ORIGINAL ARTICLE

Optimal Deep Learning based Convolution Neural Network for digital forensics Face Sketch Synthesis in internet of things (IoT)

Mohamed Elhoseny[1] · Mahmoud Mohamed Selim[2] · K. Shankar[3]

Abstract

The rapid development in 5G cellular and IoT technologies is expected to be deployed widespread in the next few years. At the same time, crime rates are also increasing to a greater extent while the investigation officers are held responsible to deal with a broad range of cyber and internet issues in investigations. Therefore, advanced IT technologies and IoT devices can be deployed to ease the investigation process, especially, the identification of suspects. At present, only a few research works has been conducted upon deep learning-based Face Sketch Synthesis (FSS) models, concerning its success in diverse application domains including conventional face recognition. This paper proposes a new IoT-enabled Optimal Deep Learning based Convolutional Neural Network (ODL-CNN) for FSS to assist in suspect identification process. The hyper parameter optimization of the DL-CNN model was performed using Improved Elephant Herd Optimization (IEHO) algorithm. In the beginning, the proposed method captures the surveillance videos using IoT-based cameras which are then fed into the proposed ODL-CNN model. The proposed method initially involves preprocessing in which the contrast enhancement process is carried out using Gamma correction method. Then, the ODL-CNN model draws the sketches of the input images following which it undergoes similarity assessment, with professional sketch being drawn as per the directions from eyewitnesses. When the similarity between both the sketches are high, the suspect gets identified. A comprehensive qualitative and quantitative examination was conducted to assess the effectiveness of the presented ODL-CNN model. A detailed simulation analysis pointed out the effective performance of ODL-CNN model with maximum average Peak Signal to Noise Ratio (PSNR) of 20.11dB, Average Structural Similarity (SSIM) of 0.64 and average accuracy of 90.10%.

Keywords 5G IoT · Deep learning · Cybercrime · Forensic sketch · Convolution Neural Network

1 Introduction

In recent years, there is a tremendous development observed in the field of information technology and electronic components while there exists an exponential increase in the growth of 5G IoT systems. These technologies helped the society to improve quality and effectiveness of different urban and suburban services such as healthcare, transport, energy, traffic, etc. At the same time, cybercrimes pose as an essential threat in day-to-day life since such crimes are not collected in a finite set of local crime scenes. Some of the probable traces of evidence are spread across a number of systems, victims, and cross in other jurisdictions than ever before [1]. Digital forensics is an essential part of cybercrime investigation process. It can be defined as the scientific acquisition, investigation, and archival of data contained in electronic media while the information in it, can be utilized as evidence in a court

✉ K. Shankar
  drkshankar@ieee.org

  Mohamed Elhoseny
  mohamed_elhoseny@mans.edu.eg

  Mahmoud Mohamed Selim
  m.selim@pssu.edu.sa

[1] Faculty of Computers and Information, Mansoura University, Dakahlia Governorate, Egypt

[2] Department of Mathematics, College of Science and Humanities in Alaflaj, Prince Sattam bin Abdulaziz University, Alaflaj, Saudi Arabia

[3] Department of Computer Applications, Alagappa University, Karaikudi, India

of law. Forensic image processing includes the computer restoration and improvement of surveillance imaging [2]. It is aimed at exploiting the information extraction from surveillance imagery, particularly for images which are noisy, incomplete or over/underexposed. Forensic imaging processing is a process of enhancing the digital image using different computer techniques. In computer vision, Face Sketch Synthesis (FSS) is a significant area which comprises of extensive range of applications like photo recovery, virtual social network, and face recognition.

Considering face recognition as a negligible issue that might be a real-time bottleneck, it might be succeeded to few extents through relative generation of training data with different variations. Even though the technological growth is visible in image synthesis, the generation of facial images with different modifications and preservation of primary identity are composite. Besides, there is a difficulty still exists in mapping, from a variation factor towards high dimensional face image. It also misleads at times when the human is highly adapted with slight details in the facial area. There are two significant difficulties present in FSS. At the same time, different modifications are also present in semantic face like lighting changes, pose, facial disguises and expressions that are difficult to synthesize within image space. It is still under discussion on how to learn a composite transformation from actual image space in an efficient manner towards appropriate latent semantic space. Even though face recognition had been modeled, the way of preserving the subject identity is difficult. FSS gained attention among the researchers in the recent years, providing importance to the research area. Generation of cross-modal image is one-directional research like FSS. The model of Bayesian synthesis of face sketch is projected in the earlier study [3] in which the task was segmented into two namely weight computation and neighbor selection models.

To speed up the synthesis procedure, a novel model of model-driven FSS was projected in the previous study, and numerous efficient techniques were introduced earlier [4] to enhance the process of neighbor selection. The purpose of an actual process is to repair the face images' occluded parts and to carry out the appearance approximation of another facial portion by employing a morphable model [5]. To enhance the functioning of morphable models, Mo et al. [6] employed a frontal face image dictionary primarily. A novel method was projected [7] by depending on Principal Component Analysis (PCA) and this study demonstrated a goal image through linear Eigenfaces combination. Sparse coding [9] and cascaded pose regression were introduced [8] for generating various types of expressions and poses. Due to artifacts, lack of details in personal facial feature and resolution, the quality of the image produced through actual methods got reduced. On the whole, it is complicated for actual methods to recognize face images.

On the other end, Convolutional Neural Networks (CNN) are applied drastically in recent years. It occupies the central place since it has the capability of massive data processing and hierarchical feature representation in numerous computer vision applications [10–14]. The first layer of convolution is employed to derive the photo features. Due to a misalignment in sketch, the photo is present within the training pair. In order to ease the misalignment, an overlapping max pooling layer [10] was associated towards the primary layer. Subsequently, the nonlinear layer known as multilayer perceptron convolution (mlpconv) layer [15] converts the feature map, towards sketch from photo modalities. To synthesize the end sketch, the convolution layer is employed. Through backpropagation technique, all the parameters and filters were learnt in the training phase [16]. The images were passed through layers in testing phase until the end outcomes were attained. Mlpconv layer is made through multiple Fully Connected layers (FC) with nonlinear activation functions (ReLU) i.e., nonlinear micro-network instantiation, in spite of traditional linear convolution layer in CNN. The mlpconv layer is convolution framework, when the hidden layers counts are 0 and are superior in learning excellent content over convolution function. It also improves the abstraction level of the local model over convolution filter. The abstraction can be defined as the feature which is invariant for a similar idea.

