An Analysis of Advanced FaceNet Deep Learning Algorithm in Facial Authentication

D. Gayathri [1] and V. Raghavendran [2]
[1]Research Scholar, Department Computer Science, VISTAS, Chennai.
[2] Assistant Professor, Department of Computer Science, VISTAS, Chennai.
gayathri.d1992@gmail.com

Abstract. Emerging Deep-Learning techniques are highly effective face recognition models, primarily due to the ability to learn complex data representations through multiple layers of processing. Convolutional-neural networks (CNNs) has proven to be the cornerstone of face-recognition systems based on deep learning and plays a central role in their success. This paper explores the architecture, feature extraction, face matching through deep learning features, and the enhancement of the Traditional FaceNet algorithm to an advanced FaceNet Deep Learning Algorithm through additional feature recognition like jaw lines and forehead lines can be used to prove optimum results in accuracy, Mean Average Precision, face verification-accuracy Receiver Operating Characteristic (ROC) Curve, Recall, F1 Score, Face Identification Accuracy.

Keywords: Deep-Learning Networks, Traditional FaceNet algorithm, Advanced-FaceNet Algorithm, Metrics.

1. Introduction

Face detection remains a prominent research area, particularly in systems such as accessing control, attendance management, and identifying vulnerable individuals. Various techniques are employed for face detection, including binary descriptors,[1]three-way decisions, and alignment-learning, all relying on image and video analysis. Deep-learning mechanisms, especially convolutional-neural networks (CNNs), has show casted important promise in feature-extraction from images and videos. CNNs are adept at object recognition tasks and have undergone continuous refinement to improve their performance[2]. They are widely used in image classification and have been integrated with other architectures such as long short-term memory networks (LSTMs) and region-based networks (R-CNNs) to enhance sequential data processing and detection speed, respectively[3]. These upgrades in CNN-based framework have eminently advanced face detection techniques based on deep learning.

N. V. Suresh and P. S. Buvaneswari (eds.), *Proceedings of the International Conference on Digital Transformation in Business: Navigating the New Frontiers Beyond Boundaries (DTBNNF 2024)*, Advances in Economics, Business and Management Research 283,
https://doi.org/10.2991/978-94-6463-433-4_41

Face recognition[4] is a complex task in biometrics, involving the identification and verification of human faces using unique natural features. It has garnered significant attention from researchers in the vision of computer and recognition pattern fields and f utilizations in various domains such as military, finance, security, and anti-terrorism[5]. Unlike other biometric methods, face recognition does not require user cooperation and can be performed non-intrusively with good concealment. It involves recognizing faces across multiple dimensions and considering various appearance variations, including lighting conditions and background data. Since faces are projected onto 2D images, robust face recognition techniques are needed to handle inherent variations and deformations. This becomes even more challenging when dealing with diverse facial characteristics related to identity, race, genetics, and other intra-personal factors. Additionally, the recognition process must address issues related to image acquisition and achieve a balance between accuracy, execution time, and space efficiency.

The traditional facial recognition pipeline typically comprises four primary phases[12].(1)Detect,(2)Align,(3)Represent,(4)Classify. Deep learning, particularly with the rise of extensive face datasets, has revolutionized 2D face-recognition. The remarkable data-learning capability inherent in deep learning has significantly boosted the accuracy and performance of face recognition systems. Deep learning, a prominent area in machine-learning research, involves the development of neural networks that mimic the workings of the human brain to interpret and analyze multimodal data effectively. Unlike traditional machine learning approaches, which heavily rely on handcrafted feature representations, deep-learning methods spontaneously learn data representations also extract features from raw input data. Among the enormous deep-learning Layout, the convolutional neural network (CNN) stands out as the most widely implemented in face-recognition applications because of its ability to derive discriminative features across multiple levels of abstraction. CNNs excel at learning rich feature representations directly from images, making them well-suited for face recognition tasks. However, training a deep CNN involves adjusting a vast number of network parameters using large labeled datasets, which can be computationally intensive and time-consuming. Typically, CNNs are trained on large datasets to initialize model weights or extract features relevant to the specific task at hand. Fine-tuning the trained CNN model on an intent dataset helps fetch the network to the intent domain. Still, excessive fine-tuning may lead to overfitting, especially while there are significant variations between the training and test datasets, such as variations in illumination, facial expressions, and viewpoints. Furthermore, changing the first data distribution for a trained CNN model is challenging, and the network may struggle to adapt to new distributions efficiently. To mitigate these challenges, careful parameter tuning, including using small learning rates and normalization techniques, is crucial to ensure optimal performance and prevent overfitting during training.

