_____

A Robust Feature Selection Using Modified Whale Optimization And Adversarial Networks Based Classifier For Age Estimation On Facial Wrinkles

Shifana E. M

Research Scholar

PG and Research Department, Department of Computer Science

Park's College, Chinnakarai, Tirupur, India.

Dr. S. Prabhu

Assistant Professor (GL)

Department of Computer Science, Government Arts & Science College

Thittamalai, Nambiyur, India

Abstract

Facial wrinkles are considered to be a facial feature, and appear as people get older. The changes of facial wrinkles depend on the nature of skin and muscle contraction. Therefore, the detection of wrinkles plays an important role in applications that depend on facial skin changes, such as face age estimation . While existing wrinkle detection algorithms focus on forehead horizontal lines, it is necessary to develop new methods to detect all wrinkles (vertical and horizontal) on whole face. However, achieving high prediction accuracy without human intervention (such as preprocessing and hand-crafted feature extraction) is currently and potentially a challenge. To solve this problem, this research work proposed a robust feature selection using modified whale optimization and deep CNN-based age estimation method that utilized by conditional GAN. In this work, the age estimation system is divided into five distinct stages. The first step of the proposed system is preprocessing performed by rotation angle. The second step is to extract local features including Gabor wavelets (GW), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradients (HOG). These attributes are then combined in the third step as a feature fusion method, which combines four different feature extraction methods. In the following fourth step, Modified Whale Optimization (MWO) as a meta-heuristic optimization algorithm is used to decrease the size of attributes and find optimal features. Finally, propose a classification and regression methods to estimate the age and age groups. Initially, GAN is used to determine the age groups, and then CNN used to estimate the ages within those groups. In an experiment using two open databases (the FG-NET database and the MORPH database), when the proposed method in this study was applied to state-of-the art methods for age estimation, the age estimation achieved a higher accuracy than when using low-resolution images.

Keywords: Age estimation, face wrinkles, Gabor wavelets (GW), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradients (HOG), Modified Whale Optimization (MWO) and convolutional Neural network (CNN).

Introduction

A recent popular study showed that more than 27.5% of adults had insufficient physical activity worldwide. The study included 358 population-based surveys in 168 countries with a total of 1.9 million participants. Numerous health risks such as hypertension, diabetes, mental health, and weight-gain are related directly to physical activity [1]. With aging, intensity of physical activity tends to decrease for older people, and this is more evident for females. Various organizations have recommended levels of physical activity for different age groups. However, the exact relationship between physical activity and aging is still unclear. For instance, there is still the question of whether physical activity can predict age. The process of aging is complex and affects all biological systems. Age has a deep connection with health and mortality [2]. In general, a younger person is expected to have a better health condition, to be physically more active, and to have lower mortality hazard in comparison with a relatively older person. But two different people of the same

_____

chronological age may have very different health profiles and mortality hazards. This brings up an important classification of age, namely, chronological age versus biological age.

Age estimation systems are generally designed to use two steps: an aging feature extraction and a feature classification. Feature extraction is very important in age estimation, since the extracted features greatly affect the classification performance [3]. For this reason, a great deal of effort has been directed towards the extraction of discriminative aging features. These features can be categorized into local and global features, and hybrid features, which are a combination of the global and local features. Local features consist of the amount and depth of wrinkles, skin aging using freckles and age spots, hair color and the geometry of facial components. The local features have been commonly used to classify people into age groups (e.g. babies, young adults and senior adults) as they possess unique characteristics that distinguish specific age groups [4]. For example, wrinkles are found in adulthood rather than in childhood, and geometric features, such as the distance ratios between features such as the eyes, the end of the nose, and the corners of the mouth, are noticeably changed in childhood rather than in adulthood. Consequently, these features are better suited to applications requiring an age group classification (e.g. the class of less than 20 years old, the class of 20–39 years old, etc.) rather than a detailed age estimation (e.g. 17, 23 years old, etc.). Conversely, unlike the local features, the global features are generally used to estimate a detailed age and contain not only aging characteristics, but other individual characteristics such as identity, expression, gender, ethnicity, among others. Therefore, individual characteristics are reflected better in global features than aging characteristics. For instance, active appearance models (AAM) [5] are mainly used to estimate an age as global features, because they offer greater amounts of information concerning the appearance and shape of a face than local features. However, AAM features do not include detailed wrinkle and skin informations, due to the dimensional reduction made by the principle component analysis (PCA). This defect in the global features can be resolved through combination with local features. In order to solve this problem, an age estimation method based on the hybrid features [6], and the defects found in each feature can then be compensated for by combining the global and local features. Therefore, hybrid features are desirable for the accurate estimation of age.

Deep learning algorithms develop rapidly recently and have achieved success in many pattern recognition tasks. In a Deep ConvNets structure for 1000 classification task and get the top performance. A deeper and wider structure compared with Alex-Net to improve the classification performance. Also increased the depth and width of CNN for classification and detection task. Some researchers have applied deep learning algorithms to other pattern recognition tasks such as speech recognition and pedestrian detection [7]. These experiments demonstrate that high-level semantic features extracted by deep learning algorithms usually perform better than handcraft features. However, as introduced in the first paragraph, traditional age estimation methods usually use low-level handcraft features to train a model. So, in this paper, we propose a novel age estimation method based on Deep ConvNets. Train Deep ConvNets to get high-level facial features to improve the age estimation performance. While existing wrinkle detection algorithms focus on forehead horizontal lines, it is necessary to develop new methods to detect all wrinkles (vertical and horizontal) on whole face. However, achieving high prediction accuracy without human intervention (such as preprocessing and hand-crafted feature extraction) is currently and potentially a challenge [8]. To solve this problem, this research work proposed a robust feature selection using modified whale optimization and deep CNN-based age estimation method that utilized by conditional GAN.

The rest of the research work is structured as follows as, section 2 reviews the some of the recent techniques for age estimation using machine learning and deep learning techniques. section 3 presents the process of the proposed methodology. section 4 provides the results and discussion. section 5 deals with the conclusion and future work.

1. Literature Review

In this section reviews the some of the recent techniques for the estimation of human age using face wrinkles.

