IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 11, 2021, accepted May 22, 2021, date of publication June 3, 2021, date of current version June 16, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3085835

Deep Face Age Progression: A Survey

MARCEL GRIMMER[1,2], RAGHAVENDRA RAMACHANDRA[1], (Senior Member, IEEE), AND CHRISTOPH BUSCH[1,2], (Senior Member, IEEE)

[1]Norwegian Biometrics Laboratory (NBL), Norwegian University of Science and Technology, 2818 Gjøvik, Norway
[2]Biometrics and Internet Security Research Group (da/sec), Hochschule Darmstadt, 64295 Darmstadt, Germany

Corresponding author: Marcel Grimmer (marceg@ntnu.no)

This work was supported in part by the German Federal Ministry of Education and Research, in part by the Hessian Ministry of Higher Education, Research, Science and the Arts, and in part by the National Research Center for Applied Cybersecurity ATHENE.

ABSTRACT Face Age Progression (FAP) refers to synthesizing face images while simulating ageing effects, thus enabling predicting the future appearance of an individual. The generation of age-progressed face images brings benefits for various applications, ranging from face recognition systems to forensic investigations and digital entertainment. In particular, the recent success achieved with deep generative networks significantly leveraged the quality of age-synthesized face images in terms of visual fidelity, ageing accuracy and identity preservation. However, the high number of contributions in recent years requires systematically structuring new findings and ideas to identify a common taxonomy, accelerate future research and reduce redundancy. Therefore, we present a comparative analysis of recent deep learning based face age progression methods for both adult and child-based face ageing, broken down into three high-level concepts: translation-based, condition-based, and sequence-based FAP. Further, we offer a comprehensive summary of the most common performance evaluation techniques, cross-age datasets, and open challenges to steer future research in the right direction.

INDEX TERMS Face age progression, semantic face editing, generative adversarial networks, biometrics.

I. INTRODUCTION

Biometric recognition refers to the automated recognition of individuals based on their biological and behavioural characteristics [1] and has steadily gained popularity in recent years. Among other applications, human forensic experts use the human face for identifying long-missing individuals or fugitive criminals. Especially for automated biometric recognition, human faces have proven to be unique, easy to capture, and non-intrusive. Based on these benefits, various large-scale border control projects have been initiated to work interoperably [2], such as the European *Entry/Exit-System (EES)* [3]–[5], the *Schengen Information System (SIS)* [6], [7], or the *Visa Information System (VIS)* [8], [9]. However, the increasing demand for face recognition also raises questions about the robustness of these systems. More precisely, the main objective of face recognition is to be robust against ''intra-subject'' variations while, at the same time, being sensitive to ''inter-subject'' differences. In this context, intra-subject variations can be caused by various factors, such as different head poses, illumination settings, face expressions (PIE factors), change of hairstyle, or face ageing.

With the emergence of deep learning based face recognition systems [10], [11], the robustness against intra-subject variations can be improved by collecting more training data that correctly reflects the distribution of a real-world scenario. While the collection of face images in an unconstrained capturing environment naturally leads to a variation of PIE factors, it is much more challenging to collect face images of the same person in the long term corresponding to the validity period of an ID document. In a recent study, Chen *et al.* [12] showed that face ageing has a tremendous effect on the performance of a face recognition system, leading to a degraded biometric performance of over 13%. The impact of face ageing has also been quantified in the recent NIST FRVT study regarding demographic differentials, which documented that the time elapsed between a reference image and probe image is highly influential on face recognition false negatives [13].

Since long-term data acquisitions of the same subjects are practically not feasible, face age progression (FAP) methods are developed to synthesize face images with ageing effects of older ages. In the early days, FAP methods could be roughly categorized into physical model-based [14]–[16] or

The associate editor coordinating the review of this manuscript and approving it for publication was Zahid Akhtar.

prototype-based [17], [18] approaches. Physical model-based methods build complex models to simulate the biological ageing mechanisms of the cranium, muscles, and facial skin. Often, they are computationally intensive and rely on long-term face sequences of the same person. On the other hand, prototype-based methods divide faces into different age groups, the average faces of which are then assumed to represent the typical age patterns. Finally, the age-synthesis can be achieved by fusing the input image with the average face of a target age group. However, prototype-based methods cannot preserve an individual's identity, making them less suitable for face recognition tasks.

From 2014, FAP methods based on deep generative networks have gained more and more attention as they significantly outperformed classical approaches in terms of visual fidelity, ageing accuracy, and identity preservation [19]–[21]. In particular, generative adversarial networks (GANs) [22] have achieved remarkable face ageing results. Besides generating photorealistic face images in high-resolution, GANs are not restricted to mated samples other than physical-model-based methods. Also, in contrast to prototype-based FAP approaches, state-of-the-art GAN architectures have addressed the problem of identity preservation.

Previous works by Fu *et al.* [23] and Ramanathan *et al.* [24] provide comprehensive introductions into state-of-the-art physical-model based and prototype-based methods. Given the increasing number of publications related to deep FAP and the high demand for robust face recognition systems, the main contribution of this survey is to conceptualize recent achievements and point out open challenges to steer future work in the right direction.

A. PROBLEM STATEMENT AND CHALLENGES

Face age progression refers to simulating the future appearance of an individual by synthesising its face image with the ageing effects of an older age. More specifically, each FAP method analyzed in this survey follows one of the following prediction schemes:

- **Age-translation between age groups:** Face images are divided into pre-defined age groups with similar ageing patterns. In this scenario, FAP methods focus on the transition between age groups by synthesizing input face images with the typical ageing effects of another age group [19], [21].
- **Age-translation to specific ages:** Instead of transitions between discrete age groups, FAP methods from this category synthesize input face images with ageing signs from specific ages (in years). The problem of missing training data samples of individual ages is solved by interpolating between ageing effects of neighbouring ages, which are more represented within the training dataset [25], [26].
- **Continuous age-translations:** Instead of synthesizing face images with ageing patterns of pre-defined ages or age groups, FAP methods from this category simulate the natural face ageing process on a continuous scale. [27], [28].

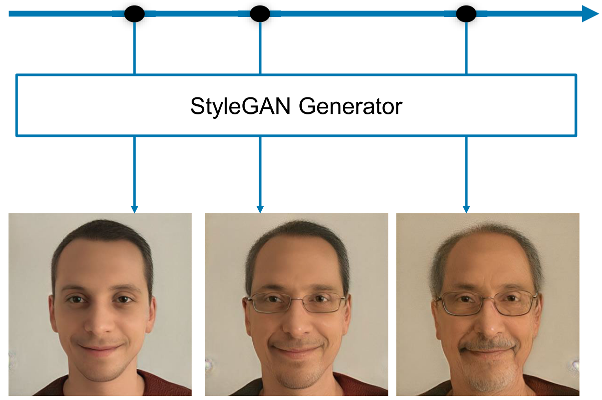

FIGURE 1. Continuous FAP with InterFaceGAN [27]: First, an encoder projects the original face image into the latent space of StyleGAN [29] (marked as black points). Then, the latent code is shifted in an age-changing direction and passed to the StyleGAN generator to reconstruct the age-progressed face image.

Figure 1 shows an example of continuous FAP with InterFaceGAN [27] - without any target age passed to the age-synthesis module. Further examples of FAP corresponding to age-translations based on age groups and specific ages are depicted in Figure 7, comparing the results of multiple FAP methods.

Despite various real-world applications that would benefit from well-performing FAP frameworks, capturing general face ageing patterns remains challenging. In particular, the complexity of the human face ageing process is due to different ageing rates varying from individual to individual, depending on genetic [30], environmental [31], and behavioural factors [32].

On a molecular basis, chronological ageing refers to the progressive degeneration of tissue, cells, and organs in the human body, which occurs throughout life and tends to be inherited [33]. The degeneration of skin tissue intensifies with exposure to ultraviolet radiation (sunlight), thus enhancing the natural chronological ageing process in local skin areas [34]. In this context, Gasperlin and Gosenca [35] emphasized the age-enhancing impact of oxidative stress caused by sunlight and pointed out the importance of a balanced diet to support the endogenous antioxidant system in the human body. Further studies on cutaneous ageing have shown the age-accelerating effects of drug abuse, such as the regular consumption of alcohol or cigarettes [32]. According to Loth and Iscan [30], emotional stress and chronic anxiety lead to intense and long-enduring muscle tensions, which increase the formation of wrinkles in various face regions. Other factors affecting skin ageing involve diseases [36], exposure to extreme climate conditions [33], or hormonal changes in the body [37].

Based on the wide range of influential ageing factors, an individual's biological age can significantly differ from the corresponding actual age (in years) [30]. Therefore, one of the main challenges of predicting future appearances is to take into account the individual ageing rates of different subjects instead of learning fixed ageing patterns [28].

Nevertheless, general ageing trends have been observed by Albert *et al.* [38], who divided human face ageing into two stages: The first stage describes the development from childhood to adulthood, which is characterized by craniofacial growth [39]. The second stage includes mainly textural changes that occur during the transition from adulthood to older ages [38]. In this context, a study by Abel *et al.* [40] emphasizes the relationship between the intensity of wrinkles and furrows with the age of an individual. Due to the differences between adult and child ageing, most research in the field of FAP either focuses on adults or children, which is why this survey presents state-of-the-art works separately in Section III and IV.

B. APPLICATIONS

Nowadays, various real-world applications benefit from successfully predicting the future appearance of individuals. Realistic age-progressed face images can be utilised to mitigate age-related biases in face recognition systems. In particular, FAP facilitates creating age-balanced datasets, which can later be used to train face recognition systems or perform biometric performance tests on existing models.

Often, several years elapse between the initial enrolment of a human face and the re-capturing of a probe sample for conducting the face verification. In this context, typical scenarios in which face recognition systems benefit from the robustness against long-term age variations include law enforcement or automatic border control. According to the Federal Bureau of Investigation, hundreds of thousands of individuals are reported as missing each year, including children, fugitive criminals, or senior citizens with dementia [41]. However, investigations can endure over many years during which the appearances of the individuals change due to natural ageing effects.

However, besides increasing the robustness of face recognition systems, other applications for FAP involve the entertainment and cosmetology sector. FAP is particularly interesting in the movie post-production, where the skin texture of actors is often retouched either digitally or physically to manipulate the perceived age. In this context, the film industry benefits from the increasing computational resources available at lower costs and advances in developing more efficient deep generative networks.

C. ORGANISATION OF THE SURVEY

This survey is structured as follows: Section II presents the taxonomy of the deep learning based FAP concepts, including a discussion about advantages and disadvantages. Next, Section III explains the basic FAP concepts with a detailed summary of state-of-the-art works. Section IV describes the differences between child vs adult face ageing and presents recent accomplishments. In Section V, the most common performance evaluation techniques are introduced based on our literature analysis and presented with face ageing examples of three recently published FAP methods (see Figure 7). Further, a crucial aspect of developing deep learning based FAP frameworks is to have a suitable dataset. Therefore, Section VI gives an overview of publicly available datasets most commonly used in the FAP literature. Finally, Section VII-B presents a summary of open challenges in the field of deep FAP.

II. DEEP FACE AGE PROGRESSION

Traditional FAP methods can be categorised into physical-model based and prototype-based approaches. More precisely, physical-model based methods focus on parametric models to simulate anatomical changes of the human face, such as muscles, skin, or cranium. However, the main drawback of those approaches is that they are very computational expensive since the model parameters lack generalisation capability and thus need to be re-learned for each face. Also, parametric models require mated samples of the same subjects over long periods, thus significantly increasing the time and costs to collect a large-scale training dataset.

On the other hand, prototype-based methods compute average faces (prototypes) from predefined age groups, the ageing patterns of which are then transferred to the younger face image. Despite avoiding capturing a sequence of mated samples, prototype-based age-synthesis often causes a loss of identity with visible ghosting effects. At this point, we refer the reader to the comprehensive survey of Fu *et al.* [23], who covers physical-model based and prototype-based FAP methods.

