Facial Age Estimation using Deep neural networks: A Survey

Marwa Mahmoud Badr   Amany Mahmoud Sarhan   Reda M. Elbasiony

Faculty of Engineering, Tanta University/Computers and Control Department, Tanta, Egypt
Email: { Marwa.badr, amany_m_sarhan, reda}@f_eng.tanta.edu.eg

Abstract— *Facial age estimation techniques are extensively used in vitality applications nowadays; however, they are a time-consuming task. Compared to other methods, deep learning algorithms showed a superior performance to solve face related age estimation problem. The current surveys are incomplete, not up-to-date and are not concerned with deep learning in particular. In this paper, a comprehensive updated summary of the deep age estimation frameworks are introduced. We review the previous research efforts in deep age estimation systems and present the popular datasets. Convolution Neural Networks (CNN) are introduced with their models. The performance criteria widely used in evaluation are declared. Finally, the challenges and the future directions are discussed to reach a robust facial age estimation system.*

Keywords— *feature analysis, deep neural networks, estimating age, facial aging datasets, Convolution Neural Network*

I. INTRODUCTION

Recently, age-related research has become a challenging need due to being very important to several applications. Age estimation is used in access control systems where the service is provided according to the customer age. For example, tobacco vending machines in japan where child and adult under 20 years old were prevented from buying a cigarette [1]. Also in marketing issues, age-based computer interaction systems (ABCI) are widely used where displaying of advertisements is based on the estimated age of customers. Many companies need to determine a customer's age related with their products. Customers are classified according to their age into groups. This will help companies to marketing their products to the suitable denomination. ABCI can be used in the security area also where the vehicle can make an alarm when the child is left alone by mistake. In addition, ABCI can prevent child under safe age from playing a danger game in a theme park [2]. Also in bars, age estimation application can detect underage adults who try to drink. Age estimation system is also applied in demographic analysis for criminal prevention. It can estimate the age of offender captured by a monitoring camera. In health systems, Robotic nurse can get benefit from age analysis to speed the first aid [1].

Age estimation can be used in information retrieval. It allows image retrieval from a big image database by age image query. Flickr.com is a web site with billions of face images uploaded by regular users with different ages. It can make benefit from age estimation application in friend search and image identification [1]. So, it is clear that age estimation is useful in many applications.

The aging growth affects greatly the human appearance which involves a change in human face, hair, length, the way of walk… etc. Figure 1 shows aging changes on face appearance. In [3], they classify the factors affected the aging process into intrinsic factors and extrinsic factors. The intrinsic factors can be Health, living style..etc. While the weather conditions and surrounding environment can be extrinsic factors. Several computational techniques are used to solve age estimation problem whereas deep learning [4] [5] shown the best results than the others. Whereas there are vast of powerful algorithms indeed for age estimation, some challenges are still unsolved. Human aging process is uncontrollable as it differs from one person to another even human can fail in expecting the exact age for someone. Wrinkles have no specific age as they can appear on an adult with a bad habit of smoking or unhealthy food system. In some times, wrinkles can be delayed to appear on a healthy sportive person. A representative large dataset with accurate training labels is another challenge in training CNNs. A representative dataset should have a wide variation in race, gender and age groups to be more expressive.

In this survey, we aim at giving a summary of the most acknowledge work about deep age estimation techniques with presenting the challenges found, current solutions, famous datasets and the performance measures used to judge the performance of these techniques. This paper consists of seven sections. Section 2 states a brief discussion of age estimation. We introduce in section 3 the deep learning approach and its models. The literature works of estimating age using deep learning methods are summarized in section 4. We present in section 5 the most popular facial aging datasets used by deep learning algorithms. Section 6 states the performance measures used in the literature. Finally, conclusions and the new insights are discussed in section 7.

II. AGE ESTIMATION

At a recent time, human aging research has gotten big interest due to its extensive participation in vitality applications. There are three important fields centered on aging researches. First one is Age Invariant Face Recognition [3] which helps in identifying faces regardless the aging changes appeared on it (faces). It has been applied in real world for regeneration of passport or drive license and biometric authentication [3]. Second one is Age Synthesis which predicts the future look of a face image at certain age [2]. Age synthesis can be used in victim identification in police work. In addition, it can simulate the aging features of lost people from an old image. In figure 2, part (a) shows a simulated old face from younger one where part (b) shows a

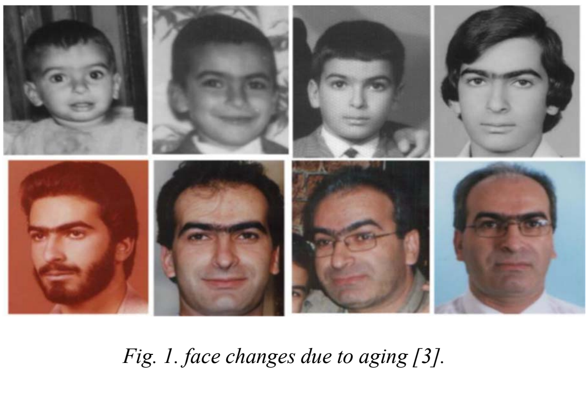
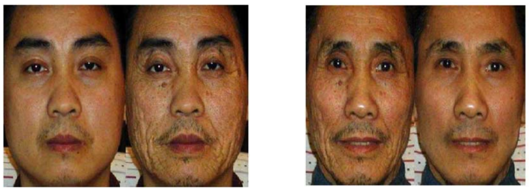


Fig. 1. face changes due to aging [3].



a- Simulated old face on the right.   b- Simulated young face on the right.

