IOP Conference Series: Materials
Science and Engineering

PAPER • OPEN ACCESS

Face Recognition Method Based on Residual Convolution Neural Network

To cite this article: Arshi Husain and Virendra P. Vishvakarma 2022 *IOP Conf. Ser.: Mater. Sci. Eng.* **1228** 012005

Face Recognition Method Based on Residual Convolution Neural Network

Arshi Husain[1] and Virendra P. Vishvakarma[2]

USICT, Guru Gobind Singh Indraprastha University

arshihusaincdac@gmail.com, virendravishwa@rediffmail.com, vpv@ipu.ac.in

Abstract. With the advancement of information technology and societal growth, social security has become more important than ever. Face recognition, as compared to other traditional recognition methods like fingerprint recognition, palm recognition, etc, has the benefit that it is contact less, and now it is becoming one among the most prominent technologies in development. Although there are numerous recognition systems that use DNNs in the field of facial expression recognition, their accuracy and practicality are still insufficient for real-world applications. A facial recognition approach based on Resnet 152 v2 has been proposed in this work. In this paper, a residual learning approach is presented to make the training of networks that are far deeper than previously employed networks easier. The proposed method, employs the AT&T face dataset, and supposing that normalization and segmentation are complete, we concentrate on the subtask of person verification and recognition, demonstrating performance using a testing database comprising illumination, pose, expression and occlusion variations. SoftMax is the activation function that has been used, which adjusts the output sum up to one allowing it to be understood as probabilities. Then, the model would generate a judgment depending on which option has a strong likelihood. This system employs Adam as an optimizer to control the learning rate through training and categorical cross entropy as its loss function. The proposed approach has a 97 percent face recognition accuracy on AT&T dataset, showing its efficacy after a significant number of analyses and experimental verification.

Keywords: Deep convolutional neural network, Face recognition, ResNet.

1. Introduction

In many computer visual identification applications, the CNN model has outperformed standard machine learning algorithms, thanks to the advancement of deep learning.
Face recognition algorithms are now being improved in three areas: face pre- processing (comprising face alignment and detection), extraction of features (mostly developing ANN structures), and classification of features.
Face confirmation and face verification are generally two subtasks in face recognition. The former categorizes faces into distinct identities, whilst the latter assesses if 2 face picture pairs are associated with the same identity.
Automatic face recognition is a multi-step process which includes face detection and localization in a messed-up environment, normalization, recognition, and verification. Some of the subtasks may be difficult to complete depending on the nature of the application, such as the size of the training and testing databases, background variability, occlusion, noise, and speed requirements. Supposing that normalization and segmentation are complete, we concentrate on the subtask of person verification and recognition, demonstrating performance using a testing database of about 400 pictures.
In the realm of image identification, CNN is the most used machine learning algorithm. Visual Geometry Group (VGG-15) network contains nineteen layers whereas LeNet comprises five layers of network. The 100-layer barrier was not broken until the development of networks like ResNet in 2016. By constructing a short connecting channel from the front layer to the rear layer, the signal is sent immediately from the one layer to the next via ResNet.
Although there are numerous recognition systems that use Neural Networks in the field of facial expression recognition, their accuracy and practicality are still insufficient for real-world applications. A facial recognition approach based on Resnet 152 v2 has been proposed in this work which has better accuracy than the existing ones. Since we have used the deep Neural networks in our system, hence the features need not to be extracted manually.

IOP Conf. Series: Materials Science and Engineering 1228 (2022) 012005 doi:10.1088/1757-899X/1228/1/012005

2. Related Work

In the past, various classification approaches were used to address the problem of facial recognition. Trunk-Branch Ensemble CNN model was given by Changxing Dingand Dacheng Tao.Onetrunk network and numerous branch networks are contained by Trunk-Branch Ensemble CNN. The trunk network was trained so that it can learn face representations for holistic face pictures, whereas the branch networks were programmed in order to make it learned face representations for image patches sliced from 1 facial component. GoogLeNet has been used to implement the trunk network. They also present a technique to generate video like training picture data in order to achieve blur-resistant face identification. They tested their findings using PaSC and YouTube Faces, two publicly available large-scale video face datasets. On the YouTube Faces Database, their mean verification accuracy using the TBE-CNN technique is 94.96 percent, and on the PaSC dataset, their verification rate using the TBE-CNN method is 95.83 and 94.80 on the control set and hand held set, respectively[1]. Vallimeena P, Uma Gopalakrishnan, Bhavana B Nair, and Sethuraman N Rao used the FERET Database to perform ethnicity-based face categorization on 447 samples (357 for training, 90 for testing) Extraction of skin color and computation of the normalised forehead area utilising the Sobel Edge Detection technique, with a 94 percent experimental accuracy [2].

