Masked Face Recognition via CNN Embeddings Optimized with a Discriminative Quadruplet Loss Function

Siham Ahmam, Yazid Safiny, Nidal Lamghari, Abdelghani Ghazdali
Laboratory of Process Engineering, Computer Science, and Mathematics, University Sultan Moulay Slimane, Khouribga, Morocco
E-mail: siham.ahmam@usms.ac.ma, yazid.safiny@usms.ac.ma, n.lamghari@usms.ma, a.ghazdali@usms.ma

Keywords: Face recognition, face mask, quadruplet loss, triplet loss, deep metric learning, convolutional neural network

Received: September 28, 2025

Masked face recognition remains a challenging problem because masks occlude key facial regions that are crucial for identity verification. In this paper, we propose a deep convolutional architecture composed of stacked Conv–BatchNorm–MaxPooling blocks followed by dropout, a flatten layer, and a dense embedding layer trained with an improved quadruplet loss. Each quadruplet consists of two images from the same identity and two from different identities, enforcing compact intra-class clusters and well-separated inter-class distributions in the embedding space. We investigate three similarity measures on the learned embeddings: Euclidean distance, Manhattan distance, and a learned similarity network. The best performance is obtained with Euclidean distance: on our ENSA-MFRD dataset of 32,186 masked face images collected from university students, the proposed model reaches an accuracy of 99.27%, outperforming the standard triplet loss and the original quadruplet loss by 0.62% and 0.72%, respectively. Using the learned similarity network, our approach also surpasses the triplet and original quadruplet losses by 37.5% and 11.87%, respectively. The model is further evaluated on four public benchmarks—COMASK20, MFRD-80K, CASIA-WebMaskedFace, and LFW-SMFD—where it consistently improves accuracy, F1-score over both baselines. These results demonstrate that the proposed architecture and enhanced quadruplet loss yield robust and discriminative representations for masked face recognition across diverse datasets and acquisition conditions.

Povzetek: Predstavljen je model globokega učenja za prepoznavanje obrazov z maskami, okrepljen z izboljšano štirikratno izgubo. Znanstveni prispevek je visoka točnost in robustnost metode, potrjena na lastnem in javnih naborih podatkov.

1 Introduction

Facial recognition technologies have increased in usage in various businesses, including shopping malls, health centres, schools, and transportation stations such as airports. The growing significance of effective and precise facial recognition systems is reflected in this broad usage. Due to COVID-19, an illness caused by the coronavirus family, the world is experiencing a profound health crisis. The main means of COVID-19 viral transmission between individuals is through respiratory droplets. To stop the pandemic from spreading, the World Health Organization recommends wearing a face mask as one of the proposed solutions to combat this disease [1]. Wearing a face mask is not limited to pandemics, but can also be used in laboratories and other polluted spaces. However, following these safety recommendations seriously puts existing security systems that are based on face recognition to the test. Although preserving lives is important, there is a pressing need to identify individuals wearing masks without revealing their identities. For instance, there are numerous places where people cooperate with cameras, such as immigration checkpoints and grounds access control, which present a challenge for face recognition since obscured sections are essential for face detection and recognition. Ngan et al.[2] tested algorithms developed prior to the COVID-19 pandemic to assess their ability to identify masked faces. However, they concluded that performance degrades with face mask datasets. Therefore, developers of facial recognition systems had to consider this challenge and improve the performance of their system by taking masked faces into account. The main contributions of this work are as follows:

- **Enhanced Quadruplet Loss Function:** We propose an improved quadruplet loss designed to improve both intra-class compactness and inter-class separability under mask occlusion and enhance the discriminative power of feature embeddings for masked face recognition.
- **Deep Learning-Based Recognition Model:** We develop a deep learning model that leverages an enhanced loss function to improve masked face recog-

nition accuracy.

– **Custom Masked Face Dataset:** We present a properly collected and curated dataset of 32,186 masked face images, specifically designed to train and evaluate our model under real-world masked conditions.

The remainder of this paper is organized as follows: Section 2 reviews related work. Section 3 details the dataset, model architecture, and the proposed improved loss function. Section 4 presents and discusses the experimental results. Finally, Section 5 concludes the study and outlines directions for future research.

2 Related work

Face recognition has attracted sustained research interest, leading to a broad spectrum of architectures and loss functions. Schroff et al. [3] introduced FaceNet, which learns 128-dimensional embeddings using a triplet loss and achieves strong performance on large-scale, unconstrained face recognition benchmarks, but without explicitly addressing heavy occlusions such as masks. Goel et al. [4] evaluated pre-trained models (e.g., VGG and FaceNet) for sibling identification and showed that, in scenarios with high inter-class similarity and partial occlusion, generic face embeddings are less discriminative.

Metric-learning losses have been extensively explored to improve intra-class compactness and inter-class separability. Bromley et al. [5] developed a siamese network with a distance-based contrastive loss for image-pair comparison. Chen et al. [6] extended this idea with a quadruplet loss to further increase inter-class margins in person re-identification, thereby improving rank-based retrieval performance. For masked face recognition, Boutros et al. [7] proposed a self-restrained triplet loss that emphasizes genuine positive pairs to better cope with the structured occlusion induced by masks. Huang et al. [8] introduced PLFace, a progressive learning framework that balances masked and unmasked samples during training to mitigate mask-related bias. Salim et al. [9] proposed a pre-processing technique that preserves facial landmarks while zeroing out the occluded region, which improves masked-face recognition accuracy without retraining the backbone network. Golwalkar et al. [10] presented FaceMaskNet-21, a deep metric-learning framework that outputs 128-dimensional encodings specifically designed for masked faces. Sikha et al. [11] combined a modified VGG16 architecture with cropping of the unmasked upper-face region to extract more reliable features under mask occlusion.

