



Deep learning approach for facial age classification: a survey of the state-of-the-art

Olatunbosun Agbo-Ajala¹ · Serestina Viriri¹

Published online: 19 June 2020
© Springer Nature B.V. 2020

Abstract

Age estimation using face images is an exciting and challenging task. The traits from the face images are used to determine age, gender, ethnic background, and emotion of people. Among this set of traits, age estimation can be valuable in several potential real-time applications. The traditional hand-crafted methods relied-on for age estimation, cannot correctly estimate the age. The availability of huge datasets for training and an increase in computational power has made deep learning with convolutional neural network a better method for age estimation; convolutional neural network will learn discriminative feature descriptors directly from image pixels. Several convolutional neural network approaches have been proposed by many of the researchers, and these have made a significant impact on the results and performances of age estimation systems. In this paper, we present a thorough study of the state-of-the-art deep learning techniques which estimate age from human faces. We discuss the popular convolutional neural network architectures used for age estimation, presents a critical analysis of the performance of some deep learning models on popular facial aging datasets, and study the standard evaluation metrics used for performance evaluations. Finally, we try to analyze the main aspects that can increase the performance of the age estimation system in future.

Keywords Age estimation · Convolutional neural network · Deep learning · Facial aging

Abbreviations

| | |
|------|---------------------------------------|
| Ad | Adience |
| AFAD | Asian Face Age Dataset |
| BERC | Biometric Engineering Research Center |
| CACD | Cross-Age Celebrity Dataset |
| CNN | Convolutional Neural Network |
| CS | Cummulative Score |
| DEX | Deep EXpectation |

✉ Serestina Viriri
viriris@ukzn.ac.za

Olatunbosun Agbo-Ajala
ajalabosun@gmail.com

¹ School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Westville, Durban 4000, South Africa

| | |
|----------|---|
| DLDL | Deep Label Distribution Learning |
| ELM | Extreme Learning Machine |
| FC | Fully Connected |
| FG | FG-NET |
| FG-NET | Face and Gesture Recognition Network |
| GA-DFL | Group-aware Deep Feature Learning |
| HOIP | Human Object Interaction Processing |
| IMW | IMDb-WIKI |
| KL | Kullback–Leibler |
| LAP2015 | Looking At People 2015 |
| LAP2016 | Looking At People 2016 |
| LHI | Lotus Hill Research Institute |
| L5 | LAP2015 |
| L6 | LAP2016 |
| MAE | Mean Absolute Error |
| MC | Multi-class Classification |
| MP | MORPH-II |
| MR | Metric Regression |
| MRCNN | Multi-Region Convolutional Neural Network |
| ODFL | Ordinal Deep Feature Learning |
| OR-CNN | Ordinal Regression Convolutional Neural Network |
| RoR | Residual Networks of Residual Networks |
| VGGNET | Visual Geometry Group Network |
| WIT-DB | Waseda human–computer Interaction Technology |
| Xception | Extreme Inception |

1 Introduction

Age classification through face, plays a very significant role in our social beings. The human face conveys important information related to individual traits and the information serves as a pivotal role in face-to-face interaction among people. These behaviours are heavily dependent on our ability to estimate these individual traits: age, gender, facial expression, pose, and ethnic group, which are from facial appearances (Gallagher and Chen 2008). Those facial features are significant in our lives while the ability to predict them correctly and dependably from the facial image, is still far from satisfying the requirements of commercial applications (Badame and Jamadagni 2017).

However, regardless of the continuous research in the field of age estimation, where both academia and industry have devoted their effort to algorithm design, modeling, data collecting, system performance test, and valid evaluation protocols, it is still a challenging job in determining the accurate age of an individual (Angulu et al. 2018). The process of facial aging is influenced not only by intrinsic factors like change in size of the face, facial feature shape, wrinkles, face contour, facial feature distribution on the face, but also by extrinsic factors, such as the manner of living, health, eating habits, sociality, environment and weather conditions (Farkas et al. 2013).

In an automatic age estimation, obtaining the facial features needed for the face representations has been very challenging through the conventional methods employed by most of the existing techniques (Gurpinar et al. 2016). Many of those methods are hand-crafted,

which requires strong prior knowledge to engineer it by hand; they cannot be relied on to accurately predict human's age. Recently, Convolutional Neural Network (CNN) methods have been employed in the current age estimation analysis and have resulted in better age estimation accuracy performance. Deep learning and CNN as a learning-based feature representation methods, have been compelled to learn discriminative feature representation directly from raw pixels and acquired linear feature filters needed to project face images into another feature space (Liu et al. 2015b). The CNN architectures possess the ability to extract facial features distinctively and robustly from the face image (Antipov et al. 2016).

However, due to the challenging nature of facial estimation task, attempts to further improve the correctness of age estimation is still very much on and researchers keep studying different (CNN) approaches to further improve results. We, therefore, present a survey of the state-of-the-art deep learning approaches for facial age estimation.

Our contributions are summarized below:

1. We present a survey of different state-of-the-art CNN architectures for facial age estimation describing their strengths and weaknesses.
2. We also study the current facial aging benchmark databases and suggest their suitability to different age estimation techniques.
3. We outline different state-of-the-art algorithms and techniques used for age estimation highlighting their strengths and weaknesses.
4. We review the standard age estimation accuracy metrics and their area of applicability in age estimation.
5. We also present a concise report of different CNN methodology employed in the current state-of-the-art and their performance.

A typical reflection of age estimation system using CNN method is presented in Fig. 1. To the best of our knowledge, this is the first paper that presents a comprehensive review of CNN methods in age estimation, others like Angulu et al. (2018) and Fu et al. (2010) focussed on the hand-crafted and other machine learning methods.

The rest of this work is arranged as follows; in Sect. 2, we present different CNN architectures associated with age estimation, in Sect. 3, we discuss the popular facial aging datasets, Sect. 4 discusses different age estimation algorithms, Sect. 5 discusses previous works in age estimation using the CNN approach, Sect. 6 presents the discussion, we then summarize our contributions and conclusions in Sect. 7.

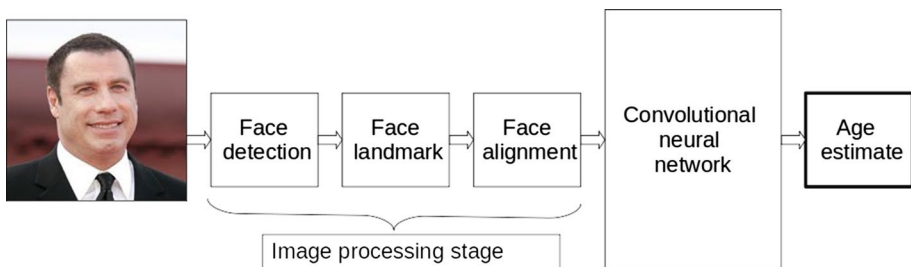


Fig. 1 A typical CNN age estimation system. An age estimation system follows a general process that includes face detection, image preprocessing (landmark detection and face alignment), features extraction (extracting the useful features from the input image), and the classification itself

2 A review of state-of-the-art CNN architectures

In recent times, deep learning and CNN have demonstrated a promising performance in feature learning and face recognition. It has the ability to learn discriminative trait descriptors directly from image pixels (Liu et al. 2017a). These traits are needed to correctly estimate the age of people. AlexNet, GoogLeNet, VGGNet, ResNet, SqueezeNet and Xception CNN architecture are generally considered as the most common architectures because of their state-of-the-art performance on different benchmarks including age estimation task. The following are the description of the architectures:

2.1 AlexNet architecture

One of the earliest CNN architecture for age estimation has been presented by Krizhevsky et al. (2017). AlexNet the winner of the 2012 edition of ImageNet Large Scale Visual Recognition Challenge (ILSVRC), was recorded as the first successful CNN architecture, trained on “imageNet” dataset with about 1.2 million labeled images of objects. AlexNet consists of a simple layout of eight layers with five convolutional and three fully-connected layers. The CNN architecture is similar to that of LeNet but deeper with stacked convolutional layers and more filters. The depth of AlexNet model contributed to the performance of works in Levi and Hassner (2015), Anand et al. (2017) and Agbo-Ajala and Viriri (2020). It is employed majorly to solve a very challenging facial analysis problem including age estimation, gender recognition etc. However, AlexNet is outperformed by more deeper models like ResNet and GoogleNet but it is computationally expensive.

Figure 2 shows a representation of the network architecture.

2.2 Visual Geometry Group (VGGNet) architecture

The authors in Karen and Andrew (2015) proposed an improved CNN architecture shown in Fig. 3. VGGNet is the most popular CNN architecture in the literature. It has a very small filter trained to increasingly higher depths (16–19 layers) to obtain a state-of-the-art result on the “ImageNet” classification challenge. The architecture is often used for transfer learning (features extractor and finetuning), as it shows an above average ability to generalize to datasets it was not trained on. VGGNet has proven to be effective in age estimation and this is observed in Rothe et al. (2015), Malli et al. (2016), Agustsson et al. (2017), Anand et al. (2017), Shara and Shemitha (2018), Qawaqneh et al. (2017), Rothe et al.

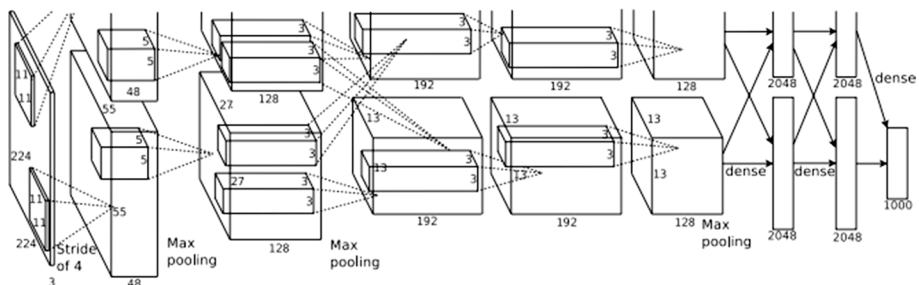


Fig. 2 The AlexNet architecture (Krizhevsky et al. 2017)

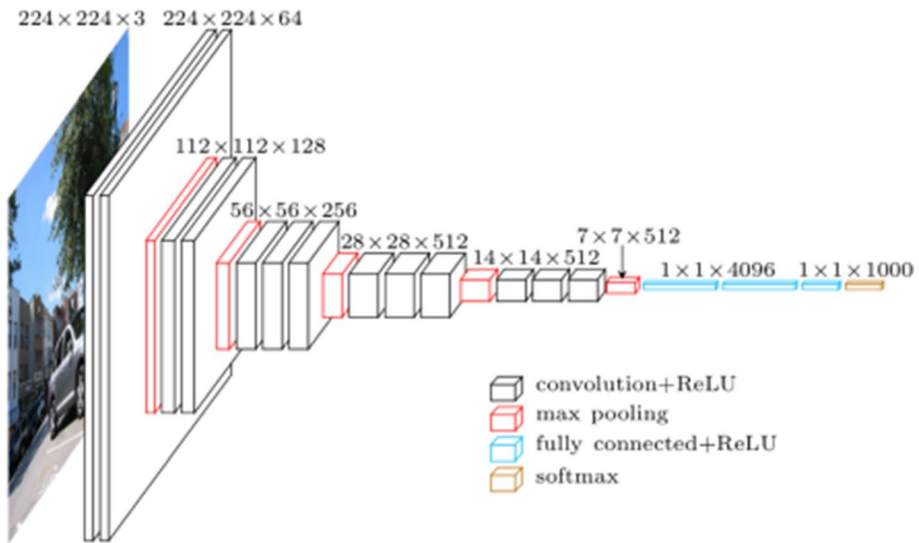


Fig. 3 The VGGNet architecture (Karen and Andrew 2015)

(2018), Nam et al. (2020), Li et al. (2019), Liu et al. (2017a) and Antipov et al. (2016). However, training VGG from scratch is time-consuming and also requires high computational power; the network is brutally slow to train, and the weights of the network architecture themselves are quite large. The network with the “fully-connected” layers and the backpropagation, contributed to its weight size. VGG architectures struggle to learn if they were randomly initialized and trained from scratch; the networks were simply too deep for basic random initialization. In order to solve the aforementioned problems, Karen and Andrew (2015) developed a “pre-training” approach, that only needed fewer weight layers for training before converging network weights as the initializations for deeper networks. However, training smaller variations of network architecture and then using the converged weights as initializations to the deeper versions of the network, is time-consuming especially for deeper networks with many fully-connected layers like VGG; it requires training and tuning the hyperparameters to achieve a good result. Nevertheless, VGG architecture has proven itself to be more suitable for generalization tasks. There is a smaller variation of VGG called “MiniVGGNet” that is well suited for smaller datasets.