With the emergence of deep generative methods, research in FAP has made remarkable progress, which essentially eliminates the disadvantages of the traditional approaches mentioned above. In particular, generative adversarial networks (GANs) [22] have proven their capability to generate photo-realistic and accurate ageing effects. Furthermore, the loss function of the network can be complemented by additional loss terms to preserve the subject's identity, meaning that from the synthetically aged facial image, biometric features can be extracted sufficiently similar to the features from the original image. Thus, biometric recognition remains possible. Also, in contrast to physical-model based approaches, the training of deep generative models does not require the collection of mated samples across different ages. To leverage future research, the main objective of this survey is to summarize the fundamental concepts of deep learning based FAP techniques, the taxonomy of which is depicted in Figure 2 and divided into three classes: translation-based, sequence-based, and condition-based.

Translation-based methods are developed to convert the style of an image into the style of another set of images. For this purpose, Zhu *et al.* [42] introduced Cycle-GAN, which captures the style-based characteristics of an image collection and translates these characteristics to another collection of images. Cycle-GAN has been a milestone in the more general task of image-to-image translation [43] since it does not require paired images from both domains. Later, the same

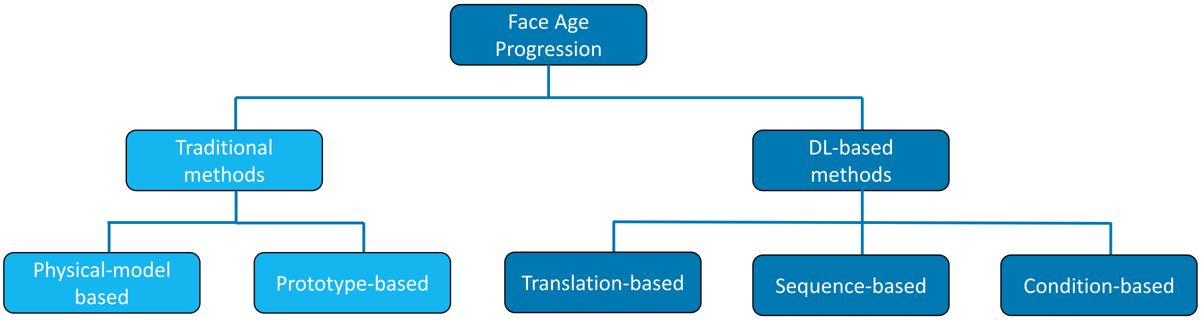

FIGURE 2. Taxonomy of FAP techniques.

idea has been exploited for FAP by using the architecture of Cycle-GAN to overcome the issue of collecting mated face images. For example, a first collection of young face images is defined (e.g. 20-30), the style of which is then converted to a second set of images, which only includes older faces (e.g. 50-60). While translation-based approaches are suitable for age translations between two age groups, their efficiency decreases for multiple age groups since a separate Cycle-GAN must be trained for each domain translation.

Other than translation-based FAP, sequence-based methods [20] are not designed to transform face images into another age group directly. Instead, multiple networks are trained separately for the translation between adjacent age groups. Each of the trained models is then concatenated in a recursive way to form a single FAP framework, where the aged output of the i-th model defines the input of the (i + 1)-th model. This strategy is motivated by the observation that the more time passes, the more complex face ageing effects occur. According to Wang *et al.* [44], even though modern deep learning approaches are getting more powerful, it is still challenging to learn age group transitions in a "one-shot" manner. Therefore, sequence-based FAP methods seek to progressively synthesize ageing effects by traversing through a chain of adjacent face ageing models. However, the main disadvantage of sequence-based methods is that for long-term age translations, the whole ageing chain must be established, including collecting training data for each age group. However, contrary to the argument of Wang *et al.* [44], most recent FAP methods focus on one-shot age-synthesis while achieving state-of-the-art performances [28], [45], [46].

Finally, condition-based FAP methods use conditional GANs [47] to control the age-synthesis with additional age labels. More precisely, age labels are constructed in a one-hot-encoded manner to indicate to which age group the given input face image will be translated. In the literature, different strategies have been developed to inject age labels into the GAN framework. While some works pass the age labels to both the generator and discriminator [19], others only feed them to the generator [26]. Also, there are different notions of how to include them in the network. For example, Wang *et al.* [19] constructed one-hot-encoded tensors directly concatenated with the input image. On the other hand, Yao *et al.* [26] designed a modulation network that fuses the age labels with the latent vectors. However, in summary, condition-based FAP methods share the same concept of guiding the generator by including extra information about the target age group. The high efficiency of condition-based FAP methods is a significant advantage compared to translation-based approaches. The inclusion of age labels enables to use a single conditional GAN framework for synthesizing face images with ageing patterns of an arbitrary age group.

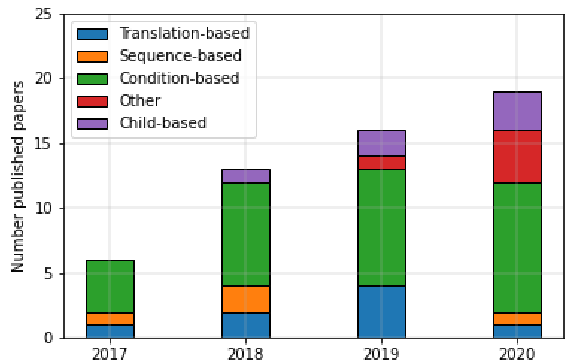

FIGURE 3. Timeline of FAP works reviewed in this survey.

Figure 3 shows the number of FAP publications covered by this survey, representing the main period between 2017 until 2020. Each year's increasing number of publications confirms the growing demand for deep learning based FAP solutions as indicated by the various application scenarios. Especially condition-based FAP methods dominate the recent research activity due to the high efficiency of conditional GANs without sacrificing the quality of age-progressed images. However, there is a slight increase of alternative FAP approaches ("Other") that could not be assigned to one of the three concepts presented in this survey. Those include FAP techniques based on feature map normalization [48], [49], ethnic-specific ageing maps [28], or latent space manipulations of existing image generation frameworks [27]. Further, the number of works dedicated to child-based FAP has grown steadily since 2018, reflecting the necessity of future research efforts in this field to counteract social issues, such as child trafficking.

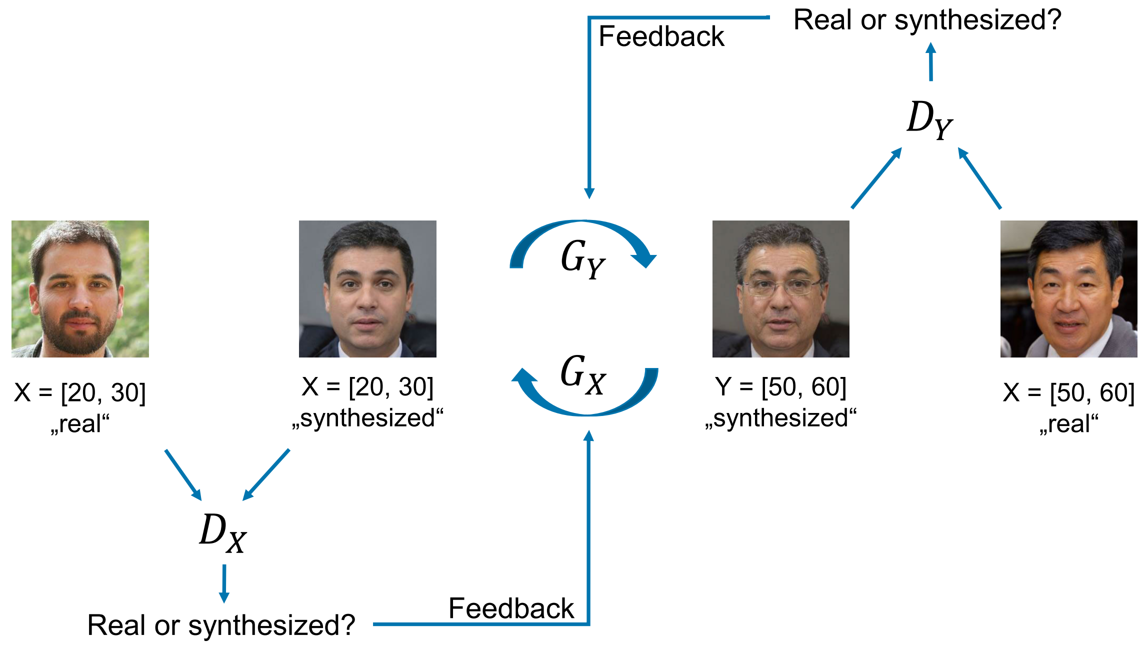

FIGURE 4. Example structure of a basic translation-based FAP framework.

III. DL-BASED CONCEPTS

This section summarises the basic principles of the FAP concepts following the taxonomy given in Figure 2. Additionally, state-of-the-art FAP approaches are introduced to familiarise the reader with the variety of ideas represented by each of the three concept domains. Finally, Section III-D presents alternative approaches that do not fit into one of the three concepts.

A. TRANSLATION-BASED

Translation-based FAP methods are based on Cycle-GANs [42], the basic idea of which Figure 4 illustrates. The framework consists of two generators G_X and G_Y with two associated discriminators D_X, D_Y, where X and Y denote face images from different age groups. While G_Y performs the transition from X to Y, G_X learns to translate a face back from Y to X. Meanwhile, D_X and D_Y are trained to discriminate between real and synthesized images, forwarding their feedback to the generators and guiding them to generate face images indistinguishable from the other age group. To further regularize the transitions between X and Y, Zhu *et al.* [42] introduced a *cycle-consistency loss*, which encourages the generators to be cycle-consistent:

$$x \approx G_X(G_Y(x)) \tag{1}$$

and

$$y \approx G_Y(G_X(y)) \tag{2}$$

with $x \in X$ and $y \in Y$. Often, the basic structure of Cycle-GANs is slightly modified or augmented by additional components. For example, Zhou *et al.* [50] proposed a FAP method conditioned on an individual's profession, reflecting the observation that the human ageing process depends on environmental factors. Another approach has been presented by Palsson *et al.* [51], who divided the faces into disjunctive age groups, followed by training a cycle-consistent GAN for each pair.

Pantraki and Kontropoulos [53] designed a method motivated by the UNsupervised Image-to-Image translation (UNIT) framework introduced by Liu *et al.* [60]. To achieve this, similar to Palsson *et al.* [51], a cycle-based GAN is trained in a pairwise manner for each age group. However, they assume that the encoded faces across all ages follow the same joint distribution. Therefore, all faces are mapped into the same latent space by forcing the last layers of the encoders and the first layers of the generators to share the same weights. In a follow-up work by Pantraki *et al.* [54], the weight-sharing logic is re-designed by dividing the encoder into three groups: while the first layers are trained individually, the intermediate layers share the weights with encoders from adjacent age groups. Finally, the last layers of all encoders share the same weights to project all face images into the same latent space.

To support the generator to pay more attention to texture information, Wang *et al.* [56] utilize Cycle-GANs for the translation into "edge maps", which capture the canny contours and landmarks of a face image. The edge maps are then translated to the aged face using a pre-trained edge-to-face generator introduced by Wang *et al.* [61]. Despite the progress of deep generative networks, a general issue of convolution operations is caused by an increasing computational power required to synthesize images with high resolutions.

TABLE 1. Summary of translation-based FAP methods.