Karthikeyan Shanmugasundaram, Sathees Kumar Ramasamy and Sharma S deployed a FAREC - CNN Based Efficient Facial Recognition Technique using Dlib. Face Recognition Grand Challenge dataset has been used in their paper [3]. Face detection is the first method employed in this study to detect human frontal faces from digital pictures or videos. For this purpose, computer vision technology was used to detect frontal faces using facial landmark identification of the nose and upper lips. Following that, the Dlib is used to align the frontal faces. Following that,face cropping is used to cut off the face with varying resolution depending on the distance between the face and the camera. Finally, CNN is used to extract features from the face images. So finally, FAREC takes 20 epochs and produces 96% accuracy for FRGC [3]. Yuxiang Zhou and Hongjun Ni developed a Face and Gender Recognition System Based on Convolutional Neural networks. The datasets used were Labeled Faces in the Wild (LFW), YouTube Face (YTF) and VGGFace2 and the highest achieved accuracy was 93.22% [4].

Ying Wen, Pengfei Shi developed a face recognition approach using PCA, 2DPCA (2D)2, PCA, IPCA as the feature extraction techniques on ORL and Yale datasets. They used Nearest Neighbour classifiers for the purpose of classification and the highest recognition rate was 90%. [5] Divya Meena, Ravi Sharan proposed a method for face detection and face recognition. In their work, viola Jones algorithm was used to detect face and principal component analysis for face recognition. The highest accuracy achieved was 90% [6].

A method proposed by W. Nan, Z. Zhigang, M. et al used ResNet and enhanced edge cosine loss transform for the purpose of face recognition and highest recognition accuracy achieved was 72.602% and verification accuracy was 85.420% [7]. P. Shamna, C. Tripti et al investigated the Difference Component Analysis (DCA) Approach on AT & T dataset for face recognition and 73.3% correct recognition rate was acquired [8].

W. Wang and W. F. Wang proposed a Grayscale Face Recognition approach, in which feature extraction was done by deriving the whole expression of gray conversion; the face classifier was trained by back propagation learning algorithm and achieved 93.33% recognition rate [9]. H. Hatimi, M. Fakir et al developed an approach for face Recognition Using Fuzzy and Multi-agent systems from Video Sequences. Face detection was done using texture color and geometrical face, multi-agent system and fuzzy approach are used in the recognition process and achieved 95% recognition rate [10].

C. Lu and X. Tang. investigated human-level face verification performance on LFW dataset using Gaussian Face model and achieved 93.73% accuracy [11]. D. Wang, H. Yu, D. Wang and G. Li explored different activation functions in CNN for Face Recognition System Based and the highest achieved accuracy was 88.41% [12].

S. Ergin and M. Bilginer Gulmezoglu developed a Face Recognition method using Face Partitions with Common Vector Approach and achieved 93.3% recognition rate [13]. Y. Wen and P. Shi proposed a method for the purpose of Face Recognition using PCA and achieved 90% accuracy [14] S. M. S. Hossain, A. Yousuf and M. S. Sadi, proposed Eigen vector and Covariance matrix-based approach for face recognition and achieved 96.25% accuracy [15].

3. Dataset and Proposed Methodology

3.1. Image Dataset

This method employs the AT&T face dataset, formerly referred as the ORL Database Of Faces. It is a collection of 400 face photos, with 10 different images of each of 40 different subjects. The pictures for people had been taken at varying conditions like different facial emotions (e.g smiling / not smiling etc), varied lighting and different facial features (with glasses / without glasses) were being used.

The people were snapped in a frontal, upright position against a background that was black and uniform. Pictures in the dataset were saved in **Portable Gray Map (pgm)** format before being converted to jpg. Each picture is 92x112 pixels in size and has 256 grey levels per pixel. A preview image of the Database of Faces is shown in Fig.1.

