International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Smart Attendance System Using Computer Vision

Chippada Ram Prasad[1], Satish Muppidi[2]

[1]Undergraduate Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India ramprasadchippada5@gmail.com
[2]Associate Professor Dept. of CSE, GMR Institute of Technology, Rajam, Andhra Pradesh, India satish.m@gmrit.edu.in

ABSTRACT:

Daily attendance marking is a common and important activity in schools and colleges for checking the performance of students. Manual Attendance maintaining is difficult to process, especially for a large group of students. Traditional attendance systems employ's methods to identify a face from the given input but the results are not usually accurate and precise as desired. The system described for aim to deviate from such traditional systems and introduce a new approach to identify a student using a face recognition system. The limitations of the traditional attendance monitoring system stimulated the adoption of "computer vision" to stand in the gap. Student's attendance can be monitored with biometric candidate's systems such as iris recognition system and face recognition system. Among these, face recognition has a greater potential because of its non-intrusive nature. Using computer vision along with "deep learning" concepts (CNN related) it can detect the faces of the student and record their attendance on the register in digitalized format. Initially, we need to register the data (faces) of the students in database .so, whenever faces recognized by computer vision uses the data to detect their faces and updates their attendance register.

Keywords: *Computer Vision, Deep Learning, Attendance management, CNN, Face Recognition, Database, Artificial Intelligence*.

Introduction:

In today's fast-paced world, managing attendance has become more than just a routine administrative task; it has evolved into a dynamic process that requires heightened security and precision. Conventional methods of attendance tracking, like manual sign-ins or card swiping, are not only outdated but also vulnerable to fraudulent practices. To address these challenges and usher in an era of unparalleled accuracy and security, the integration of Face Recognition, coupled with Deep Learning techniques, emerges as a game-changing solution.

A Smart Attendance System that leverages Face Recognition, with a specific focus on advanced Deep Learning techniques for Face Spoof Detection and Morphing Face Detection. By seamlessly combining the capabilities of modern computer vision and artificial intelligence. We will develop deep into the intricate workings of Face Recognition, exploring the fundamentals of Deep Learning, Convolutional Neural Networks (CNNs), and their applications in attendance tracking. But the spotlight of our exploration will be on the sophisticated methods used to detect not only real faces but also the cunning manipulations aimed at subverting the system.

The fusion of Deep Learning with Face Recognition brings forth a level of sophistication that enables not just face recognition but also discernment between genuine faces and fraudulent attempts. By addressing the critical challenges of Face Spoofing and Morphing Face Detection, this technology ensures not only the integrity of attendance records but also reinforces security measures across various domains.

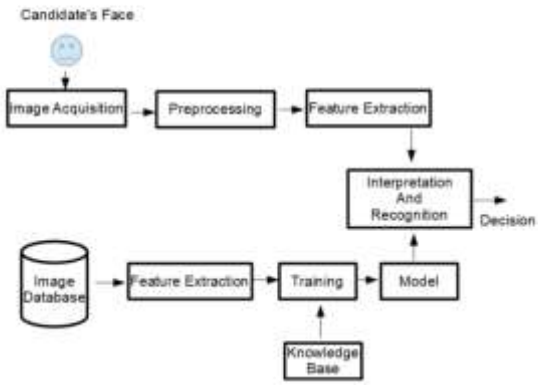

Fig. 1. Process of Attendance Making

The system relies on the robust DCNN VGG-16 and MTCNN frameworks to optimize face recognition and detection processes. The inclusion of VGG-16, renowned for its deep learning capabilities, ensures high precision in identifying faces under diverse conditions. Complementarily, the Multi-task Cascaded Convolutional Networks (MTCNN) architecture introduces multi-stage face detection, effectively filtering non-face regions and enhancing accuracy.

Literature Survey:

In recent years, the field of face recognition systems has witnessed significant advancements, primarily driven by the integration of technologies such as deep learning, machine learning, and the Internet of Things (IoT). This literature review aims to provide a comprehensive understanding of the current landscape of face recognition, covering various aspects including detection, spoof detection, attendance systems, anti-spoofing measures, and the emerging field of 3D face recognition. By synthesizing findings from multiple studies, this review seeks to illuminate the methodologies employed, key findings, and potential avenues for future research in this dynamic and rapidly evolving field.