Lu et al. [17], proposed a viable technique which comprises of two phases such as preprocessing and sketch synthesis. Broad investigations on open face sketch database confirmed that the presented strategy improves the sketch synthesis nature of exemplar-based strategies. Ye et al. [18], recommended Triple Interpretation GAN (TTGAN) with multi-layer sparse representation. The authors structured a multi-layer scanty representation model. In this model, L1-standard representation limitation was incorporated into the image age so as to improve the capacity of character safeguarding and the power of the created facial images-to-reconstruction error. The face synthesis exploratory outcomes, on benchmark face databases, unmistakably demonstrated better execution over other contending techniques. In the literature [19], a modified CNN (M-CNN) model was developed comprising of two convolutional layers i.e., a pooling layer and a multilayer perceptron (MLP) convolutional layer to learn the mapping between face photos to sketches. But the performance of this model can be improved by using parameter optimization techniques.

At present, several approaches that recognize the persons from sketches, as instructed by eyewitnesses, have been presented; however, the performance seems to be mostly degraded to be used in real-world forensic sketches and extended galleries which are based on law enforcement mug-shot galleries. Though various FSS models are available, none of the models has incorporated hyper parameter

optimization and IoT enabled applications. In this view, this research work introduces a novel IoT-enabled Optimal Deep Learning-based Convolutional Neural Network (ODL-CNN) for FSS to assist in suspect identification process. The parameters of DL-CNN such as number of hidden layers, number of hidden layer nodes, batch size and learning rate were optimized using Improved Elephant Herd Optimization (IEHO) algorithm. In the proposed method, the sketches of the images are generated from surveillance cameras. Then, the sketches undergo comparison with professional sketch provided with the help of eyewitness. The highly resembling sketches are used to identify the suspects effectively. An extensive set of experimentation was utilized to determine the effectiveness of the ODL-CNN model.

The remaining portions of the paper are organized as follows. Section 2 introduces the presented FSS model. Section 3 validates the introduced FSS model, and Sect. 4 concludes the study.

2 The proposed ODL-CNN model

The general working process of ODL-CNN model is depicted in the Fig. 1. Primarily, the proposed method captures the surveillance videos using IoT-based cameras which are then fed into the proposed ODL-CNN model. The proposed method initially involves preprocessing in which contrast enhancement process is carried out using Gamma correction method. Before the transformation of digital image to face sketch, gamma correction is employed to improve the contrast level of the input image. Then, ODL-CNN model draws the sketches of the input images following which it undergoes similarity assessment with professional sketch drawn as per the directions given by eyewitnesses. When similarity between both the sketches are high, the suspect is identified. Hence, it is a rapid process when compared to exemplar-based techniques that are required to resolve complex optimization issues.

2.1 Contrast enhancement process

The process involved in gamma correction is illustrated in the Fig. 2. In applied input image $I(x, y), x = 1, 2, \ldots, M, y = 1, 2, \ldots, N$, it is primarily identified using the statistical quantity thresholding.
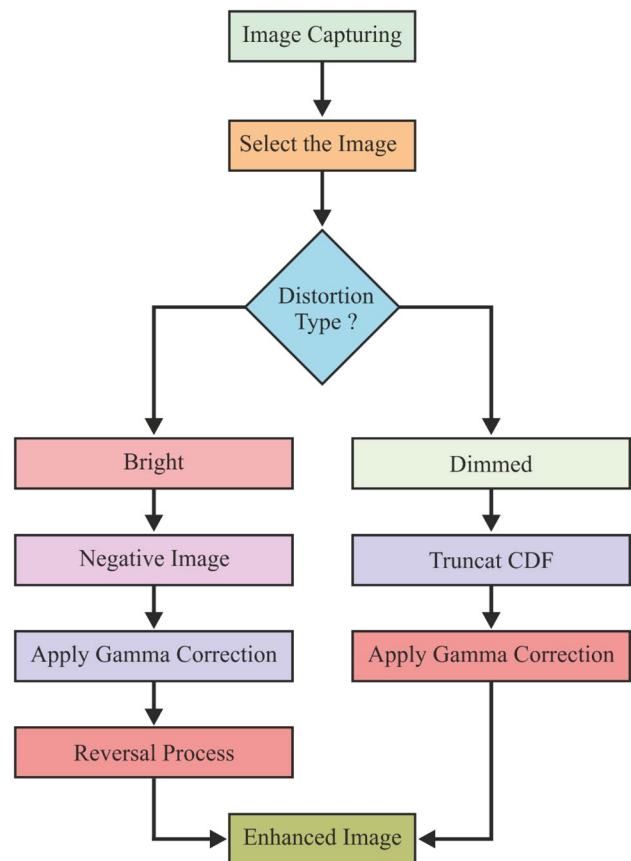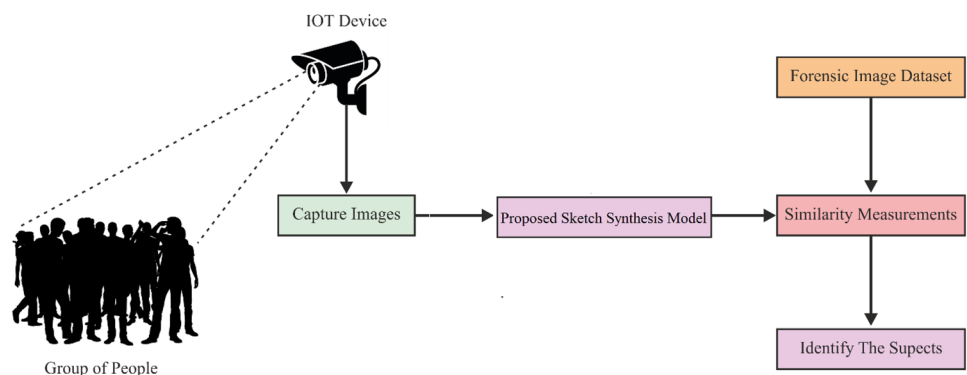
$$t = \frac{m_I - T_t}{T_t} \tag{1}$$

Fig. 2 Contrast enhancement process

Fig. 1 The block diagram of ODL-CNN model

where $m_I = \sum_x \sum_y I(x, y)/MN$. For normal natural images, constant T_t is computed as specified global average brightness. The simulated statistical values, from different benchmark image databases, indicated that T_t is appropriately configured at half of the high pixel intensity, i.e., 128 for 8-bit images. The images, provided as input, are decided as bright if $t > \tau_t$, and dimmed if $t < -\tau_t$, when the threshold τ_t is applied to distinguish the brightness-distorted images from original images. When considering the trade-off between technical applicability and enhancement quality, τ_t is configured experimentally. To restore the brightness and to improve the contrast, the ACG-dependent Cumulative Distribution Function (CDF) truncation and negative image are employed respectively with regards to the identified image types i.e., dimmed and bright ones. The images obtained from contrast enhancement process on some of the test images are given in the Fig. 3.