Reference	Year	Ageing Scheme	Dataset	Note
Zhou et al. [50]	2017	Age group transitions	CACD, FG-NET + Webcrawled DB (private)	Cycle-GAN conditioned on the profession a data subject.
Palsson et al. [51]	2018	Age group transitions	CelebA	Cycle-GAN complemented by age estimator [52].
Pantraki et al. [53]	2018	Age group transitions	CACD, UTKFace, FG-NET	Shared latent space across multiple Cycle-GANs.
Pantraki et al. [54]	2019	Age group transitions	CACD, UTKFace, FG-NET	Extension of [53] with improved shared latent space.
Thengane et al. [55]	2019	Age group transitions	UTKFace	Cycle-GAN based on image patches.
Wang et al. [56]	2019	Age group transitions	CelebA, FFHQ	Cycle-GAN based on edge maps.
Wang et al. [57]	2019	Age group transitions	CACD	
Sharma et al. [58]	2020	Age group transitions	IMDb-Wiki, CACD, UTKFace, CelebA, FG-NET	Cycle-GAN combined with super-resolution technique. [59]

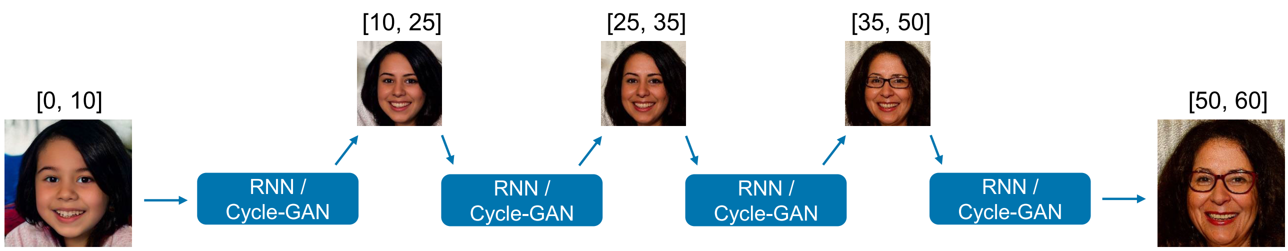

FIGURE 5. Example of sequence-based FAP framework.

Therefore, Sharma *et al.* [58] circumvented this problem by using a deep learning based super-resolution technique (ESRGAN [59]) to increase the resolution of the generated Cycle-GAN images with a scaling factor of ×4.

B. SEQUENCE-BASED

Unlike transitions between age groups in a "one-shot" manner, sequence-based FAP methods are designed to establish a chain-based face ageing framework, as shown in Figure 5. Specifically, each unit of the chain represents a deep generative network that learns to synthesize face images with the ageing effects of an adjacent age group. The units can be developed as *Recurrent Neural Networks (RNNs)* [20], [44] or generative probabilistic models [62] with the advantage of "memorizing" earlier unit states and thus taking into account correlations between age groups. Alternatively, the units can be constructed as Cycle-GANs as presented by Heljakka *et al.* [63].

The *Recurrent Face ageing (RFA)* technique by Wang *et al.* [20] has been published in 2016 and first enabled sequence-based FAP based on deep learning. The authors divided the task into two steps: face normalisation and age pattern learning. In this context, face normalisation refers to creating robust face representations by neutralising facial variations (e.g. closed eyes). First, the face images are projected into the eigenface space [66], where a separate representation space is learned for each pair of adjacent age groups. The input face image is then warped to its low-rank face representation using optical flow [67], a method emphasised by the authors to preserve facial details (e.g. wrinkles). The optimisation of the eigenface space and the optical flow estimation is conducted iteratively to minimise ghosting effects. Finally, the low-rank face representations are passed to the RFA framework to synthesise them with ageing effects. For each pair of adjacent age groups, a recurrent neural network, more precisely, a bi-layered gate recurrent unit (GRU) [64] is trained to perform the age transition. Once the low-rank age-progressed face image is predicted, the textures of the nearest neighbour in the eigenface space are adapted to transfer the fine-grained ageing details. The bi-layered GRUs of all adjacent age groups are then concatenated to generate age-progressed face images of a target age group recurrently to obtain a single face ageing framework.

The same authors have introduced an extension of RFA [44] by replacing the bi-layered GRUs with tri-layered GRUs. This adjustment is motivated by the observation that an additional hidden layer increases the network flexibility and capacity, enabling to capture more complex ageing patterns. Additionally, the face image normalisation procedure is complemented by progressively decreasing the dimensionality of the eigenface space to neutralise facial expressions further. However, training an RNN between two adjacent age groups still requires collecting mated samples from both age domains, thus significantly reducing available training data.

Another approach has been published by Heljakka *et al.* [63], who reversibly use Cycle-GANs [42] to establish a transformer chain and traverse a young face image through subsequent age groups. Instead of using a Cycle-GAN for every age group transition, they found that a single model can handle the transition between multiple age groups with a minor performance loss. In contrast to RFA [20], age transitions based on Cycle-GANs are not constraint on datasets with mated samples.

Recently, Huang *et al.* [65] highlighted the importance of training the whole transformer chain in an "end-to-end" manner. Specifically, the authors argue that training each of the FAP framework units independently causes artefacts for age progressions over long time spans since errors accumulate when being passed through the network chain. Therefore, they introduced a recursive GAN-based FAP framework that

TABLE 2. Summary of sequence-based FAP methods.

Reference	Year	Ageing Scheme	Dataset(s)	Note
Wang et al. [20]	2016	Age group transitions	LFW, MORPH-II, CACD	Recursive FAP with bi-layered GRUs. [64]
Duong et al. [62]	2017	Age group transitions	AGFW, CACD, MORPH-II, FG-NET	Recursive FAP with temporal non-volume preserving transformations.
Heljakka et al. [63]	2018	Age group transitions	CACD, IMDb-Wiki	Recursive FAP with Cycle-GANs.
Wang et al. [44]	2018	Age group transitions	LFW, MORPH-II, CACD	Recursive FAP with tri-layered GRUs.
Huang et al. [65]	2020	Age group transitions	MORPH-II, CACD, FG-NET	"End-to-end" GAN-based FAP framework.

is trained simultaneously on the whole ageing span to minimize the propagated errors.

C. CONDITION-BASED

Condition-based FAP aims to guide the age-synthesis by including a target age group as an extra condition into the GAN framework. Typically, age groups are encoded in a "one-hot-encoded" manner, also defined as age labels. Age labels can be constructed as vectors or tensors, where either the vector dimensionality or the number of input channels corresponds to the number of age groups. The decision of which shape to choose depends on where the age labels are injected into the GAN network.

Figure 6 illustrates the basic architecture of a conditional GAN complemented with an age classification loss based on a pre-trained age classifier that penalises large differences between the estimated age of the generated face image to its target age. While the age classification loss forces the network to achieve ageing accuracy, the main task of the discriminator is to support the generator to generate photo-realistic face images by learning to distinguish between real and synthesised images. Additionally, the pixel-wise $\mathcal{L}_2$ loss with $\mathcal{L}_2 = ||x - x'||_2$ motivates the network to increase the similarity between the original image x and the age-synthesized image x'. Note that condition-based FAP methods do not require mated samples since the generator learns to generalise ageing patterns of older age groups automatically during the training phase [19].

In Figure 6, the age label is constructed as a 4-channelled tensor and directly concatenated to the input image. Since only the second channel of the age label is filled with "ones", the generator is guided to synthesize the given face image with the ageing patterns of the second age group. However, the question of where to include the age labels into the network remains open: While some authors pass them to both the discriminator and generator [68], others limit the additional information feed to the generator [69]. Further, some works directly concatenate the age labels with the input image [19], whereas others inject them to intermediate network layers [45].

In 2017, Zhang *et al.* [21] were among the first to include additional age labels into the network architecture. The authors introduced a Conditional Adversarial Autoencoder (CAAE) network, assuming that all face images lie on a high-dimensional manifold. For this purpose, an input face image is first mapped to the latent space with a convolutional encoder. Once the images are projected into the latent space, the encoded samples are shifted into the direction of age changing by manipulating the age label. Afterwards, a decoder network is used to reconstruct the input image with ageing effects.

Despite the capability of CAAE to generate face images with accurate ageing effects, the personality often gets lost by traversing the encoded sample in the latent space. This problem has motivated several follow-up works to address the issue of identity preservation. For example, Antipov *et al.* [70] trained an encoder to project faces into the latent space of an age-conditioned GAN by minimizing the euclidean distance between the embeddings of a face recognition model [10]. With this idea, the path was paved for multiple contributions [19], [71], [72], augmenting the ordinary discriminator loss $\mathcal{L}_D$ with additional loss components, such as an identity-preservation ($\mathcal{L}_{ID}$) and age classification based ($\mathcal{L}_{Age}$) loss. The main motivation behind this idea is to force the generator to output aged-progressed face images that, on the one hand, belong to the target age group, and on the other hand, represent the same subject. Finally, the overall loss function is a linear combination of $\mathcal{L}_D$, $\mathcal{L}_{ID}$ and $\mathcal{L}_{Age}$, where the coefficients are adjusted to keep the balance between visual fidelity, ageing accuracy, and identity preservation.

To construct $\mathcal{L}_{ID}$, Yang *et al.* [72] utilized a pre-trained deep face descriptor [11] to extract identity-based feature vectors from both young face images and the generated older versions. The Euclidean distance is then used to measure the difference between the corresponding identity-related feature vectors, thus penalizing the network for large identity gaps. Similarly, $\mathcal{L}_{Age}$ is designed to prevent the generated face from deviating from the target age by including an age classifier that penalizes the difference between the age of the synthesized face image and the target age. Following this principle, Wang *et al.* [19] introduced an *Identity Preserving GAN (IPCGAN)*, which both integrates an identity-preserving component [73], [74], such as a pre-trained CNN [75] that serves as an age estimator.

Synthesizing face images with ageing patterns is a non-linear transformation that includes global effects (e.g. skin deformations) and local effects, such as intensifying wrinkles and furrows. This phenomenon has been observed by Li *et al.* [79], who proposed a global and consistent GAN that divides the generator into a global and three local networks. While the global network synthesizes the whole image to capture coarse-grained ageing effects (e.g. head deformations), the three local networks operate on small image patches to focus on more fine-grained ageing patterns (e.g. local furrows). The same authors [86] could further improve

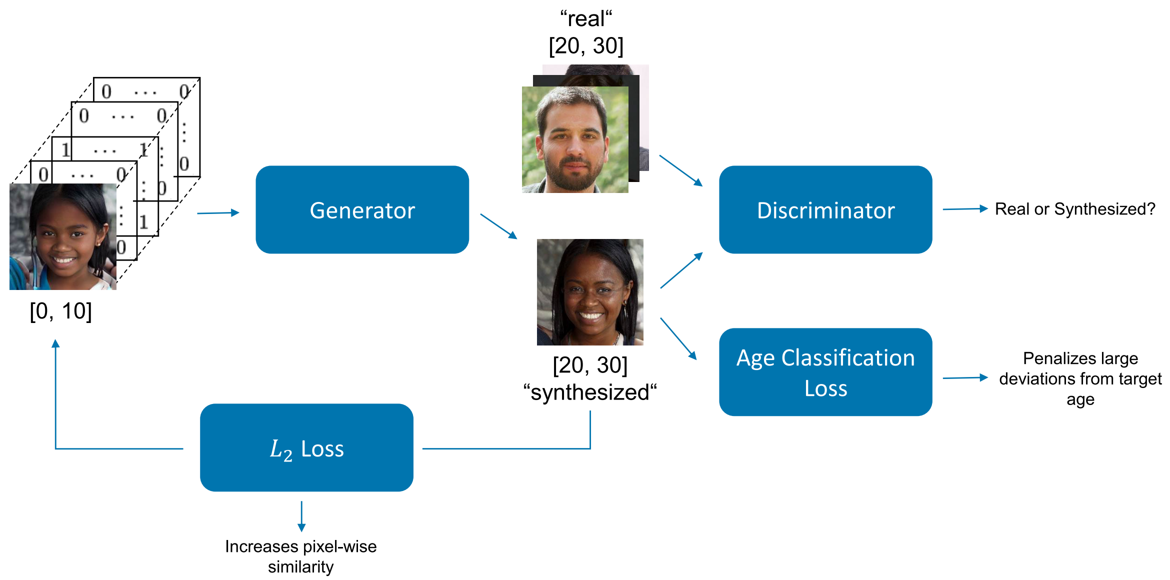

FIGURE 6. **Example of condition-based FAP framework.**

their ageing results by transforming the face images into the frequency domain, using a wavelet packet transformation. Another GAN-based framework has been introduced by Liu *et al.* [45], who also operate within the frequency domain to extract textual features at multiple scales more effective. Instead of only conditioning the GAN with target age labels, they further include facial attributes, such as ethnicity and gender, which helps to preserve these characteristics.

