A Comprehensive Survey on the Face Recognition Using CNN

Mr. T. Madhu[*1] K.V. Hemanth[*2] Sheldon Shaji[*3] V. Nikil[*4]
Associate Professor[*1] *Scholar*[*2,3,4]
Department of Information Technology, Nalla Narsimha Reddy Education Society's Group of Institutions[*1,2,3,4]

Abstract

Face recognition is one of the interesting research topics in the field of computer vision. In recent years, deep learning methods, especially the Convolutional Neural Network, have progressed. One of the successes of CNN is in face recognition. Face recognition by computer is a technique done so that the computer can automatically recognize faces in an image. This survey presents research related to face recognition based on Convolutional Neural Network that has been conducted. The studies used are studies that have been published in the last five years. It was performed to determine the renewal that emerged in face recognition based on Convolutional Neural Network. The basic theory of the Convolutional Neural Network, face recognition, and description of the database used in various researches are also discussed. Hopefully, this survey can provide additional knowledge regarding face recognition based on the Convolutional Neural Network. Keywords: convolutional neural network; deep learning; face recognition; survey.

Introduction

Humans often use faces to recognize individuals and advancements in computing capability over the past few decades now enable similar recognitions automatically. Early face recognition algorithms used simple geometric models, but the recognition process has now matured into a science of sophisticated mathematical representations and matching processes. Major advancements and initiatives in the past ten to fifteen years have propelled face recognition technology into the spotlight. Face recognition can be used for both verification and identification (open-set and closed-set). In face recognition system it identifies faces present in the images and videos automatically. It is classified into two modes:

1. Face verification (or authentication)
2. Face identification (or recognition)

Face recognition has been an active research area over last 40 years. The face recognition research has several disciplines such as image processing, machine learning approach, pattern recognition, computer vision, and neural networks. Classification is the main problem. In the process of face recognition it includes, to train the face images from the known individuals and then to classify the newly coming test images into one of the classes

In automatic face recognition system, the main complicated task is that it involves detection of faces from a cluttered background, facial feature extraction, and face recognition. A complete face recognition system has to solve all sub-problems, where each one is a separate research problem. Image template based and geometry feature-based are the two classes of face recognition system algorithms. In template-based method it (Robert J. 1981) computes the correlation between a face image and one or more model of face image templates to estimate the face image identity from the database. Brunelli and Poggio (R. Brunelli, 1993) suggest the optimal strategy for face recognition system which is holistic and corresponds to template matching. The statistical tools such as Support Vector machines (SVM) (E. Osuna, 1997), (Vladimir N, 1995) Independent component Analysis, Principal Component Analysis (PCA) (L. Siro ich, 1987), (Matthew Turk, 1991), Linear Discriminant Analysis (LDA) (Peter N. Belhumeur et.al, 1997), used to construct a suitable database of face image templates.

LITERATURE SURVEY

Zafar et al. focused on making Bayesian DCNN for face recognition [1]. The researchers formed three CNN models with different depths and a Bayesian-DCNN. The use of different models is done to test the accuracy of each model. The dataset used in the study is the AT&T Face Database When tested with 100% with the AT&T Face Database.

Bendjillali et al. in their research, is performed with the architecture Inception-v3 [2]. The Viola-Jones algorithm is used to detect the face. The authors increased the contrast of the training image to determine the impact given to the accuracy of face recognition. Testing is done using CMU PIE datasets. The results obtained in testing using the CMU PIE, the accuracy of face recognition with Inception-v3 is 99.89%.

Hu et al. used 3 Convolutional Neural Networks arranged in parallel and proposed a Diversity Combination method for face recognition [3]. Inception-ResNet-v1 was used as the basic architecture of CNN. The three CNNs were used for feature extraction, and diversity combination is a strategy used to adaptively adjust the weight value in each CNN and make joint classification decisions. CAISA Web Face were used to train the CNN. Face matching testing was performed using CASIA NIR-VIS 2.0 and Oulu-CASIA NIR-VIS dataset. The accuracy level obtained was 99.8% using the Oulu-CASIA NIR-VIS dataset.

Li et al. in their research used multi-CNN [4]. Detection of facial components (eyes, eyebrows, nose, mouth) is done using the Active Shape Model (ASM). CNN is then used to extract features from each facial component. The dataset used for training and testing is Nenu LD. The highest level of accuracy is achieved is 99.5%.

