International Journal of Thin Films Science and Technology

http://dx.doi.org/10.18576/ijtfst/110109

Real Time Face Authentication System Using Stacked Deep Auto Encoder for Facial Reconstruction

Showkat A. Dar and S. Palanivel*

Department of Computer Science and Engineering, Annamalai University, Annamalainagar, Chidambaram, India

Received: 2 Jun. 2021, Revised: 2 Oct. 2021, Accepted: 3 Nov. 2021

Published online: 1 Jan. 2022

Abstract: Any human being has unique biological traits biological characteristics which can be studied using Biometrics which encompasses an individual's characteristics like DNA, face, finger prints, voice, signatures etc. Human faces as an element of authentication are being increasingly used where biometrics add value in terms of quantifying an individual's natural data. Facial authentications validate personal identities based on facial images with 1-1 matches. These kinds of authenticating applications have been applied in a variety of areas including banking applications and personal mobile devices. RTFAs (Real Time Face Authentications) based systems are a necessity for ATMs (Automated Teller Machines) in banking for enhanced security. Several machine learning methods have been introduced RTFA based systems. The overall performance of the traditional machine learning methods is lesser due to considering the same image for authentication; noises presented in the image samples. To solve this issues, in this work authentication is performed based on the reconstructed image via deep learning method. The major novelty of the work is to apply a deep learning method for image reconstruction and authentication. None of the existing methods will apply deep learning methods for reconstructing of image with real time authentication. This work proposes a novel supervised DLT (Deep Learning Technique) based on SDAE (Stacked Deep Auto Encoder) for image reconstructions in RTFA based systems. The proposed system consists of five major steps: (1) database collection, (2) Pre-processing, (3) SDAE modelling and image reconstruction, (4). RTFA based system, and (5) Performance evaluation. For the first step, 220 faces are collected in real time from 5 persons (each 44) with image size of 92*92. In the second step, facial images, RGB (Red Green Blue) images are converted to Gray scale which is resized into 32*32-pixel images. In the third step, SDAE model reconstructs the image which is then used by the RTFR system to identify a person. The reconstructed facial image is then compared with previously registered images using threshold based NCCs (Normalized Cross Correlations). The proposed SDAE model is evaluated for reconstructions of the original facial images in terms of PSNRs (Peak Signal To Noise Ratios), MSEs (Mean Square Errors), and RMSEs (Root Mean Square Errors). The proposed SADE classifier gives lesser RMSE results of 0.1000 whereas other methods such as CNN, LSTM and VGG16 gives increased MSE results of 0.3500, 0.30741 and 0.27423.
Keywords: Deep learning Techniques, Stacked Deep Autoencoder (SDAE) model, Face authentication, real time face images, Normalized Cross Correlation (NCC), and classifiers.

1 Introduction

Personal authentications have received much attention recently, mainly due to an increasing demand for reliability of assessing identities for security purposes. Systems using traditional authentication methods such as passwords or logons do not identify the individual being authenticated. Imposters who can and do acquire these authentications fraudulently and thus overcome these weak points. Improvements and use of MLTs (Machine Learning Techniques) in data analysis has made it possible for exploring multitude methods of user authentications, specifically Biometric based authentications.

Authentications based on Biometric scan quantify human's biological characteristics where their automated methods exploit user authentications using biometric signs [1]. Many Biometric techniques have been proposed in security applications which authenticate users based [2, 3] and the use of facial based authentications are predominant [4]. The face of a human being is important to social interactions and day to day life as most people are identified with their faces. The use of facial images in identifying humans can be categorized as a classification issue which also is an

