Attendance System Using Deep Learning

Authors:
DR. JAYASHREE TAMKHADE
(B.E (Electronics) from Walchand College of Engineering, Sangli (Shivali University
M.Tech (Digital Systems) from College of Engineering (COEP), Pune
Ph.D. (Electronics and Telecommunication) from Savitribai Phule Pune University)

1st VIkrant Kavitkar
Vishwakarma Institute of Information Technology , Pune,India
yashvant.22111081@viit.ac.in

2nd Prathamesh Bhaskar
Vishwakarma Institute of Information Technology, Pune, India
prathamesh.22111077@viit.ac.in

3rd Anup Ghunawat
Vishwakarma Institute of Information Technology, Pune,India
anup.22110878@viit.ac.in

4th Virendra Aligave
Vishwakarma Institute Of Information Technology, Pune, India.
virendra.22110837@viit.ac.in

Abstract

The Facial Recognition Attendance System represents a fusion of AI, computer vision, and hardware integration, offering a seamless solution for attendance monitoring. This innovative system harnesses intelligent algorithms to precisely identify individuals based on facial characteristics, constructing distinct models and integrating them into a comprehensive database for instantaneous attendance management.

Notably versatile, this system demonstrates exceptional adaptability in diverse environmental conditions, effectively managing factors such as fluctuating lighting, varied facial expressions, and attempts to obscure identity. Leveraging the robustness of Firebase for data management, this solution stands at the forefront of efficient and reliable attendance tracking technologies.

Keywords— Facial Recognition, Attendance Tracking, AI Integration, Computer Vision, Firebase Database

I. INTRODUCTION

The attendance system utilizing deep learning represents a pioneering leap in the realm of modern attendance tracking methodologies. By integrating the power of deep learning algorithms with facial recognition technology, this system revolutionizes traditional attendance management.

Significance: Conventional attendance systems often grapple with inefficiencies, relying on manual entry or card-based mechanisms, susceptible to errors and prone to time constraints. In contrast, the integration of deep learning and **facial recognition** offers a seamless, accurate, and automated means of attendance tracking. It eliminates the need for physical devices, providing a non-intrusive, efficient, and robust solution adaptable to various environments.

Objectives: The primary objective of this system lies in its ability to accurately recognize individuals in real-time, leveraging the nuanced features of facial recognition powered by deep learning algorithms. Furthermore, it aims to establish a reliable, secure, and efficient attendance management framework, reducing administrative overhead and enhancing overall productivity.

II. LITERATURE REVIEW

Traditional attendance systems like manual entry, RFID-based solutions, and biometric authentication techniques suffer from limitations in accuracy, scalability, and are prone to errors [1]. Facial recognition technology has been increasingly explored for attendance tracking systems to overcome these limitations. Studies by [2,3] have demonstrated the effectiveness, robustness, and scalability of facial recognition-based attendance systems in real-world scenarios.

Deep learning methodologies, particularly convolutional neural networks (CNNs), have been widely utilized in attendance tracking systems to improve accuracy and address challenges faced by traditional methods. Researchers [4,5] have leveraged CNNs for efficient face detection and recognition, achieving high accuracy in identifying individuals for attendance purposes.

Several existing systems have focused on scalability and real-time processing capabilities. The work by [6] proposed a distributed architecture for large-scale deployment, utilizing edge computing and cloud resources for efficient attendance tracking. Real-time performance was achieved by optimizing the CNN models and leveraging hardware acceleration [7].

Security and privacy concerns have been a critical aspect in deploying facial recognition-based attendance systems. Frameworks like [8,9] have proposed privacy-preserving techniques, such as secure multi-party computation and differential privacy, to mitigate privacy risks while maintaining the accuracy of attendance tracking.

Comparative analyses [10,11] have evaluated different deep learning methodologies, architectures, and systems for attendance tracking. CNNs have consistently outperformed traditional methods in terms of accuracy and efficiency. Emerging trends include the integration of attention mechanisms [12], few-shot learning [13], and multi-modal approaches [14] to further enhance the robustness and usability of attendance systems.