Fig.1. Dataset Images

3.2. Proposed Methodology

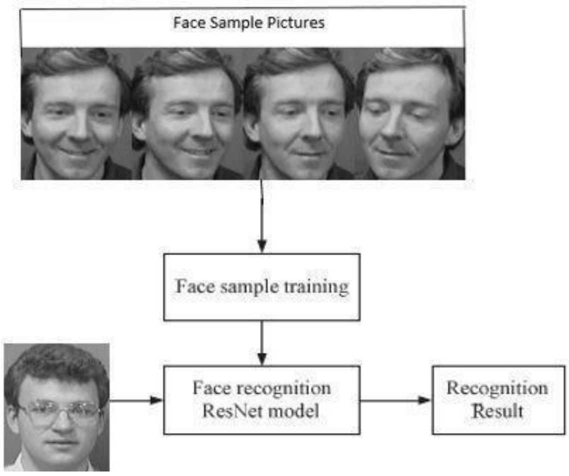

Fig.2. Resnet based Face recognition method

The proposed system mentioned below uses residual networks for the face recognition problem. Figure 2 depicts the concept of a *ResNet*-based face recognition algorithm. The *ResNet* based face recognition method is used in this work.

Step1: The grayscale images of dataset which are in pgm format are loaded into the system and then they are converted into jpg format.

Step 2: We split the original dataset which contain 400 images in the ratio of 70:30, i.e, into two sub datasets known as training dataset and testing dataset, as a result, 280 pictures are used to train the system and 120 pictures are used to test it.

Step 3: We import the required libraries, add a preprocessing layer to the front of *RESTNET152V2*.

IOP Conf. Series: Materials Science and Engineering 1228 (2022) 012005 doi:10.1088/1757-899X/1228/1/012005

Step 4: Model Compilation has been done using the three parameters that are Optimizer, Loss function and metrics

Step 5: The model has been trained by utilising the 'fit()' function with the following three parameters:

(A) number of epochs used (B) validation data (C) training data target data.

RESIDUAL NETWORK MODEL

Lately, Deep convolutional networks have led to tremendous breakthroughs for image generation, classification and so on. With development of Residual networks, the difficulty of training incredibly deep networks has been alleviated and these *Resnets* are made up from Residual Blocks. The residual learning module allows to train layers with hundreds or even thousands of them and still get great results. The *resNet* learning fundamental idea is to preserve a portion of the original input data throughout CNN unit training in order to prevent classification accuracy saturation generated by numerous convolutional layers. Simultaneously, it is not necessary for the residual module to memorize the entire output; instead, it only has to learn the differences between the input and output, which optimizes the learning objectives and complexity.

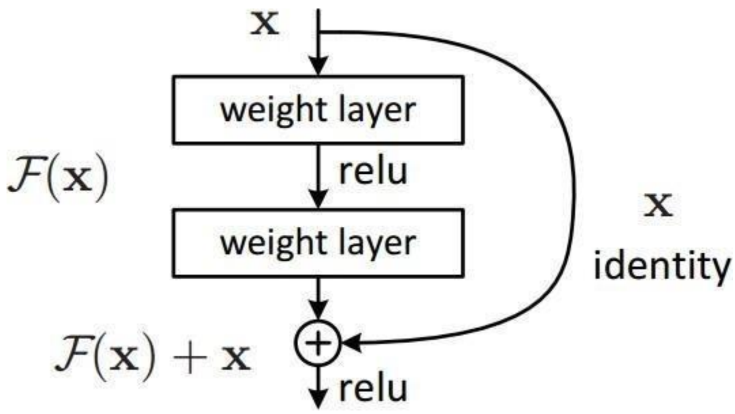

Fig.3. Residual Learning: A building block

There is a direct link in *ResNet* that skips certain layers (which may vary depending on the model) in between. The core of residual blocks is a link known as the 'skip connection.' That adds previous layer outputs to stacked layer outputs. By including skip connections to our network, we are permitting the network to leave out training for the layers that are not relevant and do not contribute value to overall accuracy, rather than using the number of layers as an essential hyperparameter to tune. Skip connections, in a sense, make our neural networks dynamic, so that it can tune the number of layers appropriately during training

The residual block shown in Fig.3. is mathematically represented as F(x)

$$y = F(x, \{W_i\}) + x \tag{1}$$

Here the input and output vectors of the layers in concern are x and y. The residual mapping to be learned is represented by the function F(x, Wi).

The residual block consists of various weight layers which are represented as Wi.The number of weight layers should be greater than one. The two layers contained in the residual block can be illustrated using the following equation:

$$F(x, \{W_i\}) = W_2\boldsymbol{\sigma}(W_1x) \tag{2}$$

where σ represents the Rectified Linear Unit activation function. so

$$H(x) = \boldsymbol{\sigma}(y) \tag{3}$$

In this research, our system utilises Resnet152V2, which employs batch normalisation (BN) processing to speed up the training pace of the entire training network in the proposed module which is based on residual learning. The *ResNet V2* contains the non-linearity which refers to an identity mapping. This mapping can be considered as the

output of an addition of the residual mapping and the identity that is to be passed on for the processing in the next block. In ResNet V1, the result of the addition operation is transmitted to the next block as the input via *ReLU activation*.

