International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-15 Issue-3, February 2026

Real-time Face Identification from Video in an Uncontrolled Environment using CNN

Patel Bhautika R., Desai Apurva A.

Abstract: ***We have witnessed a significant amount of fraud and security issues in modern life. Numerous biometric characteristics, such as the eyes, face, fingers, and palms, are used to address these problems. Among these, facial recognition is considered one of the least intrusive methods and is frequently used to identify or verify an individual. Face recognition is one of the most effective applications of computer vision, and has achieved considerable attention in recent years. Deep learning networks have achieved state-of-the-art performance in still-image-based face recognition. Video-based face recognition is a more complex task than still-image-based face recognition due to video quality, pose variation, occlusion, and illumination, and it also entails processing a large volume of data. We address these challenges by developing an efficient deep learning model trained, tested, and evaluated on the YouTube Face Dataset, designed for unconstrained face recognition in videos. In this paper, a deep learning face detection algorithm, Multi-task Cascaded Convolutional Neural Network (MTCNN), is employed to detect and localise faces in videos. Feature extraction and face recognition have been performed by using a convolutional neural network (CNN). This model has been proposed for accurate face detection and recognition from unconstrained video and performs better on the YouTube face dataset. The test accuracy of the proposed model is 93.11%. This work has been conducted to improve face recognition accuracy in the presence of intra-video variations.***

Keywords: Video-Based Face Recognition, Deep Learning, Convolutional Neural Network, Multi-Task Cascaded Convolutional Neural Network

Nomenclature:
CMS-RCNN: Contextual Multi-Scale Region-based Convolutional Neural Network
CNN: Convolutional Neural Network
DPSSD: Deep Pyramid Single Shot Face Detector
DSFD: Dual-Shot Face Detector
EBGM: Elastic Bunch Graph Matching
GPU: Graphics Processing Unit
HOG: Histogram of Oriented Gradients
LBCNN: Local Binary Convolutional Neural Network
LBP: Local Binary Patterns
LDA: Linear Discriminant Analysis
LFFD: Light and Fast Face Detector
LFW: Labelled Faces in the Wild
LSTM: Long Short-Term Memory
MTCNN: Multitask Cascaded Convolutional Networks
NLP: Natural Language Processing
PCA: Principal Component Analysis
S3FD: Single-Shot Scale-invariant Face Detector
SIFT: Scale Invariant Feature Transform
SSD: Single Shot MultiBox Detector
TBE-CNN: Trunk-Branch Ensemble CNN
TPU: Tensor Processing Unit; VGG - Visual Geometry Group
YOLO: You Only Look Once
YTF: YouTube Faces
ZF Net: Zeiler and Fergus Network

Manuscript received on 31 January 2026 | First Revised Manuscript received on 04 February 2026 | Second Revised Manuscript received on 09 February 2026 | Manuscript Accepted on 15 February 2026 | Manuscript published on 28 February 2026.

*Correspondence Author(s)
Patel Bhautika R.*, Assistant Professor, Smt. Tanuben & Dr. Manubhai Trivedi College of Information Science, Surat (Gujarat), India. Email ID: bhautika.patel@gmail.com, ORCID ID: 0000-0003-4617-4172
Desai Apurva A., Professor and Head, Department of Computer Science, Veer Narmad South Gujarat University, Surat (Gujarat), India. Email ID: aadesai@vnsgu.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open-access article under the CC-BY-NC-ND license http://creativecommons.org/licenses/by-nc-nd/4.0/

I. INTRODUCTION

Among various applications of computer vision, face recognition has received significant attention in recent years. Face recognition techniques can be classified as still-image-based or video-based, depending on whether recognition is performed on single images or video sequences. They also include appearance-based, feature-based, template-based, machine-learning-based, and modern deep-learning-based techniques, which differ in how facial features are extracted and matched. Face recognition using facial skeletal features has been proposed by Patel et al. [1], which employs a feature-based technique combined with machine learning. This approach exhibits justifiable performance but relies heavily on handcrafted features and is sensitive to significant variations in pose, illumination, and expression. In the paper by Zheng et al. [2], video-based face recognition is a popular area due to its wide range of applications, including video content analysis, surveillance, and access control. Compared with static image recognition, video-based face recognition is more challenging due to significant variations in pose, lighting conditions, and visual quality, as well as the large volume of data to be processed.