2.3 GoogLeNet architecture

The authors in Szegedy et al. (2015) proposed a deeper and wider architecture (GoogLeNet) than AlexNet. It has a model weight of 28MB and this is lighter than VGG-16 (16 layers) and VGG 19 (19-layers). GoogLeNet uses “global average pooling” instead of “fully-connected” layers found in earlier architectures, and this dramatically reduces its weight size. Although GoogLeNet is a tiny architecture when compared to AlexNet and VGGNet, it outperformed the VGG model in the 2014 edition of the “ILSVRC”; the model makes use of a “network in network” or “micro-architecture” when constructing the overall “macro-architecture”. Goodman et al. (2015) proposed an improved version of Inception (Inception V3), to further boost “ImageNet” classification accuracy and this also impact the performance of the works in Liao et al. (2018) and Liu et al.

(2015b). However, for a smaller dataset (with smaller image spatial dimensions) which might require a fewer network parameter, a simplified version of the “inception module” (with lighter layers) can be used. Figure 4 presents a diagrammatic representation of the GoogLeNet network architecture.

2.4 ResNet architecture

ResNet architecture was developed by He et al. (2016a) primarily to improve the performance of the existing CNN architecture like VGGNet, GoogLeNet and AlexNet. It is an exciting network as it introduces the concept of “residual module” and “identity mappings” not in the existing architecture but capable of achieving state-of-the-art results. ResNet is deeper than VGG network but with a substantially smaller model size due to the use of “global average pooling” in place of “fully-connected” layers found in VGGNet. The strength of ResNet is the “residual module” introduced by the authors in He et al. (2016a). The “residual module” consists of two branches: The first is simply a shortcut which connects the input to the addition of the second branch, and a series of convolutions and activations. However, it was found that “bottleneck” an extension to the “residual module”, performs better when training a deeper network. Furthermore, He et al. (2016a) in their updated publication, experimented with the ordering of convolution, activation, and “batch normalization” layers within the “residual module” and they found that by updating the “residual module” to use “identity mappings”, higher accuracy could be obtained (Zhang et al. 2017). However, ResNet is computationally expensive with large depth; it takes a long time to train. There are other shallow variations of ResNet like “ResNet-10”, “ResNet-18”, and “ResNet-34” which will also generalize on less difficult task but will not obtain as high an accuracy as the deeper ones. It is fondly used when classifying a challenging and more difficult task of facial age estimation. Figure 5a, b are the representations of the architecture.

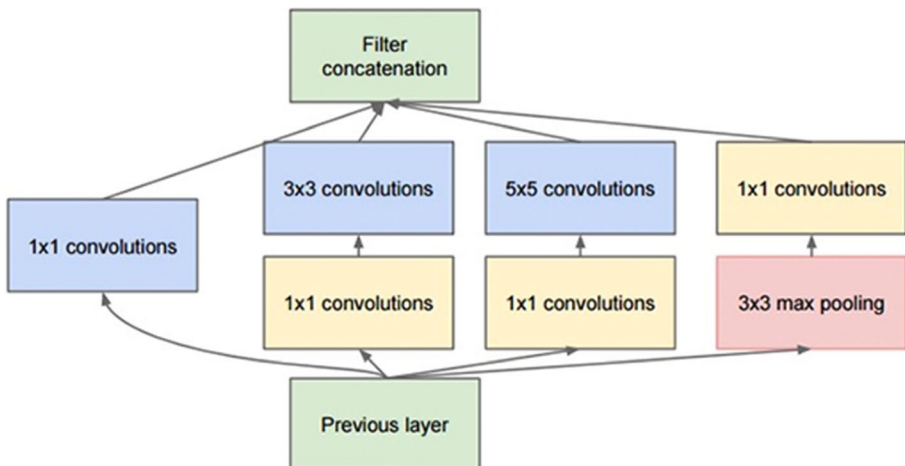
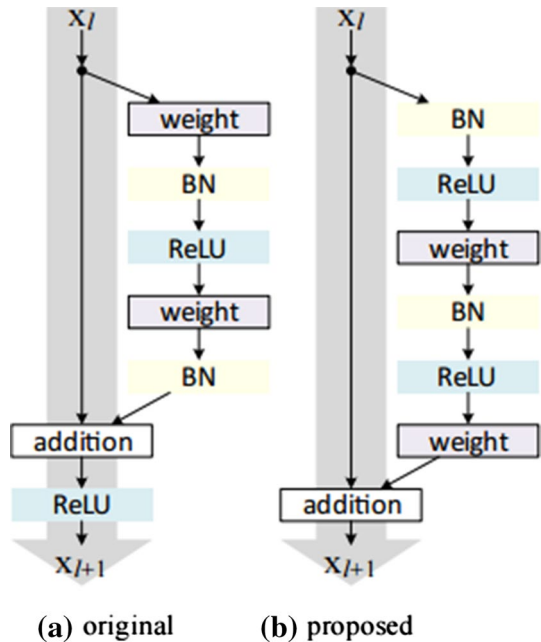


Fig. 4 The Inception module in GoogLeNet (Szegedy et al. 2015)

Fig. 5 The original and updated residual module in ResNet architecture (He et al. 2016b)



2.5 SqueezeNet architecture

Iandola et al. (2017), proposed a lighter CNN architecture that is often used when we need a tiny footprint. It is a tiny network when compared with VGGNet, ResNet, AlexNet and GoogLeNet, with model weighing in at 4.9M which can further be reduced to 0.5MB by model compression. SqueezeNet is often used when networks need to be trained and then deployed over a “network” and/or to “resource-constrained” devices. The network can be trained with a reduction in the number of parameters and still obtain a high level of accuracy. Squeezenet uses “fire module” that relies on an expansive and reduced phase of only 1×1 and 3×3 convolutions. The module reduces the spatial volume size in the network with its relatively low number of filters and the presence of “global average pooling”. “Global average pooling” acted in place of “fully-connected (FC)” layers, removing the FC layers, has the added benefit of reducing the number of parameters required by the network extensively. However, SqueezeNet is not included in the Keras core library. As shown in Fig. 6, the tiny nature of the architecture will affect its performance on the generalization.

2.6 Xception architecture

Recently, Chollet (2017) proposed “Xception” (“Extreme Inception”) network, which is an extension of the “Inception” architecture but replaces the Inception modules with “depthwise separable convolutions”. “Xception” is a convolutional neural network architecture based entirely on “depthwise separable convolution” layers with the smallest “weight serialization” at only 91MB. As presented in Fig. 7, the network architecture has 36 “convolutional layers” structured into 14 “modules”, all of which have linear “residual connections” around them, except for the first and last modules. The “Xception” architecture is a linear

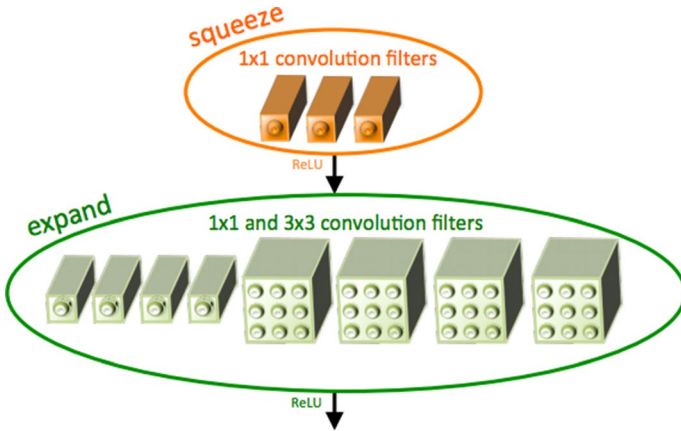


Fig. 6 The fire module in SqueezeNet architecture (Iandola et al. 2017)

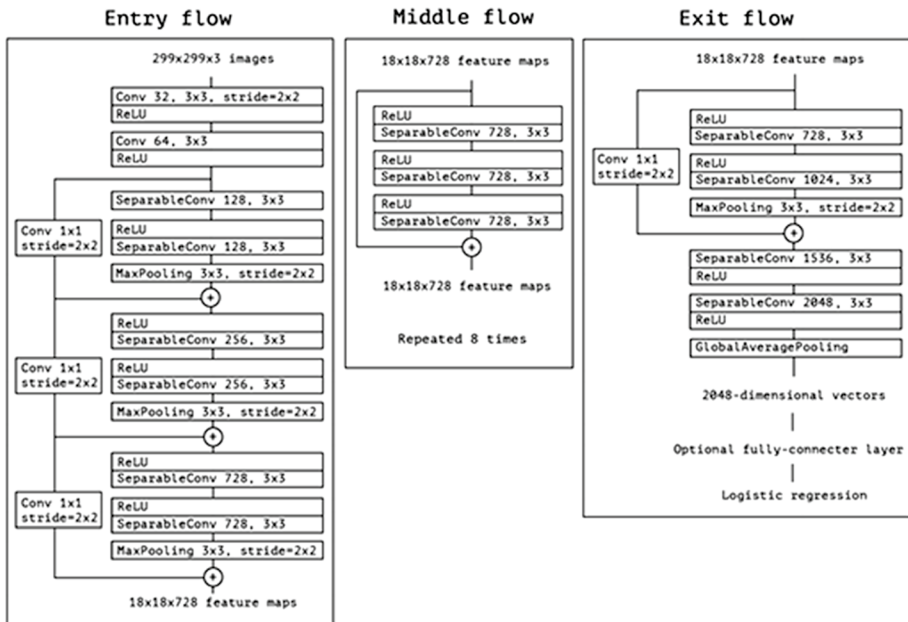


Fig. 7 The Xception architecture (Chollet 2017)

stack of layers with “residual connections” which makes the architecture easier to define and modify when compared to “Inception” architecture. On “ImageNet” dataset, “Xception” shows a better result not only on “Inception V3” but on “ResNet-50”, “ResNet-101” and “ResNet-152” (He et al. 2016a). However, “Xception” is marginally slower than “Inception” modules.

A description of the current state-of-the-art CNN architectures is presented in Table 1 while a typical CNN structure for age estimation is shown in Fig. 8.

