Final Year Project : Literature Review

A systematic literature review for face aging synthetics with age and gender prediction

Na Li (17210325)

A thesis submitted in part fulfilment of the degree of

MSc. (Master) in Computer Science

Supervisor: Prof. G. C. Silvestre

UCD School of Computer Science

University College Dublin

June 19, 2019

Table of Contents

Abstract

As we know, many elements can be captured from our human faces as significant information for identification and each individual has a unique face at different lifetime. Especially, age and gender are the most important identified information for each person in many fields. Age and gender prediction by facial images is one of the most popular tasks in computer version as it can be implemented applications for entertainment or security. Face aging is a challenging task, which its goal is to generate synthetic faces in a given age group accurately as much as possible. Both tasks have a close connection and are designed and verified by many frameworks and datasets in computer vision, for example, Convolutional Neural Networks (CNNs) have proven to be a best basic framework for age and gender prediction from images. Meanwhile, Generative Adversarial Nets (GANs) have shown a powerful structure for face aging generation. Normally, an age classifier or a gender classifier for age and gender prediction as a module is built on face aging frameworks mainly because it can improve the performance and the quality of synthetic facial images. Having investigated tremendous academic sources, this paper aims to investigate state-of-the-art frameworks based on CNNs and GANs for gender and age prediction and face aging tasks respectively and a summary of the popular eight facial datasets that applying on the two tasks.

Keywords

Face aging, Gender and Age classification, Age estimation, Gender recognition, Deep neural networks, Convolutional Neural Network (CNNs), Generative Adversarial Nets (GANs)

1 Introduction

It is no doubt that there are many benefits doing face analysis in computer version and implementing on many applications, such as user identification, social interaction, face tracking and behavior recognition people's identities such as gender or age from a certain digital face image [1]. In recent world, many researchers are pursuing design more effective and precise algorithms of facial recognition or automatic recognition of human demographic traits, but the report of performance is still hard to achieve the public expectation. Specially, facial age and gender prediction technologies based on different facial datasets have been explored by divers frameworks. In the very recent years, Convolutional Neural Networks (CNNs) make more spectacular contributions for the age and gender estimation. Moreover, CNNs in deep learning has witnessed that the performance is satisfying, for example [1] handwritten digit recognition, automatic face detection, face verification and so forth. Another popular task is face aging which means synthetic faces in a certain group of ages. The task can be applied in cross-face recognition and entertainment, such as finding lost children or predicting people 's future faces and so on [2]. Recently, Googlfellow et al (2014) delivered the first paper for Generative Adversarial Nets (GANs) [3], many researches have been attracted by GANs and designed more advanced frameworks for face aging generation in computer vision. It is fully shown that GANs is an a practical and basic proposal to figure out the relative facing aging task in the future. In fact, both two kinds of technologies have close connection; The face aging framework is more complex than gender and age classification. The gender and age classifiers as pertrained model as a module could be transferred to a structure of face aging generation in order to robust the framework and to synthetic and verify face ageing images accurately, such as IPCGANs [2] uses age classification, faceNet CNN is used to recognize a person's identities as inputs [4].

This paper intends to adequately investigate the state-of-the-art CNN techniques for facial age and gender prediction and modern techniques with GANs for face aging generation for a further study of face generation combining the two aspects, which were researched by evaluating around 50 academic articles. This paper's outline is as follow, the section 2, facial age and gender prediction for estimator; the section 3, Face aging generation; the section 4, State-of-the-art Dataset;The paper ends with a comprehensive conclusion.

2 Gender and Age prediction for estimator

In recent years, there are many areas needing to automate recognition and surveillance systems, human faces analysis is one of the most topics in computer vision fields [5] such as face detection, facial expression recognition and gender and age classification. As for gender and age prediction, past approaches [6] were to generally classify attributes of face images and they also relied on the facial feature dimensions [7] or optimized face descriptors. Moreover, these approaches have employed classification schemes to design for age or gender estimation tasks such as a distance metric learning [8], texture-based estimation [9] or fusion schemes which is widely used in biometrics [10] combining with Sector Vector Machine (SVM) classification [11] or drop-out SVM for linear SVM training [12]. Currently, deep learning has tremendous breakthroughs for image classifications [13] as it explores multiple layers for image data processing, for feature extraction and transformation. Convolutional Neural Networks (CNNs) have been witnessed as well-known of an architecture for image recognition [14], mainly because CNNs owns the high sensitiveness for facial variations [15]. In this section, three models applying CNNs are presented for Age and gender prediction, that is, CNNs with different depths [1], CNNs with feedforwards attention technology [15] and group-aware deep feature learning (GA-DFL) based on deep CNNs [16].