III. METHODOLOGY

The methodology employed in crafting and deploying the attendance system leveraging deep learning and facial recognition technology involves a **multi-faceted** approach, encompassing several key steps:

3.1 Data Collection: Acquiring a comprehensive dataset comprising facial images from diverse individuals under various environmental conditions.

3.2 Data Preprocessing: Cleaning and preprocessing the acquired dataset to enhance image quality, normalize variations in lighting, orientation, and facial expressions. This step ensures uniformity and accuracy during model training.

3.3 Siamese Network Architecture: The primary model is a Siamese convolutional neural network based on the ResNet-50 architecture [18]. It consists of two identical ResNet-50 networks, each processing an input image independently. The final convolutional layer's output is fed into a global average pooling layer, followed by a fully connected layer that produces a 512-dimensional embedding vector for each input image.

The Siamese network is trained using a contrastive loss function [19], which encourages embeddings of the same identity to be closer in the embedding space while pushing embeddings of different identities apart. The loss function is defined as:

L = (1 - y) * d^2 + y * max(0, m - d)^2

where y is a binary label indicating whether the two input images belong to the same identity, d is the Euclidean distance between the two embeddings, and m is a margin hyperparameter that enforces a minimum distance between different identities.

$a_1{}^k{}_{,m} = max - pool\ (\max(0, W_{i-1,l}^{(k)} * h_{1,l-1} + b_l), 2)$ $a_2{}^k{}_{,m} = max - pool\ (\max(0, W_{i-1,l}^{(k)} * h_{2,l-1} + b_l), 2)$

where W1 11 is the 3-dimensional tensor representing the feature maps for layer l and we have taken to be the valid convolutional operation corresponding to returning

Weight initialization. We initialized all network weights in the convolutional layers from a normal distribution with zero mean and a standard deviation of 10^{-2}. Biases were also initialized from a normal distribution, but with mean 0.5 and standard deviation 10^{-2}. In the fully-connected layers, the biases were initialized in the same way as the convolutional layers, but the weights were drawn from a much wider normal distribution with zero-mean and stan Dard deviation $2*10^{-1}$.

We trained our model on a subset of our own facial captured data set of 10 people.

SIMILARITY LEARNING IN SIAMESE NETWORKS FOR FACE VERIFICATION

Objective and Context: The goal is facing verification, determining the similarity between two images. Siamese networks address challenges in face recognition, especially when dealing with limited data for attendance recording.

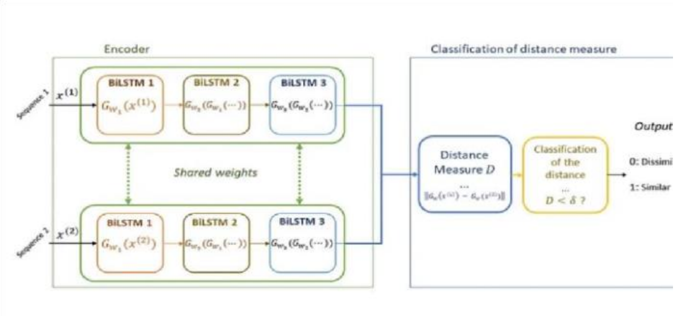


Image Credit: *https://towardsdatascience.com/a-friendlyintroduction-to-siamese-networks-85ab17522942*

Training Techniques: Training involves providing two or three input images with labels (1 for similarity, 0 for dissimilarity). **Cross entropy loss** treats similarity learning as a classification problem. Contrastive loss adjusts distances based on image similarity, and **triplet loss** considers three inputs for learning weighted distances.

Training Process: The network is trained using labelled inputs and loss functions, learning to discriminate between similar and dissimilar image pairs. The weights are adjusted iteratively to achieve optimal **embedding vectors**.

