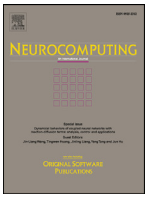
Neurocomputing 329 (2019) 300–310

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom



# On the use of DAG-CNN architecture for age estimation with multi-stage features fusion

Shahram Taheri, Önsen Toygar*

*Computer Engineering Departments, Faculty of Engineering, Eastern Mediterranean University, Famagusta, via Mersin 10, Turkey*

ARTICLE INFO

*Article history:*
Received 13 March 2018
Revised 23 October 2018
Accepted 31 October 2018
Available online 7 November 2018

Communicated by Dr. Ran He

ABSTRACT

Accurate facial age estimation is quite challenging, since ageing process is dependent on gender, ethnicity, lifestyle and many other factors, therefore actual age and apparent age can be quite different. In this paper, we propose a new architecture of deep neural networks namely Directed Acyclic Graph Convolutional Neural Networks (DAG-CNNs) for age estimation which exploits multi-stage features from different layers of a CNN. Two instants of this system are constructed by adding multi-scale output connections to the underlying backbone from two well-known deep learning architectures, namely VGG-16 and GoogLeNet. DAG-CNNs not only fuse the feature extraction and classification stages of the age estimation into a single automated learning procedure, but also utilized multi-scale features and perform score-level fusion of multiple classifiers automatically. Fine-tuning such models helps to increase the performance and we show that even "off-the-shelf" multi-scale features perform quite well. Experiments on the publicly available Morph-II and FG-NET databases prove the effectiveness of our novel method.

## 1. Introduction

Human age estimation from facial images has been historically a significant challenge in computer vision and image understanding society. Age estimation has wide potential practical applications like age-oriented commercial advertisement, audience-measurement systems, security surveillance, Age Specific Human-Computer Interaction (ASHCI) and soft biometrics. Besides, age estimation can be useful in other face-related tasks such as face detection and recognition.

The difficulty in automatic age estimation is mainly due to the speciality of ageing effects on the face compared with other facial variations. Uncontrollable nature of ageing process, personalized ageing patterns, large inter-class similarity and intra-class variation of subjects' images within the age classes, and finally, lacks of a comprehensive and representative age annotated facial images dataset for training a precise model are some of the difficulties in facial age estimation.

The inherent difficulties in age estimation from facial image, have derived research into constructing especially complex feature descriptor approaches in which most of them are either user defined multi-level and orientation bank of filters which were tried to mimic the behaviour of animal visual cortex (primary) network, or fine-grained facial regions to perform accurate alignment by using multiple facial fiducial points. In both cases, the generated feature extractor method is difficult to reuse, and in many cases it has high-dimension which takes considerable time to be extracted.

In the recent literature, deep learning and especially CNN has received much attention in the research field of machine learning and computer vision due to its superior performance in learning a series of nonlinear feature mapping functions directly from raw pixels. Deep learning have been used in many image understanding, object recognition and computer vision applications [1–3] and outperformed state-of-the-art methods. CNN has learnable parameters and there is no gap between the feature extraction step and the classification / regression step and all steps are optimized together to minimize the system's error.

Information fusion is the merging of information from heterogeneous sources with differing conceptual, contextual and typographical representations. Four different levels of fusions are sensor-level fusion, feature-level fusion, score-level fusion and decision-level fusion. In a biometric recognition system, the feature-level fusion is concatenating features, while in score- level fusion the similarity values are combined, which are obtained by different matchers. Apart from the raw data and feature vectors, the match scores contain the richest information about the input pattern. Additionally, it is relatively easy to access and combine the scores generated by different biometric matchers. Score-level fusion shows very good performance in multimodal biometric systems [4–6].

* Corresponding author.
*E-mail address:* onsen.toygar@emu.edu.tr (Ö. Toygar).

https://doi.org/10.1016/j.neucom.2018.10.071
0925-2312/© 2018 Elsevier B.V. All rights reserved.

In our previous work [6], we proposed a new age estimation system which exploits multi-stage features from a shallow CNN and precisely combined these features with a selection of age-related handcrafted features. This method utilized a decision-level fusion of estimated ages by two different approaches; the first one uses feature-level fusion of different handcrafted local feature descriptors for wrinkle, skin and facial component while the second one uses score-level fusion of different feature layers of a shallow CNN for its age estimation. However in this study, we propose a novel architecture of deep neural networks, namely Directed Acyclic Graph Convolutional Neural Networks (DAG-CNNs) for age estimation which exploits multi-stage features from different layers of a CNN. DAG-CNNs not only combine the feature extraction and classification stages of the age estimation into a single automated learning procedure, but also utilize multi-scale features and perform score-level fusion of multiple classifiers automatically. Therefore, DAG-CNN negates the necessity to extract hand-crafted features. In most of the current approaches, only the high level features which are extracted by the last layer of CNN are used. Instead of performing feature-level fusion manually and feeding the results into a classifier, the proposed multi-scale system can automatically learn different level of features, combine them and estimate the subject's age. Consequently based on the demonstrated success of CNNs for image processing, we propose a novel CNN architecture, namely, DAG-CNN for the human facial age estimation with the following contributions:

- The feature extraction and regression stages of the age estimation are fused into a single automated learning procedure.
- Instead of using the last layer's feature, it utilizes multistage learned features which are extracted from different intermediate layers of the CNN.
- Instead of manually combined different layers' features by feature-level fusion approach, it performs score-level fusion of multiple classifiers automatically.

The rest of this paper is ordered as follows: previous studies in facial age estimation have been reviewed in Section 2. In Section 3, the selected feature extractors to be used in feature-level and score-level fusion are reviewed. Section 4 describes the overall structure of the proposed method. Section 5 demonstrates the evaluation results of experiments performed by the fusion of local and global feature extractors. Finally, Section 6 provides the final conclusion and future works.

## 2. Related works

The earliest work about age estimation was published in 1994 by Kwon and da Vitoria Lobo [7], in which the age was just classified into several ranges. The first public domain database that has been used for age estimation is FG-NET [8] which contains 1002 images of 82 individuals and the age of subjects range from 0–69 years. Many researchers used FG-NET database for age estimation and classification [9]. However, a lot of attention on this topic has been drawn from the year that the large database, namely Morph-II [10] is published. The difficulty in data collection is now partly alleviated thanks to the availability of these ageing datasets.

The early works on age estimation can be classified based on their two fundamental modules of their pipeline: feature extraction methods and classification approaches. The visual feature descriptors used in these works are divided into three categories: local features, global features and hybrid features. Some researches rely on appearance models and flexible shape such as Active Appearance Model (AAM) and Active Shape Model (ASM) techniques. The AAM is a well-known method that represents faces with statistical appearance and shape models using PCA [11,12]. On the other hand, Bio-Inspired Features (BIF) scheme has been wildly used for age estimation [13]. Local neighbourhood features are also common descriptor choices used in this field [14–17].

Recently, many researchers have been using CNN for facial age estimation problems. Based on the number of architecture layers, these deep learning approaches are divided into two classes: shallow architecture and deep architecture. Modern CNN architectures such as VGGNet [18] and GoogLeNet [19] are two examples of deep architecture. The deep architecture suffers from over-fitting problem when there is a small number of training data like Morph-II dataset. Recently, the researchers used additional datasets with thousands of annotated images and transfer learning approach to overcome this problem and achieved the-state-of-the-art results [20–23]. In the work of Yi et al. [24], they used several shallow multiscale CNNs on different face regions and obtained the MAE of 3.63 on Morph-II dataset. On the other hand, in [25], the authors used CNNs and proposed using a ranking encoding for age and gender and they reported the MAE of 3.5 on Morph-II dataset. Hu et al. [20] proposed a novel learning scheme to embed the age difference information. Rothe et al. [23] proposed a deep learning solution for age estimation and introduced the IMDB-WIKI dataset which is the largest public age dataset. The authors reported the MAE of 2.68 on Morph-II dataset, which is the best reported result to the best of our knowledge. In [26] they proposed a conditional multitask learning method that architecturally factorizes an age variable into gender-conditioned age probabilities in a deep neural network. In order to overcome the lack of accurate training labels with discrete age values problem, they proposed a label expansion method that increases the number of accurate labels from weakly supervised categorical labels. Liu et al. [27] proposed an ordinal deep feature learning (ODFL) method to learn feature descriptors for face representation directly from raw pixels. They designed an end-to-end ordinal deep learning framework, where the complementary information of both feature extraction and age estimation is exploited to reinforce their model.

