Improved accuracy and precision in IECNN face recognition model in similar face datasets (SFD)

Anmoldeep Singh[1]

[1]singh.anmoldeep1997@gmail.com

Abstract— **In the field of face recognition, similar face recognition is difficult due to the high degree of similarity of the face structure. The following two factors are needed to make progress in this field: (i) the availability of large scale similar face training datasets, and (ii) a fine-grained face recognition method. With the above factors fulfilled, we make two contributions. First, we show how a large scale similar face dataset (SFD) can be assembled by a combination of automation and human in the loop, and divide the dataset into five grades according to different degrees of similarity. In this research we have optimized IECNN with BSO and results shows an improvement.**

Keywords-Face Detection, IECNN, BSO

I. INTRODUCTION

A facial recognition system is a technology capable of matching a human face from a digital image or a video frame against a database of faces, typically employed to authenticate users through ID verification services, works by pinpointing and measuring facial features from a given image.

While humans can recognize faces without much effort, facial recognition is a challenging pattern recognition problem in computing. Facial recognition systems attempt to identify a human face, which is three-dimensional and changes in appearance with lighting and facial expression, based on its two-dimensional image. To accomplish this computational task, facial recognition systems perform four steps.

Traditional: Some face recognition algorithms identify facial features by extracting landmarks, or features, from an image of the subject's face. For example, an algorithm may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features.

Human Identification at a Distance: To enable human identification at a distance (HID) low-resolution images of faces are enhanced using face hallucination. In CCTV imagery faces are often very small. But because facial recognition algorithms that identify and plot facial features require high resolution images, resolution enhancement techniques have been developed to enable facial recognition systems to work with imagery that has been captured in environments with a high signal-to-noise ratio.

3-Dimensional Recognition: Three-dimensional face recognition technique uses 3D sensors to capture information about the shape of a face. This information is then used to identify distinctive features on the surface of a face, such as the contour of the eye sockets, nose, and chin.

Thermal cameras: A different form of taking input data for face recognition is by using thermal cameras, by this procedure the cameras will only detect the shape of the head and it will ignore the subject accessories such as glasses, hats, or makeup.

Deep Learning: Convolutional Neural Networks allow us to extract a wide range of features from images. The key here is to get a deep neural network to produce a bunch of numbers that describe a face (known as face encodings).

II. CONVOLUTION NEURAL NETWORK

In deep learning, a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as shift invariant or space invariant artificial neural networks (SIANN), based on their shared-weights architecture and translation invariance characteristics.

CNNs are regularized versions of multilayer perceptrons. Multilayer perceptrons usually mean fully connected networks, that is, each neuron in one layer is connected to all neurons in the next layer. The "fully-connectedness" of these networks makes them prone to overfitting data. Typical ways of regularization include adding some form of magnitude measurement of weights to the loss function. CNNs take a different approach towards regularization: they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are on the lower extreme.

CNNs for segmentation can be categorized based on the dimension of convolutional kernel that is utilized. 2D CNNs use 2D convolutional kernels to predict the segmentation map for a single slice. Segmentation maps are predicted for a full volume by taking predictions one slice at a time. The 2D convolutional kernels are able to leverage context across the height and width of the slice to make predictions. However, because 2D CNNs take a single slice as input, they inherently fail to leverage context from adjacent slices. Voxel information from adjacent slices may be useful for the prediction of segmentation maps.

3D CNNs address this issue by using 3D convolutional kernels to make segmentation predictions for a volumetric patch of a scan. The ability to leverage interslice context can lead to improved performance but comes with a computational cost as a result of the increased number of parameters used by these CNNs.

III. IE CNN

In the field of face recognition, similar face recognition is difficult due to the high degree of similarity of the face structure. The following two factors are needed to make progress in this field:

(i) the availability of large scale similar face training datasets, and

(ii) a fine-grained face recognition method. With the above factors fulfilled, we make two contributions.

