Using Age Information as a Soft Biometric Trait for Face Image Analysis

Haoyi Wang, Victor Sanchez, Wanli Ouyang, Chang-Tsun Li

Abstract Soft biometrics refers to a group of traits that can provide some information about an individual but are inadequate for identification or recognition purposes. Age, as an important soft biometric trait, can be inferred based on the appearance of human faces. However, compared to other facial attributes like race and gender, age is rather subtle due to the underlying conditions of individuals (i.e., their upbringing environment and genes). These uncertainties make age-related face image analysis (including age estimation, age synthesis and age-invariant face recognition) still unsolved. Specifically, age estimation is concerned with inferring the specific age from human face images. Age synthesis is concerned with the rendering of face images with natural ageing or rejuvenating effects. Age-invariant face recognition involves the recognition of the identity of subjects correctly regardless of their age. Recently, thanks to the rapid development of machine learning, especially deep learning, age-related face image analysis has gained much more attention from the research community than ever before. Deep learning based models that deal with age-related face image analysis have also significantly boosted performance compared to models that only use traditional machine learning methods, such as decision trees or boost algorithms. In this chapter, we first introduce the concepts and theory behind the three main areas of age-related face image analysis and how they can be used in practical biometric applications. Then, we analyse the difficulties involved in these applications and summarise the recent progress by reviewing the state-of-the-art methods involving deep learning. Finally, we discuss the future research trends and

Haoyi Wang
University of Warwick, Coventry, UK, e-mail: h.wang.16@warwick.ac.uk

Victor Sanchez
University of Warwick, Coventry, UK, e-mail: v.f.sanchez-silva@warwick.ac.uk

Wanli Ouyang
University of Sydney, Sydney, Australia, e-mail: wanli.ouyang@sydney.edu.au

Chang-Tsun Li
University of Warwick, Coventry, UK, e-mail: c-t.li@warwick.ac.uk
Deakin University, Melbourne, Australia, e-mail: changtsun.li@deakin.edu.au

the issues that are not addressed by existing works. We also discuss the relationship among these three areas and show how solutions within one area can help to tackle issues in the others.

Keywords Soft Biometrics · Age Estimation · Age Synthesis · Age-Invariant Face Recognition · Facial Analysis · Deep Learning · Convolutional Neural Network

1 Introduction

Biometrics aim to determine the identity of an individual by leveraging the users' physiological or behavioural attributes [23]. Physiological attributes refer to the physical characteristics of the human body, like the face, iris, fingerprint, etc. On the other hand, behavioural attributes indicate the particular patterns of the behaviour of a person, which include gait, voice, keystroke dynamics, etc. Among all these biometrics attributes, the face is the most commonly used one due to its accessibility and the fact that face-based biometric systems require little cooperation from the subject.

Besides the identity information, other ancillary information like age, race and gender (often referred to as soft biometrics) can also be retrieved from the face. Soft biometrics is the set of traits that provide some information to describe individuals, but do not have the capability to discriminate identities due to their lack of distinctiveness and permanence [22]. Although soft biometric traits alone cannot distinguish among individuals, they can be used in conjunction with the identity information to boost the recognition or verification performance or be leveraged in other scenarios. For example, locating persons-of-interest based on a combination of soft biometric traits by using surveillance footage.

Compared to traditional biometrics, soft biometrics have the following merits. First, when the identity information is not available, soft biometrics can generate human-understandable descriptions to track the person-of-interest, such as in the 2013 Boston bombings [24]. Second, as the data abuse issue becomes more and more severe in the information age, using soft biometric traits to capture subjects' ancillary information can preserve their identity while achieving the expected goals. For example, companies can efficiently recommend merchandises by merely knowing the age or the gender of their potential customers. Third, collecting soft biometric traits do not require the participation of the subject, which makes them easy to compute.

Among all the soft biometric traits (age, gender, race, etc.) that can be obtained from face images, in this chapter, we focus on the age as it attracts the most attention from the research community, and can be used in various real-life applications. Specifically, the age-related face image analysis encompasses three areas: estimating the age (age estimation), synthesising younger or elder faces (age synthesis), and identifying or verifying a person across a time span (age-invariant face recognition). As to their real-life applications, the age estimation models can be widely embedded

into the security control and surveillance monitoring applications. For example, such systems can run age estimation algorithms to prevent teenagers from purchasing alcohol and tobacco from vending machines or access adult-exclusive content on the Internet. The age synthesis models can be used, for example, to predict the outcome of cosmetic surgeries, and generate special visual effects on characters of video games and films [12]. The age-invariant face recognition models can be used to efficiently track persons-of-interest like suspects or missing children over a long time span. Although the age-oriented face image analysis models can be used in a variety of applications, due to the underlying conditions of the individuals, such as their upbringing environment and genes, there are still several issues that remain unsolved. We will discuss these issues in the next section.

After Krizhevsky *et al.* [25] demonstrated the robustness of the deep convolutional neural network (CNN) [27, 26] on the ImageNet dataset [10], CNN based models have been widely deployed in computer vision and biometrics tasks. Some well-known CNN architectures are AlexNet [25], VGGNet [47], ResNet [17], and DenseNet [21]. In this chapter, we only focus on the CNN based models for age-related face image analysis and discuss their novelties and limitations.

To provide a clear layout, we present the three areas of age-related face image analysis in individual sections. For each area, we first introduce its basic concepts, the available datasets and the evaluation methods. Then, we present a comprehensive review of recently published deep learning based methods. Finally, we discuss the future research trends by discussing the unaddressed issues in the existing deep learning based methods.

2 Age Estimation

As the name suggested, the purpose of age estimation is to estimate the real age (cumulated years after birth) of the individual. The predicted age is mainly deduced based on the age-specific features extracted by the feature extractor. Since CNNs are powerful tools for extracting features, state-of-the-art age estimation methods are CNN-based. A simple block diagram of a deep learning based age estimation model can be found in Figure 1.