Despite variations in lighting, pose, occlusion, and distance, the human visual system can recognise faces within a fraction of a second, motivating advances in computer vision techniques that seek to mimic human perception. According to Goodfellow et al. [3], deep learning has made significant advances across domains such as computer vision, robotics, NLP, speech, and audio processing. In addition, deep learning models can extract deep features from images and inherently support classification tasks, as shown by Singhal et al. [4]. Computer vision is widely used in biometric applications, including face recognition.

A significant shift has been observed in video analysis toward Deep Convolutional Neural Networks, including face recognition. CNNs are biologically inspired deep learning architectures that learn spatial image features adaptively and automatically, as suggested by Yamashita et al. [5]. The study by Michalski et al. [6] reports that, unlike traditional neural networks, CNN architectures are generally

© Copyright: All rights reserved.

deeper; however, there is no strict definition specifying an exact threshold for network depth. In addition, CNN eliminates hand-crafted feature extraction by automatically learning feature representations during training. Moreover, model training is computationally expensive and typically requires a TPU or GPU.

Face detection forms the primary stage in a face recognition pipeline. It can be performed using classical approaches, despite deep learning–based face detection techniques having achieved state-of-the-art performance in the past years, such as MTCNN by Chou et al. [7], six degrees of freedom and face image pose estimation without face detection (img2pose) by Albiero et al. [8], DSFD by Li et al. [9], CMS-RCNN by Zhu et al. [10], Two-Stage cascaded convolutional neural network (Two-Stage CNN) by Yang et al. [11], S3FD by Zhang et al. [12], LFFD by He et al. [13]. For detecting small, blurred faces, a multi-scale detector yields efficient and reliable results.

The growing availability of large-scale video data in public and private domains has created a strong demand for robust, efficient, and scalable face recognition systems. Recent developments in deep learning techniques, including LSTM networks, CNNs, 3D-CNNs, and attention mechanisms, have shown promise in addressing these challenges. Therefore, this research aims to investigate approaches that effectively address the variability and constraints inherent in video data and improve accuracy and robustness. The primary goal of this research is to develop an efficient and effective real-time face recognition system for video in uncontrolled environments.

The remainder of the paper is organised as follows. Section 2 briefly reviews related works, including the use of deep learning in face recognition, face detection, and video-based face recognition. The proposed methodology, which covers face detection using MTCNN, data collection and preparation, and model architecture, is presented in Section 3. An experimental setup is given in Section 4, followed by model training and evaluation in Section 5, and concluding with a discussion of future work in Section 6.

II. RELATED WORKS

In recent years, face recognition from unconstrained video data has emerged as an active research area. Early contributions focused on hand-crafted features and statistical classifiers, but the introduction of large-scale benchmark datasets significantly accelerated progress in video-based face recognition. We review the literature in three parts: (i) deep learning for face recognition, (ii) face detection, and (iii) video-based face recognition.

A. Deep Learning for Face Recognition

With the affordable and powerful GPU and availability of a considerable amount of face databases, research on deep neural network-based face recognition has been accelerated, as mentioned by Shepley et al. [14]. With the development of highly challenging datasets such as MegaFace, WIDERFACE, and YTF, research in detection and recognition has also increased, achieving impressive accuracy. The performance of object detection and image recognition has advanced dramatically with the development of CNNs. Several CNN architectures, such as LaNet, AlexNet, ZF Net, VGG, GoogleNet, ResNet, DenseNet, and FaceNet, play a vital role in computer vision, including face recognition, as suggested by Jiwei et al. [15] and Ghosh et al. [16]. Although deep learning methods outperform traditional approaches, there remains a gap between human perception and current face recognition systems, particularly with respect to ageing effects, extreme pose variation, partial occlusion, and poor illumination.

B. Face Detection

Face detection is a primary step in face verification and recognition. There are many applications of face detection, including face tracking for surveillance purposes, as suggested by Kumar et al. [17], digital tagging on social media, and facial expression recognition. With the advent of Eigenfaces and Haar feature techniques in computer vision, researchers have been motivated to develop face recognition systems. Real-time face detection using the Viola-Jones boosting algorithm has become commonplace. Progressively, feature extraction techniques, including LBP, SIFT, and HoGs, have been developed. But these methods have limitations in extracting features under unconstrained conditions. To overcome these challenges, deep learning methods were developed, proposed by Ranjan et al. [18]. Some deep learning-based face detection approaches that achieve state-of-the-art performance include MTCNN, Two-Stage CNN, and DSFD.