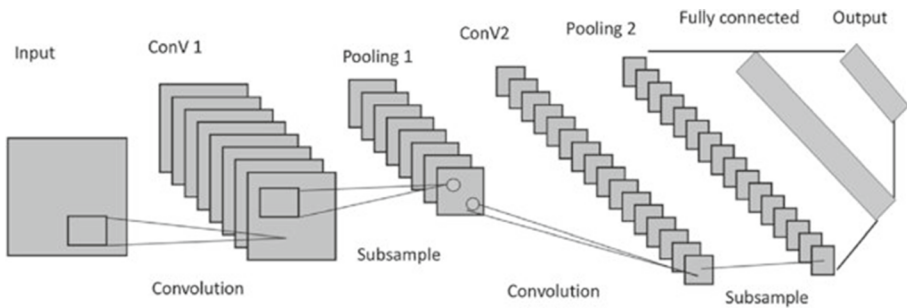


Fig. 8 A typical CNN structure for age estimation

3 An overview of facial aging databases

The availability of appropriate facial aging databases plays an important role in the evolution of a research field in age estimation. This allows researchers to get engaged in research activities quickly. In the research area related to soft biometrics, good quality age-separated face images are also of utmost importance to the success of the research, especially in the age estimation task (Panis et al. 2015). A brief description of some of the available facial aging databases are highlighted below:

3.1 Burt's Caucasian database

Burt's caucasian database (Burt and Perrett 1995) was developed by Burt and Perrett, to investigate “visual cues” to age through blending shape and colour information from many facial components. The database is made up of 147 face images of adult european males between the age of 20 and 62 years. The images reflect a neutral expression with shaved beards and without glasses or makeups. The facial images have 208 features points that encode facial shape.

3.2 3D morphable database

3D morphable database (Vetter 1999) is a powerful representation for human faces. The database contains a “3D scans of 200” adult faces (half males and half females) and 238 “teenaged faces” aged between 8 and 16 years; 125 males and 113 females with a resolution of “76,800 vertices” per scan (320 by 240).

3.3 Face and Gesture Recognition Network (FG-NET) database

FG-NET (Panis et al. 2015) is a publicly-available aging database that was collected in 2004 to support research activities in human facial aging. The database contains 1,002 face images of 82 different individuals whose ages are between 0 and 69 years. These facial images were produced by scanning the photographs found in personal collections of different subjects as such they reflect some significant variability in illumination,

Table 1 Description of state-of-the-art CNN architectures used in age estimation

| Architecture | #Parameters | Strengths | Weaknesses |
|----------------------------------|-------------|---|---|
| Krizhevsky et al. (2017) | 58,282,752 | AlexNet is a milestone CNN architecture for image classification Its overlapping pooling is a way to reduce overfitting problem | Its performance is outperformed by more deeper models like ResNet and GoogleNet It is also computationally expensive |
| Karen and Andrew (2015) | 137,479,232 | It is often used for transfer learning (features extractor and finetuning) as the network shows an above average ability to generalize to datasets it was not trained on It is also the most preferred option in the computer vision community for extracting features from images | Training VGG from scratch is time-consuming and also requires high computational power The network is too deep for basic random initialization |
| GoogLeNet (Szegedy et al. 2015) | 4,000,000 | It is lighter than VGG-16 and VGG-19 with model weight of 28 MB The network learns faster and more efficiently while increasing network depth | The network model is deeper and more complex than AlexNet or VGG-Face, will therefore demand more computational power |
| ResNet (He et al. 2016a) | 60,200,000 | It can be used for transfer learning | ResNet is computationally expensive network with huge depth It also take a long time to train the model so it demands high computational power |
| SqueezeNet (Iandola et al. 2017) | 3,753,856 | Global average pooling helps reduce and often eliminate entirely the need for FC layers; removing the FC layers has the added benefit of reducing the number of parameters required by the network substantially It can be used on a resource-constrained devices | SqueezeNet is not included in the Keras core library Also, the tiny nature of the architecture will affect its performance on the generalization |

Table 1 (continued)

| Architecture | #Parameters | Strengths | Weaknesses |
|-------------------------|-------------|---|---|
| Xception (Chollet 2017) | 21,577,728 | <p>The Xception architecture is a linear stack of layers with residual connections which makes the architecture easier to define and modify when compared to Inception architectures</p> <p>Xception is a very light architecture with weight serialization at only 91 MB</p> | <p>Xception is marginally slower when compared to Inception modules</p> |

image resolution, expression among others. Also, the images show some degree of occlusions of different forms in their appearance.

3.4 PAL database

PAL face database (Minear and Park 2004) was collected at the “university of michigan” by Meredith Minear and Denise Park to support research in the field of facial aging. It contains a total of 1142 face images of 576 subjects aged between 18 and 93. The database has four different age group comprising: 18–29 (218), 30–49 (76), 50–69 (123), and 70+ (158).

3.5 Face Recognition Grand Challenge (FRGC) database

FRGC database (Phillips et al. 2005) was collected in 2005 at the “university of notredame” as a contribution to “multi-modal” data collection for biometric applications. The database contains about 50,000 face images of 568 different subjects with an age span between 18 and 70 years old.

3.6 MORPH database

MORPH database (Ricanek and Tesafaye 2006) is a publicly-available aging database collected at the “university of north carolina” at wilmington by the face aging group. The database records linked attributes such as age, gender, ethnicity, weight, height, and ancestry. The whole database is divided into two albums; album I and album II. Album I contains a total of 1,724 face images of 515 individuals aged between 27 and 68 years. There are 1430 images of males and 294 images of females whose age gap is from 46 days to 29 years. Album-II includes 55,134 facial images gathered from more than 13,000 individual.

3.7 Waseda human–computer Interaction Technology (WIT-DB) database

WIT-DB database (Ueki et al. 2006) contains images of approximately 5500 different japanese subjects consisting of about 3000 males and 2500 females. There are 14,214 male face images and 12,008 female face images in total with a contribution of 1 to 14 image samples from each subject. WIT-DB contains facial images that are of “unoccluded frontal views” with an expression that is normal. The face images exhibit different degrees of variability in appearance that reflect a real-world image. The age label ranges from 3 to 85 years of age and divided into 11 age-groups.

3.8 The AI & R asian database

The AI & R Asian database (Fu and Zheng 2006) was collected for asian face rendering and synthesis. It contains four different databases that include: “AI and R V1.0” for expression, “AI and R V2.0” for aging, “AI and R V3.0” for viewing, and “AI and R V4.0” for illumination. “AI and R V2.0”, a database for age estimation, has a total of 34 face images with 17 subjects of age from 22 to 61 years. All the face images are “640 by 480” in dimension and “24-bit” colour depth.

3.9 Iranian face database

Iranian face database (Bastanfard et al. 2007) was collected in the year 2007 by the department of engineering, “the university of karaj”, for research activities in race detection, age estimation, facial surgery among others. The database contains 3600 coloured face images of 616 Iranian subjects comprising 487 men and 129 women with face images that are between the ages of 2 and 85. The facial images were captured using a white frame background and in a controlled environment with a high-resolution digital camera.

3.10 UIUC-IFP database

UIUC-IFP database (Guo et al. 2008b) is a non-publicly available database for age estimation. The database contains a total of 8000 high-resolution colour face images that is between 0 and 93 years. It has 1600 different asian natives of 800 males and 800 females with each subjects contributing a minimum of “five frontal images”. The frontal images were collected from an uncontrolled environment with different degrees of variation from illumination, makeup to facial expression.

3.11 Lotus Hill Research Institute (LHI) face database

LHI face database (Suo et al. 2008) contains 8,000 coloured face images of individuals from asians, half females and half males and one image per subject. This dataset covers an age range between 9 and 89 and contains an average of 100 images per age. The face images are at a resolution of about “120 by 160” pixels with a little pose and illumination variations.

3.12 Gallagher’s web-collected Database

Gallagher’s web-collected database (Gallagher and Chen 2009) is a collection of images from “flickr.com image search engine”. It contains 5080 images of 28,231 faces that are labeled with gender and age. The database has a total of seven different age groups: 0–2, 3–7, 8–12, 13–19, 20–36, 37–65, and 66+.

3.13 Ni’s web-collected database

Ni’s web-collected database (Ni et al. 2009) is one of the largest reported for facial aging. It is a collection of 219,892 faces in 77,021 images crawled from “google image search engine” and “flickr.com”. The database consists of age labels from 1 to 80 years.

3.14 Human & Object Interaction Processing (HOIP) database

HOIP face database (Fu et al. 2010) contains 306,600 images of 300 different subjects whose age range is between 15 and 64 years. The database has 10 different age-groups with an interval of five years that is distributed from 15 to 64 years.

3.15 Biometric Engineering Research Center (BERC) Database

BERC database (Choi et al. 2011) is a facial aging database collected and developed by “Biometric Engineering Research Center” (BERC). The database consists of face images of 390 individuals whose age range between 3 and 83 years. The images are of high resolution (3648×2736 pixels) and with no variations in facial expression and illumination. The images of the subjects are uniformly distributed with their age and gender.

3.16 Kyaw’s web-collected Database

Kyaw’s web-collected database (Kyaw et al. 2013) is created with the images gathered from the internet using “microsoft image search” application programming interface. The images are prepared by aligning the eye corner points and cropped to an image size of 65 by 75. The database contains a total of 963 different images divided into four age groups: 3–13, 23–33, 43–53, and 63–73.

3.17 OIU-Adience database

OIU-Adience database (Eidinger et al. 2014) is a collection of face images from an ideal real-life and unconstrained environments. It reflects all the features that are expected of an image collected from challenging real-world scenarios. They are face images that were uploaded to “flickr” website from smartphone without any filtering. “Adience” images, therefore, display a high-level of variations in noise, pose, appearance among others. It is used in studying age and gender classification system. The entire collection of OIU-Adience database is about 26,580 face images of 2284 subjects and with an age-group label of eight comprising: 0–2, 4–6, 8–13, 15–20, 25–32, 38–43, 48–53, 60+.

3.18 Cross-Age Celebrity Database (CACD)

CACD is a publicly-available cross-age face dataset, collected by Chen et al. (2015) in 2015. It is a collection of over 160,000 face images of “2000 celebrities” with their age ranging from 16 to 62 years.

3.19 Asian Face Age Database (AFAD)

AFAD, a large and publicly-available dataset, was collected by Niu et al. (2016). The dataset comprises 164,432 facial images of asian descends with an exact age ground-truths. It consists of 63,680 female images and 100,752 images of the male with the age

ranging from 15 to 40 years. AFAD images were collected from “RenRen social network”, a network of students to connect with each other and share pictures.

3.20 Aml-Face database

Aml-Face dataset (Anand et al. 2017) consists of face images captured in a laboratory at different distances using a “microsoft lifeCam HD-3000”. They collected the face images at 0°, 22°, 45°, 75°, and 90° of rotations. The images were acquired under different non-ideal conditions. The Aml-Face dataset is composed of 4535 face images of 16 different individuals.

3.21 IMDb-WIKI Database

IMDb-WIKI database (Zhang et al. 2017) is the largest publicly available dataset for age estimation of people in the wild, containing more than half a million images with accurate age labels between 0 and 100 years. For the IMDb-WIKI dataset, the images were crawled from IMDb and Wikipedia; IMDb contains 460,723 images of 20,284 celebrities and Wikipedia with 62,328 images. The images of IMDb-WIKI dataset are obtained directly from the website, as such the dataset contains many low-quality images, such as “human comic” images, sketch images, severe facial mask, full body images, multi-person images, blank images, and so on.

3.22 APPA-REAL database

APPA-REAL database (Agustsson et al. 2017) is the first state-of-the-art database with both real and apparent age labels. The images are collected using labeling application, “crowd-sourcing” data collection, data from the “AgeGuess platform”, and with the assistance of “Amazon Mechanical Turk” (AMT) workers. APPA-REAL database contains a total of 7591 images with the real and apparent age annotations. It has an age range that is between 0 and 95 of images of subjects that were taken under different conditions, which makes it more challenging for recognition purposes.