*Corresponding author E-mail: showkatme2009@gmail.com

influential task in MLT domains [5,6]. These classifications have attracted researchers into exploring this domain in computer vision [7]. There is ample proof of studies using IPMs (Image Processing Methods) in FASs (Facial Authentication Systems) [8]. Facial recognitions have been applied to wide range applications including the army, security systems, airports and social media [9,10,11]. FASs are being used for identifying humans at a personal level. Further, reconstructions of faces for FASs are also becoming increasingly popular. In spite of multiple studies and techniques prevalent in FASs, their performances when compared to human's visual identifications based on faces seem to fall short in expectations. Achieving accuracies in parallel to human accuracy levels in such applications is a difficult task. Studies have also attempted FASs for enhanced performances in personal authentications. DLT applications in the area of FASs have resulted in higher accuracies. This is mainly due to the advantage of DLTs ability to train and learn data from datasets. Large facial image datasets are available for DLT learning mainly due to the World Wide Web [12,13,14]. Auto encoders are a kind of unsupervised methods in DLTs and have the ability to address issues or errors arising from DLT back propagations [15]. Motivated by these auto encoders, this study proposes a DLT based on these encoders called SDAE for facial image reconstructions. The proposed SDAE is the base for facial authentications called RTFA. This work matches faces with five major steps: (1) database collection, (2) Preprocessing, (3) SDAE model and image reconstruction, (4). RTFA system, and (5) Performance evaluation. These steps have been clearly explained in the proposed work. This section details on related studies in line with this research work. The study by Shen [16] optimized the classification projection matrix of ℓ_1. Their evaluations on public datasets showed that optimizations of the projection matrix improved accuracy between five to seventeen percent when compared to OpenCV (Open-Source Computer Vision) algorithms. The proposed projection matrix output did not need any further re-computations when new faces were added in training. The schema applied SRCs (Sparse Representation Classifications) with the optimized projection matrix on Android smart phones. It was found that SRCs created severe bottlenecks by consuming up to 90% of the execution time. The study overcame this issue by approximating the residuals and making it 50 times faster and without compromising accuracy of recognitions. The feasibility of the new algorithm was demonstrated by implementing and evaluating it with a face unlocking app and where it recognized faces amidst occlusions and lighting variations. A two factor FAS was proposed by Kang [17]. The study authenticated faces using a user given password and matrix transformations. The proposal had a secure cancellation feature where templates created from transformation of feature vectors. Their experimental results produced notable features in transformations. Further, the study also proposed security enhancements as a part of the work. CNNs (Convolution Neural Networks) were used by Liu et al [18] in their study detected live faces using IR (Infrared Radiations) images of a Kinect camera. The scheme's positive samples were live phases while IR outputs of photos or videos were the negative samples. These samples were then trained using CNNs for differentiating spoof attacks from live faces. On detecting live faces, an enhanced FaceNet was used in the study to recognize the face corresponding to an ID or even unfamiliar face for accurate authentications. The study also proposed two methods for running their FaceNet model. CNNs were also used by Pham et al [19] in their study to detect live faces. A Microsoft Kinect camera was used to collect facial RGB (Red Green Blue) and IR images. MTCNNs (Multitask Cascaded Convolution Networks) clipped and aligned facial parts of the images. MTCNN then trained on IR images for live detections, whereas RGB images trained the FaceNet model for facial recognitions. If a live face is detected, face recognition process is continued for authenticating the face. Their experimental results performed better than other techniques when benchmarked with live faces and proving its utility for practical applications. Secure FAS was proposed in the study by Rexha et al [20]. The proposal used locations, gestures for facial recognitions where gestures and location data were treated as time series data. Their study used unsupervised learning of LSTM (Long Short-Term Memory) and RNNs (Recurrent Neural Networks) to learn and identify gestures, groups and then discriminate user gestures based on location information. The study also clustered gestures and locations in their identification process. Viola Jones face detector was used by Zulfiqar et al [21] in their study. The proposed scheme used CNNs for training input images which automatically extracted facial features. These extracted features were used by Viola Jones face detector for recognizing faces. A large database of facial images with different illumination and noise conditions were input into the CNN for its optimal training. The proposed technique's CNN model also selected hyper parameters for in-depth facial identifications. Their experimental results proved their effectiveness in automated biometric FAS systems. Auto encoders figured in the study of Huang et al [22]. The study used ADSNTs (Adaptive Deep Supervised Network Templates) with auto encoders for extracting characteristic facial features from both clean and corrupted images for reconstructions. The study mapped facial images into low-dimensional vectors using "bottleneck" Neural Networks and reconstruct corresponding facial images from the mapped vectors. ADSNT training is then employed to identify new faces by comparing reconstructed images with a gallery of images. Their scheme significantly improved facial recognitions irrespective of occlusions or illumination variations or facial poses. DCNNs (Deep Convolution Neural Network) were used by Salama AbdELminaam et al [23] in their study. The study developed a facial recognition system using transfer learning for fog and cloud computing. Facial images vary mainly due to occlusions or illuminations or expressions and pose which result in decreased performances of FASs.

DCNN extracted relevancy of facial features improved the efficiency of facial comparisons. Their experimental results show its supremacy in comparison to other algorithms. 3D human facial reconstructions were used in studies [24, 25] in their study. Their scheme modelled deep convolution auto encoders for overcoming challenges and complexities in 3D facial reconstructions. Their scheme's convolution encoder's generative models were used as decoders. This differentiable parametric based decoder encapsulated analyzed image formation based on the generative model. Further, input code vectors with precisely defined semantic meaning were decoded for exposing postures, expressions, skin reflections and illumination variations. The study proved that CNN encoders and generative models can be trained in an unsupervised way on unlabelled real world data for FASs. The study's reconstructions when compared with other approaches were highly qualitative and rich in representations. Auto encoders denoised images in the study of Le [26]. The scheme denoised images using a stacked auto encoder and built a deep architecture for extracting significant and dormant features from images. The proposed scheme used a Triplet loss function for preserving equivalent faces and enhancing auto encoder's performance in clustering tasks. The proposed scheme's experimental results were found comparable to deep and on-deep CNN methods. The study by Devi and Baskaran [27] introduced a framework called SL2E-AFRE for 3D facial reconstructions. The scheme localized landmarks of database facial images for a mapping template model. An energy assessing Auto encoder learns facial patches and key points used for predicting landmarks. The scheme reconstructs the 3D images using predicted deformation landmarks. The study's use of energy parameters in training enhanced auto encoder network's performances in face reconstruction models. The experimental results showed using Auto encoders with Simultaneous patch Learning and Landmark Estimations was an efficient way for 3D image reconstructions and could be enhanced with more iterations.