2.2 ODL-CNN model

2.2.1 Architecture of CNN

The CNN is motivated from cognitive technique of biological vision which is a deep learning technique. The initial CNN idea is made through three elements such as shared weights framework, pooling and receptive field. A neuron corresponds to small area of visual domain in the convolutional layer, as it derives majority of the visual features such as precise direction edges and corners. All the neurons employ similar weight values and formulate feature maps. It indicates that the shared weights find a similar feature in whole visual field which is advantageous to protect the translation invariance and feature continuity. Various filters contain convolutional layer due to which it creates various feature maps. A pooling layer is associated with convolutional layer to minimize the sensitivity of feature locations. Every feature map is divided into non-overlapping rectangles by pooling layer which results in higher value for all rectangles. Hence, the procedure of down sampling becomes nonlinear and therefore minimizes the sum of parameters.

The architecture of CNN comprises of a cascade of pooling and convolutional layers. The final convolutional layer creates exceeding abstract features that might be employed for recognition as shown in the Fig. 4. CNN has the capability of handling huge data and attains higher results in classifying the images with a proposal of GPU acceleration development and ReLU. CNN might be used in numerous computer vision applications like object identification, semantic segmentation, and human action recognition through modeling various loss layers. Mostly,

Fig. 3 Gamma corrected images. **a** Input image, **b** contrast-improved image
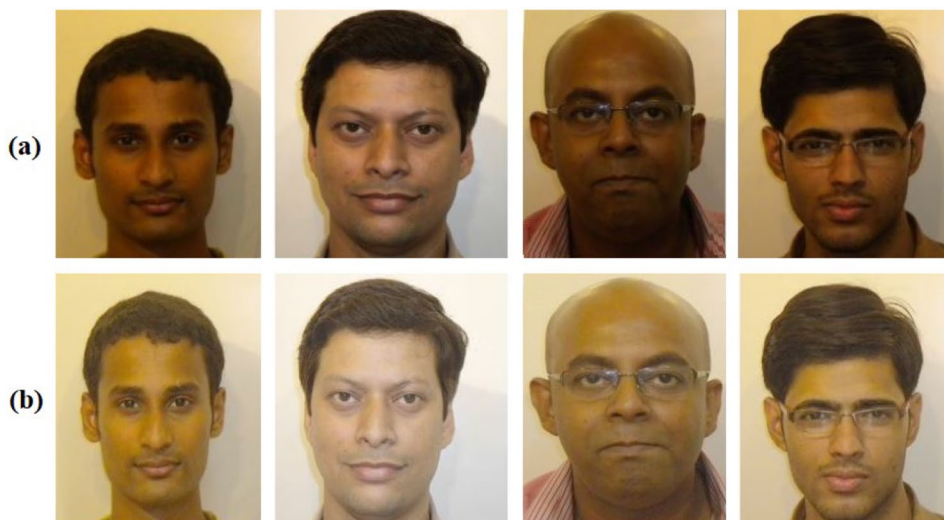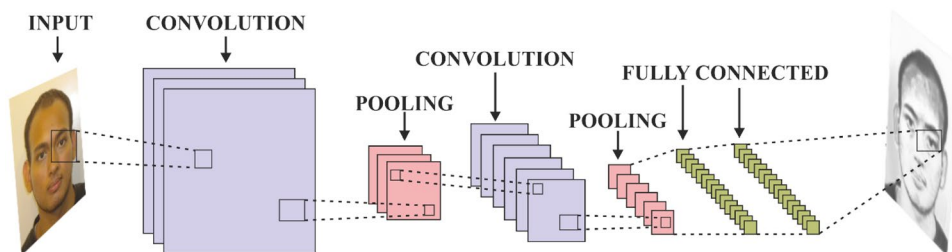

Fig. 4 Structure of CNN

the CNN-based techniques are modeled to make decisions for high-level vision tasks. There exists few types of CNN-based research works for image restoration.

SC-based techniques might be combined with CNN model. Indeed, in convolutional layer, every atom might be considered as a filter in the dictionary. While the filter scans an image in every phase, it also plays like local patches in receptive domain and creates the response accordingly. Therefore, for every patch, a set of filters would create a response vector that belongs to sparse coefficient vector within SC-based technique. However, coefficient mapping among various patches' aggregation and modalities might be considered through convolutional layers. Hence, the traditional SC-based techniques might be defined through three-layer fully-connected CNN. They highly differ in operation mechanism and in algorithm design too.

The notations are employed here. The image matrices are denoted through uppercase letters, X and Y. Vectors are denoted through lower case letters such as b, w, and f. The letters F and W denote mapping function and filter groups respectively. Scalars are denoted by lower case letters, for instance n, b. The main aim is to derive a respective sketch Y for the applied face photo X. The main challenge is to draw mapping $F(\cdot) : Y = F(X)$. The CNN framework offers highly compacted form of SC-based technique, and comprises of various benefits. For mapping approximation, a four-layer network is built.

The first conv. layer is the primary layer that employs a set of filters for input photo scanning while it returns a feature map group. This is expressed as

$$F_1(X) = W_1 * X + b_1 \tag{2}$$

in which b_1 and W_1 demonstrate biases and filters correspondingly; the convolution operation is denoted through '*' and the primary layer output is demonstrated through $F_1(X)$. With supporting $c_{in} \times f_1 \times f_1$, the n_1 filters are comprised of W_1, wherever the receptive field is referred to $f_1 \times f_1$ and input channels' counts are denoted through n_1. To derive the feature map, every filter is convoluted in the procedure of convolution. By adding respective biases, the feature map is rectified.

In general, the sketch and photo within a training pair are not completely registered. To ease the training pair misalignment, a max-pooling layer is employed since it consists of invariance towards small transformations [20]. Every feature map is segmented into nonoverlapping rectangles in traditional max-pooling. Here, overlapping max-pooling layer [21] is used to protect the actual sketch resolution that employs sliding window to output which results in higher value in the accessible domain. The next layer is demonstrated below:

$$F_2(X) = maxPooling\big(F_1(X)\big), \tag{3}$$

The layer input comprises of feature maps of n_1. Over every feature map, with $f_2 \times f_2$ receptive field, maxPooling $(\cdot)$ operator tests the operations of overlapping max-pooling. As every feature map is independently processed in the next layer, n_1 pooled feature maps are offered by $F_2(X)$.

