Comparison of Three Different CNN Architectures for Age Classification

M. Fatih Aydogdu and Vakkas Celik and M. Fatih Demirci
Department of Computer Engineering
TOBB University of Economics and Technology
Ankara, Turkey
Email: {mfaydogdu, vcelik, mfdemirci}@etu.edu.tr

***Abstract*—As one of the powerful tools of machine learning, Convolutional Neural Network (CNN) architectures are used to solve complex problems like image recognition, video analysis and natural language processing. In this paper, three different CNN architectures for age classification using face images are compared. The Morph dataset containing over 55k images is used in experiments and success of a 6-layer CNN and 2 variants of ResNet with different depths are compared. The images in the dataset are divided into 6 different age classes. While 80% of the images are used in training of the networks, the rest of the 20% is used for testing. The performance of the networks are compared according to two different criteria namely, the ability to make the estimation pointing the exact age classes of test images and the ability to make the estimation pointing the exact age classes or at most neighboring classes of the images. According to the performance results obtained, with 6-layer network, it is possible to estimate the exact or neighboring classes of the images with less than 5% error. It is shown that for a 6 class age classification problem 6-layer network is more successful than the deeper ResNet counterparts since 6-layer network is less susceptible to overfitting for this problem.**

I. Introduction

Automated age estimation from facial images is an important problem studied in several fields such as computer forensics, human computer interaction, biometrics, entertainment, pattern recognition, and computer vision. An accurate age estimation forms the basis for applications such as producing younger or older images of a person, which is needed to find a missing person or a criminal. An age estimation system may be used to prevent vendine machines from selling products, e.g., alcohol, tobacco, to underaged individuals. The fact that people's preferences change depending on their ages also yields a number of potential applications of automated age estimation. In all these applications, the individual is not required to be identified, but rather his/her age is to be estimated. Due to the importance and a number of application areas, the problem has received much attention employing a diverse set of solutions from both industry and academia [1].

The main difficulty for solving this problem lies in different aging patterns of different people, i.e., aging patterns depend on both internal and external factors such as genes, gender, lifestyle, ethnicity, and race [2]. Thus, the actual age of a person may be different from his/her appearance age. Given a face image, the objective of computer-based age estimation is to assign a label to the face with the exact age or the age group it belongs. We focus on the problem of age group classification rather than that of exact age estimation in this paper.

The relationship between age and face is studied in the context of simulating the aging effects on human faces, e.g., [3], [4], [5], [6]. To name a few, aging variations are simulated by superimposing typical aging changes in shape and color on face images in [3]. A dense surface point distribution model for expressing the shape changes with respect to growth and aging is presented in [6]. While these techniques do not perform age estimation, the mapping from age to face inspires various frameworks to do the automated age estimation.

One of the earliest age estimation frameworks is due to Lanitis et al. [7], who generate a statistical model of facial appearance used as the basis for obtaining a compact parametric description of face images. Shortest distance classifier, supervised and unsupervised neural networks are employed for designing the age estimator in this work. Although promising results have been reported, empirically determining the aging function, learning of an individual's aging pattern based on the face images of the individual only, and computing the aging function for the previously unseen face image simply as a linear combination of the known aging functions are main drawbacks associated with this approach.

To address these problems, Geng et al. [2] present the AGES framework, where the aging pattern is modeled by obtaining a representative subspace based on the sequence of a particular person's face images sorted in time order. The aging pattern of an unseen face image is then computed by the projection in the subspace that reconstructs the face image with minimum reconstruction error. The lack of complete aging patterns yields incomplete training data. To deal with this issue, an iterative learning algorithm that estimates a part of the missing personal aging pattern using the global aging pattern model is adopted.

Ricanek et al. [8] use the Active Appearance Model (AAM) to obtain relevant aging features and identify the most important ones through Least Angle Regression (LAR). AAM has also been used in various techniques ([9], [10]) to extract the appearance features, which is capable of describing the shape and texture of a face image with a set of parameters, after a proper training. The performance of AAM for facial aging is also studied in [11]. A recent survey on face estimation may be found in [12].