Numerous studies have contributed to the exploration of face detection and recognition models. Warman and Kusuma (2023) conducted an extensive evaluation of models such as RetinaFace, MTCNN, FaceNet, and ArcFace, focusing on their accuracy and speed. Gupta et al. (2020) introduced an automated attendance system that combines classical computer vision techniques with a dual approach using LBPH and Convolutional Neural Network (CNN) models for face recognition. Suresh et al. (2019) addressed the challenges of real-time multiple face detection in crowded classrooms, potentially utilizing convolutional neural networks (CNNs).

Another significant area of research is face spoof detection. Balamurali et al. (2021) explored VGG-Face architecture for face spoof detection, achieving impressive test accuracy. Qi et al. (2023) extended the focus to anti-spoofing methods, incorporating Bidirectional Long Short-Term Memory (BiLSTM) networks for temporal analysis, providing a holistic approach to facial security.

The detection of morphed face images has been extensively explored by Raoof et al. (2020), introducing a comprehensive taxonomy for Morph Attack Detection (MAD) algorithms. Tapia and Busch (2021) emphasized the role of feature selection in high-accuracy face morphing detection, combining it with advanced Deep Learning (DL) methods.

Dang (2023) introduced a smart attendance system based on improved facial recognition, leveraging the MobileNetV2 architecture and IoT technology. Hasan et al. (2021) provided foundational insights into IoT integration in facial recognition systems, highlighting the synergy between machine learning and IoT for efficient real-time face recognition.

Jabberi et al. (2023) contributed to the field by applying 3D ShapeNets to 3D face recognition, asserting its superiority over 2D counterparts. Mishra and Singh (2022) introduced the Sf3CNN framework, utilizing a 3D ResNet architecture and A-Softmax loss function for face recognition in video data.

In the domain of DeepFake detection, Malik et al. (2022) conducted a comprehensive survey, evaluating various methods and providing insights into the versatility of Convolutional Neural Network (CNN) architecture. Hamza et al. (2022) focused on the generation and detection of face morphing attacks, contributing to ongoing efforts to fortify facial recognition technologies.

Looking ahead, future research directions, as highlighted by Hasan et al. (2021) and Dang (2023), may include further exploration of IoT technology integration, enhancing system robustness, exploring novel architectures, and improving adaptability to dynamic environments. The literature collectively underscores the continuous evolution of face recognition systems, combining traditional computer vision techniques with cutting-edge deep learning frameworks, addressing challenges in real-time scenarios, and pushing the boundaries of what is possible in the field.