Fig. 2. Age progression or synthesis [6].

reverse process as a simulated young face from an older one.

The third field is age estimation which we concern with in this paper. Age estimation is used to determine the exact age with a computer-aid from face image [1]. It labels the face images with exact value of age or label group (child, adult or old). It comprises of feature extraction and age estimation stages. Many factors affect the estimation process like gender, ethnicity, wrinkles…etc. [2]. To estimate the age, firstly we extract the features then we use learning techniques to produce the estimated age or the suitable age group [1]. The main task of the learning techniques is degrading the difference between the actual age and the estimated one. Actual age refers to the real account of years that a human lives. The age that is realized by human based on the face look is called appearance age or perceived age. Estimated age is counted by the computational techniques [4].

Age estimation is a challenging problem as face characteristics change over time as hair, muscles and wrinkles. Aging is uncontrollable personalized process which differs from one person to another. Aging signs change face appearance depending on human life style, surrounding environment, race and gender. There is a study proved that wrinkles are appeared on women more and earlier than men [7]. Recent studies refer to variations in individuals' aging appearance due to differences in face deformation and emotional expression of each individual [7]. In the point of image processing, face images can differ in quality, lighting, pose and texture. These issues can be summarized as following:

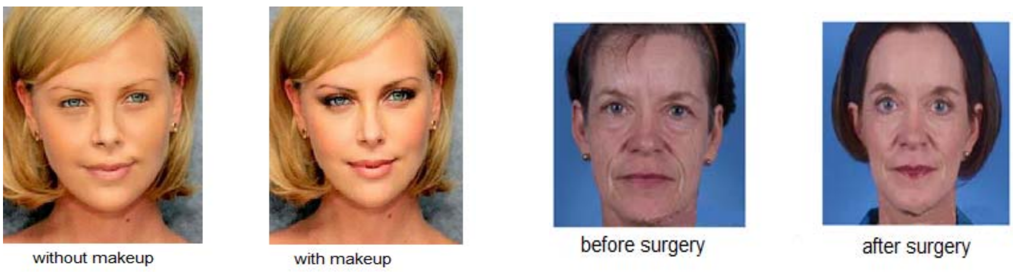
1- Age estimation depends on many factors such as gender, race, wrinkles…etc. These factors can be intrinsic or extrinsic. intrinsic factors can be gender, race, life style…etc. where extrinsic factors can be ultraviolet radiation of sun, disease, climate changes, ….etc. Aging is uncontrollable process which differs from one person to another.
2- Makeup, beauty care products and cosmetic surgeries can remove or decrease aging signs on face as shown in figure 3. This makes the age estimation process more difficult.
3- It is difficult to find a representative large dataset with accurate labels of images. There should be a low variation of pose, lighting, quality and texture in face images and a wide variation in race, gender and age groups to be more expressive. A sufficient number of face images for each individual should be introduced also to make the dataset more beneficial.
4- Describing aging periods with a sufficient number of images belong to the same individual is difficult to be accomplished even by using the internet.

III. DEEP CONVOLUTIONAL NEURAL NETWORK

Deep networks are an intelligent structure of traditional well-known neural networks that present a multimedia data with a complicated structure and good performance of results [8]. There are three types of deep networks: stacked autoencoders, deep belief networks and convolutional neural network (CNN). Recently, the most used one in feature extraction is CNN due to its high performance. CNN replaces handcrafted features by unsupervised or semi-supervised feature extraction [8]. CNN formed from three layers: input, output and one or more hidden layers in the middle. There are several components in the hidden layer to provide convolution and normalization to the data. Max-pooling and fully connected layers are examples of these components. In fact, the system performance depends mainly on the network architecture. Not only the utilization of deep networks but also the deep architectures play a main role in system performance. VGGNet [9] and GoogLeNet[10] are two examples of modern CNN architectures. In this section, VGG-16, GoogLeNet and DAG-CNN architectures are briefly discussed.

A. VGG-16 architecture

VGG-16 network [9] was developed by Oxford Visual Geometry Groups. It acts like a conventional convolution network with substantially increased depth.VGG-16 architecture consists of convolutional layers (up to 19 layers) ordered on top of each other with 3 × 3 receptive fields. The small size of receptive fields is beneficial for decreasing the number of parameters and making the decision function more distinctive. There are five hidden layers of max-pooling responsible for providing nonlinearity and spatial pooling. There are 1 × 1 convolution filters which introduced an additional non-linearity to the decision function. The spatial padding is preserved after convolution in some layers (not all layers). VGG-16 used 3 fully-connected layers with the same configuration. Figure 4 shows VGG-16 architecture.



a- Effects of makeup   *b- Effects of cosmetic surgery*

Fig.3. Effects of makeup and cosmetic surgery on face appearance [2].