Dong et al [9] proposed a deep learning based framework for age classification task. In our proposed algorithm, Deep Convolutional Neural Networks (Deep ConvNets) are used to extract high-level complex age related visual features and predict age range of input face image. Due to lack of age labeled face images, we use the transfer learning strategy to train the Deep ConvNets. In addition, to describe the relationships between labels that compose an ordered sequence, we define a new loss function in the training process of age classification task. The experiments are conducted on a widely used age estimation dataset-Images of Groups of

People. The experimental results demonstrate the excellent performance of our proposed algorithm against the state-of-the-art methods. Štern et al [10] proposed an automatic multi-factorial age estimation method based on MRI data of hand, clavicle, and teeth to extend the maximal age range from up to 19 years, as commonly used for age assessment based on hand bones, to up to 25 years, when combined with clavicle bones and wisdom teeth. Fusing age-relevant information from all three anatomical sites, our method utilizes a deep convolutional neural network that is trained on a dataset of 322 subjects in the age range between 13 and 25 years, to achieve a mean absolute prediction error in regressing chronological age of 1.01±0.74 years. Luu et al [11] introduced a novel age estimation technique that combines Active Appearance Models (AAMs) and Support Vector Machines (SVMs), to dramatically improve the accuracy of age estimation over the current state-of-the-art techniques. In this method, characteristics of the input images, face image, are interpreted as feature vectors by AAMs, which are used to discriminate between childhood and adulthood, prior to age estimation. Faces classified as adults are passed to the adult age-determination function and the others are passed to the child age-determination function. Compared to published results, this method yields the highest accuracy recognition rates, both in overall mean-absolute error (MAE) and mean-absolute error for the two periods of human development: childhood and adulthood.

Geng et al [12] proposed an automatic age estimation method named AGES (AGing pattErn Subspace). The basic idea is to model the aging pattern, which is defined as the sequence of a particular individual's face images sorted in time order, by constructing a representative subspace. The proper aging pattern for a previously unseen face image is determined by the projection in the subspace that can reconstruct the face image with minimum reconstruction error, while the position of the face image in that aging pattern will then indicate its age. In the experiments, AGES and its variants are compared with the limited existing age estimation methods (WAS and AAS) and some well-established classification methods (kNN, BP, C4.5, and SVM). Pontes et al [13] presented a new framework for the integration of Active Appearance Models (AAM), Local Binary Patterns (LBP), Gabor wavelets (GW) and Local Phase Quantization (LPQ) in order to obtain a highly discriminative feature representation which is able to model shape, appearance, wrinkles and skin spots. In addition, this paper proposes a novel flexible hierarchical age estimation approach consisting of a multi-class Support Vector Machine (SVM) to classify a subject into an age group followed by a Support Vector Regression (SVR) to estimate a specific age. The errors that may happen in the classification step, caused by the hard boundaries between age classes, are compensated in the specific age estimation by a flexible overlapping of the age ranges. The performance of the proposed approach was evaluated on FG-NET Aging and MORPH Album 2 datasets and a mean absolute error (*MAE*) of 4.50 and 5.86 years was achieved respectively. Dehshibi et al [14] a new algorithm for age-group recognition from frontal face image is presented. The algorithm classifies subjects into four different age categories in four key stages: Pre-processing, facial feature extraction by a novel geometric feature-based method, face feature analysis, and age classification. In order to apply the algorithm to the problem, a face image database focusing on people's age information is required. Because there were no such databases, we created a database for this purpose, which is called Iranian face database (IFDB). IFDB contains digital images of people from 1 to 85 years of age. After pre-processing, then primary features of the faces in the database will be accurately detected. Finally, a neural network is used to classify the face into age groups using computed facial feature ratios and wrinkle densities. Experimental results show that the algorithm identifies the age group with accuracy of 86.64%.

Liu et al [15] presented a novel multistage learning system, called grouping estimation fusion (GEF), for human age estimation via facial images. The GEF consists of three stages: (1) age grouping; (2) age estimation within age groups; and (3) decision fusion for final age estimation. In the first stage, faces are classified into different groups, where each group has a different age range. In the second stage, three methods are adopted to extract global features from the whole face and local features from facial components (e.g., eyes, nose, and mouth). Each global or local feature is individually utilized for age estimation in each group. Thus, several decisions (i.e., estimation results) are derived. In the third stage, we obtain the final estimated age by fusing the diverse decisions from the second stage. The performance of the GEF system is evaluated on the Face and Gesture Recognition Research Network and the MORPH-II databases, and it outperforms the existing state-of-the-art age estimation methods by a significant margin. Ashiqur Rahman et al [16] introduced an approach based on deep convolutional long short term memory (ConvLSTM) to predict biological age, using human physical activity as recorded by a wearable device. We also demonstrate five deep biological age estimation models including the proposed approach and compare their performance on the NHANES physical activity dataset. Results on mortality hazard analysis using both the Cox proportional hazard model and Kaplan-Meier curves each show that the proposed method for estimating biological age outperforms other state-of-the-art approaches.

2. Proposed Methodology

This research work proposed a robust feature selection using modified whale optimization and deep CNN-based age estimation method that utilized by conditional GAN. In this work, the age estimation system is divided into five distinct stages.

- The first step of the proposed system is preprocessing performed by rotation angle.
- The second step is to extract local features including Gabor wavelets (GW), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradients (HOG).
- These attributes are then combined in the third step as a feature fusion method, which combines four different feature extraction methods.
- In the following fourth step, Modified Whale Optimization (MWO) as a meta-heuristic optimization algorithm is used to decrease the size of attributes and find optimal features.
- Finally, propose a classification methods to estimate the age and age groups. Initially, low resolution facial images are restored using the conditional GAN, which was learned for SR by pairs of low- and high-resolution facial images. GAN is used to determine the age groups, and then CNN used to estimate the ages within those groups.

The proposed age estimation method in this work, which is robust to low-resolution facial images, is implemented in the following steps described in Figure 1.
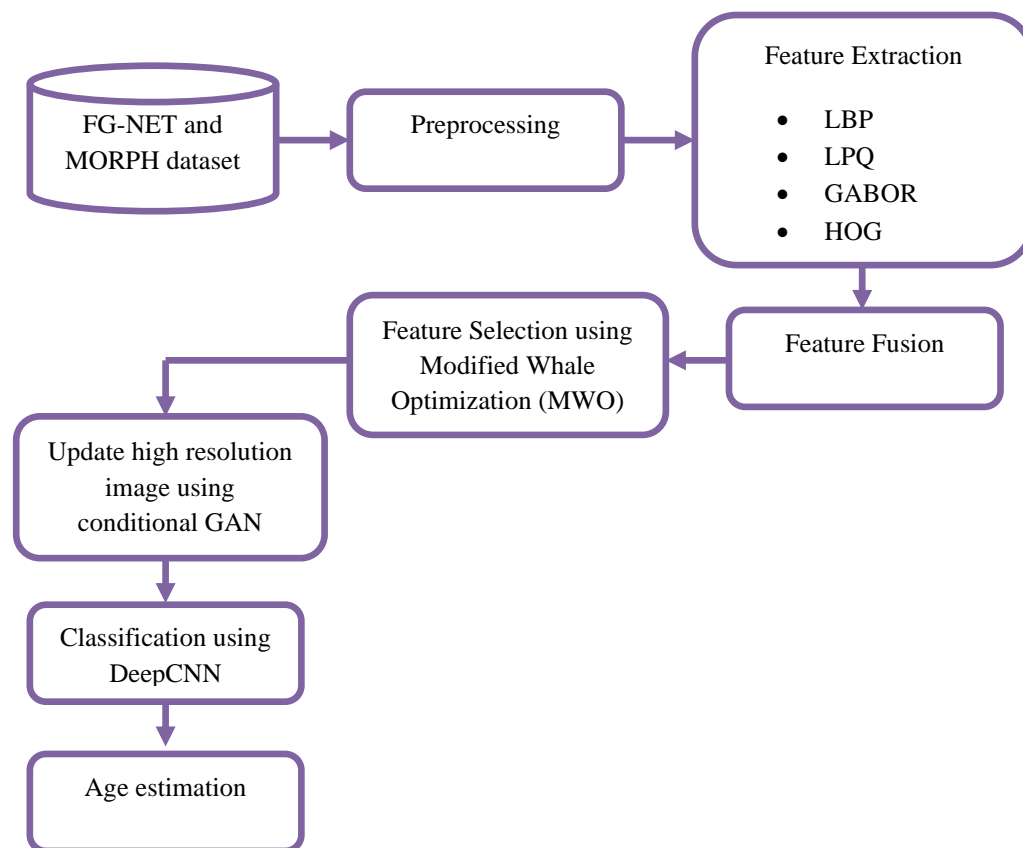