After sparse coding of the face photos, SC-based pipeline method is used, and representation coefficients are converted into sketch modalities. Therefore, the sketches are generally derived through professional artists and the conversion is non-linear. The improvised convolutional structure is known as a layer of mlpconv [22]. This is employed as nonlinear mapping between two modalities. The MLP is inserted within the filter which acts similar to convolution kernel that scans the feature maps of inputs and produces numerous novel feature maps. The mlpconv layer computation is demonstrated as follows

$$f_{i,j,k}^p = \max\Big(w_k^{p^T} f_{i,j}^{p-1} + b_k^p, 0\Big) 1 \le k \le k_p, 1 \le p \le l \tag{4}$$

Let the sum of perceptron layers is denoted through l, in pth layer, k_p refers to the node count and (i, j) refers to the center of the presently processed patch. Furthermore, $f_3 \times f_3$ denotes the receptive-field. The vector $w_k^{p^T}$ is made up of weights that associate towards node $f_{i,j,k}^p$, input from prior layer is denoted as $f_{i,j}^{p-1}$ vector and the bias is denoted through b_k^p. ReLU is used in every node. One ReLU advantage in gradient is it is unsaturated. Since the gradient dispersion problem is alleviated, the attributes of primary neural network layers might get rapidly updated in reverse propagation procedure. At the same time, ReLU computation is faster when compared to tanh and sigmoid functions as the limit is set in the process of forwarding propagation [19]. The activation function for ReLU is denoted as

$$f(x) = max(0, x) \tag{5}$$

The final layer node $f_{i,j,k}^l$ demonstrates the patch (i, j) response in kth feature map of $F_3(X)$ i.e.,

$$F_3(X)_{i,j,k} = f_{i,j,k}{}^1 \tag{6}$$

By exploiting the entire probable (i, j, k) combinations, output $F_3(X)$ is obtained. For final sketch synthesis, convolutional layer is employed. The mapping from digital photo towards sketch could be represented as follows

$$F(X) = W_4 * F_3(X) + b_4 \tag{7}$$

in which, with the support of $n_3 \times f_4 \times f_4$, W_4 contain c_{out} filters, X, the advanced IT technologies and IoT devices can be used to ease the investigation process of identifying the suspects. The IoT devices gather the digital data from crime

scene, examine the social media accounts and transmit to forensic department for further investigation.

2.2.2 Optimization of DL-CNN model using IEHO algorithm

Though the DL-CNN model yields better results, the choice of optimal architectural model for a particular process still remains as an open research problem. Generally, the effectiveness of CNN is decided on the basis of different hyper parameters like depth of CNN, number of layers, filter count, filter size, batch size, step size and learning rate. Several CNN structures were manually considered by developers and were validated to ensure the performance. To automate this process, in this paper, IEHO algorithm was employed to optimize the hyper parameters of DL-CNN model.

In IEHO algorithm, the herding nature of elephants is considered to overcome the global optimization issue and is provided in the following points [23].

1. Elephant population lives as clans, and every individual clan has few numbers of elephants. Actually, a clan has equal and predefined elephant count.
2. Some permanent male elephants would leave families and live alone in a distance at the beginning of every iteration.
3. Elephants in every clan could live along with the leadership of a matriarch. Basically, the oldest elephant is referred as a matriarch in all the clans and is assigned as the correct elephant individual in a clan for optimization issue.

2.2.2.1 Clan updating operator
As mentioned before, every elephant resides with one another jointly under matriarch leadership in all the clans. Hence, for an elephant in a clan ci, the subsequent position is inclined by a matriarch. The elephant j in clan ci can be upgraded as

$$x_{n,ci,j} = x_{ci,j} + \alpha \times (x_{b,ci} - x_{ci,j}) \times r \tag{8}$$

where $x_{n,.cij}$ and x_{cij} are said to be the newly improved and existing positions for elephant j. In clan ci, $\alpha \in [0,1]$ denotes the scaling factor that computes the impact of matriarch ci on x_{cij}. $x_{best,ci}$ shows the matriarch ci i.e., fit elephant individual in clan ci i.e., $r \in [0,1]$. It is defined as the type of stochastic distribution that improves the population diversity in advanced search space. In this study, uniform distribution was applied. Obviously, the fittest elephant in every clan could not be extended, where $x_{cij} = x_{best,ci}$. This scenario can be eliminated by the fittest elephant and has been expanded as,

$$x_{n,ci,j} = \beta \times x_{cen,ci} \tag{9}$$

where $\beta \in [0,1]$ implies the factor, which calculates the influence of $x_{center,ci}$ on $x_{new,cij}$. In fact, the novel individual $x_{n,cij}$ in Eq. (2) is produced by the data attained from every elephant in clan ci. $x_{cen,ci}$ represents the centre of clan ci, and dth dimension is determined as,

$$x_{center,ci,d} = \frac{1}{n_{ci}} \times \sum_{j=1}^{n_{Ci}} x_{ci,j,d} \tag{10}$$

where $1 \leq d \leq D$ signifies the dth dimension, and D refers to dimension. n_{ci} depicts the number of elephants present in clan ci. $x_{ci,j,d}$ is the dth dimension of elephant individual $x_{ci,j}$. The centre of clan ci, $x_{cen,ci}$ is determined by D calculations based on Eq. (10).

2.2.2.2 Separating operator
The male elephants often live separately by leaving their community, once they mature. This isolation process can be developed into separating operator, while resolving the optimization problem. The search ability for EHO model is enhanced by considering the scenario in which the elephant individuals that possess inferior fitness would be named as separating operator at every iteration as given in Eq. (11).

$$x_{w,ci} = x_{min} + (x_{max} - x_{min} + 1) \times random \tag{11}$$

where x_{max} and x_{min} are maximum and minimum limits of the position of an elephant. $x_{w,ci}$ is considered to be the worst elephant individual in clan ci. $rand \in [0,1]$ defines the type of stochastic distribution as well as uniform distribution within the range $[0,1]$ and is applied in the recent operation.

2.2.2.3 Elitism strategy
In line with the meta-heuristic approach, the type of elitism procedure is employed here to prevent the decomposition of optimal elephant individuals by clan updating and separating operators. Initially, the optimal elephant individual is protected and inferior ones are swapped by the protected optimal elephant individual in the searching procedure. It confirms that the advanced elephant population is not worst at every search process.

2.2.2.4 Improved EHO algorithm
The classical EHO algorithm fails to consider the optimal details which exist in the earlier group of distinct elephants for guiding present and future searching processes. It might result in slower convergences on complex optimization problems. So, certain information, utilized for the earlier individual elephants, was reprocessed with the intention of enhancing the searching capability of EHO algorithm.

The EHO algorithm is improved using new individual updating strategy. Theoretically, $k(k \geq 1)$ the earlier elephant individuals can be chosen. However, the selection of many individuals (> 3) makes the weight determination process difficult. So, the value of k is examined by using $k \in \{1,2,3\}$.