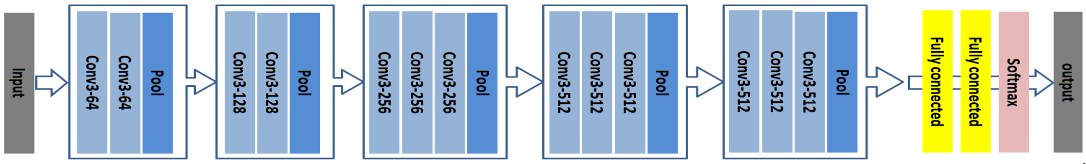


Fig. 4. VGG-16 architecture [5]

B. GoogLeNet architecture

The network architecture of GoogLeNet [10] is based on inception modules [10] but its architecture has more deep and wide layers. In GoogLeNet, the authors used average pooling in place of fully connected layers to refine the accuracy. In Inception network, the modules are ordered on top of each other where the output of one layer is clustered into groups of units and preserved as the input to the higher layers. There is an additional max-pooling layer in each stage. In figure 5, part (a) shows the inception modules with the alternative parallel pooling layer in each stage while part (b) declares dimension reductions using 1×1 convolutions. There is a tradeoff between the depth of network and gradients back propagation through the network. The network architecture of GoogLeNet consists of 22 layers with parameters only. With pooling layers, it can reach 27 layers deep. GoogLeNet is constructed from nine Inception modules divided into three groups. An extra linear layer is provided to enable fine-tuning with other labeled datasets. There is an auxiliary classifiers were added to overcome the gradient degrading problem by optimizing all parameters of low layers.

C. DAG-CNN architecture

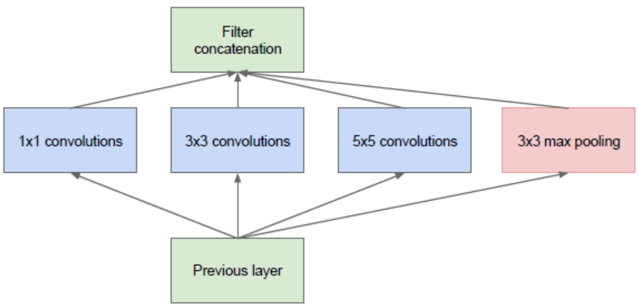
It stands for Directed Acyclic Graph CNN. The architecture of multi-scale DAG [11] is simply a feed forward CNN. This model uses learning algorithm based on gradient propagation to extract the multi-scale features. DAG-CNN is a traditional CNNs where consists of Convolution (conv.) layers, Multi-output layers (ReLU), normalization layers (Norm), pooling layers. The average pooling was performed to reduce features' dimension. ReLU layer is connected to an average pooling layer then it is normalized to feed to a fully-connected layer that computes an inner product with k outputs. The add layer is introduced to add the scores from all layers and then fed into the softmax function to produce the final predictions. To avoid the gradient degrading problem, the lower layers were directly connected with the output layer so a strong gradient signal was propagated through layers. Figure 6 illustrates multi-scale DAG-CNN architecture.

IV. Deep Learning-Based Age Estimation Techniques

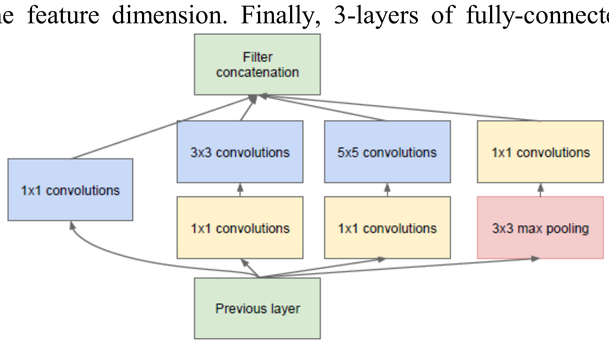
Over the past decade, deep learning techniques have been extensively exploited in several fields of image processing and produced high performance. Many researchers used deep learning approaches in estimating the human age because of its superior performance over the other methods (such as classification [12], regression [13], hierarchical age estimation [14] and ordinal ranking [15]). The process of estimating age is composed of feature extraction and fusion of classifiers' decision using learning algorithm. The early researches categorized the features into: local, global and hybrid [16]. Other researchers used image representation for age modeling that can be active shape, active appearance or hybrid models [17]. Active appearance model [18] was the extensively used one to learn shape and texture features from training images which uses principal component analysis [19] in dimension reduction. It was widely used in age estimation modeling which give precise results [13].

The other powerful model is Bio-Inspired Features (BIF) [20]. It represents the face images in a high dimension feature vector using simple (S) and complex (C) layers created by Gabor filters. Guo et al. applied BIF to estimate the human age for the first time in 2009 and produce a good performance [20]. In 2015, deep learning approach capture big attention from the age estimation community because of its superior performance over BIF and the previous methods.

In [21], Dong et al. proposed an end-to-end CNN system to estimating age from face images. They achieved MAE equals 3.63 years on MORPH2 dataset where BIF and Canonical correlation analysis in [22] achieved MAE of 3.98 years. Their CNN estimated the age from image directly by a deep learned end-to-end system. The parameters are learned instead of hand-crafted. Due to the lack of large labeled datasets, deep learning approach is used in feature extraction rather than classification. So, Yang et al. [23] tended to combine ordinal regression with a CNN to predict the age as a first one in this field. The model composed of 3-layers ScatNet to extract the features then they used PCA to reduce the feature dimension. Finally, 3-layers of fully-connected



(a) Naïve form



(b) Dimension reductions

Fig.5. Inception modules [10]