Several works focus on particular sensing configurations or fusion strategies. Du et al. [12] addressed NIR–VIS masked face recognition by combining heterogeneous semi-siamese training with 3D reconstruction to bridge cross-spectral domain gaps. Omar et al. [13] proposed a lightweight convolutional neural network (CNN) trained on HMFD using Adam optimization and sparse categorical cross-entropy, targeting efficient masked-face recognition for deployment on constrained devices. Zhang et al. [14] and Mahmoud et al. [15] provided surveys of masked face recognition, detection, and unmasking, covering feature extraction, classification, multimodal fusion, and dataset design. Huang et al. [16] designed a system that combines domain adaptation and self-attention to align features across masked and unmasked faces. Ge et al. [17] proposed CVSAN, which integrates convolutional layers and visual self-attention and is trained on simulated masked data using an angular-margin loss. Alqaralleh et al. [18] fused information from visible frontal and profile regions using BSIF descriptors and CNNs at multiple fusion levels to exploit unoccluded facial areas.

In parallel, some contributions focus primarily on data generation and mask detection. Cabani et al. [19] constructed the MaskedFace-Net dataset by applying deformable masks to FFHQ images, providing large-scale synthetic masked faces for training and benchmarking. Zhang H. et al. [20] proposed AI-YOLO, which augments YOLO with a selective kernel, spatial pyramid pooling, feature fusion, and CIoU loss for robust mask detection in complex scenes. Oulad-Kaddour et al. [21] developed a CNN-based model for mask-wearing prediction and gender classification, illustrating the importance of masked data in multi-task settings. Aly [22] combined ResNet-50, CBAM, and temporal convolutional networks for facial expression recognition in online learning environments, demonstrating that temporal modeling and attention can partially compensate for occlusions.

To provide a more compact and quantitative comparison, Table 1 summarizes the main characteristics of these approaches in terms of datasets, training objectives, and evaluation metrics.

Despite this rich body of work, several concrete limitations remain with respect to the problem addressed in this paper. FaceNet [3] is designed for fully visible faces and is neither trained nor evaluated under mask occlusion. Goel et al. [4] consider sibling recognition with partial occlusions, but the occluded regions are limited (eyes, nose, forehead) and do not correspond to realistic full-face masks. Metric-learning approaches such as the Siamese network of Bromley et al. [5] and the quadruplet loss of Chen et al. [6] achieve strong performance on signatures or person re-identification, yet they are not analyzed in the context of masked faces and do not investigate the role of different similarity metrics on heavily occluded embeddings. Mask-specific methods (e.g., self-restrained triplet loss [7], PLFace [8], zeroing-based preprocessing [9], FaceMaskNet-21 [10], VGG16-based cropping [11], lightweight HMFD CNNs [13], frontal–profile fusion [18]) typically focus on one or a few datasets, often with synthetic masks or controlled conditions, and rarely report systematic cross-dataset evaluations or robustness under challenging illumination and acquisition variability. Dataset and detection-oriented works (MaskedFace-Net [19], AI-YOLO [20], mask-wearing and gender pre-

Table 1: Summary of methods discussed in the related work: datasets, loss functions, and key evaluation metrics

Work	Datasets used	Loss/training objective	Metrics
Schroff et al. [3]	LFW, YTF	Triplet loss (metric learning)	Acc. 99.63% (LFW), 95.12% (YTF).
Goel et al. [4]	Sibling face dataset (masked/unmasked regions)	Softmax cross-entropy on deep embeddings	Acc. 98%.
Bromley et al. [5]	Signature / image-pair datasets	Siamese contrastive (distance-based) loss	Acc. 97%.
Chen et al. [6]	CUHK03, Market-1501, VIPeR, etc.	Quadruplet loss + classification loss	Rank1 81.00% on CUHK01
Boutros et al. [7]	Real/synthetic masked FR datasets	Self-Restrained Triplet (SRT) loss	-
Huang B. et al. [8]	Large-scale FR + masked benchmarks	PLFace margin-based loss (progressive learning)	Average 79.40%.
Salim & Surantha [9]	MFR2, RMFRD, SMFRD (masked variants)	Original FR loss (e.g. ArcFace); mask-zeroing preprocessing	Acc. 86.46%;
Golwalkar & Mehendale [10]	FaceMaskNet-21 (masked faces)	Softmax loss + metric comparison in 128-D space	Acc. 88.92%.
Sikha & Bharath [11]	MFR2, RMFRD, SMFRD (upper-face crops)	Softmax cross-entropy (VGG16–RF)	Acc.85.002%.
Du et al. [12]	CASIA NIR-VIS 2.0, BUAA-VisNir (with masks)	Triplet / center + cross-entropy (semi-Siamese)	Rank-1 98.53%.
Omar et al. [13]	HMFD (HSTU Masked Face Dataset)	Sparse categorical cross-entropy (CNN)	Acc. 97% (HMFD)
Alqaralleh et al. [18]	Frontal & profile masked-face datasets (RMFRD/SMFRD-like)	Cross-entropy (CNN with BSIF fusion)	Acc 99.83%.
Aly [22]	RAF-DB, FER2013, CK+, KDEF (FER)	Cross-entropy (ResNet-50 + CBAM + TCNs)	Acc. 97%

diction [21], expression recognition under partial occlusion [22]) either provide data or solve auxiliary tasks (mask detection, gender, expression) but do not directly tackle large-scale masked face identification across heterogeneous real and synthetic datasets.

