Free Patch-based Age Estimation from Facial Images by Convolutional Neutral Network

Yi Li
Supervised by Luuk Spreeuwers, Raymond Veldhuis and Joris van Hoof

Electrical Engineering, EEMCS
University of Twente, the Netherlands

February 4, 2018

Abstract

In biometric and computer vision studies, estimating age from facial images is an important task with practical value. However, due to the complexity of aging process and the lack of appropriate databases, it is hard to develop an automatic age estimation system based on facial images. Aiming to overcome these difficulties, this paper proposes a free patch-based method for age estimation. The method selects a set of skin patches extracted from facial images as the studied objects, instead of examining the holistic images. For investigating the biometric features behind those skin patches, the deep learning tool Convolutional Neural Network is deployed. Along with the implementation of several experiments on patches, the possibility of classifying people's age group by this proposed method is evaluated in this paper.

1 Introduction

Since a wide variety of biometric features have been encoded in the face, facial images are a common data source for many biometric systems, including age estimation. No biometric features are permanent over time and the appearance of the face will also change due to aging. Due to the fact that "the progress of human aging is uncontrollable, with many internal and external influence factors such as one's health state, life-cycle and extreme weather conditions," [1] estimating age based on two-dimensional facial images is still a demanding task.

Despite the difficulty mentioned above, estimating age from facial images has great practical values. This paper was initially inspired by the enforcement of the legal age limits for purchasing alcohol in the Netherlands. Legally, "alcohol can only be sold if the age of the customer has been verified over 18 years. The law also conveys that age verification is not allowed if someone is unmistakably old enough, therefore, retailers use a 'reference age' of 25 years."[2] In this case, an automatic system which is able to quickly determine if someone is older than 25 years would prevent a lot of inconvenience and improve the working efficiency.

In recent years, with the development of computer vision and artificial intelligence, there is an increasing trend in the application of deep-learning techniques such as Neural Networks in age estimation systems. This is of great potential value, but also raises new problems regarding data acquisition. The artificial neural network is a mathematical model for exploring complex relationships between inputs and outputs or finding patterns in data.[3] To achieve a fast and accurate performance, the Neural Network needs to learn a large amount of appropriate data; the new challenge is to develop the means to collect this data.

As for developing an age estimation system, the strict selection of facial images makes the appropriate data pretty limited. For example, facial images with mark-ups and big smiles should be avoided, because those non-aging factors might influence the accuracy of estimation. Additionally, the images in a dataset should be vary in resolutions and poses significantly, but that is fulfilled by most public databases of facial images, such as FG-Net and MORPH.

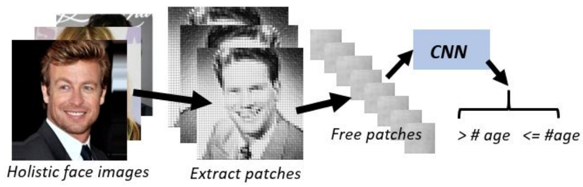

Figure 1: The basic operation of age estimation system using the proposed free patch-based method

In this paper, a free-patch based method is proposed for age estimation from facial images. The method extracts a set of small skin patches from facial images as the examined units in an attempt to solve the problem of the insufficient appropriate data. The patches extracted from each face are treated equally regardless of their original locations which is why they are called 'free'. The Convolutional Neural Network is employed to learn the characteristics behind those patches and classify them into the corresponding age groups. In this study, several experiments will be conducted to evaluate whether these small patches of skin extracted from facial images contain useful biometric features for age estimation or classification.

2 Realated work

Regarding age estimation from face images, a number of research and studies have been done. The aging process affects structure and appearance of a person in many ways. Narayanan Ramanathan et al.[4] summarized that in computer vision the facial aging problem can be categorized in Shape vs. Texture, Feature selection and Factors.

Craniofacial morphology is a science of shape of the face and skull. [1] Darcy Thompsons's study of morphogenesis (1917) made the first steps in associating the morphological changes with the growth in biological form[5]. From morphology theories, anthropometric model was developed to describe the person's head shape by mathematics. Farkas et al. [6] presented a systematic and comprehensive measurement of facial distance between various characteristic points. Those anthropometric distances can be used to characterize the face at different ages. Since the craniofacial growth is small during adulthood, the anthropometric model is not really effective for estimating exact age, but can separate the youth and adults. For example, based on the six ratios of anthropometric measurement, Kwon and Lobo (1999) classified the facial images of children and adults [7].