3.23 AgeDB database

AgeDB (Moschoglou et al. 2017) was collected for the purpose of research activities in age estimation, age-invariant face verification and face age progression. The manually collected face images reflect some degrees of variability that is expected of an image captured under completely uncontrolled, in-the-wild conditions. The variations appear in expressions, pose, occlusions etc. AgeDB contains a total of 16,488 images of actors, actresses, politicians, among others. All the images are labeled with their identity, age and gender attribute with a total of 568 different subjects of ages between 1 and 101 years.

3.24 UTKFace database

The UTKface UTKFace (2017) collected, is a very large face dataset with about 20,000 face images labeled with age, gender, and ethnicity. It has an age range of 0 and 116 years. The images cover large variation in facial expression, pose, illumination,

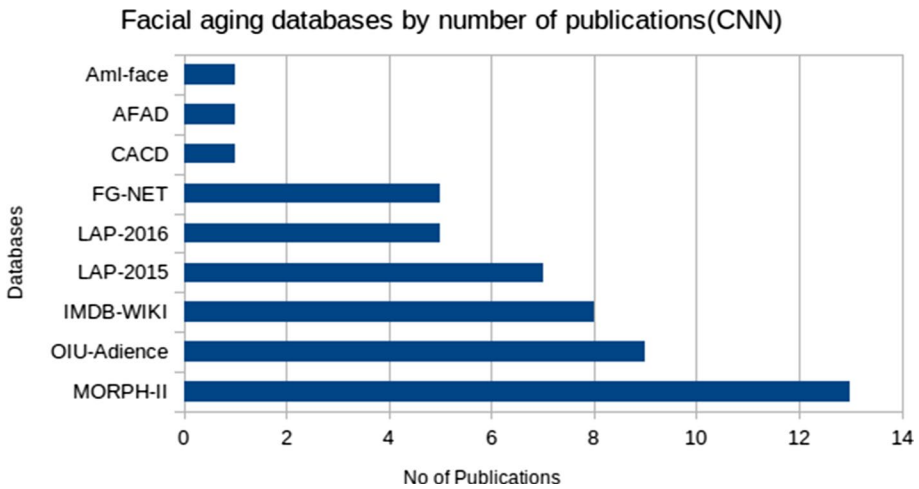


Fig. 9 A graph of aging databases by the number of publications

resolution, occlusion, among others. The dataset could be useful for research in age estimation, face detection, landmark localization, age progression/regression etc.

3.25 Summary

MORPH-II, IMDb-WIKI, OIU-Adience, CACD, AFAD, WIT-BD, HOIP, Gallagher's web-collected, Ni's web-collected, AgeDB, and UTKface databases are most suitable for age estimation using CNN techniques because of their large data size. However, Fig. 9 shows that MORPH-II has the largest number of usage in age estimation using CNN techniques; it is the most appropriate dataset while estimating ages of faces taken in a controlled environment while OIU-Adience and IMDb-WIKI are well suited for classifying age of face images from an uncontrolled real-world environment. Samples of the images in IMDb-WIKI, OIU-Adience and MORPH-II datasets are presented in Fig. 10. Table 2 also presents a summary of the suitability of all the datasets studied in this paper.

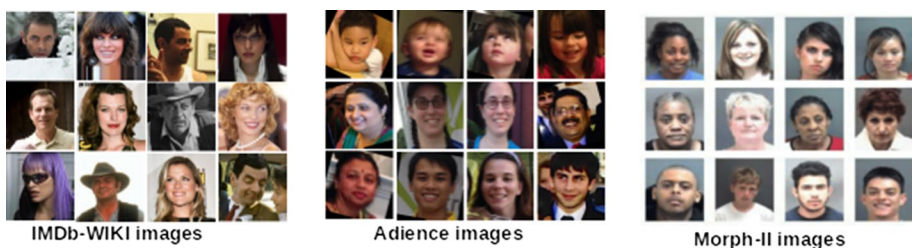


Fig. 10 Samples of images in IMDb-WIKI, Adience and MORPH-II datasets

Table 2 Summary of the Popular Facial Aging Databases

| Database | Database size | #Subjects | Age range | Age type | Year | In-the-wild? |
|---|---------------|-----------|-----------|-----------------|------|--------------|
| Burt's caucasian (Burt and Perrett 1995) | 147 | N/A | 20–62 | Real | 1995 | No |
| 3D morphable (Vetter 1999) | 438 | N/A | N/A | Real | 1999 | No |
| FG-NET (Pamir et al. 2015) | 1002 | 82 | 0–69 | Real | 2004 | Yes |
| PAL face (Minear and Park 2004) | 580 | 580 | 19–93 | Age group | 2004 | Yes |
| FRGC (Minear and Park 2004) | 50,000 | 568 | 18–70 | Real | 2005 | Partially |
| MORPH-II (Ricanek and Tesafaye 2006) | 55,134 | 13,618 | 16–77 | Real | 2006 | No |
| WIT-DB (Ueki et al. 2006) | 26,222 | 5500 | 3–85 | Age group | 2006 | Yes |
| AI&R asian (Fu and Zheng 2006) | 34 | 17 | 22–61 | Real | 2006 | No |
| Iranian face (Bastanfard et al. 2007) | 3600 | 616 | 2–85 | Real | 2007 | Yes |
| UIUC-IFP (Guo et al. 2008b) | 8000 | 1600 | 0–93 | Real | 2007 | Yes |
| LHI face (Suo et al. 2008) | 8000 | 8000 | 9–89 | Real | 2008 | No |
| Gallagher's web-collected (Gallagher and Chen 2009) | 28,231 | 28,231 | 0–66+ | Age group | 2008 | Yes |
| Ni's web-collected (Ni et al. 2009) | 219,892 | N/A | 1–80 | Real | 2009 | Yes |
| HOIP face (Fu et al. 2010) | 306,600 | 300 | 15–64 | Age group | 2010 | No |
| BERC (Choi et al. 2011) | 5910 | 95 | 3–83 | Real | 2011 | No |
| Kyaw's web-collected (Kyaw et al. 2013) | 963 | N/A | 3–73 | Age group | 2013 | Yes |
| OIU-Adience (Eidinger et al. 2014) | 26,580 | 2284 | 0–60+ | Age group | 2014 | Yes |
| CACD (Chen et al. 2015) | 160,000+ | 2000 | 16–62 | Real | 2015 | Yes |
| AFAD (Niu et al. 2016) | 164,432 | N/A | 15–40 | Real | 2016 | Yes |
| AmI-Face (Anand et al. 2017) | 4535 | 16 | N/A | Real | 2017 | No |
| IMDb-WIKI (Zhang et al. 2017) | 523,051 | 20,284 | 0–100 | Real | 2016 | Yes |
| APPA-REAL (Agustsson et al. 2017) | 7591 | 7000+ | 0–95 | Real & apparent | 2017 | Yes |
| AgeDB (Moschoglou et al. 2017) | 16,488 | 568 | 1–101 | Real | 2017 | Yes |
| UTKface (UTKFace 2017) | 20,000+ | N/A | 0–116 | Real | 2017 | Yes |

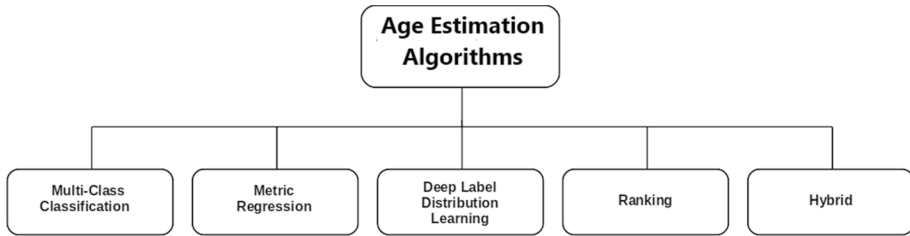


Fig. 11 Classification of facial age estimation algorithms. The typical age estimation methods are categorized into five different algorithms

4 Description of age estimation algorithms

In this section, we present different algorithms and techniques used in facial age estimation. As shown in Fig. 11, most of these techniques fall into five different categories; Age estimation can be modeled as a multi-class classification (MC), metric regression (MR), ranking, deep label distribution learning (DLDL) or an hybrid (combination of two or more techniques). We present a description of these algorithms and suggest the most effective approach in our opinion.

4.1 Multi-class Classification (MC)

Multi-class classification approach views the the ages or age groups category as an independent label and treats age value as a separate category and learns the age classifier to infer the person's age (Feng et al. 2017; Zhu et al. 2015; Malli et al. 2016). MC algorithm does maximize the probability of ground-truth class label by not even considering other classes. Nevertheless, the limited training samples and the class imbalance among most facial aging datasets can result in overfitting problem (Gao et al. 2018).

4.2 Metric Regression (MR)

Metric Regression-based algorithm views the age class as a linearly progressing relationship and does not display the diversity of the aging method. It learns the trait that most appropriate for the mapping of the age-value space from the feature space using the appropriate regularization method. Although, it is quite normal to address age estimation task as a MR problem, which does minimize the Mean Absolute Error (MAE) result, and improve the estimation accuracy performance. However, MR generates an unsteady training mode, causing a large error term, which affects accuracy performance. Some of the typical regression methods include Gaussian Process (Zhang and Yeung 2010), quadratic regression (Lanitis et al. 2004), Support Vector Regression (Guo et al. 2008a).

4.3 Deep Label Distribution Learning (DLDL)

DLDL approach converts real-value age to a discrete-age distribution to fit the entire age distribution. It is an end-to-end learning model that solves the problem of insufficient training images experienced in most age estimation task. It relaxes the demand for a large

number of training images and uneven distribution of the data by converting real age value to discrete age distribution to fit the whole age. The training instances connected with each class label will be increased without an increase in the number of training samples (Shen et al. 2017; Gao et al. 2018). However, it is usually observed that there is a lack of consistency between the employed evaluation metric and the training goals, therefore, generate an unsatisfactory result.

4.4 Ranking

The ranking-based algorithm uses age-axis tactics for age-class prediction and utilizes the relative order of the age. It uses relative age ranks instead of real age labels and ranks age class labels in the descending order using their relevance to the presented face images, to prevent making a decision for each age label that can simplify the problem (Chang et al. 2010; Li et al. 2012; Liu et al. 2017b). Nonetheless, ranking algorithm can generate sub-optimal results especially, when the training objectives and the evaluating metric, are not consistent.

4.5 Hybrid

Hybrid algorithm can be built by combining two or more algorithms in a parallel or hierarchical manner, to produce a better performance. The algorithm makes the most of the advantage of the strengths of each individual algorithm to obtain more robust system (Choi et al. 2011; Dib and El-saban 2010; Guo et al. 2008a). Unfortunately, combining two or more algorithms can result in large storage overhead and computational cost, therefore, affects its applicability in resource-constrained machines.

4.6 Summary of age estimation algorithms

In this section, we summarize the main strengths and weaknesses of the different age estimation algorithms. As presented in Table 4, most of the existing state-of-the-art methods used multi-class classification and hybrid algorithms. The hybrid algorithm combines two or more algorithms and this gives a better and more robust model compensating the weak points in the each algorithm with the strength of others. The ranking algorithm, on the other hand, solves the problem peculiar to the classification algorithm, by using the ordinal information of various ages to convert it into various binary classification problems. Metric regression makes some predictions from data by learning the relationship between the features of the data and some values. However, with deep label distribution learning, we can obtain a better model by using the adjacent ages to generate label distribution for each age, even when the label distribution of the dataset is uneven. Table 3 produces a summary of those state-of-the-art algorithms, stating their strengths and weaknesses.