In light of these observations, our work is designed to address three specific gaps. First, instead of treating quadruplet loss as a generic metric-learning tool, we propose an *improved* quadruplet-loss formulation tailored to masked face recognition and experimentally compare it against both the standard triplet loss and the original quadruplet loss under identical training and evaluation protocols. Second, we explicitly study the impact of the similarity metric on masked embeddings by evaluating Euclidean distance, Manhattan distance, and a learned neural similarity, and we show that Euclidean distance yields the best recognition accuracy in our setting. Third, we move beyond single-dataset evaluations by training and testing on a large real-mask dataset (ENSA-MFRD) and four heterogeneous public benchmarks (COMASK20, MFRD-80K, CASIA-WebMaskedFace, LFW-SMFD), including cross-dataset experiments and reporting accuracy and F1-score. This systematic analysis of loss functions, similarity metrics, and cross-dataset behavior under realistic masked conditions distinguishes our contribution from the existing literature.

3 Methodology

3.1 Dataset

Creating a powerful recognition system is based on compiling a large, high-quality dataset of masked faces. We have therefore decided to collect a large dataset by filming students from our university in various poses. We have collected a dataset named the ENSA-Masked Face Recognition Dataset (ENSA-MFRD), which contains 32,186 images of 126 students. Each student has several images, varying in number from 27 to 528. The construction of the dataset involved a multi-stage pipeline. Initially, video

recordings of students were captured under controlled conditions. These videos were subsequently decomposed into individual frames to generate a set of static images. Facial regions within these images were automatically detected using the RetinaFace algorithm, ensuring robust and accurate localization. The detected faces were then cropped and extracted to form the core dataset. Finally, the resulting face images were systematically organized into directories, with each folder corresponding to a unique individual, thereby facilitating identity-specific data management. An example of the collected dataset images is shown in Figure. 1.

Figure 1: A representative sample from our ENSA-MFRD dataset

3.2 Loss function

The quantification of the difference between the model predictions and the actual observations in the training data is measured by a loss function named quadruplet loss, which is based on a combination of the ideas of triplet loss [3] and quadruplet loss [6]. Triplet loss, which will be mentioned in Equation 1, focuses on intra-class distance reduction.

$$L_{Tiplet} = \sum_{i,j,l}^{N} [f(x_i, x_j)^2 - f(x_i, x_l)^2 + \alpha]_+ \quad (1)$$

The quadruplet loss expands the triplet loss by adding an extra term. This is included to increase the inter-class distance to further enhance class separation within the set. Their mathematical formula will be presented in Equation 2.

$$\begin{aligned} L_{Quadruplet} = &\sum_{i,j,l}^{N} \Big[g(y_i, y_j)^2 - g(y_i, y_l)^2 + \alpha \Big]_+ \\ &+ \sum_{i,j,l,k}^{N} \Big[g(y_i, y_j)^2 - g(y_l, y_k)^2 + \beta \Big]_+ \end{aligned} \quad (2)$$

Where g represents a learned metric and N denotes the total number of quadruplets. The embeddings y_i and y_j are extracted from two distinct images of the same individual, denoted as S_{ij}. Conversely, y_l and y_k are embeddings obtained from images of two different individuals, represented as S_l and S_k, respectively, ensuring that $S_{ij} \neq S_l \neq S_k$. The parameters α and β define two margin constraints.

Our proposed loss function builds upon the formulation detailed in Equation 3, retaining the primary term common to both functions while incorporating an additional constraint to further minimize intra-class variation. This modification is strategically introduced to enhance the discriminative power of the learned embeddings by reinforcing compactness within the same identity while maximizing inter-class separability. The quadruplet-based loss function utilizes four input images, each sampled from a distinct individual, to optimize feature space representations. Specifically, the quadruplet consists of an Anchor and a Positive, which correspond to two images of the same individual, alongside two Negative samples: Negative 1, sourced from a second identity, and Negative 2, obtained from a third. An illustration of the quadruplet sampling strategy is provided in Figure. 2. This structured formulation enforces a more robust separation between different identities while preserving intra-class cohesion. The primary goal of the loss function is to minimize intra-class variations and promote effective separation between different identity representations. In particular, the role defines a constraint by which the distance within the same class (the distance between embeddings for the same person) is substantially smaller than inter-class distances. The constraint can be mathematically stated as $d_1 \ll d_2$ and $d_1 \ll d_3$, where d_1 is the spatial distance between the Positive sample and the Anchor sample, both of which are images of the same individual. d_2 is the spatial distance between the Anchor and the first Negative 1 sample. and d_3 is the spatial distance between the Anchor and the Negative 2 sample. The loss function can be defined mathematically as

Figure 2: Example of the quadruplet set extracted from our dataset

$$\begin{aligned} L_{ImQuad} = &\sum_{i,j,l}^{N} \Big[h(x_i, x_j)^2 - h(x_i, x_l)^2 + \alpha \Big]_+ \\ &+ \sum_{i,j,k}^{N} \big[h(x_i, x_j)^2 - h(x_i, x_k)^2 + \beta \big] + \end{aligned} \quad (3)$$

where x_i, x_j, x_l, and x_k are the embeddings of Anchor,

Positive, Negative 1, and Negative 2, respectively. $h(\cdot,\cdot)$ is the learned metric function. N is the set of quadruplets that was employed within the training step. α and β are introduced margin parameters enabling sufficient separation of positive and negative pairs. Such margins are used as thresholds to prevent the production of trivial solutions, and, in most cases, need to be obtained empirically.

With the use of quadruplet-based learning, this model enhances the discriminative capability of the model to learn finer identity distinctions without compromising the model's robustness to variations within a class. Adding two negative samples to a quadruplet enhances the strength of the decision boundary, leading to improved generalization in real-world scenarios. The overall workflow of the proposed improved quadruplet loss computation is illustrated in Figure 3.