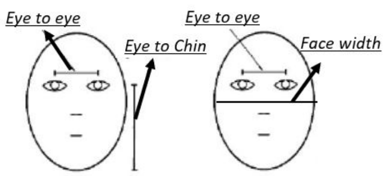

Figure 2: Some anthropometric ratios on faces

Rather than fully focus on the face shape changes, Cootes et al. [8] proposed Active Appearance Model (AAM) in 1998 as a statistical model to represent facial images. It works in a way that considering both the geometry of human face and its textures. The model labels the landmarks on face and encoded the facial structure, in this process, the Principle Component Analysis (PCA) techniques are applied. In 2002 The model has been expanded to facial aging suggesting an aging function defined by $age = f(b)$, to explain the variation in years. [9] Age is the age of a person in the picture, b is a vector containing 50 parameters learned from AAM, and f is an aging function. [1] Unlike anthropometric model, AAM is not oriented only to young people, but also deals with assessment of people of all ages.

Most previous age estimation algorithms are based on holistic images of face. However, holistic features are sensitive to illumination variations and image occlusions. The patch-based method can help to solve these issues, which has already been widely applied in face recognition field. Kim et al. [10] classified 2D facial image by small regions, noticing that this algorithm can be effectively decrease the influence of illumination variations. In 2011, Spreeuwers et al. [11] proposed a 3D face classifier based on the fusion of many dependent region classifiers for overlapping face regions. This approach provided fast and accurate recognition results with a better fit for biometric template protection. Compared with the general patch-base approaches inside coordinate frames, Lucey et al [12] stated that a free-patch representation of face can provide even more benefits, since the mismatch effects can be reduced further by synthesizing the statistical modal of free patches.

In computer vision studies, Convolutional Neural Network (CNN) is a widely used machine learning framework for classifying images. It can use relatively little pre-processing algorithm but are known to be robust against distortions. To evaluate the age estimation, mean absolute error (MAE) is a common measure, which describe the the accuracy of statistical predictions. Yang et al.[13] enforced a system of profile recognition on an advanced CNN. It achieved 4.88 of age estimation and 89.7% accuracy of gender recognition, which are competitive performances against the state-of-art algorithm on the same database. In 2015, Rothe et al [14] tackled the estimation of apparent age in face images with CNNs. The applied networks are pretrained on ImageNet for image classification. After trained by over 50,000 images crawled online, the method got 3.221 MAE as the best result.

3 Method

Inspired by the works mentioned above in section 2, a free patch-based system of estimating age from facial images will be proposed in this paper, which aims to effectively classify the face images of people who are younger or older 25 years old. The system operation consist of four main steps are stated as below:

- Determine and extract the appropriate patches from face images
- Establish the baseline experiment over the pixel variance of patches
- Train and test the Convolution Neural Network
- Fuse the patch scores from the CNN

Currently, there is no well-organized database of face images for age estimation, the quality of available face images are highly varied. As a result, preprocessing the images is an important but demanding task.

The free patch-based method, is expected to solve the problem of lacking appropriate face images, because each face can provide a number of patches for age estimation. Meanwhile, being robust against the different pose and illumination variations might be an additional advantage.

The setup of baseline experiment aims to evaluate the quality of the extracted patches in advance, and ensure the feasibility of further implement of CNN classification with same dataset.

The Convolution Neural Network is selected to be the tool of classifier, as a well-trained network can perform an effective classification over a set of patches. In final testing phases, the fusing algorithm of the patch scores are designed to improve the accuracy of the classifier.

As the classification in our experiments is always binary, the accuracies expressed in this paper is in percentage form, rather than Mean Average Error as mentioned in section2.Each step will be illustrated in depth in following sections.

3.1 Patch of Interest

3.1.1 Normalization of Face Images

As one knows, the patch is a two-dimensional square, but in reality, the face is three dimensional. The angle between face and camera will make the same patch area contain more skin area. Besides, some big facial poses might make one side of cheek too bright and the other covered by shade. As a result, the different poses in images can lead to different estimations even for the same face.

Although the patch-based method has ability to reduce the influence of different face poses, it is still necessary to normalize the face images before extracting the skin patches.

Figure 3: The original dataset with various face poses and locations

To solve these problem, one of the most important criteria of normalization is that the face in images should look into the camera as straight as possible. In addition, the face images would better have some common reference coordinates, for the convenience of next step, patch extraction.

To supply a large number of normalized face images, the Active Shape Models with Stasm was employed. Stasm is a C++ software library for finding features in face, it read the input face images and return the position of features on face[15]. In our case, the returned features are set to be eyes.

