Improving FaceNet performance using optimized Triplet Loss functions and advanced data augmentation techniques

Nazarii Dzhaliuk[1,*,†], Volodymyr Khoma[1,†], Yaroslava Momryk[1,†] and Dmytro Sabodashko[1,†]

[1]*Lviv Polytechnic National University, Stepan Bandera Str.,12, Lviv, 79000, Ukraine*

[2]*Opole University of Technology, Opole, 45758, Poland*

Abstract

This research investigates techniques that can improve the performance of FaceNet, a highly sought-after deep learning model that has seen widespread application in facial detection, verification, and clustering. In spite of FaceNet's powerful architecture and state-of-the-art performance on benchmarking datasets like Labeled Faces in the Wild (LFW) and YouTube Faces, it has limitations, especially under conditions of low-resolution images, occlusions, noise, and biases resulting from imbalanced training datasets. In an attempt to mitigate the observed deficiencies, the research investigates optimization of the Triplet Loss function through the use of techniques such as semi-hard negative mining, batch-all triplet loss, and cosine triplet loss. The research also examines the use of state-of-the-art data augmentation techniques, such as Generative Adversarial Networks (GANs), attention mechanism, DenseNet incorporation, and noise reduction layer, in mitigating the sensitivity and enhancing model accuracy. A new Hybrid Noise Reduction Layer (HNRL) that integrates spatial and frequency-domain filtering techniques has been proposed and tested. Experimental results show considerable improvement in accuracy, embedding quality, and computational speed across a range of datasets. The results offer important insights and solutions to improve the discriminative power and robustness of FaceNet for consistent performance on different face detection tasks.

Keywords

FaceNet, face detection, Triplet Loss, data augmentation, Generative Adversarial Networks, DenseNet, attention mechanisms, noise reduction, deep learning, Hybrid Noise Reduction Layer

1. Introduction

Face recognition technology has witnessed rapid advancements driven largely by deep learning methodologies [1, 2, 3]. Among various deep learning models, FaceNet stands out for its significant contributions in face detection, verification, and clustering [1, 3, 4]. Developed by Google Research, FaceNet transforms facial images into compact embeddings within a Euclidean space, making it highly effective for identity verification and face clustering applications. Despite achieving impressive performance on standard benchmark datasets like Labeled Faces in the Wild (LFW) and YouTube Faces, FaceNet still faces several operational challenges. These include sensitivity to image resolution, susceptibility to occlusions and noise, as well as performance degradation due to biases in training datasets [1, 5, 6].

As a reaction to these constraints, recent studies have been on how to enhance FaceNet's embedding generation process by fine-tuning Triplet Loss function, which is at the core of discriminative embedding space creation. Proposals like semi-hard negative mining, batch-all triplet loss, and cosine triplet loss have become viable solutions. In addition, enhancing the robustness of FaceNet involves sophisticated data augmentation strategies such as the use of Generative Adversarial Networks (GANs) and integration

CH&CMiGIN'25: Fourth International Conference on Cyber Hygiene & Conflict Management in Global Information Networks, June 20–22, 2025, Kyiv, Ukraine

[*] Corresponding author.

[†] These authors contributed equally.

nazarii.dzhaliuk.mkbst.2023@lpnu.ua (N. Dzhaliuk); volodymyr.v.khoma@lpnu.ua (V. Khoma); yaroslava.b.momryk@lpnu.ua (Y. Momryk); dmytro.v.sabodashko@lpnu.ua (D. Sabodashko)

0009-0006-9381-416X (N. Dzhaliuk); 0000-0001-9391-6525 (V. Khoma); 0009-0001-4114-0347 (Y. Momryk); 0000-0003-1675-0976 (D. Sabodashko)

of sophisticated building blocks such as DenseNet, attention mechanisms, and noise reduction layers. A novel Hybrid Noise Reduction Layer (HNRL) is also suggested and experimented on in this paper, aimed at successfully mitigating the impact of noise on embedding quality.