First, large scale similar face dataset (SFD) can be assembled by a combination of automation and human in the loop, and divide the dataset into five grades according to different degrees of similarity. Second, a new fine-grained face feature extraction method is proposed to solve this problem using the attention mechanism which combines the Internal Features and External Features.

The algorithm is divided into two branches, one called IE-CNN branch and the other called trunk branch. Given the trunk branch output T (x) with input x, the IE-CNN branch is to learn the same size mask F(x) that softly weight output features T (x). The output of algorithm H(x) is: $H_{i,c}(x) = (1 + F_{i,c}(x)) \times T_{i,c}(x)$ (1) where i ranges over all spatial positions and $c \in \{1, \ldots, C\}$ is the index of the channel.

The hyper-parameter p denotes the number of Basic Units in the IE-CNN branch. q denotes the number of Basic Units in trunk branch. In our experiments, we use the following hyper-parameters setting:p = 3, q = 2. The numbers of channels in the IE-CNN and corresponding trunk branches are the same.

Drawbacks: The biggest drawback after adding internal and external features to assist the fine-grained face recognition is that it will bring additional parameters to the model and increase the difficulty of training.

In the second one, efficiency of this technique can be improved.

IV. METHODOLOGY

Image Acquisition: First considered that the MRI scan images of a given patient are either color, Gray-scale or intensity images herein are displayed with a default size of 220×220. If it is color image, a Gray-scale converted image is defined by using a large matrix whose entries are numerical values between 0 and 255, where 0 corresponds to black and 255 white for instance. Then the brain tumor detection of a given patient consist of two main stages namely, image segmentation and edge detection.

Pre-processing stage: Pre-processing stage consists of Noise removal this can be done by using various spatial filters linear or nonlinear filters (Median filter). Other artifacts like text removed by some morphological operations. RGB to grey conversion and reshaping also takes place here. It includes median filter for noise removal. The possibilities of arrival of noise in modern MRI scan are very less. It may arrived due to thermal Effect.

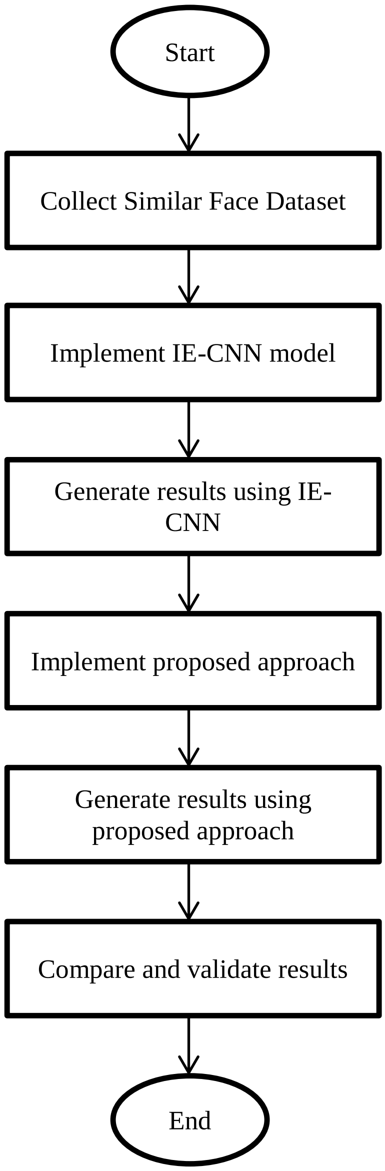

Image Smoothing: It is the action of simplifying an image while preserving important information. The goal is to reduce noise or useless details without introducing too much distortion so as to simplify subsequent analysis. Image Registration: Image registration is the process of bringing two or more images into spatial correspondence (aligning them). In the context of medical imaging, image registration allows for the concurrent use of images taken with different modalities (e.g. MRI and CT), at different times or with different patient positions. In surgery, for example, images are acquired before (preoperative), as well as during (intra-operative) surgery. Because of time constraints, the real-time intraoperative images have a lower resolution than the pre-operative images obtained before surgery. Moreover, deformations which occur naturally during surgery make it difficult to relate the highresolution pre-operative image to the lowresolution intra-operative anatomy of the patient. Image registration attempts to help the surgeon relate the two sets of images.