In this paper, we study the performance of deep convolutional neural networks (CNN) with three different architectures for age classification. Our motivation for using deep CNN comes from recent studies showing that they achieve a

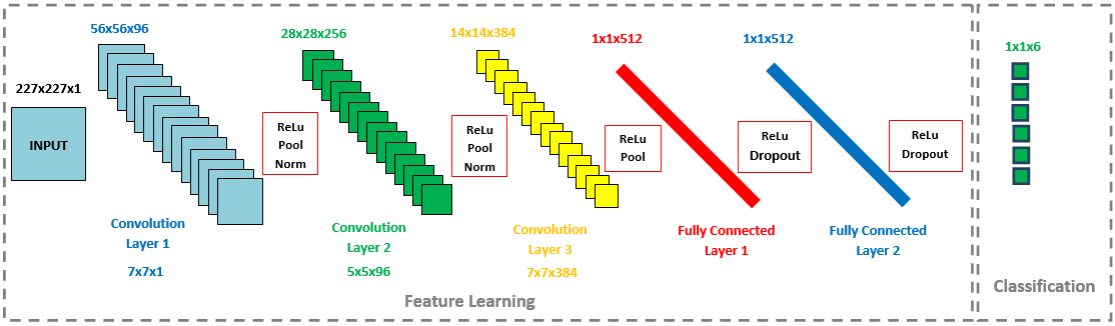

Fig. 1. The Architecture of 6 Layer Network to solve 6 age groups classification problem

tremendous success for several computer vision tasks such as image classification [13], [14] and object detection [15], [16]. CNN obtain a suitable feature vector consisting of low, mid, and high level features by training the entire system end to end. The use of CNN for age estimation is not new as it has been recently introduced in [17]. Although powerful results have been reported by this approach, its network architecture is rather simple, i.e., it consists of only three convolutional layers and two fully-connected layers with a small number of neurons. On the other hand, the network depth is shown to be important [18] and the best image classification techniques exploit much deeper models [19], [20].

To see whether deeper CNN achieve better age classification scores, we employ [17], ResNet-18 and ResNet-34 [21], consisting of 6, 18 and 34 layers, respectively. ResNet-18 and ResNet-34 are successfully used on the ImageNet classification dataset [22] consisting of 1000 classes. For the experiments, we use the MORPH dataset [23] containing more than 55k near-frontal face images.

The rest of the paper is organized as follows. After introducing the CNN architectures and the dataset in the next section, we present the image processing algorithms applied to each input face image in Section III. We then describe the experiments and report the age classification results for each architecture in Section IV. The paper is concluded in Section VI.

II. Networks and Dataset

A. 6 Layer CNN

The structure of the 6 layer CNN is shown in Fig. 1. The network is composed of three convolutional layers and three fully-connected layers. The data volumes of each layers are shown at the top of the layers in figure. Although the original network architecture uses 227x227 RGB images as input, in this study, grey scale images at the same resolution are processed in the network. While three convolutional layers are followed by rectified linear operator (ReLU) layers and local response normalization (Norm) layers taking the maximal value of 3x3 regions with two-pixel strides, the first two fully connected layers are followed by ReLu and dropout layers having a 0.5 dropout ratio. The last fully connected layer has 6 neurons corresponding to the values for each age class in the discussed study.

B. ResNet

As reported in [24], [25], [21], plain networks whose building blocks are shown in Fig. 2 tend to degrade as the networks become deeper. So as to deal with the degradation problem and use deeper networks to solve complicated problems residual networks are employed. Although the earlier implementations of residual networks exist in the literature, the advantages of the residual network is discussed in a recent work of [21]. According to the discussion in that paper, the residual networks are more resistant to the degradation with respect to plain networks. In a network, the main motivation beyond generating residual connections is to provide alternative connections to the regular ones. As shown in Fig. 2, the residual connections have a constant coefficient, which equals to 1, excluding the residual connections from train procedures. In a train procedure, if the coefficient of a regular connection tends to converge to 0, the residual shortcut assures the integrity of the network. When a regular connection is shortcutted by a residual connection the cumulative data calculated before the regular connection is forwarded to the rest of the net. The alternative connections give the opportunity to the network to be able to use these shortcuts instead of regular connections when necessary.