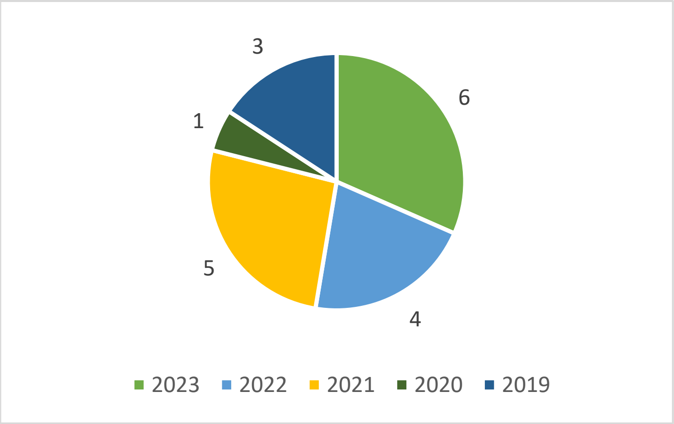

Fig. 2. Year of Publishing Vs No of Papers

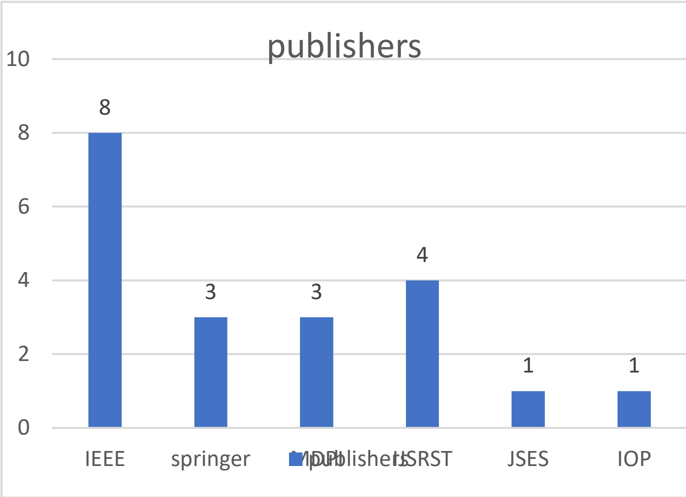

Fig. 3. Publisher Vs No of papers

Methodology:

Data Augmentation:

In a deep learning approach, the size of the data set is an important component. If the data set size is small, then the possibilities of over fitting may arise. To overcome this problem, data augmentation technique is used to increase the size of the data set. Applied an image transformation technique flipping and shifting to generate a new image from the available data set. The more generalize capability of a machine learning model can be achieved by training with the original image and with the augmented image. The images are transformed with the opensource programming language python and Keras library.

Among the various augmentation techniques, geometrical transformations and Image filters are used for better results.

Geometrical Transformation:

Image Zoom: Adjusting the scale of an image, either zooming in or out, to introduce variations in object sizes.

Translation: Shifting an image horizontally or vertically, simulating changes in object positions.

Rotation: Rotating the image at different angles, helping the model adapt to variations in orientation.

Brightness Adjustments: Modifying the brightness level of images to account for varying lighting conditions.

Contrast, Saturation, Scaling, Cropping, Adding Noise, etc….

Fig. 4. Geometrical Transformations

Image Filters:

Mean Filter: A filter that replaces each pixel's value with the average value of its neighbouring pixels, smoothing out the image.

Median Filter: A filter that replaces each pixel's value with the median value of its neighbouring pixels, effective in reducing noise.

Gaussian Filter: A filter that blurs the image by applying a convolution operation with a Gaussian kernel, reducing high-frequency noise.

Bilateral Filter: A filter that smoothens images while preserving edges, maintaining a balance between noise reduction and edge preservation.

(a) (b) (c) (d) (e)

Fig. 5. Filter operation. (a) The original face image. (b) Result of using a mean filter. (c) Result of using a median filter. (d) Result of using a Gaussian filter. (e) Result of using a bilateral filter.

Deep learning Algorithms:

Architecture of Deep CNN:

Deep convolutional neural networks (DCNNs) are a type of neural network that are commonly used to identify patterns in images and video. Deep CNN algorithm is used for image classification and face recognition along with face-Spoofing detection. It contains Input layer, Convolution layer, Activation function, Max-Pooling layer, Fully-Connected (Dense) layer, Dropout layer, SoftMax layer, Output layer. These layers work together to transform raw input data into meaningful representations and make predictions. The architecture and the number of layers may vary depending on the specific DCNN model and its application. Popular DCNN architectures include VGG-Net, Res-Net, and Inception.

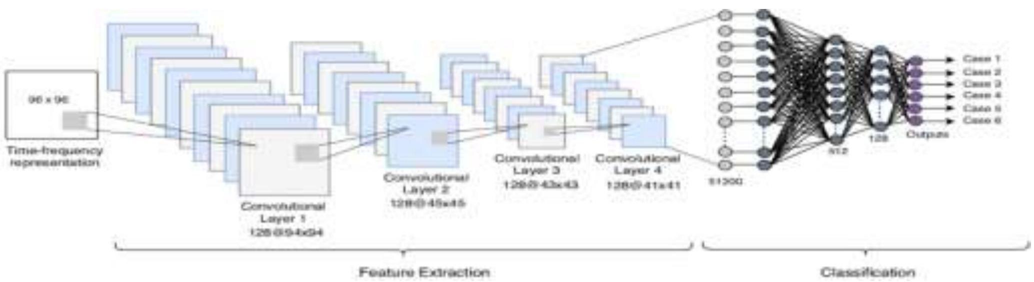

Fig. 6. Architecture of DCNN

Convolution layer: The convolution layer extract feature from the input image by applying filters. They are the foundation of CNN, and they are in charge of executing convolution operations. The Kernel/Filter is the component in this layer that performs the convolution operation. Until the complete image is scanned, the kernel makes horizontal and vertical adjustments dependent on the stride rate. The kernel is less in size than a picture, but it has more depth. This means that if the image has three (RGB) channels, the kernel height and width will be modest spatially, but the depth will span all three.