Ding proposed a CNN model called Trunk-Branch Ensemble CNN (TBE -CNN) for video-based face recognition [5]. The model was designed to be able to extract additional information from the face holistically, and facial parts took around facial components. The TBE-CNN model is based on Go o g LeNet. Data augmentation. Test are performed by COX Face; accuracy was achieved using the Pa SC database 99.33%.

Deng et al. proposed a facial recognition algorithm to recognize a masked face based on large margin cosine loss called MF Cos face [6]. In their work, MTCNN was utilized for face detection, and the base architecture used is the Inception-ResNet-v1, and the proposed large margin cosine loss was used to train the model. Some faces from the datasets are generated with a mask. The accuracy achieved on LFW m is 99.33% respectively.

Yang et al. proposed a face matching method named SR-CNN [7]. The study combines Rotation- Invariant Texture Features (RITF) and Convolutional Neural Network (CNN). Testing and training are done using the LFW dataset. The results obtained reached 98.98%. The combination of RITF, SIFT, and CNN provides an increase in the level of accuracy of face matching.

Nam et al. proposed a Pyramid-Based Scale-Invariant Convolutional Neural Network (PSI-CNN) model [8]. The method is used to overcome the problem of input image scale, for CNN to recognize faces on low-resolution images, therefore increasing the performance of CNN. Evaluation and training is done using the LFW database, which is a standard dataset for face recognition evaluation. The results obtained are that the model has a face matching accuracy level of 98.87%.

Peng et al. introduced a CNN called NIR Face Net [9]. NIR Face Net is a modification of Goo g Le Net, used for face recognition whose input is in the form of Near-Infrared (NIR) images. The researchers chose to do face recognition with a NIR image because it has advantages over lighting changes. The dataset used is the CASIA NIR database. Then, the testing with blurred and noise images achieved accuracy ranging from 96.02% to 98.48%.

Nam et al. proposed a Pyramid-Based Scale-Invariant Convolutional Neural Network (PSI-CNN) model [10]. The method is used to overcome the problem of input image scale, for CNN to recognize faces on low-resolution images, therefore increasing the performance of CNN. Evaluation and training is done using the LFW database, which is a standard dataset for face recognition evaluation. The results obtained are that the model has a face matching accuracy level of 98.37%.

Pei et al. made a student attendance system using face recognition based on deep learning [11]. There was a difficulty in getting a large amount of training data. Data augmentation was used as a solution to increase the number of images in the dataset that can be used. Researcher used the CNN VGG-16 architecture had taken data through face capture via video, and the level of accuracy increased to 98.1.

Khan et al. used a Convolutional Neural Network to detect and recognize faces [12]. R-CNN that is used for object detection, is used by researchers to detect faces in images. The dataset used for training is LFW, and the authors used the data augmentation method by reversing each image in the dataset. The accuracy achieved at around 97.9%.

Song et al. proposed the Spartan Face Mask Detection and Facial Recognition system to address the challenge of mask detection and identity recognition that emerged during the Covid-19 pandemic [13]. The proposed system utilized the MTCNN to detect faces in an image, Face Net to extract embedded facial features, The training and testing were done using a total of 2000 images. The accuracy achieved using the Face Net + SVM is 97%

Ben Fredj et al. trained a CNN to recognize faces in an uncontrolled environment, in the sense that face images have noise, or partially covered faces [14]. The researchers used the data augmentation method of flipping. Data augmentation increases the number of images in the dataset and adds variation. The research in [19] has little in common with research. when tested with the YTF dataset it gives a result of 96.83%.

Hu et al. proposed a method for increasing the number of images in the dataset through image synthesis [15]. The architecture is named CNN-L. CNN-L has more layers than CNN-S The dataset used for training is NIR-VIS 2.0 The test results show the highest accuracy obtained using CNN-L, which is 95.77% The model was designed to be able to extract additional information from the face holistically, and facial parts took around facial components.

Ullah et al. proposed a unified framework for mask detection and masked facial recognition [16]. The authors made a custom CNN called Deep Mask Net, which comprises 17 layers, including the input and classification layers. was used for training and testing the model. The Deep Mask Net achieved 100% accuracy in face mask detection and 93.33%.

Chen et al. in their study, proposed a CNN with deep transformation learning [17]. The method increases the robustness and degree of discrimination of the extracted features. The datasets used for training are CASIA-Web face. Tests were carried out by researchers using the LFW and IJB- A datasets. The test results show that the accuracy rate obtained using IJB-A is 93.1%.