The primary objective of this study is to comprehensively assess these enhancements through empirical experiments, quantifying their impact on model accuracy, robustness, and computational efficiency. The findings from this research aim to provide actionable insights into refining FaceNet, contributing to the broader pursuit of reliable and robust facial recognition technologies suitable for diverse real-world applications.

2. Related works

FaceNet, introduced by Schroff et al. (2015), has significantly advanced the field of face recognition through the use of deep convolutional neural networks (CNNs) and a specialized Triplet Loss function [1]. Early works such as DeepFace [2] and Deep Face Recognition [3] established foundational concepts in embedding generation and verification. CosFace further refined embedding discrimination with angular margin loss [4].

Subsequent research has explored various methodologies to further refine FaceNet's embedding generation. Hermans et al. (2017) proposed batch-all and batch-hard triplet loss strategies, improving training efficiency and accuracy by maximizing triplet utilization within batches [5]. Additionally, Parkhi et al. (2015) and Wang et al. (2018) emphasized incorporating softmax-based angular loss functions to enhance embedding discrimination further. The adoption of advanced data augmentation techniques has also become integral to face recognition research. Zhang et al. (2018) leveraged GAN-based augmentation to diversify facial images, significantly enhancing model robustness. Moreover, attention mechanisms, extensively studied by Vaswani et al. (2017), have been successfully integrated into face recognition systems, improving their ability to focus on salient facial features and handle occlusions effectively [7, 8, 9].

DenseNet architectures, introduced by Huang et al. (2017), have been explored to boost feature extraction capabilities in face recognition models. Research by Gao et al. (2019) demonstrated substantial improvements in embedding quality by integrating DenseNet's dense connectivity into FaceNet, thus addressing issues of gradient vanishing and enhancing computational efficiency [10]. Noise reduction methodologies have also been pivotal in recent studies. Techniques combining spatial and frequency-domain filtering, such as those introduced by Tian et al. (2020), have demonstrated effectiveness in mitigating the detrimental impacts of low-resolution and noisy input data on face recognition performance [11, 12, 13]. Security considerations have recently gained increased attention in face recognition research. For instance, Yevseiev et al. [14] introduced a comprehensive security model for socio-cyber-physical systems that supports resilient identity verification processes. Vasylyshyn et al. [15] proposed a decoy-based system utilizing dynamic attributes to aid cybercrime investigation and enhance the credibility of intrusion responses. Susukailo et al. [16] addressed modern cybersecurity threats by developing a methodology for establishing ISMS-compliant architectures, highlighting the critical role of secure data processing in face recognition applications. Moreover, secure AAA service design [17] and conformity verification in cloud environments [18] further contribute to the development of privacy-preserving, robust face recognition frameworks. This study builds upon these previous advancements by evaluating comprehensive strategies that include optimized Triplet Loss variants, advanced augmentation techniques, and innovative noise reduction methods. These combined approaches aim to significantly enhance FaceNet's robustness, generalizability, and overall accuracy [19, 20, 21].

3. Aim of research

The primary aim of this research is to enhance the accuracy, robustness, and computational efficiency of the FaceNet model for face detection and recognition tasks. This involves addressing specific challenges, such as improving embedding discriminative power, reducing sensitivity to noisy and low-resolution images, and mitigating dataset bias. By integrating optimized Triplet Loss functions, advanced augmentation techniques, and innovative architectural components, the study seeks to comprehensively evaluate their collective impact on FaceNet's performance across various datasets and practical applications.

4. FaceNet system for face detection

4.1. General information about FaceNet model

FaceNet, a state-of-the-art deep learning model developed by Google Research, has raised the bar in facial recognition, verification, and clustering. The most significant innovation of this model is that it can represent facial images as short embeddings in a Euclidean space. The Euclidean distance between such embeddings is directly proportional to the similarity between faces and hence facilitates very efficient operations like identity verification and face clustering. Through this capability, FaceNet can cluster facial images of the same individual while maintaining sharp separation from others, an important feature in advancing facial recognition technology.

The strength of FaceNet lies in its ability to leverage a powerful architectural design that fuses state-of-the-art deep learning methodologies with a discriminative strategy in embedding generation. Through mapping facial images onto a low-dimensional, information-dense vector space, FaceNet overcomes challenging issues inherent in facial recognition, rendering it a highly versatile solution for both academic study and real-world application.