The first step in a deep learning based age estimation model is the face detection and alignment as the input image can contain other objects other than the face and a large amount of background. This step can be achieved by either a traditional computer vision algorithm like the Histogram of Oriented Gradients (HOG) filter or a state-of-the-art face preprocessing model like a deep cascaded multi-task framework [56]. After the face is cropped from the original image, and normalised (the mean value is subtracted), it is fed into the CNN backbone to estimate the age. In order to attain a good performance, the CNN is often designed to employ one or more loss functions to optimise its parameters. We will see later in this section that the recent age estimation models either involve advanced loss functions or change the network architecture to improve performance.

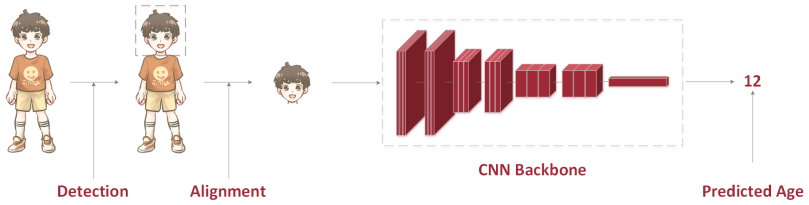

Figure 1: A simplified diagram of a deep learning based age estimation model. Since we are only interested in the face region, the face should be located and aligned from the original image before fed into the CNN model. *Illustration by Tian Tian.*

2.1 Datasets for the Age Estimation

Among all the age-oriented datasets, the MORPH II dataset [44] is the most broadly used to evaluate age estimation models. This dataset contains more than 55,000 face images from about 13,000 subjects with ages ranging from 16 to 77 with an average age of 33. Each image in the MORPH II dataset is associated with identity, age, race and gender labels. The second most commonly used dataset to evaluate age estimation models is the FG-NET [9] dataset which contains 1002 images from 82 subjects. However, due to the limited number of images, the FG-NET dataset is usually only used during the evaluation phase. Since the training of CNN-based models requires a large number of training samples, to meet this requirement, two large-scale age-oriented datasets have been built, the Cross-Age Celebrity Dataset (CACD) [7] and the IMDB-WIKI dataset [45]. The CACD contains more than 160,000 face images from 2000 individuals with ages ranging from 16 to 62. The IMDB-WIKI dataset contains 523,051 face images (460,723 images from IMDB and 62,328 images from Wikipedia) from 20,284 celebrities. However, both datasets contain noisy (incorrect) labels. The details of these four datasets are tabulated in Table 1.

Table 1: Most commonly used datasets to evaluate age estimation models.

Dataset	#images	#subjects	age range	noise-free label	Mugshot
MORPH II	55,134	13,618	16-77	Yes	Yes
FG-NET	1,002	82	0-69	Yes	No
CACD	163,446	2000	16-62	No	No
IMDB-WIKI	523,051	20,284	0-100	No	No

2.2 Evaluation Metrics for Age Estimation Models

There are two evaluation metrics commonly used for age estimation models. The first one is the Mean Absolute Error (MAE), which measures the average absolute difference between the predicted age and the ground truth:

$$MAE = \frac{\sum_{i=1}^{M} e_i}{M}, \tag{1}$$

where e_i is the absolute error between the predicted age $\hat{l}_i$ and the input age label l_i for the i-th sample. The denominator M is the total number of testing samples.

The other evaluation metric is the Cumulative Score (CS), which measures the percentage of images that are correctly classified in a certain range:

$$CS(n) = -\frac{M_n}{M} \times 100\%, \tag{2}$$

where M_n is the number of images whose predicted age $\hat{l}_i$ is in the range of $[l_i - n, l_i + n]$, and and n indicates the number of years.

2.3 Deep Learning Based Age Estimation Methods

Due to the appearance differences among different images of the same individual, extracting age-specific features and predicting the precise age can be onerous. Due to the extraordinary capability of CNN for feature extraction, [50] first employ a CNN to tackle the age estimation problem. In [50], the authors design a two-layer CNN to extract the age-specific features and use manifold learning algorithms (Support Vector Regression (SVR) and Support Vector Machines (SVMs)) to compute the final output. Their results show a dramatic improvement on the MORPH II dataset compared to the methods that use traditional machine learning [13, 57, 5].

As aforementioned, recent deep learning based attempts for age estimation can be classified into two categories. The first category is about improving the accuracy by leveraging customised loss functions rather than using conventional classification loss functions, such as the cross-entropy loss. The second category boosts the estimation performance by modifying the network architecture of a plain CNN model. We first review the recent age estimation works based on these two categories. Then, we discuss some works that involve multi-task learning frameworks to learn age information along other tasks.

2.3.1 Customised Loss Functions for Age Estimation

Traditionally, the age estimation problem can be treated as a multi-class classification problem [39] or a regression problem [37]. Rothe *et al.* [45] propose a formulation

that combines regression and classification for this particular task. Since age estimation usually involves a large number of classes (approximately 50 to 100) and based on the fact that the discretisation error becomes smaller for the regressed signal when the number of classes becomes larger, they compute the final output value by using the following equation:

$$\mathbb{E}(O) = \sum_{i=1}^{n} p_i y_i, \tag{3}$$

where O is the output from the final layer of the network after a softmax function, y_i is the discrete year representing the i-th class and n indicates the number of classes. Evaluation results demonstrate that this method outperforms both conventional regression and classification in the ChaLearn LAP 2015 apparent age estimation challenge [11] and other benchmarks.