Consider X_r^t is the *ith* individual at round t whereas xr and f_r^t denote position and fitness values, correspondingly. At this point, r is the present round, $1 \leq i \leq N_P$ is an integer number, and NP is the population size. y_i^{r+1} is the individual produced by the traditional EHO algorithm and f_i^{r+1} is fitness [24]. The presented model is derived by the individual at $(r-2)th$, $(r-1)th$, rth, and $(r+1)th$ rounds. In case of k=1, the *ith* individual X_i^{r+1} could be produced as given

$$x_i^{r+1} = \theta y_i^{r+1} + \omega x_j^r \tag{12}$$

where x_j^t represents the individual's position $j (j \in \{1,2,\dots,N_P\})$ at round r, θ and ω are weighting factors filling Θ. It could be represented in Eq. (13).

$$\theta = rand, co = 1 - rand \tag{13}$$

where *rand* is an arbitrary number that lies under the uniform distribution in [0,1]. The individual j could be computed using two criteria

1. $\dot{L} = i$;
2. $\dot{L}$=rand1, where r_{and1} is random number, $1 \leq rand_1 \leq NP$.

The individuals produced from the second point attained high population diversity over the other one.

2.3 Validation phase

Once the face sketches of the images, captured by IoT devices, are generated by ODL-CNN model, a similarity measurement is done with professional sketch image. The measures used to determine the similarity are Structural Similarity (SSIM) and Peak Signal to Noise Ratio (PSNR). The images with maximum resemblance are considered as the suspect.

3 Performance validation

The proposed method was simulated using Python Programming language in a PC with configurations such as Intel i7-7500UCPU @ 2.70 GHz, 8 GB RAM, 64-bit OS and 1 TB HDD. To evaluate the efficiency of the projected technique, a detailed simulation analysis was performed on four benchmark datasets. A detailed quantitative and qualitative analyses, using SSIM and PSNR measure, were performed. The parameter settings of the proposed method are given as follows: Batch Size: 64, Epochs: 100, Learning rate: 0.001 and Activation function ReLU.

3.1 Dataset used

The analysis of the proposed technique was conducted over four datasets namely IIIT-D Sketch database [25], CUHK [26], AR [27] and CUFSF [28] dataset. The first IIIT-D dataset comprises of a group of three types of sketches such as Semi-forensic Sketch, Forensic Sketch database, and Viewed Sketch. The used database comprises of 238 sketch-digital image set of pairs. From specialized sketch artist, the sketches were derived from different sources and provided as digital images. From FG-NET aging database, 67 sketch-digital image set of pairs were considered. A total of 72 pairs of sketch-digital image from IIIT-D staff and student database was sourced and finally from Labeled Faces in Wild (LFW) database, 99 sketch-digital images were sourced. Then, CUHK and AR datasets hold a total of 188 and 123 color photo-sketch pairs correspondingly and the CUFSF dataset has a total of 1194 images with respective sketches. Some sample images are shown in the Fig. 5.

3.2 Measures

To evaluate the visual resemblance between the viewed sketch and output from the projected technique, SSIM and PSNR were employed. To quantitatively evaluate the quality of a visual image, SSIM was employed as it is the common metric in this regard. To find out the structural resemblance between two input images, it acts as a Human Visual System (HVS)-based metric. By employing loss combination in luminance, contrast, and correlation, the image distortion is designed by it. The rate of SSIM remains in the range of 0 towards 1 while the maximum resemblance is represented through higher value. It is demonstrated as

$$SSM = \frac{(2\mu_x\mu_x + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_x^2 + C_1)(\sigma_x^2 + \sigma_x^2 + C_2)} \tag{14}$$

PSNR is a measure used to determine the quality of original and reconstructed images. The value of PSNR should be high for better performance.

3.3 Results analysis

With a given sketch image database, Fig. 6 demonstrates the comparative outcomes of projected technique. The first two rows demonstrate the actual input and viewed sketch images correspondingly. At the same time, the fifth and fourth rows demonstrate the generated image through proposed techniques and corrected gamma image. It is proved from the figure that the derived sketch image through projected

Fig. 5 Sample test images

method is visually clear when compared with the viewed sketch drawn by artist.

Figures 7 and 8 show the comparison of results offered by ODL-CNN with existing models in terms of PSNR and SSIM. A set of methods namely Markov Random Field (MRF), Markov Weight Field (MWF), Sparse Representation-based Global Search method (SRGS), Semi-Coupled Dictionary Learning method (SCDL), CNN and Modified CNN (MCNN) [19] were used.

Based on the results on AR dataset in terms of PSNR, the SCDL model failed to yield good results and attained a minimal PSNR value of 17.18 dB. Next to that, the MWF and CNN models yielded slightly better and closer PSNR values of 18.74 dB and 18.23 dB, respectively. Along with that, the MRF and SRGS models attained even higher PSNR values of 19.84 dB and 19.13 dB, respectively. In line with this, a competitive PSNR value of 20.10 dB was achieved by M-CNN model, whereas the proposed ODL-CNN model offered superior performance with higher PSNR value of 21.98 dB. While computing the simulation outcome on AR dataset with respect to SSIM, both MRF and MWF models were unable to demonstrate the best results and reached a least SSIM of 0.62 dB. Then, the CNN and SCDL methods exhibited slightly better and identical SSIM values of 0.63 dB and 0.64 dB correspondingly. In line with this, the SRGS and M-CNN approaches accomplished better SSIM values of 0.65 dB. Along with that, the newly developed ODL-CNN model attained optimal performance with maximum SSIM

value of 0.69 dB. In the determination of final outcome of CUHK dataset by means of PSNR, the MWF method seems to be inappropriate to illustrate the maximum results and achieved lower PSNR value of 14.41 dB. Followed by this, the SRGS and MRF techniques arrived at moderate and nearby PSNR values of 14.79 dB and 15.07 dB respectively. Likewise, SCDL and CNN models reached the finest PSNR values of 15.14 dB and 15.64 dB correspondingly. Similarly, a competing PSNR value of 17.15 dB was attained by M-CNN model while the presented ODL-CNN model ended up with best performance by accomplishing high PSNR value of 18.64 dB.