2.1 CNNs with different depths

Antipov et al (2017) [1] based on exiting CNN-based approaches with two pretrained model VGG-16 and RestNet-50 for age and gender estimators to evaluate which depth of CNNs has better performance. The CNNs depth which means the number of hidden layers has fundamental importance in deep learning because it can solve specific problems by an different hierarchy of image descriptors. General speaking, the more complicated problem the deeper CNNs architecture is addressed it. In recent works, such as deep residual learning [17, 18], Antipov et al point that using fully convolutional CNNs which composed of only convolutional layers without fully-connected layers cannot have trainable parameters. They explain that the discriminative power of a CNN relies on the fully-connected layers, not just convolutional layers.

Therefore, in Antipov et al (2017) [1] work, they used different convolutional layers (fast_CNN_n where n= 2, 4, 6, 8) from 2 to 8 to study the impact of different layers or depths on the quality of gender or age CNNs. The evaluating ways are to calculate the Areas Under ROC curves (AUCs) for age prediction and Classification Accuracies (CAs) of different fast_CNN_n for gender prediction. The result shows that the best performance is only four convolutional layers (fast_CNN_4); meanwhile, there is not any strong impact on gender prediction result if the layer is up to six (fast_CNN_6). What's more, when the convolutional layer is up to 8 (fast_CNN_8), it leads to overfitting issue on the training set. As the depth for aging prediction CNNs, they analyzed that fast_CNN_4 is much better than fast_CNN_6 by 10% from the data of a positive correlation AE (age estimator). By the experiments and evaluations, they also indicate that gender estimator training does not show enough information which depths of CNNs suit to gender estimator, but AE CNNs shows the discriminative enough to provide the advantage of the depth of CNNs.

The pretraining and multi-task training are the independent techniques, which are included in "Transfer Learning" [19]. Intuitively, the concept of Transfer Learning is that the knowledge learned from one problem can be reused into another one in order to no recollect data and no need to rebuild the model from scratch [19]. For this question GR and AE, "Face Recognition" (FR) is selected as a pretraining model [1]. FR as a pretraining task allows training deep CNNs from scratch; and the FR task is similar with the GR and AE task as gender is a one of a person's identity and GR is as sub-problem of FR. Meanwhlie FR CNNs model can also implicitly encode enough age information for face representation learning [20]. As for the multi-task case, a multi-task CNNs can resolve many problems at the same time (GR and AE). When the multi-task CNNs is trained, it can learn extract information from sources (images). Antipov et al [1] compared both training learning methods for the problems of GR and AE, they confirm that FR pretraining is much better than using multi-task training for gender fast_CNN. They used two pretrained models VGG-16 [21] and ResNet-50 [13] as strategies for GR an AE. The detail steps are that CNN GR and CNN AE are trained independently following both are pretrained for FR. The result shows that VGG-16 is better for AE and RestNet-50 is the best model for GR.

2.2 CNNs with feedforwards attention technology

Attention-based CNN [15] is another interesting strategy to solve AE and GE problems. Their proposal is involving a novel feedforward attention mechanism in order to discover the more informative and reliable parts of a facial image for improving the age and gender classification. In fact, the neural network with attention mode is from Recurrent Neural networks [22], attention is a powerful mechanism that a neural network can show more detail specific regions of sources images. The main purpose of using attention is to reduce the task complexity and ignore irrelevant information. Rodriguze et al [15] indicate that RNNs (Recurrent Neural Networks) use attention mechanisms effectively as the RNN integrates necessary information which is extracted at the different time-steps [22]. Because of CNN models own the high variability of unconstrained

environments, the CNNs cannot give the same important information to all regions of an image, then using attention mechanisms on CNN models can be suitable to select the specific regions of faces to do further processing and to ignore background clutter. They call "Attention-base CNN models". The model consists of three sub-modules, the first one is an attention CNN (call "Where") which predicts the best attention grid for performing the glimpses; the second one is a patch CNN (call "What") to evaluate the higher resolution patches based on the attention grid; the third one is a Multi Layer Perceptron (MLP) integrating the information which is from CNNs and performs the final classification.

The attention CNN uses pretrained model VGG-16, the patch CNN is fed with the first convolutional layers of the attention CNN in order to reduce the computational requirements for this architecture. The MLP is used for the classifier which is fed with features from the pool5 layer of the attention CNN and the weighted features of the pool4 or pool3 from patch CNN [15]. And then there are two strategies to merge the feature maps for the final classifier, that is the first one is concatenated both features after an L2 normalization and the second one is learning a projection of the patch CNN feature maps to the space of the attention CNN feature map and then add them together. Finally, they used the resulting feature maps are fed to the final classifier. At last, Adience dataset is applied to evaluate the performance, the dataset contains unconstrained facial images which are datasets for face recognition under Real-World conditions [23]. They also use the Images of Group (IoG) dataset to test the generalization capability of the models on age recognition as IoG dataset has different pose, appearance and light, and the size of the faces is much smaller than Adience dataset. The result shows that they got a relative improvement of 8.75% for gender recognition and a 7.89% on age classification with using Adience dataset comparing other models. Using IoG show the generality of the proposed model has surpassed the state-of-the-art score from [15, 24]. Lastly, using MOPRH II [25] dataset shows that the model enhances CNNs even in constrained environments with centred faces and grey backgrounds, and the result is improved 4.47% as the enhanced CNNs owns the ability to perform detailed fixations in the most discriminative patches based on the contexts.