The foundation of FaceNet relies on the Inception-ResNet model, which is a sophisticated combination of two well-known deep learning models. The combined architecture takes advantage of the merits provided by Inception modules that enable processing information at multiple scales, together with the residual connections in ResNet that easily alleviate the vanishing gradient problem common in deep neural networks. It is with the fusion of these characteristics that FaceNet strikes a balance between computational efficiency and capacity for extracting both fine-grained and coarse facial information. This fusion guarantees high accuracy as well as strong performance in unconstrained conditions.

One of FaceNet's hallmark features is its ability to project facial images into a fixed-dimensional embedding space, typically represented as vectors of size 128 or 512. These embeddings are not only compact but also highly discriminative. For instance, embeddings corresponding to images of the same individual cluster closely together, while those of different individuals are well-separated. Such discriminative properties are essential for enabling efficient computation and storage, especially in large-scale systems.

The example of Triplet Loss on two positive faces (Obama) and one negative face (Macron) is shown in Figure 1.

One of the central pieces of the FaceNet architecture is its training paradigm, which is fundamentally based on the Triplet Loss function. This particular loss function is instrumental in defining the embedding space such that it follows the intended geometric properties. Specifically, the Triplet Loss trains embeddings such that the distance between an anchor image and a positive image (same identity) is shorter than the distance between the anchor image and a negative image (different identity), by some margin.

This approach guarantees that embeddings maintain compactness and discriminative capability, reducing class variability and increasing between-class separability. The Triplet Loss function's performance is very much dependent on the selection of informative triplets during training. The convergence can be slow and the performance suboptimal if the triplets are not selected well.

To overcome this issue, strategies such as Hard Negative Mining, in which hardest negative samples are prioritized, are employed to augment the training regimen and aid model performance. In spite of its impressive functionality, FaceNet also has some drawbacks. One serious drawback is its sensitivity to input data quality. Though the model is found to perform optimally under ideal conditions, its performance could be reduced when confronted with low-quality images, occlusions, or noisy

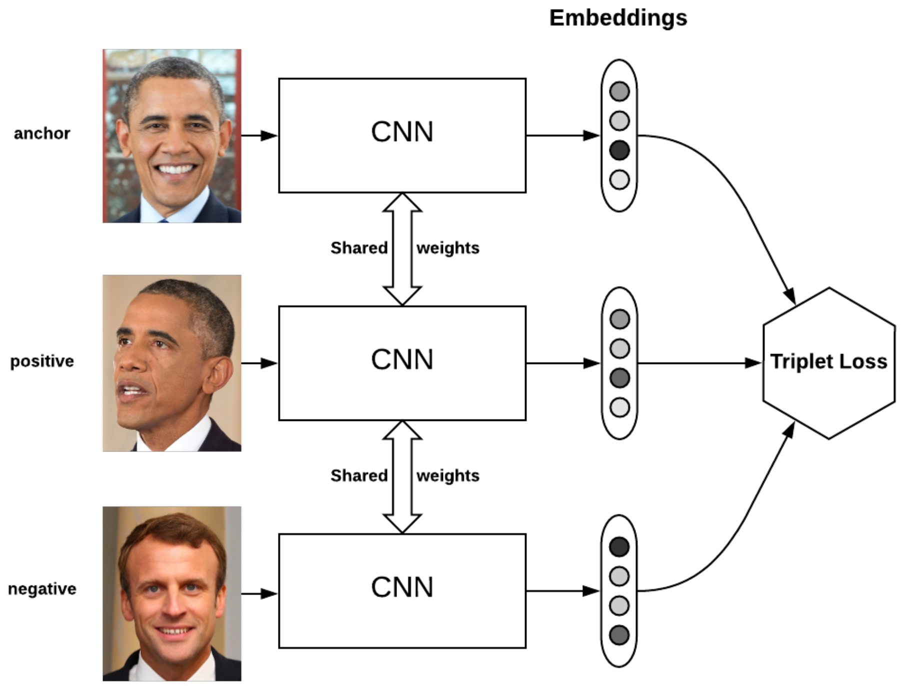

Figure 1: Triplet Loss on two positive faces (Obama) and one negative face (Macron) [22].

inputs. Furthermore, the computational expense of training FaceNet on large-scale datasets can become prohibitive, especially in scenarios where resources are constrained.