Figure 1. The overall process of the proposed methodology

3.1. Dataset

To assess the efficacy of the age estimation algorithm, we need a database with a high statistical population. In this work, two popular datasets are used: FG-NET and MORPH. The FGNET dataset includes 1002 facial images of 82 persons between the ages of 0 and 69 years, with an average of 12 images for each person. This database is unbalanced because half of the instances are younger than 13 years old. The photos, which are both color and greyscale, have an average resolution of 384x487 pixels. Also, there is extreme diversity in the pose, facial expression, blur, and lighting.

_____

The Age Group at the University of Northern California created the MORPH dataset, which is freely available to the public. This dataset is split into two parts. Race, sex, birth date, and age are all included in both categories. The MORPH includes 55,608 images of 13000 persons ranging in age from 16 to 77. The average dimension of these images is 400x480 pixels.

3.2. Image Preprocessing

In many practical applications, it is difficult to ask users to keep frontal face postures, which will further affect the accurate localization of local facial regions and the computation of facial geometric measurements. To decrease the influence of non-frontal facial posture, conduct sample rotation in-plane. First, the rotation angle θ in-plane is computed using the location of two eyes. If the angle θ is larger than a specific threshold, rotation correction with θ will be added for the sample. Considering marking deviations of facial landmarks, set the threshold as 5^0. The computation of rotation angle θ in-plane is shown as follows,

$$\theta = argtan\frac{y_L - y_R}{x_L - x_R} \tag{1}$$

where (x_L, y_L) and (x_R, y_R) are the locations of the left and right eyes, respectively.

3.3. Feature Extraction

Local features are obtained using the LBP, LPQ, HOG, and Gabor Algorithms.

1) Local Binary Pattern (LBP)

The LBP algorithm is widely applied for texture classification, face analysis, motion analysis, facial description, sex, and identifying an age group. Aging skin has very details, so the texture features are extracted using the LBP method. LBP technique can detect microstructure patterns such as spots, lines, edges, and flat areas on the skin. Based on LBP, each pixel is assigned a value that is compared with its neighbors. The LBP code can be defined as follows:

$$LBP_{P,R} = \sum_{p=0}^{p-1} s(g_p - g_c) 2^P \tag{2}$$

Where P denotes the number of neighboring pixels, and the distance between the center and the neighboring pixels is represented by R. Also, g_p and g_c show gray-level values of the central pixels and P surrounding pixels, respectively. The function s is expressed by:

$$s(x) = f(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \tag{3}$$

2) Local Phase Quantization (LPQ)

The LPQ is utilized as an image descriptor that is based on the blur invariance property in Fourier processing. The LPQ strategy uses local phase data that phase is inspected over a square neighborhood N_x of size M by M:

$$F(u,x) = \sum_{y \in N_x} f(x-y)e^{-j2\pi u^T y} = w_u^T f_x \tag{4}$$

Where w_u is the vector of the 2D DFT at a frequency of u, and f_x is a vector including whole images from N_x. Four frequency points are considered in $LPQ : u_1 = [a,0]^T, u_2 = [0,a]^T, u_3 = [a,a]^T, and\ u_4 = [a,-a]^T$, where a is the adequately small scalar. An F_x vector is generated for each pixel position:

$$F_x = [F(u_1,x) + F(u_2,x) + F(u_3,x) + F(u_4,x)] \tag{5}$$

Afterward, the phase information is obtained by considering the sign of the real and imaginary part of each Fourier coefficient as a binary coefficient. As a result, these binary values are expressed as integer numbers ranging from 0 to 255.

3) Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradient (HOG) is one of the most significant feature extractors that is used to recognize objects in many different applications. It is employed to extract the magnitude and edge orientation of

_____

an image. Therefore, the general method of HOG is to first divide an image into smaller parts called cells, then a histogram of edge orientation is computed for each one. In the first step, the image gradients are computed by using the following filter kernel:

$$g_x = [1,0,-1] \qquad (6)$$

$$g_y = [1,0,-1]^T \qquad (7)$$

The next stage is to create cell histograms. The cells could have a rectangular or radial shape. Furthermore, histograms are equally distributed between 0 and 180 degrees as well as 0 and 360 degrees, depending on whether the gradient is "unsigned" or "signed." Histogram counts are normalized to offset the illumination, then use the results obtained to normalize every cell within the block. The blocks are normalized as follows:

$$L_1 - norm: u \rightarrow {}^u\!/_{(\|u\|)_1 + \in} \qquad (8)$$

$$L_1 - sqrt: u \rightarrow \sqrt{{}^u\!/_{(\|u\|)_1 + \in}} \qquad (9)$$

$$L_2 - norm: u \rightarrow {}^u\!\Big/_{\sqrt{{}^u\!/_{(\|u\|_2{}^2 + \in^2)}}} \qquad (10)$$

$$L_2 - hys: L_2 - norm, plus\ clipping\ at\ .2\ and\ renormalizing$$

Where u is the non-normalized vector containing all histograms in a given block, ||v||k is k-norm for k=1, 2 and $\in$ is a small constant. Finally, the integration of these histograms illustrates the HOG model.

4) Gabor Wavelet

For estimating age, a facial wrinkle is one of the most valuable features. Wrinkles form for a variety of reasons and can develop in a variety of directions. For example, it may create wrinkle lines based on muscle movements or facial expressions in a smiling mood. The essential point is that each wrinkle is unique. As a result, it could be used in an age estimation system. Therefore, the Gabor filter is used to extract wrinkle and texture characteristics. A further benefit of Gabor wavelets is their ability to extract specific information about the facial area. Gabor filters produce features with considerably large dimensional space, making all subsequent computations costly. Solving this issue can be achieved by reducing dimensionality with the MWO algorithm.