Zangeneh et al. [18] uses the model based on the VGG net model. Model training uses the UMIST and FERET datasets. Testing is done using FERET face datasets. The accuracy obtained using the FERET dataset reached 92.1%.

Moon et al. trained CNN to recognize a person's face at different distances [19]. The used dataset is a dataset formed by the researchers themselves (private), with 12 individuals, and 300 images for each individual. Everyone's faces in the dataset are taken from 5-meter to 15 meters, with 30 images taken at each distance. The average level of accuracy obtained from this study was 88.9%.

Deep Learning Methods for Feature Extraction

Deep Learning:

Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers. The Figure (a) discusses about Deep Learning. These neural networks attempt to simulate the behaviour of the human brain albeit far from matching its ability allowing it to "learn" from large amounts of data. While a neural network with a single layer can still make approximate predictions, additional hidden layers can help to optimize and refine for accuracy. Deep learning drives many artificial intelligence (AI) applications and services that improve automation, performing analytical and physical tasks without human intervention. Deep learning technology lies behind everyday products and services as well as emerging technologies.

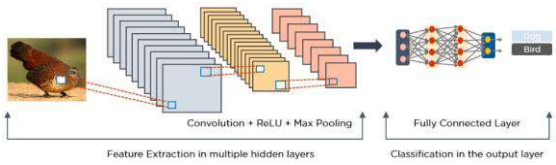


Fig (a): Deep Learning

OPEN FACE:

Open Face is an open-source tool intended for computer vision and machine learning researchers, the affective computing community and people interested in building interactive applications based on facial behaviour analysis. It is the first open-source tool capable of facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation. The computer vision algorithms which represent the core of Open Face demonstrate state-of-the-art results in all of the above-mentioned tasks.

ARC FACE:

Arc Face is a specific approach or loss function used in face recognition tasks. The figure (b) describes the how the arc face will work. It's designed to enhance the discriminative power of features extracted from facial images, improving the accuracy of face recognition systems. Arc Face is particularly known for its effectiveness in handling the challenging problem of large variations in lighting, pose, and facial expressions.

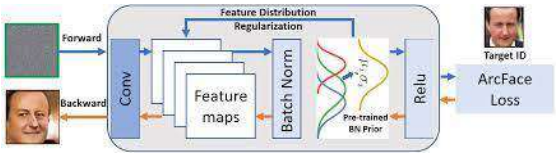


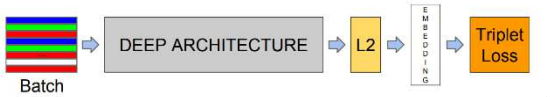
Fig (b): Arc Face

VGG FACE:

There are many pre-trained deep learning models so that by using those architectures we can train our model instead of creating a new architecture every time It is based on the VGG architecture, which is known for its deep convolutional neural network (CNN) layers. The VGG Face model is capable of extracting rich and discriminative features from facial images, enabling accurate face recognition. Multiple layer detection on the last feature map is used to detect varied face sizes. Details of how to crop the face given detection can be found in vgg face misconnect package below in class face Crop.

FACE NET:

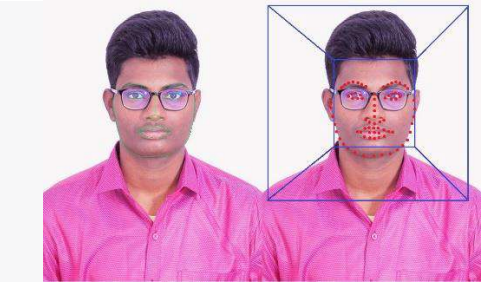
Face Net is a deep learning model for face recognition developed by researchers at Google. The below fig (C) explains the working of Face net. It's designed to directly learn a mapping from facial images to a compact Euclidean space, where the distance between points in this space corresponds to the similarity between the faces. Face Net leverages a triplet loss function to learn discriminative face embeddings that can be used for accurate face recognition and verification tasks.



Fig(c): Face Net

Dlib Face Recognition:

Dlib is an open-source suite of applications and libraries written in C++ under a permissive Boost license and algorithms for various computer vision tasks, including face recognition. It's known for its efficiency and effectiveness through Dlib also features robust tools for object pose estimation, object tracking, face detection and face recognition (identifying a perceived face). Dlib's face recognition capabilities are built upon deep learning techniques and classic machine learning algorithms. The below fig (d) describes about the Dlib's Face Recognition.