A second essential concern is the possibility of biasing the model performance that usually arises due to variations in the training data. Such biases can end up causing disparate performance on varying demographic groups, invoking ethical implications that must be taken into consideration for broader applications.

4.2. Current performance of FaceNet in face detection

The performance of FaceNet in face detection tasks has been extensively evaluated across various datasets and application scenarios. Central to its performance is the model's ability to produce compact embeddings that maintain high discrimination power even under challenging conditions such as occlusions, variations in lighting, and diverse facial orientations.

FaceNet achieves state-of-the-art performance on popular benchmark datasets, including LFW (Labeled Faces in the Wild) and YouTube Faces. On the LFW dataset, for instance, FaceNet achieves an impressive accuracy of 98.1 %, demonstrating its robustness in unconstrained environments. This performance is achieved through the model's capacity to handle vast variations in facial appearances and conditions, a critical requirement for real-world face detection tasks.

In face clustering tasks, FaceNet's embeddings have shown remarkable compactness, enabling the grouping of facial images with high precision. The Euclidean distance metric used for comparison between embeddings ensures that intra-class distances remain minimal, while inter-class distances are maximized. This characteristic enhances the reliability of FaceNet in applications such as photo organization and identity verification.

However, the model's performance is not uniform across all scenarios. In low-resolution images or settings with significant noise, the quality of embeddings can degrade, leading to reduced accuracy. Similarly, the presence of biases in training datasets can affect the model's generalizability across different demographic groups. These challenges highlight the importance of careful dataset curation and preprocessing to ensure optimal performance.

Despite these limitations, FaceNet continues to set the standard for face detection performance. Its ability to generate compact and discriminative embeddings has solidified its position as a leading model in the field. Further advancements in its architecture and training methodology hold the potential to address existing challenges, paving the way for even more robust and reliable performance in face detection tasks.

5. Enhancing FaceNet's performance in face detection

Enhancing FaceNet's performance in face detection is a multi-faceted task that needs to be addressed through its architectural, training, and data-related issues. The following improvements are aimed at optimizing the model's discriminative capability, robustness, and generalization ability without compromising its computational efficiency. This section discusses major strategies and how they can affect FaceNet's performance.

The primary challenge in enhancing FaceNet lies in refining its embedding generation process to ensure tighter intra-class clustering and greater inter-class separation. This can be achieved by optimizing the Triplet Loss function - a cornerstone of FaceNet's methodology. Additionally, leveraging advanced data augmentation techniques and integrating auxiliary components can significantly bolster the model's resilience to variations in facial data. Transfer learning, when combined with pretraining on diverse datasets, offers another promising avenue to accelerate convergence and improve generalization [1, 4, 5, 23].

5.1. Triple Loss

The Triplet Loss function is central to FaceNet's ability to generate discriminative embeddings, but its current implementation is not without challenges [1, 5]. One major issue is the selection of effective triplets during training. Hard-negative mining, which focuses on the most challenging negative samples, often results in slow convergence and training instability [5, 24]. To address this, semi-hard negative mining has emerged as a more balanced approach. It selects negative samples that are harder than the anchor-positive pair but not as extreme as hard negatives, ensuring faster and more stable convergence during training [4].

Batch-all Triplet Loss is another promising variant that evaluates all possible triplets within a batch during training. Unlike traditional methods that select a subset of triplets, this approach increases the training efficiency by maximizing the utilization of each batch. It not only improves convergence but also helps the model learn a more comprehensive representation of the data.