One of the age relevant topics is age progression which is defined as aesthetically re-rendering the ageing face at any future age for an individual face. In [28] the authors proposed novel bi-level dictionary learning based personalized age progression method. For each age group, they learned an ageing dictionary to reveal its ageing characteristics (e.g., wrinkles), based on face pairs from neighbouring age groups. Shu et al. [29] used a set of age-group specific dictionaries and a linear combination of these patterns to express a particular personalized ageing process. In [30] the authors presented a novel generative probabilistic model with a tractable density function for age progression. Their model inherits the strengths of both probabilistic graphical model and recent advances of ResNet.

In the few past years, some researchers tried to utilize the multi-scale features learned by different layers of a CNN for different problems. Tang et al. [31] proposed GoogLeNet based multi-stage feature fusion (G-MS2F) for scene recognition. The GoogLeNet model is employed and divided into three parts and the output features from each of the three parts are applied for final decision. In [32] the authors presented an object detection framework based on multi-stage convolutional features for pedestrian detection. Their framework extended the Fast R-CNN framework for the combination of several convolutional features from different stages of the used CNN to improve the network's detection accuracy.

## 3. Background

The fundamental components of DAG-CNN are given in the following subsections. We shortly introduce VGG-16 and GoogLeNet

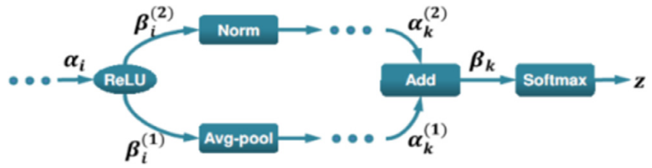


**Fig. 1.** Parameter setup at i-th ReLU [33].

architectures and explain the method of expected age estimation and score-level fusion for age estimation.

### 3.1. Directed Acyclic Graph-Convolutional Neural Network (DAG-CNN)

Deep Learning is a subfield of machine learning concerned with algorithms inspired by the structure and function of the brain called artificial neural network. Recently, deep artificial neural networks (including recurrent ones) have outperformed numerous state-of-the-art methods in pattern recognition and machine learning. The Directed Acyclic Graph (DAG) networks can represent more complex network architectures compared to simple ones which consist of a linear chain of layers. DAG architecture for neural networks (NNs) has emerged from the idea of recurrent NNs that have some feedback connections from forward layers to backward ones, which give them the ability of capturing dynamic states. The main advantage of DAG-structured networks is that their forward layers can have multiple input parameters from several backward layers. In this way, they can achieve different levels of image representations. A fundamental feature of the deep learning neural networks is the use of connections between their layers, called “skip connection”, that is similar to DAG-CNNs main idea, and it is shown that these skip connections can improve the accuracy of the classification tasks significantly.

DAG-CNN was proposed by Yang and Ramanan [33] to learn a set of multi-scale image features that are successfully used for classification of three standard scene benchmarks. They showed that the multi-scale model can be implemented as a DAG-structured feed forward CNN. By this approach, it is possible to use an end–to-end gradient-based learning for automatically extracting multi-scale features using generalized back propagation algorithm over the layers that have more than one input. In fact, all the required equations for training the network are standard CNN equations except for the Add and ReLU layers since they have multiple inputs or outputs. Considering the $i$th ReLU layer in Fig. 1, let $\alpha_i$ be its input, $\beta_i^{(j)}$ be the output for its $j$th output branch (its $j$th child in the DAG), and assume that $z$ is the final output of the softmax layer. The gradient of $z$ with respect to the input of the $i$th ReLU layer can be computed as in Eq. (1):

$$\frac{\partial_z}{\partial_{\alpha_i}} = \sum_{j=1}^{C} \frac{\partial_z}{\alpha \beta_i^{(j)}} \frac{\alpha \beta_i^{(j)}}{\partial \alpha_i} \tag{1}$$

where $C$ is the number of output edge of the $i$th ReLU.

For the Add layer, let $\beta_k = g(\alpha_k^{(1)}, \ldots, \alpha_k^{(N)})$ represents the output of an Add layer with multiple inputs. The gradient along the layer can be computed by applying the chain rule as in Eq. (2):

$$\frac{\partial_z}{\partial \alpha_i} = \frac{\partial_z}{\partial \beta_k} \frac{\partial \beta_k}{\partial \alpha_i} = \frac{\partial_z}{\partial \beta_k} \sum_{j=1}^{C} \frac{\partial \beta_k}{\partial \alpha_k^{(j)}} \frac{\partial \alpha_k^{(j)}}{\partial \alpha_i} \tag{2}$$

In the convolutional layers, the convolution operation is computed by Eq. (3) as follows:

$$X_n = \sum_{k=0}^{N-1} y_k f_{n-k} \tag{3}$$

where $y$ and $f$ are the input image and applied filter, respectively and $N$ is the number of elements in the input image. The convolution layer output is represented by vector $X$. For all layers of the DAG-CNN architecture except ReLU and Add layers, the equations Eqs. (4)–(5) are used to update biases and weights as follows:

$$\Delta W_t(t+1) = -\frac{x_\lambda}{r} W_l - \frac{x}{n} \frac{\partial C}{\partial W_l} + m \Delta W_l(t) \tag{4}$$

$$\Delta B_l(t+1) = -\frac{x}{n} \frac{\partial C}{\partial B_l} + m \Delta B_l(t) \tag{5}$$

where $W$, $B$, $l$, $\lambda$, $x$, $n$, $m$, $t$, and $C$ denote the weight, bias, layer number, regularization parameter, learning rate, total number of training samples, momentum, updating step, and cost function, respectively.

In DAG-CNNs, since lower layers are directly connected to the output layer through multi-scale connections, it is guaranteed that these layers' neurons receive a strong gradient signal during learning and do not suffer from the problem of vanishing gradients. In CNNs, the size of the learned features in intermediate layers can be very large and combining these features may cause the curse of dimensionality problem. In order to overcome this problem, marginal activations by performing average pooling on the learned features of some layers which are used for score-level fusion.

In this paper, two different DAG-CNN architectures are proposed to improve the discrimination capability of a deep neural network by allowing its layers to share their learned features and work collaboratively for classification. The proposed multi-scale CNN topologies employ learned features with different level of complexity in order to estimate the subject's age with high precision.

### 3.2. VGG-16 architecture

VGG-16 architecture had been proposed by the Oxford Visual Geometry Groups' model in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [18]. VGG-16 is deeper and wider than former CNN structure and it has five batches of convolution operations, each batch consisting of 2–3 adjacent convolution layers. Adjacent convolution batches are connected via max-pooling layers. The size of kernels in all convolutional layers is 3 × 3 convolutional layers and the number of kernels within each batch is the same (increases from 64 in the first group to 512 in the last one). Fig. 2 illustrates the VGG-16 architecture. This architecture has been used in many researches and it was the first one that outperformed human-level performance on ImageNet.

### 3.3. GoogLeNet architecture

The GoogLeNet [19] model is the winner of ILSVRC in 2014. A new module named Inception is introduced in [19] which apply various sizes of convolutional kernels to be composed to form more discriminative feature representations. The depth of GoogLeNet reaches to 22 and the number of convolutional layers reaches to 60, so that the errors always vanish with back propagation, and the parameters of low layers may not be optimized sufficiently. To address this issue, two auxiliary classifiers are employed by GoogLeNet to optimize the parameters of low layers. In GoogLeNet, there are 9 Inception modules employed to construct the architecture. The Inception module consists of a few of convolutional kernels with small sizes (such as 1 × 1, 3 × 3 and 5 × 5), which are conducive to limit the scale of parameters and model complexity. To learn efficiently, GoogLeNet introduced 1 × 1 convolutions for feature dimension reduction. In order to overcome the problems of gradient vanishing and over-fitting, these 9 Inception modules are divided into 3 groups, and three objective functions are added on every 3 Inception modules (Fig. 3).