Recent solutions for age estimation have shown that there is an ordinal relationship among ages and leveraged this relationship to design customised loss functions. The ordinal relation indicates that the age of an individual increase as time elapses since ageing is a non-stationary process. Specifically, in [31], the authors construct a label ordinal graph based on a set of quadruplets from training batches and use a hinge loss to force the topology of this graph to remain constant in the feature space. On the other hand, [37] treats the age estimation problem as an ordinal regression problem [30]. The ordinal regression is a type of classification method which transforms the conventional classification into a series of simpler binary classification subproblems. In [37], each binary classification subproblem is used to determine whether the estimated age is younger or elder than a specific age. To this end, the authors replace the final output layer with n binary classifiers, where n equals the number of classes. Let us assume that there are N samples $\{x_i, y_i\}_{i=1}^{N}$, where x_i is the i-th input image and y_i is the corresponding age label, and T binary classifiers (tasks). The loss function to optimise the multi-output CNN can then be formulated as:

$$\mathbb{E}_m = -\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \lambda^t 1\{o_i^t = y_i^t\} w_i^t log(p(o_i^t \mid x_i, W^t)), \tag{4}$$

where o_i^t indicates the output of the t-th binary linear layer, y_i^t indicates the label for the t-th task of the i-th input, and w_i^t indicates the weight of the i-th image for the t-th task. Moreover, W^t is the weight parameter for the t-th task, and λ^t is the importance coefficient of the t-th task. Chen *et al.* [8] take a step further by training separate networks for each age group so that each network can learn specific features for the target age group rather than sharing the common features as in [37]. Experiments show that this separate training strategy leads to a significant performance gain on the MORPH II dataset under both evaluation metrics. Li *et al.* [29] also consider the ordinal relation among ages in their work. However, instead of applying the age estimation model on the entire dataset, they take the different ageing pattern of different races and genders into consideration and leverage the domain adaptation methodology to tackle the problem. As stated in their paper, it is difficult to collect

and label sufficient images of every population (one particular race or gender) to train the network. Therefore, an age estimation model that is trained on the population with an insufficient number of images would have lower accuracy than models trained on other populations. In their work, they first train an age estimation model under the ranking based formulation on the source population (the population with sufficient images). Then, they fine-tune the pre-trained model on the target population (the population with a limited number of images) by adopting a pairwise loss function to align the age-specific features of the two populations. The loss function used for feature alignment is:

$$\sum_{i=1}^{N^s}\sum_{j=1}^{N^t}\{1 - l_{ij}(\eta - d(\hat{x_i^s}, \hat{x_j^t})) \cdot \omega(y_i^s, y_j^t)\}, \tag{5}$$

where $\hat{x_i^s}$ and $\hat{x_j^t}$ are the high-level features extracted from the network, y_i^s and y_j^t are the labels of the images from the source and target populations, respectively. $d(\cdot)$ is the Euclidean distance. η and $\omega(\cdot)$ are a predefined threshold value and a weighting function, respectively. l_{ij} is set to 1 if $y_i^s = y_j^t$ or -1 otherwise. The basic idea behind this function is that when the two images have the same age label, the model tries to minimise:

$$d(\hat{x_i^s}, \hat{x_j^t})) - 1, \tag{6}$$

which reduces the Euclidean distance between two features. When the two images have different labels, i.e. $y_i^s \neq y_j^t$, the model tries to minimise:

$$\frac{3}{\omega(y_i^s, y_j^t)} - d(\hat{x_i^s}, \hat{x_j^t})), \tag{7}$$

where $\omega(y_i^s, y_j^t)$ is a number smaller than one. This pushs the two features away from each other with a large distance value. In addition, the distance value is proportional to the age difference between the two images.

Another research trend based on customised loss functions is to involve joint loss functions to optimise the age estimation model. Current works that involve joint loss functions include [20] and [40]. [20] studies the problem where the labelled data are not sufficient. In that work, the authors use the Gaussian distributions as the labels rather than specific numbers, which allows the model to learn the similarity between adjacent ages. Since the labels are distributions, they use the Kullback-Leibler (KL) divergence to minimise the dissimilarity between the output probability and the label. The KL divergence can be formulated as:

$$D_{KL}(P \parallel Q) = \mathbb{E}_{x\sim P}[log(P) - log(Q)], \tag{8}$$

where P and Q are two distributions. Besides the KL divergence, their model also involves an entropy loss and a cross-entropy loss. The entropy loss is used to make sure the output probability only has one peak since an image can only be associated with one specific age. The cross-entropy loss is used to consider the age difference

between images for the non-labelled datasets. Moreover, for the non-labelled datasets, their model accepts two images as input simultaneously. For example, for two images a and b, where a is K years younger than b, then the age of a should not be larger than K. For the image a, the authors split the output layer into two parts, the first part is the neurons with the indices 0 to K, and the second part is the neurons with the indices K to M, where M is the total number of classes. Based on the aforementioned assumption, the sum of the values in the second part should be 0 while the sum of the values in the first part should be a positive number. The authors treat this problem as a binary classification problem and use the cross-entropy loss to minimise the probability error.

[40] also uses the Gaussian distribution to represent the age label. In addition, it proposes a mean-variance loss to penalise the mean and variance value of the predicted age distribution. The mean-variance loss is used alongside the classification loss to optimise the model, which currently achieves the best performance on the MORPH II dataset and the FG-NET dataset under the MAE metric.

Other worth noting works that also use customised loss function are [32] and [18]. [32] considers both the ordinal relation among ages and the age distribution and involve the metric learning method to cluster the age-specific features in the feature domain. On the other hand, [18] adopts the triplet loss [46] from the conventional face recognition task and uses it for age estimation.

2.3.2 Modifying the Network Architecture for Age Estimation

Instead of using plain CNN models (a stack of convolutional layers), some works modify the network architecture to design efficient age estimation models, which is another trending research topic to boost the estimation performance.