C. Video-Based Face Recognition

The LBCNN was introduced by Ferraz et al. [19] to reduce the number of trainable parameters in CNN architectures for face recognition. A new method based on multi-modal feature fusion for iQIYI celebrity video identification was proposed by Chen et al. [20]. Face recognition using CNNs was proposed by Yi et al. [21] and evaluated on the CASIA-WebFace and LFW datasets. Zheng et al. [2] proposed face recognition from unconstrained video using techniques like SSD, DPSSD, and subspace-to-subspace similarity measures. A TBE-CNN model that can handle occlusion, pose variation, and blurred images in video-based face recognition has been developed by Ding et al. [22]. A study by Jiwei et al. [15] presented video-based face recognition using the deep learning algorithm FaceNet. Face recognition using techniques such as MTCNN for face detection, PCA, LDA, and EBGM for feature extraction, and deep learning for face recognition has been proposed by Acuña-Escobar et al. [23]. Lin et al. [24] have proposed a CNN-based model that uses the signReLU activation function to improve face recognition accuracy.

Although recent advancements in deep learning have significantly improved the performance of video-based face recognition, several critical challenges remain unresolved. First, many existing approaches use constrained, small-scale datasets to evaluate performance, which may be difficult during deployment. Second, recognition performance often degrades when dealing with low-quality or unconstrained videos that contain blurred images, occlusion, significant pose variations, and varying illumination. Third, most state-of-the-art models are heavily reliant on large-scale annotated datasets, which

© Copyright: All rights reserved.

limits their scalability and adaptability to real-world scenarios where labelled data are scarce. Finally, the accuracy is not sufficient for human perception.

III. PROPOSED METHODOLOGY

The proposed methodology presents a systematic framework for developing a video-based face recognition model using CNN. The approach aims to effectively capture spatial and temporal facial features from continuous video frames to enhance recognition accuracy under unconstrained environments. The workflow comprises several stages, including video data acquisition, frame extraction, preprocessing, feature learning through CNN architectures, and model evaluation. Each stage is designed to ensure robustness, computational efficiency, and adaptability to real-time recognition scenarios.

A. Data Collection

To conduct this study, a benchmark video dataset with significant intra-class variation is necessary. We therefore used the YTF database, created by Wolf et al. [25], for research on video-based face recognition. The dataset consists of 3,425 videos representing 1,595 distinct celebrities, all collected from YouTube. Table 1 presents the distribution of individuals in the dataset by the number of videos available per person. The video frames exhibit substantial variability, including differences in background, image size, illumination, and brightness conditions, as well as variations in facial pose and tilt.

Table 1: Distribution of Individuals Based on the Number of Available Video Clips

Videos	1	2	3	4	5	6
People	591	471	307	167	51	8

B. Face Detection using MTCNN

Face detection is a crucial preliminary stage in the face recognition process. An effective face recognition system requires that each module operate accurately and efficiently, particularly the face detection stage, which often limits overall performance. Errors in detection directly affect recognition accuracy. It can be performed using the classic feature-based cascade classifiers like Haar Cascade, LBP, HOG, etc., as well as modern deep learning-based approaches like DSFD, MTCNN, YOLO, SSD, RetinaFace, etc.

The proposed model employs MTCNN for face detection. This method is widely adopted for its state-of-the-art performance on several benchmark datasets and its effectiveness in unconstrained environments with variations in pose, illumination, and occlusion. In addition to face detection, MTCNN can identify key facial features, such as the eyes and mouth, via landmark detection, as described by Zhang et al. [26]. The architecture follows a cascaded design consisting of three networks: the Proposal Network (P-Net), the Refine Network (R-Net), and the Output Network (O-Net). Initially, the input image is resized to multiple scales to form an image pyramid. The P-Net generates candidate face regions; the R-Net refines and filters these bounding boxes; and the O-Net produces the final face detections along with facial landmark locations.

Figure 1 represents sample face detection results on the YTF dataset, showing correctly detected faces (dashed boxes) and missed detections using the MTCNN face detection method in the first row and second row, respectively. Some reasons that account for missed detection are shadow, occlusion, substantial pose variation, and blurriness of the face in the frame.