Image Segmentation: The segmentation is the most important stage for analysing image properly since it affects the accuracy of the subsequent steps. However, proper segmentation is difficult because of the great verities of the lesion shapes, sizes, and colors along with different skin types and textures. In addition, some lesions have irregular boundaries and in some cases there is smooth transition between the lesion and the skin. To address this problem, several algorithms have been proposed. They can be broadly classified as thresholding, edge-based or region-based, supervised and unsupervised classification techniques

Threshold segmentation

Water shed segmentation

Gradient Vector Flow (GVF)

K-mean Clustering

Fuzzy C-means Clustering

Morphological Operations: after segmentation morphological processing is applied to remove unwanted part. It consists of image opening, image closing, dilation, erosion operations. At the end the decision has taken weather that MRI image consists of any tumor or not and weather it normal or abnormal.

V. RELATED WORK

Song, et al. [1] show how a large scale similar face dataset (SFD) can be assembled by a combination of automation and human in the loop, and divide the dataset into five grades according to different degrees of similarity. Second, a new fine-grained face feature extraction method is proposed to solve this problem using the attention mechanism which combines the Internal Features and External Features.

Li, et al. [2] introduced the related researches of face recognition from different perspectives and described the development stages and the related technologies of face recognition.

An, et al. [3] proposed a new face alignment method for pose-invariant face recognition, called adaptive pose alignment (APA), which can greatly reduce the intra-class difference and correct the noise caused by the traditional method in the alignment process, especially in unconstrained settings.

Zhang, et al. [4] developed an improved fusion method for calculation of optimal weight value for multiplication fusion applicable to sparse representation. The fusion scheme not only is easy to use but also does not need to be artificially set weights.

Zhou, et al. [5] assumed a new deep neural architecture search pipeline combined with NAS technology and reinforcement learning strategy into face recognition. We quote the framework of NAS, which trains the child and controller networks alternately and optimize NAS by incorporating evaluation latency into rewards of reinforcement learning and utilize the policy gradient algorithm to search the architecture automatically with the cross-entropy loss.

Min, et al. [6] aimed a scheme combined transfer learning and sample expansion in feature space. First, it uses transfer learning by training a deep convolutional neural network on a common multi-sample face dataset and then applies the well-trained model to a target data set. Second, it proposes a sample expansion method in feature space called k class feature transfer (KCFT) to enrich intra-class variation information for a single-sample face feature.

Yang, et al. [7] aimed to design a face recognition attendance system based on real-time video processing. It sets four directions to consider the problems: the accuracy rate of the face recognition system in the actual check-in, the stability of the face recognition attendance system with real-time video processing, the truancy rate of the face recognition attendance system with real-time video processing and the interface settings of the face recognition attendance system using real-time video processing.

Liu, et al. [8] developed Cross-Pose Generative Adversarial Networks(CP-GAN) to frontalize the profile face with unaltered identity by learning the mapping between the profile and frontal faces in image space. The generator is an encoder-decoder U-net, and generate frontal face image by fusing multiple profile images to achieve a better performance in PIFR. The siamese discriminative network attends to extract the deep representations of the generated frontal face and the ground truth without introducing extra networks in verification and recognition.

Hsu, et al. [9] proposed a framework for dual-view normalization that generates a frontal pose and an additional yaw-45◦ pose to an input face of an arbitrary pose. The proposed Dual-View Normalization (DVN) framework is designed to learn the transformation from a source set to two normal sets. The source set contains faces collected in the wild and covers a wide scope of variables.