Max-Pooling Layer: It was used to reduce the dimension of the filtered image at each convolution layer. Therefore, this layer focused on the concerned object in the image and best feature of the image. In this study at each convolution layer, a max-pooling layer of size 2x2 has been applied. This layer is in charge of reducing dimensionality. It aids in reducing the amount of computing power required to process the data.

Flatten Layer: This layer reduced the 2- dimensional feature vector into an array that is feed to a fully connected layer.

Fully Connected Layer: This layer is also known as fully dense layer. The fully connected layer works with a flattened input, which means that each input is coupled to every neuron. After that, the flattened vector is sent via a few additional FC layers, where the mathematical functional operations are normally performed. The classification procedure gets started at this point. FC layers are frequently found near the end of CNN architectures.

Activation Function: The last fully connected layer's activation function is frequently distinct from the others. Each activity necessitates the selection of an appropriate activation function. The Softmax function, which normalizes output real values from the last fully connected layer to target class probabilities, where each value ranges between 0 and 1 and all values total to 1, is an activation function used in the multiclass classification problem.

The activation function ReLU has been implemented in each layer. In the dense layer two activation functions SoftMax have been used to one at a time.

ARCHITECTURE OF VGG-16:

VGG16 is a convolutional neural network model that's used for image recognition. It's unique in that it has only 16 layers that have weights, as opposed to relying on a large number of hyper-parameters. VGG-16 is the most commonly used version of CNN. It has a total of 16 layers with 13 convolutional and 3 fully connected layers. VGG-16 has introduced the deeper way of designing the CNN. It uses ReLU as an activation function to improve the nonlinearity in the model, whereas the Softmax function is used at the final layers for classification. The main contribution of the VGG16 architecture is the introduction of small kernel size. This network stacks more layers to make the architecture deeper and uses small-sized filters. Training dataset is VGG-Face dataset with 2.6 million RBG images with 224*224 pixels. This VGG-16 is a Deep CNN algorithm with 13 convolution layers, 5 pooling layers, 3 fully-connected layers. Where, last fully-connected has 54 channels because, there are 54 identities in the class.

In VGG-16 architecture Spatial resolution is preserved after convolution with spatial padding of 1 and convolutional stride of 1. Downsampling is carried by 5 max pooling. ReLU activation functions are used in convolutional and fully-connected layers. Softmax function is used in final layer.

Process of the Attendance Taking:

1. The instructor takes photos of students at the beginning of several classes in a term.
2. After each class, a single image with all students' faces is captured.
3. The instructor submits the images to an attendance-taking website.
4. Students log in to the website and choose their faces in the images.
5. Students annotate their chosen faces with their IDs.
6. This annotation process simplifies training data creation and helps identify students in subsequent class images.

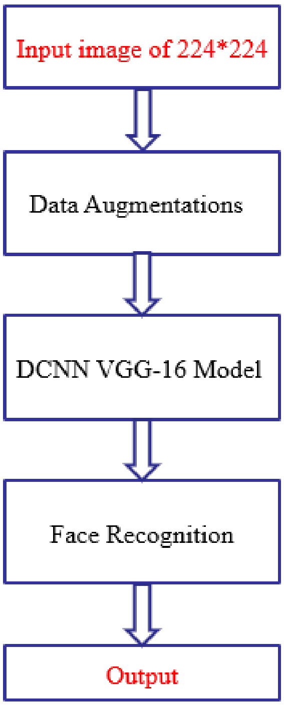

Fig. 7. Process of Face Recognition

Process of the Face Detection:

1. Original training sample from class pictures are acquired using face detection algorithm.
2. Data augmentation is used.
3. DCNN model is trained.
4. Unknown image is input and used in model then output will be the name of the student.