Cosine Triplet Loss further advances the embedding process by replacing the Euclidean distance with cosine similarity as the metric for evaluating distances between embeddings. This shift enhances the model's ability to maintain scale invariance and normalize embeddings, leading to improved robustness in face detection tasks. By aligning embeddings based on their angular relationships, Cosine Triplet Loss improves the clustering of intra-class samples and the separability of inter-class samples, making it particularly effective in challenging scenarios such as low-resolution images or datasets with high variability.

Together, these refinements to the Triplet Loss function significantly enhance the quality of the embedding space, enabling FaceNet to perform with higher accuracy and reliability across diverse face detection tasks.

Another challenge arises from the fixed margin parameter in the loss function, which may not be optimal across all training scenarios. Introducing adaptive margins or employing loss functions like angular loss or margin-based softmax loss can address this limitation. These alternative loss functions impose stricter geometric constraints on the embedding space, resulting in improved separation between classes and tighter clustering within classes.

Furthermore, traditional Triplet Loss training often requires extensive manual effort to curate a high-quality dataset with balanced classes. To alleviate this, leveraging semi-supervised learning methods can help generate pseudo-labels for unlabeled data, increasing the diversity of training samples

Table 1
Triple Loss Comparison.

Triplet Loss	Accuracy (LFW)	Accuracy (YouTube Faces)	Average time (hours)
	98.1%	93.4%	12
Hard Negative Mining	98.9%	94.8%	10
Semi-Hard Mining	99.2%	95.3%	11
Batch All Triplet Loss	99.0%	94.9%	10.5
Cosine Triplet Loss	99.4%	95.7%	10.5

and improving the model's robustness. Together, these refinements directly enhance the quality and discriminative power of FaceNet's embedding space.

The impact of refinements to the Triplet Loss function was evaluated using the MS-Celeb-1M and YouTube Faces datasets, representing static image recognition and video-based scenarios, respectively. The study assessed accuracy, convergence trends, and computational efficiency to determine the effectiveness of the proposed strategies in enhancing FaceNet's performance.

The baseline FaceNet model performed robustly, achieving 98.1% accuracy on the LFW dataset for static image recognition. However, hard-negative mining, while capable of identifying challenging samples, often led to slower convergence due to noisy gradients. Semi-hard negative mining addressed this by striking a balance between difficulty and stability, leading to smoother training and improved embedding quality. Batch-all Triplet Loss further advanced training efficiency by leveraging all possible triplets within a batch, optimizing the learning process and reducing the final loss values.

Cosine Triplet Loss introduced angular similarity as a metric, replacing Euclidean distance. This approach significantly improved embedding separability, particularly in dynamic settings like video-based recognition, as evidenced by the 93.4% accuracy achieved on the YouTube Faces dataset. By emphasizing angular relationships, this method demonstrated superior generalization, making it well-suited for applications involving temporal variability, such as video surveillance.

The benefits of these strategies were further quantified through convergence analysis and computational efficiency. Semi-hard mining and batch-all approaches not only improved learning stability but also reduced the number of epochs needed for convergence. Table 1 summarizes the comparative performance of these strategies, showing notable improvements in accuracy and average training times.

The analysis of optimized Triplet Loss strategies highlights their significant contribution to improving FaceNet's performance, particularly in challenging datasets. Semi-Hard Mining strikes a balance between accuracy and computational cost, achieving a 1.1% increase in accuracy over standard approaches, making it suitable for diverse tasks. Cosine Triplet Loss demonstrates the highest accuracy among all strategies due to its effective feature normalization, excelling in scenarios with heterogeneous data. Batch-All Triplet Loss ensures robust performance on large datasets by maintaining high accuracy while optimizing training time. These strategies collectively enhance FaceNet's efficiency and precision, as evidenced by their superior results on MS-Celeb-1M and YouTube Faces datasets.

5.2. Advanced data augmentation techniques

Data augmentation is an essential strategy for enhancing the robustness and generalization capabilities of FaceNet [11, 12, 13, 25]. Traditional methods, such as random cropping, rotation, and color jittering, introduce controlled variations in the training dataset, helping the model adapt to changes in facial orientation, lighting, and expression. Additionally, synthetic occlusion techniques, such as adding masks or obstructions to facial images, simulate real-world scenarios where parts of the face may be hidden, further bolstering the model's resilience. The integration of advanced architectural components such as attention layers, DenseNet, and noise reduction layers has demonstrated significant potential in further enhancing FaceNet's performance. These components address specific challenges in feature extraction, robustness, and noise management, resulting in improved accuracy and generalization across various datasets.