2.3 Group-aware deep feature learning (GA-DFL)

Group-aware deep feature learning (GA-DFL) [16] approach is only used for facial age estimation under deep convolutional neural networks. In GA-DFL, discriminative face representations are learned by images as input sources, meanwhile the aging order information also are utilized. Because of lacking face images for the same person covering a wide range of ages in the most facial dataset, GA-DFL model separates the chronological aging progress. Then face pairs in the same age groups are projected more closely, vice versa. In the GA-DFL approach, a multi-path deep CNN architecture based on VGG-16 architecture [21] with ReLu non-linearity to integrate different scales information from face images into the learned face presentation. The detail multi-path network architecture is that each face image starts with three kinds of scales (224 x 224 for VGG-16 Net [21], 64 x 64 and 32 x 32 for downsampling scales), in the two lower scales, they use convenet1 and FC1 (Fully connection), convnet2 and FC2. There are three L2 normalizations for the three scales. 4096 is the output dimension of each sub-net. At last, they also use L2 normalization for embedded feature, and going into the top layer face descriptor. Three different public datasets (FG-NET [26], MORPH (II) [25] and Chalearn Challenge Dataset [27]) are used for training and test, Liu et al (2017) [16] show that their model is powerful enough to the nonlinear problem which an nonlinear relationship between face images and age values comparing existing shallow models and hand-crafted features.

The model are powerful enough to solve nonlinear problems comparing with existing simple models and hand-crafted features [28]. The nonlinear problems are nonlinear relationships between facial images and age values. What's more, compared with other deep learning models, GA-DFL's

learning facial characterization can explore the ordinal relationship of facial similarity by integrating aging level and age difference information.

3 Face aging generation

Many image processing problems, computer graphics and computer vision can be set to translate from input images to corresponding output images, also call image-to-image translation. Face aging generation is a type of image-to-image translation, it is defined that face aging can render aesthetically a face image with natural aging and rejuvenating effects on the same as individual person's face [29]. There are two type of approaches for automatic face aging which are prototype approaches and physical mode approaches [30, 31]. The purpose of prototype approaches is that building an average face as prototypes for young and old groups and then transferring the texture difference among the prototypes to the test image.

From the technical point of view, CNNs have been to learn a parametric translation function through the input/output examples [32]; RNNs have been used for face aging, which is called Recurrent face aging (RFA) [29] to ensure generated face aging images are smooth and natural. Even though the most advantage of RNNs is that RNNs can memory the previous information in each time step. [29] point that evolution of new faces is slow leading to strong negative influence for the quality of images. Other research suggests that the face aging generation with a target age is related to style transfer [33]. The first researcher for style transfer is Gatys et al [34] who shows combing two kinds of images, one is the input images while another is artistic style image, then generating a new image from the input image with the same as artistic style from the artistic style image. However, facing age generation is different from style transfer, it needs to generate faces following the target age group, obviously, style transfer is difficult to apply face aging problem [2].

Currently, there are many state-of-the-art methods of face aging or new face generation with Generative Adversarial Net (GANs) [3], such as Conditional GANs [2,4,35], Conditional Adversarial Autoencoder (CAAE) [35], CycleGANs [31], Deep Convolutional GAN [36], style-based GANs [37] that is the most latest framework for synthetics of new faces from two different people's faces. In this section, comparing the above approaches (frameworks) corresponding to performances, two main latest approaches Conditional GANs, CycleGANs are presented and compared by their performances within different models.

3.1 Generative Adversarial Nets

It is an arduous task to product a generative model [31] as the rich distributions need to be captured by such model from which natural images come from. Generative Adversarial Nets (GANs) [3] is the most popular algorithm for image generation or image processing and also have proven that it is excellent to capture distributions from natural images and produce high visual images. GANs have two components [3, 35]: a generative G model that captures distributions of training samples and tries to generate new samples which are similar with training samples without detection. Whereas a discriminative D model that learns to clarify a image or sample is from the generative model G distributions or the image (data) distributions. In other words, the generator algorithm generates new data instances and the discriminator evaluates authenticity which means whether each instance of data belongs to the training data [3]. The two models (G and D) are multi-layer perceptrons. A noise vector z which is sampled from a normal distribution is given and then getting the input noise variables $p_z(Z)$, the generator maps z to a synthesized image x. The

discriminator takes x which is sampled from real image distribution $P_{data}(x)$ as input and tries to make the image x apart. So the generator G is trained to make the discriminator be unable to discriminate them, the discriminator D can increase the probability of identifying correct labels to the real examples from a training set and to generate new samples from G accurately. The formula of GANs is given as follows [3],

$$\min_G \max_D V(D,\ G)\ =\ E_{x\sim\ P_{data}(x)}\left[logD(x)\right] + E_{z\sim\ P_z(z)}\left[log(1-D(G(z)))\right]. \tag{1}$$

Apparently, two networks G and D are iteratively optimized against each other so this is a reason why to emphasize the name "adversarial" [38].