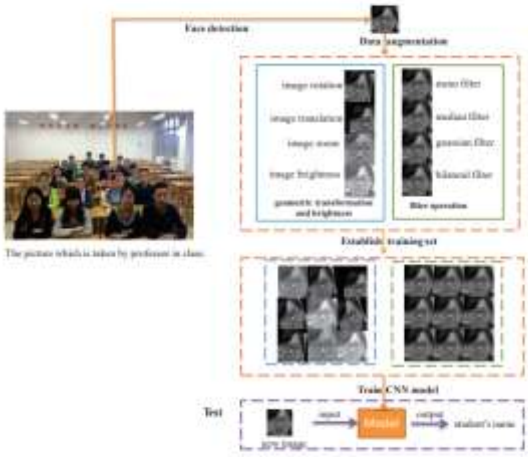

Fig. 8. The workflow of our class attendance-taking method. First, the instructor takes a photo of the class with all the students' faces. Then, face detection is used to capture each student's face. Data augmentation is performed to increase the number of training images in the dataset. Finally, the convolutional neural network (CNN) can be trained to generate a model that can be used to predict a student's name.

Process:

The output feature map of the convolutional layer as C,

$$C = \phi(H(x,y)),$$

where φ(.) denotes the ReLU function φ(H(x, y)) = max(0, H(x, y)) and

$$H(x,y) = \sum_{m,n \in s} W(m,n) I_i'(x+m, y+n) + b,$$

W=Weight matrix of kernel

b=Bias

Φ(.)=activation function

Activation function can reduce the computation and accelerate the convergence of the network.

A pooling layer is used to reduce the spatial size and the number of parameters in the network. It can prevent overfitting

The output map of the pooling layer denoted as P can be calculated by

P = g(C),

g(.) denotes the function to calculate the max value.

As the window moves across C, g(.) selects the largest value in the window and discards the rest.

Droupout layer neurons are "drooped out" and do not contribute to the forward propagation and back-propagation, which helps prevent overfitting.

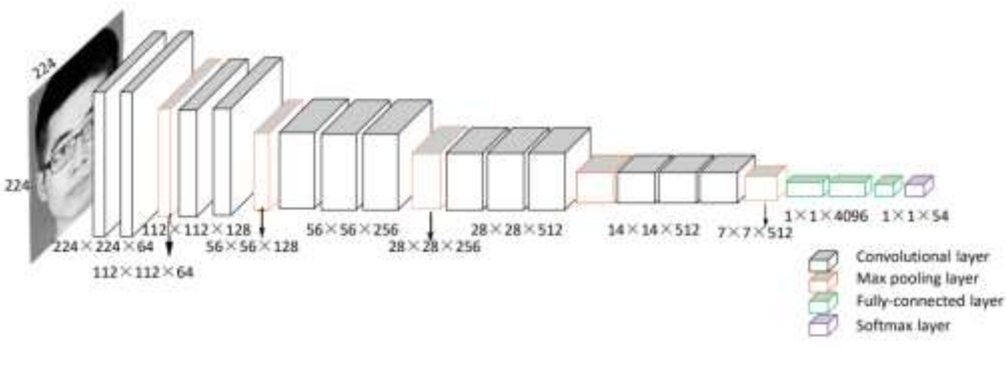

Fig. 9. VGG-16 Architecture demonstration with its layers

Output of the fully-connected layer at neuron q denoted as Fq is computed as

$$F_q = \phi(\sum_{m,n \in s} W(m,n)P(x,y) + b).$$

SoftMax-loss function denoted as L is used as the network's loss function, and our model is trained with the MBGD (Mini-Batch Gradient Descent) method

$$L = -\frac{1}{M}\sum_{i=1}^{M}\sum_{q}^{J} T_i \log(p_q),$$

M = number of images in a batch of one iteration

pq= output of the network at neuron q

Probability of the model's prediction which can be calculated by the SoftMax function,

$$p_q = \frac{\exp(F_q)}{\sum_{Z=1}^{J} \exp(F_Z)}.$$

The VGG-16 model is pretrained with VGG-Face dataset. They use 3538 students' face images for fine-tuning and 372 face images for validating and addition to this 5-fold validation set is used for more accuracy.