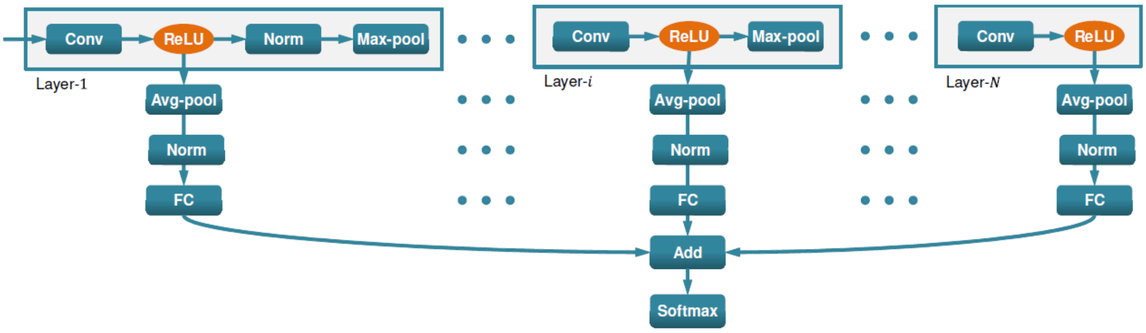


Fig. 6. Multi-scale DAG-CNN architecture [11]

network produced the final decision. ScatNet is a deep convolutional network with special characteristics. They minimized MAE as it equals 3.49, 5.19 and 7.04 years on MORPH, Lifespan, and FACES respectively. The training time is reduced by using ScatNet as it requires only the fully-connected layers to be learned. But in the other side, the computational cost may increase due to the hierarchical strategy of ordinal regression that involved more layers to be introduced.

Wang et al. investigated a new approach to extract features based on deep learning [24]. They used CNN to extract the features and Support Vector Regression (SVR) to learn aging patterns. They didn't concern with the influence of gender and ethnicities where they used only the age information. In [25], Chang and chen reduced MEA to 3.74 years on the large dataset MORPH Album 2 and outperformed the compared methods. They proposed an age ranking approach, CSOHR which based on ordinal regression. They didn't consider the changes of neutral facial expressions which have effects on age estimation.

For the first time, Han et al. showed that a machine bit the human performance according to a crowdsourced study [26]. Their approach acted better than human on MORPH II and PCSO datasets in race, gender and age estimation tasks. They proposed a generic framework enabling machine to reduce MAE on MORPH II to 3.49 years while 4.3 years for human. They used BIF algorithm for extracting the demographic features from a face image. They used a hierarchical approach where classification stages were followed by a regression stage for demographic estimation. Although they outperformed human, they showed poor results in some samples of PCSO and MORPH databases. This raises the need of fine-tuning with a large representative database to improve the feature extraction stage and solve the over fitting problem.

The estimated age can be modeled into age groups or an exact age. In 2016, Rothe et al. tended to determine the age group before detecting the exact value. Estimating the exact age value from a narrow range is more accurate and fast than from a wide range of all ages. Rothe et al. proposed a deep age estimation method called DEX which refers to Deep Expectation of apparent age [27]. Without the use of facial landmarks, DEX predicted the exact age by first classifies images into age groups then uses a softmax expected value to refine the final decision of exact age. DEX achieved MEA of 3.25 years and 4.63 years on MORPH2 and FG-NET respectively. DEX achieved MEA of 2.68 years and 3.09 years on MORPH2 and FG-NET respectively with fine-tuning on IMDB-WIKI. DEX with a single CNN failed to estimate the age in some cases such as dark images, wearing glasses, old photographs.

Niu et al. [28] preferred Ordinal Ranking-CNN (OR-CNN) than softmax classifier to address age estimation problem as it can be transformed into sub-problems of binary classification. OR-CNN is based on OHRank [29] and deep networks for age estimation. They trained CNN using binary ordinal age labels, one for each age group. The ordinal regression was converted to sub-problems. Each sub-problem was attached to the input layer of multiple output CNN to solve it. Their experiment achieved MEA of 3.27 years on MORPH and 3.34 years on AFAD dataset. In contrast, the produced features had large dimension which affecting the performance and consuming large time. In [4], the proposed approach achieved MEA of 2.96 years on MORPH. The authors proposed a ranking CNN-based framework which was trained with ordinal age labels. The age is predicted by aggregating the binary outputs of these CNNs in the fully connected layers. The training of CNN models is consuming large time because of the existence of many sub-models.

Most researches concerned with faces with only neutral expressions and this affect the estimation accuracy. Researches proved the interrelationships between age and appearance expressions. Yang et al. [30] returned and enhanced their estimation accuracy by considering facial expression in estimation process. They estimated the age and facial expression independently as two tasks. Their framework consists of two parallel columns; CNN and ScatNet; to extract features from image. Then they fed into two fully connected layers to estimate age and classify facial expression. But their approach was failed to recognize the correct expression in some cases. This may be resulted from ignoring the race and gender in the estimation process.

In [31], ordinal deep feature learning was used to deal directly with raw pixels for extracting features. They presented an ordinal deep learning approach where the feature extraction and estimating age processes were optimized using back-propagation. They showed that ODFL achieved the efficiency even with the mislabeling training data because they used the correlation between face aging features, instead of considering little label correlation into account. In [32], the authors used a deep residual network to detect the difficult mislabeled samples and deal with label-similarity problems. They accomplished that by combining deep residual networks with OHRANK. They achieved MEA of 3.08 years on MORPH and 2.89 years when using multi-source datasets. Although this approach claimed reducing the consuming time but it did not use the nonlinear information of face features with different facial expressions and ethnicity.

As a matter of fact, the aging progress is different between man and woman as some recent studies stated[7]. For example, wrinkles are appeared on women more and earlier than men. Recent works tend to determine gender before estimating age. In [33], dropout-SVM was used by the authors in the training step of traditional support vector classifiers to avoid over-fitting problem. Firstly, they aligned the face image before extracting features to determine gender and age of each sample. They used Adience dataset with unfiltered images captured in unconstrained environment. They achieved 95.3 % 1-off accuracy on Adience dataset.