Yi *et al.* [55] design a multi-column CNN for age estimation. They take the facial attributes (the eyes, nose, mouth, etc.) into consideration and train several sub-networks for each attribute. All the features extracted from different attributes are then fused before the final layer. [55] is also one of the earliest works that uses a CNN for age estimation.

Recently, Wang *et al.* [49], inspired by advances in Neuroscience [6], have designed the fusion network for age estimation. Neuroscientist have discovered that when the primate brain is processing the facial information, different neurons respond to different facial features [6]. Based on this discovery, the authors intuitively assume that the accuracy of the age estimation problem may be largely improved if the CNN learns from age-specific patches. Specifically, their model takes the face and several age-specific facial patches as successive inputs. The aligned face, which provides most of the information, is the primary input that is fed into the lowest layer to have the longest learning path. The selected age-specific patches are subsequently fed into the CNN, in a sequential manner. The patch selection is based on the Adaboost algorithm. Moreover, the input feeding scheme at the middle-level layers can be viewed as shortcut connections that boost the flow of the age-specific features. The architecture of their proposed model can be found in Figure 2.

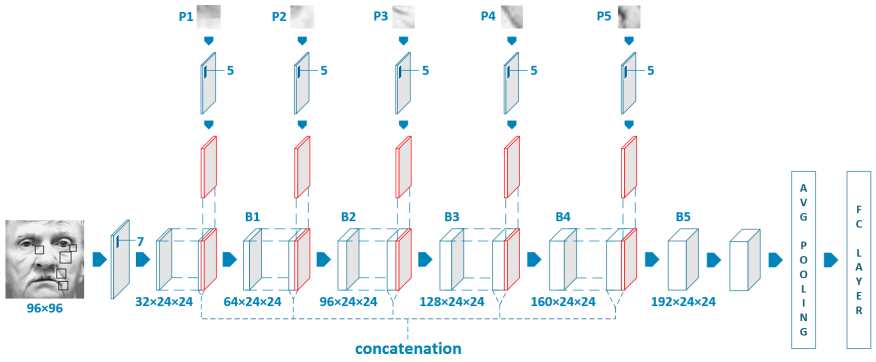

Figure 2: The architecture of the fusion network in [49]. The selected patches (P1 to P5) are fed to the network sequentially as the secondary learning source. The input of patches can be viewed as shortcut connections that enhance the learning of age-specific features.

Taheri and Toygar [48] also fuse the information during the learning process. They design a fusion framework to fuse the low-level features, the middle-level features, and the high-level features from a CNN to estimate the age.

2.3.3 Age Estimation with Multi-Task Learning

Another challenging research area is multi-task learning, which combines age estimation with other facial attribute classification problems or with face recognition. Multi-task learning is a learning scheme that can learn several tasks simultaneously, which allows the network to learn the correlation among all the tasks and saves training time and computational resources.

Levi and Hassner [28] first design a three-layer CNN to classify both the age and the race. Recently, Hsieh *et al.* [19] design a CNN with ten layers for age estimation, gender classification and face recognition. Results show that this joint learning scheme can boost the performance of all three tasks. Similarly, Ranjan *et al.* [43] propose an all-in-one face analyser which can detect and align faces, detect smiles, and classify age, gender and identity simultaneously. They use a pre-trained network for face recognition and fine-tune it using the target datasets. Authors argue that the network pre-trained for the face recognition task can capture the fine-grained details of the face better than a randomly-initialised one. Each subnetwork used for each task is then branched out from the main path based on the level of features on which they depend. Experimental results demonstrate a robust performance on all the tasks.

Lately, Han *et al.* [16] also involve age estimation in a multi-task learning scheme for the face attribute classification problem. Different from the aforementioned works, they group attributes based on their characteristics. For example, since the age is an ordinal attribute, it is grouped with other ordinal attributes like the hair

length. Rather than sharing the high-level features among all the attributes, each group of attributes has independent high-level features.

Results of the aforementioned methods on the MORPH II dataset are tabulated in Table 2. The results are only reported based on the MAE metric since some of the works do not involve the CS metric. Note that although some works have reported better results by using a pre-trained network, for a fair comparison, we do not include those in the table.

Table 2: State-of-the-art age estimation results on the MORPH II dataset. The results are based on the MAE metric (the lower, the better).

Method	Result
Yi *et al.* [55]	3.63
Niu *et al.* [37]	3.27
Rothe *et al.* [45]	3.25
Liu *et al.* [31]	3.12
Han *et al.* [16]	3.00
Chen *et al.* [8]	2.96
Liu *et al.* [32]	2.89
Taheri and Toygar [48]	2.87
Wang *et al.* [49]	2.82
Li *et al.* [29]	2.80
Hu *et al.* [20]	2.78
He *et al.* [18]	2.71
Pan *et al.* [40]	2.51

2.4 Future Research Trends on Age Estimation

Although deep learning based age estimators have achieved much better results than models that use traditional machine learning methods, there are still some issues that have not been addressed yet. First, existing age-oriented datasets like the MORPH II dataset and the FG-NET dataset involve other variations like pose, illumination, expression (PIE) and occlusion. With these unexpected factors, extracting age-specific features is onerous. [1] shows that the expression can downgrade the performance of the age estimation models, and proposes a graphical model to tackle the expression-invariant age estimation problem. Such disentangled age estimation problem has not been studied by using a CNN yet, which could be a possible future research trend.

Another possible topic is to build large-scale noise-free datasets. Recent datasets for face recognition have several millions of training samples [15, 4]. However, the largest noise-free dataset for age estimation (the MORPH II dataset) has only 40,000 to 50,000 images for training based on different data partition strategies. Therefore,

a larger noise-free dataset is needed to help to boost the age estimation performance further.