So far ResNet [21] is the most successful residual network used in different objectives. Based on ResNet, different residual CNN architectures are generated and submitted to ILSVRC and COCO competitions. In these competitions, the submitted networks won ImageNet classification, ImageNet detection, ImageNet localization, COCO detection and COCO segmentation tasks. An example of ResNet architectures composing of 18 convolutional layers and modified within this study is shown in Fig. 3. In this architecture, Norm and ReLu operations are used after each convolution layer. The residual connections are realized by sum operations and the curved arrows indicate the shortcut residual connections to

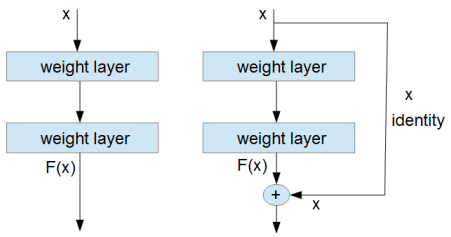

Fig. 2. Plain Building Blocks (left) and Residual Building Blocks (right)

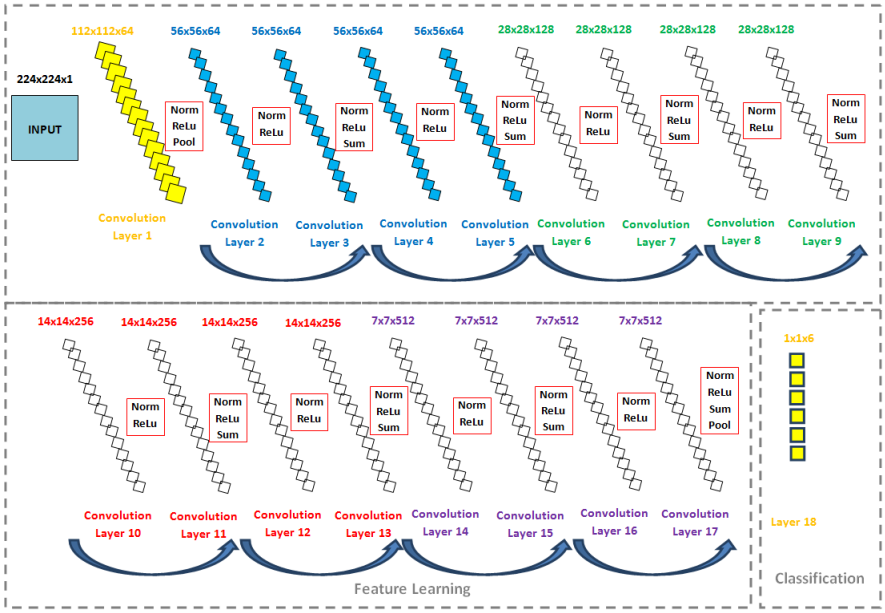

Fig. 3. The Architecture of ResNet-18 to solve 6 age groups classification problem

the sum operations. After 4 groups of convolution layers the final convolution layer placed to realize classification operation to the input images. Different than the 6 layer network discussed previously, the ResNet architecture does not include any dropout layers. There are different versions of ResNet architecture, some of which have more than 1000 convolution layers. Among the ResNet versions, ResNet-18 and ResNet-34 are considered to be suitable in order to compare with the 6 layer counterpart in terms of depth vs. success.

C. Dataset

The face database used in this paper is Album-2 of Craniofacial Longitudinal Morphological Face Database (MORPH) [23]. The images in the dataset belong to individuals photographed from front within five year period as shown in Fig. 4. Dataset contains 55,134 images belonging to more than 13,000 persons whose ages range from 16 to 77. The median age in the dataset is 33 and each person in the dataset have 4 images on average. It is one of the largest datasets available for research purposes and the size of the dataset make it suitable for learning experiments. Moreover, since the persons in the dataset represent different races and moods, the dataset provides sufficient image diversity for learning.