Most FAP methods are based on learning how to traverse between different age groups. However, each age group must have sufficient representative data to enable the GAN framework to learn the individual age patterns. The shorter the time intervals are chosen, the less training data represent each age group. This data scarcity makes short-term FAP a challenging task recently addressed by Sun *et al.* [69], who presented an ordinal ranking adversarial network. In addition to deciding whether an input image is real or not, three discriminators are further trained to output binary ranking vectors, which are used to calculate a rank estimation loss, which ensures that the generated face images are translated to the target age group.

Zhu *et al.* [98] introduced an attention-based GAN framework motivated by the findings of [43], who state that the utilization of a pixel-wise loss results in blurring or ghosting effects. These artefacts are avoided by training a generator to output an attention mask and a colour mask: While the attention mask learns to mark the image areas relevant to the age synthesis, the colour mask learns how to modify those regions. Following this strategy, both background area and personal identity are well preserved.

Most contributions made for FAP are limited to synthesize face images with low resolutions since processing larger images require adequate computational resources. Recently, this issue has been tackled by Yao *et al.* [26], who designed a GAN-based architecture able to synthesize high-definition face images (1024 $\times$ 1024 pixels). In contrast to most previous works, no age labels are fed into the discriminator, which reduces its task to discriminate whether an image is photo-realistic or not. Further, the authors used a feature modulation layer, which connects the latent vectors with source and target age labels by applying a fully connected neural network. To obtain the source age labels, they utilize a pre-trained CNN [52] for age classification. The same age classifier is finally reused to penalize age differences between the generated face images and the target ages, forcing the network to achieve ageing accuracy.

A typical disadvantage of splitting age into discrete bins is emphasized by Fang *et al.* [46], who highlighted the importance of taking into account the correlation between adjacent age groups. To capture these inter-correlations, they propose a triple translation loss, which forces the generator to generate age-progressed face images stemming from different age groups simultaneously.

Most state-of-the-art FAP methods focus on either short-term or adult to elderly face ageing, dominated by texture changes (e.g. wrinkles and furrows). However, it becomes more challenging once FAP is conducted as lifespan ageing since the generative network must learn more complex ageing patterns. In this context, Or-El *et al.* [25] proposed a lifespan FAP method based on a multi-domain image-to-image conditional GAN framework. Instead of defining equidistant age groups, they designed them to represent life phases, where the most significant changes to the facial biometric characteristic occur: 0-2, 3-6, 7-9, 15-19, 30-39, and 50-69. The network structure includes both an identity-based encoder and a mapping network, which is constructed to project age vectors into a latent space optimized for

TABLE 3. Summary of condition-based FAP methods.

Reference	Year	Ageing Scheme	Dataset(s)	Note
Antipov et al. [70]	2017	Age group transitions	Cleaned IMDb-Wiki	First to include an identity preservation loss. [10]
Chen et al. [76]	2017	Age group transitions	Adience, CACD	
Liu et al. [77]	2017	Age group transitions	CACD, FG-NET, LFW, MORPH-II, IMBd-Wiki, Adience	Additional discriminator to distinguish between real and synthesized images of adjacent age groups.
Zhang et al. [21]	2017	Age group transitions	MORPH-II, CACD, FG-NET	FAP with Conditional Adversarial Autoencoders (CAAEs) by manipulating age labels in the latent space.
Chen et al. [78]	2018	Age group transitions	UTKFace	Extension of [21], which addresses saturation problem during network training.
Jia et al. [68]	2018	Age group transitions	Cleaned IMDb-Wiki, CACD, UTKFace	Two subencoders to separate identity and age related features.
Li et al. [79]	2018	Age group transitions	MORPH-II, CACD, FG-NET	Generator divided into global network and three local networks.
Liu et al. [80]	2018	Age group transitions	MORPH-II, CACD, IMDb-Wiki	
Song et al. [81]	2018	Age group transitions	UTKFace, CACD, MORPH-II, FG-NET, IMDb-Wiki	Dual cGAN: One cGAN trained for age-synthesis and another one to reverse ageing effects.
Wang et al. [19]	2018	Age group transitions	CACD	First to include both identity-preservation and age classification loss.
Zeng et al. [82]	2018	Age group transitions	UTKFace, FG-NET	
Zhu et al. [71]	2018	Year-accurate ageing	MORPH-II, UTKFace, FG-NET	
Gou et al. [83]	2019	Age group transitions	Webcrawled DB (private)	
Li et al. [84]	2019	Age group transitions	MORPH-II, CACD, UTKFace	Dual cGAN (see [81]) with spatial attention mechanism.
Liu et al. [85]	2019	Age group transitions	MORPH-II, CACD, IMDb-Wiki	cGAN conditioned on age and gender.
Liu et al. [45]	2019	Age group transitions	MORPH-II, CACD	1) Generator conditioned on various facial attributes. 2) Wavelet Packet Transformation applied to synthesized images before passing them to discriminator.
Li et al. [86]	2019	Age group transitions	MORPH-II, CACD, FG-NET	
Roy et al. [87]	2019	Age group transitions	UTKFace	
Yang et al. [88]	2019	Age group transitions	MORPH-II, CACD, FG-NET	Extension of [72] based on cGAN with a pyramidal-structured discriminator.
Zhang et al. [89]	2019	Age group transitions	UTKFace, Adience, FG-NET	cGAN with additional Autoencoder trained on residual images.
Zeng et al. [90]	2019	Age group transitions	MORPH-II, UTKFace	Controls age progression via style transfer [91].
Fang et al. [46]	2020	Age group transitions	CACD, MORPH-II, CALFW	cGAN with triple-translation loss.
Ning et al. [92]	2020	Age group transitions	CACD, private DB (Webcrawled)	
Or-El et al. [25]	2020	Continuous ageing	FFHQ-ageing	Enables continuous ageing by interpolating between discrete age groups.
Pham et al. [93]	2020	Age group transitions	UTKFace, FG-NET	
Sheng et al. [94]	2020	Age group transitions	CACD	cGANs with rank-based discriminators [95].
Sun et al. [96]	2020	Age group transitions	MORPH-II	Age labels are modeled as distributions instead of one-hot-encoded vectors.
Sun et al. [69]	2020	Age group transitions	MORPH-II	cGANs with rank-based discriminators. [95]
Wang et al. [97]	2020	Age group transitions	MORPH-II, CACD	cGAN with pool of discriminators.
Yao et al. [26]	2020	Year-accurate ageing	FFHQ augmented with synthetic face images [29]	1) Feature modulation layer fuses age labels and latent vectors. 2) Data augmentation with synthetic face images.
Zhu et al. [98]	2020	Age group transitions	MORPH-II	Generator divided into two attention-based autoencoders.
Bian et al. [99]	2021	Age group transitions	CACD, MORPH-II, UTKFace, FG-NET	
Huang et al. [100]	2021	Age group transitions	CACD, MORPH-II	Generator with age-specific feature maps constructed via Dropout [101].
Alaluf et al. [102]	2021	Year-accurate ageing	FFHQ, CelebA (only high-quality)	Exploitation of pre-trained StyleGAN2 [103] generator.

continuous age transformations. Finally, a decoder combines age encodings and identity features with the modulated convolutions introduced by Karras *et al.* [103].

D. OTHER

This section summarizes FAP methods that could not be assigned to the categories defined in Section II (i.e. condition-based, translation-based, sequence-based). In this context, Shen *et al.* [27] introduced InterFaceGAN, which is designed to manipulate facial attributes of a given face image. Instead of proposing a new FAP architecture, InterFaceGAN operates in the latent space of an existing face image generation model, such as StyleGAN [29]. More precisely, InterFaceGAN exploits the well-structured latent space by finding linear boundaries that divide the latent space into two subspaces in terms of a binary semantic (e.g. ''younger than 50 years'' vs ''older than 50 years''). Finally, an individual's age is manipulated continuously by shifting a latent vector into the perpendicular direction of the boundary. However, the further the latent vector is moved into one direction, the more the identity of the original data subject changes.

One of the main challenges associated with predicting the future appearance of an individual is to take into account both personalized ageing factors and common ageing trends. To address this issue, He *et al.* [104] proposed a GAN-based FAP architecture (S^2GAN) that learns to extract personalized ageing patterns for each individual. Given the personalized features, the age-synthesis is conducted in the encoded domain to synthesize the features with common ageing trends of different age groups. Finally, the resulting features are passed to a decoder to reconstruct the age-progressed face images. Unlike conditional GANs, S^2GAN simultaneously learns ageing trends for each pre-defined age group during the network training, thus eliminating the need for age labels in the testing phase. Further, continuous face ageing is achieved by interpolating between age-progressed features stemming from adjacent age groups.

The lack of available face images belonging to extreme age groups (e.g. 0-5 or 90-100) motivated Georgopoulos *et al.* [48] to present a style-based FAP method: Instead of conditioning the generative adversarial network with age labels, the style of a target face image is transferred via Style Transfer [91] for transferring ageing effects to the input

TABLE 4. Summary of uncategorized FAP methods.

Reference	Year	Ageing Scheme	Dataset(s)	Note
He et al. [104]	2019	Continuous ageing	CACD, MORPH-II	FAP with a combination of personalized and common age patterns.
Despois et al. [28]	2020	Continuous ageing	FFHQ, private DB containing clinical ageing signs [105] [106] [107] [108] [109]	FAP based on ageing maps defined by ethnic-specific clinical age-signs.
Georgopoulos et al. [48]	2020	Year-accurate ageing	MORPH-II, CACD, FG-NET	FAP based on Style Transfer [91].
Shen et al. [27]	2020	Continuous ageing	No training data required	Latent space manipulations of existing face image generation network [29].
Shi et al. [49]	2020	Age group transitions	CACD, FG-NET, MORPH-II	FAP via Conditional-Attention Normalization.

face image. Specifically, the layers of the discriminator and decoder are constructed identical but in reverse order. In order to achieve the age-synthesis, a target face image is passed to the discriminator, where the statistics (column-wise mean and standard deviation) of each feature map is forwarded to the corresponding layer of the decoder in order to transfer the style via Adaptive Instance Normalization (AdaIN) [91].

The main problem associated with AdaIN operations is emphasised by Shi *et al.* [49], who state that local age-relevant face regions are smoothed out caused by the equal normalisation of convolution feature maps. To address this issue, the authors proposed a Conditioned-Attention Normalised GAN framework. More precisely, the AdaIN operations are replaced by Conditional-Attention Normalization (CAN) layers, which control the age transition between different age groups with learned attention maps. The main advantage of including CAN-layers is to focus more on local face regions relevant to the age synthesis. Further, the authors utilise a *Contribution-Aware Age Classifier*, which measures the contribution of the elements of the discriminator's feature vectors to the age classification, yielding a more fine-grained age assessment.

Recently, Despois *et al.* [28] presented a novel approach for high-resolution FAP on a continuous age scale. The authors argue that smooth face age translations cannot be achieved with domain transitions between discrete age groups because of the individual nature of face ageing due to data subject specific factors, such as genetic, ethnicity, or lifestyle. Therefore, Despois *et al.* [28] utilized ethnic-specific skin atlases [105]–[109] each of which captures clinical age-signs of a specific face region expressed as a numerical score. Instead of conditioning the GAN framework with "one-hot-encoded" age vectors, the authors introduced ageing maps that summarize information from 15 age-relevant face zones. The authors collected a private database of 6,000 high-resolution (3000×3000) face images labelled based on the ethnic-specific age atlases.

IV. CHILD VS ADULT FACE AGEING

As described in Section II, human face ageing can be divided into two stages: While craniofacial growth occurs from childhood to adulthood, the remaining ageing process is dominated by texture changes. Consequently, most FAP methods focus either on facial ageing of children [115]–[117] or adults [19], [45], [70] in order to reduce the complexity of patterns a deep generative network must learn. However, the amount of research spent on child face ageing is still limited compared to research conducted for adults. This research gap exists because children are either not included in common cross-age datasets (MORPH-II, CACD) or extremely underrepresented (UTKFace, FG-NET). The under-representativeness of children in cross-age datasets is associated with collecting face images from social media or web search engines, where adults are naturally more represented.