2 Materials and Methods

The proposed system has five major steps: (1) database collection, (2) Pre-processing, (3) Deep Learning model and image reconstruction, (4). Face Authentication System, and (5) Performance evaluation. For the first step, the system collects the faces in real time. It consists of five different persons with totally 220 images (Each person 44 images). In the second step, pre-processing is carried based on two major steps such as color conversion and image resizing. Converting the collected image into grayscale and image manipulation where the image can be resized from grayscale image. Thirdly resized images are modelled using SDAE modelling and results reconstructed image. Fourthly RTFA, reconstructed image and auto encoder model are compared to produce threshold based NCC that indicates the probability of a facial image belonging to the same person. The proposed scheme is evaluated for reconstructions of the original facial images in terms of PSNRs, MSEs, and RMSEs. The proposed methodology of RTFA system is shown in the Figure.

2.1 Image Preprocessing

Preprocessing is carried based on two major steps such as color conversion and image resizing

2.1.1 RGB to Grayscale Conversion

An Input color image with RGB channels is first converted to Gray scale image J. The converted images are then processed by face detection methods. A color image I has three channels (R, G, B) for the output J. This linearly weighted transformation can be depicted mathematically as equation (1),

$$J(x,y) = \alpha.R(x,y) + \beta.G(x,y) + \gamma.B(x,y) \quad (1)$$

Where α - Red Channel Weight, β - Blue Channel Weight, γ- Green channel weight (x, y) – location of pixels of input facial images [28,29]. When $\alpha > \beta >$, facial image detections are enhanced by suppressing noises. The proposed RTFA system found optimality is the values: $\alpha^* = 0.85, \beta^* = 0.10$ and $\gamma^* = 0.05$. The RGB to gray scale conversion is explained as a procedure below,

Load Images and Labels from the dataset Load the image data into Images array as float32 value.

Load Labels of Images as int32 values.
########features adjustment -
From color to RGB(GREYSCALE)

Divide images by 255.0

Call training and testing models

Split the dataset into training and testing with values test_size=0.1 and random_state=42

2.1.2 Image Resizing

In the image processing, need to resize the image to perform the particular operation. Images are generally stored in Numpy ndarray (array). The ndarray. Shape is used to obtain the dimension of the image. Get the width, height, and numbers of the channels for each pixel by using the index of the dimension variable. The resizing of image means changing the dimension of the image, its width or height as well as both. Also, the aspect ratio of the original image could be retained by resizing an image. It takes inputs and makes them go through Downscale which are supposed to give an output similar to the input with reduced size (32*32). Image resizing is described as follows,

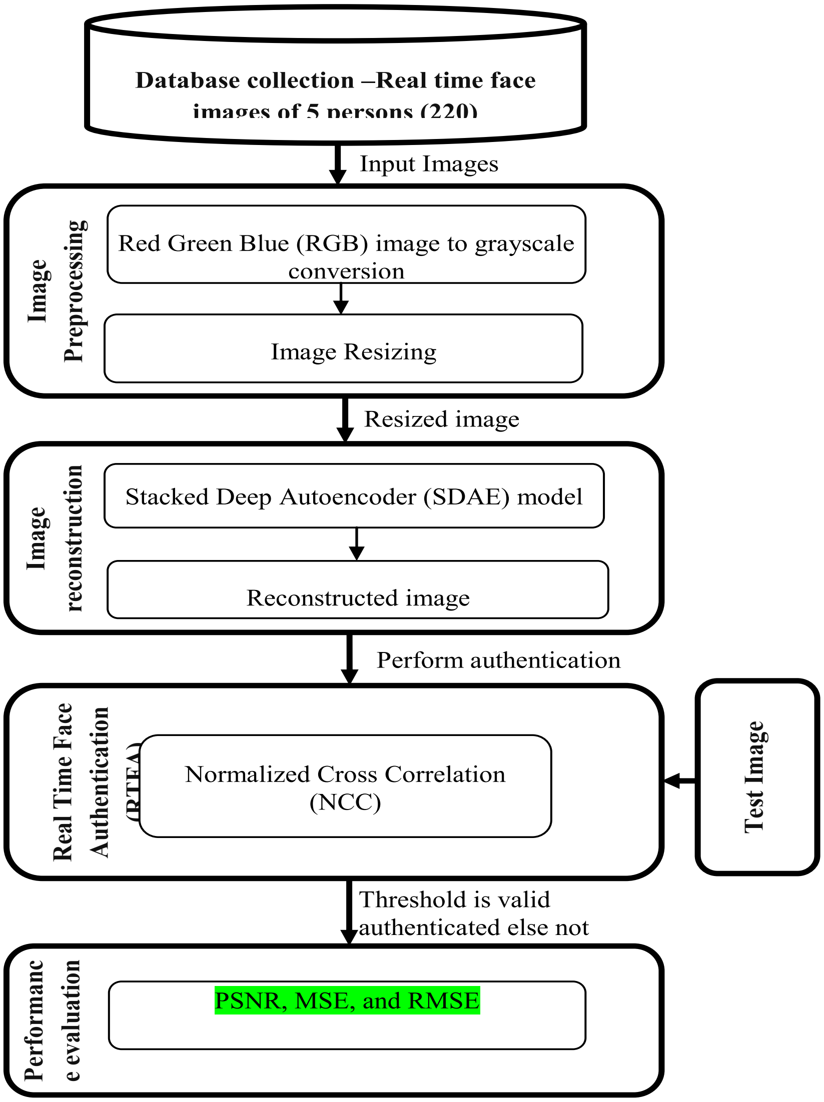


Fig. 1: Real Time Face Authentication (RTFA) System.