4. Experimental Result and Discussion

The TASK was performed using a 64-bit operating system laptop. The machine had 16 GB of RAM installed. The CPU utilized was an Intel i5-9300H with a clock speed of 2.5 GHz. The operations were performed on anaconda software running on python 3.8 and jupyter notebook version2.1 with libraries used for image classification like *tensorflow*. Initially, the grayscale images in pgm format are loaded into the system and then they are converted into jpg format. We have split the original dataset containing 400 images into two sub datasets known as training dataset and testing dataset in a ratio of 70:30, as a result, 280 pictures are used to train the system and 120 pictures are used to test it.

We import the required libraries, add a preprocessing layer to the front of *RESTNET152V2*. SoftMax is the activation function that has been used, which adjusts the output sum up to one allowing it to be understood as probabilities.

Maximum Entropy (MaxEnt) Classifier is another term for *Softmax Regression* and the formula for *Softmax function* is as follows:

$$\sigma(\vec{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \tag{4}$$

where,

$\vec{z}$	The input vector to the softmax function, made up of (z0, ...zK)
z_i	All the Z_i values are the elements of the input vector to the softmax function, and they can take any real value, positive, zero or negative.
e^{z_i}	The standard exponential function is applied to each element of the input vector, this gives a positive value above 0, which will be very small if the input was negative, and very large if the input was large.
$\sum_{j=1}^{K} e^{z_j}$	It is the normalization term. It ensures that all the output values of the function will sum to 1 and each be in the range (0,1), thus constituting a valid probability distribution.
K	The number of classes in the multi-class classifier.

Table.1. Symbols used in the Softmax Formula

Then, the model would generate a judgment depending on which option has a strong likelihood.

4.1. Model Compilation

Loss, Optimizer and metrics are the parameters using which the model was then compiled. The system utilizes Adam Optimizer which governs the rate of learning during the training phase.

The impact of learning rate affects the speed of computation of optimum weights. Although the lower speed increases the computation time but yields better results. For our loss function, this system uses *categorical cross entropy* and a lower score means our model is doing better. The system uses 'accuracy' metric to view accuracy score on validation set when we train the model which will make things much easier to comprehend.

4.2. Training the model

In this system, the 'fit()' function has been utilized with the following parameters to train our model: 1. number of epochs used 2. validation data 3. training data target data. The test set created for us for the AT & T dataset was used for validation data. In this model, 25 epochs the no.of epochs is set 25. After 25 epochs, we have gotten to 97% accuracy on our validation set. Fig.4. and Fig.5. shows the comparison graph between Training Accuracy and Validation Accuracy respectively on AT&T Dataset.

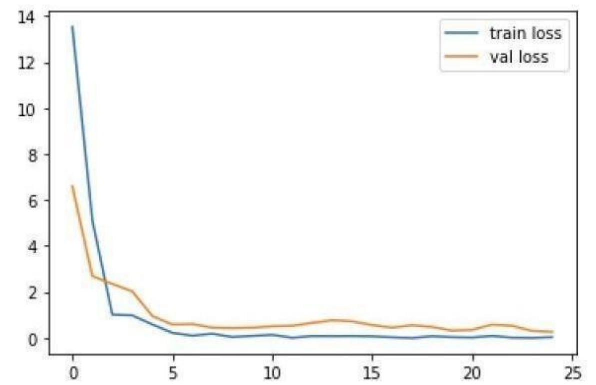

Fig 4. *Comparison of Training & Validation Loss*

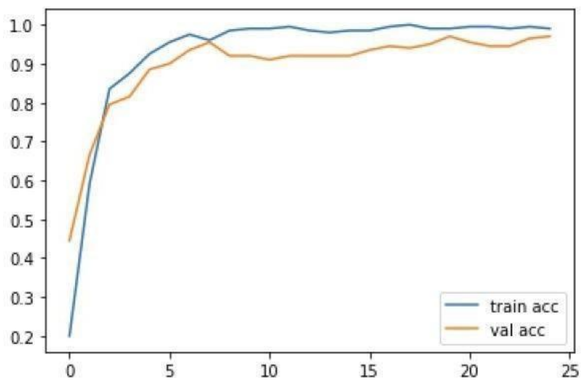

Fig. 5. *Comparison of Training & Validation Accuracy*

5. Conclusion

A facial recognition model using ResNet is proposed in this paper. Resnet 152v2 is the residual network variation that we selected since, despite the additional depth, the 152-layer ResNet has lesser complexity and is more accurate than other networks such as VGG-16/19 networks. The softmax activation function, adam optimizer and categorical cross entropy as a loss function have been employed in our system. We haven't noticed any degradation issue; therefore, we get large accuracy benefits from the greater depth. This paper's facial recognition algorithm has been trained and tested on the AT&T dataset. The findings show our suggested algorithm is superior and has a lot of promise in the open face recognition problem. In our future work, larger size databases will be explored for the investigation of proposed methods.