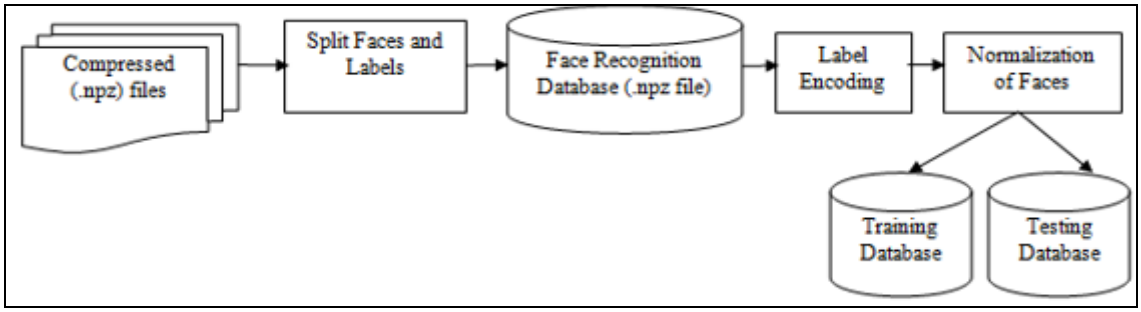


[Fig.1: Sample face Detection Outputs on the YTF Dataset, Illustrating Correctly Detected Instances (First Row) and Misdetections (Second Row)

C. Data Processing

We detected faces in each video frame using MTCNN and then resized them to 32 × 32. All these faces are compressed and stored in the form of an array. According to Hancock et al. [27], neural networks process real-valued vectors; qualitative attributes must be converted to numerical form before being used as network inputs. We performed label encoding to convert categorical variables into numeric form, and subsequently normalised all faces. Figure 2 shows data processing steps.

We used videos of 225 celebrities, comprising 84,868 video frames, which were further split into training and testing sets for model development and evaluation.



[Fig.2: Data Processing]

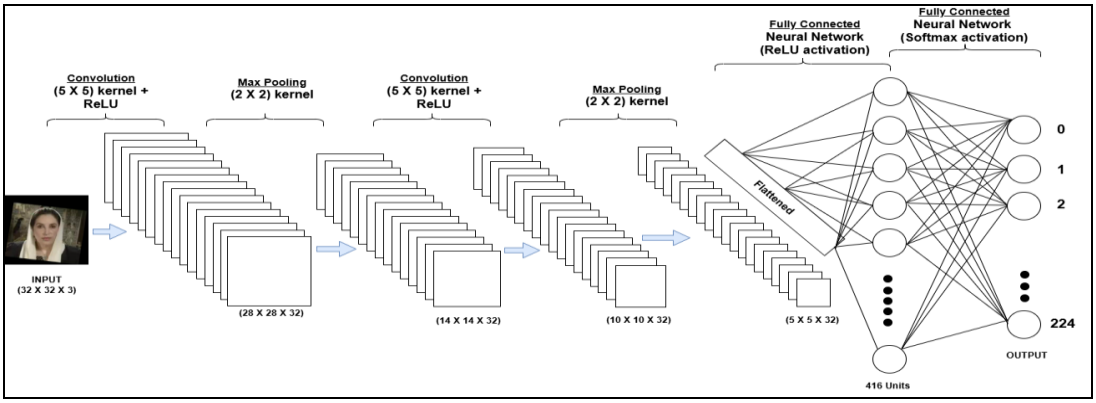
D. Model Architecture

The proposed system is built using a deep learning approach. In the study by Yamashita et al. [5], CNNs are particularly effective for addressing computer vision tasks. The proposed face recognition model comprises

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

an input layer, two convolutional layers, two pooling layers, and two fully connected layers. The input layer accepts 32×32×3 images, while the output layer is a classification layer with 225 classes, corresponding to YouTube celebrities. A non-linear activation function, ReLU, is applied to each convolutional layer along with Linear, Sigmoid, and Tanh as stated by Banerjee et al. [28]. According to Ghosh et al. [16], ReLU is particularly popular because it activates only a subset of neurons at a time, thereby reducing computational cost. Max pooling is employed during the pooling layers, which significantly affects model performance. The final fully connected layer serves as the classification output layer.