III. Image Processing

In order to increase the success of the learning process, the objective function binding the input images to output classes is simplified. It is intended to equalize the useful information coming from each image to the network. To do so, grey scale conversion, image rotation, visage detection, and visage rescaling processes are applied to the images in database as shown in Fig. 5.

A. Image Rotation

After all raw images in the dataset are converted to grey scale, image rotation is applied. With this step, it is intended to

Fig. 4. Sample face images from the Morph Dataset

make the direction of the visages parallel to the vertical image edge. To make image rotation, a method composing of two phases is applied. In the first phase, the rotation is applied by finding the eye pairs and rotating the images with respect to the direction of the line connecting the eye pairs. In case, the eye pairs are not found and thus, the images are not rotated in the first phase successfully, we find the face directions and rotate the images with respect to the direction of the faces in the second phase of the method.

More specifically, in the first phase of the method, both eyes are detected separately and the necessary rotation angles are determined with respect to the vertical positions of the

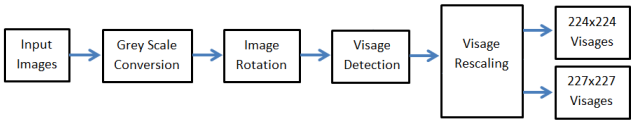

Fig. 5. Image Processing

eyes. To detect the eyes, two different variants of Viola Jones algorithm [26] are used. Classification and regression tree (CART) [27] based eye detection algorithm gives the correct eye positions in 54121 images of the 55134 images in the dataset. Decision stump based eye detection algorithm [28] gives the eye positions of 500 images of the remaining ones. For the remaining 523 images of the dataset, eye detection algorithms do not give the correct positions of the eye features. Therefore, face detection algorithms are used in second phase for image rotation. In the second phase, 2 variants of Viola Jones algorithm to detect faces are used to obtain the rotation angles. The main idea of this phase relies on the fact that face detection with Viola Jones algorithm is sensitive to image rotations. More precisely, while detecting the face features, Viola Jones algorithm locates only the features, which do not have a big leaning angle with respect to the vertical edge of the image. The following steps are, thus, followed to estimate the rotation angle. For each image, its rotated versions are generated starting from a rotation angle of -45° to 45° with 1° steps. Local binary patterns (LBP) based face detection [29] and CART based face detection [30] algorithms are then applied to the rotated images. When successful face detections occur in rotated images with these face detection algorithms, the rotation angles are recorded and mean of the recorded angles for each image is used to make image rotation.

B. Visage Detection

The next image processing technique applied to each rotated image in the database is the visage detection. We note that while training a CNN for age classification, the pixels outside the visages do not provide valuable information. Therefore, such pixels positioning outside of the image visages are cropped out from the rotated images. The visages detected in this step are square crops whose vertical and horizontal pixels counts are equal. In order to detect the visages in the images, two variants of Viola Jones face detection algorithms are employed similar to the previous image rotation step. First, CART based face detection [30] algorithm is applied to the images. If this algorithm is not successful in finding the visages, LBP based face detection is then used. According to the visage detection results, CART based face detection algorithm finds visages in 55038 rotated images of the dataset. LBP based face detection algorithm finds visages in 35 of the remaining images and in the rest of the 61 of the images no visages are detected with any of these algorithms. Moreover, we note that 41 of the detections made with CART based face detection and 9 of the detections made with LBP based face detection are incorrect. In order to use the complete dataset, visages are cropped manually in the 61 images in which no face detection is achieved and in the 50 images in which incorrect visages are detected. Since the networks used in this study require 224x224 and 227x227 square images as input, the detected face crops are rescaled to match with the input sizes of the networks.