In 2019, liu et al. avoided the deep convolution neural networks due to their heavy computing and learning [34]. They considered a specified framework of feature extraction (scattering transform) and estimating algorithm (support vector machine). They investigated a framework composed of four processes defined as follows : 1) classification of gender, 2) detect age label based on gender classification, 3) determine the exact age value from the age label and 4) the fusion stage. They achieved a reduced MEA of 3.92 years on FG-NET dataset compared with PLO [35], NDF [36] and DEX [27]. In this approach, the computational cost is increased due to the hierarchical strategy and the race effect wasn't concerned. After a while, Taheri et al. outperformed the pervious methods by achieving MEA of 2.81 by using DAG-VGG-16 and 2.87 years by DAG-GoogLeNet on Morph-II [37]. They utilized multi-scale features of face images to be collected in the fusion stage by two ways. The first method uses handcrafted parameters on each layer to be fused using fully-connected layers (facial expression and wrinkle). The other method generates a score vector from each layer representing the given features then aggregates these vectors to generate the final decision. This score vector has large dimension according to the fusion stage. This is hard to handling as it requires more space, more training data and processing time.

Some studies tended to combine race with gender as they have an influence on age estimation process. In [5], a new ensemble structure was proposed called RAGN which used three networks to determine age, gender and race. RAGN used CNN to extract features of race and gender to limit the range values of age. Extreme learning machine was used to decide the final result and estimate ages. They achieved MEA of 2.61 years on Morph-II. Although, they considered the gender and race properties but the varied facial poses and neutral expression weren't mentioned.

The main obstacle in deep learning approach is the missing of textual representations of images in massive image dataset. In 2016, Dong et al. [38] used transfer learning strategy in training a deep network to predict the unlabeled images. By using the deep networks, the age labels are classified, ordered and the relationships between them were addressed using a loss function. Only the age difference information was used in the loss function to predict the age by the deep network regardless the race and gender. In [39], the authors overcome the lack of accurate labels by converting categorical labels into discrete labels considering the age and gender. They proposed a conditional multitask deep learning approach. Firstly, they computed the probability of the image I is of an age a1 under condition the gender type is g1. Secondly, they introduced an expansion algorithm for unlabeled or mislabeled samples where classified labels were converted to discrete labels. Considering the race information may enhance the performance.

In [40], the authors tended to pre-train their CNNs on large benchmarks to refine the extracted features from unlabeled images. So, CNNS were fine-tuned with Adience dataset which contains unfiltered images captured in unconstrained environment. They tended to increase the depth of CNN till reached eight layers. In addition, there are 3- layers of fully connected type to increase the performance and solve the over fitting problem. They showed that their method outperformed other approaches with 92.63% 1-off accuracy. The architecture of network is compact and complex as they increase the depth of network. In addition, the gradient vanishing problem appeared with increasing depth of network architecture.

Another inspired contribution, the authors transferred learned aging features from one population dataset with large labels to another population with small amount of labeled data [41]. The authors trained a deep CNN on the source population with large sized labeled dataset to extract transferable low-level aging features that can be learned to the target population with weak labels. MEA was 3.18 years when source population was black men and the target was white female on Morph II. MEA was 3.13 years when source population was white men and the target was white female on Morph II. In other words, the estimation accuracy depended on the feature-relation between source and target population. Table 1 presents a comparison between the literature works of age estimation based on deep learning methods. From figure 7, the recent researches are based on deep learning rather than regression and classification methods. Deep learning is mostly used to extract features or

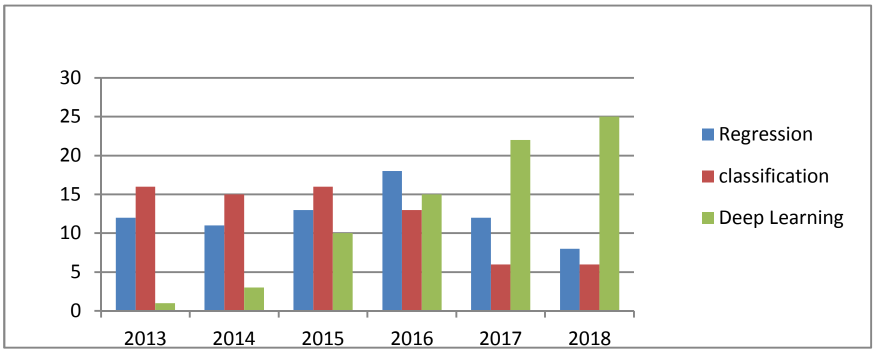


Fig. 7. Comparison between classification, regression and deep learning methods in number of researches [16]