3 Age Synthesis

Compared to age estimation, age synthesis has not gained much attention from the research community yet. Age synthesis methods aim to generate elder or younger faces by rendering facial images with natural ageing or rejuvenating effects. The synthesis is usually conducted between age categories (e.g. the 20s, 30s, 40s) rather than specific ages (e.g. 22, 25, 29) since there is no noticeable visual change of a face over a several-year span. A simplified block diagram of an age synthesis model can be found in Figure 3.

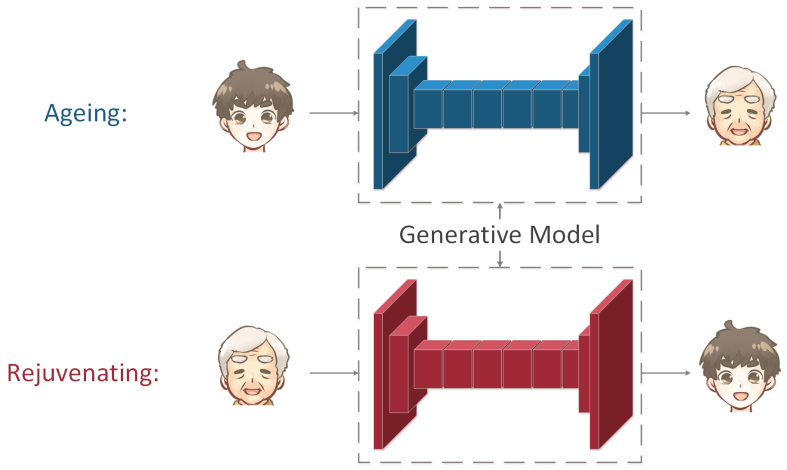

Figure 3: A simplified block diagram of an age synthesis model. An age synthesis model usually comprises two processes: the ageing process and the rejuvenating process. *Illustration by Tian Tian.*

In Figure 3, the Generative Model is usually an adversarial autoencoder (AAE) [34] or a Generative Adversarial Network (GAN) [14] in deep learning based methods. The original GAN, which is introduced by Goodfellow *et al.*, is capable of generating realistic images by using a minimax game. There are two components in the original GAN: a generator used to generate expected outputs and a discriminator used to discriminate the real images from the fake (generated) ones. The loss function used in the original GAN is:

$$V(D,G) = \min_{G}\max_{D} \mathbb{E}_{x\sim P_{data(x)}} log[D(x)] + \mathbb{E}_{z\sim P_z} log[1 - D(G(x))], \qquad (9)$$

where D and G, respectively, denote the discriminator and generator learning functions; and x and z, respectively, denote the real data and the input noise. In this model, the discriminator usually converges faster than the generator due to the saturation problem in the log loss. Several variations have been introduced to tackle this problem, including the Wasserstein GAN (WGAN) [3], the f-GAN [38] and the Least Squares GAN (LSGAN) [35].

Since the age synthesis models also require age information for the training phase, they can also rely on the datasets mentioned in Section 2.1 for training and evaluation. The most broadly used datasets to evaluate age synthesis models are the MORPH II dataset, the CACD and the FG-NET dataset. Typically, the MORPH II dataset and the CACD are used for both training and evaluation, and the FG-NET dataset is only involved in the evaluation phase due to its limited number of samples.

3.1 Evaluation Methods for Age Synthesis Models

Although age synthesis methods have attracted important attention from the research community, several challenges make the synthesis process hard to achieve. First, age synthesis benchmark datasets like the CACD involve other variations like the PIE and occlusion. With these unexpected factors, extracting age-specific features is onerous. Second, existing datasets do not have enough images covering a wide age range for each subject. For example, the MORPH II dataset only captures a time span of 164 days, on average, which may make the learning of long-term personalised ageing and rejuvenating features an unsupervised task. Third, the underlying conditions of the individuals, such as their upbringing environment and genes, make the whole synthesis process a difficult prediction task.

Based on these aforementioned challenges, researchers have established two criteria to measure the quality of synthesised faces. One is the synthesis accuracy, under which synthesised faces are fed into an age classification model to test whether the faces have been transformed into the target age category. Another criterion is the identity permanence, which relies on face verification algorithms to test whether the synthesised face and the original face belong to the same person [54].

3.2 Deep Learning Based Age Synthesis Methods

With the increasing popularity of deep learning, several age synthesis models have been proposed using various network architectures. Antipov *et al.* [2] first leverage a conditional GAN [36] to synthesise elderly faces. In their work, the authors first pre-train an autoencoder-shaped generator to reconstruct the original input. During the pre-training, they add an identity-preserving constraint on the latent features to force the identity information to remain constant during the transformation. The identity-preserving constraint is an L2 norm which can be formulated as:

$$Z^*_{IP} = argmin \parallel FR(x) - FR(\bar{x}) \parallel, \tag{10}$$

where x is the input image and $\bar{x}$ is the reconstructed image, and $FR(\cdot)$ is a pre-trained face recognition model [46] used to extract identity-specific features. After pre-training the generator, they fine-tune the network by using the age labels as conditions.

Zhang *et al.* [58] also use the conditional adversarial learning scheme to synthesise elder faces by using a conditional adversarial autoencoder. Different from [2], they do not use a pre-trained face recognition model. Instead, they implement an additional discriminator to discriminate the latent features that belong to different subjects. Therefore, their model can be trained end-to-end.

Wang *et al.* [52] recently propose the Identity-Preserving Conditional GAN (IPC-GAN). They use a similar strategy in [2], which tries to minimise the two identity-specific features from the input and the output in the feature space. To increase the synthesis accuracy, they pre-train an age estimator to estimate the age of the generated face and use the gradient from this pre-trained model to optimise the latent features through backpropagation. In this way, the latent features can learn more accurate age information. Yang *et al.* [54] use a GAN with a pyramid-shaped discriminator for age synthesis. The pyramid-shaped discriminator can discriminate multi-level age-specific features extracted from a pre-trained age estimator while conventional discriminators can only discriminate the high-level feature from the images. Following the previous works, they employ a pre-trained face recognition model to preserve the identity information. Experimental results show that their method can generate realistic images with rich ageing and rejuvenating characteristics.