3.2 Conditional GANs

Basically, Conditional GANs (cGANs) is an extension of GANs model in order to allow the model to condition the extra information. Figure 1 demonstrates the basic structure of cGANs [38], the extract information y could be any kind of information, such as class labels (age groups), data from other styles etc. In the discriminator, x and y are presented as inputs and the formula of cGANs is updated from formula (1) as follows [38, 39],

$$\min_G \max_D V(D,\ G)\ =\ E_{x,y\sim\ P_{data}(x,y)}\left[\log D(x,y)\right] + E_{y\sim\ P_y, z\sim\ P_z(z)}\left[log(1-D(G(z,y),y))\right]. \tag{2}$$

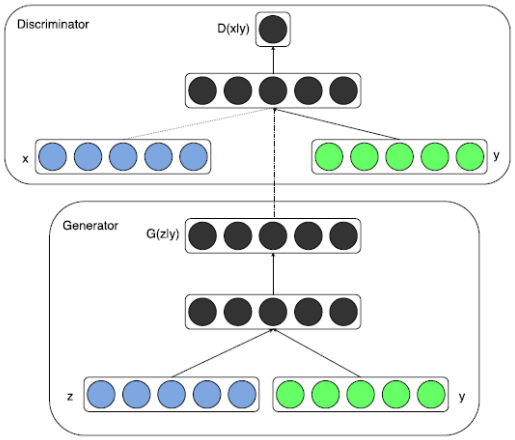

Figure 1: Conditional GANs framework [38]

Moreover, three different frameworks with the main concept of cGANs are illustrated for face aging generation.

Age-cGANs, Antipov et al (2017) [4] discussed one key issue at the GANs model for face generation is the original person's identity is usually lost during the generating fake faces. So they involved the concept of "identity-preserving" face aging and designed a new framework "Age-cGANs" (Age Conditional Generative Adversarial Network), then while an input face image is reconstructed, the original person's identity will be preserved. The first contribution that high-quality aging faces should require age categories which are defined six age categories: 0-18, 19-29, 30-39, 40-49, 50-59 and 60+ years old are generated. In the Age-cGANs framework (formula 2), y contains many information associated with the target face image such as facial pose or level of illumination. Another contribution is a novel "Identity-Preserving" latent vector optimization [4] to solve two issues of latent vector optimization [4, 40] which are an increasing of blurriness for reconstructions and unnecessary identities of input images may lead to strong impact on the level of the pixel. To solve the two issues, Antipov et al (2017) [4] involve a FaceNet CNNs

classifier [41] as an internal implementation of face recognition to recognize a person's identity from an input and use Euclidean distant can express difference identities from the original images and generated images. Thus, the distance is be minimized to improve identity preservation. L-BFGS-B algorithm [42] can solve the second issue as it can take advantage of a form of a limited memory approximation and initialled it with initial latent vector approximation. According to figure 2 2, it shows the two steps for face aging on the Age-cGANs model, the first step is reconstruct inputs with a centre age category in latent vector approximation; the second step is generating new face by the new target age simply switching the age at the input G in figure 2-b. At last, IMDB-Wiki_cleaned dataset [43] as training data contains around 120K images and is the subset of IMDB-Wiki dataset.

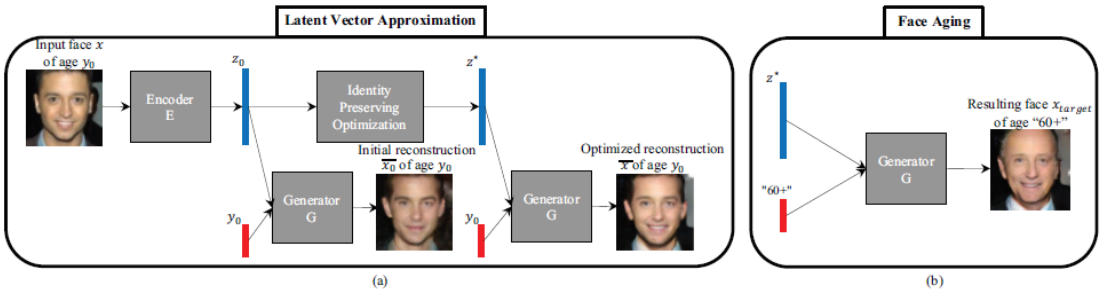

Figure 2: Two steps of face aging [4]

Conditional Adversarial Autoencoder (CAAE) [35] adds the concept of autoencoder which is a family of neural networks for which the "Input is the same as the output" based on cGANs. The main concept is that in the figure 3 the face images lie on a manifold, based on different ages and personalities, the images are grouped by the different directions on the manifold. Apparently, doing the age progression and regression at the same time is flexible through controlling the property of age. So a latent vector within a convolutional encoder can be projected the high-dimensional manifold with ages by a deconvolutional generator (decoder). The latent vector space is a compressed representation for images between encoder and decoder and the decoder only use the information from the space. Obviously, the network by the latent vector space (call bottleneck) may learn more and the most relevant features. Therefore, the synthetic faces generating by the generation is more realistic as the two adversarial networks are on the encoder and generator.