IMG_SHAPE = images. Shape [1:]

2.2 Image Reconstruction

Auto encoders are MLNNs (Multi-Layered Neural Networks) [28, 29, 30]. The basic model includes an input, hidden and output layers. Auto encoders are feed forward networks with a hidden layer. Assuming resized input facial images are x_i and a_i^j is the activation of unit of i in a layer j, and w_i is the weights matrix of a function that maps layer i to subsequent layer $i+1$. Assuming there are s_i units in i layer, then there are s_{i+1} units in $i+1$ layers and the matrix w_i will be of size $s_i * s_{i+1}$. The activation function can be depicted as equation (2), where a_1^2 represents first unit of the second layer and $x_0 - x_3$ are 4 input features.

$$a_1^2 = g(w_{10}^1 x_0 + w_{11}^1 x_1 + w_{12}^1 x_2 + w_{13}^1 x_3) \quad (2)$$

An Auto encoder's outputs and inputs which are similar can be expressed as equation (3),

$$\begin{aligned} a = h(x) = f(W^E x + b), x' = h'(x) \\ = g(W^D a + b') \\ = g(W^D h(x) + b') \end{aligned} \quad (3)$$

The Auto Encoder's network is divided into 2 namely an encoder and a decoder where input and hidden layers form the encoder and it converts an input image x into its corresponding feature vector a. Decoders formed from the hidden and output layers transform a into a reconstructed facial image x'. W^E and $W^D$are enoder's and decoder's weight matrices. Activation functions depicted by $f(.)$ and $g(.)$ can be sigmoid or tanh functions and activate each layer's units. $f(.)$ maps x to a while g(·) maps reconstructed and a to x' approximating x. Thus, an input face feature can be reconstructed and compressed to an output feature vector a. The generic cost function can be defined as equation (4),

IMG_SHP = images. shape [1]

per_encoder, per_decoder = build_Auto_Encoder (IMG_SHP, 10)

$$J(W,b) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left\|x_i' - x_i\right\|^2 + \lambda\sum_{l=1}^{N_i-1}\sum_{i=1}^{M_i}\sum_{j=1}^{M_{i+1}}\left(W_{ij}^l\right)^2 \quad (4)$$

The stacked Auto Encoder model used in the study with 3 hidden layers and total of five layers is depicted in Figure 2.

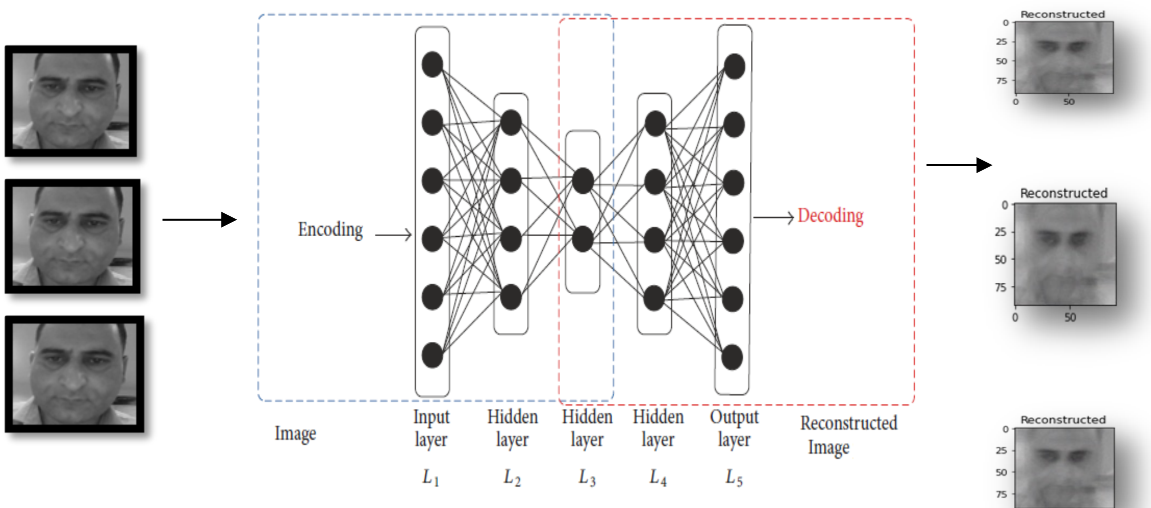
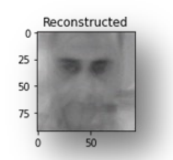