It is worth noting that both [52] and [54] leverage the GAN function of the LSGAN. In the original GAN, when the distribution of the real data and the generated data are separated from each other, the gradient of the Jensen-Shannon Divergence vanishes. LSGAN replaces the log loss of the original GAN by the L2 loss. The optimisation in the LSGAN can be seen as minimising the Pearson χ^2 divergence, which efficiently solves the saturation problem in the original GAN loss while converging much faster than other distance metrics, such as the Wasserstein distance. Taking the ageing process as an example, the loss function in LSGAN is:

$$\mathcal{L}_D = \mathbb{E}_{x \sim P_{old_X}}[(D(x) - 1)^2] + \mathbb{E}_{x \sim P_{young_X}}[D(G(x))^2], \tag{11}$$

$$\mathcal{L}_G = \mathbb{E}_{x \sim P_{young_X}}[(D(G(x) - 1)^2], \tag{12}$$

where $\mathcal{L}_D$ is used to optimise the discriminator and $\mathcal{L}_G$ is used to optimise the generator.

Examples of ageing result of [54] can be found in Figure 4, The authors divide the data into four categories according to the following age ranges: 30-, 31 – 40, 41 – 50, and 51+. In the figure, the left entry of each set of images is the original face from the dataset and the the other three images are the generated results.

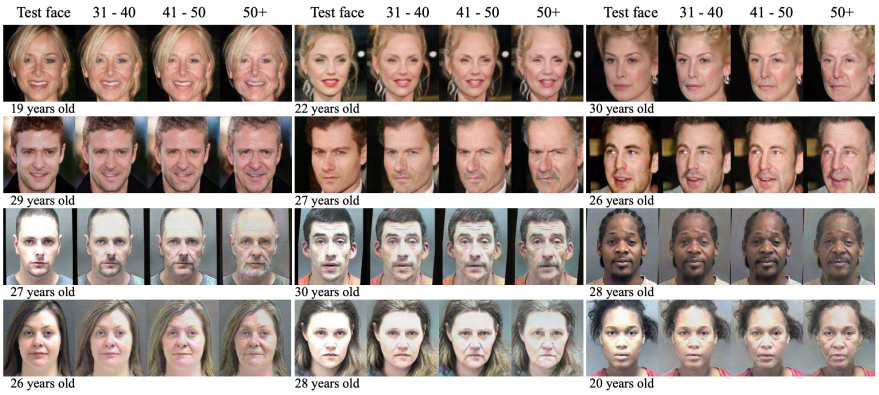

Figure 4: Ageing results of [54]. The first two rows are obtained on the CACD and the bottom two rows are obtained on the MORPH II dataset.

3.3 Future Research on Age Synthesis

The most important topic that none of the above works cover is standardising the evaluation methods of age synthesis models. Early attempts [2, 58] mainly use subjective evaluation methods by taking surveys. Recent works [52, 54] evaluate their model based on the two criteria mentioned in Section 3.2, but they use different evaluation models. Specifically, [54] uses a commercial face recognition and age estimation tool, while [52] uses their pre-trained face recognition and age estimation model. Such differences make related works hard to compare, which may hinder the development of further research.

Moreover, from the previous section, we can see that it is common to use a pre-trained face recognition model or an age estimation model to guide the training process. However, those models may be noisy. According to [52], the age estimation accuracy of their age estimator is only about 30%. Due to the fact that the classification error is high (the classifier is noisy), the gradient for the age information is not accurate. The performance can then be boosted by developing other methods to guarantee the synthesis accuracy and keep the identity information simultaneously. New methods could also make the whole training process end-to-end instead of pre-training several separate networks, which can save training time and computational resources.

4 Age-Invariant Face Recognition

Although the accuracy of the conventional face recognition models (do not explicitly consider the intra-class variations, like the pose, illumination and expression

variations, among the images of the same individual) is relatively high [46, 42], age-invariant face recognition (AIFR) is still a challenging task.

The datasets commonly used for evaluation of AIFR models are the MORPH II dataset and the FG-NET dataset. Moreover, the CACD-VS, which is a noise-free dataset derived from the CACD for cross-age face verification, is also used for AIFR. The CACD-VS contains 2,000 positive cross-age image pairs and 2,000 negative pairs. In addition, researchers also test their AIFR models on the conventional face datasets such as the Labeled Faces in the Wild (LFW) dataset to demonstrate the generalisation ability of their models.

The evaluation criteria for AIFR models are the same as those for the conventional face recognition models, which are the recognition accuracy and the verification accuracy.

4.1 Deep Learning Based Age-Invariant Face Recognition Methods

Different from conventional face recognition methods, which need to consider only the inter-class variation (the appearance and feature difference among different subjects), AIFR models also need to consider the intra-class variation, which is the age difference among the images of the same subject.

[53] is the first work that involves a CNN for AIFR. In this work, the authors propose the latent feature fully-connected layer (LF-FC) and the latent identity analysis (LIA) to extract the age-invariant identity-specific features. The LIA is formulated as:

$$v = \sum_{i=1}^{d} U_i x_i + \bar{v}, \tag{13}$$

where U_i is the corresponding matrix in which the columns span the subspace of different variations that need to be learned, x_i is the normalised latent variables from the CNN, and $\bar{v}$ is the mean of all the facial features. The output v is the set of age-invariant features. As stated in [53], each set of facial features can be decomposed into different components based on different supervised signals. Therefore, Equation (13) can be rewritten as:

$$v = U_{id} x_{id} + U_{ag} x_{ag} + U_e x_e + \bar{v}, \tag{14}$$

where $U_{id} x_{id}$ represents the identity-specific component used to achieve AIFR, $U_{ag} x_{ag}$ represents the age-specific component which encodes the age variation, and $U_e x_e$ represents the noise component. The authors then use the expectation-maximization (EM) algorithm to learn the parameters of the LIA.