Architecture of MTCNN:

MTCNN (Multi-task Cascaded Convolutional Networks) is a popular deep learning model designed for face detection and facial feature alignment.

Face Detection:

MTCNN is primarily used for face detection. The model is capable of detecting multiple faces in an image, providing bounding box coordinates and confidence scores for each detected face.

Facial Landmark Detection:

In addition to face detection, MTCNN can predict facial landmarks such as eye corners, nose tip, and mouth corners. This feature is valuable for tasks like face alignment.

The main advantage of MTCNN is The size of input images of MTCNN can be any size. Given an image, we often resize it to different scales to build an image pyramid as the inputs of the following three-stage cascaded framework.

The main tasks for MTCNN Model is:

1. Face classification
2. Bounding box regression.
3. Facial landmark localization.

Stages in MTCNN:

1. Proposal Network (P-Net):

Task: The primary function of P-Net is to generate candidate bounding boxes (proposals) likely to contain faces.

Architecture: P-Net is a fully convolutional network (FCN) that scans the input image with a small convolutional kernel to predict face/non-face scores and bounding box regression offsets.

Output: It produces a set of bounding box candidates, each associated with a face probability score.

2. Refinement Network (R-Net):

Task: R-Net takes the bounding box candidates from P-Net and refines them. It also performs face/non-face classification and bounding box regression.

Architecture: Similar to P-Net, R-Net is also a fully convolutional network but with a larger receptive field. It refines the bounding box proposals and further improves the accuracy of face detection.

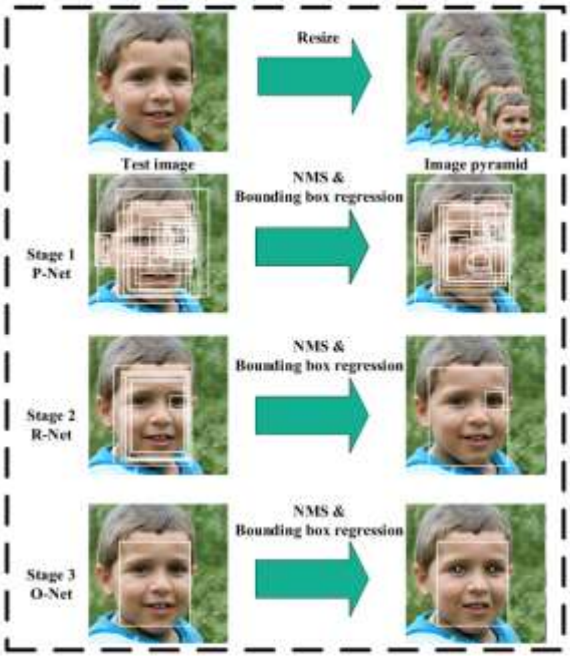

Fig. 10. Stages of MTCNN

Output: R-Net outputs refined bounding boxes and face probability scores.

3. Output Network (O-Net):

Task: O-Net is the final stage, responsible for refining and filtering the bounding boxes further. It also performs facial landmark localization, identifying key facial points.

Architecture: O-Net is the most complex of the three networks. It combines both convolutional and fully connected layers to handle more detailed tasks such as facial landmark localization.

Output: O-Net produces the final set of bounding boxes, facial landmarks, and face probabilities.

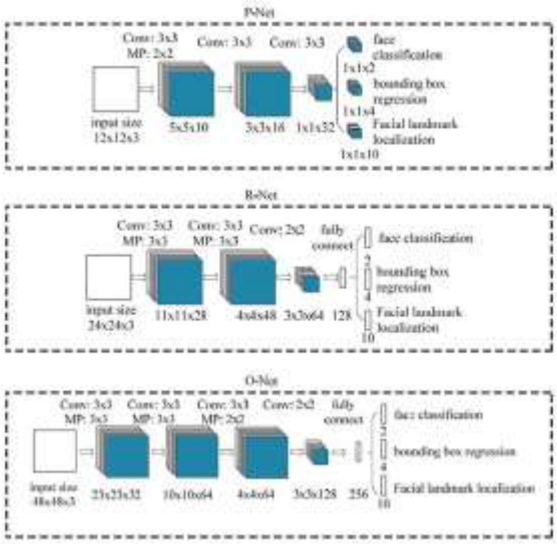

Fig. 11. Architecture of MTCNN.

