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[6],MORPH-II,LAP,FG-NET provided more samples per each class. 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.

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 [4], 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 [6] 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.

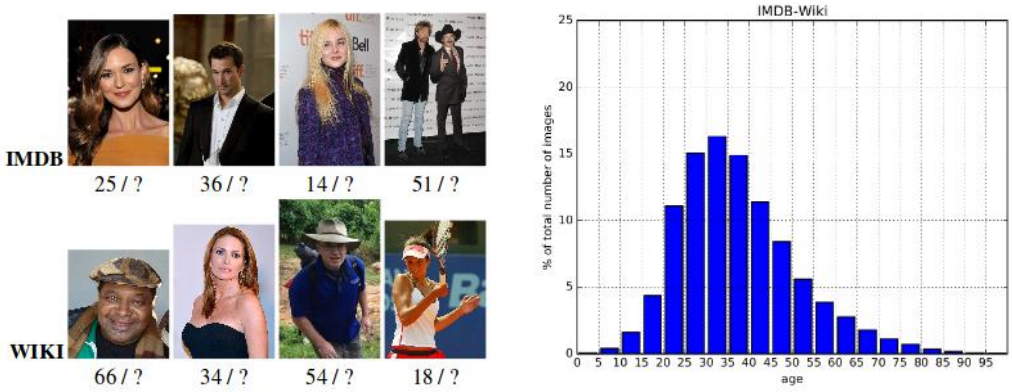
The rest of the paper is organized as follows: In section 2, the datasets like IMDB-WIKI and LAP are described. Section 3, describes different data augmentation techniques, Section 4, discusses about approach followed for implementation for real/apparent age estimation, mainly DEX. In section 5, evaluation 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.[3], trained a large deep convolutional neural network 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 of them are GoogleNet[8], VGG Net[7] and Microsoft ResNet[2]. Among them, VGG-16 is a simple and deep architecture with significant less number of parameters.

2 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.



(a) (b)

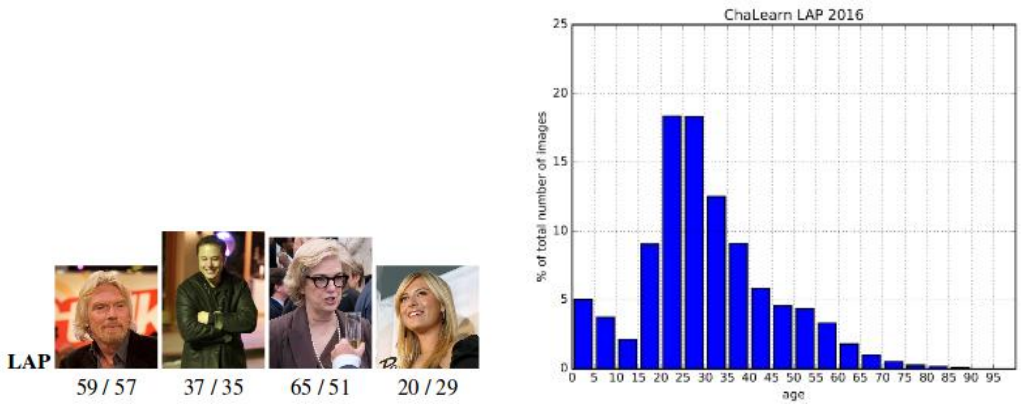
Fig. 1. IMDB-WIKI dataset with sample images and distribution [6][1]

2.1 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.

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 with sample images and distribution [6][1]

2.2 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 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).

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[6] which attained excellent results and won LAP-2015 challenge for apparent age estimation. So we chose to implement this method.