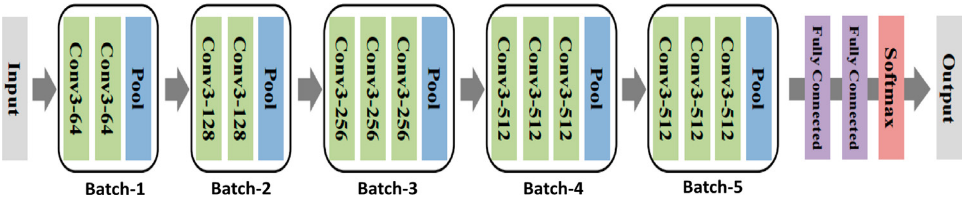


**Fig. 2.** VGG-16 Architecture.

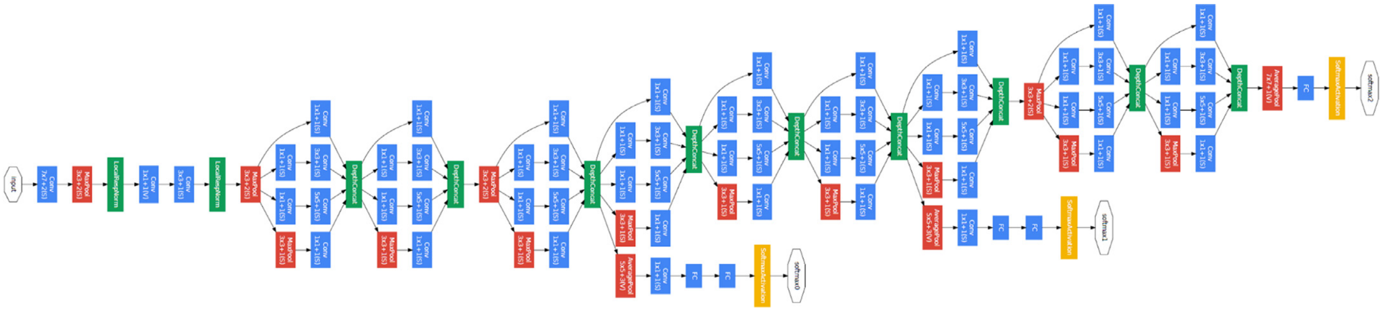


**Fig. 3.** GoogLeNet Architecture (check the online-version for more details).

### 3.4. Expected age value

Age estimation can be considered as a discrete classification with multiple discrete value labels. For Morph-II dataset, it is a one dimensional regression problem with the age being sampled from a continuous range between 16 and 77 and for FG-NET dataset it is between 0 and 70. For computing the expected age from CNN, instead of using simple softmax, we multiply each softmax output probabilities by the corresponding class year label and add them together as follows:

$$Expected\ Age = \sum_{i=L}^{H} p_i \times year\_label_i \tag{6}$$

where $L, H$ are lower and upper bound of subject's age, $p_i$ is softmax output probability and $year\_label_i$ is an integer age year value of output neuron i in the last layer of CNN.

### 3.5. Score-level fusion of multi-stage CNN learned features

CNNs can be used as automatic feature extractors and the cost-free mid-level features extracted from intermediate layers of a CNN can be discriminative for classifying different patterns with varying complexities. There are two ways in order to use these features: feature-level fusion and score-level fusion. In the feature-level fusion approach, different layer features are concatenated to create final feature vector, then it is fed into a classifier or regression. One of the common problems with feature-level fusion is the size of the feature vectors. In CNNs, the size of the learned features in intermediate layers can be very large and combining these features may cause the curse of dimensionality problem. Dimensionality reduction methods such as Principal Component Analysis (PCA) or Discrete Cosine Transform (DCT) can be used to overcome this problem [37]. Another approach for overcoming the dimensionality problem is score-level fusion. In this method, features of each layer are given to a separate classifier to generate a score vector for the test sample, then these scores are combined together to generate the final decision.

The whole process of score-level fusion is given in Algorithm 1. This algorithm can be expanded for more than two feature sets. In score-level fusion, the distance between each test sample and all the training samples is computed and assumed to be the score of that test sample in the corresponding classification/regression system. These scores are normalized by Min-Max normalization [34] method as follows:

**Algorithm 1** Score-level fusion of two feature sets ($FS^m, FS^n$) for age estimation.

```
Input:
Trainset $X_{train} = \{(X^i_{tr}, y^i_{tr})\}, i = 1, \dots, N_{train}$
Testset $X_{test} = \{(X^i_{te}, y^i_{te})\}, i = 1, \dots, N_{test}$
Output:
        Estimated_age_vector, MAE
$F_m{}^{X_{tr}{}^i} \leftarrow$ Compute $FS^m$ $\forall$ i = 1, ..., $N_{train}$
$F_n{}^{X_{tr}{}^i} \leftarrow$ Compute $FS^n$ $\forall$ i = 1, ..., $N_{train}$
For $j = 1$ To $N_{test}$
    $F_m{}^{X_{te}{}^j} \leftarrow$ Compute $FS^m$
    $F_n{}^{X_{te}{}^j} \leftarrow$ Compute $FS^n$
    For $i = 1$ To $N_{train}$
        $score^m_i = compute\_distance\ (F_m{}^{X_{tr}{}^i}, F_m{}^{X_{te}{}^j})$
        $score^n_i = compute\_distance\ (F_n{}^{X_{tr}{}^i}, F_n{}^{X_{te}{}^j})$
    For $i = 1$ To $N_{train}$
        normalized $score^m_i$ according to (7)
        normalized $score^n_i$ according to (7)
        $fusion_i = score^m_i + score^n_i$
    $minIndex = Find\_Min\_index(\boldsymbol{fusion})$
    Estimated_age_vector$_i = y^{minindex}_{tr}$
    $Error += abs$ (Estimated_age_vector$_i$ - $y^j_{te}$)
MAE=$Error/\ N_{test}$
```

$$x' = \frac{x - \mathrm{Min}(x)}{\mathrm{Max}(x) - \mathrm{Min}(x)} \tag{7}$$

where $x$ is the raw score, Max($x$) and Min($x$) are the maximum and minimum values of the raw scores respectively and $x'$ is the normalized score.

After normalization, the score vectors are combined by Sum rule-based fusion method [34] to generate a single scalar score which is then used to make the final decision.

## 4. Proposed method

In this paper, we propose two new DAG-CNN architecture for estimating the accurate age from facial image by exploiting multi-stage learned features from different layers of a VGG-16 CNN and GoogLeNet models. Convolutional neural networks can be used as automatic feature extractors and the learned features can be fed to classifiers like SVMs or NNs to predict the output labels. Mid-level features at intermediate layers of the CNN can be discriminative for classifying different patterns with varying complexities. However, in CNN architectures used in literature so far, these cross-layer heterogeneity features are ignored. It is clear to see that these mid-level features are already computed when the system is trained to extract high-level features, and hence, their usage does not bring any extra computational burden within our proposed model.

According to the success rate of using score-level fusion in face and multimodal biometric recognition systems [4–6], it is believed that accuracy can be improved when the information of different types of feature descriptors and classifiers are consolidated. The reason is that different layers of CNN learn different level of information which varies from local and detailed information to more abstract one. In order to test this hypothesis, we construct a DAG-CNN network which automatically performs score-level fusion of different selected layers. Therefore, instead of manually performing feature-level or score-level fusion and feeding the results to a classifier, we propose a multi-scale system by using a CNN with Directed Acyclic Graph (DAG) topology. Our proposed model can automatically learn different level of features, combine them by score-level fusion method and estimate the final age.

In order to investigate the suitability of DAG-CNN, we propose two different models based on two well-known CNN architecture, namely VGG-16 and GoogLeNet and named these proposed systems as DAG-VGG16 and DAG-GoogLeNet, respectively. We use these architectures as the backbone of our proposed DAG topology and employed some branches from their intermediate and last layers. These links are connected to an average pooling layer to reduce their dimensionality, then are normalized and given to a separate fully connected MLP layers. Each of these fully connected layers have the same number of neurons in their last layer and that is equal to number of age classes(54 and 70 different age classes for Morph-II and FG-NET datasets, respectively) and generate a score vector for each input image. These score-vectors are added with each other, element by element and the result vector is normalized such that its components' summation becomes 1. Then the normalized vector is fed into the final decision layer in which the subject's age is estimated by Eq. (6).

Both VGG-16 and GoogLeNet contain many layers and due to the possible redundancy among these layers' features, it is improper to fuse all of these features. VGG-16 has five batches of convolution operations with 2 or 3 adjacent convolution layers. Adjacent convolution batches are connected via max-pooling layers and these locations are suitable candidate points for multi-stage score-level fusion. Therefore, by considering the trade off between model accuracy and complexity, the VGG-16 is partitioned into five parts and the optimal selection of these parts' features should be considered for the final age estimation. Finally, for DAG-VGG16, we select Batch 2, 3 and 5 as the DAG branch positions by using a greedy algorithm which is explained in Section 5.5. For GoogLeNet based system, in order to select layers for combining in DAG-CNN model in an effective way, with the aid of the auxiliary classifiers defined by GoogLeNet, we select the position of the auxiliary classifiers for getting branches and perform score-level fusion of these three stages.