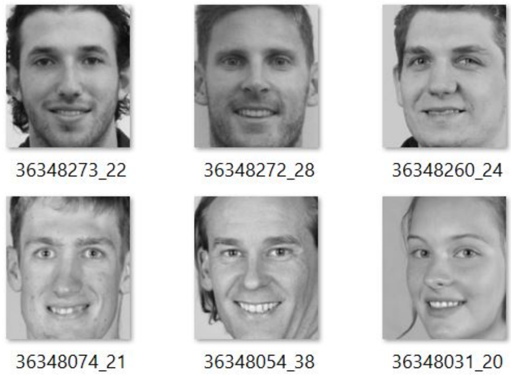

Figure 4: Some sample face images after the normalization with same eye coordinates

After the normalization, all images are cropped to focus more on face, and all those face images are registrated along the same eyes position. The photos with bad resolution and poses which cannot be normalized have been disposed automatically.

3.1.2 Locate and Extract the Patch

In order to extract the skin patches from certain locations on each face, the anthropometric model theory is applied, which has been mentioned in above sections.

Being benefit from the normalization step, the fixed locations of two eyes in images helped to define the whole face into a coordinate first. And then the distance between the eyes as well as other facial components can be easily measured. Moreover, according to biometric theories, the distance between two eyes of an adult is around $6cm$ in reality, which can leads to a scale ratio between the real face and the photographs. With the help of those distance and ratio, any points of the face can be easily located in the coordinate of the image.

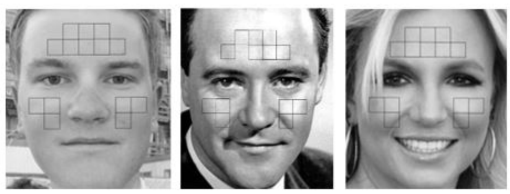

Figure 5: Faces image with the extracted patches

The size of the patches is an important parameter, which should be neither too small to lose some important features, nor too large that encode many noisy factors. In this experiment, the original face images come from the IMDB+Wiki database, which will be discussed more specifically in implement section. After the normalization, most of those face images are around 380*380 pixels, and there are 80 to 90 pixels between two eyes. The final defined patch size was chose to be 17*17 pixels, a middle size ensures that each patch contains sufficient biometric features to analyze and each face can provide a set of such patches.

3.1.3 Final Choice of Patches

In the model of our age-estimation system, there are 14 skin patches extracted from each face in the beginning. The first 8 patches are located on the forehead in two rows. It is not a wise choice to extract more rows in this area, because a lot of people have a haircut which somehow covers a part of forehead, more patches tend to include those hairs. The influence of the hair on patch-based classifier can be significant. Because the computer vision can hardly distinguish the hair from the wrinkles, sometimes, those patches containing hairs will be treated as the wrinkle patches from old people.

The other 6 skin patches are extracted from the front cheek, 3 for each side. These areas are unlikely to be covered by shade due to the poses, and inside enough to avoid the edge of face. The skin around mouth is not considered, because even a small smile will cause many obvious wrinkles in those area, besides, for some males' photographs, the influence of beard is also unneglectable.

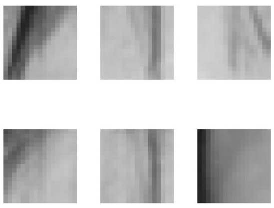

Figure 6: Some sample patches containing hair or edge of face

Although the selection of patch positions has already tried to avoid those sensitive areas. It is still quite common that for each face, there are one or two patches contain some non-aging-related factors, that disturbing the process of age estimation. As mentioned above, things like hairs and edge of face might significantly influence the computer vision analysis on skin texture. In order to ensure the accuracy of the patch classification, those disturbing patches should be better deleted from the dataset. Generally, those factors will make the variance of pixels in the whole patches become really high. By comparing the pixel variance with other patches from the same face, most of these disturbing patches are detected and removed. Since then there are 12 or 13 patches in total obtained from each face.

3.2 Baseline experiment

After acquiring the dataset of face patches, a baseline experiment is established first, which focuses on the variance of the pixels in each patch. By deploying the histogram and ROC curve analysis, this experiment can be regarded as a simple but independent age classifier. Some insights into the face patches from different age groups can be derived in advance. As a result, the potentials of free patch-based method are somehow examined at this early stage.

3.2.1 Variance and normalization

To examine the skin texture from image patches, the variation of the pixels is an important parameter. All of the extracted patches in this study are resized as 17*17 unit 8 pixels. Since pixel is the smallest controllable and addressable element of a digital image, each 8-units pixel encodes the image in its certain position. After applying the double-transform function, each pixels can be represented into a real number. And then the variance of these pixel numbers can indicate how far they are spread out from their average. One can expect a plain image results in zero variance, because all of its pixels are same. Thus, it is reasonable to assume that a smooth skin from young people will show a smaller variance than a coarse skin from older people.