IV. Experiments

A. Errors in Dataset Meta Data

Before proceeding with the train and test phases, the whole dataset is examined and the documented ages in the meta data of the images are checked. Since some obvious errors in the provided ages are noticed, born dates and image dates are checked for the whole dataset. We observe that there are different born dates for some images belonging to the same person in the dataset. More specifically, 10163 of the images have unreliable born dates. To make the dataset more reliable, the following correction is applied. When face images belonging to the same person are associated with more than one born date and one of these born dates is observed more frequently than others, the all of the born dates for this person is set to this frequent born date. With this method, the born dates for 9913 images are corrected. Secondly, if the face images of the same person have more than one born date and no frequent born date is observed, we apply no change. For the images with unreliable born dates, the age meta data of the images are calculated again and the new computed ages are used in the experiments.

B. Tests

In order to increase the number of the images used in the experiments, horizontal flips of the images are generated. With these flips the number of the images used is doubled to 110268. In the experiments, 80% of the images are used in training the CNNs and the remaining 20% is used for testing. The images in the dataset are divided into 6 different classes whose age boundaries and number of images are shown in Table I.

After the images are divided with respect to the age classes, train and test sets are generated. In each class, images are randomly distributed among train and test sets while preserving the train-test ratio. When distributing the images, in case an image belonging to a person is randomly placed in one of the train or test sets, then all of the images belonging to the same person and their horizontal flips are also placed to the same set. With this principle, the data diversity of train and test sets is preserved.

While training the CNNs, the weight decay and momentum values are selected as 0.0001 and 0.9 are selected, similar to [21]. The learning rate is started from 0.001 and decreased according to the progress in learning. The train procedures continue until precise stabilizations are observed for test errors. These stabilizations do not occur before 60 epochs for the CNNs trained. On a NVIDIA Quadro K4000 192 bit GPU with 3GB memory, the train process of 6-layer network lasts about 1 day. The train procedures for ResNet-18 and ResNet34 last 8 times and 16 times longer than the 6-layer network, respectively. Note that to make a fair comparison in the experiments, the same train and test sets are used for all of the train procedures.

V. Results

In order to compare the success of the performance of the trained CNNs, two different criteria are used. The first criterion is to see whether an input image can be correctly classified, while the second criterion is defined as the 1-off success, showing whether an input image can be correctly

TABLE I. The Number of Images in Classes

16-20	21-27	28-34	35-41	42-48	49-
18910	23638	18990	22784	16654	9292

classified or it can be classified in one of the neighbors of the correct class. The 1-off success is useful especially for cases where the age of a person is close to the boundary between two age classes. When this occurs, the networks should not be penalized for classifying the input image as one of its neighbors. Therefore, 1-off success should be considered as the main success criterion in the discussions throughout this paper.

The overall age classification results are presented in Tables II, III and IV respectively. The rows in the tables correspond to the real age classes, while columns correspond to the estimated age classes of test images. The diagonal cells colored as dark green indicate the number of images whose ages are estimated correctly. For each row, non-diagonal elements contain the number of images whose ages are estimated incorrectly and the columns of these elements indicate the age class estimated incorrectly. In the non-diagonal cells for each row, the light green colors show the number of neighboring class estimations increasing the 1-off success rate.

As shown in Tables II, III and IV, for each row, while moving away from diagonal cells the numbers in the cells converge to 0. This shows that all of the networks are suitable to be trained for 6-class age classification problem. In Table V, the overall exact and 1-off successes of the networks is shown. According to the overall results, for both exact success and 1-off case, we observe that 6-layer network is more successful than the other deeper residual networks. Precisely, the 6-layer network has more than 95% success for 1-off criteria and more than 53% success for the exact criteria. In terms of standard deviation among success rates in different age classes, 6-layer network has the minimum value. This presents that the

TABLE II. TEST RESULTS OF 6-LAYER NETWORK W.R.T CLASSES

Image Classes	Number of Estimations w.r.t. Classes					
	16-20	21-27	28-34	35-41	42-48	49-
16-20	1790	1730	74	6	0	0
21-27	405	3309	895	128	5	0
28-34	22	867	1596	990	87	2
35-41	3	183	1171	2719	653	37
42-48	1	31	176	1348	1409	337
49-	0	1	9	138	690	792