5 A review of age estimation studies using CNN

In 2015, Levi and Hassner (2015) developed a shallow CNN architecture that used “three convolutional layers and two fully connected layers” to learn features representations. The simple convolutional neural network architecture was proposed out of the desire to reduce

Table 3 Description of state-of-the-art algorithms used in age estimation

| Algorithms | Brief description | Strengths | Weaknesses |
|---------------------------------|--|--|--|
| Multi-class Classification (MC) | MC method considers the age value as an independence category and later learns a classifier for the age classification task; It neglects the internal relationships between those age values | <p>It works well with small label classes</p> <p>MC maximizes the expectation of the ground-truth set without considering other classes</p> <p>It also presents age value as a separate category and later learns a classifier for age estimation</p> | <p>MC algorithm can easily lead to overfitting due to the imbalance problem among the age classes and insufficient training images</p> <p>The method can also lead to unstable training</p> |
| Metric Regression (MR) | MR models the relationship between a set of features and a continuous target variable. It makes some predictions from data by learning the relationship between the features of the data and some continuous-valued answer | MR minimizes the MAE; The smaller the MAE, the better the performance of an estimator | <p>Some outliers in the input data can cause a large error term which can lead to an unstable training process thereby producing an unsatisfactory performance</p> <p>MR presents age category as a linearly growing dependence rather than displaying the diversity of the aging process</p> <p>MR is sensitive to imbalanced training data</p> |
| Ranking | Ranking uses the age-axis approach for age estimation. It employs the relative order of age to solve the problem of classification bias caused by the unevenness of the sample images in the dataset | <p>The ranking model employs an age-axis approach that used the relative order of age for the age classification</p> <p>The method also transform age estimation to a series of binary classification problems where the output of the rankers is aggregated directly from those binary outputs for age classification</p> | <p>Ranking produces inconsistency in the training objectives and evaluation metric</p> <p>Ranking method may be suboptimal at times</p> <p>It achieves less accurate result than DLDDL and Hybrid</p> |

Table 3 (continued)

| Algorithms | Brief description | Strengths | Weaknesses |
|---|---|---|---|
| Deep Label Distribution Learning (DLDL) | DLDL is an end-to-end learning model that solves the problem of insufficient training images experienced in most age estimation task. It relaxes the demand for a large amount of training images and uneven distribution of the data by converting real age value to discrete age distribution to fit the whole age distribution | DLDL overcome the problem of uneven age label distribution by converting real age value to discrete age distribution to fit the whole age distribution DLDL also ease the necessity for a large number of images during training | DLDL may be suboptima. There might be an inconsistency during the training stage |
| Hybrid | A hybrid combines two or more modeling methods in a parallel or hierarchical manner to produce a better performance. It makes the most of the advantage of the strengths of each technique employed and is exacted to not only defeats other individual approaches but also make it robust | A hybrid model makes the most of the advantage of the strengths of each technique used, and it is expected to outperform other individual approaches Hybrid also produces a robust classifier | Combining two or more models might result in storage overhead and a huge computational cost It might be hard to deploy a hybrid model on resource-constrained devices It does inherit the weaknesses of the combined models |

Table 4 Summary of state-of-the-art methods and performance

| Algorithm | Publication | CNN architecture | Method | Database | Performance | | | |
|-----------|------------------------------|------------------|--|---------------------------|------------------|-----------|----------------------|------|
| | | | | | MAE | Exact (%) | 1-off (%) e-error | |
| MC | Levi and Hassner (2015) | AlexNet | CNN + dropout | Ad | N/A | 50.7 | 84.7 | N/A |
| | Rothe et al. (2015) | VGGNet | DEX | IMW, L5 | 3.22 | N/A | N/A | 0.26 |
| | Malli et al. (2016) | VGGNet | Ensembled model including Age-shifted grouping | IMW | N/A | N/A | N/A | 0.37 |
| | Agustsson et al. (2017) | VGGNet | Residual DEX | L5 | 4.45; 5.30 | N/A | N/A | N/A |
| | Anand et al. (2017) | VGGNet + AlexNet | Dimensionality reduction + FFNNs | WIKI, Aml-Face, Ad | 3.30 | N/A | N/A | N/A |
| | Aydogdu and Demirci (2017) | CNN | optimized CNN (4C2FC) | MP | N/A | 46.39 | N/A | N/A |
| | Qawaqneh et al. (2017) | VGGNet | VGG + dropout | Ad | N/A | 59.9 | 90.57 | N/A |
| | Zhang et al. (2017) | ResNets. | RoR | IMW-101, Ad | N/A | 67.3 | 97.51 | N/A |
| | Shara and Shermitha (2018) | VGGNet | CNN + KL loss | Private | 0.25 | N/A | N/A | N/A |
| | Rothe et al. (2018) | VGGNet | DEX | MP, FG, CACD, IMW, L5, Ad | 2.68; 3.09; 6.52 | 64.0 | 96.6 | 0.26 |
| MR | Nam et al. (2020) | VGGNet (DEX) | CNN + GAN | MP; PAL; FG | 9.42; 8.33; 8.56 | N/A | N/A | N/A |
| | Agbo-Ajala and Viriri (2020) | AlexNet | 4C2FC + dropout | Ad | N/A | 84.8% | 89.7% | N/A |
| | Ranjan et al. (2015) | 3NNR | CNN + Gaussian loss function | Ad, MP, L5 | N/A | N/A | N/A | 0.37 |
| | Niu et al. (2016) | CNN (MR) | OR-CNN | AFAD, MP | 3.34; 3.27 | N/A | N/A | N/A |
| | Li et al. (2019) | VGGNet | CNN + BridgeNet | MP; FG; L5 | 2.38; 2.56; 2.98 | N/A | N/A | 0.26 |

Table 4 (continued)

| Algorithm | Publication | CNN architecture | Method | Database | Performance | | | |
|-----------|------------------------|---|--|---------------------|------------------|-----------|----------------------|--------------|
| | | | | | MAE | Exact (%) | 1-off (%) e-error | |
| HYBRID | Liu et al. (2015b) | GoogLeNet | AgeNet | MP, CACD | 3.33 | N/A | N/A | 0.27 |
| | Gurpinar et al. (2016) | VGGNet | Kernel ELM + CNN | L6 | N/A | N/A | N/A | 0.37 |
| | Antipov et al. (2016) | VGGNet | Label distribution + classification encoding | IMW, L6, Private | N/A | N/A | N/A | 0.24 |
| DLDL | Liu et al. (2015a) | MRCNN (Multi-Region Convolutional Neural Network) | MRCNN | MP | 3.48 | N/A | N/A | N/A |
| | Liu et al. (2017a) | VGGNet | GA-DFL | MP, FG, L5 | 3.25; 3.93; 4.21 | N/A | N/A | 0.37 |
| | Duan et al. (2018a) | CNN | CNN + ELM | MP, Ad | 3.44 | 52.3 | N/A | N/A |
| RANKING | Duan et al. (2018b) | CNN | RAGN | MP, IMW, Ad, L6 | 2.61 | 66.5 | N/A | 0.37 |
| | Liao et al. (2018) | GoogLeNet | AgeNet + divide-&-rule | FG, MP, IMW | 4.02; 3.48; 3.29 | N/A | N/A | N/A |
| | Liu et al. (2020) | ShuffleNetV2 | MA-SFV2 | MP; FG | 2.68; 3.81 | N/A | N/A | N/A |
| DLDL | Huo et al. (2016) | VGGNet + new CNN | CNN (Gaussian + KL divergence) | MP, Private, FG, Ad | N/A | N/A | N/A | 0.31 |
| | Gao et al. (2018) | VGG-Net (ThinAgeNet) | DLDL-v2 | L5, L6, MP | 3.14; 3.45; 1.97 | N/A | N/A | 0.272; 0.267 |
| | Zhang et al. (2019) | Inception v4 | CNN | MP; FG | 1.32; 2.19 | N/A | N/A | N/A |
| RANKING | Chen et al. (2017) | CNN | ranking-CNN | MP | 2.96 | N/A | N/A | N/A |
| | Liu et al. (2018) | VGGNet | ODEL + OHRanker | MP, FG, L6 | 3.12; 3.89; 4.12 | N/A | N/A | 0.34 |
| | Liu et al. (2018) | 3C2FC | CNN + Multi-class Focal Loss | Ad | N/A | 54.0 | 88.2 | N/A |
| RANKING | Liu et al. (2019) | VGGNet | ODL (cross-entropy) | MP, FG, L6 | 2.92; 3.71; 3.95 | N/A | N/A | 0.312 |

3NVR 3-layer neural network regression, Ad Adience, DLDL Deep Label Distribution Learning, FFMV Feed-Forward Neural Network, FG FG-NET, GAV Generative Adversarial Network, IMW IMD6-WIKI, KL Kullback-Leibler divergence loss function, L5 LAP2015, L6 LAP2016, MA-SFV2 Mixed Attention-ShuffleNet, MC Multi-class Classification, MR Metric Regression, MP MORPH-II, MRCNN Multi-Region Convolutional Neural Network, N/A Not Applicable, RAGN Race-Net + Age-Net + Gender-Net + ELM classifier + ELM regressor

overfitting which can occur when the amount of training data is limited due to the limited availability of large facial aging databases. They also applied a “dropout regularization” and “data augmentation” methods to further limit the risk of overfitting. In spite of the simplicity of their network layout, it outperformed the existing state-of-the-art methods when it was evaluated on the challenging “adience benchmark for age and gender estimation”.

Liu et al. (2015b), proposed an “AgeNet”; an “end-to-end learning approach” for apparent age estimation. The designed network addressed the apparent age estimation problem by fusing “Gaussian label distribution based” classification models and “real value based regression” models. For the two models, a large-scale deep CNN was utilized to learn age feature representations. They also developed a deep transfer learning scheme to overcome the problem that might arise from overfitting. Consequently, the experimental result won a second place position at the “ChaLearn 2015 apparent age competition”, this demonstrated that the network (AgeNet) achieved the state-of-the-art performance in apparent age estimation.

To obtain more reliable system, trained, fine-tuned under unconstrained environment, Rothe et al. (2015) employed deep learning method “Deep EXpectation (DEX) of apparent age”, to tackle the estimation of apparent age in still face images. The architecture is a VGG-16 network that was initially pretrained on ImageNet dataset for image classification. They developed the largest public facial aging database by collecting about 500,000 face images of celebrities from the IMDb website and Wikipedia. They also fine-tuned the proposed model on apparent age labeled face images for better accuracy. Their method defined the age regression problem as a deep classification problem followed by a “softmax expected value refinement” and this shows an improvement over direct regression training of CNNs. DEX network ensembles the prediction of “20 networks” on the cropped face image and does not explicitly employ facial landmarks. The proposed method dramatically won the 1st place position at the ChaLearn LAP 2015 challenge on apparent age estimation outperforming other state-of-the-art methods.

Ranjan et al. (2015) approached an age estimation from unconstrained images with deep convolutional neural networks. The approach employed four different steps to carry out the task, this includes “face detection”, “alignment”, “deep feature extraction”, and “3-layer neural network regression”. The method obtained the needed features from the pool of a pretrained DCNN model and also adopted a “gaussian loss function” with a “3-layer neural network regression” model for age estimation task before adopting a “hierarchical learning” approach to enhance the method. The result showed that the “gaussian loss function” and the proposed “3-layer neural network regression” model outperformed the conventional “linear” model for age estimation.