By estimating the final outcome on CUHK dataset, with respect to SSIM, the MRF and SRGS models are found to be ineffective to produce optimal results since it obtained lower SSIM value of 0.58 dB. Next, the CNN, MWF and SCDL frameworks attained gradual and identical SSIM value of 0.59 dB. In line with this, the M-CNN model reached far better SSIM value of 0.60 dB. Likewise, the newly projected ODL-CNN model yielded best performance with good SSIM value of 0.63 dB. During the performance evaluation of CUFSF dataset in terms of PSNR, the SCDL models failed to display better results and accomplished a less PSNR value of 12.40 dB. Besides, the MWF and CNN models provided considerable and identical PSNR values of 14.34 dB and 14.36 dB correspondingly. In line with this, the MRF and SRGS models achieved better PSNR values of 15.72 dB and 15.34 dB respectively. Along with this, a competing

Fig. 6 Sample visualization of different sketches. **a** Input image, **b** Viewed sketch, **c** forensic image, **d** proposed sketch
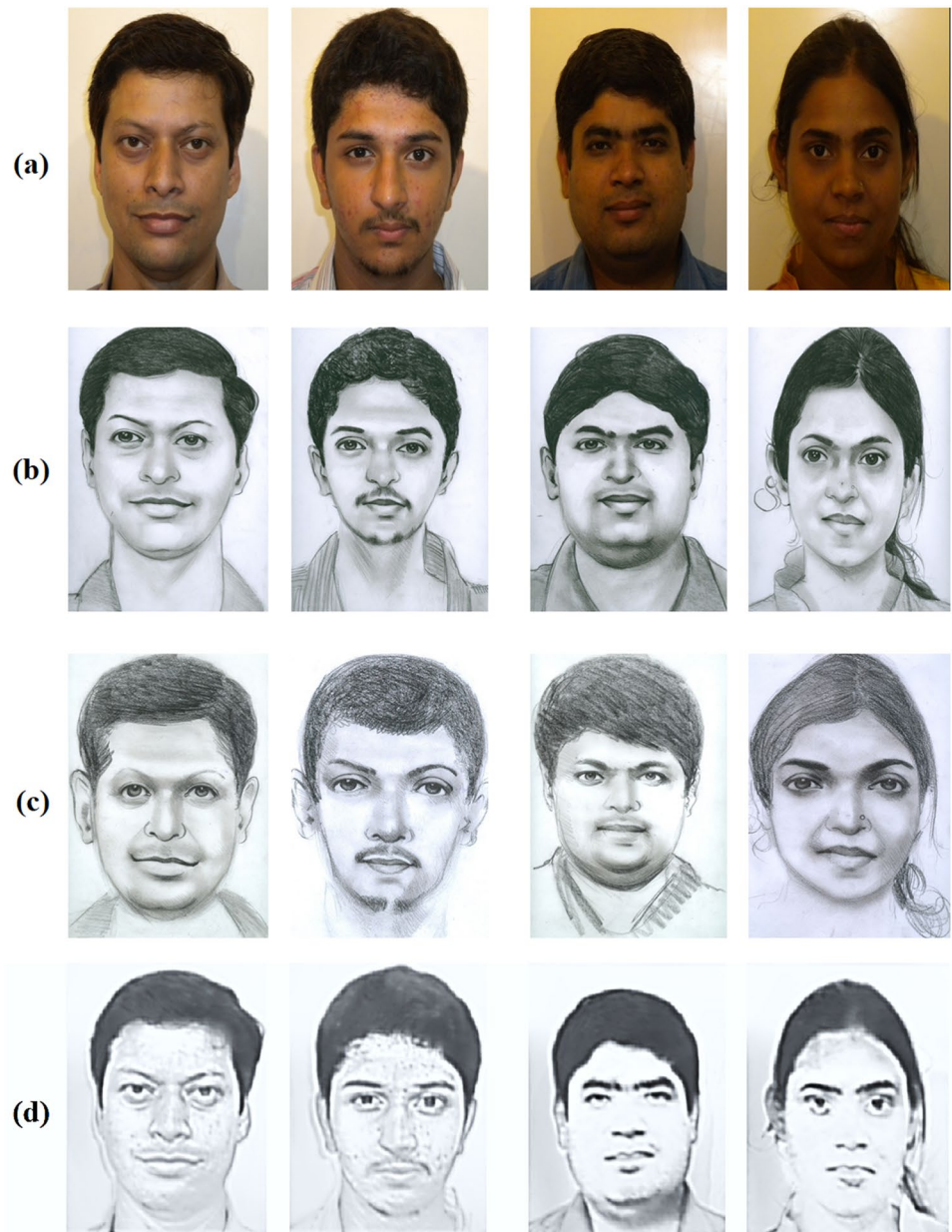
(a) Input image, (b) Viewed Sketch, (c) Forensic image, (d) Proposed Sketch

PSNR value of 17.15 dB was produced by M-CNN model and the deployed ODL-CNN model offered qualified performance with the maximum PSNR value of 18.07 dB.

By computing the results on CUFSF dataset, with respect to SSIM, the SCDL model is found to be inappropriate to produce better results since it reached a minimum SSIM value of 0.37 dB. Then, the CNN and MRF models ended with moderate and closer SSIM value of 0.38 dB. Similarly, the MWF and SRGS models accomplished even better SSIM values of 0.39 dB and 0.40 dB correspondingly. Likewise, a competitive SSIM value of 0.46 dB was achieved by M-CNN model while the proposed ODL-CNN model performed extremely well with the best SSIM value of 0.54 dB. When determining the results on IIIT dataset, by means of PSNR, the MWF model failed to produce the best results and reached a lower PSNR value of 17.20 dB. Following this, both SRGS and SCDL models concluded with manageable and closer PSNR values of 18.46 dB and 18.33 dB correspondingly. On the same way, the MRF and CNN models achieved even better PSNR values of 19.26 dB and 19.62 dB respectively. Similarly, a competitive PSNR value of 20.98 dB was yielded by M-CNN model while the projected ODL-CNN model depicted qualified performance with maximum PSNR value of 21.74 dB. When calculating the outcome
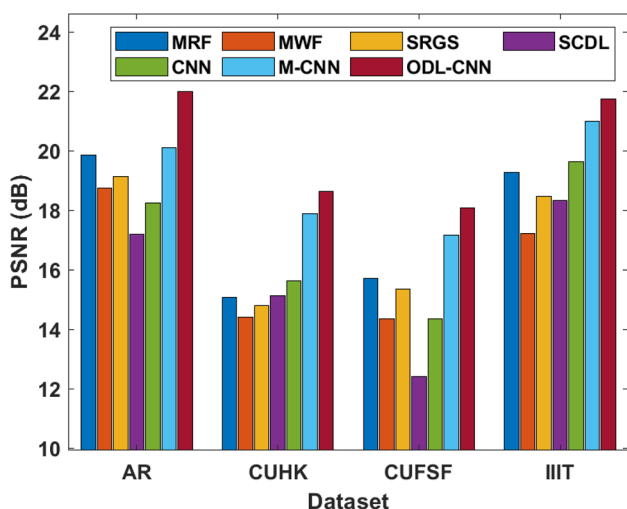
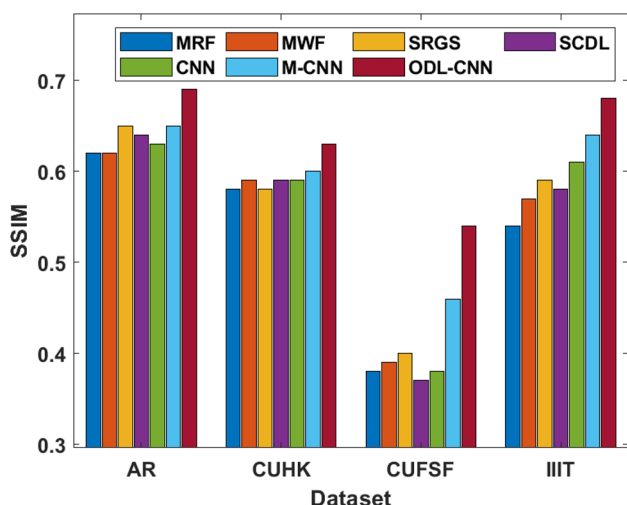**Fig. 7** PSNR analysis of ODL-CNN with different models