3.4. Feature Fusion

The integration and combination of features have been used effectively in several fields such as biometric security, facial recognition, photos, etc. Due to that, each feature vector contains specific information, combining characteristics can create a comprehensive set of features. Therefore, a feature fusion method is employed to combine the extracted features and provide an informative set of normalized local features.

$$f_{fused} = \left(f'_{LBP} f'_{LPQ} f'_{HOG} f'_{Gabor}\right) \qquad (11)$$

$$= (c_1, \dots \dots c_k, w_1, \dots \dots w_l, t1_1, \dots \dots t1_m, t2_1, \dots \dots t2_n) \qquad (12)$$

Where f_{fused} refers to the combined feature vector that will be utilized in the reduction of dimensionality stage, as discussed in the following section.

3.5. Feature selection using Improved Weight based Whale Optimization Algorithm (IWWOA)

The WOA is a meta-heuristic optimization algorithm. It is a nature-inspired approach that mimics the real-life behavior of a group of the largest mammals on the planet [17]. WOA is a swarm-based technique that is designed based on the social behavior of humpback whales and takes inspiration from the bubble-net strategy unique to them for hunting in the ocean. Humpback whales are the largest group of baleen whales and they usually spend their days as a group. They hunt small groups of krill and small fishes close to the surface by

_____

creating bubbles along a spiral path around their prey and then they swim up to the surface following this path. Here, consider the three typical behaviours of Humpback whales and model them mathematically.

Encircling of Prey: The hyper parameters (Maximum epochs, mini batch size, initial learning rate, l_2 regularization parameter, shuffle type, and momentum) are considered as population of whales X_i ($i= 1, 2, 3, ..., n$).). The recognition error is considered as fitness function. The Humpback whales can recognize the location of prey and encircle them. The WOA algorithm assumes that the current best candidate solution is the target optimal parameter values or is close to the optimum. After the best search agent is defined, the other search agents will hence try to update their positions towards the best search agent. This behaviour is represented by

$$\vec{D} = |\vec{C}.\overrightarrow{X^*}(t) - \vec{X}(t)|, \qquad (13)$$

$$\vec{X}(t + 1) = \overrightarrow{X^*}(t) - \vec{A}.\vec{D} \qquad (14)$$

In order to improve the performance of the WOA, the time-varying inertia weight is proposed and updation in equation (14)

$$\text{Inertia weight } \omega = (\omega_1 - \omega_2)\left(\frac{Iter_{max}-Iter}{Iter}\right) + \omega_2 \quad (15)$$

$$\vec{X}(t + 1) = \omega.\overrightarrow{X^*}(t) - \vec{A}.\vec{D} \qquad (16)$$

Where, ω_1 is initial value of the inertia weight, ω_2 final values of the inertia weight, iter current iteration, the $Iter_{max}$ is the maximum number of allowable iteration, t represent the present iteration $\vec{A}$ and $\vec{C}$ refer to the coefficient vectors. The $X*$ indicates the location vector of the best solution which is acquired till now and $\overrightarrow{X^*}$ stands for the location vector. If a better solution exists, then X *has* to be updated through iteration.

The vectors $\vec{A}$ and $\vec{C}$ are computed as below:

$$\vec{A} = 2\vec{a}.\vec{r} - \vec{a} \qquad (17)$$

$$\vec{C} = 2.\vec{r} \qquad (18)$$

Where $\vec{a}$ is linearly reduced from two to zero with the number of iterations and r stands for the random vector in the range [0, 1]. As mentioned in the previous section, the humpback whales also attack the prey with the bubble-net strategy. This method is mathematically formulated as follows.

Exploitation phase: Also termed as bubble net attack and its operation is defined by two mechanisms given as below:

Shrinking encircling technique: We can locate the search agent, the randomness of $\vec{A}$ in [-1, 1], in between the agent earlier position and the current best position.

Spiral updating position: Here, firstly, the distance between the whale and the prey located at (X, Y) and (X*,Y*), respectively, is calculated. A spiral equation is then created between the position of whale and prey as follows

$$\vec{X}(t + 1) = \overrightarrow{D'}.e^{bl}.cos\ cos(2\pi l) + \omega.\overrightarrow{X^*}(t) \qquad (19)$$

Exploration phase: WOA helps to achieve the global optimization objective. The $\vec{A}$ is set randomly from [-1,1] to indulge the search agent for navigating from the reference whale, which means that $\vec{A}$ has to be either greater than 1 or less than the -1. Moreover, we can perform the updated position of a search agent, by randomly choosing an agent that allows the WOA to perform global search. The modelling of this exploration mechanism is mathematically expressed as follows:

$$\vec{X}(t + 1) = \overrightarrow{X_{rand}} - \vec{A}.\vec{D} \qquad (20)$$

_____

Where $\overrightarrow{X_{rand}}$ refers to a random position for a random whale, which is selected from the present population.

Algorithm 1. Inertia Weight Whale Optimization (IWWOA)

1.   Initialize the hyper parameters X_i (i= 1, 2, 3, ..., n).
2.   Compute the fitness of parameters.
3.   Set X^*= *best* hyper parameter vaues.
4.   while (t <*maximum number of iterations*) do
5.   for (each hyper parameter) do
6.   Update a, A, C, $l$and p.
7.   if (p <0.5) then
8.   if ($|A|$ < 1) then
9.   The position of current search is updated by the Eq. (16).
10.   Else if ($|A| \geq 1$) then
12.   Select the random parameter Xrand
13.   The position is updated using Eq (20)
14.   End if
16.   else if ($p \geq 0.5$) then
18.   Change the position using the Eq.(19).
19.   End if
20.   End for
22.   Verify whether any search agent goes outside the search space and amend it.
23.   Compute the fitness of every search agent.
24.   Update X^* in case of existence of a better solution
25.   Get near optimal hyperparameter values as output
26.   t = t+1
27.   End while

3.6. Classification and Age estimation

The low resolution facial images are restored using the conditional GAN [18], which was learned for SR by pairs of low- and high-resolution facial images. In the final step, age estimation is conducted for the reconstructed facial images by using various networks like ResNet, VGG, and DEX.