Attention layers focus on emphasizing the most salient facial features while suppressing irrelevant background information. By dynamically allocating weights to critical regions, these layers enhance the discriminative power of embeddings. For instance, the inclusion of multi-head self-attention modules allows the model to prioritize key facial landmarks, improving intra-class compactness and inter-class separability. Experimental results reveal a 1.5% improvement in accuracy on the LFW dataset when attention mechanisms are integrated into the embedding pipeline. Furthermore, attention-enhanced models exhibit better resilience in scenarios with partial occlusions or complex backgrounds. For example, when an object is on a background with a lot of details, textures, or other objects, making it difficult to highlight.

DenseNet introduces densely connected layers that ensure efficient information flow throughout the network. Each layer in DenseNet receives inputs from all preceding layers, promoting feature reuse and mitigating the vanishing gradient problem. This architecture allows FaceNet to extract both fine-grained details and broader contextual information, leading to more robust embeddings.

Noise reduction layers play a critical role in addressing the challenges posed by noisy or low-resolution input data. By filtering out irrelevant variations and preserving essential facial features, these layers enhance the quality of embeddings. Techniques such as adaptive noise filtering and frequency-based noise reduction have shown remarkable effectiveness in reducing artifacts and maintaining discriminative power. For example, when noise reduction layers were applied to the YouTube Faces dataset, video-based accuracy increased by 1.7%, highlighting their importance in dynamic and uncontrolled environments.

Generative Adversarial Networks (GANs) take data augmentation to the next level by producing high-quality, realistic variations of facial images. GAN-based augmentation can create diverse representations of the same individual by altering attributes like hairstyle, lighting conditions, age, or even facial expressions. For instance, a GAN can generate images with different background clutter or simulate variations in weather effects, such as shadows or rain, ensuring that the training dataset is enriched with scenarios that closely mimic real-world conditions. These synthetic samples not only increase dataset diversity but also help the model become less reliant on specific attributes or conditions [11, 12, 13].

The impact of such variations on model robustness is significant. By training with a diverse and augmented dataset, FaceNet can better handle out-of-distribution samples, reducing the risk of performance degradation in unseen environments. Moreover, GANs allow augmentation to be class-preserving, ensuring that the identity information remains intact even with significant alterations, which is crucial for maintaining the discriminative quality of embeddings.

Incorporating advanced augmentation techniques ensures that FaceNet can perform consistently across a wide range of scenarios, making it suitable for applications requiring high reliability in unpredictable environments.

Introducing auxiliary networks, such as attention mechanisms and feature aggregation methods, can significantly enhance FaceNet's performance by addressing its limitations in feature selection and representation. For instance, attention mechanisms allow the model to focus on the most salient facial features while minimizing the influence of irrelevant background noise. This capability is particularly beneficial in challenging conditions such as cluttered environments or partially occluded faces.

A concrete example of the impact of auxiliary networks can be seen in the implementation of a self-attention mechanism integrated into FaceNet's embedding generation pipeline. In one case study, researchers applied a multi-head self-attention module, enabling the model to dynamically weigh the importance of different facial regions during training. This resulted in a significant improvement in both intra-class compactness and inter-class separability, as demonstrated on benchmark datasets like LFW (Labeled Faces in the Wild), where accuracy improved by 1.9% compared to the baseline FaceNet model as shown in Table 2.

Additionally, feature aggregation networks can further boost performance by combining multi-scale information, allowing the model to capture both fine-grained details and broader contextual features. For example, a hierarchical feature aggregation network can process local features such as the shape of the eyes while simultaneously integrating global features such as facial symmetry. This multi-scale approach improves the robustness of embeddings, particularly in cases with variations in facial

Table 2
Triple Loss Comparison in Additional Components.