In order to find the aging effects behind patches, it is necessary to compare the variances of the patch from different faces and ages. However, since the resolution of the photographs in database is different, the variances of the extracted patches are significantly varying from image to image. Adding with the different levels of illumination, those patch variances are not within a same comparable scale. Hereby, normalizing patches is an important step that reduces non-aging-effects among a large number of photos before the comparisons.

$$Var_{normalized} = \frac{Var(pixels)}{Mean(pixels)} \quad (1)$$

One of easiest way of normalizing applied in our case is dividing the variance by the mean number of pixels. In image processing view, the value of each pixel stands for the illumination level of this pixel, the mean number of all pixels is the average brightness of this patch. Thus, this step decreases the influence of different brightness on patches, and normalizes the variance.

3.2.2 ROC-curve and histogram

The Receiver Operating Characteristic (ROC) curve is a graphical plot tool for evaluating the diagnostic ability of a binary classifier. In this experiment, the patches from younger and older than 25 years old are labeled as $False$ and $True$ separately.

Before the plotting of ROC curve, a set of threshold of variance are defined. If the variance of the patch is higher than the threshold, then this patch is called $Positive$ sample in contrast to $Negative$ sample.

Based on the definitions above, there are $2 * 2$ statements can be derived. They are True-positive, False-positive, True-negative and False-negative separately. An optimal distribution of them has been shown in Figure 7 left. For example, the patches from older people have higher variances than that from the younger people is a True-positive statement.

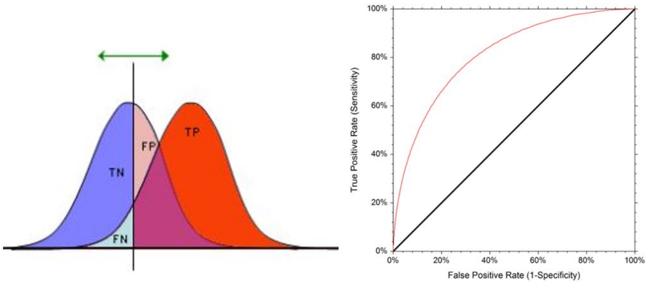

Figure 7: LEFT:distribution example of 4 states with a threshold, RIGHT:a ROC curve example with a diagonal line

The coordinate system of ROC curve uses $Sensitivity$ and $1 - specificity$ as the x and y axises 7. The $Sensitivity$ and $specificity$ are defined as:

$$Sensitivity = \frac{TP}{TP + FN} \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

Here, the T is true, F is false, N is negative and P is positive.The straight line in middle is called diagnoal line, which means there is no difference between two classes. Therefore, the ROC curve is expected to be away from the diagonal line as far as possible in this experiment. Even though it is a slight curve, the potentials of distinguishing two classes by patches are revealed.

Histogram is a much simpler comparing with ROC curve. The graph can be found in Experiment section. The histogram divides the variances of patches into many small ranges, also called bins. For each range, it shows the proportion of each class. As a result, people can easily learn in which class of patches has a higher variance possibility in a certain range. Some trends of patch variance along age can estimated.

3.3 CNN classification

To be brief, the Convolution Neural Network(CNN) is a highly non-linear mathematical function that maps the input data to a list of classification scores.[16] In this study, the CNN is applied to perform a binary classification, older or younger than a specific age, denoted as 1 or 0. The input data is the skin patches extracted from face, and the output is a two-value score for each patch, older or younger.

CNN architectures make the explicit assumption that the inputs are images, which allows people to encode certain properties into the architecture.[17] Similar to general neural networks, CNN transforms the input data through a series of mathematical operations, which called hidden layers. Each hidden layer is made up of a set of neurons, where each neuron is fully connected to all neurons in the previous layer, and where neurons in a single layer function completely independently and do not share any connections.

One of the most important hidden layer in CNN is the Conv Layer. The layer consists of a set of spatial filters with height and width, during the forward pass of image data, those filters will slide with a specific stride number. Consequently, each filter will produce a 2-dimensional activation map, which will let the network learn some visual features in images intuitively. The final output of Conv Layer will be the stack of those activation maps from filters.