Figure 3 illustrates the detailed architecture of the proposed model. A detailed breakdown of the proposed model's layers is provided in Table 2, and an overall summary is given in Table 3.



[Fig.3: Architecture of the Proposed Model]

Table II: Layer-By-Layer Architecture Details of the Model

Type	Kernel size	Stride	Filters	Activation Function	Output Size
Convolution	5X5	1	32	ReLU	28X28X32
Max Pooling	2X2	2			14X14X32
Convolution	5X5	1	32	ReLU	10X10X32
Max Pooling	2X2	2			5X5X32
Vectorization			800		1X1X800
Fully Connected 1			416	ReLU	1X1X416
Fully Connected 2			225	Softmax	1X1X225

E. Hyper-parameter Tuning

In machine learning, hyperparameters are model parameters that are set before training and are not updated during training. Hyperparameter optimisation can be considered both the final phase of model design and the initial step in training a neural network. These parameters are mainly associated with either the training process or the model architecture. Yu et al. [29] identify several common deep learning hyperparameters, including the number of layers, the number of neurons per layer, the optimiser type, the learning rate, the loss and evaluation functions, the activation functions, and the weight initialisation methods. In this work, hyperparameter tuning is carried out using the Keras Tuner, as recommended by Joshi et al. [30]. Specifically, the Random Search tuner is employed to optimise the proposed model. Two hyperparameters are considered during tuning, namely the number of units in the first fully connected layer and the learning rate.

Table III: Model Summary

Layer Type	Output Shape	#Param
Convolution	(None, 28, 28, 32)	2432
MaxPooling	(None, 14, 14, 32)	0
Convolution	(None, 10, 10, 32)	25632
MaxPooling	(None, 5, 5, 32)	0
Flatten	(None, 800)	0
Dense	(None, 416)	333216
Dense	(None, 225)	93825
Total params: 4,55,105 Trainable params: 4,55,105 Non-trainable params: 0		

IV. EXPERIMENTAL SETUP

The proposed model was implemented in Python, and all experiments were conducted on Google Colab Premium, which provides access to high-performance cloud-based GPUs for efficient training and evaluation. The computational environment comprised 25 GB of RAM, an Intel Xeon CPU backend, and an NVIDIA Tesla T4 GPU with 16–32 GB of memory. This configuration enabled faster model training and enabled effective processing of large datasets without being constrained by local hardware resources.

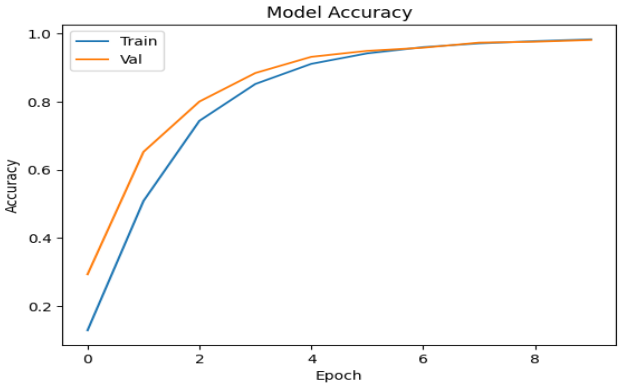
The implementation was performed using TensorFlow as the main deep learning framework. TensorFlow was selected for its robust support for CNN architectures, built-in GPU acceleration, and suitability for large-scale model training, as noted by Abadi et al. [31]. Additional libraries, including Keras, NumPy, Pandas, and Matplotlib, were used to facilitate data preprocessing, model development, and visualisation of results.

V. MODEL TRAINING AND EVALUATION

The dataset was preprocessed and partitioned into training, validation, and testing subsets to ensure an unbiased evaluation of the model. The proposed model was trained on 65,842 images of YouTube celebrities using a batch size of 256 over 10 epochs. The model was trained using the Adam optimiser with categorical cross-entropy as the loss function, with a learning rate of 0.0001. The total number of trainable parameters, including weights and biases updated during training, is 4,55,105. The validation accuracy of the proposed model is shown in Figure 4, achieving a best training accuracy of 0.9833 and a validation accuracy of 0.9814.

Retrieval Number: 100.1/ijitee.A371016010326
DOI: 10.35940/ijitee.A3710.15030226
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.