Zhang, et al. [10] introduced an algorithm to avoid the complex process of explicit feature extraction in traditional facial expression recognition, a face expression recognition method based on a convolutional neural network (CNN) and an image edge detection. Firstly, the facial expression image is normalized, and the edge of each layer of the image is extracted in the convolution process. The extracted edge information is superimposed on each feature image to preserve the edge structure information of the texture image. Then, the dimensionality reduction of the extracted implicit features is processed by the maximum pooling method.

Year	Author	Tools	Algorithm and Technique
2020	Song, et al. [1]	Matlab	IE- Convolutional Neural Network Model Deep Convolutional Neural Network
2019	An, et al. [3]	Open CV	APA: Adaptive Pose Alignment
2019	Zhang, et al. [4]	Open CV	Sparse Representation
2020	Zhou, et	AutoML	Neural Architecture

	al. [5]		Search (NAS) Deep Neural Architecture Search
2019	Min, et al. [6]	Softmax Classifier	k class feature transfer (KCFT) Histogram Of Oriented Gradient (HOG) Support Vector Machine (SVM)
2020	Yang, et al. [7]	OpenCV	Linear Discriminate Analysis (LDA) Support Vector Machine (SVM)
2020	Liu, et al. [8]	ADABOOST	Cross-Pose Generative Adversarial Networks (CP-GAN) Convolutional Neural Network
2020	Hsu, et al. [9]	MATLAB	Dual-View Normalization (DVN)
2019	Zhang, et al. [10]	Keras framework (Python)	Convolutional Neural Network Image Edge Computing
2020	Rong, et al. [11]	MultiPIE	Feature-Improving Generative Adversarial Networks (GANs)
2020	Liu, et al. [12]	ResNet26, ResNet50	Patch-Attention Generative Adversarial Network (PA-GAN) Patch-Attention Generative Adversarial Network (PA-DAN)
2020	Li, et al. [13]	PCA	Principal Component Analysis (PCA) Linear Discriminate Analysis (LDA)
2020	Zhu, et al. [14]	Resnet	RELU function NAS (Neural Architecture Structure) algorithms
2020	Min, et al. [15]	OpenCV	k class feature transfer (KCFT) softmax classifier

VI. BSO

The bees algorithm mimics the foraging strategy of honey bees to look for the best solution to an optimization problem. Each candidate solution is thought of as a food source (flower), and a population (colony) of n agents (bees) is used to search the solution space. Each time an artificial bee visits a flower (lands on a solution), it evaluates its profitability (fitness).

The bees algorithm consists of an initialisation procedure and a main search cycle which is iterated for a given number T of times, or until a solution of acceptable fitness is found. Each search cycle is composed of five procedures: recruitment, local search, neighbourhood shrinking, site abandonment, and global search.

Pseudocode

```
1 for i=1,…,ns
    i  scout[i]=Initialise_scout()
    ii flower_patch[i]=Initialise_flower_patch(scout[i])
2 do until stopping_condition=TRUE
    i   Recruitment()
    ii  for i =1,...,na
          1 flower_patch[i]=Local_search(flower_patch[i])
          2
flower_patch[i]=Site_abandonment(flower_patch[i])
          3
flower_patch[i]=Neighbourhood_shrinking(flower_patch[i])

    iii for i = nb,...,ns
          1 flower_patch[i]=Global_search(flower_patch[i])}
```

In the initialisation routine ns scout bees are randomly placed in the search space, and evaluate the fitness of the solutions where they land. For each solution, a neighbourhood (called flower patch) is delimited.

In the recruitment procedure, the scouts that visited the nb ⩽ns fittest solutions (best sites) perform the waggle dance. That is, they recruit foragers to search further the neighbourhoods of the most promising solutions. The scouts that located the very best ne⩽nb solutions (elite sites) recruit nre foragers each, whilst the remaining nb-ne scouts recruit nrb ⩽ nre foragers each. Thus, the number of foragers recruited depends on the profitability of the food source.