TABLE 1. COMPARISON BETWEEN DIFFERENT FACE AGE ESTIMATION ALGORITHMS BASED ON DEEP LEARNING

Used datasets	Authors	Used algorithm	MEA		CS %	1-off %
MORPH	Dong et al, 2014	fully learned end-to-end system based on CNN	3.63		-	-
	Yang et al.,2015	CNN with ordinal regression	3.49		-	-
	Wang et al, 2015	CNN with Support Vector Regression	4.77		-	-
	Niu et al., 2016	OR-CNN	3.27		-	-
	Chen et al., 2017	ranking CNN-based	2.96		-	-
	Liu et al., 2017	ODFL	3.12		-	-
	Liu et al., 2017	LSDML	3.08	2.89 using multi-source datasets	96	-
MORPH2	Chang and Chen, 2015	CSOHR	3.74		-	-
	Han et al., 2015	boosting algorithm + regression classification	3.49		-	-
	Rothe et al., 2016	DEX	3.25	2.68 with fine-tuning on IMDB-WIKI	-	-
	Duan et al., 2017	RAGN	2.61		-	-
	Yoo et al., 2018	weak label expansion method	2.91		99.3	92.63
	Li kai et al., 2018	deep CNN	3.13		-	-
	Yang Yi. et al, 2018 [42]	SSR-Net (Soft Stagewise Regression Network)	3.16		-	-
	Zighem et al,2019 [43]	TSE (Two-stages age estimation)	3.21		-	-
	Taheri et al., 2019	DAG-CNNs	2.81 DAG-VGG-16		-	-
FG-NET	Wang et al, 2015	CNN with Support Vector Regression	4.26		-	-
	Rothe et al., 2016	DEX	4.63	3.09 with fine-tuning on IMDB-WIKI	3.09	-
	Yoo et al., 2018	weak label expansion method	3.43		99.3	92.63
Lifespan	Yang et al.,2015	CNN with ordinal regression	5.19		-	-
	Yang et al., 2018	CNN and ScatNet	4.01		-	-
FACES	Yang et al.,2015	CNN with ordinal regression	7.04		-	-
Asian face age	Niu et al., 2016	OR-CNN	3.34		-	-
PAL	Zighem et al,2019 [43]	TSE (Two-stages age estimation)	4.49		-	-

fine-tune the CNN with large datasets due to the missing of labeled datasets.

V. POPULAR DATASETS

The efficiency of age estimation algorithm depends highly on the database used in training and testing. In this section, we summarize the facial aging datasets that are public and utilized by deep learning. The desired dataset should have low variation in conditions such as pose, lighting and texture. In addition, it should contain different race, gender and age groups to be more expressive. The desired dataset indeed should have a sufficient number of face images for each individual in different ages to be more beneficial [7]. Some of popular datasets used recently in age estimation algorithms are FG-NET [44], FERET [45], PAL [46] MORPH [47], Cross-Age Celebrity Dataset (CACD) [48], Adience [33] and LAP [49]. Figure 8 shows face images from some popular datasets.

a) FG-NET [44] contains 1002 color and gray scale images with age range from infants to old persons during different aging stages. It contains multi-race subjects with information about exact age, ethnicity and facial expression. There are different illumination, pose and background. Some noise is appeared as some images are scanned.

b) FERET [45] contains 14,126 gray scale face images that include 1199 individuals. There is over two years between the first and last images of some individuals. The quality of images is better as they were collected in a controlled environment but with little age variations. There are a variation of pose, race and gender. This makes FG-NET and MORPH better than it.

Fig. 8. Samples of datasets (a) FG-NET, (b) FERET, and (c) MORPH II [26]

c) PAL [46] contains 1,142 face images ranging from 19 to 93 years (225 males and 350 females). The images have good resolution and they represent wide range of ages which is benefit for age estimation algorithms.

d) MORPH [47] is the largest facial database. There are two sets: Album 1 and Album2. Album 1 is small compared with Album 2. It contains 1690 face images. It includes information about ethnicity, gender, age, individual‘s height and weight. Album 2 includes 78,207 images. There is a variation in expression, illumination and resolution. The age range is wide from 15 to 80 years. Although the face images per individual are not representative enough but it is the most popular one used in aging researches.

e) Cross-Age Celebrity Dataset [48] includes images of 2000 celebrities. It may not have by necessary age progression images for all celebrities or it may be inaccurate. The images are collected from internet across ten years.

f) Adience [33] is unconstrained datasets. It is labeled for age and gender. The images are captured in unconstrained environment with different poses and scene conditions from real life. It includes 26,580 images. It includes information about age, gender and label classifications.

g) Apparent Age Estimation dataset (LAP) [49] was introduced in 2016. It consists of 7,591 images with unconstrained environment. The images were labeled by human observers. The observers’ votes on each image are collected for computing the mean and the standard deviation to determine the final label of the image. Unconstrained datasets have been recently introduced and extensively used in aging research.

VI. Performance Measures

For performance evaluation, there are different metrics for evaluating the age estimation systems.

a) Mean Absolute Error (MAE):

It can be defined as the summation of errors between the predicted age and the actual age divided by the number of testing images [27] [37]. It is the most popular one utilized to evaluate the performance. It is defined by equation (1) as follows:

$$\text{MEA} = \frac{1}{N}\sum_{k=0}^{N} a^k - s^k \ , \qquad (1)$$

where N refers to the number of testing samples, a refers to actual age of a sample k and s refers to estimated age of the same sample. Some authors tend to evaluate the performance of age estimation algorithms in specific stage or decade [20]. MEA does not label accurately the error in the voted ground truth datasets such as lAP [27]. The cumulative score (CS) is driven from MEA which computes the proportional of test images that have absolute errors less than specific number of years. CS is defined by equation (2) as follows:

$$\text{CS(L)} = \frac{n^L}{N} \times 100\ \% \ , \qquad (2)$$

where n^L refers to the number of testing samples that have errors less than specific number of years (L).

b) Normal Score (ε):

As the LAP data set was labeled by several observers, the mean and standard deviation were computed of labels for each sample. The ε score is better than MEA to measure the accuracy of the labeled ground truth datasets by taking into account the variation of the observers’ vote for each image. It can be defined by equation (3):

$$\varepsilon = 1 - e^{-\left(\frac{(i-\mu)^2}{2\sigma^2}\right)}, \qquad (3)$$

where i refers to the tested image, μ refers to the mean of the observers’ vote for i and σ refers to the standard deviation of the observers’ vote for i. Therefore, the value of ε score can be between 0 and 1[37].