[Fig.4: Validation Accuracy of the Proposed Model]

We evaluated our video-based face recognition model on 19,026 video frames, achieving a test accuracy of 0.9311 and a loss of 0.2024. Based on the classification report, the average precision, recall, and F1-score are 0.9239, 0.9159, and 0.9193, respectively.

The face recognition results produced by the proposed system are shown in Figure 5. Successful identification of faces across a wide range of poses and lighting conditions demonstrates the model's robustness. Nevertheless, a few misclassifications were observed, particularly in frames that were heavily occluded or of low resolution. The images on the left demonstrate cases in which the model successfully recognised the individual despite variations in facial pose, lighting, and background conditions. Conversely, the photos on the right highlight misclassified samples in which the model incorrectly assigned the input face to another identity. These errors mainly occurred under challenging conditions, such as partial occlusion, low-resolution video frames, or extreme variations in facial appearance.



[Fig.5: Sample Recognition Outputs on the YTF Dataset, Illustrating Correctly Recognized Instances (left) and Misclassifications (Right)]

VI. CONCLUSION AND FUTURE WORK

This paper proposes a new video-based face recognition framework using a CNN and evaluates it on the unconstrained YTF dataset. Firstly, we employed MTCNN, a deep learning-based face detection model, to detect faces because it gives high accuracy and is robust to different poses, occlusion, facial expression, and illumination changes. Secondly, we applied a CNN, the backbone of modern computer vision applications, to perform face recognition. We have employed deep learning techniques for both face detection and recognition, and as a result, we achieve improved face recognition performance on the challenging YTF face dataset. The effectiveness of our proposed model is comparable to that of similar frameworks that use the YTF dataset for face recognition. The proposed model achieves competitive results compared to existing methods. While earlier models laid the foundation for video-based face recognition, the proposed model achieves strong performance in both accuracy and F1-score, highlighting its robustness to intra-video variations, including occlusion, pose, and illumination. In the future, we can enhance this model by incorporating liveliness detection to distinguish between fake and real faces, and, after recognition, we can also track the person and their activities.

ACKNOWLEDGMENT

The authors acknowledge the use of publicly available face-related video resources from YouTube for dataset preparation in this research. Google Scholar was used extensively to access relevant scholarly literature. The computational experiments were conducted in Google Colab, which provided the necessary computing infrastructure. Grammarly was utilised to assist in the language refinement of the manuscript. The authors also express their sincere gratitude to their respective institutions for providing the academic environment and support required to conduct this research.

DECLARATION STATEMENT

The authors declare that the work presented in this manuscript is original and has not been published previously, nor is it under consideration for publication elsewhere. All data, figures, and results are obtained through honest research practices, and appropriate references have been cited wherever necessary. The authors confirm that there is no conflict of interest related to this work. Additionally, the reference [25] is older and explicitly noted as such. Nonetheless, this work remains essential to the current study, as it is pioneering in its field.

After aggregating input from all authors, I must verify the accuracy of the following information as the article's author.

- **Conflicts of Interest/ Competing Interests:** Based on my understanding, this article has no conflicts of interest.
- **Funding Support:** This article has not been funded by any organizations or agencies. This independence ensures that the research is conducted objectively and free from external influence.
- **Ethical Approval and Consent to Participate:** The content of this article does not necessitate ethical approval or consent to participate with supporting documentation.
- **Data Access Statement and Material Availability:** The adequate resources of this article are publicly accessible.
- **Author's Contributions:** The authorship of this article is contributed equally to all participating individuals.

REFERENCES

1. Bhautika, R. Patel, and A. Desai Apurva. "Face Identification through Facial Skeletal Features." Futuristic Trends in Networks and Computing Technologies: Select Proceedings of Fourth

© Copyright: All rights reserved.