Results and Discussions:

The ordered data augmentation techniques, including image rotation, zoom, brightness, translation, and various filters, contribute to enhancing the performance of a deep CNN model for face recognition. Notably, the study identifies specific levels of augmentation, emphasizing the effectiveness of the 3rd level of image zoom, 1st level of image translation, 1st level of image rotation, and 3rd level of image brightness for optimal results.

The reported average accuracy of 86.3% in 5-fold cross-validation demonstrates the positive impact of data augmentation on face recognition accuracy, particularly when working with a limited number of training samples. The study introduces a Convolutional Neural Network (CNN) as a superior alternative to traditional methods like PCA and LBPH, showcasing its exceptional accuracy, especially in scenarios with a small dataset.

Despite the promising outcomes, the study acknowledges certain limitations, such as a small and lower-quality dataset acquired in an uncontrolled environment. However, these drawbacks do not diminish the effectiveness of the CNN model with data augmentation. The approach proves to be a promising strategy for face recognition, highlighting its advantages over traditional methods in specific contexts.

Moreover, the statement about additional training of images yielding more accurate results, particularly with 0.11 million samples, emphasizes the scalability and robustness of the proposed deep CNN model. Achieving an impressive accuracy of 98.1% further underscores the potential of this approach for applications like class attendance prediction.

Method	Accuracy
PCA method	(18/54) 33.3%
LBPH method	(19/54) 35.2%
CNN with geometric transformation and brightness augmentation method	(45/54) 83.3%
CNN with filter operation augmentation method	(41/54) 75.9%

Fig. 12. Performance for Various Methods

MTCNN (Multi-task Cascaded Convolutional Networks) and VGG16 (Visual Geometry Group 16-layer) are two distinct but complementary components commonly employed in face recognition systems. MTCNN specializes in face detection, efficiently locating faces within images. On the other hand, VGG16 is a deep convolutional neural network specifically designed for image classification tasks.

In the context of face recognition, a prevalent strategy involves integrating MTCNN for the initial face detection stage and subsequently utilizing VGG16 or similar networks for the recognition tasks. This two-step process optimizes the accuracy of the overall face recognition system. The success of this approach is contingent on various factors, including the quality and diversity of the dataset, the fine-tuning of training parameters, and the seamless integration of these individual components.

MTCNN plays a pivotal role in the pipeline by effectively extracting faces from images, providing a crucial input for subsequent recognition tasks. VGG16, with its deep learning capabilities, contributes to the intricate process of recognizing and classifying the identified faces. The synergy between MTCNN and VGG16 underscores the importance of a well-designed and integrated pipeline for achieving high accuracy in face recognition applications. The overall effectiveness of the system is a result of the seamless coordination between these specialized components, emphasizing the significance of a comprehensive and strategic approach to face recognition.

Fig. 13. Accuracy Vs Training Data Size

Conclusion:

This research on Smart Attendance System described integrates cutting-edge technologies, employing the powerful DCNN VGG-16 and MTCNN architectures to enhance face recognition and detection. VGG-16, with its deep learning capabilities, proves instrumental in accurately recognizing faces across various conditions. Complementing this, the Multi-task Cascaded Convolutional Networks (MTCNN) architecture ensures multi-stage face detection, efficiently filtering non-face regions. The system goes beyond mere attendance tracking, incorporating robust anti-spoofing measures and moderate morphing detection to address security concerns. This comprehensive approach seeks to mitigate identity fraud risks, ensuring the integrity of attendance records. The user-friendly interface facilitates seamless interaction for administrators and users, streamlining attendance management.