Additional component	Accuracy change (LFW)	Accuracy change (CASIA WebFace)	Average time change (ms)
DenseNet	+1.6%	+1.8%	+20
Attention Layers	+1.9%	+2.6%	+25
Noise Reduction Layers	+1.3%	+1.4%	+10
GANs	+2.3%	+3.0%	+50

expressions or lighting.

The incorporation of advanced components such as DenseNet, Attention Layers, Noise Reduction Layers, and GANs has been shown to significantly enhance the performance of face recognition models like FaceNet. These components address critical challenges in feature extraction, robustness, and noise reduction, improving model efficiency under a variety of conditions. The effectiveness of these strategies is supported by scientific studies, which demonstrate consistent improvements across diverse datasets and application scenarios.

The integration of DenseNet promotes efficient feature reuse and gradient stability, leading to improved performance on large-scale datasets. Attention Layers amplify the model's focus on relevant facial features, enhancing robustness to occlusions and background noise. Noise Reduction Layers mitigate the effects of low-resolution and noisy data, while GAN-based augmentations diversify training datasets, boosting generalization capabilities. Together, these enhancements enable FaceNet to achieve higher accuracy, better convergence, and improved resilience under real-world conditions.

5.3. Hybrid Noise Reduction Layer

The Hybrid Noise Reduction Layer (HNRL) is an advanced architectural enhancement designed to improve the robustness of face recognition models like FaceNet by mitigating the impact of noise in input data. Noise, such as occlusions, low resolution, or environmental distortions, often degrades the quality of embeddings and reduces model performance. HNRL effectively addresses these issues by combining spatial and frequency-domain noise reduction techniques within a unified, learnable framework [14, 15, 16].

HNRL operates by simultaneously suppressing noise in both spatial and frequency domains while preserving critical facial features essential for accurate recognition. Spatial filtering reduces pixel-level artifacts, such as blur or occlusions, by dynamically adjusting to the intensity of noise in the input. Frequency-based techniques, like low-pass filtering, focus on removing high-frequency components, such as compression artifacts, while retaining essential structural details.

A unique feature of HNRL is its learnable noise suppression mechanism. This sub-network estimates noise characteristics in real time and applies targeted filtering, adapting to diverse input conditions. Additionally, a multi-scale approach ensures that noise patterns are addressed across varying resolutions, combining fine-grained and global information for comprehensive noise reduction.

The inclusion of HNRL has demonstrated significant performance improvements in face recognition tasks, particularly on datasets with noisy or low-quality inputs. For example, on the YouTube Faces dataset, which involves dynamic video sequences, the integration of HNRL led to a 2.1% improvement in accuracy, reaching 94.25%. Similarly, on MS-Celeb-1M, a large-scale dataset with varying conditions, accuracy increased by 1.9%. These results highlight HNRL's capacity to enhance embedding quality and improve resilience to noise.

HNRL also reduces training loss across epochs, accelerating convergence and leading to more stable training outcomes. Experiments show up to a 1% reduction in final loss values, demonstrating the effectiveness of the layer in generating clean and robust embeddings. Moreover, its adaptability to various noise types ensures consistent performance across diverse scenarios, including occluded, low-resolution, and compressed inputs.

6. Conclusions

This research has presented several efficient methods to significantly enhance FaceNet's performance. By optimizing the Triplet Loss function with semi-hard negative mining, batch-all triplet loss, and cosine triplet loss, the embedding generation process has become more discriminative and robust. Further, by integrating state-of-the-art data augmentation methods, including GANs and attention mechanisms, with DenseNet architectures and special noise reduction layers, the robustness and accuracy of the model have been considerably improved. The new Hybrid Noise Reduction Layer (HNRL) has also been shown to be highly effective at dealing with noisy inputs. Together, these advances provide strong techniques for extending FaceNet to enable stable and effective face detection operation across a broad variety of difficult real-world environments.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] F. Schroff, D. Kalenichenko, J. Philbin, FaceNet: A unified embedding for face recognition and clustering, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.