Fig. 2: Stacked Image Reconstructing Deep Auto encoders**.**



Supplementary and symmetrical layers are created by the activation function for encoders. The inputs are then mapped to reconstructed output of decoders thus building a model. *adam* optimizer is used as the compiler where MSE values assess prediction accuracy. The model then uses a training image with 250 epochs for learning which is then tested with test image data and thus reconstructing the image. The Auto Encoder procedure is detailed below:

Call the Dense, Reshape, Flatten and Input layers from Keras Module Image reconstruction with Auto Encoder is listed below

inpts = Input (IMG_SHP)

codes = per_encoder(inpts)

reconstr = pe_decoder(codes)

Auto_Encoder = Model (inpts, reconstr)

Auto_Encoder.compile(optimizer='adam', loss='MSE')

print (Auto_Encoder.summary())

Model:
"model______________________________________
____

Layer_type Output Size (Shape) Parameters

===

input_3 (InputLayer) [(None, 92, 92, 3)] 0

---

sequential (Sequential) (None, 10) 253930

---

sequential_1 (Sequential) (None, 92, 92, 3) 279312

===

Total parameters: 533,242

Train parameters: 533,242

Non-trainable parameters: 0

In reconstructions, pixel values of facial images are altered into feature vectors which are discriminative and compact (Template). In a near ideal condition faces with same subjects get mapped into reconstructed image's feature vectors.

2.3 Real Time Face Authentication (RTFA)

In RTFA systems, the comparison of original and reconstructed images based on similarity values indicate the probability that they identify with the same subject. It is measured using NCC measures [31, 32]. NCC measure is performed by subtracting the mean from the image, computing the cross correlation of the result and dividing that by the standard deviations of the centered images. NCC is a simple and efficient measure of similarity and is invariant to varying brightness or contrasts in images. It uses a function t, within a 2D image function f in its template matches where the location of a desired pattern is represented in the template. A correlation plane is formed in the template by shifting pixels across the entire image and this plane provides the best image matches in a template. In motion tracking applications correlations are a images dependent on time say Im_{t_o} and $\mathrm{Im}_{t_o+\Delta t}$ which are matched on a pixel-by-pixel basis. The correlation matrix $\gamma_{u,v}$ used to store NCC coefficients is defined by equation (5),

$$\gamma_{u,v} = \frac{\sum_{x,y}\left(f(x,y)-\bar{f}_{u,v}\right)\left(t(x-u,y-v)-\bar{t}\right)}{\sqrt{\sum_{x,y}\left(f(x,y)-\bar{f}_{u,v}\right)^2 \sum_{x,y}\left(t(x-u,y-v)-\bar{t}\right)^2}}$$

Where $u \in \{0,1,2,\ldots,M_x-N_x\}$ and $v \in \{0,1,2,\ldots,M_y-N_y\}$ and $\bar{f}_{u,v}$ - mean of f (x, y) within a template t and shifted by (u, v) steps and depicted by equation (6),

$$\bar{f}_{u,v} = \frac{1}{N_x N_y}\sum_{x=u}^{u+N_x-1}\sum_{y=v}^{v+N_y-1} f(x,y) \tag{6}$$

Where, $\bar{t}$ – Template's mean value.

3 Experimentation Results and Discussions

This section evaluates the performance of reconstruction methods such as CNN, LSTM, VGG16 and proposed SDAE. The Database used in the study consists of 220 images (92 x 92 pixels) for five persons. Five original faces and reconstructed images are shown in sections 4.1 and 4.2 and evaluated with PSNR, MSE, and RMSE values.

3.1 Real Time Face Authentication Results

In this section shows the results of five persons with original and reconstructed image. In the figure 3 shows the facial image results of five different persons with two categories: original facial image and reconstructed facial images via the SDAE algorithm. Two rows are shown in the figure 3, in the first row shows the five person facial images with original registered images from figure 3(a), (b), (c), (d) and (e). The second row shows the reconstructed image results of five person facial images via the SDAE algorithm from figure 3(f), (g), (h), (i) and (j).

In the figure 4 shows the Real Time Face Authentication (RTFA) system, figure 4(a) the autoencoder results of the person 1 is shown, if I passed the test image as person 4 then it shows the authentication results as reject

3.2 Evaluation Metric

The evaluation metrics used for assessing the proposed work's model is detailed below

3.2.1 PSNR

PSNR values given as Equation (7), measure the quality between original and reconstructed images where higher values indicate enhanced quality of the reconstructed image.

$$\text{Peak Signal To Noise Ratio (PSNR)} = 10\log_{10}\left(\frac{M^2}{MSE}\right) \quad (7)$$

Where, M – original image's maximum value

3.3.2 MSE

MSE values are estimated using Equation (8) and depicts errors in the reconstructed image,

$$\text{Mean Square Error (MSE)} = \frac{\left(\Sigma(f - f_p)^2\right)}{N} \quad (8)$$

Where, f - new image, f_p –reconstructed image and N - image length.