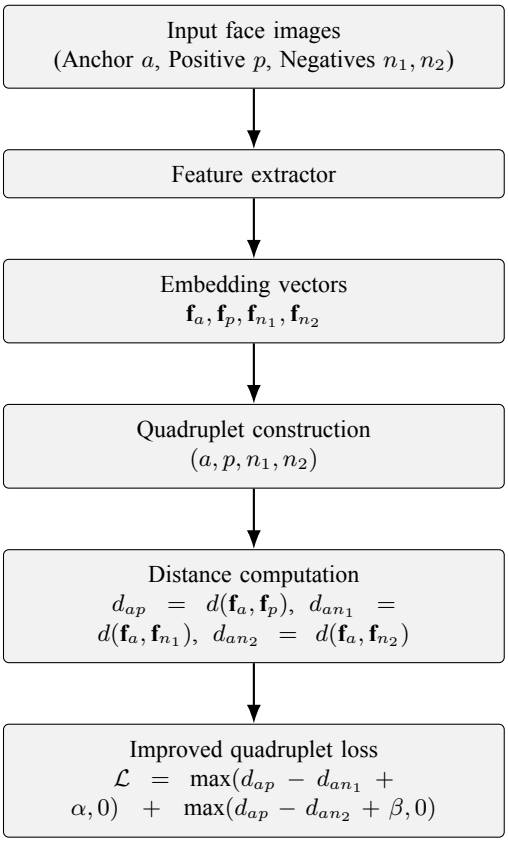

Figure 3: Flowchart of the improved quadruplet loss computation

3.3 The process of our method

The proposed model is trained from scratch, with randomly initialized weights. First, we prepared the dataset, which was divided into two subsets: 20% for testing and 80% for model training. Since our dataset was already cropped, no additional preprocessing was applied. However, before being input into our neural network, every picture was normalized to a consistent size of $220 \times 220 \times 3$. The model extracts 128-dimensional embeddings using a convolutional neural network (CNN) architecture.

The proposed feature extractor is a convolutional neural network that maps an input face image to a compact 128-dimensional embedding. The backbone is deliberately kept lightweight while still capturing multi-scale discriminative information, and it consists of a sequence of five convolutional blocks followed by two fully connected layers. In each convolutional block, a two-dimensional convolution with 3×3 kernels, stride 1×1, and *same* padding is applied, using He-normal weight initialization and ℓ_2 kernel regularization with a factor of 10^{-2}. All convolutional layers use the ReLU activation function. The number of filters is progressively increased across the blocks as 8, 16, 32, 64, and 128, respectively, in order to gradually enrich the representational capacity while controlling the overall model size. Each convolutional layer is followed by a 2×2 max-pooling layer to reduce spatial resolution, a batch-normalization layer to stabilize and accelerate training, and a dropout layer with a rate of 0.4 to mitigate overfitting. This sequence of operations yields a hierarchy of feature maps that encode increasingly abstract facial patterns under mask occlusion. After the final convolutional block, the resulting feature maps are flattened into a one-dimensional feature vector, which is then processed by a fully connected layer with 128 units, ReLU activation, He-normal initialization, and the same ℓ_2 regularization. This dense layer aggregates the spatially distributed convolutional responses into a more compact, high-level representation. A final dense layer with 128 units and linear activation is then applied to produce the embedding vector used by the metric-learning component. For clarity and reproducibility, the complete layer-by-layer specification of the backbone network, including filter sizes, padding, activation functions and regularization settings, is summarized in Table 2. A 40% dropout rate was found to be effective in reducing overfitting and improving generalization without notably harming performance. Figure.3 illustrates the architecture of the proposed model.

For classification, we tested Euclidean distance, Manhattan distance, and a neural network.

Euclidean distance: Calculate the distance between two points in Euclidean space. Mathematically, this normalization procedure is given in Equation 4 as follows:

$$\text{Distance}_E(p,m) = \left(\sum_{i=1}^{n} (p_i - m_i)^2 \right)^{1/2} \tag{4}$$

where p and m are two points in an n-dimensional space, where p_i and m_i represent the coordinates of p and m, respectively, for each dimension $i = 1, 2, \ldots, n$. The value of n defines the dimensionality of the space in which these points are located.

Manhattan distance is calculated by taking the sum of the absolute values between the coordinates p and m in the space n, as defined in Equation 5.

Table 2: CNN architecture with separate columns for kernel size, number of filters/units, padding, and parameters

Layer (type)	Output shape	Kernel / pool	Filters / units	Padding / rate	Param #
input_layer (InputLayer)	(None, 200, 200, 3)	–	–	–	0
conv2d (Conv2D)	(None, 200, 200, 8)	3×3	8 filters	padding = same	224
max_pooling2d (MaxPooling2D)	(None, 100, 100, 8)	2×2	–	padding = valid	0
batch_normalization (BatchNormalization)	(None, 100, 100, 8)	–	–	–	32
dropout (Dropout)	(None, 100, 100, 8)	–	–	rate = 0.4	0
conv2d_1 (Conv2D)	(None, 100, 100, 16)	3×3	16 filters	padding = same	1,168
max_pooling2d_1 (MaxPooling2D)	(None, 50, 50, 16)	2×2	–	padding = valid	0
batch_normalization_1 (BatchNormalization)	(None, 50, 50, 16)	–	–	–	64
dropout_1 (Dropout)	(None, 50, 50, 16)	–	–	rate = 0.4	0
conv2d_2 (Conv2D)	(None, 50, 50, 32)	3×3	32 filters	padding = same	4,640
max_pooling2d_2 (MaxPooling2D)	(None, 25, 25, 32)	2×2	–	padding = valid	0
batch_normalization_2 (BatchNormalization)	(None, 25, 25, 32)	–	–	–	128
dropout_2 (Dropout)	(None, 25, 25, 32)	–	–	rate = 0.4	0
conv2d_3 (Conv2D)	(None, 25, 25, 64)	3×3	64 filters	padding = same	18,496
max_pooling2d_3 (MaxPooling2D)	(None, 12, 12, 64)	2×2	–	padding = valid	0
batch_normalization_3 (BatchNormalization)	(None, 12, 12, 64)	–	–	–	256
dropout_3 (Dropout)	(None, 12, 12, 64)	–	–	rate = 0.4	0
conv2d_4 (Conv2D)	(None, 12, 12, 128)	3×3	128 filters	padding = same	73,856
max_pooling2d_4 (MaxPooling2D)	(None, 6, 6, 128)	2×2	–	padding = valid	0
batch_normalization_4 (BatchNormalization)	(None, 6, 6, 128)	–	–	–	512
dropout_4 (Dropout)	(None, 6, 6, 128)	–	–	rate = 0.4	0
flatten (Flatten)	(None, 4608)	–	–	–	0
dense (Dense)	(None, 128)	–	128 units	ReLU	589,952
dense_1 (Dense)	(None, 128)	–	128 units	linear	16,512