In the local search procedure, the recruited foragers are randomly scattered within the flower patches enclosing the solutions visited by the scouts (local exploitation). If any of the foragers in a flower patch lands on a solution of higher fitness than the solution visited by the scout, that forager becomes the new scout. If no forager finds a solution of higher fitness, the size of the flower patch is shrunk (neighbourhood shrinking procedure). Usually, flower patches are initially defined over a large area, and their size is gradually shrunk by the neighbourhood shrinking procedure. As a result, the scope of the local exploration is progressively focused on the area immediately close to the local fitness best. If no improvement in fitness is recorded in a given flower patch for a pre-set number of search cycles, the local maximum of fitness is considered found, the patch is abandoned (site abandonment), and a new scout is randomly generated.

VII. EXPERIMENT EVALUATION INDEX

There are two types of evaluation methods for our experimental results, one is judged by the currently widely used comparison method. The formulas are given below.

$$TPR = TP / (TP + FN) \quad (1)$$

$$ACC = (TP + TN) / TP + FP + TN + FN \quad (2)$$

The second custom evaluation method called Top1&Top5 precision was also used in our experiments. Traversing the N samples, if the total number of Top1 hits is N1 and the total number of Top5 hits is N5, precision is calculated as follows:

$$Precision_{top1} = N1 / N \quad (3)$$

$$Precision_{top5} = N5 / N \quad (4)$$

Below are screens for our results using various images

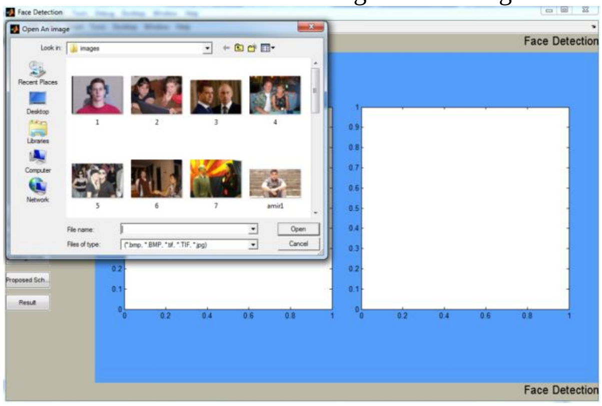

Fig 1: Image upload

Fig 1 is describing about the image that is to be uploaded for face detection and matching.

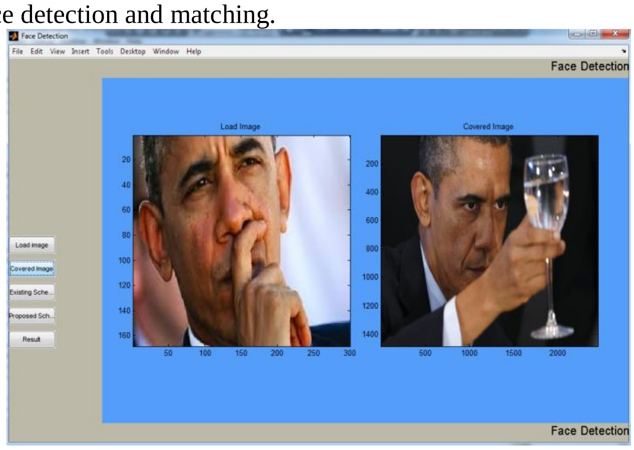

Fig 2: Face Detection and Matching

In fig 2 images are uploaded and then faces are detected and then matched if the images are same or not using proposed approach.