VII. Conclusion and Future Directions

In this survey, an exhaustive summary of the recent age estimation frameworks based on deep networks was introduced. Feature extraction and learning algorithm are the main processes of any age estimation system. In general, deep learning present multimedia data with a complicated structure with high levels of abstraction and good performance of results. CNN replaces handcrafted features by unsupervised or semi-supervised feature extraction.

Although the superior performance yielded from using deep learning approach but there are some obstacles to be solved in the future. The lack of accurate training labels has an influence on the performance. Deep learning needs a representative large dataset for training the deep convolution neural networks. A suitable dataset should have varsity in race, gender and age range. A low variation in pose, lighting and environment changes of face images will accelerate and empower the age estimation system. There should be sufficient number of face images for each individual so the dataset will be more expressive.

The deep convolution neural networks consume high cost of computing and learning. Alignment of face images can be a step before feature analysis to speed up processing. The dimension of extracted features needs to be reduced to release the reduction step and save memory and processing time. To refine the age estimation system, there are directions to use CNN with another technique instead of end to end system based on CNN. Scattering transform, transfer learning, ordinal ranking and CBIF can be combined with deep learning.

REFERENCES

[1] R. Angulu, J. Tapamo, and A. Adewumi, "Age estimation via face images: a survey," EURASIP Journal on Image and Video Processing 2018, Springer International Publishing, vol. 42, June 2018.

[2] M. Yehia, Automatic facial age estimation, Ph.D. dissertation, Faculty of Computers & Information, Cairo University, Cairo, Egypt, 2011.

[3] M. Sawant and K. Bhurchandi, "Age invariant face recognition: a survey on facial aging databases, techniques and effect of aging," Artificial Intelligence Review 52.2, Springer, pp. 981-1008, October 2019.

[4] Sh. Chen, C. Zhang, M. Dong, J Le, and M. Rao, "Using ranking-cnn for age estimation," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5183-5192, 2017.

[5] S. Taheri, and Ö. Toygar, "On the use of DAG-CNN architecture for age estimation with multi-stage features fusion," Neurocomputing, vol. 329, pp. 300-310, 2019.

[6] Z. Liu, Z. Zhang, and Y. Shan, "Image-based surface detail transfer," IEEE Computer Graphics and Applications, vol. 24, no. 3, pp. 30-35, 2004.

[7] K. Sveikata, I. Balciuniene, and J. Tutkuviene, "Factors influencing face aging. literature review," Stomatologija, vol. 13, no. 4, pp. 113–115, 2011.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015.

[9] H.A. Song, and S. Y. Lee, Hierarchical Representation Using NMF, Neural Information Processing. Lectures Notes in Computer Sciences, Springer Berlin Heidelberg, pp. 466–473, 2013.

[10] K. Simonyan, and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv preprint arXiv:1409.1556 ,2014.

[11] S.Yang, D. Ramanan, "Multi-scale recognition with DAG-CNNs," In Computer Vision (ICCV), 2015 IEEE International Conference, pp. 1215-1223, 2015.

[12] Y. Zheng, H.Yao,Y. Zhang, and P. Xu, "Age classification based on backpropagation network," ACM 5th Int. Conf. Internet Multimedia Comput. Serv., pp. 319–322, 2013.

[13] K. Luu, K. Ricanek Jr, T. D. Bui, and C. Y. Suen, "Age estimation using active appearance models and support vector machine regression," in Proc. 3rd Int. Conf. Biometrics, Theory, Appl., Syst., pp. 1–5, 2009.

[14] W. Li, Y. Wang, and Z. Zhang, "A hierarchical framework for imagebased human age estimation by weighted and OHRanked sparse representation-based classification," 5th IAPR Int. Conf. Biometrics, pp. 19–25, 2012.

[15] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, "Ranking with uncertain labels," in Proc. IEEE Int. Conf. Multimedia Expo., pp. 96–99, 2007.

[16] O. F. Osman, and M. H. Yap, "Computational intelligence in automatic face age estimation: A survey," IEEE Transactions on Emerging Topics in Computational Intelligence, pp. 271-285, 2018.

[17] R. Angulu, J. R. Tapamo, and A. O. Adewumi,"Age estimation via face images: a survey," EURASIP Journal on Image and Video Processing, vol. 42, December 2018.

[18] T. F. Cootes, E. J. Gareth, and T. J. Christopher, "Active appearance models," IEEE Transactions on Pattern Analysis & Machine Intelligence, vol. 6, pp. 681-685, 2011.

[19] I. Jolliffe, Principal component analysis, Springer, Berlin Heidelberg, 2011.

[20] G. Guo, G. Mu, Y. Fu , and T. S. Huang, "Human age estimation using bio-inspired features," IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Miami, FL, USA, pp. 112–119, 2009.

[21] D. Yi, L. Zhen, and Z. L. Stan , "Age estimation by multi-scale convolutional network," Asian conference on computer vision, Springer, Cham, 2014.

[22] G. Guo, and G. Mu, "Joint estimation of age, gender and ethnicity: CCA vs. PLS," 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), Shanghai, China, pp. 1–6, May 2013.

[23] H. Yang, B. Lin, K. Chang, and C. Chen, "Automatic age estimation from face images via deep ranking," In Proc. BMVC, vol. 55, pp.1–55, 2015.