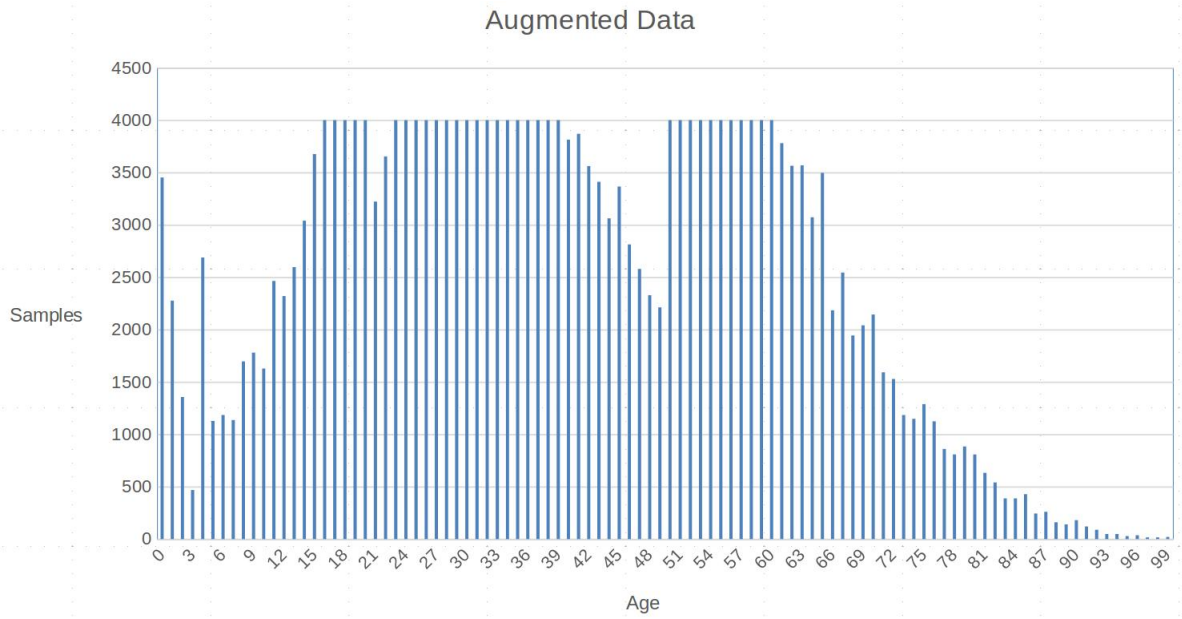


Fig. 3. Augmented training data of real age dataset

4.1 DEX: Deep Expectation of Apparent Age

Age estimation using CNNs follows the process of face detection, face alignment and training. DEX uses VGG-16 CNN architecture for training. Fig.3. shows the pipeline in detail and the same is explained below. The DEX[6] provides already detected faces for training.

Training Pipeline One of the most simple and deep CNN architecture VGG-16 is used for training. 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.

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

Table 1. Training Details

For apparent age estimation, the LAP dataset augmented as discussed in section. Then we divided the dataset into 8 splits with equal age 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.

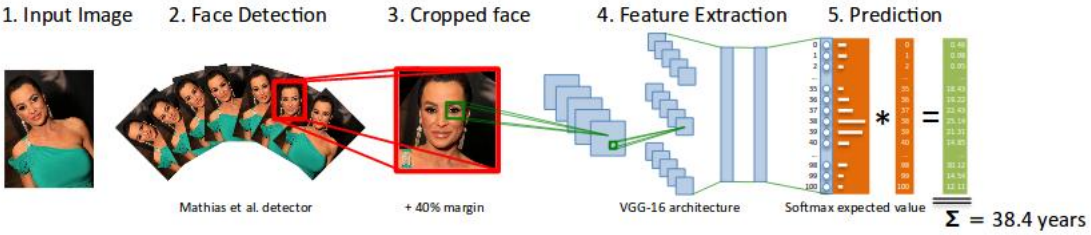


Fig. 4. Training pipeline of DEX method for apparent age estimation[6]

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 33% as shown in Fig.5. 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%.