In this study, we have shown that combining different level features can improve age estimation accuracy significantly. Particularly, the accuracy is improved when we add features learned by intermediate layers, with the exception of the low-level features of early layers that cause a decrease in estimation accuracy. For the purpose of testing different combinations of feature layers and finding the best one experimentally, features of the last layer are considered as of necessary and intermediate layer features are added layer-by-layer, one at a time, in a backward fashion until no improvement observed in classification accuracy. This greedy approach ignores the features of layers closer to the input layer. Experimental evaluations as illustrated within the next section exhibited that the proposed system's capability of fusion of multi-scale features improves the accuracy of age estimation. The overall schematic of the proposed methods are illustrated in Fig. 4(a) and (b).

**Table 1**
The age and gender information of samples from Morph-II.

| | <20 | 20–29 | 30–39 | 40–49 | >50 | Total |
|---|---|---|---|---|---|---|
| **Male** | 6638 | 14,016 | 12,448 | 10,062 | 3482 | 46,646 |
| **Female** | 831 | 2309 | 2909 | 1988 | 453 | 8490 |
| **Total** | 7469 | 16,325 | 15,357 | 12,050 | 3935 | 55,136 |

## 5. Experiments

Many experiments are conducted to evaluate the performance of the proposed method over the Morph-II and FG-NET ageing datasets. The datasets, metrics, parameters and experimental setup details are given in the following subsections.

### *5.1. Morph-II ageing database*

Morph-II [10] is an ageing database which contains more than 55,000 face images of about 13,000 subjects. These images are captured during 2003–2007. Age ranges in this database vary from 16 to 77 years. The age and gender information of the samples that are used in the experiments are shown in Table 1.

In order to use the database in a systematic way, we follow the way described in [35] to split the database into five non-overlapped subsets randomly with a very important criterion: all of the sample images from a specific subject should be in one and only one unique fold each time.

### *5.2. FG-NET ageing database*

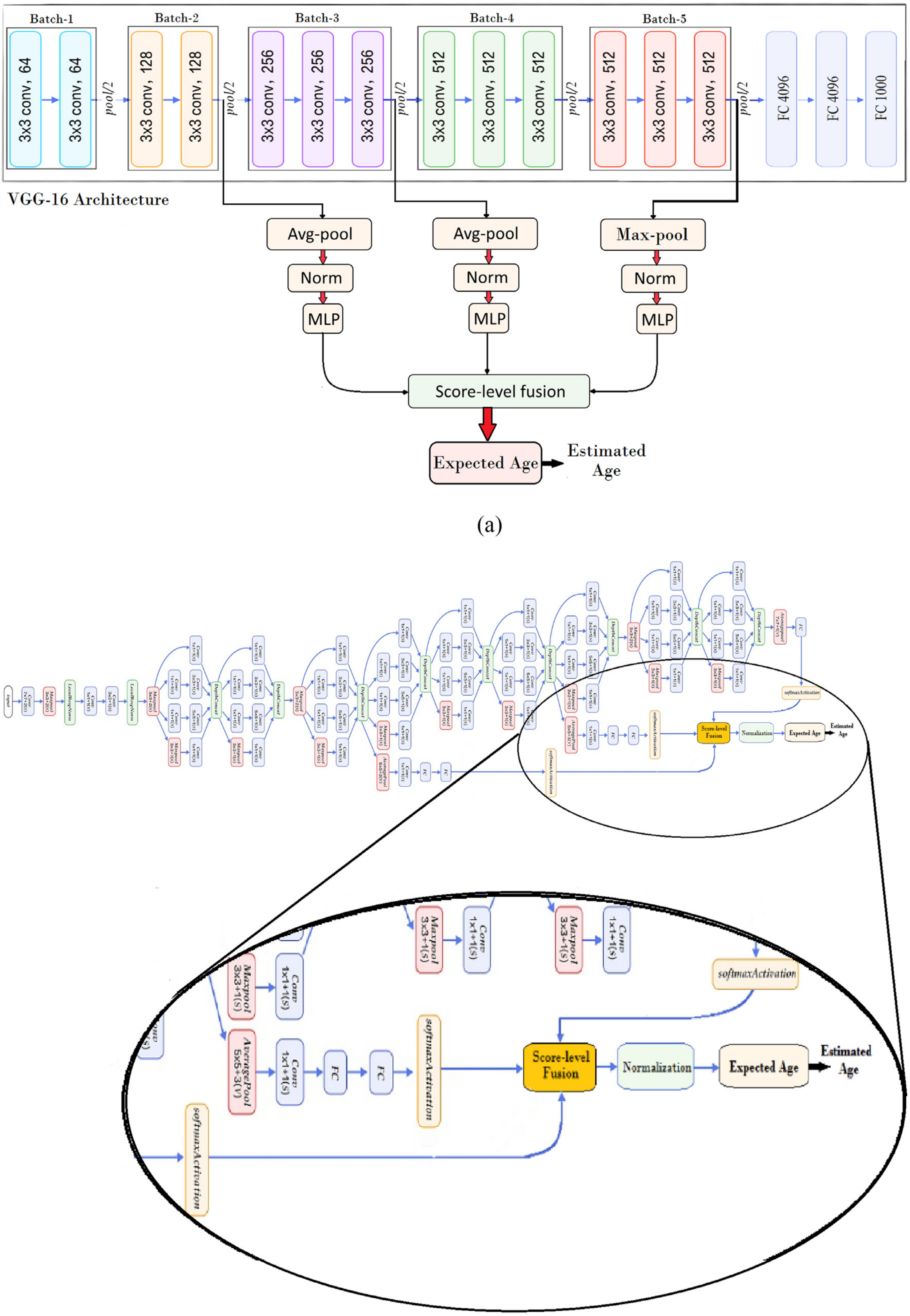
In 2004, the FG-NET ageing database (The Face and Gesture recognition NETwork) was released in order to help researchers who try to understand the effect of ageing on facial appearance. After that, FG-NET was used in many studies in different domains such as in age estimation, age-invariant face recognition, gender classification and age progression.

The FG-NET consists of 1002 images from 82 different people with ages varying between 0 and 69 years old. The subjects' ages are not equally distributed and most of the subjects' ages are less than 40 years old in the database. The age distribution is shown in Fig. 5(a) and the ageing faces example of one subject is shown in Fig. 5(b). These images were collected by scanning personal photographs of subjects so they display considerable variability in resolution, image sharpness, and illumination in combination with face viewpoint and expression variation. Occlusions in the form of spectacles, facial hair and hats also exist in a number of images.

Most of the studies which evaluate systems using FG-NET employ leave one person out (LOPO) method. This approach is optimum for the databases including small number of images. In the experiments, for each of 82 people in the dataset, a separate age estimator model is trained by using 81 remaining images and finally the average results of these models are reported.

### *5.3. Metrics*

The most common metric for accuracy evaluation of the age estimators is the Mean Average Error (MAE) which is also used in



(a)

(b)