IOP Conf. Series: Materials Science and Engineering 1228 (2022) 012005 doi:10.1088/1757-899X/1228/1/012005

References

[1] r C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 1002-1014, 1 April 2018

[2] P. Vallimeena, U. Gopalakrishnan, B. B. Nair and S. N. Rao, "CNN Algorithms for Detection of Human Face Attributes – A Survey," 2019 International Conference on Intelligent Computing and Control Systems (ICCS), 2019, pp. 576-581

[3] S. Sharma, K. Shanmugasundaram and S. K. Ramasamy, "FAREC — CNN based efficient face recognition technique using Dlib," 2016 International Conference on Advanced Communication Control and Computing Technologies (ICACCCT), 2016, pp.192-195

[4] Y. Zhou, H. Ni, F. Ren and X. Kang, "Face and Gender Recognition System Based on Convolutional Neural networks," 2019 IEEE International Conference on Mechatronics and Automation (ICMA), 2019, pp. 1091-1095

[5] Y. Wen and P. Shi, "Image PCA: A New Approach for Face Recognition," 2007 IEEE International Conference on Acoustics, Speech and Signal Processing - ICASSP '07, 2007, pp. I-1241-I-1244

[6] D. Meena and R. Sharan, "An approach to face detection and recognition," 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE), 2016, pp. 1-6

[7] W. Nan, Z. Zhigang, M. Jingqi, L. Huan, L. Junyi and Z. Zhenyu, "Face Recognition Method Based on Enhanced Edge Cosine Loss Function and Residual Network," 2019 Chinese Control and Decision Conference (CCDC), 2019, pp. 3320-3324

[8] P. Shamna, C. Tripti and P. Augustine, "DCA: An Approach for Face Recognition through Component Analysis," 2013 Third International Conference on Advances in Computing and Communications, 2013, pp. 34-38

[9] W. Wang and W. F. Wang, "A GrayScale Face Recognition Approach," 2008 Second International Symposium on Intelligent Information Technology Application, 2008, pp. 395-398

[10] H. Hatimi, M. Fakir and M. Chabi, "Face Recognition Using a Fuzzy Approach and a Multi-agent System from Video Sequences," 2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV), 2016, pp. 442-447

[11] C. Lu and X. Tang. Surpassing human-level face verification performance on LFW with GaussianFace. arXiv Preprint arXiv:1404.3840, 2014.

[12] D. Wang, H. Yu, D. Wang and G. Li, "Face Recognition System Based on CNN," 2020 International Conference on Computer Information and Big Data Applications (CIBDA), 2020, pp. 470-473

[13] S. Ergin and M. Bilginer Gulmezoglu, "Face Recognition Based on Face Partitions Using Common Vector Approach," 2008 3rd International Symposium on Communications, Control and Signal Processing, 2008, pp. 624-628

[14] n S. M. S. Hossain, A. Yousuf and M. S. Sadi, "Towards an efficient face recognition approach," 2015 International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), 2015, pp. 1-5

[15] Bhawna Ahuja, Virendra P Vishwakarma, "Deterministic Multi-kernel based extreme learning machine for pattern classification," Expert System with Applications,2021, pp. 115308.

[16] Sahil Dalal, Virendra P Vishwakarma, "Classification of ECG signals using multi-cumulants based evolutionary hybrid classifier", Scientific Reports, 2021

[17] Virendra P Vishwakarma, Varsha Sisaudia, "Self-adjustive DE and KELM-based image watermarking in DCT domain using fuzzy entropy", International Journal of Embedded Systems, 13 (1), 74-84

[18] Sahil Dalal, Virendra P Vishwakarma, "A Novel Approach of Face Recognition Using Optimized Adaptive Illumination–Normalization and KELM", Arabian Journal for Science & Engineering (Springer Science & Business Media BV), 2020, Vol.45, issue 12

[19] VP Vishwakarma, S Dalal, "A novel non-linear modifier for adaptive illumination normalization for robust face recognition", Multimedia Tools and Applications, 2020, 79 (17), 11503-11529