International Conference on FTNCT 2021. Singapore: Springer Nature Singapore, (2022): 921-933.
URL: https://doi.org/10.1007/978-981-19-5037-7_66
2. Zheng, Jingxiao, et al. "An automatic system for unconstrained video-based face recognition." IEEE Transactions on Biometrics, Behaviour, and Identity Science 2.3 (2020): 194-209.
DOI: https://doi.org/10.1109/TBIOM.2020.2973504
3. Goodfellow, Ian, et al. Deep learning. Vol. 1. No. 2. Cambridge: MIT Press, 2016. DOI: https://doi.org/10.4258/hir.2016.22.4.351
4. Singhal, Soniya, Madasu Hanmandlu, and Shantaram Vasikarla. "Video-based face recognition with new classifiers." Journal of Modern Physics 12.03 (2021): 361.
DOI: https://doi.org/10.4236/jmp.2021.123026
5. Yamashita, Rikiya, et al. "Convolutional neural networks: an overview and application in radiology." Insights into imaging 9.4 (2018): 611-629. DOI: https://doi.org/10.1007/s13244-018-0639-9
6. Michalski, Paweł, Bogdan Ruszczak and Michał Tomaszewski. "Convolutional neural networks implementations for computer vision." International Scientific Conference BCI 2018 Opole. Cham: Springer International Publishing, 2018.
DOI: https://doi.org/10.1007/978-3-319-75025-5_10
7. Chou, Kuan-Yu, et al. "Multi-task cascaded and densely connected convolutional networks applied to human face detection and facial expression recognition system." 2019 International Automatic Control Conference (CACS). IEEE, 2019.
DOI: https://doi.org/10.1109/CACS47674.2019.9024357
8. Albiero, Vitor, et al. "img2pose: Face alignment and detection via 6dof, face pose estimation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021.
DOI: https://doi.org/10.48550/arXiv.2012.07791
9. Li, Jian, et al. "DSFD: dual shot face detector." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019. DOI: https://doi.org/10.1109/CVPR.2019.00520
10. Zhu, Chenchen, et al. "CMS-RCNN: contextual multi-scale region-based CNN for unconstrained face detection." Deep learning for biometrics. Cham: Springer International Publishing, 2017. 57-79.
DOI: https://doi.org/10.1007/978-3-319-61657-5_3
11. Yang, Shuo, et al. "Wider face: A face detection benchmark." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. DOI: https://doi.org/10.1109/CVPR.2016.596
12. Zhang, Shifeng, et al. "S3FD: Single shot scale-invariant face detector." Proceedings of the IEEE international conference on computer vision. 2017. DOI: https://doi.org/10.1109/ICCV.2017.30
13. He, Yonghao, et al. "Lffd: A light and fast face detector for edge devices." arXiv preprint arXiv:1904.10633 (2019).
DOI: https://doi.org/10.48550/arXiv.1904.10633
14. Shepley, Andrew Jason. "Deep learning for face recognition: a critical analysis." arXiv preprint arXiv:1907.12739 (2019).
DOI: https://doi.org/10.48550/arXiv.1907.12739
15. Jiwei, Xue, et al. "Research on video face retrieval method based on deep learning and key frame." Proceedings of the 2020 4th International Conference on Digital Signal Processing. 2020.
DOI: https://doi.org/10.1145/3408127.3408199
16. Ghosh, Anirudha, et al. "Fundamental concepts of convolutional neural network." Recent trends and advances in artificial intelligence and the Internet of Things. Cham: Springer International Publishing, 2019. 519-567. DOI: https://doi.org/10.1007/978-3-030-32644-9_36
17. Kumar, Deepak, et al. "Automated panning of video devices." 2017 International Conference on Signal Processing and Communication (ICSPC). IEEE, 2017.
DOI: https://doi.org/10.1109/CSPC.2017.8305867
18. Ranjan, Rajeev, et al. "Deep learning for understanding faces: Machines may be just as good, or better, than humans." IEEE Signal Processing Magazine 35.1 (2018): 66-83.
DOI: https://doi.org/10.1109/MSP.2017.2764116
19. Ferraz, Carolina Toledo, and José Hiroki Saito. "A comprehensive analysis of local binary convolutional neural network for fast face recognition in surveillance video." Proceedings of the 24th Brazilian Symposium on Multimedia and the Web. 2018.
DOI: https://doi.org/10.1145/3243082.3267444
20. Chen, Jianrong, et al. "A novel deep multi-modal feature fusion method for celebrity video identification." Proceedings of the 27th ACM International Conference on Multimedia. 2019
DOI: https://doi.org/10.1145/3343031.3356067.
21. Yi, Xin, and Xingcheng Luo. "A system for real-time detecting and recognising a person." Proceedings of the 4th International Conference on Communication and Information Processing. 2018.
DOI: https://doi.org/10.1145/3290420.3290452
22. Ding, Changxing, and Dacheng Tao. "Trunk-branch ensemble convolutional neural networks for video-based face recognition." IEEE transactions on pattern analysis and machine intelligence 40.4 (2017): 1002-1014. DOI: https://doi.org/10.1109/TPAMI.2017.2700390
23. Acuña-Escobar, Diego, Julio Ibarra-Fiallo y Monserrate Intriago-Pazmiño. "Real-time face identification from video surveillance cameras." Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems. 2019. DOI: https://doi.org/10.1145/3368691.3368737
24. Lin, Guifang, and Wei Shen. "Research on convolutional neural network based on improved Relu piecewise activation function." Procedia Computer Science 131 (2018): 977-984.
DOI: https://doi.org/10.1016/j.procs.2018.04.239
25. Wolf, Lior, Tal Hassner, and Itay Maoz. "Face recognition in unconstrained videos with matched background similarity." CVPR 2011. IEEE, 2011. DOI: https://doi.org/10.1109/CVPR.2011.5995566, works remain significant, see the declaration
26. Zhang, Kaipeng, et al. "Joint face detection and alignment using multitask cascaded convolutional networks." IEEE signal processing letters 23.10 (2016): 1499-1503.
DOI: https://doi.org/10.1109/LSP.2016.2603342
27. Hancock, John T., and Taghi M. Khoshgoftaar. "Survey on categorical data for neural networks." Journal of Big Data 7.1 (2020): 28.
DOI: https://doi.org/10.1186/s40537-020-00305-w
28. Banerjee, Chaity, Tathagata Mukherjee, and Eduardo Pasiliao Jr. "An empirical study on generalizations of the ReLU activation function." Proceedings of the 2019 ACM Southeast Conference. 2019.
DOI: https://doi.org/10.1145/3299815.3314450
29. Yu, Tong, and Hong Zhu. "Hyper-parameter optimization: A review of algorithms and applications." arXiv preprint arXiv:2003.05689 (2020).
DOI: https://doi.org/10.48550/arXiv.2003.05689
30. Joshi, Sharmad, et al. "Analysis of preprocessing techniques, Keras tuner, and transfer learning on cloud street image data." 2021 IEEE International Conference on Big Data (Big Data). IEEE, 2021.
DOI: https://doi.org/10.1109/BigData52589.2021.9671878
31. Abadi, Martín, et al. "TensorFlow: a system for Large-Scale machine learning." 12th USENIX symposium on operating systems design and implementation (OSDI 16). 2016.
DOI: https://doi.org/10.48550/arXiv.1605.08695