[24] X Wang, R Guo, and C Kambhamettu, "Deeply-learned feature for age estimation," in Proceedings of IEEE Winter Conference on Applications of Computer Vision, Waikoloa, pp. 534–541, 2015.

[25] K. Chang, and C. Chen. "A learning framework for age rank estimation based on face images with scattering transform," IEEE Transactions on Image Processing, vol. 24, pp. 785-798, 2015.

[26] H. Han, C. Otto, X. Liu, and A. K. Jain, "Demographic estimation fromface images: Human vs. machine performance," IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, pp. 1148–1161, 2015.

[27] R. Rothe, R. Timofte, and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks,"International Journal of Computer Vision, vol. 126, no. 2-4, pp. 144–157, 2018.

[28] Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua,"Ordinal regression with multiple output CNN for age estimation," In Proc. CVPR, pp. 4920–4928, 2016.

[29] K. Chang, C. Chen, and Y. Hung,"Ordinal hyperplanes ranker with cost sensitivities for age estimation," In Proc. CVPR, pp. 585–592, 2011.

[30] H. F. Yang, B. Y. Lin, K. Y. Chang, and C. S. Chen, "Joint estimation of age and expression by combining scattering and convolutional networks," ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 14, January 2018.

[31] H. Liu, J. Lu, J. Feng, and J. Zhou, "Ordinal Deep Learning for Facial Age Estimation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, pp. 486 – 501, 2017.

[32] H. Liu, J. Lu, J. Feng, and J. Zhou, "Label- sensitive deep metric learning for facial age estimation," IEEE Transactions on Information Forensics and Security, vol. 13, no. 2, pp. 292-305, 2017.

[33] E. Eidinger, R. Enbar, and T. Hassner, "Age and gender estimation of unfiltered faces," IEEE Transactions on Information Forensics and Security, vol. 9, no. 12, pp. 2170-2179, 2014.

[34] K. Liu, and T. Liua, "A Structure-Based Human Facial Age Estimation Framework under a Constrained Condition," IEEE Transactions on Image Processing, vol. 28, pp. 5187 – 5200, October 2019.

[35] C. Li, Q. Liu, J. Liu, and H. Lu, "Learning ordinal discriminative features for age estimation," in Computer Vision and Pattern Recognition (CVPR), IEEE Conference, pp. 2570–2577, 2012.

[36] N. Fan, "Learning nonlinear distance functions using neural network for regression with application to robust human age estimation," Computer Vision (ICCV), IEEE International Conference, pp. 249–254, 2011.

[37] M. Duan, K. Li, and K. Li, "An ensemble cnn2elm for age estimation," IEEE Transactions on Information Forensics and Security, vol. 13, no. 3, pp. 758-772, 2017.

[38] Y. Dong, Y. Liu, and S. Lian, "Automatic age estimation based on deep learning algorithm," Neurocomputing, vol. 187, pp. 4-10, 2016.

[39] B. Yoo, Y. Kwak, Y. Kim, C. Choi, and J. Kim, "Deep facial age estimation using conditional multitask learning with weak label expansion," IEEE Signal Processing Letters, vol. 25, pp. 808-812, 2018.

[40] A. A. Mallouh, Z. Qawaqneh, and B. D. Barkana, "Utilizing CNNs and transfer learning of pre-trained models for age range classification from unconstrained face images," Image and Vision Computing, vol. 88, pp. 41-51, 2019.

[41] K. Li et al., "Deep cost-sensitive and order-preserving feature learning for cross-population age estimation," Proceedings of the

IEEE Conference on Computer Vision and Pattern Recognition, pp. 399-408, 2018.

[42] T. Y. Yang, Y. H. Huang, Y. Y. Lin, P. C. Hsiu, and Y. Y. Chuang, "SSR-Net: A Compact Soft Stagewise Regression Network for Age Estimation," IJCAI, vol. 5, no. 6, July 2018.

[43] A. Ouafi, A. Zitouni, Y. Ruichek, and A. Taleb-Ahmed "Two-stages based facial demographic attributes combination for age estimation," Journal of Visual Communication and Image Representation, vol. 61, pp. 236-249, May 2019.

[44] J. L. Crowley, and T. Cootes, FGNET, Face and Gesture Recognition Working Group, 2009, http://www-prima.inrialpes.fr/FGnet/

[45] P. J. Phillips, S. Z. Der, P. J. Rauss, and O. Z. Der, FERET (face recognition technology) recognition algorithm development and test results. Adelphi, MD: Army Research Laboratory, 1996.

[46] M. Minear and D. C. Park, “A lifespan database of adult facial stimuli,” Behavior Research Methods, Instruments, & Computers, vol. 36, no. 4, pp. 630–633, 2004.

[47] K Ricanek, and T Tesafaye, “Morph: A longitudinal image database of normal adult age-progression,” in Proceedings of IEEE 7th International Conference on Automatic Face and Gesture Recognition,IEEE, Southampton, pp. 341–345, 2006.

[48] B.C. Chen, C.S. Chen, and W. H. Hsu, “Cross-age reference coding for age-invariant face recognition and retrieval,” European conference on computer vision, Springer, Cham, pp. 768–783, 2014.

[49] S. Escalera, X. Baro, H. J. Escalante, and I. Guyon, “ChaLearn Looking at People: A Review of Events and Resources,” Proceedings of the International Joint Conference on Neural Networks (IJCNN), IEEE, 2017.