The first step towards overcoming the lack of available child face images has been made by Chandaliya *et al.* [110], who collected a private dataset (*Children Longitudinal Face (CLF)*) that consists of 8,581 face images of Indian children and covers an age span from 2-20. Based on CLF, the authors re-trained an already existing FAP method (CAAE [21]) and compared the age-synthesis results to the performance of the original CAAE model. Although the performance could be slightly improved, the identity loss caused by the CAAE remained the main drawback. Therefore, the same authors augmented the architecture of CAAE in a follow-up work [112] with a perceptual loss based on VGG-19 [118]. More precisely, a perceptual loss measures the difference between high-level semantic features extracted with a well-trained image classification network, thus minimizing the spatial differences between input and synthesized face. Motivated by this work, Xiao and Zhao [117] further developed the CAAE architecture by including gender labels in addition to age labels. This strategy is based on the observation that the distinction between male and female toddlers can be challenging during early childhood, which causes gender inconsistencies after the age synthesis. Therefore, both age labels and gender labels are concatenated with the latent face representation to support the encoder to better cluster the faces according to these facial attributes.

Following the idea of IPCGAN [19], Chandaliya *et al.* [115] adopted the same architecture but with a multi-scale discriminator structure. Additionally, the VGG19-based perceptual loss is complemented with an age-based loss constructed with LightCNN [119], penalizing large age gaps between the age of the synthesized face and the target age.

Recently, Dhar *et al.* [120] found that age-related information is highly coupled with identity-salient features in the latent space of a well-trained face recognition model [121]. The entanglement of these attributes has been exploited by Deb *et al.* [113], who trained an autoencoder that operates directly within the latent space of a pre-trained face recognition model, such as CosNet [114]. Given that the latent space

TABLE 5. Summary of child-based FAP methods.

Reference	Year	redAgeing Scheme	Dataset(s)	Note
Chandaliya et al. [110]	2018	Age group transitions	The Children Longitudinal Face [111] (private)	CAAE [21] tested for private dataset.
Chandaliya et al. [112]	2019	Age group transitions	The Children Longitudinal Face [111] (private)	Extension of [110] complemented with perceptual loss based on VGG-19.
Deb et al. [113]	2019	Year-accurate ageing	Children's Face Aging (private), ITWCC	Latent space manipulations of existing face recognition model [114].
Chandaliya et al. [115]	2020	Age group transitions	Children Longitudinal Face [111](private)	Extension of IPCGAN [19] with multi-scale discriminators.
Deb et al. [116]	2020	Year-accurate ageing	Children's Face Aging (private), ITWCC, FG-NET, CACD	Extension of [113], complemented by decoder to reconstruct face images from deep features.
Xiao et al. [117]	2020	Age group transitions	UTKFace	Extension of [112] with gender information included.

is well structured in terms of facial attributes, age manipulations can be achieved by traversing the latent vector into the corresponding age direction. In summary, the proposed face ageing module significantly improved rank-1 identification rates due to its direct link to the underlying face recognition model. Following the same idea, Deb *et al.* [116] further developed a decoder that is trained to reconstruct face images from the deep features obtained by the face ageing module.

V. PERFORMANCE EVALUATION

FAP refers to the task of synthesizing face images with ageing patterns of older faces to simulate the future appearance of a data subject. However, the question of how to assess the performance of FAP methods, such that multiple works can be compared objectively, remains open. In general, the main objective of FAP can be summarized as that all of the following three objectives must be achieved simultaneously: *visual fidelity*, *ageing accuracy*, and *identity preservation*. There is no standardized way for evaluating the performance of FAP methods according to these criteria. Therefore, this section summarizes the most commonly used evaluation techniques identified in the works examined as part of this survey.

- The **visual fidelity** of a synthesized face image is typically evaluated in terms of human perception. More precisely, the ageing results of a handful of representative face images are compared to previous state-of-the-art FAP methods. The primary motivation behind this manual assessment is to exploit the well evolved human visual system of the brain, which is effective in recognizing artefacts caused by the generator. However, recent FAP methods include additional metrics to support a quantitative analysis of the visual fidelity, such as the *Frèchet Inception Distance (FID)* [122]. More precisely, the FID assesses the fidelity of the generated face image to its source image by measuring the differences in the density of two distributions based on the high-dimensional features extracted with an InceptionV3 [118] classifier, as given in equation 3:

$$FID = |\mu - \mu_w|^2 + tr(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{\frac{1}{2}}) \quad (3)$$

where μ and μ_w denote the mean of the InceptionV3 features extracted from both real and generated images. Further, Σ and Σ_w refer to the covariance matrices of the extracted features and $tr()$ describes the trace matrix operation (sum of elements on the matrix main diagonal).
- **Ageing accuracy** refers to whether a synthesised face image belongs to the target age group. To achieve this, two quantitative evaluation techniques are frequently observed: age estimators [45], [46], [88] or user studies [19]–[21]. Specifically, pre-trained Convolutional Neural Networks (CNNs) [52] are utilised to estimate the age of a given face image in years. Alternatively, user studies are conducted where human experts estimate ages from both synthetic and real faces. Once the estimated ages are available, the mean values can be used to analyse whether they are within the target age span. Further, histograms can be utilised to compare the age distributions of different age groups for both synthesised vs ground-truth face images [45].
- Finally, besides generating realistic face images of the target age group, the last objective is to also **preserve the identity** of the data subject. Again, this can be evaluated with three techniques: automatic face verification [45], [72], [96], automatic face identification [115], [116], or user studies [19]–[21]. Face verification describes one-to-one comparisons in order to verify whether two face images stem from the same individual. More precisely, the comparison score (CS) between two face images is determined by measuring the distance or similarity between their face embeddings extracted with a pre-trained face recognition model, such as ArcFace [123]. On the other hand, face identification refers to one-to-many comparisons, where the CS of a biometric probe to all references contained in a database is measured. Typically, the result is a ranked list of CSs, with the first entry representing the biometric reference that is most similar to the biometric probe (*rank-1*). In the literature, especially child-based FAP methods report *rank-1 identification rates* since they are best suited to reflect future application scenarios, where face images of missing children are compared to large-scale databases. According to Chandaliya *et al.* [115], ArcFace [123] and FaceNet [10] are particularly suited for evaluating face identification metrics. Finally, user studies are also employed for measuring the

capability to preserve identities. For a given face image that belongs to age group 0 (AG0), the age-synthesised faces are generated for the predefined age groups (e.g. AG1, AG2), which are used to create pairs: (AG0, AG1), (AG0, AG2), (AG1, AG2). Additionally, the pairings are supplemented with random impostor comparisons of non-mated face image pairs. Finally, the paired images are handed to the participants, who decide whether the images are mated or non-mated presentations.

- An **Attribute Consistency** analysis is conducted by a small number of authors [26], [45], the idea of which is to check whether facial attributes change after the age synthesis. In particular, attribute inconsistency can be caused by an unbalanced dataset. For example, if a FAP model is applied to a female face after training it only on male faces, the likelihood of a gender switch after the age synthesis increases significantly. Therefore, measuring the attribute consistency enables to view the robustness of the FAP model against various types of biases. In this context, DeepFace [124] provides a tool for predicting facial attributes, such as the ethnicity or emotion of an individual. Also, measuring the consistency of non-facial related quality aspects, such as the degree of blurriness induced by the proposed FAP method, offers an interesting evaluation criterion.

In addition to the above described evaluation techniques, Figure 7 provides an impression on the results of three recently published FAP methods with publicly available implementations [25], [26], [102]. Following the evaluation techniques presented in this section, all images are annotated with FIDs and CSs. Note that all FIDs and CSs are computed based on the original image, preprocessed with the pipeline given by the authors, and the corresponding age-progressed image.

The qualitative analysis of the age-synthesized images reveals some interesting characteristics unique to each of the tested FAP methods. In particular, the proposed FAP method by Alaluf *et al.* [102] demonstrates the high photo-realism (1024×1024) of the generated images by exploiting the remarkable face image generation capability of the StyleGAN2 generator [103]. However, the low CSs indicate that the identity after the age-synthesis deviates from the identity of the original individual. This identity loss is a typical effect of projecting real face images into the latent space of StyleGAN2, which does not represent the full range of possible real-world identities.

On the other hand, the high-resolution (1024×1024) age-progressed images by Yao *et al.* [26] appear less realistic due to unnatural ageing effects caused by the generator (see Figure 7 - (f)). This observation emphasizes the importance of assessing the generation quality from a human perspective as a complement for quantitative metrics since the human brain is well-suited to detect disturbing factors in human faces.

Other than the previous FAP approaches, the method by Or-El *et al.* [25] operates in a lower resolution of 256×256 pixels. The face images are preprocessed by replacing the noisy background with a plain grey colour to support the generator to focus on the age-synthesis of the face region only. Note that the grey discolouration of the beard and the main hair is more pronounced compared to the other approaches, especially regarding the male individuals.

VI. DATASETS

Since it is well known that the performance of deep neural networks scales with the amount of available training data, this section presents public datasets commonly used in the FAP literature. As depicted in Table 6, the *Cross-Age Celebrity Dataset (CACD)*, as well as the *Academic MORPH Database* (MORPH-II) have been chosen by researchers most frequently. The celebrity images in CACD have been crawled from the internet, therefore representing an unconstrained capturing environment with a high variation in terms of PIE factors. However, the associated age labels are only estimations since they were obtained by subtracting the publication year from the birth year. On the other hand, the images in MORPH-II were captured in a controlled scenario, including near-frontal faces with neutral expressions, uniform illumination, and simple backgrounds. Other than with CACD, the age labels given by MORPH-II were annotated accurately, which prevents learning distorted age patterns. In practice, many authors exploited the advantages of both CACD (unconstrained) and MORPH-II (constrained) by selecting images from both datasets.

In general, we identified two main problems with existing datasets during our analysis:

- **Age Bias:** Although many datasets cover wide age spans, the number of young adults between 20-40 is typically over-represented. Simultaneously, there is a significant lack of face images from toddlers, children, and older people, which leads to deep neural networks being biased towards specific age groups. To represent each age group with the same number of samples, many authors collected images from various public datasets. Alternatively, Yao *et al.* [26] followed the idea of generating synthetic face images with StyleGAN [29] to augment and balance their dataset.
- **Ethnicity Bias:** As described in the comprehensive study of Drozdowski *et al.* [137], deep neural networks are prone to bias effects caused by non-uniform distributions of demographic factors. For example, datasets are typically over-represented by ethnic groups most common in the country where the images were acquired. This problem has been addressed by Karras *et al.* [29] who introduced the *Flickr-Faces-HQ Dataset (FFHQ)*, which covers facial images with a wide variety of ethnic groups. Since the default version of FFHQ is not annotated with labels, an extension was published by Or-Él *et al.* [25], who supplemented the dataset with various labels.