The authors in Huo et al. (2016) proposed a method using deep CNN with “distribution-based (KL divergence) loss functions”. The architecture consists of two deep CNNs of different architectures as two streams: VGG-16 and a novel architecture. The VGG-16 was fine-tuned on three different datasets, the second model utilized different types of inputs of different augmentation methods to train this novel CNN model. Each deep CNN model was pre-trained on different datasets before they finetune the deep CNN models on the competition dataset. They have to fuse the results of two models to achieve the final predicted ages. Consequently, their approach achieved an ϵ -error of 0.3057, this won a fourth position at the “ChaLearn 2015 apparent age competition”. The model achieved a stronger prediction because it utilized extra 119,539 face images and other public facial datasets for training.

Gurpinar et al. (2016) proposed an apparent age estimation method with the use of deep learning for apparent age estimation from facial images. The method classified samples into different “overlapping age-groups”. The estimation of these age-groups was carried

out with a “local regressors” method before finally fusing all the age-groups for the final estimate. They utilized “Kernel extreme learning” machines for the classification. The proposed model was evaluated on the apparent age estimation dataset for “ChaLearn LAP 2015 challenge” and it achieved the 7th position with 0.374 normal scores on the “sequestered test set”. This showed that “local regressors” can perform better than the “global regressor” for almost all groups.

Antipov et al. (2016) developed a solution using a pretrained VGG-16 convolutional neural network. They trained the network on the huge IMDB-WIKI dataset and then fine-tuned it on the small dataset provided at the competition. They showed that the actual age estimation of children is the keystone of the competition. In view of that, they developed a separate VGG-16 network for children between 0 and 12 and trained it for apparent age estimation; the “children network” was separated from the “general” one. They employed separate “age encoding” strategies for training the “children” and the “general” networks; the strict one for the “children” network and “label distribution encoding” for the “general” network. The result of the experiment won a first place position at the 2016 edition of the ChaLearn LAP apparent age estimation.

Liu et al. (2015a) proposed a “Multi-Region Convolutional Neural Network” (MRCNN) for facial age estimation. The proposed method makes use of the “multiple sub-regions” which contain rich information on the age. It does this, by joining the “multiple face sub-regions” together for age estimation. The method utilized “8 network” and then constructed “8 sub-network” structures before fusing them at the feature level. The proposed model has two benefits: the “8 sub-networks” learned the unique age characteristics of the corresponding sub-region and the “8 networks” are packaged together to complement age-related information. Dramatically, the experimental result achieved a state-of-the-art performance when evaluated on MORPH-II database.

Malli et al. (2016) developed an ensemble of deep learning networks to achieve an apparent age estimation. They employed a fine-tuned VGG-16 convolutional neural networks architecture that was pretrained on the IMDB-WIKI database. They discovered that the apparent age is different from the real age of individuals. The real age is associated with a single age label while apparent age has multiple age labels associated with the face image. To solve the problem, they classified the face images that are within a defined age range together. They trained an ensemble of deep learning models using these age groups and their age shifted groupings, then combined the outputs of those model to achieve the final estimation. They solved the problem of imbalance in age distribution associated with the dataset by using an “adaptive data augmentation”.

Based on ordinal regression and deep learning, Niu et al. (2016) proposed an “end-to-end learning” approach to address the difficulty associated with “ordinal regression problems”; the first work to address “ordinal regression” problems through CNN. The approach employed deep CNN which could conduct “feature learning” and “regression” modeling at the same time. The proposed method is a multiple output CNN learning algorithm that collectively solves the series of “ordinal regression” sub-problems. As part of the solutions, they developed a dataset “Asian Face Age Dataset (AFAD)” with about 160K facial images with exact age. The approach achieved the state-of-the-art performance when evaluated on MORPH-II and AFAD datasets.

Agustsson et al. (2017) proposed a “deep Residual Deep Expectation” (DEX) method that posses the capacity to improve the performance of the original “DEX regressors” on age estimation tasks. The “original regressors” give a rough estimate of the age, by extracting the robust features from the input face image but the proposed “residual regressors” tackled the residuals between the “rough DEX estimation” and the “ground truth labels”

through a “specialized” model. The new “regressor” model allowed correction and also improved the performance of the “original DEX” showing improvements in age estimation task. As part of their solutions, they developed a large face image dataset “APPA-REAL”, with both real and apparent age annotations.

Anand et al. (2017) applied “post-processing” approaches to improve the performance of pre-trained deep networks. The method used the strategies to extract features from the input face image using pre-trained CNN. The proposed method implemented a “feature level fusion”, decreased the dimension of the feature space before finally estimating the age of the individual using a “Feed-Forward Neural Network” (FFNN). The age estimation method achieved better results than the state-of-the-art techniques when evaluated on “adience benchmark of unfiltered faces for gender and age estimation” and on a private (AmI-Face) dataset.

Aydogdu and Demirci (2017) proposed an “optimized deep CNN” architecture for the age estimation task. The proposed CNN architecture consists of four convolutional layers and two fully connected layers. The performance of the architecture was evaluated on a MORPH-II database and it outperformed other CNN architectures in their study using “exact success”, “top-3”, “1-off” criteria, and “standard deviation” values.

Based on ranking approach, Chen et al. (2017) developed a novel CNN-based architecture “ranking-CNN” for age estimation. The architecture has a series of fundamental CNNs trained on “ordinal age labels”. The binary outputs of the those CNNs are collected for the final age estimation. Through extensive empirical experiments, they demonstrated that their proposed method resulted in smaller estimation errors when compared with “multi-class” classification techniques. Consequently, “ranking-CNN” method remarkably outperformed other state-of-the-art age estimation models on benchmark datasets.

To overcome the problem of strong invariance of the model caused by datasets with various sale of data, a multi-path CNN model was proposed by Liu et al. (2017a) proposed a “Group-Aware Deep Feature Learning” (GA-DFL) technique for facial age estimation. “GA-DFL” method extracted the features required for face description by learning a “discriminative feature descriptor” directly from the raw pixels. In order to smoothen the adjacent age groups, they introduced an overlapped coupled learning method. They also employed a “multi-path” deep CNN architecture to integrate multiple scale information into the learned face presentation which further improved the performance of the method. They assessed the effectiveness of the proposed method on three publicly-available datasets on facial age estimation that were obtained in both controlled and uncontrolled conditions and it achieved a better performance when compared with most state-of-the-art facial age estimation methods.

Liu et al. (2018, 2019) proposed an “Ordinal Deep Feature Learning” (ODFL) method for facial age estimation. ODFL developed deep CNN to study “age-adaptive” face descriptors with CNN to utilize the “topology-aware ordinal relation” for face description. To achieve this, they ensured the “topology-aware ordinal relation” of face images is maintained in the learned feature place, they also ensured that the age distinction information of the embedded feature representation is employed in a “ranking-preserving” way. They evaluated the empirical result on four publicly-available datasets on facial age estimation and it showed an encouraging performance when compared with the current state-of-the-art methods.

Also, the authors in Qawaqneh et al. (2017) employed a VGG-Face network model that was trained on a database for face identification task. The deep CNN architecture consists of 11 layers; eight convolutional layers and three fully connected layers with each “convolutional” layer followed by a “rectification” layer, and a “max-pool” layer featuring at the

end of each convolutional block. The study also investigated a “GoogLeNet” architecture and was also trained on a very large database with millions of training images, but unfortunately, it could not outperform the proposed VGG model. The VGG-Net CNN was further finetuned and modified to perform age estimation task.

Zhang et al. (2017), proposed a new CNN based method “Residual Networks of Residual Networks (RoR)” for age group and gender estimation in the wild. The RoR model was first pretrained on ImageNet dataset before finetuning on the IMDB-WIKI-101 and adience database to achieve better learning ability of face images. They evaluated the effectiveness of the proposed RoR method for age and gender estimation on popular adience benchmark.

To overcome the problem of insufficient training data, Gao et al. (2018), proposed a CNN-based method that is based on label distribution. They designed a “lightweight” network architecture that is devoid of a large number of network parameters, reducing the computation cost and storage overhead; the model parameters is 0.9M. The proposed method is a “unified” structure which can collectively learn age distribution and regress age. The model was designed by unifying two existing current state-of-the-art age estimation methods into a single DLDL framework. They also proposed a “DLDL-v2” framework which eases the discrepancy between training and evaluation stages through collectively learning age distribution and regressing single age with shallow and deep network structure. This approach created a new state-of-the-art result on apparent and real age estimation tasks achieving comparable results as the state-of-the-art methods when evaluated on LAP2016 and MORPH-II datasets.

Duan et al. (2018a), introduced a “hybrid structure” of CNN and “Extreme Learning Machine” (ELM) in a “hierarchical” style for age estimation. The “hybrid architecture” utilized CNN to extract the features from the input images while the ELM classified the “intermediate results”. Subsequently, they established the performance of their hybrid structure on two popular datasets: MORPH-II and adience benchmark, the experiments showed that the hybrid framework achieved a better performance when compared with other results on the same face aging datasets.

Furher, in Duan et al. (2018b), the authors proposed an ensemble structure referred to as “CNN2ELM”, which includes CNN and “Extreme Learning Machine” (ELM) for age estimation. The model is a modification of the method used by Duan et al. (2018a). It consists of a three-level model including “feature extraction” and fusion, age grouping through an “ELM classifier”, and age estimation through an “ELM regressor”. They trained three networks to extract traits resembling age, gender, and race from the same image of a person during test and validation stages. Features associated with the age property were improved by fusing race and gender features. Then, to obtain a narrow age span, the ELM classifies the fusion results into one of the age groups. Subsequently, an age determination is executed using an “ELM regressor”. They pretrained the network on “ImageNet” database and then finetuned on the “IMDb-WIKI” database. They evaluated the effectiveness of the proposed network on adience benchmark, ChaLearn LAP2016, and MORPH-II, it outperformed the current state-of-the-art methods on age estimation tasks. It achieved the sixth position in “ChaLearn Looking At People 2016 apparent age estimation challenge” final results.

Liao et al. (2018) proposed an “AgeNet” and “divide-and-rule” architecture to estimate age. The “AgeNet” is a CNN-based network. The network was utilized to extract face age descriptor while the divide-and-rule strategy was employed for face age estimation. The “AgeNet” model used an approach based on regression and classification to construct an age estimated deep CNN. The network is a sturdy face age feature extractor model that possess a superior image representation capacity. The proposed “divide-and-rule” learning

model was to resolve the “ordinal regression” problem associated with the age estimation task. The experimental result on FG-NET, MORPH-II, and IMDB-WIKI showed that the “AgeNet” method and “divide-and-rule” age estimator achieved a better result than the conventional age estimation methods.

Also in Shara and Shemitha (2018), the authors proposed a multiple deep CNN that is based on VGG-face network for the facial age estimation. The age estimation method involved three different phases: a “training” phase, “feature extraction” phase, and “testing” phase. They also collected more than 10,000 age-labeled face images. They exploited the age information from the face images of age difference through the deep CNN-based model. The method utilized the “symmetric Kullback–Leibler divergence loss function” at the top layer of the model and used the “label distribution” for the loss function. The performance of the method was evaluated on the privately-collected images.