**Fig. 4.** Overview of proposed methods: (a) DAG-VGG16, (b) DAG-GoogLeNet.

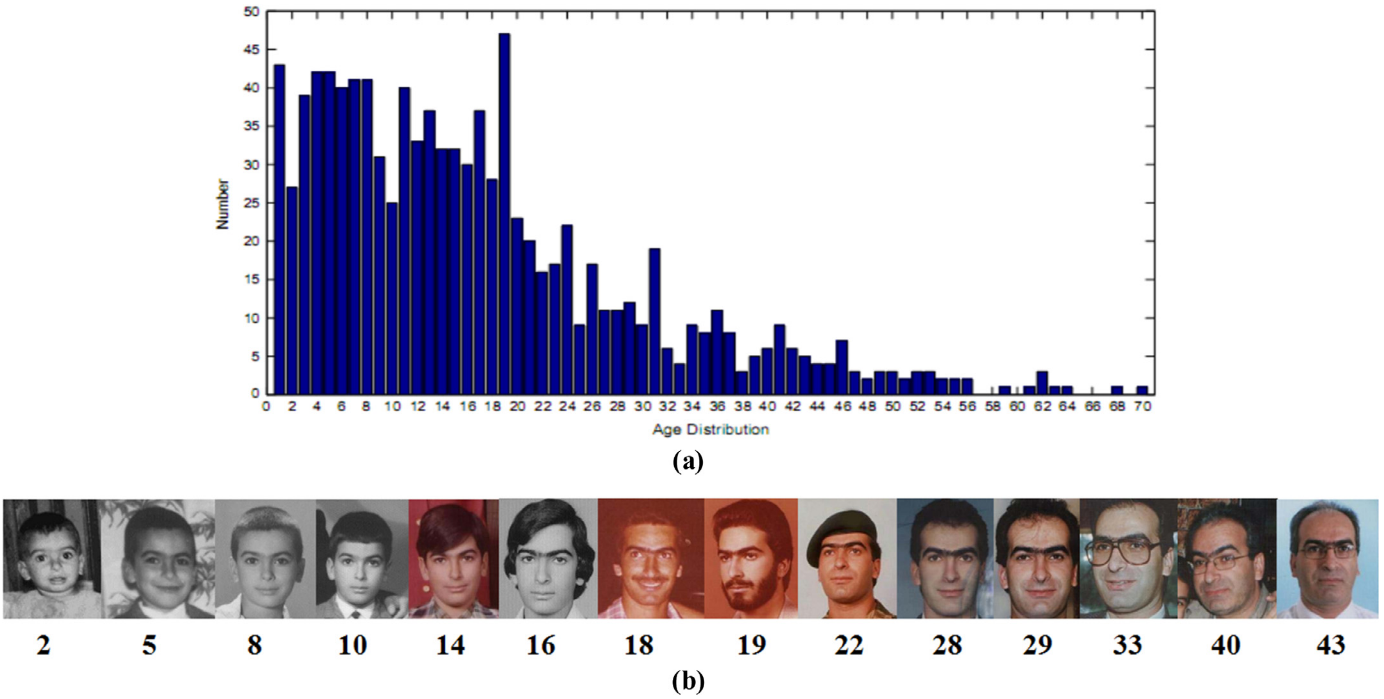


**Fig. 5.** (a) Age distributions of FG-NET dataset. (b) Example of ageing faces of one subject.

this study. MAE computes the average age deviation error in absolute terms as follows:

$$\mathrm{MAE} = \sum_{\mathrm{i}=1}^{N} \frac{|\hat{\alpha}_{\mathrm{i}} - \alpha_{\mathrm{i}}|}{N} \tag{8}$$

where $\hat{\alpha}_{\mathrm{i}}$ is the computed age of the i subject, $\alpha_{\mathrm{i}}$ is its actual annotated age and $N$ is the total number of test subjects.

Another common metric is the cumulative score (CS) which quantitatively shows the evaluation of age estimation approach by a curve. The CS accuracy at the error ε is computed as follows:

$$CS(\theta) = \frac{N_{\varepsilon \le \theta}}{N} \times 100\% \tag{9}$$

where $N_{\varepsilon \le \theta}$ is the number of test samples in which their estimated age error $\theta$ is not less than ε.

### 5.4. Preprocessing

Preprocessing is an essential step in image processing systems in order to enhance the quality of the input images. In this study, we used face detection method described in [36] for the detection of facial images. Then the facial image will be aligned by using geometric transformation such that the eyes have been symmetrically placed at 25% and 75% of the aligned image. All the input images are in RGB format and resized to 256 × 256. Five different cropped size of 227 × 227 and their flip are fed to the DAG-VGG16 network while the size of cropped input images for DAG-GoogLeNet architecture is 224 × 224. The same data augmentation methods are deployed for the offline multi-stage feature fusion systems.

### 5.5. Offline multi-stage feature fusion

In order to investigate the effectiveness of our multi-stage features fusion approach, we manually combined the features from different layers of trained and fine-tuned VGG-16 and GoogLeNet models. In order to implement score-level fusion manually for VGG-16 based system, a pre-trained model on IMDB-WIKI dataset is selected from [23] and fine-tuned on FG-NET and Morph-II datasets. For GoogLeNet based system, we firstly pre-train the model using CASIA-WebFace database. Afterwards, it is fine-tuned on IMDB-WIKI and finally fine-tuned on one of the target datasets, FG-NET and Morph-II. The numbers of neurons in the last layer of the models are changed to output 54 or 70 age labels for Morph-II or FG-NET dataset, respectively. Additionally, expected age obtained by Eq. (6) is used as the cost functions. In fine-tuning phase for VGG-16 based system, the weights of early layers are frozen and only the Batch-3 to Batch-5 and fully connected layers' weights are updated. These two configurations are considered as the baselines for comparison with offline and online multi-stage score-level fusion approaches. For VGG-16 based system, the baseline MAE is 2.91 for Morph-II and 3.36 for FG-NET databases and for GoogLeNet based system, the baseline MAE are 3.13 and 3.29, for Morph-II and FG-NET databases, respectively.

Afterwards, different layers' features are extracted and are fused together by Algorithm-1 for score-level fusion. In this computation, different layers' features are extracted for all training samples separately. In the test phase, for each test sample and for each separate layer's features, the corresponding feature vector is computed and its distance is compared with all of the feature vectors in the training set and these distances are stored in an array, namely score vector. After computing the score vector for all of the feature sets, the scores are normalized and combined together by adding them element by element. The training sample's age with minimum distance is considered as the estimated age.

For VGG-16 based system, in order to test different combinations of feature layers and finding the optimal configuration experimentally, the features of the last batch layer (Batch-5) are considered as of necessary and features from different intermediate batch layers are added, one at a time, in a backward fashion until no improvement is observed in age estimation accuracy. When a new batch layer's feature is added, if the accuracy is decreased, we ignore the new batch layer's feature and backtracking and select another batch from the remaining ones. This greedy approach ignores the features of batch layer close to the input layer (Batch-1) and Batch-4 and as a result, the optimal subset of

**Table 2**
The experimental results summary.

| Method | MAE (years) | |
|---|---|---|
| | Morph-II | FG-NET |
| VGG-16 Baseline | 2.91 | 3.36 |
| Offline multi-stage features fusion | 2.86 | 3.22 |
| Proposed DAG-VGG16 | 2.81 | 3.08 |
| GoogLeNet Baseline | 3.13 | 3.29 |
| Offline multi-stage features fusion | 2.99 | 3.17 |
| Proposed DAG-GoogLeNet | 2.87 | 3.05 |

**Table 3**
Comparison with the state-of-the-art methods on Morph-II dataset.

| Reference | Method/feature | MAE |
|---|---|---|
| Geng et al. [40] | AAS/AAS + BIF | 10.10 |
| Chang et al. [15] | OHRank/AAM | 6.07 |
| Guo and Mu [39] | kPLS/BIF | 4.04 |
| Geng et al. [40] | CPNN/AAM + BIF | 4.87 |
| Guo and Mu [41] | kCCA/BIF | 3.98 |
| Geng et al. [40] | MFOR/PCA + LBP + BIF | 4.20 |
| Han et al. [13] | SVM + SVR/BIF + ASM | 4.20 |
| Fernández [42] | SVR.HOG | 4.83 |
| Huerta et al. [43] | CNN/CNN | 3.88 |
| Yang et al. [25] | Deeprank/deep network | 3.57 |
| Han et al. [44] | DIF/demographic | 3.80 |
| Huerta et al. [45] | Fusion | 4.25 |
| Niu et al. [21] | OR-CNN/CNN | 3.27 |
| Rothe et al. [23] | DEX(IMDB-WIKI)/CNN | 2.68 |
| Wang et al. [46] | DLA/CNN | 4.77 |
| Yi et al. [24] | CNN | 3.63 |
| Duan et al. [47] | CNN + ELM | 3.44 |
| Hu et al. [20] | CNN | 2.78 |
| Ng et al. [48] | CNN | 3.88 |
| Antipov et al. [49] | CNN | 2.99 |
| **Proposed methods** | **DAG-VGG16** | **2.81** |
| | **DAG-GoogLeNet** | **2.87** |

features for score-level fusion is features extracted from batch layers 2, 3, and 5. For GoogLeNet based system, we select the position of the auxiliary classifiers for performing score-level fusion.