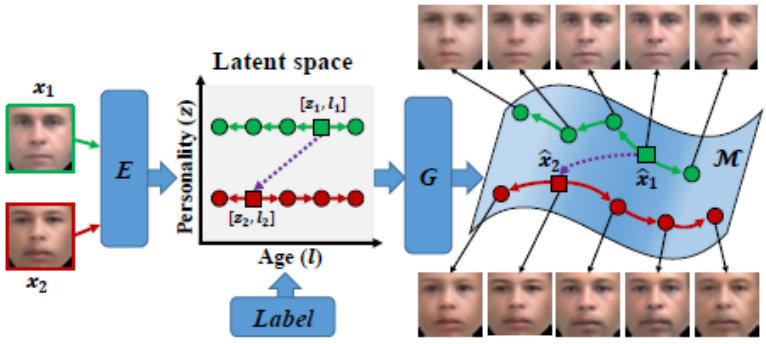

Figure 3: Instruction of traversing on the face manifold M [35]

Because of CAAE has the age progression and regression abilities [35], so it can generate more realistic face images. What's more, it does not use paired samples in the training data or labelled face for test data, which is more flexible and general. The latent vector space can preserve more specific personality to ensure accurate and clear synthetic faces for each age groups which separated ten groups of categories: 0-5, 6-10, 11-15, 16-20, 21-30, 31-40, 41-50, 51-60, 61-70, and 71-80. At last, they collect two face datasets from Morph dataset [44] and CACD dataset [45],

because the two datasets do not have enough image from newborn and the really old faces, so they have to grab the face images with different age groups from Bing and Google [35].

Identity-Preserved Conditional Generative Adversarial Networks (IPCGANs) [2] is the latest approach related to cGANs framework and the concept of "Identity-Preserved"which is similar with Age-cGANs but more complex. Figure 4 shows that three models of IPCGANs, a CGANs module can create a synthesized face following the target age and ensure the generated face looks realistic. so they use a Conditional LSGANs (Least Squares Generative Adversarial Networks) [46] model for face generation in the CGANs module as LSGANs have an ability to generate high quality face images and the process of training is stable. As for Identity-Preserved module, it is necessary to store identity information from the original face to the synthesized faces, they use pretrained AlexNet model as age classifier which has been trained model on ImageNet dataset. The age classification module is to ensure the synthesized faces belong to one target age group in five defined age groups corresponding to aged 11-20, 21-30, 31-40, 41-50 and 50+ respectively. Moreover, the age classifier is also fine-tuned AlexNet by adding two fully connected layers and a softmax layer and preventing overfitting problem by dropout. The Cross-Age-Celebrity Dataset (CACD) [45] dataset is used for training and evaluation.

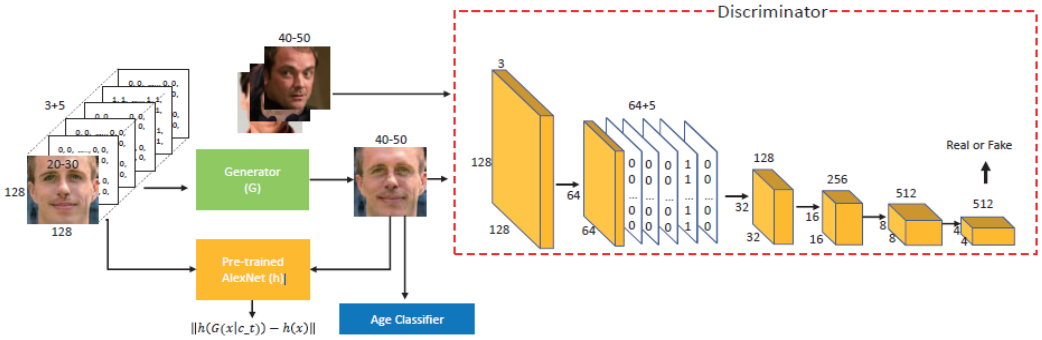

Figure 4: The pipeline of IPCGANs [2]

At last, Wang et al [2] compared performance of the IPCGANs model with other two frameworks which are CAAE and Age-cGANs from different aspects (face verification, image quality, age classification, VGG-face score and time cost). In the table 1, the best performance is IPCGANs.