3.3.3 RMSE

The square root of the MSE is RMSE and estimated from de- reconstructed and original images as given equation (9)

$$\text{Root Mean Square Error (RMSE)} = \sqrt{\frac{\left(\Sigma(f - f_p)^2\right)}{N}} \quad (9)$$

These metrics has been evaluated for image reconstruction results comparison between the methods.

Table 1: Performance Evaluation Metrics Under Reconstruction Methods.

Methods	PSNR (dB)	MSE	RMSE
CNN	55.26	0.1225	0.3500
LSTM	58.15	0.0945	0.30741
VGG16	62.75	0.0752	0.27423
SDAE	68.49	0.0100	0.1000

Figure 5 shows the PSNR comparison results of four reconstruction methods such as CNN, LSTM, VGG16 and SDAE. The proposed SADE classifier gives higher PSNR results of 68.49 dB whereas other methods such as CNN, LSTM and VGG16 gives reduced PSNR results of 55.26 dB, 58.12 dB and 62.75 dB (See Table 1).

Figure 6 shows the MSE results of four reconstruction methods such as CNN, LSTM, VGG16 and SDAE. The proposed SADE classifier gives lesser MSE results of 0.01 whereas other methods such as CNN, LSTM and VGG16 gives increased MSE results of 0.1225, 0.0945 and 0.0752 (See Table1).

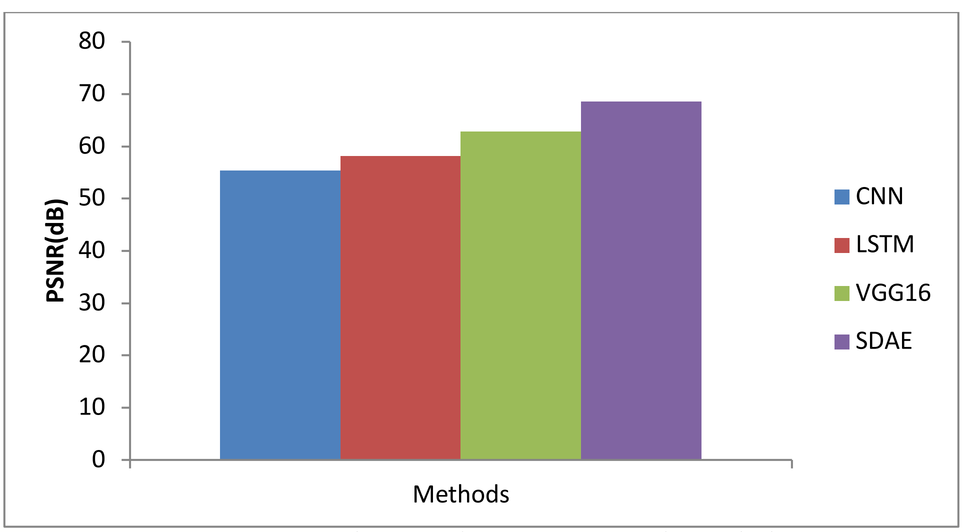


Fig.5: PSNR Results Comparison VS. Reconstruction Methods

4 Conclusion and Future Work

Recently, DLTs have been used in FASs, but require enhancements in while reconstructing images. This