The experimental results show that combining the intermediate features with last layer features with score-level fusion causes meaningful improvement in MAE. The MAE of manually multi-stage score-level fusion for VGG-16 based system on Morph-II and FG-NET datasets are improved to 2.86 and 3.22 years old, respectively. The results of this approach shows improvement for GoogLeNet based system too. As shown in Table 2, the MAE of GoogLeNet based system on Morph-II and FG-NET datasets are improved to 2.99 and 3.17 years old, respectively.

### 5.6. DAG-CNN architecture for age estimation

DAG-CNN consists of a normal CNN and some branches from its different layer to fuse multi-stage learned features. We selected two publicly available CNN, VGG-16 and GoogLeNet architectures, as the chain-structured or backbone of the DAG-CNN architectures due to their impressive result on the ILSVRC. For DAG-VGG16 system, we started with VGG-16 deep CNN models from [25] which is pre-trained on the IMDB-WIKI dataset and performed the following modifications: instead of rectified linear unit (ReLU), we utilized S-shaped ReLU [38]. Additionally, we use batch normalization between convolution layers to reduce the internal covariate shift. It helps the network to learn how to combine colour features in an optimal way. For DAG-GoogLeNet system, we used the aforementioned baseline system as the back-bone structure.

For the purpose of testing different combinations of feature layers and finding the best DAG-CNN architecture experimentally, for both DAG-VGG16 and DAG-GoogLeNet systems, we selected the DAG branch places according to the configuration result obtained in Section 5.5. Each branch is given to an average pooling layer with 5 × 5 kernel size to reduce its dimensionality and is normalized and finally, is fed into a multi-layer perceptron (MLP) classifier to compute its score. The scores from different MLP are fused together by addition rule [34] . This score-level fusion result is normalized to sum 1 and used to compute the estimated-age.

All of the MLP classifiers contain three layers in which the first and the second layers have 500 neurons and the last fully connected layer of the model to output 54 or 70 classes corresponding to different age labels in Morph-II and FG-NET datasets, respectively. The MLP neuron's weights are initialized to small normally-distributed numbers. For computing the estimated age, instead of using simple softmax, the estimated age is computed by using Eq. (6). Finally, the modified model is employed for fine tuning on the Morph-II dataset. In order to avoid overfitting, we used different learning rate policy for different layers. Therefore we used small learning rate for the feature extraction layers and froze the early layers' weights, but for the fully connected layers we utilized higher learning rate. Additionally, we performed data augmentation by cropping five different regions of 224 × 224 pixels (227 × 227 for DAG-GoogLeNet) from the 256 × 256 input image (four corners and the centre one) and their mirror version in the training phase. All the input images are in RGB format.

Our proposed DAG-CNN architecture was fine-tuned through the standard backpropagation technique with a batch size of 32. In order to obtain optimum performance, the other learning parameters are set as follows: to prevent overfitting of training data, the regularization ($\lambda$) is set to 0.1, momentum parameters which adjust the speed of learning during training is set to 0.9, and learning rate that control the convergence of the training data are set to 0.001 and linearly changed according to the mean-squared error values in each ten iteration. The training was performed for 50 epochs. The MAE of DAG-VGG16 is 2.81 years for Morph-II and 3.08 years for FG-NET, while the MAE of DAG-GoogLeNet system are 2.81 and 3.08 years for Morph-II and FG-NET datasets, respectively. All of the DAG related results are better than the results of offline multi-stage feature fusion method. The reason is that, in DAG-CNN and during the training phase, different features from different layers are combined and the model learned more discriminative features with respect to the simple CNN model in the baseline and its offline score-level fusion counterpart. All experiment results are summarized in Table 2. Additionally, in Figs. 6 and 7, we showed the MAE separately for each age in order to investigate the effectiveness of each aforementioned systems on Morph-II dataset. These results show that in both cases of DAG-CNN and offline score-level fusion, for all subjects' age, the integration of several layers' features caused the improvement in age estimation's accuracy. Furthermore, it is clear that, in most of the age values, the features learned by DAG-CNN have more discriminative power than features obtained by offline score-level fusion. All the experiments are performed on a machine with Intel Xeon E5-2683 - 2.0 GHz processor and 16 GB Ram. All the codes are written with Matlab 2017b platform.

### 5.7. Comparison with the state-of-the-art methods

We compare the proposed methods' accuracy with the state-of-the-art methods that presented the results on Morph-II and FG-NET datasets. The robustness and effectiveness of the proposed methods are studied in terms of MAEs in Tables 3 and 4. Additionally, The CS curves of the proposed methods compared with state-of-the-art methods on Morph-II and FG-NET datasets are illustrated in Figs. 8 and 9. The cumulative score diagrams of our

