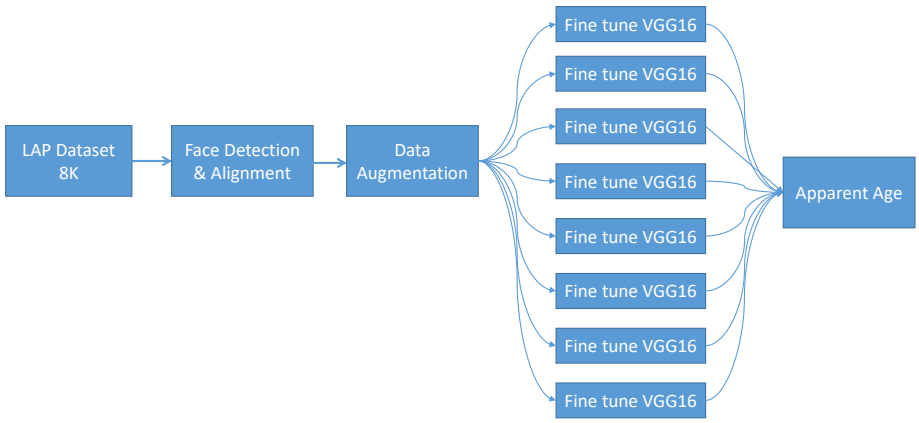


Fig. 5. Ensemble of 8 VGG-16 models pre-trained on IMDB-WIKi for apparent age estimation

4.2 Over-fitting

We faced the problem of over-fitting while training the model. The training accuracy was around 57% and validation accuracy was around 29% as shown in Fig.6. We applied L2 regularization techniques like weight decay and weighted cross entropy. Weight decay value is set to 0.00001. The weight vector for weighted cross entropy loss function is calculated as Vec(w)= $\frac{T}{N*10}$, Where T is total number of images and N is the number of images per class. We also used dropout layers to limit over-fitting. A custom dropout layer with probability of 0.7 is added to the vgg-16 network. After applying regularization techniques, the validation accuracy was increased around 6%.

5 Experiments and Results

The results are evaluated by using the standard mean absolute error (MAE) and Gaussian error(ϵ). MAE is computed as the average of absolute errors between the estimated age(ex) and the ground

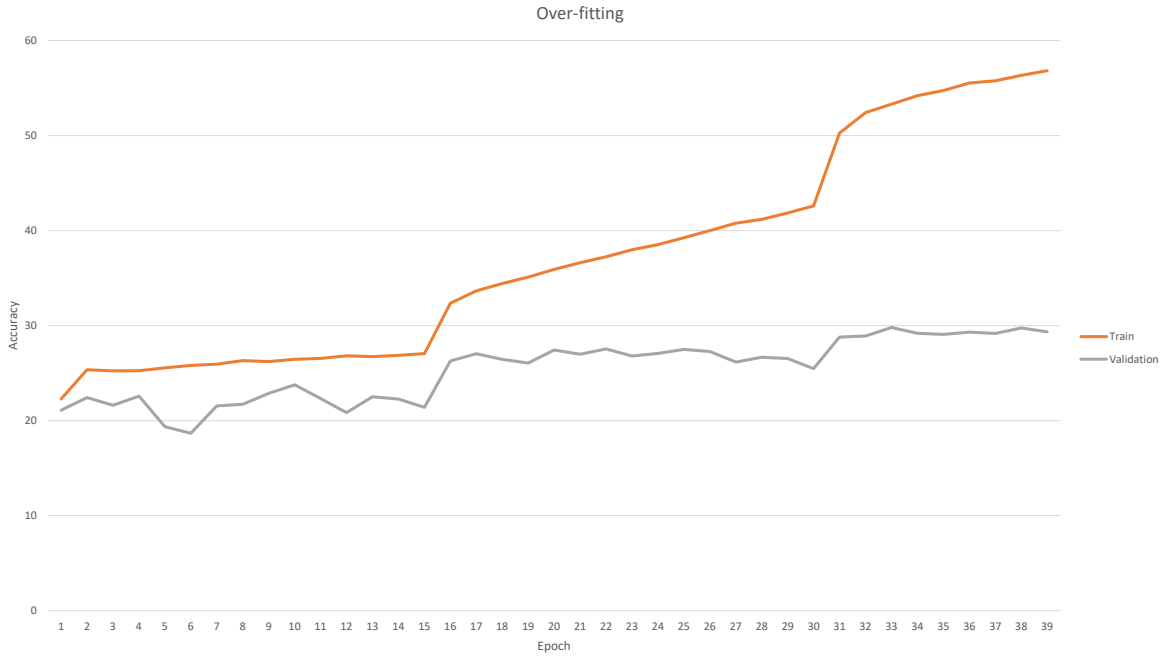


Fig. 6. Over-fitting problem: the training accuracy is around 57% and validation accuracy of 29%

truth ages (x). The Gaussian error fits the normal distribution with mean μ and standard deviation σ of the votes for each image. Usually, MAE is used to evaluate real age estimation and ϵ is used to evaluate apparent age estimation

$$MAE{=}\frac{1}{N}\sum_{i=1}^{N}|ex_i - x| \quad \text{and} \quad \epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

5.1 Results

*T*he network is experimented to train for classification, as this problem is considered as classification with calculation of expected value. Also, as per DEX[8]the softmax expected value on the network trained for classification worked better than regression. The test images of IMDB-WIKI and LAP are is then passed to the trained model on LAP dataset and the results are placed in Table.2, which reports the MAE for real age estimation and e-error apparent age estimation for DEX too.