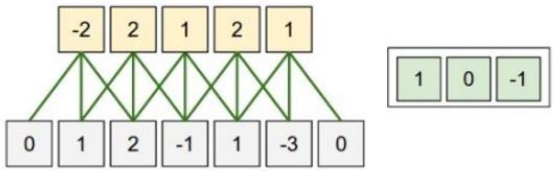

Figure 8: One-dimensional convolution example with stride of 1. Left: Multiplying the bottom input data with filter result in the activations on top. Right: The filter

ReLu layer and Pooling layer are commonly placed after the Conv layer. ReLu layer is referred as Rectified Linear Units, adding non-linearity to the operations. It performs as $max(0, x)$ thresholding at zero, and unchange the size of data. In contrast, pooling layer is a down-sampling operation, making the spatial size of data smaller. It can perform Mean, Average and many other functions.

Fully-connection layers is called the "output layer", it computes the input activations with a matrix multiplication. And in classification settings, it represents the class scores. The output score will be evaluated by a loss function, where the SoftMax is the widest applied[17]:

$$P_{class-A} = \frac{e^{x_A}}{e^{x_{older}} + e^{x_{younger}}} \tag{4}$$

In the formula, x_A is the score of class A, older or younger. x_{older} and $x_{younger}$ are the scores for each class. And $P_{class-A}$ is the probability of the class A derived by the SoftMax algorithm.

With the evaluation of the loss function, the back-propagation can be implemented to optimize the weights of multiplications between neurons. The network will perform a better classification with the tuned weights. This process is achieved by making partial derivation of loss function respects to each weight.

3.4 Fusion of Patch scores

After a proper training, the CNN model can be deployed to test the patches. A set of skin patches will be extracted from the new testing faces, just following the same algorithm as used in collecting the dataset. For each of those patches, the pre-trained CNN will make a classification score, scaled and expressed as a percentage.

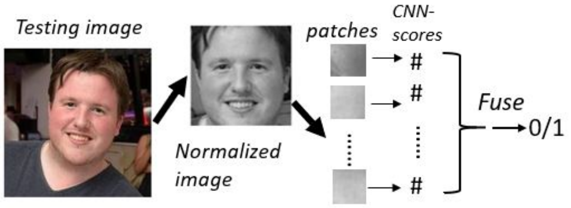

Figure 9: The work-flow of testing a image with fusion

The fusing step is responsible for combining those separate patch scores and determine the age class of the people in photographs. In this process, the algorithm of fusing plays an important role. It will take 12 to 13 patch scores from one face as the inputs, and return a final classification result 1 or 0 as the output.

In contrast to directly determining the age from one single patch by CNN, the fusion step is expected to compensate the potential errors in middle steps, and improve the final accuracy of the whole age estimation system.

4 Implement and Results

4.1 Database

The evolution of biometric systems relies on the availability of suitable datasets that allow the development, experimentation, benchmarking and performance evaluation of biometric systems.[19] For developing age estimation systems, collecting sufficient and appropriate data is always an important but challenging step.

Although there are a number of individual or multi-modal biometric datasets publicly available, the suitable datasets of face images for aging analysis are quite a few, because the selection of facial images is pretty strict. For example, the faces should better look straight into the camera and has no big facial expressions, since the resulted shades and wrinkles on face will influence the accuracy of classifier. Besides, the resolutions of images in a database should also be restricted.

Summarily, an appropriate database for age estimation from face images should contain following features:

1. It should contain a large number of facial images with age labels. For each age or a certain range of age, there should be multiple samples presented.

2. The non-aging-related variations among the photographs should be neglectable in dataset, such as face poses, wearing accessories and the different photographic types.

3. Possibly, the dataset could contain photos of fixed persons in different ages, which will significantly ease the implement of multimodal techniques, for example, the construction of age patterns.

At the beginning of experiments, there are two databases available, the FG-Net and the IMDB+Wiki. Although FG-Net is a widely-used database for age estimation, it is not appropriate for this study. Firstly, the most images are old photographs with relatively low resolutions. Secondly, it contains a lot of youth photos, which are not fit for the aiming age group, around 25 years old. Hereby, the IMDB+Wiki dataset is the left option. This dataset is created by the Computer Vision Lab, ETH Zurich, Switzerland. They crawled 0.5 million images of celebrities from IMDB and Wikipedia that they make public. This is the largest public dataset for age prediction to date. [4] However, the crawled database can hardly avoid containing some irrelevant photographs, and a lot of face images inside have inappropriate poses. So the images from IMDB+Wiki database still need to be preprocessed, as described in 3.1.1

4.2 Baseline experiment

The baseline experiments made use of two manually-picked datasets for eliminating the non-aging-effects as much as possible. The first dataset used has 242 male images and the second dataset contains 389 images with both female and male.From each face image, one noisy patch, which has the highest pixel variance is detected and deleted. As a result, there are $(14-1)*242$ and $(14-1)*389$ patches available from two datasets respectively. The ROC plot of two datasets have been shown in10, and the 25 years old is the boundary of binary classification.