Fig. 8 SSIM analysis of ODL-CNN with different models

on IIIT dataset in terms of SSIM, the MRF model failed to provide good results and achieved a lower SSIM value of 0.54 dB. Then, the MWF and SCDL methods yielded considerable and closer SSIM values of 0.57 dB and 0.58 dB correspondingly. In line with this, the SRGS and CNN models reached even greater SSIM values of 0.59 dB and 0.61 dB correspondingly. In line with this, the competitive SSIM value of 0.64 dB was attained by M-CNN model while

the deployed ODL-CNN model yielded better performance with maximum SSIM value of 0.68 dB.

Table 1; Figs. 9 and 10 show the average analyses of the proposed ODL-CNN model with existing ones in terms of PSNR and SSIM. The proposed ODL-CNN model showcased effective results with maximum PSNR value of 20.11dB and SSIM of 0.64.

A detailed accuracy analysis of ODL-CNN model with existing models is shown in Table 2; Fig. 11. Also, Fig. 12 shows the average accuracy analysis of different models on the applied four datasets. It is depicted that both MWF and SCDL models attained ineffective and closer average accuracy values of 69.14% and 69.08%, respectively. Simultaneously, the MRF model produced slightly higher average accuracy of 70.26%. In line with this, the SRGS model outperformed previous methods with average accuracy of 71.14%. Concurrently, the CNN model produced moderate results with average accuracy of 78.64%. Then, the M-CNN model outperformed all the other methods except ODL-CNN with an average accuracy of 85.61%. However, the proposed ODL-CNN model performed extraordinarily with maximum average accuracy of 90.10%.

Based on the above presented graphs, it is evident that the proposed ODL-CNN model outperformed all the existing methods on four applied dataset. Therefore, it can be found as an appropriate tool for FSS and it helps in suspect identification process effectively.

4 Conclusion

In this paper, a new FSS approach is proposed using ODL-CNN model where the parameter optimization of DL-CNN was performed using IEHO algorithm. Before transforming the digital images to face sketches, gamma correction was employed to improve the contrast level of the input image. The ODL-CNN model drawn the sketches of the input images which underwent similarity assessment with professional sketch for proper suspect identification. The proposed technique was assessed using four datasets. To evaluate the visual resemblance between viewed sketch and output from the projected technique, SSIM and PSNR were employed. The detailed simulation analysis pointed out the effective performance of ODL-CNN model with maximum average PSNR of 20.11 dB, average SSIM of 0.64 and average accuracy of 90.10%. The presented model exhibited

Table 1 Average PSNR and SSIM analyses of existing models with ODL-CNN method

Average	Methods						
	MRF	MWF	SRGS	SCDL	CNN	M-CNN	ODL-CNN
PSNR	17.47	16.17	16.93	15.77	16.96	19.03	20.11
SSIM	0.53	0.54	0.56	0.55	0.55	0.59	0.64

Fig. 9 Average PSNR analysis of ODL-CNN with different models

Fig. 11 Accuracy analysis of ODL-CNN with different models
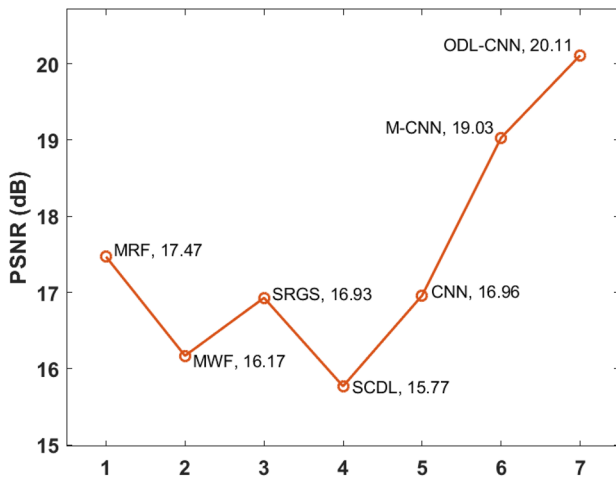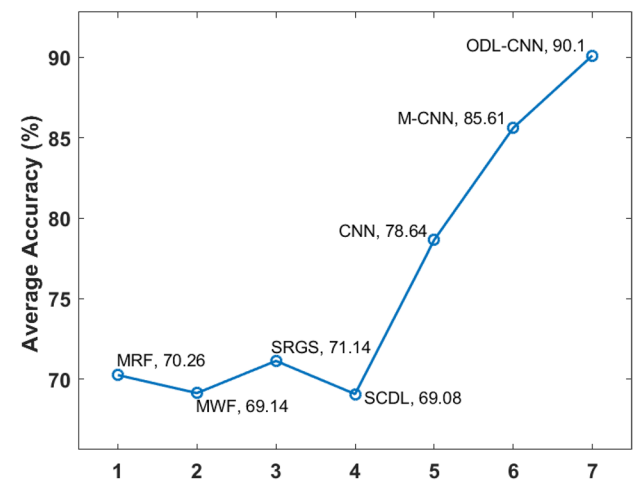

Fig. 10 Average SSIM analysis of ODL-CNN with different models

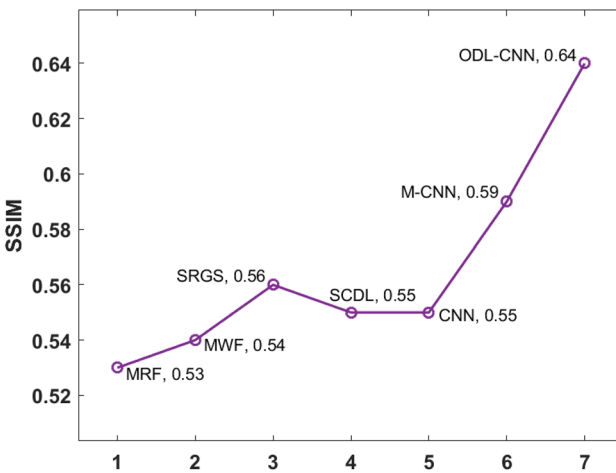**Fig. 12** Average accuracy analysis of ODL-CNN with different models