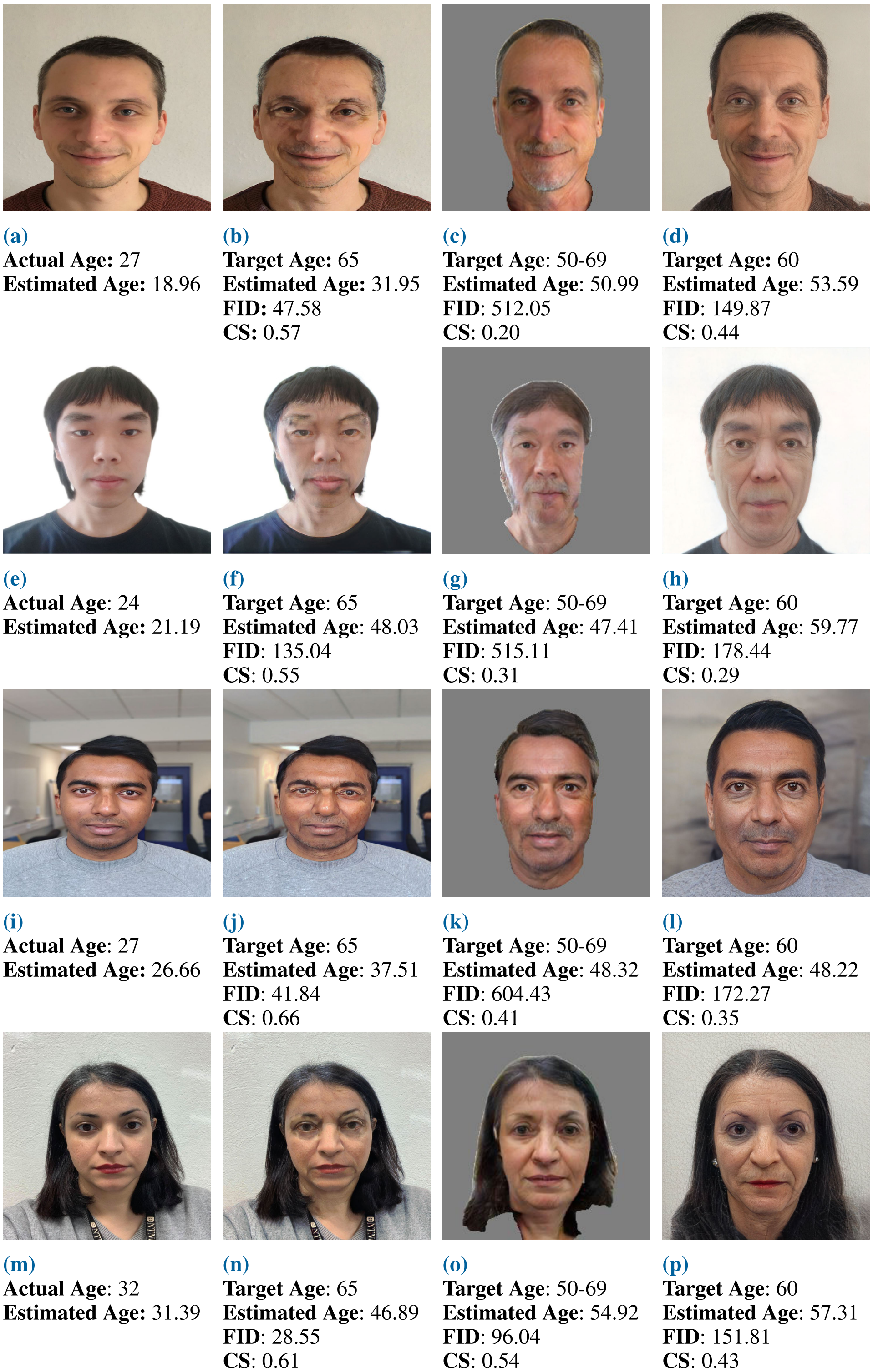

FIGURE 7. FAP examples with following columns: Source images (first), Yao *et al.* [26] (second), Or-El *et al.* [25] (third), Alaluf *et al.* [102] (fourth). The age-progressed images are annotated with: Target age, estimated age [125], Frèchet-Inception Distance (FID), and comparison score (CS) measured as Cosine Similarity between face embeddings extracted with ArcFace [123].

VII. OPEN CHALLENGES AND FUTURE WORK

Despite the many milestones achieved with deep generative networks, there are still open challenges to be addressed by future works. Therefore, the following subsections describe several promising research directions categorized as either data- or concept-based.

TABLE 6. Summary of publicly available cross-age datasets.

Reference	#Citations	#Images	Age Span	Labels	Mated Samples?	Note
CACD [126]	34	163k	16-62	Age	Yes	Webcrawled collection of celebrities with high variation of poses, illuminations, and expressions, estimated age labels.
MORPH-II [127]	26	55k	16-77	Age, Ethnicity, Gender, Glasses, Facial Hair	Yes	Average Age span of 164 days, colored, near-frontal, neutral expression, uniform illumination, simple background, accurate age-labels.
FG-NET [128]	21	1k	0-69	Age, Gender, Glasses, Hat, Facial Hair, Pose	Yes	Captured in uncontrolled environments.
UTKFace [21]	16	20k+	0-116	Age, Gender, Ethnicity	No	Subset of MORPH, CACD, and images webcrawled from Google/Bing.
IMDb-Wiki [52]	5	524k	5-86	Age, Gender	Yes	Collection of celebrities from IMDb and Wikipedia.
CelebA [129]	5	203k	Unknown	40 labels (not including age)	Yes	Webcrawled collection of celebrities with high variation of poses, illuminations, expressions, and accessories.
LFW [130]	3	13k	Unknown	Metadata only	Yes	Webcrawled with low-resolution (250x250).
Adience [21]	3	27k	0-60+	Age, Gender	Yes	Webcrawled images from Flickr.
FFHQ [29]	4	70k	0-70+	Metadata only	No	High-quality Flickr images (1024x1024), high variation in ethnicity
ITWCC [131]	3	8k	0-32	Age, Gender	Yes	Webcrawled collection of images from child celebrities with high variation in pose, illumination, and expression.
IMDb-Wiki (Cleaned) [132]	2	250k	5-90	Age, Gender	Yes	Image collection of celebrities from IMDb and Wikipedia.
CALFW [133]	1	6k	Unknown	Metadata only	Yes	Subset of LFW optimized for cross-age verification.
FFHQ-Aging [25]	1	70k	0-70+	Age, Gender, Pose, Glasses, Eye Occlusion, Semantic Maps	No	Extension of FFHQ with additional labels.
AGFW [134]	1	1̃9k	10-64	Age	No	Images webcrawled and collected from "The Productive Aging Laboratory" [135].
AgeDB [136]	1	1̃7k	9-95	Age, Gender	Yes	Image collection of celebrities with manually annotated ages.

A. DATA-BASED CHALLENGES

The performance of deep neural networks is directly correlated with the quality and number of data samples available for training. Therefore, the development of task-specific cross-age datasets remains a crucial pre-condition to enable deep generative models to learn relevant ageing patterns.

- One major challenge is to collect face images from age groups that are typically underrepresented in existing public datasets, such as young children and elderly individuals. For example, most current child-based FAP methods are based on private datasets (see Table 5), which limits the reproducibility and comparability to other works. Therefore, the establishment of new public datasets focused on underrepresented age groups accelerates new research and improves existing FAP methods by enabling them to learn patterns from the whole lifetime age span.
- Since the human face ageing process also depends on external aspects, such as lifestyle, nutrition, or working conditions [138], the collection of face images labelled with these factors allows for conducting interesting experiments to establish a further understanding of the relationship between the human face ageing process and external factors.
- Modern FAP methods focus more and more on synthesizing images with higher resolutions. However, the most popular cross-age datasets (CACD [126], and MORPH-II [127]) only include images with a resolution of up to 400 × 480 pixels. Although FFHQ [29] contains 70,000 images with a resolution of 1024 × 1024, the collection of more data will leverage the generation capability of deep generative networks.

B. CONCEPTUAL CHALLENGES

Recently, the increasing attention for deep FAP led to many interesting new concepts, paving the way for future research.

- FAP is often treated as a domain translation problem, where a given input face is synthesized with ageing patterns of another age group. However, age groups are often defined as intervals with more than ten years, which means that transitions from one age group to another cause significant age gaps. Therefore, a few recent FAP approaches switch from discrete age group transitions to age progressions on year-accurate [26] or continuous scales [28], representing a promising future research direction.
- With the introduction of FFHQ [29], it is becoming more interesting to synthesize high-resolution face images with ageing effects. The higher the quality of the face images, the more fine-grained ageing patterns can be learned by the parameters of a deep generative network. However, the processing of large images requires extensive computational resources, thus demanding more cost-efficient GAN architectures.
- The application of FAP in large-scale projects, such as the Entry/Exit-System [3], requires an unbiased age-synthesis in terms of different ethnicities. However, according to our study, most works attribute a subordinate role to the model's bias, which can lead to changes in the ethnicity after the age-synthesis (compare Figure 7 - (c) vs (h)). Therefore, a comprehensive comparison of state-of-the-art FAP methods concerning their biases towards different facial attributes is considered beneficial and should be addressed by future works. Besides including an age estimator to achieve ageing accuracy, the loss function of a GAN architecture can

be complemented by other facial attribute estimators as well in order to prevent inconsistencies caused by unbalanced datasets.

- An interesting approach has been presented by Despois *et al.* [28]: Instead of synthesizing face images with general ageing patterns, each face image is divided into distinct zones, the age of which is assessed individually with a numerical score and passed to the GAN network in the form of ''ageing maps''. With this approach, the authors take into account that human face ageing rates typically differ between individuals, an observation that future works should take into account.
- Our literature survey has shown that most FAP contributions focus on adult face ageing (88%) while only a few works are dedicated to child face ageing. In contrast to adult face ageing, where the facial changes are mainly texture-based (e.g. wrinkles and furrows), the craniofacial deformations occurring while growing up are much more challenging to simulate, thus offering great potential for future research.
- Motivated by the remarkable face image generation capabilities of StyleGAN [29], the lack of available child or elderly-based face images can be compensated by synthesizing ageing patterns of a single reference image via Style Transfer [91]. The works of Georgopoulos *et al.* [48] and Shi *et al.* [49] demonstrate the effectiveness of FAP based on (attention-based) instance normalization, thus inspiring further work in this direction.

VIII. SUMMARY

In this survey, a comprehensive analysis of deep face age progression literature has been conducted. As the high number of recent publications indicates, FAP is still an active and emerging field of research relevant for various applications, such as the European Entry-Exit System. In this context, the growing attention for FAP methods can be explained by the remarkable progress achieved with deep generative networks, which enable the generation of photo-realistic ageing effects. The Conceptualisation of the methods analysed as part of this survey has resulted in three categories: translation-based, sequence-based, and condition-based. Translation-based approaches are based on the principle of Cycle-GAN [42] and focus on the translation between two age domains. On the other hand, sequence-based techniques create chain-like FAP frameworks to progressively synthesise face images with ageing effects, where the output of the $i-th$ unit defines the input of the $(i+1)-th$ unit. Finally, condition-based FAP methods inject target age labels as additional information into the network to control the age synthesis. Comparing the number of contributions associated with each concept reveals that the majority can be classified as condition-based, which is due to its high efficiency using a single generator capable of synthesising face images with ageing patterns of an arbitrary age group.

As outlined in Section VII-B, open challenges in the field of deep FAP are either considered data-based or concept-based. In particular, the collection of face images stemming from underrepresented age groups (elderly and children) will be beneficial for a vast number of GAN-based models, the performance of which scales with the number of training images. Further, the annotation of face images with additional information (e.g. profession or nutrition type) helps to establish FAP methods that are tailored to the specific conditions of an individual. Finally, from the conceptual point of view, future efforts are recommended to be directed towards continuous age progressions [25], while taking into account the individual ageing rates among different individuals [28]. Additionally, Section III-D shed light on alternative FAP approaches that are either based on manipulating the latent representations of a pre-trained GAN-model [27] or treating the age-synthesis as a style transfer problem [48], [49], thus bypassing the lack of training data from underrepresented age groups.