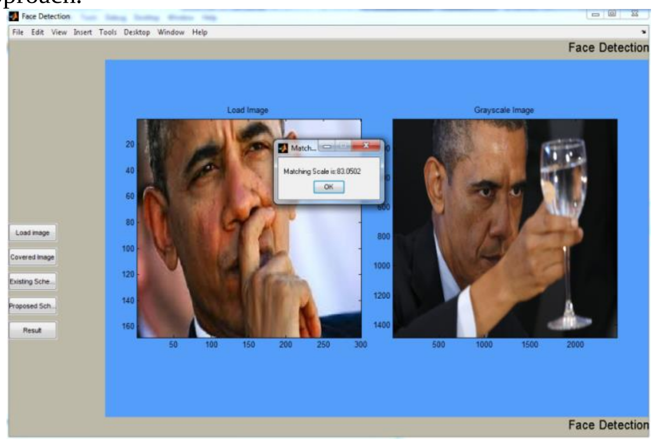

Fig 3: Matching Scale Value

Fig 3 is matching scale value that is calculated by comparing two images by using proposed approach.

VIII. RESULTS

In the proposed approach 165 images taken from which 100 images are considered for testing purpose that are used for classification as normal and abnormal. To classify images into normal and abnormal 65 images are used as training set.

Table 1: Testing using True positive, true negative, false positive and false negative using Proposed Technique

N=100	**Normal**	**Abnormal**	
Normal	TN = 10	FP = 1	11
Abnormal	FN = 2	TP = 87	89
	12	88	

Table 2: Testing using True positive, true negative, false positive and false negative using Existing Technique

N=100	**Normal**	**Abnormal**	
Normal	TN = 8	FP = 3	11
Abnormal	FN = 4	TP = 85	89
	12	88	

Table 3: Comparative Study of Existing and Proposed technique

Techniques / **Metrics**	**Existing**	**Enhanced**
Specificity	0.955	0.977
Sensitivity	0.727	0.909
Accuracy	93%	97%

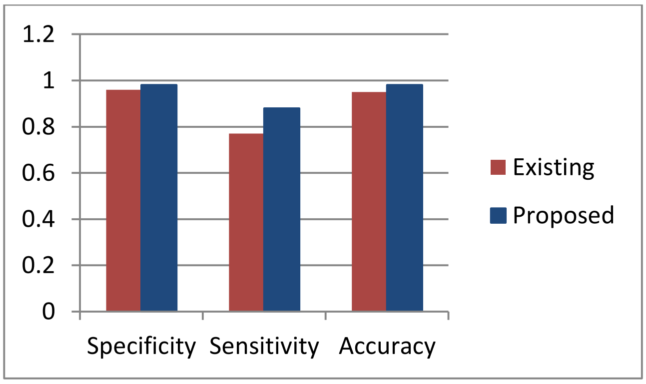

Fig 4: Comparative Study for Existing and Proposed

This section had described about the results that are generated using the proposed approach and it is validated by comparing with the existing approach. From this it is clear that proposed approach is far better than that of existing approach.

IX. Conclusion

This research collected a new similar face datasets (SFD) and introduced a robust face recognition method combining internal features and external features of face. By combining the external feature and the internal feature, our method not only ensures that face matching accuracy is improved, but also achieves fine-grained recognition effect. Our method improves the face features extraction effectiveness of the traditional face recognition model while maintaining the advantages of the original deep CNN model, which tackles the recognition accuracy problem caused by the traditional methods when there are very similar face images here. The higher the similarity between the two face images is, the lower the recognition accuracy is, which is inevitable and will be further studied in the future. There is approximately 8% improvement in the matching scale in proposed technique.

Future Scope: The biggest drawback after adding internal and external features to assist the fine-grained face recognition is that it will bring additional parameters to the model and increase the difficulty of training. Therefore, the method of training model proposed in this paper is an end-to-end training mode, which greatly reduces the difficulty of model training.