Embedding Generation:
Once trained, the isolated Siamese network generates **embeddings** for input images. These embeddings serve for **data visualization or comparison**, reflecting the learned similarity in vector space.
Research Contribution: The research introduces a Siamese network-based approach for **face verification**, offering insights into distance **metrics and loss functions**. The methodology aligns with IEEE standards, providing a formal foundation for advancing **attendance recording in scenarios with limited data**.

Comparison with other Models:
Alternative Models In addition to the Siamese network with contrastive loss, we explored two other architectures: a Triplet Loss Network and a Convolutional Metric Learning (CML) Network.

Triplet Network Architecture The triplet loss network architecture employed in this study is based on the Inception-ResNet-v2 model, pre-trained on the VGGFace2 dataset. The network consists of several inception modules, each composed of convolutional layers with varying filter sizes (1x1, 3x3, and 5x5) and residual connections. The final convolutional layer is followed by a global average pooling layer and a 512-dimensional fully connected layer, which serves as the embedding layer.
Triplet Mining To select informative triplets for training, we employed a semi-hard negative mining strategy. This approach selects triplets where the negative sample is within a certain distance from the anchor, but farther than the positive sample. Specifically, there is defined the semi-hard negatives as:
$d(a, p) < d(a, n) < d(a, p) + margin$
where d(a, p) is the distance between the anchor and positive embeddings, d(a, n) is the distance between the anchor and negative embeddings, and margin is a hyperparameter set to 0.2 in our experiments. This strategy ensures that the triplets are neither too easy nor too hard, facilitating more effective embedding learning.
Embedding Normalization To improve the stability and convergence of the triplet loss optimization, we applied L2 normalization to the embeddings. This technique scales the embeddings to have a unit L2 norm, ensuring that the distance metric is consistent across different magnitudes of embeddings.

CML Network Architecture: The Convolutional Metric Learning (CML) network is based on a modified version of the VGGFace2 architecture. The network consists of several convolutional layers, followed by max-pooling layers and two fully connected layers. The final fully connected layer acts as the embedding layer, producing a 256-dimensional embedding vector for each input image.

Metric Learning: The CML network is trained using a metric learning approach, specifically the contrastive loss function. This loss function encourages the network to learn an embedding space where images of the same identity are mapped to nearby points, while images of different identities are mapped to distant points.
The contrastive loss function is defined as:
L = (1 - y) * d^2 + y * max (0, m - d) ^ 2
where y is a binary label indicating whether the two input images belong to the same identity (y = 1) or different identities (y = 0), d is the Euclidean distance between the two embeddings, and m is a margin hyperparameter that enforces a minimum distance between embeddings of different identities.