	CAAE	Age-cGANs	IPCGANs
Face verification (%)	91.53	85.83	**96.90**
Image quality (%)	68.85	39.67	**71.74**
Age classification (%)	24.84	32.70	**31.74**
VGG-face score	19.53 ± 1.76	23.42 ± 1.82	$\mathbf{36.33 \pm 1.85}$
Time cost (s)	0.71	38.68	**0.28**

Table 1: The comparison of different frameworks [2]

3.3 Cycle-GAN: Group-GAN and FA-GAN

Cycle-GAN [31]. In fact, paired images are difficult for data collection and expensive for training data [47]. Thus it is also hard to obtain the paired face image with corresponding to a certain age for face aging generation. Cycle-consistent Generative Adversarial Networks (CycleGANs) [47] could fix this flaw of dataset. (G, D_x) and (F, D_y) are included in CycleGANs as two paired neural networks, then G and F are translators from X to Y and from Y to X respectively, D_x is a function to discriminate between input real images $\{x\}$ and generated images $F(x)$,

conversely, D_y is a another function to discriminate between images $\{y\}$ and $\{G(x)\}$, refer to CycleGAN figure 5. The two processes can be expressed by two cycle consistency losses by Zhu et al (2017) [47], the two losses can transfer from G to F and back to F to G, that is, a forward cycle-consistency loss: $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$ and a backward cycle-consistency loss $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$.

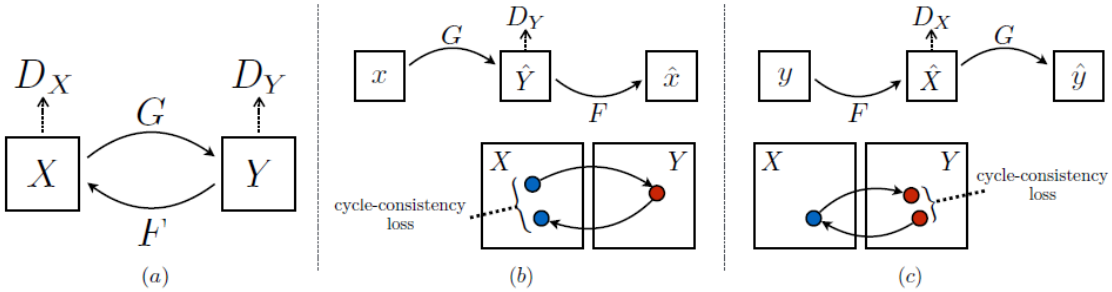

Figure 5: Pipeline of Cycle-Consistency [47]

Meanwhile, Palsson et al (2018) [31] show that the system is trained by an adversarial loss (3) which use a least square GAN loss [46] and a cycle consistency loss (4) in order to produce high quality facial images.

$$\mathcal{L}_{SGAN}(G,\ D_Y,\ X,\ Y)\ =\ \mathbb{E}_{y\sim\ p_{data}(y)}\left[(D_Y(y)\ -\ 1)^2\right] + \mathbb{E}_{x\sim\ p_d data(x)}\left[D_Y(G(x))^2\right]. \quad (3)$$

$$\mathcal{L}_{cyc}(G,\ F)\ =\ \mathbb{E}_{x\sim\ p_{data}(x)}\left[\|F(G(x)) -\ x\|_1\right] + E_{y\sim\ p_{data}(y)}\left[\|G(F(y)) - y\|_1\right]. \quad (4)$$

The full objective of CycleGANs is

$$\mathcal{L}(G,\ F,\ D_x,\ D_y)\ =\ \mathcal{L}_{LSGAN}(G,\ D_Y,\ X,\ Y) + \mathcal{L}_{LSGAN}(F,\ D_x,\ Y,\ X) + \lambda\mathcal{L}_{cyc}(G,\ F). \quad (5)$$

Group-GAN [31], the face dataset is separated small and large age groups which are 00-20, 20-40, 40-60, 60+ and 30-40, 40-50, 50-60, 60-70. Then cycleGANs model is trained winthn each groups. Finally, Palsson et al(2018) [31] verify that Group-GAN is effective for small age groups. CelebA dataset [20] have 202,599 face images of 10,177 different people for the experiment.

Face Aging-GAN (FA-GAN) [31] is another framework based on cycleGAN by Palsson et al (2018) [31] mainly because it is different to generate face aging images from a young face and an old face. So it is better to recognize the current face age from input real image. The FA_GAN adds a pretrained age prediction model by Deep Expectation of Apparent Age (DEX) [31, 48, 49] that applys VGG-16 architecture instead of using predefined age groups in Group-GAN framework.

As for a comparison of performances of age progression and age regression within these frameworks, Two kinds of ways to quantitatively compare face aging systems as a fair comparison should be decided to application [31]. Therefore, one way is that a survey is designed and participants vote on the best result, while another way is using a cross-age face verification system with generated facial images [29]. As a result, Palsson et al 2018 [31] compares other state-of-the-art GAN models, that is, CAAE 201 [35] and RFA (Recurrent Face Aging) [29], the F-GANs which is a fusion of FA-GANs and Group-GANs has better performance.