$$\text{Distance}_M(p, m) = \sum_{i=1}^{n} |p_i - m_i| \tag{5}$$

Neural Network : To evaluate the similarity between two images, we utilize a neural network comprising four fully connected layers. The network takes as input two 128-dimensional embeddings; we first compute their element-wise absolute difference and then concatenate $\mathbf{e}_a$, $\mathbf{e}_b$, and $|\mathbf{e}_a - \mathbf{e}_b|$, resulting in a 384-dimensional input vector. This vector is passed through a fully connected multilayer perceptron with three hidden layers of sizes 128, 32, and 8, respectively, each using a ReLU activation function, He-uniform weight initialization, and ℓ_2 kernel regularization with a factor of 10^{-2}. The final output layer is a single neuron with linear activation that produces the similarity score. No dropout is applied in this module. In total, the metric network comprises approximately 5.37×10^4 trainable parameters (53 681). The structure of the learned metric is illustrated in Fig. 5.

Instead of raw input, it receives the concatenation of two embeddings $A, B \in \mathbb{R}^n$ extracted from feature encoders:

$$X = \begin{bmatrix} A \\ B \end{bmatrix} \in \mathbb{R}^{2n}$$

Each hidden layer applies a linear transformation followed by ReLU activation: $h_1 = \sigma(W_1 X + b_1)$, $h_2 = \sigma(W_2 h_1 + b_2)$, $h_3 = \sigma(W_3 h_2 + b_3)$ where $W_i \in \mathbb{R}^{d \times d}$ are weight matrices, $b_i \in \mathbb{R}^d$ are biases, and $\sigma(x) = \max(0, x)$ is the ReLU function. The final similarity score is computed as:$D(A, B) = W_4 h_3 + b_4$ where $W_4 \in \mathbb{R}^{1 \times d}$ and $b_4 \in \mathbb{R}$.

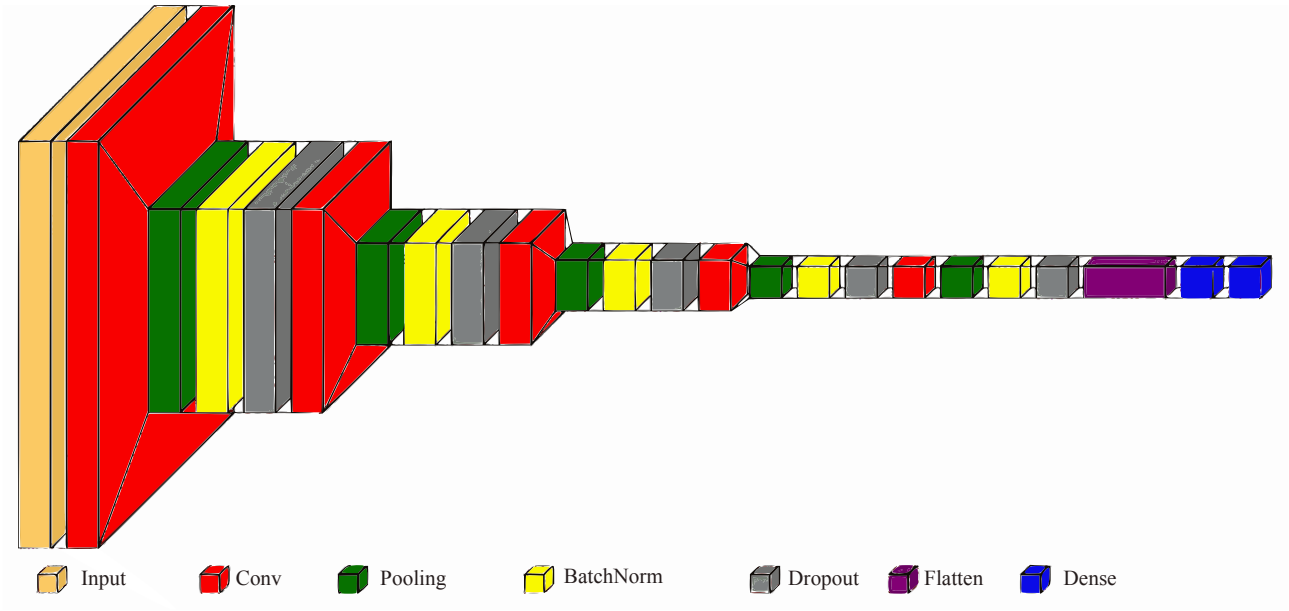

Figure 4: Architecture of the proposed model: the main layers and their connections