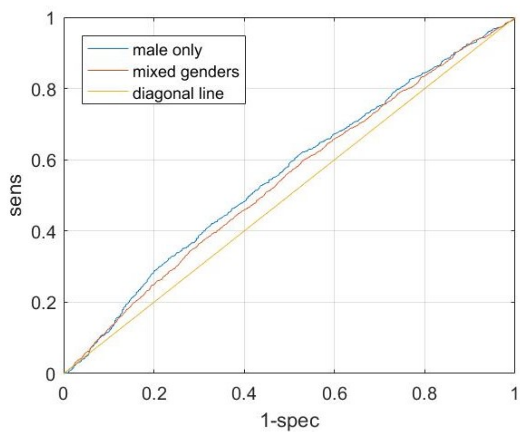

Figure 10: The Matlab plots of ROC curves from two face image datasets

As one can see, both ROC curves are above the diagonal line, but the male dataset shows a slightly better shape than the other. When the *True-positive* rate is 50%, the sensitivity of male database reaches to around 62% in contrast to 56% of the database with mixed genders. since the different skin textures between male and female will undoubtedly influence the age classification, it is not a surprise to see this phenomenon.

Moreover, although the ROC curve looks like indicating some potential difference between the skin patches from two age groups, it cannot prove the correlation between the age and patch is certainly exist, because the classifying trend of ROC is not obvious enough, and the used datasets is in small-scale.

However, this experiment is simply focused on the variance of pixels, which is a small feature in skin patches. More biometric features can still be expected from the patches by other algorithms, and they will be able to contribute to age classification further.

Based on the male datasets, two histograms in different forms are derived as below11:

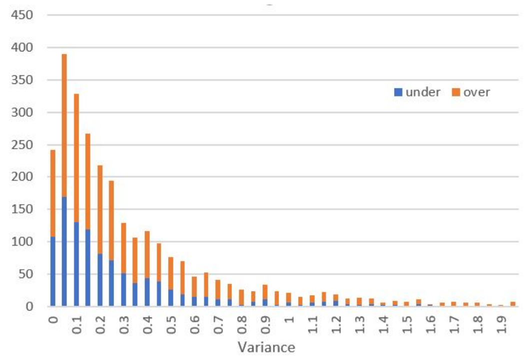

Figure 11: The histogram of data data numbers from the male dataset

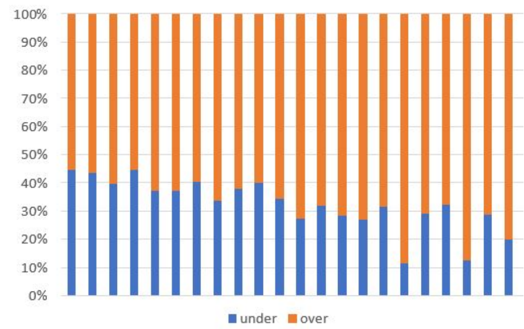

Figure 12: The histogram of percentage from the dataset of two genders

In histograms, under and over classes are defined by younger and older than 25 years old respectively. The distribution shape in first graph looks clear that, the number of 'over' group data is higher than the other group for each variance range. Meanwhile, the second graph reveals that the ratio of older group over younger group is slightly decrease with the increase of variance. Since the trend is not obvious and the samples are limited, this feature cannot to be confirmed for all skin patches. Next, it is worthwhile to plot the datasets with two genders for comparison13, for getting more insight into the correlation between variance and age.

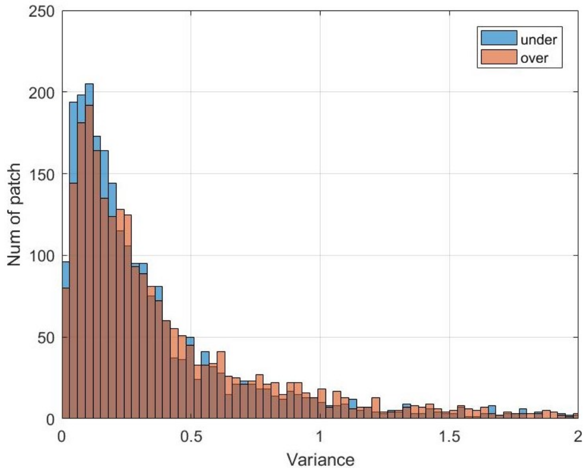

Figure 13: The histogram for dataset of two genders

Due to the dataset consists of more images and the factors of gender cannot be neglected, In the new hidtogram, the difference between two group is pretty small, the distributions of the 30 bins for each class are pretty similar. However, it is still noticeable that in small range of variances(between 0 and 0.25), the proportion of patches from under class is higher. In contrast, for patches with larger variances above 0.25, most are from over class.