Fig(d): Dlib Face Recognition

DEEP ID:

"Deep ID" refers to a series of deep learning-based face recognition models developed by the Chinese University of Hong Kong (CUHK). Deep ID model has shown very promising results for 2D face recognition, Multi-task learning face recognition and attribute prediction does not improve performance. Deep ID models aim to learn compact and discriminative representations of faces, it seems reasonable to somehow integrate parts of this model in our general model. The below fig (e) describes about the Deep ID.

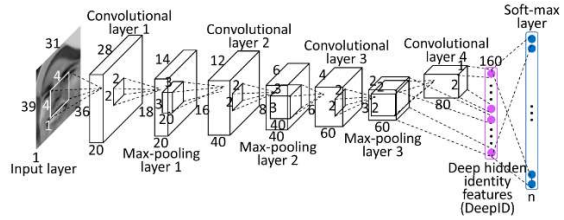


Fig (e): Deep ID

Table 1: Advantages and Disadvantages of Deep Learning Methods

Methods	Advantages	Disadvantages
Deep Learning	High Accuracy, Adaptability and Real-Time Processing	Data Requirements, Computational Resources
Open Face	Trained models, Scalability	Maintenance, Complexity and the Resource Intensive
Arc Face	Variations, Performance	Data Requirements, Complexity and Computationally Intensive.
VGG Face	Features, Transfer Learning	Efficiency, Limited Context and Computational Requirements
Face Net	Variations, Learning	Parameter Tuning, Training Data and Computational Demands
Dlib Face Recognition	Efficiency, Accuracy, Pre-Trained Models	It does not detect small faces in the image
Deep Id	Variations, Recognition, Pre-Trained Models	Over Fitting

The Table-1 tells us about the advantages and Disadvantages of the Deep Learning Methods.

Table 2: Based on the CNN- Based Face Recognition Research

Research	CNN Model	Face Database	Accuracy
Zafar et al[1]	B-DCCN	AT&T Face Database	100%
Bendjillali et al[2]	Inception-v3	CMU PIE database	99.89%.
Hu et al[3]	Inception-ResNet-v1	Oulu-CASIA NIR-VIS dataset	99.8%
Li et al[4.]	Multi-CNN	Nenu LD.	99.5%
Ding [5]	(TBE -CNN)	Pa SC database,	99.33%
Deng et al[6]	Inception-ResNet-v1	L F Wm	99.33%
Yang et al[7]	SR-CNN	LFW database	98.98%.
Nam et al [8]	(PSI-CNN)	LFW database	98.87%.
Peng et al[9]	Google Net	CASIA NIR database	98.48%
Nam et al[10]	(PSI-CNN)	LFW database and YTF dataset	98.37%.
Pei et al[11]	CNN VGG-16	Private	98.1%
Khan et al[12]	R-CNN	LFW database	97.9%.
Song et al [13]	Face-Net + SVM	Private	97%
Ben Fredj et al[14]	CNN	YTF dataset	96.83%
Hu et al[15]	CNN-L and CNN-S	NIR-VIS 2.0	95.77%
Ullah et al[16]	Deep Mask Net	MDMFR	93.33%
Chen et al[17]	CASIA-Web face	LFW and IJB- A database	93.1%
Zangeneh et al[18]	VGG-net model	UMIST and FERET datasets.	92.1%.
Moon et al[19]	CNN	Private	88.9%

The Table-2 shows the different researchers about the CNN Based Face recognition and also their accuracy by using different CNN model and Datasets.

Conclusion

This survey discusses face recognition based on the Convolutional Neural Network. Face recognition is one of the challenges in pattern recognition and computer vision and one of the many studies conducted in recent years. Several studies have been discussed trying to find renewal for face recognition. The renewed face recognition proposed various kinds of architecture used, modifying the images in the dataset or combining several methods. The main goal is to obtain a high level of accuracy so that the face recognition system has a high performance. This survey also discusses some face databases used in several studies. The face databases available have data from hundreds of images to millions. It is hoped that through this survey, readers can gain additional knowledge about face recognition based on Convolutional Neural Networks.

References

[1] U. Zafar et al., “Face recognition with Bayesian convolutional networks for robust surveillance systems,” EURASIP J. Image Video Process., vol. 2019, no. 1, p. 10, Dec. 2019.