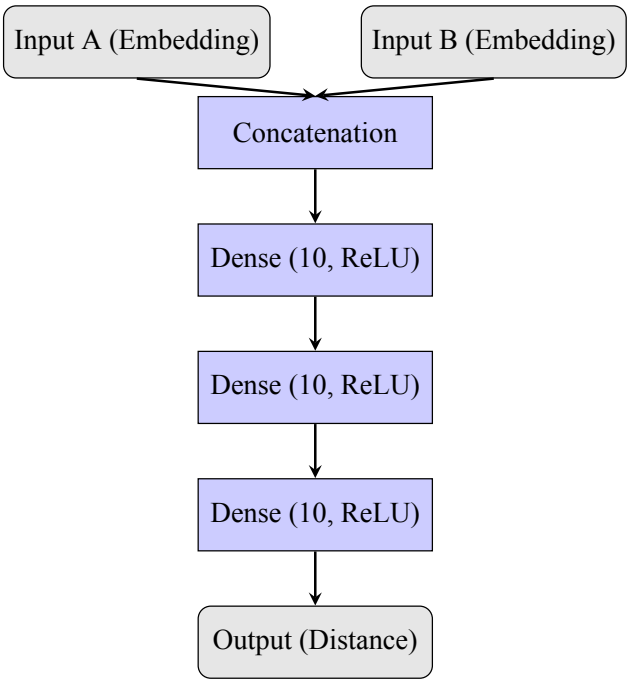

Figure 5: Neural network architecture for similarity measurement

The full expression is compactly written as in Equation 6.

$$D(A,B) = W_4\,\sigma\Big(W_3\,\sigma\big(W_2\,\sigma(W_1X + b_1) + b_2\big) + b_3\Big) + b_4 \quad (6)$$

This formulation enables the network to learn a continuous similarity metric between embeddings.

4 Experimental Results

To provide transparency regarding computational requirements, we report both the training efficiency and the inference cost of the proposed approach. All experiments were conducted in a Kaggle environment equipped with two NVIDIA T4 GPUs, and the end-to-end training time was approximately 4 hours, depending on the dataset scale. The trained model achieved a real-time inference speed of 379.3 face images per second (FPS) while requiring only 1.30 GB of GPU memory, as measured after GPU warm-up in forward-pass mode. In addition, the proposed network was designed to remain lightweight, containing 705,344 trainable parameters, which helps reduce memory usage and accelerate optimization compared to heavier face-recognition backbones. Together, these characteristics make the proposed method well suited for real-time masked face recognition under constrained computational budgets. Our model was trained on our dataset with input images resized to 200×200 before being processed by the network, using a batch size of 64. The margin parameters, α and β, were automatically optimized using Optuna to achieve optimal performance. Hyperparameter optimization using Optuna is performed as an offline tuning step and therefore does not impact inference-time complexity. In our implementation, the Optuna study consists of 10 trials, each trained for 5 epochs, resulting in a total tuning budget of 50 training epochs per dataset. This limited budget was chosen to efficiently explore the margin space while keeping the computational overhead moderate. Once the optimal margins (α, β) are selected, they are fixed and used for all subsequent experiments.

For the final training with the selected margins, the model typically converges within approximately 20–30 epochs, as indicated by stabilization of the validation loss

Table 3: Evaluation results of the proposed model using three similarity metrics reported as mean ± standard deviation

Distance	α	β	V_Acc (%)	V_Loss	V_Prec (%)	V_Rec (%)	V_F1 (%)
Euclidien	0.0771	0.2418	99.28 ± 0.01	0.294 ± 0.41	99.33 ± 0.03	99.30 ± 0.03	99.32 ± 0.03
Manhattan	0.0572	0.1705	98.58 ± 0.18	0.088 ± 0.06	98.69 ± 0.18	98.61 ± 0.18	98.65 ± 0.18
NN	0.6563	0.3047	96.84 ± 1.81	0.120 ± 0.06	98.62 ± 0.26	98.34 ± 0.62	98.48 ± 0.43

and recognition metrics. Training beyond this point yields marginal performance gains. These observations confirm that the additional cost introduced by Optuna remains reasonable relative to the overall training process, while providing consistent performance improvements. To assess the similarity between the extracted image encodings, we employed three different approaches: Euclidean distance, Manhattan distance, and a neural network trained for classification. The evaluation results on the test dataset are presented in Table 3. The reported metrics include accuracy (Acc), loss (Loss), precision (Prec), recall (Rec), and F1-score (F1), expressed as mean ± standard deviation. The Euclidean distance metric achieves the best overall performance, with an accuracy of 99.28 ± 0.01%, precision of 99.33 ± 0.03%, recall of 99.30 ± 0.03%, and F1-score of 99.32 ± 0.03%, while maintaining a low loss value. The small standard deviations indicate highly consistent performance. The Manhattan distance shows slightly lower performance, with an accuracy of 98.58 ± 0.18% and F1-score of 98.65 ± 0.18%, while preserving competitive precision and recall values. The NN-based similarity metric presents comparable precision and recall; however, its validation accuracy (96.84 ± 1.81%) exhibits higher variability compared to the distance-based approaches. Overall, the Euclidean distance metric demonstrates superior and more stable performance among the evaluated similarity measures. To rigorously assess the performance of our proposed model, we conducted evaluations on multiple benchmark datasets specifically designed for masked face recognition. COMASK20 [23], CASIA-WebFaceMasked [24], LFW-SMFD [25], and MFRD-80K [26] are some of these datasets. The proposed model is trained independently on each dataset, using the same architecture and hyper-parameters but with randomly initialized weights (He initialization). For every dataset, we train the model from scratch on its training split and evaluate it on the corresponding test split. For all experiments, the proposed model was trained for 100 epochs using the Adam optimizer with a fixed learning rate of 0.001 and gradient clipping set to 1.0, without any additional learning rate scheduling or decay. For each dataset considered in this study, the available images were randomly partitioned into 80% for training and 20% for testing, and this 80/20 split was kept fixed across all experiments so that the different loss configurations and ablation settings were evaluated under identical data conditions. Due to the variety of these datasets, our method is tested in a range of scenarios that encompass different occlusion levels, image qualities, and subject variations. The datasets selected for this research provide a strong foundation for evaluating masked face recognition under diverse conditions. They differ in terms of identities, image counts, image resolution, mask type (real or synthetic), lighting, and pose — all of which help assess the robustness of our model. Together, these datasets provide a diverse evaluation framework, helping ensure our model is not overfit to a specific scenario. Testing on a mix of real and synthetic masks, as well as small and large datasets, reveals both the strengths and limitations of our approach. The performance evaluation results in Table 4 further validate that our deep metric learning model, with an improved quadruplet loss, performs consistently across all datasets, demonstrating its adaptability and effectiveness in real-world masked face recognition scenarios.