3.6.1. SR by conditional GAN

For robust age estimation of low-resolution facial images, this study conducted SR using a conditional GAN that performed adversarial learning between generator and discriminator. CNN-based SR matches high-resolution patches to features extracted using a filter. Accordingly, the features of low-resolution facial images are extracted through the encoder of the generator. Features thus extracted are matched to the corresponding high-resolution patches by the decoder, thereby restoring the resolution. The existing studies on GAN received a random noise vector z as input and created I^{Out} as a fake image. The image thus created is a model trained for mapping to I^{Target}. Here, the discriminator learns to distinguish I^{Out}, which is the fake image, from I^{Target}, which is the real image. The generator learns to deceive the discriminator into taking I^{Out} as the real image. Accordingly, the loss function can be expressed by Equation (21) below.

$$\mathcal{L}_{GAN}(G,D) = E_{I^{Target}}[logD(I^{Target})] + E_z\left[\log\left(1 - D\big(G(z)\big)\right)\right] \qquad (21)$$

However, the conditional GAN receives a random noise vector **z** and an input image I^{In}as input, and creates I^{Out}, which is a fake image. The image thus created is a model that learns the mapping to I^{Target}. As the goal of this study is SR by adversarial learning, the pairs of low-resolution facial images and high-resolution original facial images are input as $I^{Low}$and I^{High}, respectively. Thus, the network can learn to map from I^{Recons}: to I^{High}. This process of conditional GAN is illustrated in Figure 2, where G and D denote the generator and the

_____

discriminator respectively. In this study, the generator G learns to map high-resolution restored facial images I^{Out} (I^{Recons}:), which correspond to low-resolution facial images I^{In} (I^{Low}), to high-resolution original facial images I^{Target} (I^{High}).
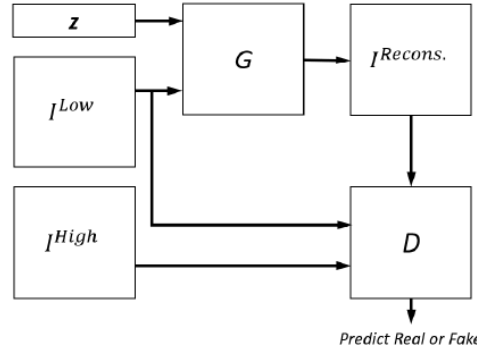

Figure 2.process of conditional GAN

The discriminator D does not simply distinguish between facial images but concatenates I^{Out} and I^{Target} to I^{In}. Thus, the mapping according to I^{In} is reinforced. Consequently, the loss function can be expressed by Equation (22) below.

$$\mathcal{L}_{GAN}(G,D) = E_{I^{In}I^{Target}}[logD(I^{In},I^{Target})] + E_{I^{In},z}\left[\log\left(1 - D\left(I^{In},G(I^{In},z)\right)\right)\right] \qquad (22)$$

The existing loss function to that of GAN for the generator. Here, the discriminator played the same role, but the generator created sharper images by calculating the *L2* distance between I^{Out} and I^{Target}. However, the *L2* distance tends to create more blurred images than the *L1* distance. For this reason, added the *L*1 distance expressed by Equation (23) to the GAN loss.

$$\mathcal{L}_{L1}(G) = E_{I^{In},I^{Target},z}[\|I^{Target} - G(I^{In},z)\|] \qquad (23)$$

Consequently, the ultimate loss function used can be expressed by equation (24).

$$G = arg\ \min_{G}\max_{D}\mathcal{L}_{cGAN}(G,D) + \lambda\mathcal{L}_{L1}(G) \qquad (24)$$

1) Generator

The CNN-based SR is a mapping process where features are extracted from low-resolution images, and corresponding high-resolution images are obtained. Accordingly, SR needs to be conducted so that the existing outer details and shape can be retained as far as possible. The majority of previous studies utilized an encoder-decoder network to create and convert images. This study configured a U-net structure by adding a skip connection to the encoder decoder network. Thus, the feature of *i*th encoder layer was concatenated with that of *n-i*th decoder layer so that the existing outer details and shape could be retained as much as possible. The generator had an encoder-decoder architecture that consisted of eight encoder blocks and eight decoder blocks. Each encoder block used convolution, batch normalization, and leaky-ReLU (no batch normalization was included in the first convolutional layer). Each decoder block used transpose convolution and realized a random noise vector by using dropout for batch normalization. Unlike the encoders, the decoders used ReLU not leaky-ReLU. Finally, the tanh function was applied to the features obtained from the decoders.

2) Discriminator

A discriminator is trained to distinguish real images from fake images. This study extracted features through convolution after the concatenation of the input image I^{In}, and created image I^{Out} or input image I^{In} and target image I^{Target}. Here, in order to distinguish real images from fake ones, a feature map with a size of $30 \times 30 \times 1$, which had been extracted from the last layer, was not checked against the $\mathcal{L}_{L1}$ loss and $\mathcal{L}_{L2}$loss, but instead, each gridwas individually judged. Thus, the details and shapes of each image could be examined.

_____

Moreover, any blurry result, which is an image problem caused by $\mathcal{L}_{L1}$ loss and $\mathcal{L}_{L2}$ loss, could be minimized. Here, each grid had a 70×70 receptive field. Real images were distinguished from fake images using a Markov random field. This is defined as the patchGAN. The patch of this patchGAN moves in an overall image and makes a judgement whether a local region is real or fake. The image is effectively modeled as a Markov random field by the discriminator because each patch can be regarded as being independent and assume the independence between pixels separated by more than a patch diameter. The output of the discriminator is a matrix of probability, where each element provides the probability of being real for a pair of corresponding patches sampled using Markov random field or PatchGAN.

A generator proceeds with learning to deceive a discriminator by using a created image I^{Out}. As the learning time increases, the generator learns not to create an image that is similar to the real image, but to simply deceive the discriminator. Accordingly, the discriminator can also be wrongly trained. This study made the discriminator learn target images and thus maintain the characteristics of real images. Moreover, I^{Out} and I^{Target} were not simply input but were concatenated with I^{In}. Consequently, the discriminator could be trained to express well the details and shape of I^{In}.

3.6.2. Age estimation

In this step, age estimation was conducted by training the deep CNN with restored facial images.

1) RESNET

ResNet is a network that has achieved excellent classification performance [19]. ResNet consists of a bottleneck structure using continuous filters of 3×3 and 1×1 size and a skin connection structure, which concatenates the feature map of the previous layer with the feature map after the residual block. For this reason, the dimension and complexity of feature maps can be reduced. Additionally, as batch normalization is applied, the feature map of data of mini batch size is normalized according to the mean and standard deviation. The learning speed is also improved. ResNet has various depths depending on the repetition of the residual block. In experiment, ResNet-50 and ResNet-152 layers were used. After the fully connected layers (FCL) in the last column of the network, classification was performed by applying the softmax function to obtain the sum of probabilities for the whole class.

2) DEX

DEX [20] is a network that ranked first in the competition for age estimation. DEX has the same architecture as VGG-16 [21], and was constructed by pretraining a model that was already pre-trained by the databases. For age estimation, this study did not apply the existing CNN, which used the probabilities of classes that were obtained by the softmax function, but output the sum of the products of each class label and probabilities as an age after the softmax function, as expressed in Equation (25).

$$Y(X) = \sum_1^n c_i o_i \qquad (25)$$

If there are n classes, c_i and o_i denote the label and probability of the ith class, respectively. Y is the estimated age for the input image X. All the convolution layers and FCLs adopted ReLU as an activation function. In addition, after the first and second max pooling, local response normalization was conducted, and the FCLs were used a dropout. The weights were initialized by zero mean Gaussian random values, of which the standard deviation was 0.01.