This chapter conducts a thorough examination of deep learning-based face recognition models, considering both algorithmic and data-related aspects. The investigation begins by outlining a multi-stage approach for assembling extensive face datasets comprising thousands of images representing numerous unique identities. These datasets are sourced from various online repositories and serve as the foundation for training and evaluating the models under study. The chapter proceeds to categorize different methods employed in the face recognition process, distinguishing between face identification and verification techniques. Various deep network architectures are explored in terms of their approaches to face alignment, metric learning, and the utilization of lossy functions. Numerous face-recognition frameworks leverage different variety of deep systems, and it evaluate their modeling choices based on their relevance to the task at hand.

2. Related Work

FaceNet serves as a consolidated embedding approach for both face-recognition and clustering-tasks. It leverages a vast dataset containing approximately 260 million face images and operates as a deeply structured model, with images closely cropped but lacking additional alignment beyond cropping. FaceNet directly learns to map face-images into a compact Euclidean Space, where spaces between embeddings align with the measures of face-similarity. Once this space is established, standard techniques for face-recognition, verification, and clustering can be implemented, treating FaceNet embeddings as feature vectors. Training in FaceNet involves optimizing the embedding itself rather than any intermediary layer or bottleneck. The FaceNet working model is depicted in Fig 1.1 . During training, FaceNet employs sets of facial patches, both matching and non-matching, that are approximately aligned. These sets are created using online triplet mining. Triplet loss, guided by distance margin considerations, aids in separating positive from negative pairs in the embedding space.

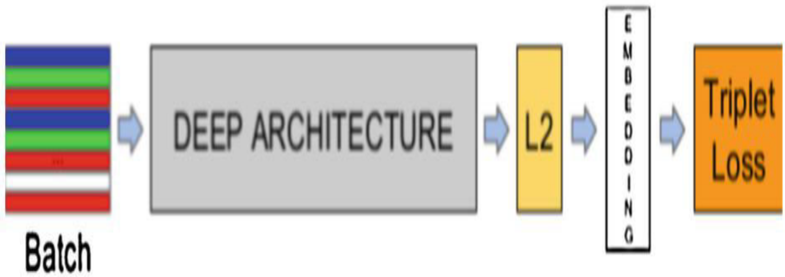

Fig 1.1 Model of FaceNet

This method offers improved representational efficiency, achieving considerable recognition performance with just 128 bytes per face. FaceNet demonstrates impressive accuracy levels, achieving 99.63%[19] accuracy on the LFW dataset and 95.12% on the YouTube images database. For further details on FaceNet's performance, refer to [18]. FaceNet employs harmonic triplet loss, depicted in Figure 1.2, which illustrates distinct face embedding versions obtained from different networks. These embeddings are mutually compatible, enabling direct comparisons between them.

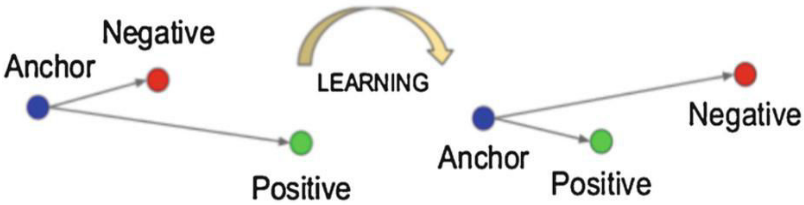

Fig 1.2 Harmonic Triplet loss in FaceNet

Selecting triplets can be challenging, especially when employing curriculum learning. It's often recommended to choose hard positive or negative examples within each mini-batch, particularly when dealing with large mini-batches containing thousands of examples [17]. Within a mini-batch, argmin and argmax calculations are performed. To enhance clustering accuracy, hard-positive mining is employed [21], which promotes clusters with a spherical-nature for single-person embeddings. Fig 1.3 illustrates typical examples of face clustering.