References:

1. Pei, Z., Xu, H., Zhang, Y., Guo, M., & Yang, Y. H. (2019). Face recognition via deep learning using data augmentation based on orthogonal experiments. *Electronics*, *8*(10), 1088.
2. Warman, G. P., & Kusuma, G. P. (2023). Face recognition for smart attendance system using deep learning. *Commun. Math. Biol. Neurosci.*, *2023*, Article-ID.
3. Balamurali, K., Chandru, S., Razvi, M. S., & Kumar, V. S. (2021, June). Face spoof detection using vgg-face architecture. In *Journal of Physics: Conference Series* (Vol. 1917, No. 1, p. 012010). IOP Publishing.
4. Gupta, N., Sharma, P., Deep, V., & Shukla, V. K. (2020, June). Automated attendance system using OpenCV. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)* (pp. 1226-1230). IEEE.
5. Mery, D., Mackenney, I., & Villalobos, E. (2019, January). Student attendance system in crowded classrooms using a smartphone camera. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)* (pp. 857-866). IEEE.
6. Sharma, D., Sharma, H., & Panchal, D. (2020, December). Automatic Office Environment System for Employees Using IoT and Computer Vision. In *2020 IEEE 17th India Council International Conference (INDICON)* (pp. 1-6). IEEE. Raja, S. V. R. R. K., & Busch, C. Face Morphing Attack Generation & Detection: A Comprehensive Survey.
7. Scherhag, U., Rathgeb, C., & Busch, C. (2022). Face morphing attack detection methods. In *Handbook of Digital Face Manipulation and Detection: From DeepFakes to Morphing Attacks* (pp. 331-349). Cham: Springer International Publishing.
8. Policepatil, S., & Hatture, S. M. (2021). Face liveness detection: An overview. *International Journal of Scientific Research in Science and Technology*, *8*(4), 22-29.
9. Suresh, V., Dumpa, S. C., Vankayala, C. D., Aduri, H., & Rapa, J. (2019). Facial recognition attendance system using python and OpenCv. *Quest Journals-Journal of Software Engineering and Simulation*, *5*(2), 18-29.
10. Dang, T. V. (2023). Smart attendance system based on improved facial recognition. *Journal of Robotics and Control (JRC)*, *4*(1), 46-53.
11. Qi, H., Wu, C., Shi, Y., Qi, X., Duan, K., & Wang, X. (2023). A real-time face detection method based on blink detection. *IEEE Access*, *11*, 28180-28189
12. Tapia, J. E., & Busch, C. (2021). Single morphing attack detection using feature selection and visualization based on mutual information. *IEEE Access*, *9*, 167628-167641.
13. Malik, A., Kuribayashi, M., Abdullahi, S. M., & Khan, A. N. (2022). DeepFake detection for human face images and videos: A survey. *Ieee Access*, *10*, 18757-18775.
14. Jabberi, M., Wali, A., Neji, B., Beyrouthy, T., & Alimi, A. M. (2023). Face ShapeNets for 3D Face Recognition. *IEEE Access*.
15. Mishra, N. K., & Singh, S. K. (2022). Face Recognition Using 3D CNN and Hardmining Loss Function. *SN Computer Science*, *3*(2), 155.
16. Yuan, M., Nikouei, S. Y., Fitwi, A., Chen, Y., & Dong, Y. (2020, August). Minor privacy protection through real-time video processing at the edge. In *2020 29th International Conference on Computer Communications and Networks (ICCCN)* (pp. 1-6). IEEE.
17. Hamza, M., Tehsin, S., Karamti, H., & Alghamdi, N. S. (2022). Generation and detection of face morphing attacks. *IEEE Access*, *10*, 72557-72576.
18. Hasan, M. K., Ahsan, M. S., Newaz, S. S., & Lee, G. M. (2021). Human face detection techniques: A comprehensive review and future research directions. *Electronics*, *10*(19), 2354.
19. Kumar, A., Kalumbi, S., Gupta, M. K., & RaoMV, P. Smart Face Recognition Using IOT and Machine Learning. *Journal homepage: www. ijrpr. com ISSN*, *2582*, 7421.
20. Saleem, I., & Shukr, B. K. (2023). Techniques and Challenges for Generation and Detection Face Morphing Attacks: A Survey. *Iraqi Journal of Science*, 385-404.