X. References

[1]. Song, A.P., Hu, Q., Ding, X., Di, X. and Song, Z., "Similar Face Recognition Using the IE-CNN Model", IEEE, Vol. 8, 2020, pp. 45244-45253

[2]. Li, L., Mu, X., Li, S. and Peng, H., "A Review of Face Recognition Technology", IEEE, Vol. 8, 2020, pp. 139110- 139120

[3]. An, Z., Deng, W., Hu, J., Zhong, Y. and Zhao, Y., "APA: Adaptive Pose Alignment for Pose-Invariant Face Recognition", Vol. 7, 2019, pp. 14653- 14670

[4]. Zhang, Y., Wang, Q., Xiao, L. and Cui, Z., "An Improved Two-Step Face Recognition Algorithm Based on Sparse Representation", IEEE, Vol. 7, 2019, pp. 131830-131838

[5]. Zhu, N., Yu, Z. and Kou, C., "A New Deep Neural Architecture Search Pipeline for Face Recognition", IEEE, Vol. 8, 2020, pp. 91303-91310

[6]. Min, R., Xu, S. and Cui, Z., "Single-Sample Face Recognition Based on Feature Expansion", IEEE, Vol. 7, 2019, pp. 45219-45229

[7]. Yang, H. and Han, X., "Face Recognition Attendance System Based on Real-Time Video Processing", IEEE, Vol. 8, 2020, pp. 159143-159150

[8]. Liu, J., Li, Q., Liu, M. and Wei, T., "CP-GAN: A Cross-Pose Profile Face Frontalization Boosting Pose-Invariant Face Recognition", Vol. 8, 2020, pp. 198659- 198667

[9]. Hsu, G.E. and Tang, C.H., "Dual-View Normalization for Face Recognition", IEEE, 2020, Vol. 8, pp. 147765- 147775

[10]. Zhang, H., Jolfaei, A. and Alazab, M., "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing", IEEE, Vol. 7, 2019, pp. 159081- 159089

[11]. Rong, C., Zhang, X. and Lin, Y., "Feature-Improving Generative Adversarial Network for Face Frontalization", Vol. 8, 2020, pp. 68842- 68851

[12]. Liu, M., Liu, J., Zhang, P. and Li, Q., "PA-GAN: A Patch-Attention Based Aggregation Network for Face Recognition in Surveillance", IEEE, Vol. 8, pp. 152780- 152789

[13]. Li, L., Mu, X., Li, S. and Peng, H., "A Review of Face Recognition Technology", IEEE, Vol. 8, 2020, pp. 139110- 139120

[14]. Zhu, N., Yu, Z. and Kou, C., "A New Deep Neural Architecture Search Pipeline for Face Recognition", IEEE, Vol. 8, 2020, pp. 91303- 91310

[15]. Min, R., Xu, S. and Cui, Z., "Single-Sample Face Recognition Based on Feature Expansion", Vol. 7, 2019, pp. 45219- 45229

[16]. Teoh, K.H.; Ismail, S.Z.M.; Husain, R.; Isa, M.N.M.; Basir, M.S.S.M.; "Face Recognition and Identification using Deep Learning Approach", Journal of Physics: Conference Series, Vol. 1755, 2021, pp. 1-9

[17]. Sharma, S.; Bhatt, M.; Sharma, P.; "Face Recognition System Using Machine Learning Algorithm", 5th International Conference on Communication and Electronics Systems (ICCES), IEEE, Coimbatore, 2020

[18]. Suganthi, S.T.; Ayoobkhan, M.U.A.; Kumar, K.; Bacanin, N.; Venkatachalam, K.; Stepan, H.; Pavel, T.; "Deep learning model for deep fake face recognition and detection", PeerJ Computer Science, 2022, pp. 1-20

[19]. Velan, E.; Fontani, M.; Carrato, S.; Jerian, M.; "Does Deep Learning-Based Super-Resolution Help Humans With Face Recognition?", Front. Signal Process, https://doi.org/10.3389/frsip.2022.854737, Vol. 2, 2022, pp. 1-10

[20]. Tabassum, F.; Islam, M.I.; Khan, R.T.; Amin, M.R.; "Human face recognition with combination of DWT and machine learning"