Later, Rothe et al. (2018), used “Deep EXpectation” (DEX); a deep learning solution that is based on VGG-16 architecture, to solve real and apparent age estimation from a single face image without the use of facial landmarks. They also introduced the IMDB-WIKI dataset, the largest public dataset of face images with age and gender annotations. The DEX model was initially pre-trained on both “ImageNet” and “IMDb-WIKI” datasets in order to achieve better performance. DEX learned from large data, utilizing a robust face alignment before formulating an expected value for age regression. They validated the DEX method on standard benchmarks: MORPH-II, FG-NET, and LAP2015 and it achieved a state-of-the-art result for both real and apparent age estimation.

The authors in Liu et al. (2018) developed a CNN architecture based on the “multi-class focal loss function” to increase the achievement of age estimation. Precisely, they designed an approach that approached the class inequality through reshaping the standard “cross entropy loss” that it down-weights the loss attached to well-classified samples; they studied the problem of excessive class imbalance amongst different age categories. They validated the approach on an audience benchmark and it showed that the proposed model achieved a significant improvement in performance for age estimation.

Li et al. (2019) proposed a CNN based technique, BridgeNet, for age estimation. The proposed model comprises two components; local regressors and gating networks that can jointly be learned in an end-to-end way. The first component (local regressors) addressed heterogeneous data by partitioning the data space. In contrast, the second one (gating networks), employed a bridge-tree structure that learns the continuity-aware weights used by the local regressors. Experimental results on the MORPH II, FG-NET, and Chalearn LAP 2015 datasets proved the CNN model to be effective, outperforming the state-of-the-art methods.

Liu et al. (2019) then developed a method that is an extension of their work in Liu et al. (2018). The work is an end-to-end ordinal deep learning (ODL) framework, including two ordinal regression loss functions; Square loss and Cross-Entropy loss. The proposed ranking-based ordinal deep feature learning (ODFL) method learns features needed for face representation directly from raw image pixels and then learn the procedures of feature extraction and age estimation independently. The work was evaluated on state-of-the-arts face aging datasets, and it achieves superior performance when compared with the state-of-the-art methods in age estimation.

Zhang et al. (2019), also proposed a novel method; recurrent age estimation (RAE). The CNN-based method makes use of the appearance features and the personalized aging patterns of input face images. RAE used an architecture that combines CNN and Long Short-Term Memory networks (LSTM); CNN is trained to extract discriminative appearance features from face images, while the LSTM network learns the personalized aging

patterns from sequences of face features. Furthermore, to exploit the ambiguity from the real age and adjacent ages, the authors employed Label distributed learning (LDL), and this consequently improved the experimental result by overcoming the problem of overfitting caused by small datasets. The experimental results show that RAE outperformed the existing approaches when it was evaluated on MORPH-II and FG-NET datasets.

Nam et al. (2020) solved the problem of age estimation of low-resolution facial images with a deep CNN-based model that reconstruct low-resolution faces as high-resolution faces. The CNN-based solution consists of a conditional generative adversarial network (GAN) that pre-processed low-resolution facial images before used as input. The model then used the state of the art CNN network architecture like ResNet, VGG, and DEX, for the age estimation of the reconstructed facial images. The experimental results on PAL, MORPH, and FG-NET databases, demonstrate the proposed method effectiveness in high-resolution reconstruction. It achieves state-of-the-art results in age estimation of low-resolution images.

Further, in Liu et al. (2020), the authors developed a lightweight CNN network (ShuffleNetV2), based on the mixed attention mechanism (MA-SFV2). The model; Mixed Attention-ShuffleNetV2, transforms the output layer, that model age estimation as a classification problem(that classify age as a separate label), regression problem (that rank the age of the human face having a particular order) and distribution learning (that consider the age correlation between adjacent ages). The model includes image pre-processing that reduces the influence of noise vectors and a data augmentation method like filtering, sharpening, histogram enhancement, etc., that increase the image size and alleviate the overfitting of the network. The model combines classification and regression and distributed learnings algorithms for the age estimation task. The experimental results on MORPH-II and FG-NET datasets, prove the applicability of the model in real-life situations, especially in mobile terminals.

Recently, Agbo-Ajala and Viriri (2020) developed a CNN-based model to classify unconstrained real-life face images into age and gender. The approach included an image preprocessing algorithm that prepares the input images and also a CNN architecture that does the feature extraction and the classification of the images to age and gender group. The experimental results were evaluated on the OIU-Adience dataset, and it confirms the effectiveness of their approach, outperformed other studies on the same dataset.

6 Discussion

Age estimation can be addressed as an exact age estimation or age-group estimation. Exact age estimation assigns an exact age label to a face image while age-group estimates the age range in which a facial image can fall. As presented in Table 3, the current techniques for age (groups) estimation fall into five different classes which include: “multi-class” classification, “metric regression”, “ranking”, “Deep Label Distribution Learning” (DLDL), and “hybrid” (combining two or more modeling methods); Age (group) estimation can be approached as “multi-class” classification, “metric regression”, “ranking”, DLDL or “hybrid” of two or more methods. Although, choosing among any of these approaches may be guided by the complexity of the problem, size of the dataset and the age distribution of the dataset used, it is very evident from the literatures that no concise conclusion can be given on the best class of deep learning approach for age estimation task. For huge and “evenly-distributed” datasets, any of the

approaches can be employed. For datasets with imbalance (uneven) age label distribution, age-group labels or insufficient training images, ranking and DLDL based approach may be more suitable. Combining two or more modeling approaches in a parallel or hierarchical manner produces a better performance when compared with a single CNN method. The hybrid method makes the most of the advantage of the strengths of each technique used and is expected to not only outperforms other individual approaches but also make it robust.

It is important to use established metrics to measure the extent to which alternative solutions, models, or approaches are able to meet expectations or goals through direct measurements of their strengths, deficiencies, and trade-offs. The standard evaluation metrics associated with facial age estimation are:

Mean Absolute Error (MAE) is one of the most commonly used metrics to evaluate the performance of each age estimation and the most suitable for measuring age estimation when there is a missing images while training data. The MAE is the average of the absolute errors between the estimated ages and ground truth ages (Onifade and Akinyemi 2014). As shown in mathematical Eq. (1), it is defined as an average performance of the age estimation techniques. The smaller the MAE, the better the age estimator effectiveness.

$$MAE = \sum_{k=1}^N \frac{|I_k - I_k^*|}{N} \quad (1)$$

where I_k : the estimated age, I_k^* : the ground truth age for the test image k , N : the total number of test images.

Cumulative Score (CS) is a useful metric when measuring the performance of an age estimator (Lu et al. 2015). It is most suitable when the training data has images at nearly every age. It can also be used as an indicator of the efficiency of an age-group classifier. Eq. (2) presents the mathematical function for CS. The larger the CS, the better the age estimator effectiveness.

$$CS(j) = \frac{N_{e \leq j}}{N} \times 100\% \quad (2)$$

where j : error level, N : the total number of test images, $N_{e \leq j}$: the amount of test images on which the age estimation performs an absolute error not greater than j (years).

Exact accuracy metric is also used to define the effectiveness of an age estimator (Levi and Hassner 2015). It is calculated as the percentage of face images that were classified into correct age-groups. Eq. (3) presents its mathematical equation.

$$Exact\ accuracy = \frac{no\ of\ accurate\ prediction}{total\ no\ of\ prediction\ made} \quad (3)$$

1-off evaluation metric measures whether the ground-truth class label matches the predicted class label (Aydogdu and Demirci 2017). It allows for a deviation of at most one bucket from the real age range. 1-off is calculated as a ratio of the correct predictions to the total number of data points.

Normal Score (ϵ -error) metric calculates the proportion of inaccurate predictions over the total number of instances evaluated (Zhu et al. 2015). The smaller the ϵ -error, the better the age estimator performance.

Equation (4) gives the mathematical definition of ϵ -error metric

$$\epsilon = 1 - e^{-\frac{(x-\sigma)^2}{2\mu^2}} \quad (4)$$

where x : the estimated age, σ : apparent age label provided for a given face image, μ : standard deviation of all estimated ages for the given face image.

However, there are major observations from earlier research work in the estimation which help in making some justifiable conclusions. Hence, we highlight those observations stating some of the points which we deem very important:

- It was observed that most of the existing work assume a constrained scenarios of the face images; the face in the input image are normalized with a frontal view. While other models employed a face pre-processing step that make face localization and alignment that prepare the images for the next steps
- It was also observed that the image processing method employed for face detection, facial landmark and face alignment has an impact on the performance of an age estimator.
- It is also important to assert that the performance of learning algorithms is determined by many factors, among which are the size and label distribution of the employed dataset, the degree of image variability, etc. The deep learning algorithms perform differently on different datasets, this is most likely due to the peculiarities of each dataset.
- Data augmentation improves the performance of an age estimation models, especially on an unevenly distributed not-too large datasets.
- We also observed from the literature that models pre-trained on large scale datasets before fine-tuning on the original dataset performed better than training the model on just the original dataset.
- From this review, we observed that multi-class classification algorithmn have been the most popularly used individual algorithm for age estimation in the literature.
- We also observed that ranking and DLDD are the most suitable algorithms for the estimation when the label distribution of the dataset is uneven.

7 Conclusion and future directions

In this paper, we present a comprehensive survey of various CNN architectures with their strengths and weaknesses. A thorough analysis of the existing deep CNN state-of-the-art methods used in age estimation was also discussed. We have also studied different facial aging dataset benchmarks available, their suitability and performance on different CNN models when measured on standard evaluating metrics. In Table 4, a summary of the performance evaluations of age estimation CNN methods on different databases is given.

There are a series of encouraging future research study that may see improvement in age estimation performance. Among them are:

- Age estimation of an individual from an unconstrained real-life face image is rapidly gaining more popularity, because of its many possible application. Although there has been a number of efforts towards achieving a result to a high level of accuracy, the results are quite not good enough because of the challenges experienced in an ideal world image with varying degrees of variations.
- A deeper network architecture, with sufficient and evenly-distributed training face images, should be investigated in the future works.

- Also, there is a need to investigate a study that focuses more on predicting apparent age (how old does the person look like?) rather than biological age (the real age of a person). This is useful in “face beauty product development”, “movie” and “theater role” casting, “plastic surgery”, “age-specific employment”, etc.
- Huge datasets with apparent age labels annotation rather than annotation with real age will also help improve accuracy in apparent age estimation research.

Comparative analysis of different approaches helps in understanding the implementation of a project in a better way. Therefore, this work is expected to serve as a guide in choosing the right method and approach for facial age estimation to further improve on the existing state-of-the-art results in the field.

Author contributions All authors had equal contribution towards this work. All authors read and approved the final manuscript.

Funding Not applicable.