REFERENCES

[1] *Information Technology—Vocabulary—Part 37: Biometrics*, Standard ISO/IEC JTC1 SC37 Biometrics, ISO/IEC 2382-37:2017, International Organization for Standardization, 2017.
[2] Council of European Union. (2019). *Council Regulation (EU) No 817/2019: Establishing a Framework for Interoperability Between EU Information Systems*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2019:135:TOC
[3] A. Orav and A. D'Alfonso, ''Smart borders: EU entry/exit system,'' Eur. Parliamentary Res. Service, Members' Res. Service, Rue Wiertz, Belgium, Tech. Rep. PE 586.614, 2016.
[4] Council of European Union. (2017). *Council Regulation (EU) No 2226/2017: Establishing An Entry/Exit System (EES)*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R2226
[5] (2011). *Council Document (EU) 52011dc0680: Smart Borders*. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52011DC0680
[6] J. Galbally, P. Ferrara, R. Haraksim, A. Psyllos, and L. Beslay, ''Study on face identification technology for its implementation in the Schengen information system,'' Publication Office Eur. Union, Luxembourg City, Luxemburg, Tech. Rep. EUR 29808 EN, JRC116530, 2019, doi: 10.2760/661464.
[7] Council of European Union. (2018). *Council Regulation (EU) No 1862/2018: Establishment, Operation and Use of the Schengen Information System (SIS)*. [Online]. Available: https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:32018R1862
[8] European Commission. *Visa Information System (VIS)*. Accessed: Nov. 12, 2020. [Online]. Available: https://ec.europa.eu/home-affairs/what-we-do/policies/borders-and-visas/visa-information-system_en
[9] Council of European Union. *Council Regulation (EU) No 767/2008: Concerning the Visa Information System (VIS)*. (2008). [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex:32008R0767
[10] F. Schroff, D. Kalenichenko, and J. Philbin, ''FaceNet: A unified embedding for face recognition and clustering,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 815–823.
[11] O. M. Parkhi, A. Vedaldi, and A. Zisserman, ''Deep face recognition,'' in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, Sep. 2015, Art. no. 41, doi: 10.5244/C.29.41.
[12] B.-C. Chen, C.-S. Chen, and W. H. Hsu, ''Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset,'' *IEEE Trans. Multimedia*, vol. 17, no. 6, pp. 804–815, Jun. 2015.
[13] National Institute for Science and Technology. (2019). *Face Recognition Vendor Test Part 3: Demographic Effects*. [Online]. Available: https://www.nist.gov/publications/face-recognition-vendor-test-part-3-demographic-effects
[14] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, ''A compositional and dynamic model for face aging,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.
[15] X. Shu, J. Tang, Z. Li, H. Lai, L. Zhang, and S. Yan, ''Personalized age progression with bi-level aging dictionary learning,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 905–917, Apr. 2018.

[16] X. Shu, J. Tang, H. Lai, L. Liu, and S. Yan, "Personalized age progression with aging dictionary," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3970–3978.
[17] X. Shu, J. Tang, H. Lai, Z. Niu, and S. Yan, "Kinship-guided age progression," *Pattern Recognit.*, vol. 59, pp. 156–167, Nov. 2016.
[18] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3334–3341.
[19] X. Tang, Z. Wang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7939–7947.
[20] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2378–2386.
[21] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5810–5818.
[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[23] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.
[24] N. Ramanathan, R. Chellappa, and S. Biswas, "Age progression in human faces: A survey," *J. Vis. Lang. Comput.*, vol. 15, pp. 3349–3361, 2009.
[25] R. Or-El, S. Sengupta, O. Fried, E. Shechtman, and I. Kemelmacher-Shlizerman, "Lifespan age transformation synthesis," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 739–755.
[26] X. Yao, G. Puy, A. Newson, Y. Gousseau, and P. Hellier, "High resolution face age editing," 2020, *arXiv:2005.04410*. [Online]. Available: http://arxiv.org/abs/2005.04410
[27] Y. Shen, J. Gu, X. Tang, and B. Zhou, "Interpreting the latent space of GANs for semantic face editing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 9243–9252.
[28] J. Despois, F. Flament, and M. Perrot, "AgingMapGAN (AMGAN): High-resolution controllable face aging with spatially-aware conditional GANs," 2020, *arXiv:2008.10960*. [Online]. Available: http://arxiv.org/abs/2008.10960
[29] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4401–4410.
[30] S. Loth and M. Iscan, "Morphological indicators of skeletal aging: Implications for paleodemography and paleogerontology," in *Biological Anthropology and Aging: Perspectives on Human Variation Over the Life Span*. Oxford, U.K.: Oxford Univ. Press, 1994, pp. 394–425.
[31] J. Uitto, "Understanding premature skin aging," *New England J. Med.*, vol. 337, no. 20, pp. 1463–1465, 1997.
[32] D. Kadunce, R. Burr, R. Gress, R. Kanner, J. Lyon, and J. Zone, "Cigarette smoking: Risk factor for premature facial wrinkling," *Ann. Internal Med.*, vol. 114, no. 10, pp. 840–844, 1991.
[33] J. Uitto, M. Fazio, and D. Olsen, "Molecular mechanisms of cutaneous aging: Age-associated connective tissue alterations in the dermis," *J. Amer. Acad. Dermatol.*, vol. 21, no. 3, pp. 614–622, 1989.
[34] S. Cho, M. H. Shin, Y. K. Kim, J.-E. Seo, Y. M. Lee, C.-H. Park, and J. H. Chung, "Effects of infrared radiation and heat on human skin aging *in vivo*," *J. Invest. Dermatol. Symp. Proc.*, vol. 14, no. 1, pp. 15–19, Aug. 2009.
[35] M. Gašperlin and M. Gosenca, "Main approaches for delivering antioxidant vitamins through the skin to prevent skin ageing," *Expert Opinion Drug Del.*, vol. 8, no. 7, pp. 905–919, Jul. 2011.
[36] A. Sreedhar, L. Aguilera-Aguirre, and K. K. Singh, "Mitochondria in skin health, aging, and disease," *Cell Death Disease*, vol. 11, no. 6, pp. 1–14, Jun. 2020.
[37] M. Farage, K. Miller, P. Elsner, and H. Maibach, "Characteristics of the aging skin," *Adv. Wound Care*, vol. 2, no. 1, pp. 5–10, 2013.
[38] A. M. Albert, K. Ricanek, Jr., and E. Patterson, "A review of the literature on the aging adult skull and face: Implications for forensic science research and applications," *Forensic Sci. Int.*, vol. 172, no. 1, pp. 1–9, Oct. 2007.
[39] L. S. Mark, J. B. Pittenger, H. Hines, C. Carello, R. E. Shaw, and J. T. Todd, "Wrinkling and head shape as coordinated sources of age-level information," *Perception Psychophys.*, vol. 27, no. 2, pp. 117–124, Mar. 1980.
[40] W. Abel, W. Enke, T. Fürst, E. Hanhart, T. Kemp, S. Koller, E. Kretschmer, O. Kroh, L. Loeffler, H. Luxenburger, and M. V. Pfaundler, *Methodik; Genetik der Gesamtperson: Zweiter Band*. Berlin, Germany: Springer-Verlag, 2019.
[41] M. A. Taister, S. D. Holliday, and H. I. M. Borrman, "Comments on facial aging in law enforcement investigation," *Forensic Sci. Commun.*, vol. 2, no. 2, pp. 1–11, 2000.
[42] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 2223–2232.
[43] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1125–1134.
[44] W. Wang, Y. Yan, Z. Cui, J. Feng, S. Yan, and N. Sebe, "Recurrent face aging with hierarchical AutoRegressive memory," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 3, pp. 654–668, Mar. 2019.
[45] Y. Liu, Q. Li, and Z. Sun, "Attribute-aware face aging with wavelet-based generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 11877–11886.
[46] H. Fang, W. Deng, Y. Zhong, and J. Hu, "Triple-GAN: Progressive face aging with triple translation loss," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 804–805.
[47] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*. [Online]. Available: http://arxiv.org/abs/1411.1784
[48] M. Georgopoulos, J. Oldfield, M. A. Nicolaou, Y. Panagakis, and M. Pantic, "Enhancing facial data diversity with style-based face aging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020, pp. 14–15.
[49] C. Shi, J. Zhang, Y. Yao, Y. Sun, H. Rao, and X. Shu, "CAN-GAN: Conditioned-attention normalized GAN for face age synthesis," *Pattern Recognit. Lett.*, vol. 138, pp. 520–526, Oct. 2020.
[50] S. Zhou, W. Zhao, J. Feng, H. Lai, Y. Pan, J. Yin, and S. Yan, "Personalized and occupational-aware age progression by generative adversarial networks," 2017, *arXiv:1711.09368*. [Online]. Available: http://arxiv.org/abs/1711.09368
[51] S. Palsson, E. Agustsson, R. Timofte, and L. Van Gool, "Generative adversarial style transfer networks for face aging," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 2084–2092.
[52] R. Rothe, R. Timofte, and L. V. Gool, "DEX: Deep EXpectation of apparent age from a single image," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 10–15.
[53] E. Pantraki and C. Kotropoulos, "Face aging as image-to-image translation using shared-latent space generative adversarial networks," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Nov. 2018, pp. 306–310.
[54] E. Pantraki, C. Kotropoulos, and A. Lanitis, "Leveraging Image-to-image translation generative adversarial networks for face aging," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2019, pp. 8370–8374.
[55] V. G. Thengane, M. B. Gawande, A. A. Dudhane, and A. B. Gonde, "Cycle face aging generative adversarial networks," in *Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS)*, Dec. 2018, pp. 125–129.
[56] Z. Wang, Z. Liu, J. Huang, S. Lian, and Y. Lin, "How old are you? Face age translation with identity preservation using GANs," 2019, *arXiv:1909.04988*. [Online]. Available: http://arxiv.org/abs/1909.04988
[57] C.-C. Wang, H.-H. Liu, S.-C. Pei, K.-H. Liu, and T.-J. Liu, "Face aging on realistic photos by generative adversarial networks," in *Proc. IEEE Int. Symp. Circuits Syst.*, May 2019, pp. 1–5.
[58] N. Sharma, R. Sharma, and N. Jindal, "An improved technique for face age progression and enhanced super-resolution with generative adversarial networks," *Wireless Pers. Commun.*, vol. 114, no. 3, pp. 2215–2233, Oct. 2020.
[59] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. Eur. Conf. Comput. Vis. Workshops*, 2018, pp. 1–16.
[60] M. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 700–708.
[61] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[62] C. N. Duong, K. G. Quach, K. Luu, T. H. N. Le, and M. Savvides, "Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 3735–3743.

[63] A. Heljakka, A. Solin, and J. Kannala, "Recursive chaining of reversible image-to-image translators for face aging," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.* Cham, Switzerland: Springer, 2018, pp. 309–320.

[64] K. Cho, B. van Merriënboer, C. Gulcehre, D. Bahdanau, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014, *arXiv:1406.1078*. [Online]. Available: http://arxiv.org/abs/1406.1078

[65] Z. Huang, S. Chen, J. Zhang, and H. Shan, "PFA-GAN: Progressive face aging with generative adversarial network," *IEEE Trans. Inf. Forensics Security*, vol. 16, pp. 2031–2045, 2021.

[66] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jan. 1991, pp. 586–587.

[67] I. Kemelmacher-Shlizerman and S. M. Seitz, "Collection flow," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1792–1799.

[68] L. Jia, Y. Song, and Y. Zhang, "Face aging with improved invertible conditional GANs," in *Proc. 24th Int. Conf. Pattern Recognit.*, Aug. 2018, pp. 1396–1401.

[69] Y. Sun, J. Tang, Z. Sun, and M. Tistarelli, "Facial age and expression synthesis using ordinal ranking adversarial networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2960–2972, 2020.

[70] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 2089–2093.

[71] H. Zhu, Q. Zhou, J. Zhang, and J. Z. Wang, "Facial aging and rejuvenation by conditional multi-adversarial autoencoder with ordinal regression," 2018, *arXiv:1804.02740*. [Online]. Available: http://arxiv.org/abs/1804.02740

[72] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning face age progression: A pyramid architecture of GANs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 31–39.

[73] A. Krizhevsky, I. Sutskever, and E. Geoffrey, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[74] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.

[75] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[76] J. Chen, J. Qian, and Z. Lai, "Coarse-to-fine age progression based on conditional GANs," in *Proc. 4th IAPR Asian Conf. Pattern Recognit. (ACPR)*, Nov. 2017, pp. 173–177.

[77] S. Liu, Y. Sun, D. Zhu, R. Bao, W. Wang, X. Shu, and S. Yan, "Face aging with contextual generative adversarial nets," in *Proc. 25th ACM Int. Conf. Multimedia*, Oct. 2017, pp. 82–90.

[78] L. Chen, X. Hu, and Z. Zhang, "Face aging with boundary equilibrium conditional autoencoder," *IEEE Access*, vol. 6, pp. 54834–54843, 2018.

[79] P. Li, Y. Hu, Q. Li, R. He, and Z. Sun, "Global and local consistent age generative adversarial networks," in *Proc. 24th Int. Conf. Pattern Recognit.*, Aug. 2018, pp. 1073–1078.

[80] X. Liu, C. Xie, H. Kuang, and X. Ma, "Face aging simulation with deep convolutional generative adversarial networks," in *Proc. 10th Int. Conf. Measuring Technol. Mechatronics Autom.*, Feb. 2018, pp. 220–224.