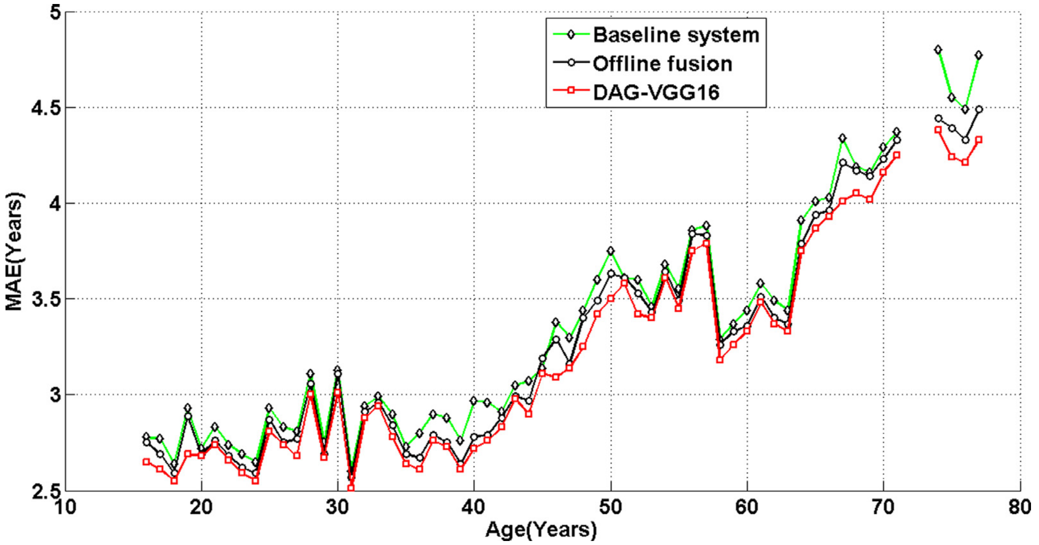


**Fig. 6.** MAE of DAG-VGG16 system for Morph-II dataset.

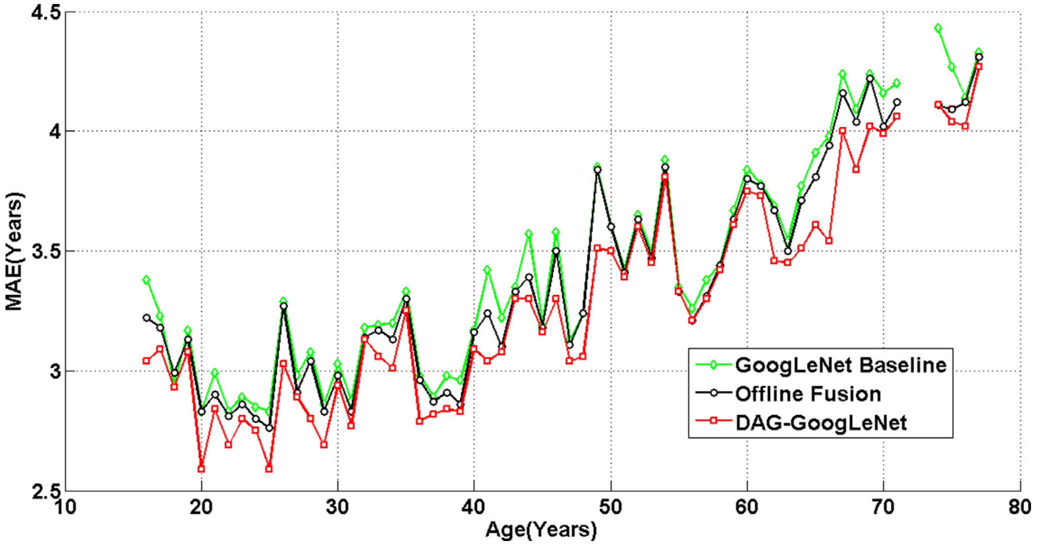


**Fig. 7.** MAE of DAG-GoogLeNet system for Morph-II dataset.

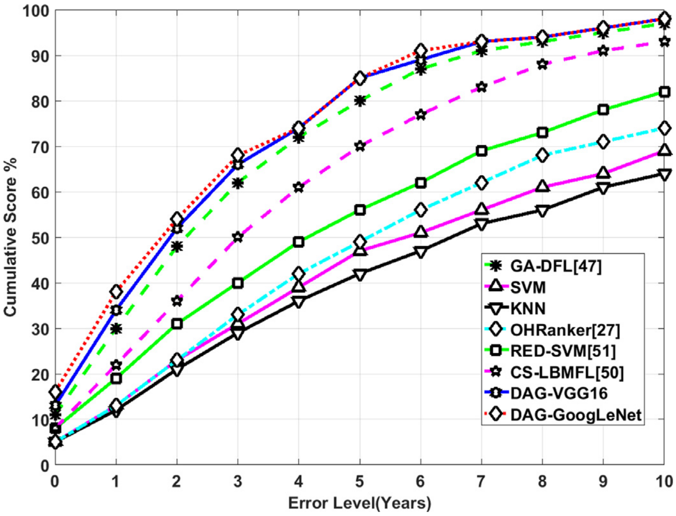


**Fig. 8.** The CS curves of the proposed method compared with state-of-the-art methods on Morph-II dataset.

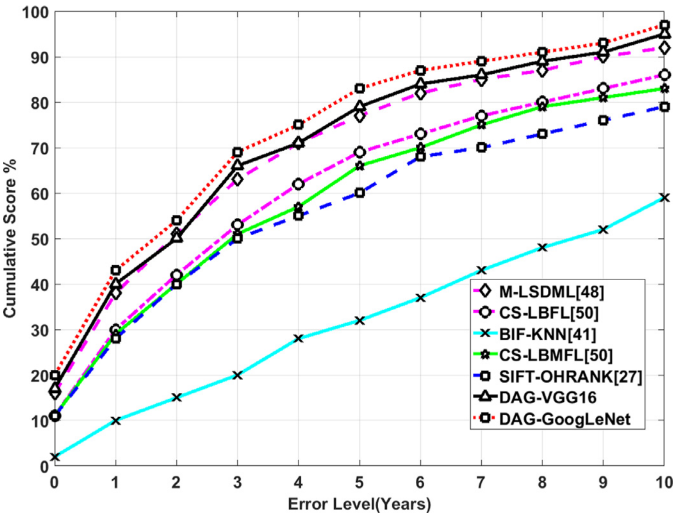


**Fig. 9.** The CS curves of the proposed method compared with state-of-the-art methods on FG-NET dataset.

**Table 4**
Comparison with the state-of-the-art methods on FG-NET dataset.

| Reference | Method/feature | MAE |
|---|---|---|
| El Dib et al. [50] | BIF | 3.17 |
| Han et al. [13] | Component and holistic BIF | 4.6 |
| Geng et al. [40] | Label distribution(CPNN) | 4.76 |
| Liang et al. [51] | Hierarchical framework | 4.97 |
| Lu et al. [52] | CS-LBFL | 4.43 |
| Lu et al. [52] | CS-LBMFL | 4.36 |
| Chang et al. [53] | CS-OHR | 4.70 |
| Chen et al. [54] | CA-SVR | 4.67 |
| Chen et al. [55] | Cascaded-CNN | 3.49 |
| Liu et al. [56] | M-LSDML | 3.31 |
| **Proposed methods** | **DAG-VGG16** | **3.08** |
| | **DAG-GoogLeNet** | **3.05** |

proposed methods outperform all of the other compared methods in all different levels of error. It can be seen that for MORPH-II dataset, both DAG-VGG16 and DAG-GoogLeNet systems' are within 1-year error for 40% of samples and within 5-year error for near 80% of them. These results for FG-NET dataset and in case of DAG-VGG16 system are 34% and 83%, and in case of DAG-GoogLeNet system are 38% and 83%, respectively. These figures illustrate that our method is better than all of the compared methods in any error level. Consequently, the experimental results, the MAE value and CS curve, show that our proposed method outperforms most of the state-of-the-art methods. It is clearly seen that our proposed method outperforms many other methods such as hand-crafted and CNN-based approaches. This improvement is caused by different factors of the proposed method such as advanced architecture, using additional dataset and transfer learning method, fusion of different layers' features and the expected age formula. The results demonstrate that combining different features by score-level method in both offline and DAG-CNN version enhances the performance of the age estimation system.

## 6. Conclusions

A novel CNN architecture for age estimation method based on multi-stage fusion of information is proposed in this paper. Multi-stage learned features from different layers of a CNN are automatically combined together by score-level fusion method by using DAG-CNN architecture. We showed that DAG-CNN can improve the discrimination capability of a deep neural network by allowing its layers to share their learned features and work collaboratively for classification. Compared with the state-of-the-art methods, our proposed approach obtained significant lower MAE on Morph-II and FG-NET datasets. The experimental results showed that the score-level fusion of global learned features provide a higher accuracy than the other algorithms. The proposed method will be applied on other ageing databases under different conditions as future work.

## References

[1] P.N. Druzhkov, V.D. Kustikova, A survey of deep learning methods and software tools for image classification and object detection, Pattern Recognit. Image Anal. 26 (1) (2016) 9–15.

[2] Z. Li, J. Tang, Weakly supervised deep matrix factorization for social image understanding, IEEE Trans. Image Process. 26 (1) (2017) 276–288.

[3] Z. Li, J. Tang, Weakly supervised deep metric learning for community-contributed image retrieval, IEEE Trans. Multimed. 17 (11) (2015) 1989–1999.

[4] M. Eskandari, Ö. Toygar, Fusion of face and iris biometrics using local and global feature extraction methods, Signal Image Video Process. 8 (6) (2014) 995–1006.

[5] H.M. Sim, H. Asmuni, R. Hassan, R.M. Othman, Multimodal biometrics: weighted score level fusion based on non-ideal iris and face images, Expert Syst. Appl. 41 (11) (2014) 5390–5404.

[6] S. Taheri, O. Toygar, Multi-stage age estimation using two level fusions of handcrafted and learned features on facial images, IET Biom. (2018) 12, doi:10.1049/iet-bmt.2018.5141. Available online: 01 October 2018.

[7] Y.H. Kwon, N. da Vitoria Lobo, Age classification from facial images, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR, 1994, pp. 762–767.

[8] A. Lanitis, C.J. Taylor, T.F. Cootes, Toward automatic simulation of aging effects on face images, IEEE Trans. Pattern Anal. Mach. Intell. 24 (4) (2002) 442–455.

[9] G. Panis, A. Lanitis, N. Tsapatsoulis, T.F. Cootes, Overview of research on facial ageing using the FG-NET ageing database, IET Biom. 5 (2) (2016) 37–46.

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

[11] T.C. Hsu, Y.S. Huang, F.H. Cheng, A novel ASM-based two-stage facial landmark detection method, in: Proceedings of the Pacific-Rim Conference on Multimedia, Springer, Berlin, Heidelberg, 2010, pp. 526–537.

[12] K. Luu, K. Ricanek, T.D. Bui, C.Y. Suen, Age estimation using active appearance models and support vector machine regression, in: Proceedings of the IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems, 2009. BTAS'09, IEEE, 2009, pp. 1–5.

[13] H. Han, C. Otto, A.K. Jain, Age estimation from face images: human vs. machine performance, in: Proceedings of the 2013 International Conference on Biometrics (ICB), IEEE, 2013, pp. 1–8.

[14] S.E. Bekhouche, A. Ouafi, A. Taleb-Ahmed, A. Hadid, A. Benlamoudi, Facial age estimation using BSIF and LBP, First International Conference on Electrical Engineering (ICEEB'14), Biskra, December 7-8, 2014.

[15] K.Y. Chang, C.S. Chen, Y.P. Hung, Ordinal hyperplanes ranker with cost sensitivities for age estimation, in: Proceedings of the 2011 IEEE conference on Computer vision and pattern recognition (cvpr), IEEE, 2011, pp. 585–592.

[16] K.H. Liu, S. Yan, C.C.J. Kuo, Age estimation via grouping and decision fusion, IEEE Trans. Inf. Forensics Secur. 10 (11) (2015) 2408–2423.

[17] R. Weng, J. Lu, G. Yang, Y.P. Tan, Multi-feature ordinal ranking for facial age estimation, in: Proceedings of the 2013 10th IEEE international conference and workshops on Automatic face and gesture recognition (FG), IEEE, 2013, pp. 1–6.