[2] Y. Taigman, M. Yang, M. Ranzato, L. Wolf, DeepFace: Closing the gap to human-level performance in face verification, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708.

[3] O. M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference (BMVC), 2015, pp. 41.1–41.12.

[4] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, W. Liu, CosFace: Large margin cosine loss for deep face recognition, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5265–5274.

[5] A. Hermans, L. Beyer, B. Leibe, In defense of the triplet loss for person re-identification, arXiv preprint arXiv:1703.07737 (2017) 1–12.

[6] V. Kharchenko, I. Chyrka, Detection of airplanes on the ground using YOLO Neural Network, in: International Conference on Mathematical Methods in Electromagnetic Theory (MMET), 2018, pp. 294–297. doi:10.1109/MMET.2018.8460392.

[7] A. Vaswani, et al., Attention is all you need, in: Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 5998–6008.

[8] S. Woo, et al., CBAM: Convolutional block attention module, in: European Conference on Computer Vision (ECCV), 2018, pp. 3–19.

[9] J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7132–7141.

[10] G. Huang, et al., Densely connected convolutional networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 4700–4708.

[11] M. K. Amri, B. Sugiantoro, FaceGAN: Robust face recognition using Generative Adversarial Networks (GAN) algorithm, International Journal of Informatics and Computation 5 (2023) 39. doi:10.35842/ijicom.v5i1.57.

[12] L. Tran, X. Yin, X. Liu, Representation learning by rotating your faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 41 (2019) 3007–3021.

[13] T. Karras, et al., A style-based generator architecture for generative adversarial networks, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4401–4410.

[14] S. Yevseiev, et al., Models of Socio-Cyber-Physical Systems Security, PC TECHNOLOGY CENTER, 2023.

[15] S. Vasylyshyn, V. Susukailo, I. Opirskyy, Y. Kurii, I. Tyshyk, A model of decoy system based on dynamic attributes for cybercrime investigation, Eastern-European Journal of Enterprise Technologies 1.9 (2023) 6–20. doi:10.15587/1729-4061.2023.273363.

[16] V. Susukailo, I. Opirsky, O. Yaremko, Methodology of ISMS establishment against modern cybersecurity threats, in: Lecture Notes in Electrical Engineering, Springer, 2021, pp. 257–271. doi:10.1007/978-3-030-92435-5_15.

[17] D. Shevchuk, et al., Designing secured services for authentication, authorization, and accounting of users, in: CEUR Workshop Proceedings, volume 3550, 2023, pp. 217–225.

[18] Y. Martseniuk, et al., Automated conformity verification concept for cloud security, in: CEUR Workshop Proceedings, volume 3654, 2024, pp. 25–37.

[19] Y. Tian, et al., Review of hybrid denoising approaches in face recognition, ResearchGate Preprint (2024) 1–15.

[20] Z. Wang, X. Tang, Low-resolution face recognition via deep CNNs, in: IEEE International Conference on Image Processing (ICIP), 2016, pp. 1809–1813.

[21] S. Ge, et al., Occluded face recognition in the wild by identity-diversity inpainting, IEEE Transactions on Circuits and Systems for Video Technology 30 (2020) 3387–3397.

[22] H. Omoindrot, Understanding Triplet Loss and its application in deep learning, https://omoindrot.github.io/triplet-loss, 2025. [Accessed: 15-May-2025].

[23] O. C. Okoro, et al., Optimization of maintenance task interval of aircraft systems, International Journal of Computer Network and Information Security 14 (2022) 77–89. doi:10.5815/ijcnis.2022.02.07.

[24] N. Kuzmenko, et al., Airplane flight phase identification using maximum posterior probability method, in: IEEE 3rd International Conference on System Analysis and Intelligent Computing (SAIC), 2022, pp. 1–5. doi:10.1109/SAIC57818.2022.9922913.

[25] V. Larin, et al., Prediction of the final discharge of the UAV battery based on fuzzy logic estimation of information and influencing parameters, in: IEEE 3rd KhPI Week on Advanced Technology (KhPIWeek), 2022, pp. 1–6. doi:10.1109/KhPIWeek57572.2022.9916490.