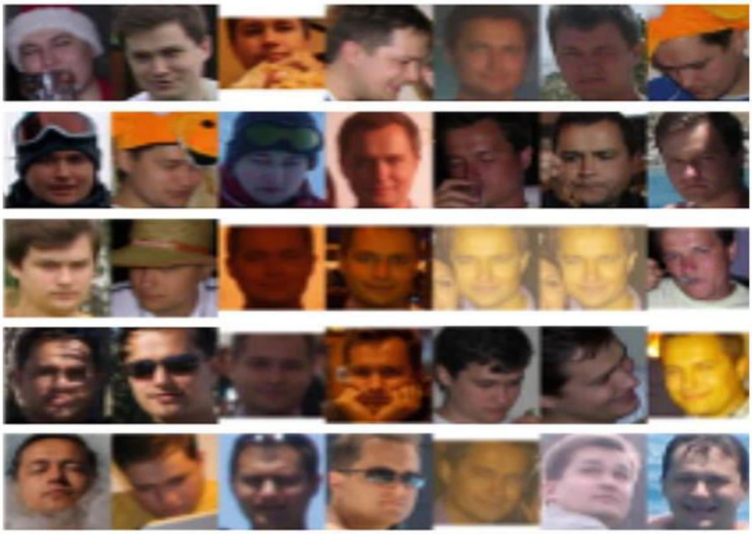

Fig 1.3 Examples of Face Clustering

3. Methodologies

Feature-extraction methodologies target to identify distinctive features from each face image for matching towards various faces. The FaceNet-model is designed to produce more distinctive embeddings for face images, facilitating accurate verification and identification. It's a already trained [20] deep-convolutional neural network that achieves facial-recognition using just 128 bytes per-face. According to [11] FaceNet differs from previous approaches by optimizing its embedding using deep convolutional networks rather than intermediate bottleneck layers. They describe FaceNet as a one-shot learning method [18] that calculates similarity distances for each face in Euclidean space.

The model acquires a mapping variable [12] that directly transforms input face-images into a good-dimensional [15] Euclidean space, where the length between embeddings reflect dissimilarity or similarity. This is facilitated by a[20] deep-CNN architecture Fig 1.4 incorporating the inception module, allowing for feature capture at multiple scales. During training[22], FaceNet employs triplet-loss, encouraging closer distances between embeddings of matching faces than unmatched ones. This fosters the acquiring of highly distinctive features resilient to lighting, pose, and also other facial-variations. FaceNet has demonstrated outstanding effectiveness on benchmark face-recognition datasets such as Labeled Faces in the Wild and MegaFace, surpassing previous methods in various tasks. One notable advantage of FaceNet is its generalization across diverse domains and datasets, maintaining resilient effectiveness even amidst significant facial appearance variations.

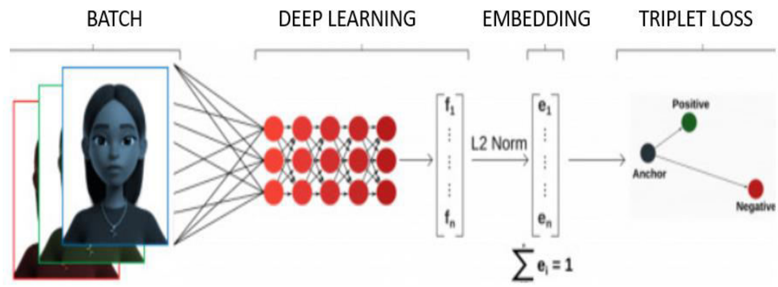

Fig 1.4 Architecture of FaceNet

The Euclidean distance loss function, as a form of metric learning, operates by embedding images into Euclidean space, reducing intra-class variance while increasing inter-class variance. Another common loss function used for this purpose is the contrastive loss, along with triplet loss.

$$\mathcal{LOSS} = \boldsymbol{b_{ij}} \max\left(\mathbf{0}, \left\|\boldsymbol{h}\left(\boldsymbol{a_i}\right) - \boldsymbol{h}\left(\boldsymbol{a_j}\right)\right\|_2 - \varepsilon^+\right) + \left(\mathbf{1} - \boldsymbol{b_{ij}}\right) \max\left(\mathbf{0}, \varepsilon^- - \left\|\boldsymbol{h}\left(\boldsymbol{a_i}\right) - \boldsymbol{h}\left(\boldsymbol{a_j}\right)\right\|_2\right)$$

Feature Extraction concept can be improved by training the FaceNet-model to extract so more additional features such as jaw lines, and forehead lines.