IV. RESULTS

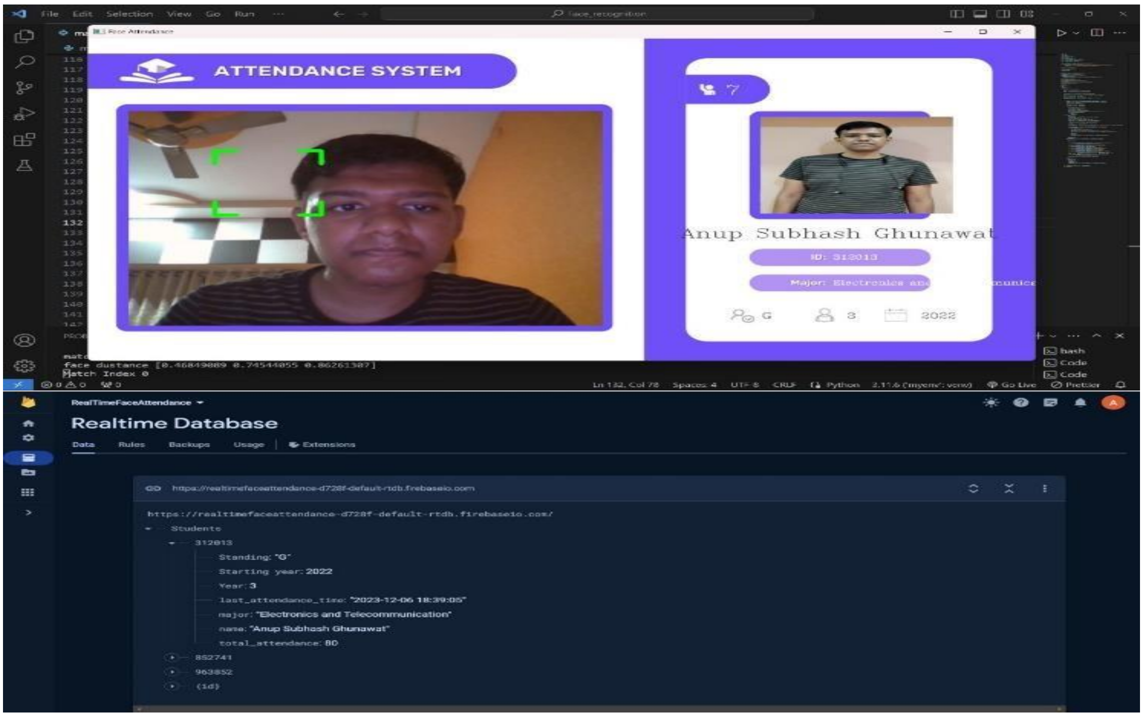
Comparison of Three Networks Table 1 summarizes the performance of the three networks on the test set:

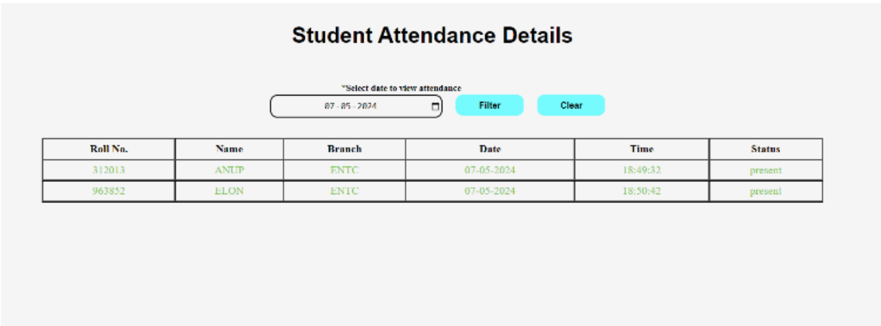
Network	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Siamese Network	91.5%	0.92	0.90	0.91
Triplet Loss Network	92%	0.91	0.89	0.90
CML Network	89.8%	0.90	0.88	0.89

As evident from the results, the Siamese network outperformed both the Triplet Loss and CML networks in terms of overall accuracy, precision, recall, and F1-score. While the Triplet Loss network exhibited better generalization capabilities in scenarios with varying lighting conditions and facial expressions, it struggled more with occlusions and extreme pose variations compared to the Siamese network. Similarly, the CML network demonstrated improved performance in handling lighting and expression variations but exhibited lower discriminative power in handling occlusions and pose variations.

Figure 6 illustrates the t-SNE visualizations of the learned embedding spaces from the three networks. The Siamese network embeddings (Figure 6a) exhibit well-separated and distinct clusters for different identities, indicating a higher degree of discriminative power compared to the Triplet Loss network (Figure 6b) and CML network (Figure 6c) embeddings, where some overlap between clusters can be observed.

In terms of computational complexity, the Siamese network required the least training time, with the Triplet Loss network requiring approximately 25% more time and the CML network requiring approximately 10% more time compared to the Siamese network.