From Table 4, the Euclidean distance consistently outperforms the Manhattan and NN-based approaches across all datasets, demonstrating its effectiveness in representing face features. The highest accuracy is observed with the ENSA-MFRD dataset (99.27%), followed by COMASK20 and MFRD. The NN-based distance shows notably lower performance, particularly on CASIA and MFRD, suggesting that it may require further optimization or additional training data for improved generalization. The loss values follow an inverse trend to accuracy, with lower losses corresponding to higher accuracy. Euclidean distance yields the lowest loss across datasets, confirming its stability. Manhattan distance performs slightly worse than Euclidean but better than the NN-based approach, suggesting it may be viable in certain scenarios. An important part of the evaluation is the impact of the margins α and β, which are optimized per dataset using Optuna. These hyperparameters determine the space between positive and negative pairs and determine how the model will learn. Lower values of α and β are generally achieved with higher accuracy (e.g., ENSA-MFRD with Euclidean: α = 0.0771, β = 0.2418, Acc = 99.27%). This demonstrates the discriminative ability of the embedding space. On the other hand, larger separations (e.g., CASIA and Manhattan: α = 0.3181, β = 0.2744) give clear performance deterioration, and poor separation is expected. This suggests the necessity of dataset-oriented tuning, as choosing a poor margin could lead to a decline in model performance. Overall, these results validate the effectiveness of Euclidean distance for face recognition tasks, highlight the importance of margin optimization using Optuna, and suggest potential areas for improvement in the neural network-based metric to enhance its generalization capabilities.

Table 4: Performance metrics across different datasets (MFRD-80K, CASIA, LFW-SMFD, COMASK20, ENSA-MFRD) (V_Acc, V_Prec, V_Rec and V_F1 score in %)

Distance	Dataset	α	β	V_Acc	V_Loss	V_Prec	V_Rec	V_F1
Euclidien	MFRD-80K	0.1990	0.1570	97.93	0.1122	98.41	98.29	98.36
	CASIA	0.4181	0.2144	73.19	2.9196	83.79	80.88	82.30
	LFW-SMFD	0.1151	0.6757	94.20	0.3211	94.50	94.28	94.38
	COMASK20	0.4025	0.5754	97.84	0.3299	98.55	98.35	98.46
	ENSA-MFRD	0.0771	0.2418	**99.27**	**0.0593**	**99.29**	**99.26**	**99.28**
Manhattan	MFRD-80K	0.1511	0.2554	87.21	1.2850	90.89	90.14	90.51
	CASIA	0.3181	0.2744	63.29	3.0196	73.69	70.78	72.50
	LFW-SMFD	0.1917	0.6531	85.92	2.0699	88.60	88.02	88.31
	COMASK20	0.6470	0.9196	96.91	1.3096	97.07	96.70	96.88
	ENSA-MFRD	0.05721	0.1704	98.54	0.0767	98.70	98.62	98.66
NN	MFRD-80K	0.1641	0.3785	49.95	4.7912	50.04	33.40	40.06
	CASIA	0.0184	0.0532	46.59	4.4553	49.28	32.54	39.14
	LFW-SMFD	0.2057	0.8826	84.67	2.0729	87.03	86.69	86.86
	COMASK20	0.0767	0.1363	82.98	0.2431	87.84	86.68	87.26
	ENSA-MFRD	0.6563	0.3047	98.54	0.0767	98.70	98.62	98.66