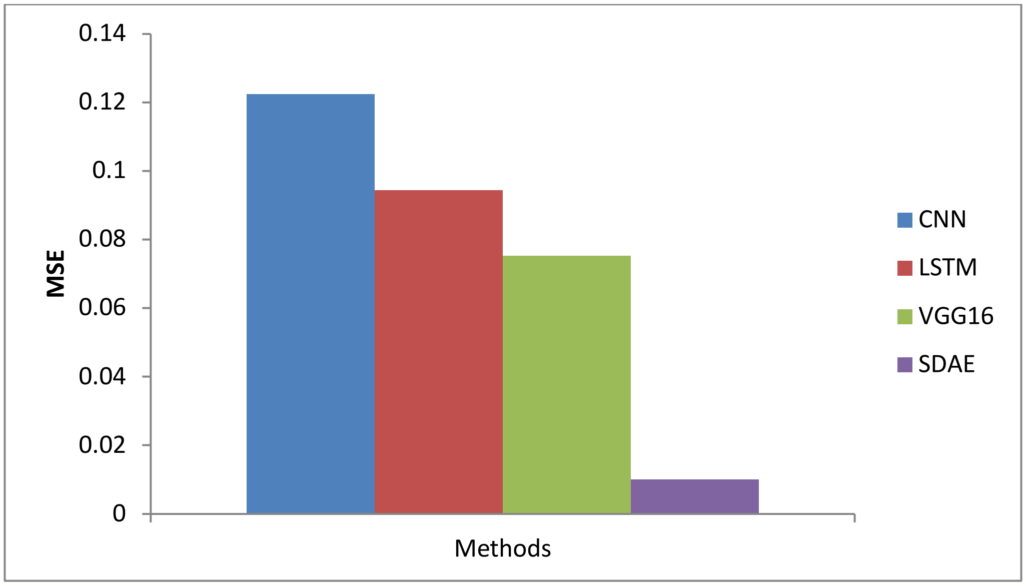


Fig. 6: MSE Results Comparison VS. Reconstruction Methods

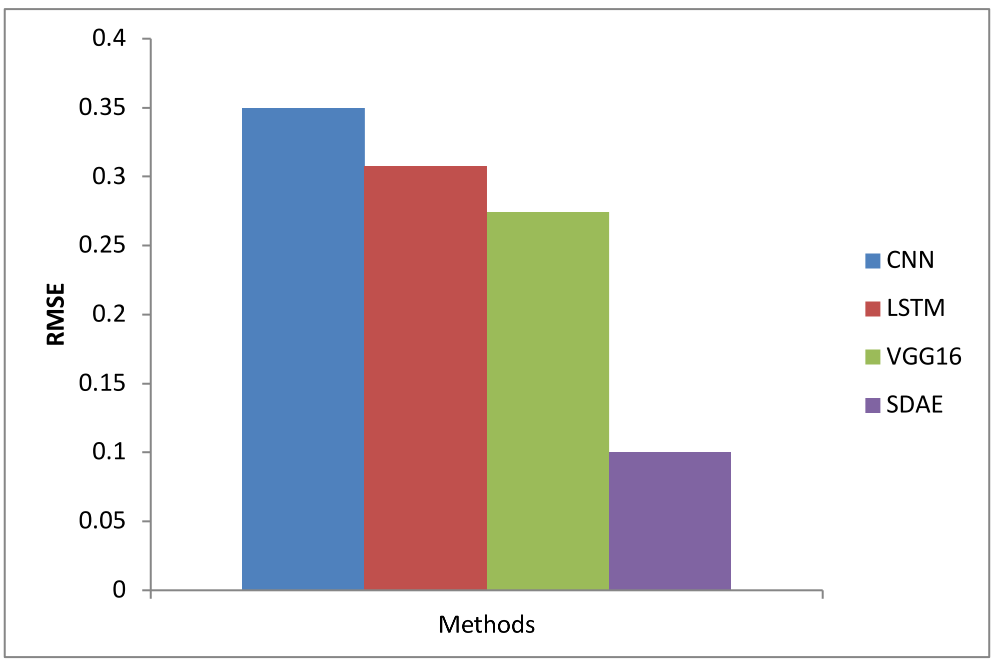


Fig. 7: RMSE Results Comparison VS. Reconstruction Methods.

Figure 7 shows the RMSE results of four reconstruction methods such as CNN, LSTM, VGG16 and SDAE. The proposed SADE classifier gives lesser RMSE results of 0.1000 whereas other methods such as CNN, LSTM and VGG16 gives increased MSE results of 0.3500, 0.27423 (See Table 1).

work presents a RTFA system using SDAE modelling. The proposed system consists of five major steps, including (1) collection of facial images (2) image pre-processing, (3) image reconstruction, (4) face matching, and (5) performance evaluation. The image samples are collected from real time with five persons. From the collected images, color conversion from Red Green Blue (RGB) image to grayscale conversion, then image is resized via the size of 32*32. Image reconstruction, Autoencoder model is constructed with an input, output and hidden layer. The input layer is responsible for generations of feature vectors from the input layer's image inputs. The decoder transforms these vectors to output a reconstructed image using the hidden layer. Face matching; NCC method is used to measure a threshold between the reconstructed image and original image. Results are evaluated with the metrics of PSNR, MSE, and RMSE. This dataset used in the study are quite straight forward in the sense, the images do not have variation and thus the future scope would be to use images with occlusions, postures and illumination variations in images. In future, improving the RTFA model by experimenting on big dataset thus it can be applied to real application in intelligent monitoring systems.

Acknowledgements

The authors express Sincere thanks to the Supervisor (Dr.S. Palanivel) who pave a way in framing this research paper.

Conflict of interest

The authors declare that there is no conflict of interest

References

[1] Yeom, S.K., Suk, H.I. and Lee, S.W., 2013. Person authentication from neural activity of face-specific visual self-representation. Pattern Recognition., **46(4)**, 1159-1169(2013).

[2] Ramya R, Srinivasan K (2020). Real time palm and finger detection for gesture recognition using convolution neural network. In: Human Behaviour Analysis Using Intelligent Systems. Springer, pp. 1–19.

[3] Rzecki K, Pławiak P, Niedźwiecki M, Sośnicki T, Leśkow J, Ciesielski M (2017) Person recognition based on touch screen gestures using computational intelligence methods. Inf Sci (Ny)., 415–416, 70–84(2017).

[4] Moustafa AA, Elnakib A, Areed NFF (2020). Age-invariant face recognition based on deep features analysis. Signal, image video process., 1–8(2020) ...

[5] Pławiak P, M Abdar, UR Acharya (2019). Application of new deep genetic Cascade ensemble of SVM classifiers to predict the Australian credit scoring; Elsevier, applied soft computing; 84(2019), pp.1-14

[6] Tadeusiewicz R (2015) Neural networks as a tool for modeling of biological systems. Bio-Algorithms and Med-Systems., **11(3),** 135– 144(2015).