Availability of data and materials Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

- Agbo-Ajala O, Viriri S (2020) Face-based age and gender classification using deep learning model. In: Dabrowski JJ, Rahman A, Paul M (eds) *Image and video technology*. Springer, Cham, pp 125–137
- Agustsson E, Timofte R, Escalera S, Baro X, Guyon I, Rothe R (2017) Apparent and real age estimation in still images with deep residual regressors on appa-real database. In: *Proceedings—12th IEEE international conference on automatic face and gesture recognition, FG 2017—1st international workshop on adaptive shot learning for gesture understanding and production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge*, pp 87–94
- Anand A, Labati RD, Genovese A, Munoz E, Piuri V, Scotti F (2017) Age estimation based on face images and pre-trained convolutional neural networks. In: *2017 IEEE Symposium Series on Computational Intelligence, SSCI 2017—Proceedings 2017*, pp 1–7
- Angulu R, Tapamo JR, Adewumi AO (2018) Age estimation via face images: a survey. *Eurasip J Image Video Process* 2018(1):42
- Antipov G, Baccouche M, Berrani SA, Dugelay JL (2016) Apparent age estimation from face images combining general and children-specialized deep learning models. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp 801–809
- Aydogdu MF, Demirci MF (2017) Age classification using an optimized CNN architecture. In: *Proceedings of the International Conference on Compute and Data Analysis—ICDA '17, September*, pp 233–239
- Badame V, Jamadagni M (2017) Study of approaches for human facial age. *Int J Innov Res Sci Eng Technol* 6(8):2347–6710
- Bastanfard A, Nik MA, Dehshibi MM (2007) Iranian face database with age, pose and expression. In: *Proceedings—International Conference on Machine Vision, ICMV 2007, January*, pp 50–55
- Burt DM, Perrett DI (1995) Perception of age in adult Caucasian male faces: computer graphic manipulation of shape and colour information. *Proc R Soc B Biol Sci* 259:137–143
- Chang KY, Chen CS, Hung YP (2010) A ranking approach for human age estimation based on face images. In: *Proceedings—International Conference on Pattern Recognition, August*, pp 3396–3399
- Chen BC, Chen CS, Hsu WH (2015) Face recognition and retrieval using cross-age reference coding with cross-age celebrity dataset. *IEEE Trans Multimed* 17(6):804–815

- Chen S, Zhang C, Dong M, Le J, Rao M (2017) Using ranking-CNN for age estimation. In: Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, vol 2017-January, pp 742–751
- Choi SE, Lee YJ, Lee SJ, Park KR, Kim J (2011) Age estimation using a hierarchical classifier based on global and local facial features. *Pattern Recognit* 44:1262–1281
- Chollet F (2017) Xception: deep learning with depthwise separable convolutions. In: Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017 January, pp 1800–1807
- Dib MYE, El-saban M (2010) Human age estimation using enhanced bio-inspired features (EBIF). *Faculty of Computers and Information, Cairo University, Cairo, Egypt*, pp 1–4
- Duan M, Li K, Yang C, Li K (2018a) A hybrid deep learning CNNELM for age and gender classification. *Neurocomputing* 275:448–461
- Duan M, Li K, Li K (2018b) An ensemble CNN2ELM for age estimation. *IEEE Trans Inf Forensics Secur* 13(3):758–772
- Eidinger E, Enbar R, Hassner T (2014) Age and gender estimation of unfiltered faces. *IEEE Trans Inf Forensics Secur* 9(12):2170–2179
- Farkas JP, Pessa JE, Hubbard B, Rohrich RJ (2013) The science and theory behind facial aging. *Plast Reconstr Surg Glob Open* 1(1):0–8
- Feng S, Lang C, Feng J, Wang T, Luo J (2017) Human facial age estimation by cost-sensitive label ranking and trace norm regularization. *IEEE Trans Multimed* 19(1):136–148
- Fu Y, Zheng N (2006) M-Face: an appearance-based photorealistic model for multiple facial attributes rendering. *IEEE Trans Circuits Syst Video Technol* 16(7):830–842
- Fu Y, Guo G, Huang TS (2010) Age synthesis and estimation via faces: a survey. *IEEE Trans Pattern Anal Mach Intell* 32(11):1955–1976
- Gallagher AC, Chen T (2008) Estimating age, gender, and identity using first name priors. In: 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, Section 3
- Gallagher AC, Chen T (2009) Understanding images of groups of people. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, vol 2009. IEEE, pp 256–263
- Gao BB, Zhou HY, Wu J, Geng X (2018) Age estimation using expectation of label distribution learning. In: IJCAI International Joint Conference on Artificial Intelligence, vol 2018-July, pp 712–718
- Goodman P, Namdeo A, Galatioto F, Bell M, Foster E, Shield C (2015) Rethinking the inception architecture for computer vision Christian. In: HARMO 2014—16th international conference on harmonisation within atmospheric dispersion modelling for regulatory purposes, Proceedings, pp 2818–2826
- Guo G, Fu Y, Dyer CR, Huang TS (2008a) Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans Image Process* 17(7):1178–1188
- Guo G, Fu Y, Dyer CR, Huang TS (2008b) A probabilistic fusion approach to human age prediction. In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp 1–6
- Gurpinar F, Kaya H, Dibeklioglu H, Salah AA (2016) Kernel ELM and CNN based facial age estimation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp 785–791
- He K, Zhang X, Ren S, Sun J (2016a) Deep residual learning for image recognition. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016 December, pp 770–778
- He K, Zhang X, Ren S, Sun J (2016b) Identity mappings in deep residual networks. In: Lecture notes in computer science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol 9908 LNCS, pp 630–645
- Huo Z, Yang X, Xing C, Zhou Y, Hou P, Lv J, Geng X (2016) Deep age distribution learning for apparent age estimation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp 722–729
- Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, Keutzer K (2017) Medical Image Computing and Computer-Assisted Intervention MICCAI. *ICLR* 8150:1–13
- Karen S, Andrew Z (2015) Very deep convolutional networks for large-scale image recognition. *Inf Softw Technol* 51:769–784
- Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. *Commun ACM* 60(6):84–90
- Kyaw SP, Wang JG, Teoh EK (2013) Web image mining for facial age estimation. In: ICICS 2013—conference guide of the 9th international conference on information, communications and signal processing

- Lanitis A, Draganova C, Christodoulou C (2004) Comparing different classifiers for automatic age estimation. *IEEE Trans Syst Man Cybern B Cybern* 34(1):621–628
- Levi G, Hassner T (2015) Age and gender classification using convolutional neural networks. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol 2015-October, pp 34–42
- Liao H, Yan Y, Dai W, Fan P (2018) Age estimation of face images based on CNN and divide-and-rule strategy. *Math Probl Eng* 2018:1–8
- Li C, Liu Q, Liu J, Lu H (2012) Learning ordinal discriminative features for age estimation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp 2570–2577
- Li W, Lu J, Feng J, Xu C, Zhou J, Tian Q (2019) Bridgenet: a continuity-aware probabilistic network for age estimation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2019-June*, pp 1145–1154
- Liu KH, Yan S, Kuo CC (2015a) Age estimation via grouping and decision fusion
- Liu X, Li S, Kan M, Zhang J, Wu S, Liu W, Han H, Shan S, Chen X (2015b) AgeNet: deeply learned regressor and classifier for robust apparent age estimation. In: *Proceedings of the IEEE International Conference on Computer Vision 2015*, pp 258–266
- Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE (2017a) A survey of deep neural network architectures and their applications. *Neurocomputing* 234:11–26
- Liu H, Lu J, Feng J, Zhou J (2017b) Ordinal deep feature learning for facial age estimation. In: *Proceedings—12th IEEE international conference on automatic face and gesture recognition, FG 2017—1st international workshop on adaptive shot learning for gesture understanding and production, ASL4GUP 2017, Biometrics in the Wild, Bwild 2017, Heteroge, May 2017*, pp 157–164
- Liu W, Chen L, Chen Y (2018) Age classification using convolutional neural networks with the multi-class focal loss. In: *IOP Conference Series: Materials Science and Engineering*, vol 428, no 1
- Liu H, Lu J, Feng J, Zhou J (2019) Ordinal deep learning for facial age estimation. *IEEE Trans Circuits Syst Video Technol* 29(2):486–501
- Liu X, Zou Y, Kuang H, Ma X (2020) Face image age estimation based on data augmentation and lightweight convolutional neural network. *Symmetry* 12(1):146
- Lu J, Liong VE, Zhou J (2015) Cost-sensitive local binary feature learning for facial age estimation. *IEEE Trans Image Process* 24:5356–5368
- Malli RC, Aygun M, Ekenel HK (2016) Apparent age estimation using ensemble of deep learning models. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pp 714–721
- Miner M, Park DC (2004) A lifespan database of adult facial stimuli. *Behav Res Methods Instrum Comput* 36(4):630–633
- Moschoglou S, Papaioannou A, Sagonas C, Deng J, Kotsia I, Zafeiriou S (2017) AgeDB: the first manually collected, in-the-wild age database. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, vol 2017-July, pp 1997–2005
- Nam SH, Kim YH, Truong NQ, Choi J, Park KR (2020) Age estimation by super-resolution reconstruction based on adversarial networks. *IEEE Access* 8:17103–17120
- Ni B, Song Z, Yan S (2009) Web image mining towards universal age estimator. In: *Proceedings of the seventeen ACM international conference on Multimedia—MM '09*, p 85
- Niu Z, Zhou M, Wang L, Gao X, Hua G (2016) Ordinal regression with multiple output CNN for age estimation. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2016 December, pp 4920–4928
- Onifade OFW, Akinyemi JD (2014) A GW ranking approach for facial age estimation. *Egypt Comput Sci J* 38(3):63–74
- Panis G, Lanitis A, Tsapatsoulis N, Cootes TF (2015) Overview of research on facial ageing using the FG-NET ageing database. *IET Biom* 5(2):37–46
- Phillips PJ, Flynn PJ, Scruggs T, Bowyer KW, Chang J, Hoffman K, Marques J, Min J, Worek W (2005) Overview of the face recognition grand challenge. In: *Proceedings—2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, vol 1, pp 947–954
- Qawaqneh Z, Mallouh AA, Barkana BD (2017) Deep convolutional neural network for age estimation based on VGG-face model
- Ranjan R, Zhou S, Chen JC, Kumar A, Alavi A, Patel VM, Chellappa R (2015) Unconstrained age estimation with deep convolutional neural networks. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp 351–359
- Ricanek K Jr, Tesafaye T (2006) MORPH: a longitudinal image age-progression, of normal adult. In: *Proc. 7th Int. Conf. Autom. Face Gesture Recognit*, pp 0–4

- Rothe R, Timofte R, Van Gool L (2015) DEX: deep expectation of apparent age from a single image. In: Proceedings of the IEEE International Conference on Computer Vision, vol 2015-February, pp 252–257
- Rothe R, Timofte R, Van Gool L (2018) Deep expectation of real and apparent age from a single image without facial landmarks. *Int J Comput Vis* 126(2–4):144–157
- Shara MS, Shemitha PA (2018) Facial age estimation based on multiple CNN. *Ijirt* 4(11):593–599
- Shen W, Zhao K, Guo Y, Yuille A (2017) Label distribution learning forests. In: *Nips*, pp 1–10
- Suo J, Wu T, Zhu S, Shan S, Chen X, Gao W (2008) Design sparse features for age estimation using hierarchical face model. In: 2008 8th IEEE International Conference on Automatic Face and Gesture Recognition, FG 2008
- Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 7–12 June. IEEE, pp 1–9
- Ueki K, Hayashida T, Kobayashi T (2006) Subspace-based age-group classification using facial images under various lighting conditions. In: FGR 2006: Proceedings of the 7th international conference on automatic face and gesture recognition, 2006 May, pp 43–48
- UTKFace (2017)
- Vetter T (1999) A morphable model for the synthesis of 3D faces f g. In: *Faces*, pp 187–194
- Zhang Y, Yeung DY (2010) Multi-task warped Gaussian process for personalized age estimation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, January 2014, pp 2622–2629
- Zhang K, Gao C, Guo L, Sun M, Yuan X, Han TX, Zhao Z, Li B (2017) Age group and gender estimation in the wild with deep RoR architecture. *IEEE Access* 5(X):22492–22503
- Zhang K, Liu N, Yuan X, Member S, Guo X, Gao C (2019) Fine-grained age estimation in the wild with attention LSTM networks. *IEEE Trans Circuits Syst Video Technol* PP(c):1
- Zhu Y, Li Y, Mu G, Guo G (2015) A study on apparent age estimation. In: Proceedings of the IEEE International Conference on Computer Vision, vol 2015-February, pp 267–273

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.