[18] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, International Conference on Learning Representations (ICLR'15), San Diego, CA, May 7-9, 2015.

[19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), Boston, MA, June 7-12, 2015.

[20] Z. Hu, Y. Wen, J. Wang, M. Wang, R. Hong, S. Yan, Facial age estimation with age difference, IEEE Trans. Image Process. 26 (7) (2017) 3087–3097.

[21] Z. Niu, M. Zhou, L. Wang, X. Gao, G. Hua, Ordinal regression with multiple output CNN for age estimation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4920–4928.

[22] R. Ranjan, S. Sankaranarayanan, C.D. Castillo, R. Chellappa, An all-in-one convolutional neural network for face analysis, in: Proceedings of the 2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017), IEEE, 2017, pp. 17–24.

[23] R. Rothe, R. Timofte, L. Van Gool, Deep expectation of real and apparent age from a single image without facial landmarks, Int. J. Comput. Vis. (2016) 1–14.

[24] D. Yi, Z. Lei, S.Z. Li, Age estimation by multi-scale convolutional network, in: Proceedings of the Asian Conference on Computer Vision, Cham, Springer, 2014, pp. 144–158.

[25] H.F. Yang, B.Y. Lin, K.Y. Chang, C.S. Chen, Automatic age estimation from face images via deep ranking, Networks 35 (8) (2013) 1872–1886.

[26] B. Yoo, Y. Kwak, Y. Kim, C. Choi, J. Kim, Deep facial age estimation using conditional multitask learning with weak label expansion, IEEE Signal Process. Lett. 25 (6) (2018) 808–812.

[27] H. Liu, J. Lu, J. Feng, J. Zhou, Ordinal deep learning for facial age estimation, IEEE Trans. Circ. Syst. Video Technol. (2017), doi:10.1109/TCSVT.2017.2782709.

[28] J. Tang, Z. Li, Lai H., L. Zhang, S. Yan, Personalized age progression with Bi-level aging dictionary learning, IEEE Trans. Pattern Anal. Mach. Intell. 40 (4) (2018) 905–917.

[29] X. Shu, J. Tang, H. Lai, L. Liu, S. Yan, Personalized age progression with aging dictionary, in: Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3970–3978.

[30] C.N. Duong, K.G. Quach, K. Luu, T.H.N. Le, M. Savvides, Temporal non-volume preserving approach to facial age-progression and age-invariant face recognition, in: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 3755–3763.

[31] P. Tang, H. Wang, S. Kwong, G-MS2F: GoogLeNet based multi-stage feature fusion of deep CNN for scene recognition, Neurocomputing 225 (2017) 188–197.

[32] M. Farrajota, J.M. Rodrigues, J.H. du Buf, Using multi-stage features in fast R-CNN for pedestrian detection, in: Proceedings of the 7th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion, 2016, pp. 400–407.

[33] S. Yang, D. Ramanan, Multi-scale recognition with DAG-CNNs, in: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1215–1223.

[34] M. He, S.J. Horng, P. Fan, R.S. Run, R.J. Chen, J.L. Lai, K.O. Sentosa, Performance evaluation of score level fusion in multimodal biometric systems, Pattern Recognit. 43 (5) (2010) 1789–1800.

[35] T. Gehrig, M. Steiner, H.K. Ekenel, Draft: evaluation guidelines for gender classification and age estimation, Karlsruhe Institute of Technology, 2011 Technical report.

[36] M. Mathias, R. Benenson, M. Pedersoli, L. Van Gool, Face detection without

bells and whistles, in: Proceedings of the European Conference on Computer Vision, Springer, 2014, pp. 720–735.
[37] L. Nanni, S. Ghidoni, S. Brahnam, Handcrafted vs. non-handcrafted features for computer vision classification, Pattern Recognit. 71 (2017) 158–172.
[38] X. Jin, C. Xu, J. Feng, Y. Wei, J. Xiong, S. Yan, Deep learning with S-shaped rectified linear activation units, in: Proceedings of the AAAI, 2016, pp. 1737–1743.
[39] G. Guo, G. Mu, Simultaneous dimensionality reduction and human age estimation via kernel partial least squares regression, in: Proceedings of the 2011 IEEE conference on Computer vision and pattern recognition (cvpr), IEEE, 2011, pp. 657–664.
[40] X. Geng, C. Yin, Z.H. Zhou, Facial age estimation by learning from label distributions, IEEE Trans. Pattern Anal. Mach. Intell. 35 (10) (2013) 2401–2412.
[41] G. Guo, G. Mu, Joint estimation of age, gender and ethnicity: CCA vs. PLS, in: Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), IEEE, 2013, pp. 1–6.
[42] C. Fernández, I. Huerta, A. Prati, A comparative evaluation of regression learning algorithms for facial age estimation, Face and Facial Expression Recognition from Real World Videos, Springer, 2015, pp. 133–144.
[43] I. Huerta, C. Fernández, C. Segura, J. Hernando, A. Prati, A deep analysis on age estimation, Pattern Recognit. Lett. 68 (2015) 239–249.
[44] H. Han, C. Otto, X. Liu, A.K. Jain, Demographic estimation from face images: human vs. machine performance, IEEE Trans. Pattern Anal. Mach. Intell. 37 (6) (2015) 1148–1161.
[45] I. Huerta, C. Fernández, A. Prati, Facial age estimation through the fusion of texture and local appearance descriptors, in: Proceedings of the European Conference on Computer Vision, Springer, 2014, pp. 667–681.
[46] X. Wang, R. Guo, C. Kambhamettu, Deeply-learned feature for age estimation, in: Proceedings of the 2015 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2015, pp. 534–541.
[47] M. Duan, K. Li, C. Yang, K. Li, A hybrid deep learning CNN–ELM for age and gender classification, Neurocomputing 275 (2018) 448–461.
[48] C.C. Ng, Y.T. Cheng, G.S. Hsu, M.H. Yap, Multi-layer age regression for face age estimation, in: Proceedings of the 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), IEEE, 2017, pp. 294–297.
[49] G. Antipov, M. Baccouche, S.A. Berrani, J.L. Dugelay, Apparent age estimation from face images combining general and children-specialized deep learning models, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2016, pp. 96–104.
[50] M.Y. El Dib, M. El-Saban, Human age estimation using enhanced bio-inspired features (ebif), in: Proceedings of the 2010 IEEE International Conference on Image Processing, Hong Kong, 2010, pp. 1589–1592.
[51] Y. Liang, X. Wang, L. Zhang, Z. Wang, A hierarchical framework for facial age estimation, Math. Probl. Eng. (2014) 1–8.
[52] J. Lu, V.E. Liong, J. Zhou, Cost-sensitive local binary feature learning for facial age estimation, IEEE Trans. Image Process. 24 (12) (2015) 5356–5368.
[53] K.Y. Chang, C.S. Chen, A learning framework for age rank estimation based on face images with scattering transform, IEEE Trans. Image Process. 24 (3) (2015) 785–798.
[54] K. Chen, S. Gong, T. Xiang, C.C. Loy, Cumulative attribute space for age and crowd density estimation, in: Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2013, pp. 2467–2474.
[55] J.C. Chen, A. Kumar, R. Ranjan, V.M. Patel, A. Alavi, R. Chellappa, A cascaded convolutional neural network for age estimation of unconstrained faces, in: Proceedings of the 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), IEEE, 2016, pp. 1–8.
[56] H. Liu, J. Lu, J. Feng, J. Zhou, Label-sensitive deep metric learning for facial age estimation, IEEE Trans. Inf. Forensics Secur. 13 (2) (2018) 292–305.

**Shahram Taheri** received the B.Eng. and M.Sc. degrees in computer engineering from the Shiraz University, Iran, in 2001 and 2004, respectively. He is currently a Ph.D. candidate in the Eastern Mediterranean University, North Cyprus. His research interests include machine learning, biometrics and neural networks.

**Önsen Toygar** received her B.S., M.S. and Ph.D. degrees in 1997, 1999 and 2004, respectively from Computer Engineering Department of Eastern Mediterranean University, Northern Cyprus. Since September 2004, she worked in Computer Engineering Department of Eastern Mediterranean University. She is currently an Associate Professor in the department and served as the Vice Chair of the department between September 2011 and January 2013. Her current research interests are in the area of biometrics, computer vision, image processing and digital forensics.