[81] J. Song, J. Zhang, L. Gao, X. Liu, and H. T. Shen, "Dual conditional GANs for face aging and rejuvenation," in *Proc. IJCAI*, Jul. 2018, pp. 899–905.

[82] J. Zeng, X. Ma, and K. Zhou, "CAAE++: Improved CAAE for age progression/regression," *IEEE Access*, vol. 6, pp. 66715–66722, 2018.

[83] D. Gou, S. Zhang, X. Ning, and W. Wang, "A face aging network based on conditional cycle loss and the principle of homology continuity," in *Proc. Int. Conf. High Perform. Big Data Intell. Syst.*, May 2019, pp. 264–268.

[84] Q. Li, Y. Liu, and Z. Sun, "Age progression and regression with spatial attention modules," 2019, *arXiv:1903.02133*. [Online]. Available: http://arxiv.org/abs/1903.02133

[85] X. Liu, Y. Zou, C. Xie, H. Kuang, and X. Ma, "Bidirectional face aging synthesis based on improved deep convolutional generative adversarial networks," *Information*, vol. 10, no. 2, p. 69, 2019.

[86] P. Li, Y. Hu, R. He, and Z. Sun, "Global and local consistent wavelet-domain age synthesis," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 11, pp. 2943–2957, Nov. 2019.

[87] S. Roy, "Applying aging effect on facial image with multi-domain generative adversarial network," *Int. J. Image, Graph. Signal Process.*, vol. 11, no. 12, p. 14, 2019.

[88] H. Yang, D. Huang, Y. Wang, and A. K. Jain, "Learning continuous face age progression: A pyramid of GANs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 499–515, Feb. 2021.

[89] X. Zhang, P. Wei, and N. Zheng, "Face age transformation with progressive residual adversarial autoencoder," in *Proc. Int. Joint Conf. Neural Netw.*, Jul. 2019, pp. 1–8.

[90] H. Zeng, H. Lai, and J. Yin, "Controllable face aging," 2019, *arXiv:1912.09694*. [Online]. Available: http://arxiv.org/abs/1912.09694

[91] X. Huang and S. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2017, pp. 1501–1510.

[92] X. Ning, D. Gou, X. Dong, W. Tian, L. Yu, and C. Wang, "Conditional generative adversarial networks based on the principle of homologycontinuity for face aging," *Concurrency Comput., Pract. Exp.*, p. e5792. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.5792, doi: 10.1002/cpe.5792.

[93] Q. Pham, J. Yang, and J. Shin, "Semi-supervised FaceGAN for face-age progression and regression with synthesized paired images," *Electronics*, vol. 9, no. 4, p. 603, 2020.

[94] M. Sheng, Z. Ma, H. Jia, Q. Mao, and M. Dong, "Face aging with conditional generative adversarial network guided by ranking-CNN," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr.*, Aug. 2020, pp. 314–319.

[95] S. Chen, C. Zhang, M. Dong, J. Le, and M. Rao, "Using ranking-CNN for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 5183–5192.

[96] Y. Sun, J. Tang, X. Shu, Z. Sun, and M. Tistarelli, "Facial age synthesis with label distribution-guided generative adversarial network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 2679–2691, 2020.

[97] H. Wang, V. Sanchez, and C.-T. Li, "Age-oriented face synthesis with conditional discriminator pool and adversarial triplet loss," 2020, *arXiv:2007.00792*. [Online]. Available: http://arxiv.org/abs/2007.00792

[98] H. Zhu, Z. Huang, H. Shan, and J. Zhang, "Look globally, age locally: Face aging with an attention mechanism," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2020, pp. 1963–1967.

[99] X. Bian and J. Li, "Conditional adversarial consistent identity autoencoder for cross-age face synthesis," *Multimedia Tools Appl.*, vol. 80, pp. 14231–14253, Jan. 2021.

[100] Z. Huang, J. Zhang, and H. Shan, "RoutingGAN: Routing age progression and regression with disentangled learning," 2021, *arXiv:2102.00601*. [Online]. Available: http://arxiv.org/abs/2102.00601

[101] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," 2012, *arXiv:1207.0580*. [Online]. Available: http://arxiv.org/abs/1207.0580

[102] Y. Alaluf, O. Patashnik, and D. Cohen-Or, "Only a matter of style: Age transformation using a style-based regression model," 2021, *arXiv:2102.02754*. [Online]. Available: http://arxiv.org/abs/2102.02754

[103] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 8110–8119.

[104] Z. He, M. Kan, S. Shan, and X. Chen, "S2GAN: Share aging factors across ages and share aging trends among individuals," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9440–9449.

[105] R. Bazin and E. Doublet, *Skin Aging Atlas: Caucasian Type*. Paris, France: Editions MED'COM, 2007.

[106] R. Bazin and F. Flament, *Skin Aging Atlas: Asian Type*. Paris, France: Editions MED'COM, 2010.

[107] R. Bazin and F. Giron, *Skin Aging Atlas: African-American Type*. Paris, France: Editions MED'COM, 2012.

[108] R. Bazin and V. Rubert, *Skin Aging Atlas: Indian Type*. Paris, France: Editions MED'COM, 2015.

[109] R. Bazin, F. Flament, and H. Qiu, *Skin Aging Atlas: Face Body*. Paris, France: Editions MED'COM, 2017.

[110] P. K. Chandaliya, P. Garg, and N. Nain, "Retrieval of facial images re-rendered with natural aging effect using child facial image and age," in *Proc. 14th Int. Conf. Signal-Image Technol. Internet-Based Syst. (SITIS)*, Nov. 2018, pp. 457–463.

[111] D. Deb, N. Nain, and A. K. Jain, "Longitudinal study of child face recognition," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 225–232.

[112] P. K. Chandaliya and N. Nain, "Conditional perceptual adversarial variational autoencoder for age progression and regression on child face," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–8.

[113] D. Deb, D. Aggarwal, and A. K. Jain, "Finding missing children: Aging deep face features," 2019, *arXiv:1911.07538*. [Online]. Available: http://arxiv.org/abs/1911.07538

[114] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[115] P. Chandaliya, A. Sinha, and N. Nain, "ChildFace: Gender aware child face aging," in *Proc. Int. Conf. Biometrics Special Interest Group*, Sep. 2020, pp. 1–5.

[116] D. Deb, D. Aggarwal, and A. K. Jain, "Identifying missing children: Face age-progression via deep feature aging," in *Proc. Int. Conf. Pattern Recognit.*, Jan. 2021, pp. 10540–10547.

[117] Y. Xiao and Y. Zhao, "Enhance gender and identity preservation in face aging simulation for infants and toddlers," 2020, *arXiv:2011.07431*. [Online]. Available: http://arxiv.org/abs/2011.07431

[118] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[119] X. Wu, R. He, Z. Sun, and T. Tan, "A light CNN for deep face representation with noisy labels," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 11, pp. 2884–2896, Nov. 2018.

[120] P. Dhar, A. Bansal, C. D. Castillo, J. Gleason, P. J. Phillips, and R. Chellappa, "How are attributes expressed in face DCNNs?" 2019, *arXiv:1910.05657*. [Online]. Available: http://arxiv.org/abs/1910.05657

[121] M. Q. Hill, C. J. Parde, C. D. Castillo, Y. I. Colon, R. Ranjan, J.-C. Chen, V. Blanz, and A. J. O'Toole, "Deep convolutional neural networks in the face of caricature," *Nature Mach. Intell.*, vol. 1, no. 11, pp. 522–529, 2019.

[122] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6626–6637.

[123] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4690–4699.

[124] S. I. Serengil and A. Ozpinar, "LightFace: A hybrid deep face recognition framework," in *Proc. Innov. Intell. Syst. Appl. Conf.*, Oct. 2020, pp. 1–5.

[125] C. Zhang, S. Liu, X. Xu, and C. Zhu, "C3AE: Exploring the limits of compact model for age estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 12587–12596.

[126] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2014, pp. 768–783.

[127] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, 2006, pp. 341–345.

[128] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 442–455, Apr. 2002.

[129] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3730–3738.

[130] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," in *Proc. Workshop Faces Real-Life Images, Detection, Alignment, Recognit.*, E. Learned-Miller, A. Ferencz, and F. Jurie, Eds., Marseille, France, Oct. 2008, pp. 1–15.

[131] N. Srinivas, K. Ricanek, D. Michalski, D. S. Bolme, and M. King, "Face recognition algorithm bias: Performance differences on images of children and adults," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2019, pp. 1–9.

[132] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2016, pp. 96–104.

[133] T. Zheng, W. Deng, and J. Hu, "Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments," 2017, *arXiv:1708.08197*. [Online]. Available: http://arxiv.org/abs/1708.08197

[134] C. N. Duong, K. Luu, K. G. Quach, and T. D. Bui, "Longitudinal face modeling via temporal deep restricted Boltzmann machines," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5772–5780.

[135] M. Minear and D. C. Park, "A lifespan database of adult facial stimuli," *Behav. Res. Methods, Instrum., Comput.*, vol. 36, no. 4, pp. 630–633, Nov. 2004.

[136] S. Moschoglou, A. Papaioannou, C. Sagonas, J. Deng, I. Kotsia, and S. Zafeiriou, "AgeDB: The first manually collected, in-the-wild age database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 51–59.

[137] P. Drozdowski, C. Rathgeb, A. Dantcheva, N. Damer, and C. Busch, "Demographic bias in biometrics: A survey on an emerging challenge," *IEEE Trans. Technol. Soc.*, vol. 1, no. 2, pp. 89–103, Jun. 2020.

[138] T. Alley, *Social and Applied Aspects of Perceiving Faces*. Hove, U.K.: Psychology Press, 2013.

MARCEL GRIMMER received the B.Sc. and M.Sc. degrees in applied mathematics and data science from the University of Applied Sciences Darmstadt, in 2018 and 2020, respectively. He is currently pursuing the Ph.D. degree with the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology (NTNU). His active research is focused on the generation and manipulation of synthetic face images. Further, his studies include the quality assessment of the synthetically generated face images in the context of face recognition and the evaluation of demographic bias.

RAGHAVENDRA RAMACHANDRA (Senior Member, IEEE) received the Ph.D. degree in computer science and technology from the University of Mysore, Mysore, India, and the Institute Telecom, and the Telecom SudParis, Evry, France (carried out as a collaborative work), in 2010. He was a Researcher with the Istituto Italiano di Tecnologia, Genoa, Italy, where he worked with video surveillance and social signal processing. He is currently appointed as a Full Professor with the Institute of Information Security and Communication Technology (IIK), Norwegian University of Science and Technology (NTNU), Gjøvik, Norway. He has authored several articles, holds patents, and is a reviewer for several international conferences and journals. His main research interests include deep learning, applied machine learning, statistical pattern recognition, data fusion schemes, and random optimization, with applications to biometrics, multimodal biometric fusion, human behavior analysis, and crowd behavior analysis. He was/is involved in various conference organizing and program committees and serving as an associate editor for various journals. He was/is participating (as a PI/a Co-PI/a contributor) in several EU projects, IARPA, USA, and other national projects. He received several best paper awards. He has served as an Editor for ISO/IEC 24722 standards on multimodal biometrics.

CHRISTOPH BUSCH (Senior Member, IEEE) received the Diploma degree from the Technical University of Darmstadt (TUD), Darmstadt, Germany, and the Ph.D. degree in computer graphics from TUD, in 1997. He joined the Fraunhofer Institute for Computer Graphics, Darmstadt, in 1997. He is currently a member of the Norwegian Biometrics Laboratory, Norwegian University of Science and Technology, Norway. He holds a joint appointment with the Faculty of Computer Science, Hochschule Darmstadt. He has been lecturing a course on biometric systems with DTU, Copenhagen, since 2007. He coauthored over 500 technical articles and has been a speaker at international conferences. His research interests include pattern recognition, multimodal and mobile biometrics, and privacy enhancing technologies for biometric systems. He is a Co-Founder of the European Association for Biometrics and a Convener of WG3 in ISO/IEC JTC1 SC37 on Biometrics.