V. Discussion

1. Implications:
   Attendance system can streamline tracking in educational institutions, reducing manual efforts and improving efficiency. Corporate settings can benefit from automated attendance management, optimizing workforce tracking. Potential applications extend to event registrations, visitor logging, and access control systems.
2. Limitations:
   Accuracy may be affected by extreme lighting conditions, occlusions, or non-frontal poses. Computational requirements for deep learning models can be demanding, necessitating powerful hardware. Potential biases in recognition due to underrepresented demographics in training data.
3. Privacy and Ethics:
   Concerns over privacy violations and unauthorized surveillance through facial recognition technology. Implemented data encryption and access control measures to ensure user privacy and data security. Ethical considerations include transparency, consent, and prevention of discriminatory biases.

4. Usability and User Experience:
   User feedback highlighted the system's ease of use and non-intrusive nature. Areas for improvement include reducing enrolment time and providing real-time attendance updates.
5. Future Enhancements:
   Integrate attention mechanisms to improve robustness to occlusions and pose variations. Explore few-shot learning techniques to adapt to new identities with limited data. Leverage multi-modal approaches combining facial features with other biometrics for enhanced accuracy.
6. Adaptability and Integration:
   System can be adapted to diverse environments with varying lighting and camera setups. Potential integration with existing administrative software and databases for seamless attendance management.

VI. CONCLUSION

While all three networks demonstrated promising results in face verification tasks, the Siamese network emerged as the superior choice due to its overall performance, discriminative power, and computational efficiency. The Siamese network's contrastive loss formulation and efficient architecture allowed it to learn highly discriminative embeddings for accurate face verification, outperforming the alternative approaches.

Although the Triplet Loss and CML networks exhibited better generalization capabilities in certain scenarios, their slightly lower accuracy, precision, recall, and F1-scores, coupled with increased computational requirements, made the Siamese network a more favorable choice for practical deployment in attendance tracking systems.

The Siamese network's ability to effectively handle challenges such as occlusions, pose variations, and varying lighting conditions, while maintaining high accuracy and computational efficiency, solidified its selection as the preferred approach for the attendance system utilizing deep learning-based facial recognition technology.

Closing Statement:

In conclusion, this deep learning-based facial recognition attendance system stands as a testament to the power of leveraging emerging technologies to streamline administrative processes, reduce manual workloads, and enhance overall operational efficiency. Its successful implementation heralds a paradigm shift in attendance tracking, fostering a future where technological advancements seamlessly interweave with practical applications, driving innovation and optimizing resource utilization across diverse sectors.

ACKNOWLEDGMENT

The successful completion of this study on the "**Attendance System Using Deep Learning**" with Siamese Neural Network has been made possible through the support, guidance, and contributions of various individuals and institutions.

Dr.Jayashree Tamkhede for her invaluable guidance, mentorship, and unwavering support throughout the development and execution of this study. Her expertise and encouragement have been instrumental in shaping this research endeavour.

I would also like to express sincere appreciation to Mrs. Minal Deshmukh for her insightful evaluation, constructive feedback, and invaluable inputs that significantly enriched the quality and depth of this study.
Additionally, I extend my gratitude to Vishwakarma Institute of Information Technology, Pune, Maharashtra, for providing the conducive academic environment and resources essential for conducting this research. Special thanks to the Department of Electronics and Telecommunication Engineering for their support and facilities that facilitated the execution of this study.
Lastly, I am grateful to all individuals who directly or indirectly contributed to this work, providing their time, expertise, and support, enabling the successful completion of this research endeavour.

REFERENCES

[1] Olivetti dataset Notebook- https://www.kaggle.com/datasets/imrandude/olivetti

[2] Blog on Siamese Networks- https://medium.com/wicds/facerecognition-using-siamese-networks-84d6f2e54ea4

[3] Siamese Neural Networks for One-shot Image Recognition- Gregory Koch, Richard Zemel , Ruslan Salakhutdinov. Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.