4 State-of-the-art Dataset

Server popular benchmark dataset of face images can be used in diversity of facing age approaches, gender and age classification. The summary of 8 benchmark datasets is in Table 2 follows below,

Datasets	Comments
LFW (The Labelled Faces in the Wild)	it is collected 13,233 photos, which is standard benchmark for face and gender recognition systems. [50]
MORPH-II	It is the largest public dataset of non-celebrities, and contain 50k face images, collected by American law enforcement services. [1, 25, 35]
Morph	It is the biggest dataset with multiple ages for each individual, including 55,000 images of 13,000 subjects from 16 to 77 years old [35, 44].
FG-Net	The database is smaller than others but has accurate ages for 60 subjects with 500 images. What's more, 45 subjects provided their images at the ages of 2, 5, 11, 13, 15, 20, 21, 22, 27, 28, and 30 years old [26].
CACD (Cross-Age-Celebrity Dataset)	It contains more than 160,000 face images of 2000 celebrities within the age ranging from 16 to 62 [2,35,45].
CelebA	It is a large-scale face images dataset and contains more than 200K celebrity images, each with 40 attribute annotations [20].
IMDB-Wiki	It contains face images with gender and age labels for training. [51].
IMDB-Wiki_cleaned	It is subset of IMDB-Wiki and contains 250,367 images without ambiguous images. [43].
Adience	It contains unconstrained facial images and is captured in the wild with 26.5 K images [12, 15].
IoG (Images of Groups)	It contains around 5.1k groups of people having 28.2k face images with age and gender labels. Meanwhile, the face images have different pose, appearance but it is smaller than Adience. [15, 24, 52].
Chalearn Challenge Dataset	It contains about 4112 images for training and 1500 images for validation with age ranging between 1 and 100. All images are collected in the unconstrained condition with divers poses, aspect ratio and low quality of images [16, 27].

Table 2: Summary of Datasets

5 Conclusion

In summary, this paper has mainly studied popular frameworks, technologies and comparison of datasets to solve age and gender prediction and face aging tasks. Firstly, three frameworks relating to CNNs, CNNs with different depths, CNNs with feedforwards attention technology and group-aware deep feature learning (GA-DFL) based on deep CNNs are demonstrated for age and gender prediction. Secondly, this paper presents varieties of frameworks under two main frameworks (cGANs, cycle-GANs) corresponding five variant frameworks (Age-cGANS, CAAE and IPCGANs; Group-CAN and FA-GAN) based on the GANs model. Through different strategies to compare performance among these frameworks, IPCGANs and F-GAN get the best performance comparing others in different basic frameworks. At last, eight datasets which both tasks involved have been summarized. It's even more remarkable that there is a strong close relationship between the two tasks, an age classifier usually is a module integrating into face aging model to enhance stability of framework, improve quantitative results and decrease computational cost. All in all, this comprehensive review is a fundamental study and summary for the further creative research and design for the relative facial identity and generation.

Bibliography

[1] G. Antipov, M. Baccouche, S.-A. Berrani, and J.-L. Dugelay, "Effective training of convolutional neural networks for face-based gender and age prediction," *Pattern Recognition*, vol. 72, pp. 15 – 26, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320317302534

[2] Z. Wang, X. Tang, W. Luo, and S. Gao, "Face aging with identity-preserved conditional generative adversarial networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[4] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adverarial networks," in *IEEE International Conference on Image Processing (ICIP)*, June 2017.

[5] S. Zaghbani, N. Boujneh, and M. S. Bouhlel, "Age estimation using deep learning," *Computers & Electrical Engineering*, vol. 68, pp. 337 – 347, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0045790617334298

[6] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2015.

[7] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1 – 21, 1999. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S107731429790549X

[8] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognition*, vol. 46, no. 3, pp. 628 – 641, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320312004037

[9] A. Unnikrishnan, F. Ajesh, and J. J. Kizhakkethottam, "Texture-based estimation of age and gender from wild conditions," *Procedia Technology*, vol. 24, pp. 1349 – 1357, 2016, international Conference on Emerging Trends in Engineering, Science and Technology (ICETEST - 2015). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2212017316302341

[10] K.-H. Liu, S. Yan, and J. Kuo, "Age estimation via grouping and decision fusion," *IEEE*, vol. 10, 2015.

[11] M. Castrilln-Santana, J. Lorenzo-Navarro, and E. Ramn-Balmaseda, "Descriptors and regions of interest fusion for in- and cross-database gender classification in the wild," *Image and Vision Computing*, vol. 57, pp. 15 – 24, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S026288561630186X

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

[13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[14] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Massachusetts Institute of Technology*, p. 2352 2449, 2017.

[15] P. Rodriguez, G. Cucurull, J. Gonfaus, X. Roca, and J. Gonzlez, "Age and gender recognition in the wild with deep attention," *Pattern Recognition*, 07 2017.

[16] H. Liu, J. Lu, J. Feng, and J. Zhou, "Group-aware deep feature learning for facial age estimation," *Pattern Recognition*, vol. 66, pp. 82 – 94, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0031320316303417

[17] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, K. Andrej, K. Aditya, M. Bernstein, C. A. Berg, and F. Li, "Imagenet large scale visual recognition challenge," *Springer Science+Business Media New York*, 12 2015.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[20] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *The IEEE International Conference on Computer Vision (ICCV)*, December 2015.