Note that the LIA is only used to optimise the linear layer in the network, i.e. the LF-FC layer. Parameters in the convolutional layers are optimised by using the stochastic gradient descent (SGD) algorithm. Since the convolutional layers and the LF-FC layer are trained to learn different features (the convolutional layers learn the conventional facial features, and the LF-FC layer learns the age-invariant features),

the authors use a coupled learning scheme to optimise the network. Concretely, when optimising the convolutional layers, they freeze the LF-FC layer (fix the parameters), and when optimising the LF-FC layer, they freeze the convolutional layers.

Zheng *et al.* [59] propose the age estimation guided convolutional neural network (AE-CNN) for AIFR. The basic idea of this work is to obtain the age-specific features from an age estimation loss and remove them from the global facial features. The remove operation is done by subtraction.

Recently, Wang *et al.* [51] propose the orthogonal embedding CNN in which the global features from the last fully-connected layer are decomposed into two components, the age-specific component (features) x_{age} and the identity-specific component (features) x_{id}. Instead of considering the global features as a linear combination of x_{age} and x_{id}, they model these two components in an orthogonal manner which is inspired by the A-Softmax [33].

The state-of-the-art results on three benchmarks can be found in Tables 3-5. The reported numbers are accuracies in percentage. By using an advanced architecture (a ResNet-like model) and a customised loss function, [51] achieves the best performance on the MORPH II dataset, the CACD-VS datasets, and the LFW dataset.

Table 3: State-of-the-art results of various AIFR models on the MORPH II dataset.

Method	#Test Subjects	Accuracy
[53]	10,000	97.51%
[51]	10,000	98.55%
[59]	3,000	98.13%
[51]	3,000	98.67%

Table 4: State-of-the-art results of various AIFR models on the CACD-VS dataset.

Method	Accuracy
[53]	98.5%
[51]	99.2%

Table 5: State-of-the-art results of various AIFR models on the LFW dataset.

Method	Accuracy
[53]	99.1%
[51]	99.4%

4.2 Future Research Trends on Age-Invariant Face Recognition

Although recent AIFR models can attain good results, these results could be further improved if larger age-oriented datasets are available for training and testing. Instead of building the dataset from the ground up, age synthesis methods can be used to enlarge and augment existing datasets by generating the images of each subject at different ages or age groups. As a result, the training process could benefit from more training samples, and higher accuracy could be achieved.

According to [41], there are two types of approaches for AIFR. One is the generative approach in which synthesised faces are generated to match the target age. Then the recognition is performed based on the synthesised faces. Another is the discriminative approach which aims to discriminate faces at different ages by discovering the hidden relation among ages. Existing deep learning based methods belong to the second category. Therefore, the benefits of using a deep learning based generative approach are twofold (enlarge the existing datasets and tackle the problem from a different perspective).

5 Conclusions

Age is the most commonly used soft biometric trait in computer vision and biometrics tasks. The age-oriented models can be employed in a variety of real-life applications. However, due to the complexity of the ageing pattern and the diversity among individuals, age-related face image analysis remains challenging.

In this chapter, we divided the age-related face image analysis into three areas based on their applications. The three areas are age estimation, age synthesis, and age-invariant face recognition. We discussed each area in detail by first discussing their main concepts and the commonly used datasets for evaluation. Then, we discussed the recent research works and analysed the remaining issues and unsolved problems. We also presented the possible future research topics.

For the case of age estimation, researchers currently tackle the problem from two different angles. Existing works either design customised loss functions to compute the estimated age or modify a basic CNN architecture. Important issues in the area of age estimation are disentangling other unexpected variations like the PIE and occlusion and constructing large-scale noise-free datasets to help to boost the estimation performance further. In the case of age synthesis, researchers often adopt GANs or AAEs to generate aged or rejuvenated faces based on the input faces. However, existing works do not have a unified approach to evaluate their models. Therefore, an evaluation standard needs to be proposed. In the case of age-invariant face recognition, researchers usually design models to discover the hidden relation among ages. In other words, existing works follow a discriminative approach, thus opening opportunities for the development of models that follow a generative approach.

Acknowledgement: This work is supported by the EU Horizon 2020 - Marie Sklodowska-Curie Actions through the project Computer Vision Enabled Multimedia Forensics and People Identification (Project No. 690907, Acronym: IDENTITY).

References

1. F. Alnajar, Z. Lou, J. M. Álvarez, T. Gevers, et al. Expression-invariant age estimation. In *BMVC*, 2014.
2. G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 2089–2093. IEEE, 2017.
3. M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.
4. Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on*, pages 67–74. IEEE, 2018.
5. K.-Y. Chang, C.-S. Chen, and Y.-P. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. In *Computer vision and pattern recognition (cvpr), 2011 ieee conference on*, pages 585–592. IEEE, 2011.
6. L. Chang and D. Y. Tsao. The code for facial identity in the primate brain. *Cell*, 169(6):1013–1028, 2017.
7. B.-C. Chen, C.-S. Chen, and W. H. Hsu. Cross-age reference coding for age-invariant face recognition and retrieval. In *European conference on computer vision*, pages 768–783. Springer, 2014.
8. S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao. Using ranking-cnn for age estimation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
9. T. Cootes and A. Lanitis. The fg-net aging database, 2008.
10. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. Ieee, 2009.
11. S. Escalera, J. Fabian, P. Pardo, X. Baró, J. Gonzalez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 1–9, 2015.
12. Y. Fu, G. Guo, and T. S. Huang. Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 32(11):1955–1976, 2010.
13. 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*, 29(12):2234–2240, 2007.
14. I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
15. Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European Conference on Computer Vision*, pages 87–102. Springer, 2016.
16. H. Han, A. K. Jain, F. Wang, S. Shan, and X. Chen. Heterogeneous face attribute estimation: A deep multi-task learning approach. *IEEE transactions on pattern analysis and machine intelligence*, 40(11):2597–2609, 2018.
17. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
18. Y. He, M. Huang, Q. Miao, H. Guo, and J. Wang. Deep embedding network for robust age estimation. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 1092–1096. IEEE, 2017.