[2] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, “Illumination-robust face recognition based on deep convolutional neural networks architectures,” Indones. J. Electr. Eng. Comput. Sci., vol. 18, no. 2, p. 1015, May 2020.Loss and Diversity Combination,” IEEE Access, vol. 7, pp. 75305–75317, 2019.

[3] G. Hu, X. Peng, Y. Yang, T. M. Hospedales, and J. Verbeek, "Frankenstein: Learning Deep Face Representations Using Small Data," IEEE Trans. Image Process., vol. 27, no. 1, pp. 293–303, 2018.

[4] P. Li, J. Xie, W. Yan, Z. Li, and G. Kuang, "Living Face Verification via Multi-CNNs," Int. J. Comput Intell. Syst., vol. 12, no. 1, p. 183, 2018

[5] C. Ding and D. Tao, "Trunk-Branch Ensemble Convolutional Neural Networks for Video-Based Face Recognition," IEEE Trans. Pattern Anal. Mach. Intell., vol. 40, no. 4, pp. 1002–1014, 2018.

[6] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "MFCosface: a masked-face recognition algorithm based on large margin cosine loss," Appl. Sci., vol. 11, no. 16, 2021.

[7] Y. X. Yang, C. Wen, K. Xie, F. Q. Wen, G. Q. Sheng, and X. G. Tang, "Face recognition using the SR-CNN model," Sensors (Switzerland), vol. 18, no. 12, 2018.

[8] G. P. Nam, H. Choi, J. Cho, and I. J. Kim, "PSI-CNN: A Pyramid-based scale-invariant cnn architecture for face recognition robust to various image resolutions," Appl. Sci., vol. 8, no. 9, 2018.

[9] M. Peng, C. Wang, T. Chen, and G. Liu, "NIRFaceNet: A Convolutional Neural Network for Near-Infrared Face Identification," Information, vol. 7, no. 4, p. 61, Oct. 2016.

[10] G. P. Nam, H. Choi, J. Cho, and I. J. Kim, "PSI-CNN: A Pyramid-based scale-invariant cnn architecture for face recognition robust to various image resolutions," Appl. Sci., vol. 8, no. 9, 2018.

[11] M. Z. Khan, S. Harous, S. U. Hassan, M. U. Ghani Khan, R. Iqbal, and S. Mumtaz, "Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing," IEEE Access, vol. 7, pp. 72622–72633, 2019.

[12] G. P. Nam, H. Choi, J. Cho, and I. J. Kim, "PSI-CNN: A Pyramid-based scale-invariant cnn architecture for face recognition robust to various image resolutions," Appl. Sci., vol. 8, no. 9, 2018.

[13] Z. Song, K. Nguyen, T. Nguyen, C. Cho, and J. Gao, "Spartan Face Mask Detection and Facial Recognition System," Healthc., vol. 10, no. 1, 2022.

[14] H. Ben Fredj, S. Bouguezzi, and C. Souani, "Face recognition in unconstrained environment with CNN," Vis. Comput., no. 0123456789, Jan. 2020.

[15] Z. Pei, H. Xu, Y. Zhang, M. Guo, and Y.-H. Yang, "Face Recognition via Deep Learning Using Data Augmentation Based on Orthogonal Experiments," Electronics, vol. 8, no. 10, p. 1088, Sep. 2019

[16] N. Ullah, A. Javed, M. Ali Ghazanfar, A. Alsufyani, and S. Bourouis, "A novel DeepMaskNet model for face mask detection and masked facial recognition," J. King Saud Univ. - Comput. Inf. Sci., no. 2022.

[17] G. Chen, Y. Shao, C. Tang, Z. Jin, and J. Zhang, "Deep transformation learning for face recognition in the unconstrained scene," Mach. Vis. Appl., vol. 29, no. 3, pp. 513–523, Apr. 2018.

[18] E. Zangeneh, M. Rahmati, and Y. Mohsenzadeh, "Low resolution face recognition using a two-branch deep convolutional neural network architecture," Expert Syst. Appl., vol. 139, p. 112854, 2020.

[19] H.-M. Moon, C. H. Seo, and S. B. Pan, "A face recognition system based on convolution neural network using multiple distance face," Soft Comput., vol. 21, no. 17, pp. 4995–5002, Sep. 2017.