4.3 CNN classification

4.3.1 Training network

After the baseline experiment showed an uncertain correlation between human age and skin patches variance, the Convolutional Neural Network is implemented to investigates more aging features behind the patches. In order to let CNN capture more characteristics from the patch, a large dataset is applied with both male and female images. The total number of skin patches is 13*5842. By using Matlab toolbox, those data together with age their labels are transformed into LMDB format, which is compilable for the Caffe model. The first 80% of those data was employed for deriving the weights, and the left 20% was for back-propagation to optimize those weights.

From the experiment, the build-in function of Caffe gives an accuracy of 59.3%. In this process, three Convolution Layers, two ReLu functions and Softmax algorithms 3.3 are operated. For comparison, the same Caffe model is applied again on the same the database, but the boundary of classification at this time is 28 rather than 25, resulting the 68.2% accuracy.

4.3.2 Testing and fusing patches

Although the accuracy of examining the individual patch is not really high, the fusing step can be applied to improve the final estimation accuracy. A set of new face images are selected for providing test patches, the Matlab function is responsible to forward them to the trained modal. The deployed CNN model gives a 1*2 single score for each patch by loss function, and one stands for younger, the other for older.

Unfortunately, the results showed the Caffe model in not well-trained. For 25 years old classification, the testing score is always around 0.98 vs 0.02, and 0.548 vs 0.452 for 28-years-old classification. Therefore, the step of fusing 13 patches from each face is meaningless to implemented, since it will not improve the final accuracy then.

5 Evaluation

The CNN model gives same or similar same testing score for any input patches, meaning that classifying all images into one age group is the best strategy it can found during the training. To investigate this possibility, the training accuracy 4.3.1 and the ratio of age groups in database have been compared in a table. Moreover,

age boundary	group ratio	training accuracy	testing score
25	50.2%	59.3%	0.98/0.02
28	59.4%	68.2%	0.548/0.452

As one can see, the ratio of age groups in database is closed to the training accuracy of the CNN, the around 10% difference might because the CNN shuffles training data and only select a part of it for evaluation every time. Since the progress of age prediction or estimation depends heavily on the available databases[20], especially when the Neural Network is used. In our experiments, the bad input training data cannot help to derive the optimal weights, although the training and testing structure of CNN operated as expected.

Assuming that CNN cannot learn so much age-related features from those 17*17 pixels, at very end of this study, the author also tried to investigate whether the larger skin patches are better input for CNN. Fortunately, the skin patches 72*72 from the same normalized database3.1.1, is able to let CNN make a better classification. The testing score is not always indicating the same age group anymore, but the accuracy can only reach to 54.3% for classification of 28 years old. And due to the large size, only 5 patches can be extracted from each face, the benefits of fusing is declined.

6 Improvement

6.1 Optimize Aging Database

No matter for the baseline experiment or CNN training in this study, the image data always played an important role. The bad input data can hardly provide sufficient biometric features for CNN to learn, which directly determines the results of experiments. As mentioned in method part3.1.2, the final size of patch was chosen to be 17*17 pixels is because is compromises the general resolution of the images in database, however, the small patches will unavoidably filter out some features which might be useful for age estimation.

The preferred training data should be appropriate both in quality and amount, which is contradicted for many current database of face images. Although the original images have been normalized before the experiments, the various resolution and facial poses still widely exist among the images. Thus, one can expect that the performance of age estimation/classification system will be significantly improved, if there is a better database provided face images with stable quality of pose, expression and resolution.

6.2 Age estimation with gender

The gender should has been an important factor during the selection of image data. And in the baseline experiment, this factor was taken into account as stated in Section4.2, the manually selected male dataset provided a better ROC curve than the mixed dataset. However, for training the CNN, manually separate image data by genders is not possible anymore, because a large number of image data is required.

Since the main feature encoded in patches is the skin, the gender factors disturb the learning process of CNN. From our experience, it is noticeable that females have smoother skin, comparing with males at the same age. In addition to the existence of natural difference between male skin and female skin, more females have the habit of using make-ups, which hides a lot of biometric features of aging. There, to improve the accuracy of patch-based system further,let it be trained by female and male image separately could be necessary.