[21] S. Karen and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Vision and Pattern Recognition*, 2014.

[22] V. Mnih, N. Heess, A. Graves, and k. kavukcuoglu, "Recurrent models of visual attention," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2204–2212. [Online]. Available: http://papers.nips.cc/paper/5542-recurrent-models-of-visual-attention.pdf

[23] L. Lenc and P. Král, "Unconstrained facial images: Database for face recognition under real-world conditions," in *Advances in Artificial Intelligence and Its Applications*, O. Pichardo Lagunas, O. Herrera Alcántara, and G. Arroyo Figueroa, Eds. Cham: Springer International Publishing, 2015, pp. 349–361.

[24] S. E. Bekhouche, A. Ouafi, A. Benlamoudi, A. Taleb-Ahmed, and A. Hadid, "Facial age estimation and gender classification using multi level local phase quantization," in *2015 3rd International Conference on Control, Engineering Information Technology (CEIT)*, May 2015, pp. 1–4.

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

[26] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Toward automatic simulation of aging effects on face images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, April 2002.

[27] S. Escalera, J. Gonzlez, X. Bar, P. Pardo, J. Fabian, M. Oliu, H. J. Escalante, I. Huerta, and I. Guyon, "Chalearn looking at people 2015 new competitions: Age estimation and cultural event recognition," in *2015 International Joint Conference on Neural Networks (IJCNN)*, July 2015, pp. 1–8.

[28] L. Nanni, S. Ghidoni, and S. Brahnam, "Handcrafted vs non-handcrafted features for computer vision classification," *Pattern Recognition*, vol. 71, 06 2017.

[29] W. Wang, Z. Cui, Y. Yan, J. Feng, S. Yan, X. Shu, and N. Sebe, "Recurrent face aging," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[30] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 1955–1976, Nov 2010.

[31] S. Palsson, E. Agustsson, R. Timofte, and L. Van Gool, "Generative adversarial style transfer networks for face aging," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2018.

[32] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[33] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016. [Online]. Available: http://arxiv.org/abs/1603.08155

[34] L. A. Gatys, A. S. Ecker, and M. Bethge, "A neural algorithm of artistic style," *CoRR*, vol. abs/1508.06576, 2015. [Online]. Available: http://arxiv.org/abs/1508.06576

[35] Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[36] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Computer Vision and Pattern Recognition*, 2016.

[37] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," *CoRR*, vol. abs/1812.04948, 2018. [Online]. Available: http://arxiv.org/abs/1812.04948

[38] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014. [Online]. Available: http://arxiv.org/abs/1411.1784

[39] J. Gauthier, "Conditional generative adversarial nets for convolutional face generation," *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition*, 2014. [Online]. Available: https://www.foldl.me/uploads/2015/conditional-gans-face-generation/paper.pdf

[40] J.-Y. Zhu, P. Krähenbühl, E. Shechtman, and A. A. Efros, "Generative visual manipulation on the natural image manifold," in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 597–613.

[41] F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[42] R. H. Byrd, P. Lu, J. Nocedal, and C. Zhu, "A limited memory algorithm for bound constrained optimization," *SIAM J. Sci. Comput.*, vol. 16, no. 5, pp. 1190–1208, Sep. 1995. [Online]. Available: http://dx.doi.org/10.1137/0916069

[43] G. Antipov, M. Baccouche, S. Berrani, and J. Dugelay, "Apparent age estimation from face images combining general and children-specialized deep learning models," in *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, June 2016, pp. 801–809.

[44] I. Kemelmacher-Shlizerman, S. Suwajanakorn, and S. M. Seitz, "Illumination-aware age progression," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014.

[45] B.-C. Chen, C.-S. Chen, and W. H. Hsu, "Cross-age reference coding for age-invariant face recognition and retrieval," in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 768–783.

[46] X. Mao, Q. Li, H. Xie, R. Y. K. Lau, and Z. Wang, "Least squares generative adversarial networks," *CoRR*, vol. abs/1611.04076, 2016. [Online]. Available: http://arxiv.org/abs/1611.04076

[47] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[48] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image[c]," *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 10–15, 01 2015.

[49] R. "Rothe, R. Timofte, and L. Van Gool, ""deep expectation of real and apparent age from a single image without facial landmarks"," *International Journal of Computer Vision*, vol. 126, no. 2, pp. 144–157, Apr 2018. [Online]. Available: https://doi.org/10.1007/s11263-016-0940-3

[50] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments," *Tech. rep.*, 10 2008.

[51] R. Rothe, R. Timofte, and L. Van Gool, "Dex: Deep expectation of apparent age from a single image[c]," *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pp. 10–15, 01 2015.

[52] A. C. Gallagher and T. Chen, "Understanding images of groups of people," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, June 2009, pp. 256–263.