3) Inception-V2 With Random Forest

Inceptionv2 has the same architecture as the previous Inception-v1. In particular, a wide network was constructed using filters of diverse sizes rather than deep layers. Inception-v2 was created by adding batch normalization to the inception block of Inception-v1. Inception-v2 was trained, and then features were extracted to train Inception-2 again by using the feature map as a random forest. After that, the ages were estimated.

3. Results and Discussion

In this section, evaluate the effectiveness of the proposed system and existing methods on the widely used FG-NET aging dataset and MORPH dataset. In addition, the following meta-information is available for all the images in the dataset namely: image size, age, gender, spectacles, hat, mustache, beard, horizontal pose and

_____

vertical pose. Since the images were retrieved from real-life albums of different subjects, the face images involve all possible variations including illumination, pose, expression, beards, moustaches, spectacles, etc.

4.1. Performance evaluation metrics

The results of methods are measured via the evaluation metrics such as Structural Similarity Index (SSIM), Jaccard Similarity Index (JSI), Mean Absolute Error (MAE), Cumulative Score (CS), and accuracy.

a) Structural Similarity Index (SSIM)

Structural Similarity Index (SSIM) is used to compute the similarity between the ground truth and the new faces.

b) Jaccard Similarity Index (JSI)

Jaccard Similarity Index (JSI) is used to measure the reliability of wrinkle detection method. The Jaccard index J is calculated by the intersection of A and B divided by the union of A and B. A and B are annotations of two different coders, respectively. It is defined by equation (26),

$$J(\text{A}, B) = \frac{|A \cap B|}{|A \cup B|} \tag{26}$$

c) Accuracy

In order to validate the correctness of the proposed method, accuracy is defined by equation (27),

$$Acc = \sum_{i=1}^{N} w_i, w_i = \begin{cases} 1 & if \; JSI_i > 40\% \\ 0 & else \end{cases} \tag{27}$$

where N is the total number of images in the experiment and w is true if JSI > 40%. Any overlap between A and B larger or equal to 40% is considered as correct detection

d) Mean Absolute Error (MAE)

A MAE is defined as the average of the absolute errors between the estimated age labels and the ground truth age labels. It is described by equation (28),

$$MAE = \sum_{i=1}^{N} \left(\frac{|A_i - B_i|}{N} \right) \tag{28}$$

where B_k is the ground truth age for the k-th test image, A_k is the estimated age, and N is the total number of test images.

e) Cumulative Score

A Cumulative Score (CS), is defined by equation (29),

$$CS(j) = \frac{N_{e \leq j}}{N} * 100 \tag{29}$$

where $N_{e \leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years. As a classification problem, the age estimation performance can also be revealed by the classification accuracy. If the ordinal label is available in a large training data set with dense distribution of different ages, CS could be the best way to reflect the performance.