TABLE III. TEST RESULTS OF RESNET-18 W.R.T. CLASSES

Image Classes	Number of Estimations w.r.t Classes					
	16-20	21-27	28-34	35-41	42-48	49-
16-20	2497	1002	94	7	0	0
21-27	1135	2707	751	136	13	0
28-34	137	923	1333	1023	146	2
35-41	28	205	1042	2593	841	57
42-48	3	35	192	1206	1486	380
49-	1	2	16	172	700	739

TABLE IV. TEST RESULTS OF RESNET-34 W.R.T. CLASSES

Image Classes	Number of Estimations w.r.t Classes					
	16-20	21-27	28-34	35-41	42-48	49-
16-20	2231	1217	123	28	1	0
21-27	923	2808	766	226	19	0
28-34	122	924	1216	1108	192	2
35-41	25	249	897	2568	958	69
42-48	4	30	215	1175	1498	380
49-	2	6	16	191	730	685

TABLE V. EXACT AND 1-OFF SUCCESS OF NETWORKS

Image Classes	6-Layer Success%		ResNet-18 Success%		ResNet-34 Success%	
	Exact	1-off	Exact	1-off	Exact	1-off
16-20	49.72	97.78	69.36	97.19	61.97	95.78
21-27	69.78	97.20	57.09	96.86	59.22	94.83
28-34	44.78	96.89	37.40	92.00	34.12	91.13
35-41	57.05	95.32	54.41	93.92	53.88	92.80
42-48	42.67	93.70	45.00	93.03	45.37	92.46
49-	48.59	90.92	45.34	88.28	42.02	86.81
Mean	53.76	95.82	52.56	94.23	50.94	92.96
STD	9.98	2.61	11.29	3.31	10.76	3.17

success of 6 layer network is less dependent to the age classes. Although ResNet-34 is deeper than ResNet-18 and have more classification potential, the shallower is more successful for both of the criteria.

While discussing the classification performances for the networks, the ResNet [21] paper must be revisited. In the paper, after the classification power of ResNet versions are discovered the authors explore the success of ResNet prototype by experimenting a 1202-layer version for a 10 class problem. They end up with the fact that ResNet-1202 is not as successful as ResNet-110 and they argue that ResNet-1202 is unnecessarily large and more susceptible to overfitting than the shallower ones. While analyzing the classification performances for ResNet-18 and ResNet-34 in this paper, we observe a similar situation. ResNet-34 is unnecessarily large with respect to ResNet-18 in a 6 class age classification problem having similar input images, i.e., very few number of pixels give necessary information for age classification. Since 6-layer network is more successful than ResNets, we argue that in the learning procedure, ResNets learn irrelevant data from visages, which in turn, causes reductions in the performance. As pointed out in [21], the lack of dropout layers can be considered as one of the reasons reducing the success of ResNet architectures.

Finallly, we analyse the images correctly classified according to the 1-off success using all three architectures (Figure 6). We observe that 88.44% of images are classified correctly by all networks. Moreover, in 1.12% of the test images none of the networks is successful. In addition, all of the networks have the ability to classify some test images in which no other network is successful. Specifically, 1.83% of the images are only classified correctly by 6-layer, while 0.79% and 0.57% of the images are classified by Resnet-18 and Resnet-34. respectively. These results are consistent with the previous performances and depict the improved classification rate offered by 6-layer.