19. H.-L. Hsieh, W. Hsu, and Y.-Y. Chen. Multi-task learning for face identification and attribute estimation. In *Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on*, pages 2981–2985. IEEE, 2017.
20. Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, and S. Yan. Facial age estimation with age difference. *IEEE Transactions on Image Processing*, 26(7):3087–3097, 2017.
21. G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.
22. A. K. Jain, S. C. Dass, and K. Nandakumar. Soft biometric traits for personal recognition systems. In *Biometric authentication*, pages 731–738. Springer, 2004.
23. A. K. Jain, A. A. Ross, and K. Nandakumar. *Introduction to biometrics*. Springer Science & Business Media, 2011.
24. J. C. Klontz and A. K. Jain. A case study on unconstrained facial recognition using the boston marathon bombings suspects. *Michigan State University, Tech. Rep*, 119(120):1, 2013.
25. A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
26. Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436, 2015.
27. Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551, 1989.
28. 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*, pages 34–42, 2015.
29. K. Li, J. Xing, C. Su, W. Hu, Y. Zhang, and S. Maybank. Deep cost-sensitive and order-preserving feature learning for cross-population age estimation. In *IEEE International Conference on Computer Vision*, 2018.
30. L. Li and H.-T. Lin. Ordinal regression by extended binary classification. In *Advances in neural information processing systems*, pages 865–872, 2007.
31. H. Liu, J. Lu, J. Feng, and J. Zhou. Ordinal deep feature learning for facial age estimation. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 157–164. IEEE, 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*, 13(2):292–305, 2018.
33. W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, page 1, 2017.
34. A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey. Adversarial autoencoders. *arXiv preprint arXiv:1511.05644*, 2015.
35. X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. P. Smolley. Least squares generative adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2813–2821. IEEE, 2017.
36. M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.
37. Z. Niu, M. Zhou, L. Wang, X. Gao, and G. Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4920–4928, 2016.
38. S. Nowozin, B. Cseke, and R. Tomioka. f-gan: Training generative neural samplers using variational divergence minimization. In *Advances in Neural Information Processing Systems*, pages 271–279, 2016.
39. G. Ozbulak, Y. Aytar, and H. K. Ekenel. How transferable are cnn-based features for age and gender classification? In *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of the*, pages 1–6. IEEE, 2016.
40. H. Pan, H. Han, S. Shan, and X. Chen. Mean-variance loss for deep age estimation from a face. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5285–5294, 2018.
41. U. Park, Y. Tong, and A. K. Jain. Age-invariant face recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(5):947–954, 2010.

42. O. M. Parkhi, A. Vedaldi, A. Zisserman, et al. Deep face recognition. In *BMVC*, volume 1, page 6, 2015.
43. R. Ranjan, S. Sankaranarayanan, C. D. Castillo, and R. Chellappa. An all-in-one convolutional neural network for face analysis. In *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on*, pages 17–24. IEEE, 2017.
44. K. Ricanek and T. Tesafaye. Morph: A longitudinal image database of normal adult age-progression. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on*, pages 341–345. IEEE, 2006.
45. 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*, 126(2-4):144–157, 2018.
46. F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, 2015.
47. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
48. S. Taheri and Ö. Toygar. On the use of dag-cnn architecture for age estimation with multi-stage features fusion. *Neurocomputing*, 2018.
49. H. Wang, X. Wei, V. Sanchez, and C.-T. Li. Fusion network for face-based age estimation. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 2675–2679. IEEE, 2018.
50. X. Wang, R. Guo, and C. Kambhamettu. Deeply-learned feature for age estimation. In *2015 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 534–541. IEEE, 2015.
51. Y. Wang, D. Gong, Z. Zhou, X. Ji, H. Wang, Z. Li, W. Liu, and T. Zhang. Orthogonal deep features decomposition for age-invariant face recognition. *arXiv preprint arXiv:1810.07599*, 2018.
52. Z. Wang, X. Tang, W. Luo, and S. Gao. Face aging with identity-preserved conditional generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7939–7947, 2018.
53. Y. Wen, Z. Li, and Y. Qiao. Latent factor guided convolutional neural networks for age-invariant face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4893–4901, 2016.
54. H. Yang, D. Huang, Y. Wang, and A. K. Jain. Learning face age progression: A pyramid architecture of gans. *arXiv preprint arXiv:1711.10352*, 2017.
55. D. Yi, Z. Lei, and S. Z. Li. Age estimation by multi-scale convolutional network. In *Asian conference on computer vision*, pages 144–158. Springer, 2014.
56. K. Zhang, Z. Zhang, Z. Li, and Y. Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.
57. Y. Zhang and D.-Y. Yeung. Multi-task warped gaussian process for personalized age estimation. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2622–2629. IEEE, 2010.
58. Z. Zhang, Y. Song, and H. Qi. Age progression/regression by conditional adversarial autoencoder. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017.
59. T. Zheng, W. Deng, and J. Hu. Age estimation guided convolutional neural network for age-invariant face recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 12–16, 2017.