_____
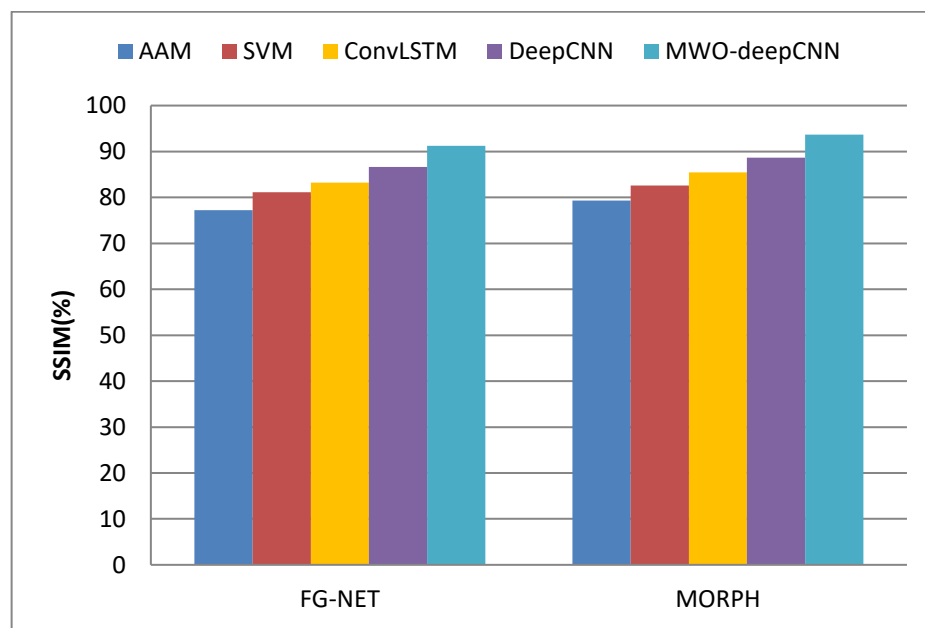
Figure 3. Structural similarity index (SSIM)

Figure 3 shows the performance comparison results of age estimation methods with respect to SSIM metric. The methods like AAM,SVM, ConvLSTM, DeepCNN and proposed MWO-DeepCNN is evaluated under FG-NET database. The proposed MWO-DeepCNN classifier gives higher SSIM value of 91.27%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser SSIM value of 77.22%, 81.12%, 83.23%, and 86.65% for the FG-NET dataset. The proposed MWO-DeepCNN classifier gives higher SSIM value of 93.68%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser SSIM value of 79.35%, 82.57%, 85.47%, and 88.67% for the MORPH dataset.
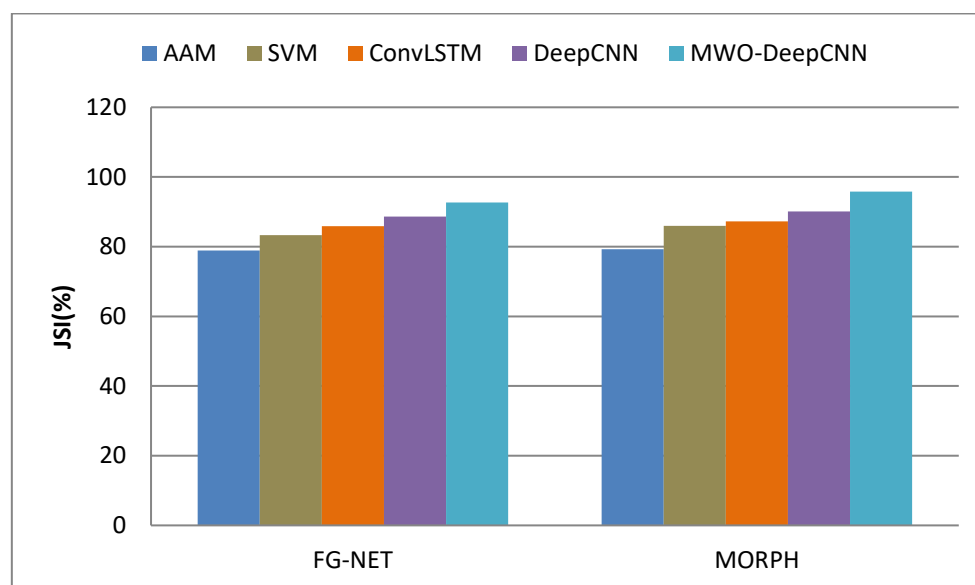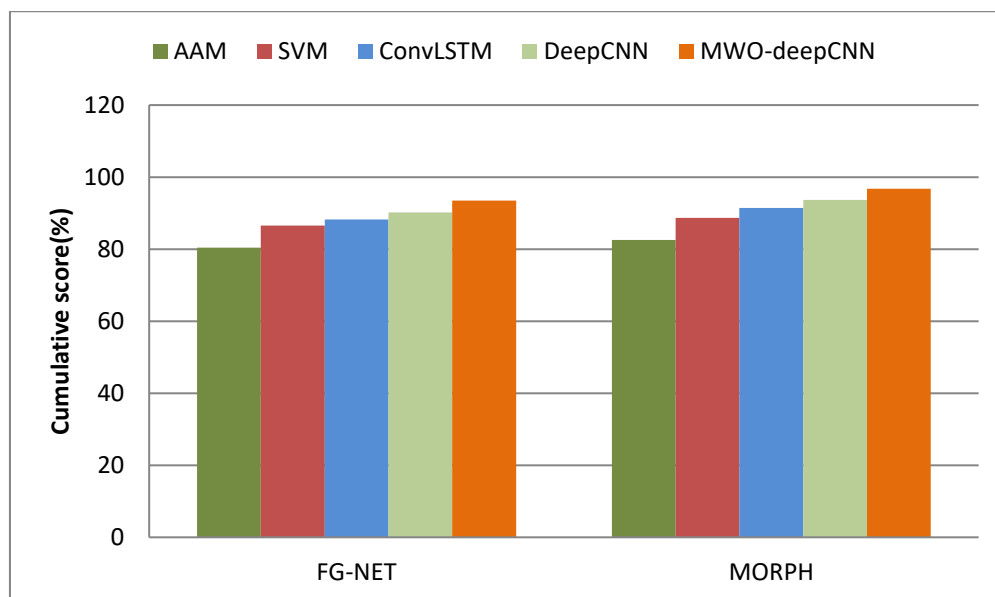

Figure 4. Jaccard similarity index (JSI)

JSI results comparison with respect to age estimation methods are illustrated in figure 4. The methods like AAM,SVM, ConvLSTM, DeepCNN and proposed MWO-DeepCNN is evaluated under FG-NET database. The proposed MWO-DeepCNN classifier gives higher JSI value of 92.72%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser SSIM value of 78.92%, 83.28%, 85.92%, and 88.63% for FG-NET dataset. The proposed MWO-DeepCNN classifier gives higher JSI value of 95.78%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser SSIM value of 79.24%, 85.97%, 87.24%, and 90.14% for MORPH dataset

_____
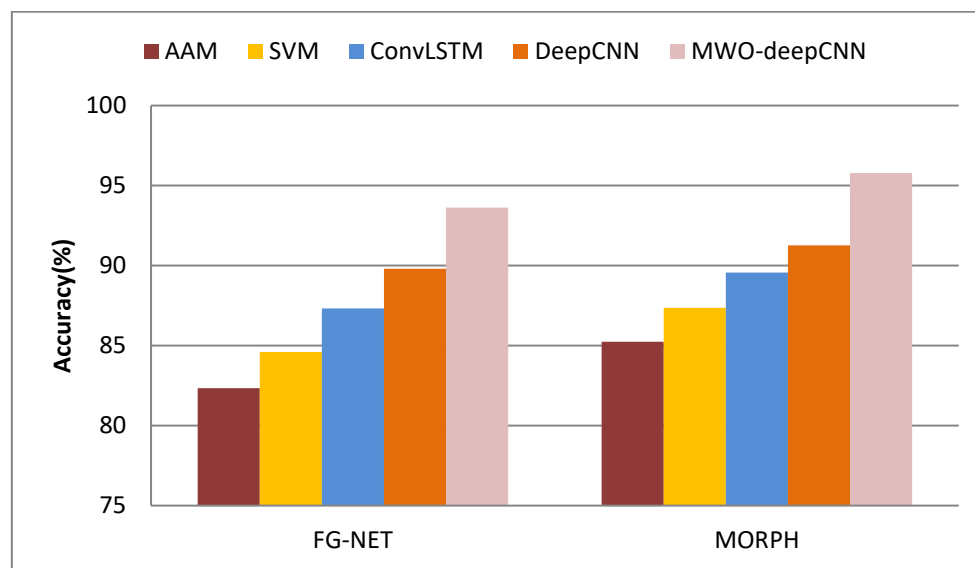
Figure 5. Accuracy

Overall accuracy with respect to age estimation methods are illustrated in figure 5. The estimation methods like AAM,SVM, ConvLSTM, DeepCNN and proposed MWO-DeepCNN is evaluated under FG-NET database. The proposed MWO-DeepCNN classifier provides higher accuracy of 93.62%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser accuracy value of 82.35%, 84.61%, 87.32%, and 89.81% for FG-NET dataset. The proposed MWO-DeepCNN classifier provides higher accuracy of 95.78%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser accuracy value of 85.24%, 87.36%, 89.57%, and 91.27% for MORPH dataset.

Figure 6. Cumulative score

Age estimation methods with respect to cumulative score are illustrated in figure 6. The age estimation methods like AAM,SVM, ConvLSTM, DeepCNN and proposed MWO-DeepCNN is evaluated under FG-NET database. The proposed MWO-DeepCNN classifier provides CS of 93.47%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser CS value of 80.41%,86.58%,88.25%, and 90.16% for FG-NET dataset. The proposed MWO-DeepCNN classifier provides CS of 96.78%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser CS value of 82.57%,88.65%,91.47%, and 93.68% for MORPH dataset.