3.1.Face-matching through other Deep-features

The utilization of deep-features for face processing is crucial in face recognition tasks. 3D reconstruction enhances training data diversity by reconstructing faces in three dimensions, utilizing deep methods. CNN models directly generate 2D images, allowing analysis of face images with variant poses through multi-view perceptron. GAN refines images by combining [15]data distribution's prior knowledge with facial features, with several versions available. [20]Many-to-one normalization aims to produce front-facing images and reduce appearance variability in test data for aligned and comparative analysis. It includes stacked progressive autoencoders (SAE), [17]CNN, and GAN techniques. SAE directs non-frontal faces to frontal views using stacked autoencoders. CNN extracts identity-preserving features for reconstructing face images in canonical views. GAN, consisting of landmark-located patch[8] and global encoder-decoder networks[9], generates frontal views while preserving global structures and local details through symmetry, adversarial and identity-preserving[18] losses.

4. Conclusion and other discussions

Achieving high performance in face recognition algorithms requires robustness to various challenges present in the situations of real-world. Deep-learning frameworks, such as convolutional-neural networks (CNNs), have depicted promise in addressing these challenges, but there are still limitations and opportunities for improvement. One area of focus is the enhancement of CNN architectures and learning methods to handle different covariates that affect performance. Covariates include factors such as image quality (blur, JPEG compression, occlusion, noise), image-characteristics (brightness, contrast, missing pixels), and model properties (architecture, color information, descriptor computation). Studies have analyzed the impact of these covariates on the improvement of popular CNN models. It has been seen that more levels of noise, blur, missing pixels, and also brightness have a different impact on verification performance, while changes in contrast and compression artifacts have a lesser effect. Interestingly, the computation strategy for descriptors and color information do not significantly influence performance. In summary, while CNNs have shown effectiveness in face recognition, there is ongoing research to enhance their robustness to various real-world challenges, ultimately aiming for improved performance across different datasets and environments. The high-level trained model can be developed by adding new features like jaw-lines recognition and forehead lines recognition so that the face recognition metrics like Accuracy, Mean Average Precision, face verification accuracy, Receiver Operating Characteristic (ROC) Curve, F1 Score, Face Identification Accuracy can be improved by the Advanced FaceNet Deep Learning Algorithm.

References

1. Y. Sun, D. Liang, X. Wang, X. Tang, DeepID3: face recognition with very deep neural networks. arXiv:1502.00873v1 (2015)
2. E. Zhou, Z. Cao, Q. Yin, Naive-deep face recognition: touching the limit of LFW benchmark or not? arXiv:1501.04690v1 (2015)
3. Suganya, V., & Suresh, N. V. (2024). Potential Mental and Physical Health Impacts of Spending Extended Periods in the Metaverse: An Analysis. In Creator's Economy in Metaverse Platforms: Empowering Stakeholders Through Omnichannel Approach (pp. 225-232). IGI Global.
4. Joseph A. Mensah,Justice K. Appati, Elijah K.A Boateng,Eric Ocran,Louis Asiedu: FaceNet recognition algorithm subject to multiple constraints: Assessment of the performance, Scientific African (2024)
5. Deng X., Da F., Shao H., Jiang Y : A multi-scale three-dimensional face recognition approach with sparse representation-based classifier and fusion of local covariance descriptors ,Comput. Electr. Eng., 85 (2020)
6. Pain C.D., Egan G.F., Chen Z, Deep learning-based image reconstruction and post-processing methods in positron emission tomography for low-dose imaging and resolution enhancement Eur. J. Nucl. Med. Mol. Imaging, 49 (9) (2022).
7. Sundararajan K., Woodard D.L.: Deep learning for biometrics: A survey ACM Comput. Surv., 51 (3) (2018)