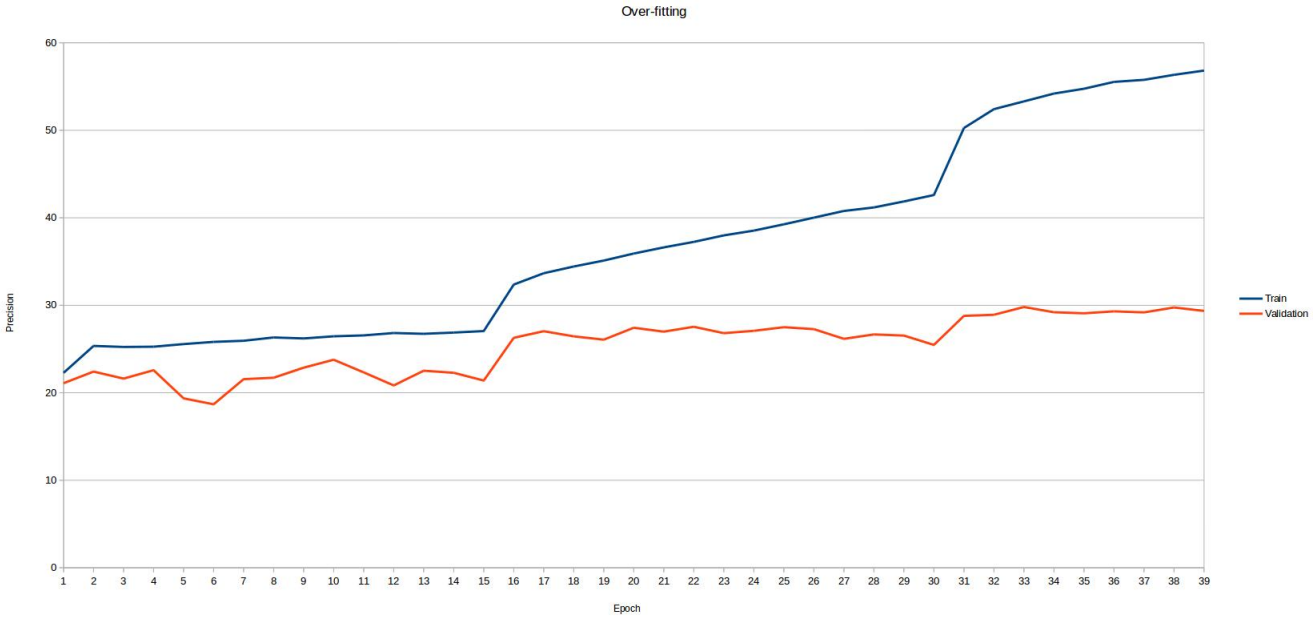


Fig. 5. Over-fitting problem

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 truth ages (x). The Gaussian error fits the normal distribution with mean μ and standard deviation σ of the votes for each image.

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

5.1 Results

DEX: The network is experimented to train for classification as this problem is considered as classification with calculation of expected value. The softmax expected value on the network trained for classification worked better than regression. Table.1 reports the MAE for real age estimation and e-error apparent age estimation.

CNN	Train Accuracy	Validation Accuracy	MAE	ϵ-error
Real Age	75%	32%	4.9	-
Apparent Age	65%	40%	-	0.38

Table 2. Results

Also, Fig.6 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.

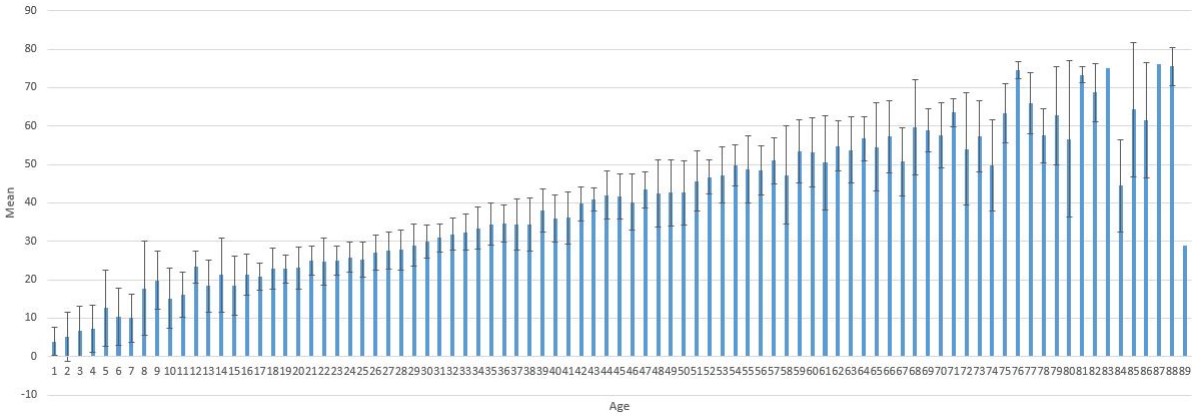


Fig. 6. Deviation for predicted ages of each class

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. Although the results of this project are not equal to the SoA, but they are very near to

the SoA results. The results can be improved by training more to set the hyper parameters properly. Also it is evident from children specialized DEX that, a pretrained model on face images[5] 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. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. *Deep Residual Learning for Image Recognition*, volume abs/1512.03385. 2015.
3. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. *Imagenet classification with deep convolutional neural networks*. 2012.
4. Young Ho Kwon and N. da Vitoria Lobo. *Age classification from facial images*. Jun 1994.
5. O. M. Parkhi, A. Vedaldi, and A. Zisserman. *Deep Face Recognition*. 2015.
6. Rasmus Rothe, Radu Timofte, and Luc Van Gool. *DEX: Deep EXpectation of apparent age from a single image*. December 2015.
7. K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition*, volume abs/1409.1556. 2014.
8. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. *Going Deeper with Convolutions*. 2015.