_____
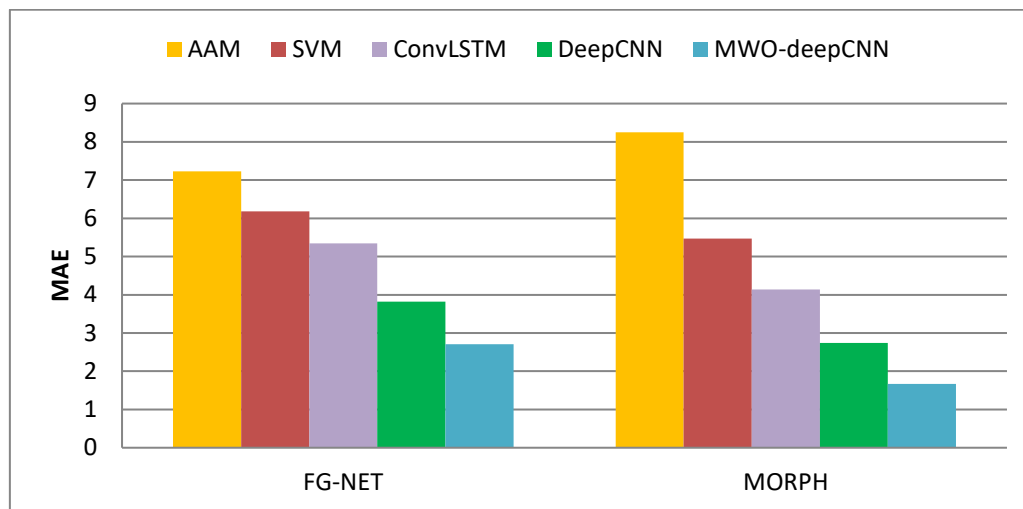
Figure 6. Mean Absolute Error (MAE) results

The Mean Absolute Error (MAE) results comparison of age estimation methods such as AAM,SVM, ConvLSTM, DeepCNN and proposed MWO-DeepCNN with respect to FG-NET database is shown in figure 6. The proposed MWO-DeepCNN classifier provides lesser MAE value of 2.71%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser CS value of 7.23%,6.18%,5.34%, and 3.82% for FG-NET dataset. The proposed MWO-DeepCNN classifier provides lesser MAE value of 1.67%, whereas other methods such as AAM,SVM, ConvLSTM, DeepCNN has lesser CS value of 8.25%,5.47%,4.14%, and 2.74% for MORPH dataset.

4. Conclusion

If low-resolution facial images are acquired, information such as facial wrinkles and facial texture is lost, which degrades the age estimation performance. To solve this problem, this research work proposed a robust feature selection using modified whale optimization and deep CNN-based age estimation method that utilized by conditional GAN. To extract local features including Gabor wavelets (GW), Local Binary Pattern (LBP), Local Phase Quantization (LPQ), and Histogram of Oriented Gradients (HOG). These attributes are then combined in the third step as a feature fusion method, which combines four different feature extraction methods. In the following fourth step, Modified Whale Optimization (MWO) as a meta-heuristic optimization algorithm is used to decrease the size of attributes and find optimal features. Finally, propose a classification methods to estimate the age and age groups. Initially, low resolution facial images are restored using the conditional GAN, which was learned for SR by pairs of low- and high-resolution facial images. GAN is used to determine the age groups, and then CNN used to estimate the ages within those groups. In an experiment using two open databases (the FG-NET database and the MORPH database), when the SR proposed in this study was applied to state-of-the art methods for age estimation, the age estimation achieved a higher accuracy than when using low-resolution images. In addition, the proposed SR based on the conditional GAN showed better SR performance than the existing VDSR and DCSCN. A further study will examine a method to improve the SR and age estimation performance by utilizing video images. The applicability of the proposed method to low-resolution images, which include both optical and motion blurring, must be examined.

REFERENCES

1. Angulu, R., Tapamo, J. R., & Adewumi, A. O. (2018). Age estimation via face images: a survey. *EURASIP Journal on Image and Video Processing*, *2018*(1), 1-35.
2. Rhodes, M. G. (2009). Age estimation of faces: A review. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(1), 1-12.
3. Yap, M. H., Batool, N., Ng, C. C., Rogers, M., & Walker, K. (2021). A survey on facial wrinkles detection and inpainting: datasets, methods, and challenges. *IEEE Transactions on Emerging Topics in Computational Intelligence*, *5*(4), 505-519.
4. Fu, Y., Guo, G., & Huang, T. S. (2010). Age synthesis and estimation via faces: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *32*(11), 1955-1976.
5. Silvia, S., & Supangkat, S. H. (2020, November). A literature review on facial age estimation researches. In *2020 International Conference on ICT for Smart Society (ICISS)* (pp. 1-6). IEEE.

_____

6. Atallah, R. R., Kamsin, A., Ismail, M. A., Abdelrahman, S. A., & Zerdoumi, S. (2018). Face recognition and age estimation implications of changes in facial features: A critical review study. *IEEE Access*, *6*, 28290-28304.
7. Patel, K., & Namdev, K. A Review of Different Techniques of Age Estimation from Human Face.
8. Guo, G., Mu, G., Fu, Y., Dyer, C., & Huang, T. (2009, September). A study on automatic age estimation using a large database. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 1986-1991). IEEE.
9. Dong, Y., Liu, Y., & Lian, S. (2016). Automatic age estimation based on deep learning algorithm. *Neurocomputing*, *187*, 4-10.
10. Štern, D., Payer, C., Giuliani, N., & Urschler, M. (2018). Automatic age estimation and majority age classification from multi-factorial MRI data. *IEEE Journal of Biomedical and Health Informatics*, *23*(4), 1392-1403.
11. Luu, K., Ricanek, K., Bui, T. D., & Suen, C. Y. (2009, September). Age estimation using active appearance models and support vector machine regression. In *2009 IEEE 3Rd international conference on biometrics: theory, applications, and systems* (pp. 1-5). IEEE.
12. Geng, X., Zhou, Z. H., & Smith-Miles, K. (2007). Automatic age estimation based on facial aging patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *29*(12), 2234-2240.
13. Pontes, J. K., Britto Jr, A. S., Fookes, C., & Koerich, A. L. (2016). A flexible hierarchical approach for facial age estimation based on multiple features. *Pattern Recognition*, *54*, 34-51.
14. Dehshibi, M. M., & Bastanfard, A. (2010). A new algorithm for age recognition from facial images. *Signal Processing*, *90*(8), 2431-2444.
15. Liu, K. H., Yan, S., & Kuo, C. C. J. (2015). Age estimation via grouping and decision fusion. *IEEE Transactions on Information Forensics and Security*, *10*(11), 2408-2423.
16. Ashiqur Rahman, S., Giacobbi, P., Pyles, L., Mullett, C., Doretto, G., & Adjeroh, D. A. (2021). Deep learning for biological age estimation. *Briefings in bioinformatics*, *22*(2), 1767-1781.
17. Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51-67.
18. P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, ``Image-to-image translation with conditional adversarial networks,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, Jul. 2017, pp. 5967_5976.
19. K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, Jun. 2016, pp. 770_778.
20. R. Rothe, R. Timofte, and L. Van Gool, ``DEX: Deep expectation of apparent age from a single image,'' in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Santiago, Chile, Dec. 2015, pp. 252_257.
21. K. Simonyan and A. Zisserman, ``Very deep convolutional networks for large-scale image recognition,'' in *Proc. 3rd Int. Conf. Learn. Represent.*, San Diego, CA, USA, May 2015, pp. 1_14.