8. F. Tang et al., "An end-to-end face recognition method with alignment learning," Optik, vol. 205, Mar. 2020
9. Catherine, S., Kiruthiga, V., Suresh, N. V., & Gabriel, R. (2024). Effective Brand Building in Metaverse Platform: Consumer-Based Brand Equity in a Virtual World (CBBE). In Omnichannel Approach to Co-Creating Customer Experiences Through Metaverse Platforms (pp. 39-48). IGI Global
10. M. Tzelepi, A. Tefas, Exploiting supervised learning for finetuning deep CNNs in content-based image retrieval, in Proceedings of 23rd International Conference on Pattern Recognition, (IEEE, Piscataway, 2016)
11. W. Rawat, Z. Wang, Deep convolutional neural networks for image classification: a comprehensive review. Neural Comput. 29(9), 2352–2449 (2017)
12. X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks. J. Mach. Learn. Res. 9, 249–256 (2010)
13. Z. Lu, X. Jiang, A. Kot, An effective color space for face recognition, in Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, (IEEE Service Center, Piscataway, 2016), pp. 849–856
14. W. Niu, Y. Zhao, Z. Yu, Y. Liu, and Y. Gong, "Research on a face recognition algorithm based on 3D face data and 2D face image matching," Journal of Visual Communication and Image Representation, vol. 91, Mar. 2023
15. F. Zhao, J. Li, L. Zhang, Z. Li, and S.-G. Na, "Multi-view face recognition using deep neural networks," Future Generation Computer Systems, vol. 111, pp. 375–380, Oct. 2020
16. K. B. Pranav and J. Manikandan, "Design and evaluation of a real-time face recognition system using convolutional neural networks," Procedia Computer Science, vol. 171, pp. 1651–1659, 2020
17. Suresh, N. V., & Rexy, V. A. M. (2024, February). An Empirical Study on Empowering Women through Self Help Groups. In 3rd International Conference on Reinventing Business Practices, Start-ups and Sustainability (ICRBSS 2023) (pp. 957-964). Atlantis Press.
18. Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," Prepr. arXiv.1411.7766, Nov. 2014
19. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Communications of the ACM, vol. 60, no. 6, pp. 84–90, May 2017
20. C. Szegedy et al., "Going deeper with convolutions," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Jun. 2015
21. R. Rameswari, S. N. Kumar, M. A. Aananth, and C. Deepak, :"Automated access control system using face recognition," Materials Today: Proceedings, vol. 45, pp. 1251–1256, 2021
22. A. Shah, B. Ali, M. Habib, J. Frnda, I. Ullah, and M. S. Anwar, :"An ensemble face recognition mechanism based on three-way decisions," Journal of King Saud University - Computer and Information Sciences, vol. 35, no. 4, pp. 196–208, Apr. 2023
23. A. M. S. Aradhya, A. Ashfahani, F. Angelina, M. Pratama, R. F. de Mello, and S. Sundaram, :"Autonomous CNN (AutoCNN): A data-driven approach to network architecture determination," Information Sciences, vol. 607, pp. 638–653, Aug. 2022
24. G. Rajeshkumar : "Smart office automation via faster R-CNN based face recognition and internet of things," Measurement: Sensors, vol. 27, Jun. 2023.
25. S. R. Mishra, T. K. Mishra, G. Sanyal, A. Sarkar, and S. C. Satapathy, :"Real time human action recognition using triggered frame extraction and a typical CNN heuristic," Pattern Recognition Letters, vol. 135, pp. 329–336, Jul. 2020

26. Hong S., Lynn H.S.:Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction BMC Med. Res. Methodol., 20 (1) (2020)
27. Solaro N., Barbiero A., Manzi G., Ferrari P: A simulation comparison of imputation methods for quantitative data in the presence of multiple data patterns J. Stat. Comput. Simul., 88 (18) (2018)
28. William I., Rachmawanto E.H., Santoso H.A., Sari C.A.:Face recognition using facenet (survey, performance test, and comparison)2019 Fourth International Conference on Informatics and Computing (ICIC), IEEE (2019)