6.3 Threshold of age classification

The unsatisfying performance of the age classification system does not mean the patch-based method is not applicable for age estimation. In this paper, the classification system uses 25 years old as the boundary, however, the difference around this age is pretty small and hard to be detected. Other choice of boundary setting might leads the patch-based method to have a better performance. More specifically, distinguish people who are 24 and 25 years old is much more difficult than classifying people who are younger than 20 and older than 25.

In further evaluation of patch-based age estimation from face images, it worth to try more classification over other age groups. And for binary classification, the boundary could be set wider for clear separations.

7 Conclusion

In this paper, the free patch-based method is proposed for age estimation from facial images by Convolutional Neural Network. To verify the feasibility of this method, an age classification system based on skin patches is built. Along with several experiments, it attempted to effectively classify the facial images under age groups of younger and older than 25. However, within the time frame of the study, the effective classification results with high accuracy are not derived.

The first experiment examined the variance of pixels in patches. By making use of ROC and histogram, it seems that some correlations between the variance and the age are exist but cannot be certainly proved. The second experiment showed the provided patches cannot provide enough information for CNN to make an accurate mathematic model. Although in evaluation part, the larger patch of skin might lead to more logic model, the accuracy is just around 55%.

Overall, the free patch-based age estimation method by Convolutional Neural Network did not work well in this study for classifying 25-years-old groups, but it does not mean that the free patch-based method by CNN is not applicable for age estimation. Several improvements over database, gender factor and age boundaries can be made to optimize the system in the future. Considering the implicit features encoded by biometric patches and the additional benefit of expanding data source, the method still has more potential values in the field of age estimation/classification.

8 Reference

[1] A. Montagu (1960), *Introduction to Human Age Estimation Using Face Images. A Handbook of Anthropometry*

[2] J.Van Hoof(2016), *The effectiveness of ID readers and remote age verification in enhancing compliance with the legal age limit for alcohol*

[3] C. Sharma(2014) *Big Data Analytics Using Neural networks*, San Jose State University

[4] N. Ramanathan, R. Chellappa, and S. Biswas. *Age progression in human faces : A survey.* Journal of Visual Languages and Computing, 2009. 1.2 24(4), pp. 442-455.

[5] D. W. Thompson (1992). *On Growth and Form. Dover Publications.*

[6] L. G. Farkas. *Anthropometry of the Head and Face in Medicine.* Elsevier Science Ltd, 1981. 3.3

[7] Y. H. Kwon and N. da Vitoria Lobo. *Age Classification From Facial Images.* Computer Vision and Image Understanding Journal, 74(1):1:21, 1999. 1.2, 1.2

[8] A. Lanitis, C. J. Taylor, and T. F. Cootes. Modeling The Process Of Ageing In Face Images. Proceedings of the Seventh International Conference on Computer Vision, 1:131-136, 1999. 4.3, 4.7

[9] A. Lanitis, C.Taylor, T. Cootes. (2002). *Toward Automatic Simulation of Aging Effects on Face Images.* IEEE Transactions on Pattern Analysis and Machine Intelligence,

[10] T.Kim, H.Kim , W.Hwang, S.Kee (2003) *Independent Component Analysis in a Facial Local Residue Space*

[11] Spreeuwers, L. J. (2011). *Fast and Accurate 3D Face Recognition Using Registration to an Intrinsic Coordinate System and Fusion of Multiple Region classifiers.* 93(3), 389-414. DOI: 10.1007/s11263-011-0426-2

[12] S. Lucey and T. Chen (2004). *A GMM Parts Based Face Representation for Improved Verification through Relevance Adaptation.* CVPR, pp. 855-861,

[13] Ming Yang1, Shenghuo Zhu1, Fengjun Lv2, Kai Yu1 (2014) *Correspondence Driven Adaptation for Human Profile Recognition*

[14] R. Rothe, R. Timofte, L. Gool, (2015) *DEX: Deep EXpectation of apparent age from a single image.* Computer Vision Lab, D-ITET, ETH Zurich, Switzerland

[15] Active Shape Models with Stasm, Stephen Milborrow, http://www.milbo.users.sonic.net/stasm/

[16] CS231n: Convolutional Neural Networks for Visual Recognition (2017) http://cs231n.github.io/convolutional-networks/

[17] S.Boer (2017) *Patch-based Approach to Finger Vein Verification with Convolutional Neural Network* University College Twente.

[19] Andreas Lanitis, Nicolas Tsapatsoulis, Anastasios Maronidis *Review Of Ageing With Respect To Biometics An Diverse Modalities*

[20] G.Guo (2013) Age Prediction in Face iamges