[7] Abdullah MFA, Sayeed MS, Sonai Muthu K, Bashier HK, Azman A, Ibrahim SZ (2014) Face recognition with symmetric local graph structure (SLGS). Expert Syst Appl., **41**, 6131–6137(2014).

[8] Masi I, Wu Y, Hassner T, Natarajan P (2019). Deep face recognition: A Survey. In: proceedings - 31st conference on graphics, Patterns and Images, SIBGRAPI., 471–478(2014).

[9] Gupta S, Gandhi T (2020). Identification of neural correlates of face recognition using machine learning approach. In: Advances in Intelligent Systems and Computing. Springer., 13–20(2020).

[10] Li Z, Yu P, Yan H, Jiang Y (2020) Face recognition based on local binary pattern auto-correlogram. In: Smart innovation. Springer, Systems and Technologies., 333–340(2020).

[11] Ou W, Luan X, Gou J, Zhou Q, Xiao W, Xiong X, Zeng W (2018) Robust discriminative nonnegative dictionary learning for occluded face recognition. Pattern Recogn Lett., **107**, 41–49(2018).

[12] Sun Y., X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1891–1898, 2014.

[13] Guo Y., L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in European Conference on Computer Vision, 87–102, Springer., 2016

[14] Bansal A., A. Nanduri, C. D. Castillo, R. Ranjan, and R. Chellappa, "Umdfaces: An annotated face dataset for training deep networks," IEEE International Joint Conference on Biometrics (IJCB)., 464–473, 2017.

[15] Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y. and Dobaie, A.M., 2018. Facial expression recognition via learning deep sparse autoencoders. Neurocomputing, **273**,643-649(2018).

[16] Shen, Y. Face recognition on smartphones via optimized Sparse Representation Classification. In Proceedings of the Information Processing in Sensor Networks, Berlin, Germany April .,173–180(2014).

[17] Kang, J. Two-factor face authentication using matrix permutation transformation and a user password. Inf. Sci., **269**, 1–2(2014).

[18] Liu, S., Song, Y., Zhang, M., Zhao, J., Yang, S. and Hou, K., 2019. An identity authentication method combining liveness detection and face recognition. Sensors., **19(21)**, 1-14(2019).

[19] Pham, V.H., Tran, D.P. and Hoang, V.D., 2019. Personal identification based on deep learning technique using facial images for intelligent surveillance systems. International Journal of Machine Learning and Computing., **9(4)**, 465-470(2019).

[20] Tham, K.Y., Cheam, T.W., Wong, H.L. and Fauzi, M.F.A., 2020. Development of a Bi-level Web Connected Home Access System using Multi-Deep Learning Neural Networks. In Computational Science and Technology (227-236). Springer, Singapore.

[21] Zulfiqar, M., Syed, F., Khan, M.J. and Khurshid, K., 2019, Deep face recognition for biometric

authentication. In 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE)., 1-6(2019).
[22] Huang, R., Liu, C., Li, G. and Zhou, J., 2016. Adaptive deep supervised autoencoder based image reconstruction for face recognition. Mathematical Problems in Engineering., **6795352**, 1-14(2016)..
[23] Salama AbdELminaam, D., Almansori, A.M., Taha, M. and Badr, E., 2020. A deep facial recognition system using computational intelligent algorithms. Plos one., **15(12)**, p.e0242269.
[24] Tewari, A., Zollhofer, M., Kim, H., Garrido, P., Bernard, F., Perez, P. and Theobalt, C., 2017. Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1274-1283.
[25] Tewari, A., Zollhoefer, M., Bernard, F., Garrido, P., Kim, H., Perez, P. and Theobalt, C., 2018. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE transactions on pattern analysis and machine intelligence., **42(2)**, 357-370(2018).
[26] Le, X.T., 2019, Building Face Recognition System with Triplet-based Stacked Variational Denoising Autoencoder. In Proceedings of the Tenth International Symposium on Information and Communication Technology., 106-110(2019).
[27] Devi, P.S. and Baskaran, R., 2020. SL2E-AFRE: Personalized 3D face reconstruction using autoencoder with simultaneous subspace learning and landmark estimation. Applied Intelligence.,1-16(2020).
[28] Lu, J. and Plataniotis, K.N., 2009, On conversion from color to gray-scale images for face detection. In 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops., 114-119(2009).
[29] Kumar, T. and Verma, K., 2010. A Theory Based on Conversion of RGB image to Gray image. International Journal of Computer Applications., **7(2)**, 7-10(2010).
[30] Mao, K., Tang, R., Wang, X., Zhang, W. and Wu, H., 2018. Feature representation using deep autoencoder for lung nodule image classification. Complexity., no. **3078374**, 1-11(2018).
[31] Wu, P.; Li, W.; Song, W.L. Fast, accurate normalized cross-correlation image matching. J. Intell. Fuzzy Syst., **37**, 4431–4436(2019).
[32] Duong, D.H.; Chen, C.S.; Chen, L.C. Absolute Depth Measurement Using Multiphase Normalized Cross-Correlation for Precise Optical Profilometry. Sensors., **19**,1–20(2019).