SoA	MAE	ϵ-error
DEX	3.3	0.264
Our	4.0	0.384

Table 2. Performance on test set of IMDB-WIKI for real age estimation and on LAP test data for apparent age estimation for DEX and our model

It is evident from the Table.2. that our results did not meet DEX, but they are comparable. But we can further train the network by setting proper hyper-parameters will give better results. We have stopped training due to the time-constraints. If pretrained model on any facial dataset used rather than on ImageNet, we can further improve the results. Since the dataset is highly imabalanced, we can add more samples to make it balanced and train the model using balanced dataset.This will also increases the accuracy furthermore.

Also, Fig.7. shows the mean and deviation for each class of predicted age. It shows that, children and senior citizen categories have more deviation than adult ages. Hence, more samples in the children and senior citizen category will lead to better accuracy. One fundamental problem with treating age as classification is that, the features are not that much discriminatory as there is not much difference between the features of two any close ages like 20 and 21. This can be overcome by using more diverse images per class of many subjects

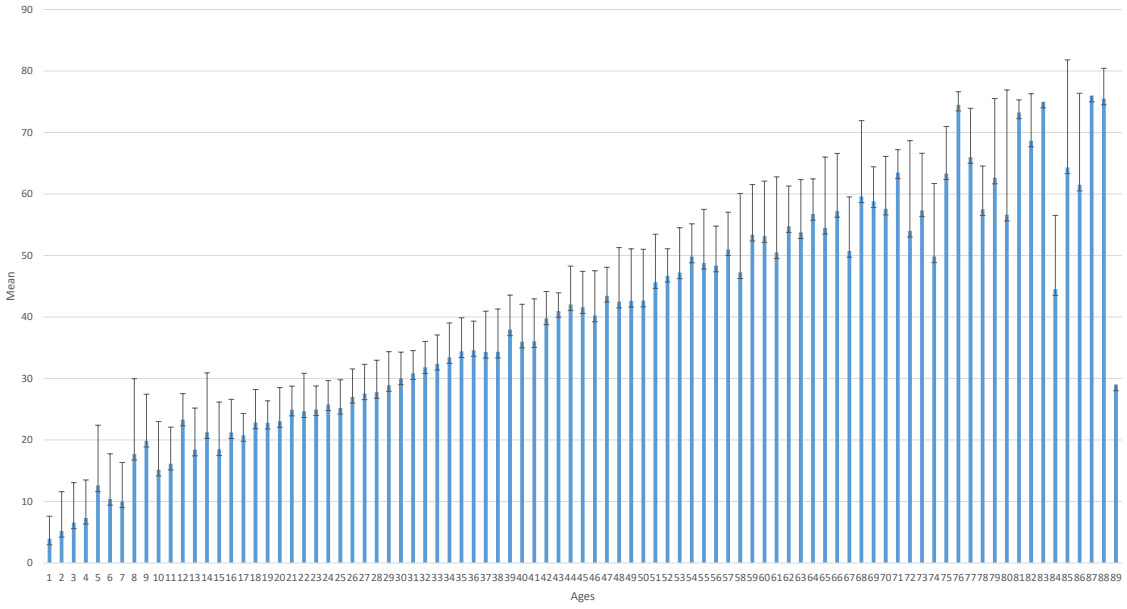


Fig. 7. Deviation for predicted ages of each class shows that there is less deviation for the category of adults compared to children and senior citizens

6 Conclusions and Future Work

In this work, the estimation of both real age and apparent age in still face images using CNNs is presented. The results of this project are not that of SoA, although they are comparable. The results can be improved by training more to set the hyper parameters properly. Also it is evident from children specialized DEX[1] that, a pretrained model on face images[**?**] helps to increase the accuracy than a model pretrained on ImageNet. So, if we use any pre-trained facial model to fine-tune IMDB-WIKI dataset, the accuracy can be improved. As Fig.6 shows the high deviation for

children and sinior citizens, it is better to have balanced dataset for all classes(0-99) as IMDB-WIKI is highly imbalanced and concentrates on adults. Hence, collecting and training more images in the category of children and senior citizens can further improve accuracy of the model.

References

1. Grigory Antipov, Moez Baccouche, Sid-Ahmed Berrani, and Jean-Luc Dugelay. *Apparent Age Estimation from Face Images Combining General and Children-Specialized Deep Learning Models*. Las Vegas, USA, 2016.
2. X. Geng, Z. H. Zhou, and K. Smith-Miles. *Automatic Age Estimation Based on Facial Aging Patterns*, volume 29. Dec 2007.
3. G. Guo, Guowang Mu, Y. Fu, and T. S. Huang. *Human age estimation using bio-inspired features*. June 2009.
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*, volume abs/1512.03385. 2015.
5. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks*. 2012.
6. Young Ho Kwon and N. da Vitoria Lobo. *Age classification from facial images*. Jun 1994.
7. O. M. Parkhi, A. Vedaldi, and A. Zisserman. *Deep Face Recognition*. 2015.
8. Rasmus Rothe, Radu Timofte, and Luc Van Gool. *DEX: Deep EXpectation of apparent age from a single image*. December 2015.
9. K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, volume abs/1409.1556. 2014.
10. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. *Going Deeper with Convolutions*. 2015.
11. Dong Yi, Zhen Lei, and Stan Z. Li. *Age Estimation by Multi-scale Convolutional Network*. Springer International Publishing, Cham, 2015.