AUTHOR'S PROFILE



Bhautika Patel is an Assistant Professor in the Smt. Tanuben & Dr. Manubhai Trivedi College of Information Science, Surat. She received her M.Phil, M.Sc (I.C.T.), and B.Sc (C.S.) degrees in Computer Science from Veer Narmad South Gujarat University (V.N.S.G.U.) and is currently pursuing a PhD at the same university. She also qualified for the SLET (State Level Eligibility Test) in 2013. She has over 17 years of teaching experience in undergraduate computer science courses. Her research interests include deep learning, computer vision, and face recognition. She has published 3 research papers in reputable journals and conference proceedings and has contributed to an academic book. Her strengths include strong conceptual knowledge, effective classroom management and student mentoring, and a keen interest in research and educational development.



Prof. Desai has been serving as a Professor and Head in the Department of Computer Science at Veer Narmad South Gujarat University. He has over 35 years of teaching experience. He earned his PhD in 1997 and has successfully supervised more than 15 research scholars for their PhD. He has also

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)
© Copyright: All rights reserved.

delivered several lectures as a resource person in refresher and orientation courses, workshops, training programs, etc. and also delivered invited talks at national and international conferences. Dr Desai has three books and about ninety research papers to his credit. He had designed a university campus network and had also extended his services to "Safe City Surat," a citywide CCTV project owned by the Surat Police. Dr Desai served as dean of the Computer Science and Information Technology faculty, as a member of the Academic Council of Veer Narmad South Gujarat University, and as Chairman of the Board of Studies. Dr Desai has completed four funded projects. He has visited Canada, Italy, and the USA to attend various academic programs. His research area includes Natural Language Processing, Optical Character Recognition, Image Processing, Computer Graphics, and Machine Learning.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of the Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)/ journal and/or the editor(s). The Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP) and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.