Table 2 Accuracy analysis of existing with proposed ODL-CNN method

Methods	AR	CUHK	CUFSF	IIIT	Average
MRF	68.00	71.30	70.38	71.34	70.26
MWF	67.87	70.84	69.53	68.30	69.14
SRGS	68.23	72.45	71.46	72.40	71.14
SCDL	64.20	69.85	70.53	71.75	69.08
CNN	76.38	78.53	79.42	80.21	78.64
M-CNN	85.28	86.04	84.62	86.48	85.61
ODL-CNN	89.27	91.35	87.31	92.45	90.10

better performance over other methods in all the applied test images. In future, the proposed model can be implemented in real time scenario to investigate surveillance videos in railway stations, airports and other public places.
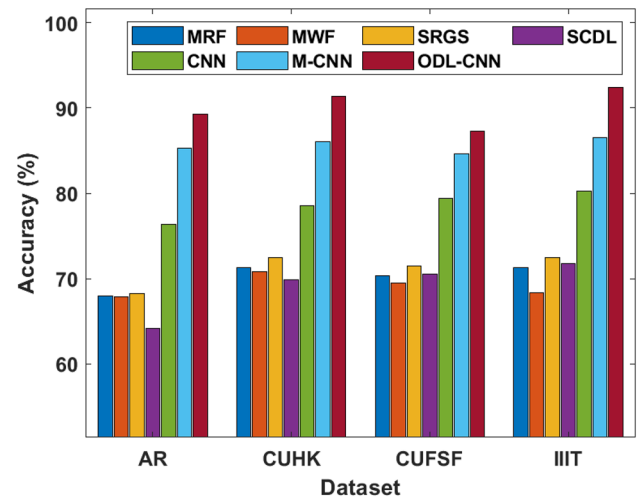
Compliance with ethical standards

Conflict of interest All the authors of the manuscript declared that there are no potential conflicts of interest.

References

1. Reyes A, Brittson R, O'Shea K, Steele J (2011) Cybercrime investigations: bridging the gaps between security professionals, law enforcement, and prosecutors. Elsevier, Amsterdam

2. Brown CS (2015) Investigating and prosecuting cyber crime: forensic dependencies and barriers to justice. Int J Cyber Criminol 9(1):55

3. Klum S, Han H, Jain AK, Klare B, (2013) Sketch based face recognition: forensic vs. composite sketches. In: 2013 international conference on biometrics (ICB), IEEE, pp 1–8

4. Wang N, Gao X, Sun L, Li J (2017) Bayesian face sketch synthesis. IEEE Trans Image Process 26(3):1264–1274

5. Wang N, Gao X, Sun L, Li J (2018) Anchored neighborhood index for face sketch synthesis. IEEE Trans Circuits Syst Video Technol 28(9):2154-2163

6. Wang N, Gao X, Li J (2018) Random sampling for fast face sketch synthesis. Pattern Recognit 76:215–227

7. Wang N, Zhu M, Li J, Song B, Li Z (2017) Data-driven vs. model-driven: fast face sketch synthesis. Neurocomputing 257:214–221

8. Hwang B-W, Lee S-W (2003) Reconstruction of partially damaged face images based on a morphable face model. IEEE Trans Pattern Anal Mach Intell 25(3):365–372

9. Mo Z, Lewis J, Neumann U (2004) Face inpainting with local linear representations. In: British Machine Vision Conference (BMVC), pp. 347–356. https://doi.org/10.5244/C.18.37

10. Wang Z-M, Tao J-H (2007) Reconstruction of partially occluded face by fast recursive PCA. In: Proceedings of the International Conference on Computational Intelligence and Security Workshops (2007) CISW 2007, IEEE, pp 304–307

11. Burgos-Artizzu XP, Perona P, Dollár P (2013) Robust face landmark estimation under occlusion. In: Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV), IEEE, pp 1513–1520

12. Mairal J, Bach F, Ponce J, Sapiro G (2009) Online dictionary learning for sparse coding. In: Proceedings of the Twenty-sixth Annual International Conference on Machine Learning, ACM, pp 689–696

13. Li Y, Liu S, Yang J, Yang M-H (2017) Generative face completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1, p 6

14. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Advances in neural information processing systems, pp 2672–2680

15. Mao X, Wang S, Zheng L, Huang Q (2018) Semantic invariant cross-domain image generation with generative adversarial networks. Neurocomputing 293:55–63

16. Wang T-C, Liu M-Y, Zhu J-Y, Tao A, Kautz J, Catanzaro B, High-resolution image synthesis and semantic manipulation with conditional GANS, arXiv preprint arXiv:1711.11585

17. Lu D, Chen Z, Wu QJ, Zhang X (2019) FCN based preprocessing for exemplar-based face sketch synthesis. Neurocomputing 365:113–124

18. Ye L, Zhang B, Yang M, Lian W (2019) Triple-translation GAN with multi-layer sparse representation for face image synthesis. Neurocomputing 358:294–308

19. Licheng J, Zhang S, Li L, Liu F, Ma W (2018) A modified convolutional neural network for face sketch synthesis. Pattern Recognit 76:125–136

20. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z et al. (2016) photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp, 21–26

21. Lin G, Wu Q, Qiu L, Huang X (2018) Image super-resolution using a dilated convolutional neural network. Neurocomputing 275:1219–1230

22. Zhou F, Li X, Li Z (2018) High-frequency details enhancing densenet for super-resolution. Neurocomputing 290:34–42

23. Wang GG, Deb S, Gao XZ, Coelho LDS (2016) A new metaheuristic optimisation algorithm motivated by elephant herding behaviour. Int J Bio-Inspired Comput 8(6):394–409

24. Li J, Guo L, Li Y, Liu C (2019) Enhancing elephant herding optimization with novel individual updating strategies for large-scale optimization problems. Mathematics 7(5):395

25. Sketch IIIT-D, Database, http://www.iab-rubric.org/resources/sketchDatabase.html

26. Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell 31(11):1955–1967

27. Mart´ınez A, Benavente R (1998) The ar face database. In: Tech. Rep. 24, Computer 565 Vision Center, Bellatera

28. Zhang W, Wang X, Tang X (2011) Coupled information-theoretic encoding for face photo-sketch recognition. In: Proceedings of the 24th IEEE Conference, Comput. Vis. and Pattern Recognit. (CVPR), pp, 513–520. https://doi.org/10.1109/CVPR.2011.5995324