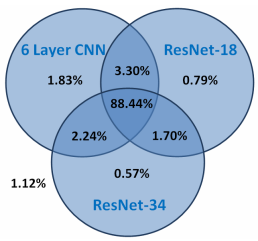

Fig. 6. 1-off success ratios of CNNs

VI. Conclusion

In this paper, the exact and 1-off success for three CNNs with different depths and architectures are examined for an age classification problem consisting of 6 classes. Prior to the train and test stages, the input images are rotated and the visages in the images are detected and cropped. After the long train procedures run on GPU, test images are classified by the trained networks. It is observed that although ResNet18 and ResNet-34 are perfect architectures to solve complicated classification problems, they are unnecessarily large with respect to 6-layer network for the relatively simpler classification problem discussed in this paper. When both the success and duration of training periods are considered, 6-layer network is more suitable than its deeper competitors for this problem. Although, with residual networks, it is sometimes possible to overcome the degradation problem occurring in deeper CNNs, it is not the case for this age classification problem since overfitting makes ResNet-34 less successful than ResNet-18.

For future work of this study, networks can be trained when the images in the dataset are divided into more age classes. The classification performance of networks deeper than 6-Layer network and shallower than ResNet-18 can be examined. Moreover, in order to understand the effect of image rotation and visage detection to the success of the networks, the images in the dataset can be used without the preprocessing. Since the objective function will be more complicated for both of these cases it is possible for the ResNet version to be more successful than the 6-layer network.

References

[1] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 11, pp. 1955–1976, 2010.

[2] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.

[3] D. M. Burt and D. I. Perrett, "Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information," *Proceedings of the Royal Society of London B: Biological Sciences*, vol. 259, no. 1355, pp. 137–143, 1995.

[4] B. Tiddeman, M. Burt, and D. Perrett, "Prototyping and transforming facial textures for perception research," *IEEE computer graphics and applications*, vol. 21, no. 5, pp. 42–50, 2001.

[5] A. J. O'toole, T. Vetter, H. Volz, and E. M. Salter, "Three-dimensional caricatures of human heads: distinctiveness and the perception of facial age," *Perception*, vol. 26, no. 6, pp. 719–732, 1997.

[6] T. J. Hutton, B. F. Buxton, P. Hammond, and H. W. Potts, "Estimating average growth trajectories in shape-space using kernel smoothing," *IEEE transactions on medical imaging*, vol. 22, no. 6, pp. 747–753, 2003.

[7] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 34, no. 1, pp. 621–628, 2004.

[8] K. Ricanek, Y. Wang, C. Chen, and S. J. Simmons, "Generalized multi-ethnic face age-estimation," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–6.

[9] K. Luu, K. Ricanek, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in *Biometrics: Theory, Applications, and Systems, 2009. BTAS'09. IEEE 3rd International Conference on*. IEEE, 2009, pp. 1–5.

[10] K. Luu, T. Dai Bui, C. Y. Suen, and K. Ricanek, "Spectral regression based age determination," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 103–107.

[11] A. Sethuram, K. Ricanek, and E. Patterson, "A hierarchical approach to facial aging," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 100–107.

[12] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *2013 International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–8.

[13] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *European Conference on Computer Vision*. Springer, 2014, pp. 818–833.

[14] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[15] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[17] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2015, pp. 34–42.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1026–1034.

[20] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, 2015.

[22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015.

[23] K. Ricanek and T. Tesafaye, "Morph: A longitudinal image database of normal adult age-progression," in *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*. IEEE, 2006, pp. 341–345.

[24] K. He and J. Sun, "Convolutional neural networks at constrained time cost," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5353–5360.

[25] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," *arXiv preprint arXiv:1505.00387*, 2015.

[26] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.

[27] S. Yu. (2016) Shiqi Yu's Homepage. [Online]. Available: http://http://yushiqi.cn/research/eyedetection/

[28] M. Castrillón, O. Déniz, C. Guerra, and M. Hernández, "Encara2: Real-time detection of multiple faces at different resolutions in video streams," *Journal of Visual Communication and Image Representation*, vol. 18, no. 2, pp. 130–140, 2007.

[29] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[30] R. Lienhart, A. Kuranov, and V. Pisarevsky, "Empirical analysis of detection cascades of boosted classifiers for rapid object detection," in *Joint Pattern Recognition Symposium*. Springer, 2003, pp. 297–304.