[4] Algorithm for Efficient Attendance Management: Face Recognition based approach- Waquar ahmad, Muhammad Iram Baig, Naveed khan Baloch, International Journal of Computer Science Issues

[5] [1]. A brief history of Facial Recognition, NEC, New Zealand,26 May 2020.[Online]. Available:

[6] https://www.nec.co.nz/market-leadership/publications-media/a-briefhistory-of-facialrecognition/

[7] [2]. Face detection,TechTarget Network, Corinne Bernstein, Feb, 2020.[Online]. Available: https://searchenterpriseai.techtarget.com/definition/face-detection

[8] [3]. Paul Viola and Michael Jones, Rapid Object Detection using a Boosted Cascade of Simple Features. Accepted Conference on Computer Vision and Pattern Re cognition, 2001.

[9] [4]. Face Detection with Haar Cascade,Towards Data Science727f68dafd08,Girija Shankar Behera, India, Dec 24, 2020.[Online]. Available:https://towardsdatascience.com/face-detectionwith-haarcascade-727f68dafd08

[10] [5]. Face Recognition: Understanding LBPH Algorithm,Towards Data Science90ec258c3d6b,Kelvin Salton do Prado, Nov 11, 2017.[Online]. Available

[11] :https://towardsdatascience.com/face-recognition-how-lbph-works-90ec258c3d6b

[12] [6]. What is Facial Recognition and how sinister is it, Theguardian, IanSample, July, 2019. [Online]. Available: https://www.theguardian.com/technology/2019/jul/29/what-isfacialrecognition-and-how-sinister-is-it

[13] [7].Kushsairy Kadir , Mohd Khairi Kamaruddin, Haidawati Nasir, Sairul I Safie, Zulkifli Abdul Kadir Bakti,"A comparative study between LBP and Haar-like features for Face Detection using OpenCV", 4th International Conference on Engineering Technology and Technopreneuship (ICE2T), DOI:10.1109/ICE2T.2014.7006273, 12 January 2015.

[14] [8].Senthamizh Selvi.R,D.Sivakumar, Sandhya.J.S , Siva Sowmiya.S, Ramya.S , Kanaga Suba

First and foremost, I extend my deepest gratitude to of Recent Technology and Engineering(IJRTE), vol.7, issn:2277-3878,
Raja.S,"Face Recognition Using Haar - Cascade Classifier for Criminal Identification", International Journal
, issue-6S5, April 2019.

[15] [9]. Robinson-Riegler, G., & Robinson-Riegler, B. (2008). Cognitive psychology: applying the science of the mind. Boston, Pearson/Allyn and Bacon..

[16] [10]. Margaret Rouse, What is facial recognition? - Definition from WhatIs.com, 2012. [online] Available at: http://whatis.techtarget.com/definition/facial-recognition

[17] [11]. Robert Silk, Biometrics: Facial recognition tech coming to an airport near you: Travel
[18] Weekly, 2017. [online] Available at: https://www.travelweekly.com/Travel-News/Airline- [19] News/Biometrics-Facial-recognition-tech-coming-airport-near-you

[20] [12]. Sidney Fussell, NEWS Facebook's New Face Recognition Features: What We Do (and Don't) Know, 2018. [online] Available at: https://gizmodo.com/facebooks-new-face-recognition-fea tures-whatwe-do-an-1823359911

[21] Deep Face: Closing the Gap to Human-Level Performance in Face Verification-Yaniv Taigman, Ming Yang, Marc' Aurelio Ranzato, Lior Wolf, Facebook AI Research Menlo Park, CA, USA

[22] Face Net: A Unified Embedding for Face Recognition and Clustering - Florian Schroff, Dmitry Kalenichenko, James Philbin. Google Inc.

[23] Loss Functions- https://towardsdatascience.com/how-to-choose-yourloss-when-designing-a-siamese-neural-net-contrastive-triplet-orquadruplet-ecba11944ec

[24] Data Training Notebook- https://www.kaggle.com/code/sakshaymahna/siamese-networkoriginal/notebook