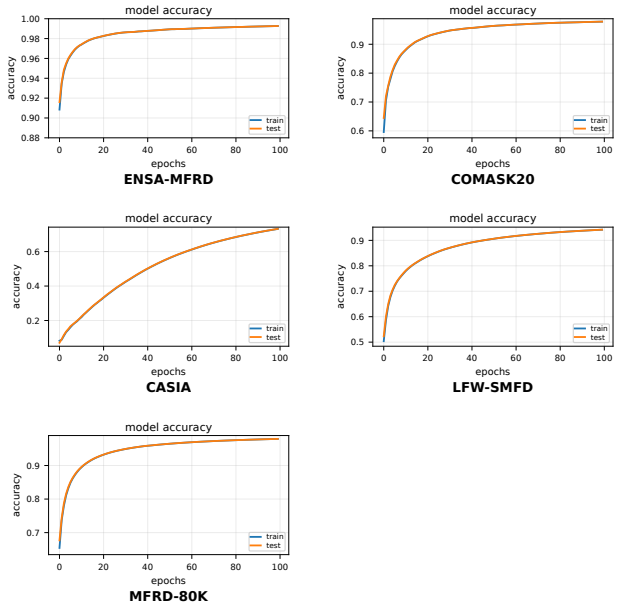

Figure 6: Accuracy convergence comparison across five datasets

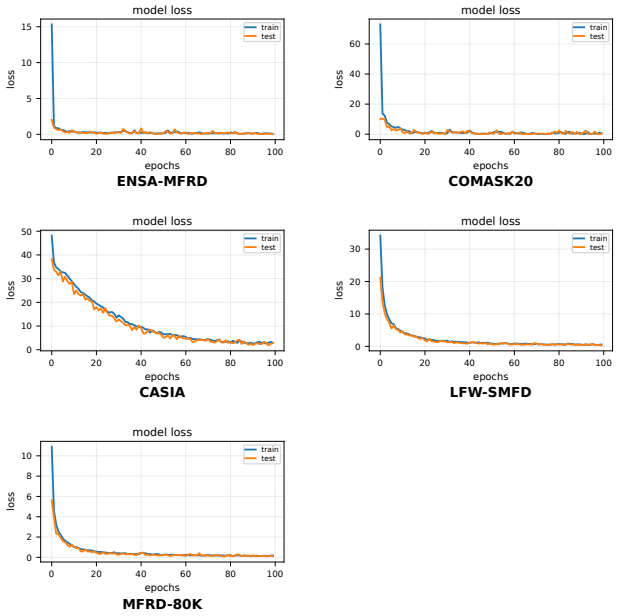

Figure 7: Loss convergence comparison across five datasets

The convergence of the model is assessed through the acuuracy and loss curves in Figure. 6 and 7 respectively, which demonstrate the training stability and optimization progress across five benchmark datasets: ENSA-MFRD, COMASK20, CASIA, LFW-SMFD, and MFRD-80K. A smooth, continuously declining loss curve indicates successful feature learning and minimal overfitting. All the subfigures show loss evolution with training in a clear downward trend, indicating successful minimization of the objective function. Convergence rate differences mirror differences in dataset size, intricacy, and intra-class diversity.

Additionally, datasets with higher intra-class variability, such as LFW-SMFD and MFRD-80K, exhibit a relatively slower convergence rate compared to structured datasets like CASIA and ENSA-MFRD. This observation aligns with the expectation that complex datasets require more iterations to effectively generalize discriminative features. The comparison highlights the robustness of our approach in adapting to different data distributions. Table 5 presents the evaluation of different distance metrics—Euclidean, Manhattan, and NN—across multiple datasets, while maintaining a fixed α and β ($\alpha = 0.1371$, $\beta = 1.9010$)

Table 5: Performance metrics across different datasets with fixed α and β parameters (MFRD-80K, CASIA, LFW-SMFD, COMASK20, ENSA-MFRD) (V_Acc, V_Prec, V_Rec and V_F1 score in %)

Distance	Dataset	α	β	V_Acc	V_Loss	V_Prec	V_Rec	V_F1
Euclidean	MFRD-80K	0.1371	1.9010	95.83	1.6137	96.18	95.85	96.01
	CASIA	0.1371	1.9010	52.22	18.122	70.02	62.92	66.28
	LFW-SMFD	0.1371	1.9010	88.3	1.5271	87.7	87.05	87.37
	COMASK20	0.1371	1.9010	96.49	1.6096	97.06	96.69	96.87
	ENSA-MFRD	0.1371	1.9010	**98.6**	**0.4025**	**98.57**	**98.47**	**98.52**
Manhattan	MFRD-80K	0.1371	1.9010	85.67	7.6483	88.7	87.17	87.93
	CASIA	0.1371	1.9010	40.78	33.409	64.85	55.11	59.58
	LFW-SMFD	0.1371	1.9010	84.81	2.4877	84.75	83.71	84.23
	COMASK20	0.1371	1.9010	94.28	0.7191	95.00	94.20	94.60
	ENSA-MFRD	0.1371	1.9010	97.27	1.6467	97.49	97.25	97.37
NN	MFRD-80K	0.1371	1.9010	50.46	65.192	50.36	33.85	40.49
	CASIA	0.1371	1.9010	16.17	63.266	50.17	33.80	40.39
	LFW-SMFD	0.1371	1.9010	50.18	64.747	50.33	33.75	40.41
	COMASK20	0.1371	1.9010	31.76	64.054	45.48	27.20	34.04
	ENSA-MFRD	0.1371	1.9010	27.44	65.078	48.1	29.97	36.93

Table 6: Evaluation results of different loss functions using various distance metrics (V_Acc, V_Prec, V_Rec and V_F1 score in %)

Method	Distance	α	β	V_Acc	V_Loss	V_Prec	V_Rec	V_F1
L_{Tiplet}	Euclidean	0.7990	-	98.65	0.1827	98.66	98.66	98.65
	Manhattan	1.4181	-	97.61	2.9196	97.70	97.70	97.61
	NN	2.1151	-	61.04	3.3211	61.05	61.10	61.05
$L_{Quadruplet}$	Euclidean	0.1771	0.2418	98.55	0.1993	98.60	98.65	98.55
	Manhattan	0.2572	0.1705	98.39	0.1467	98.39	98.40	98.41
	NN	1.6563	0.3047	86.67	0.5767	86.68	86.67	86.67
L_{ImQuad}	Euclidean	0.0771	0.2418	99.27	0.0593	99.29	99.26	99.28
	Manhattan	0.0572	0.1705	98.54	0.0767	98.70	98.62	98.66
	NN	0.6563	0.3047	98.54	0.0767	98.70	98.62	98.66

for all datasets. The results demonstrate the impact of using a uniform margin on overall performance metrics for both training and validation phases. These results emphasize the critical role of α and β hyperparameters in determining the model's ability to separate positive and negative pairs effectively. While a fixed margin may simplify model tuning, it results in suboptimal performance on datasets with varying distributions. This suggests that dataset-specific fine-tuning of α and β is necessary to achieve optimal accuracy and generalization. The performance gap observed between ENSA-MFRD and CASIA can be attributed primarily to dataset-specific acquisition and occlusion characteristics rather than to model instability. In ENSA-MFRD, mask types and coverage are relatively controlled and consistent, whereas CASIA includes a broader diversity of mask shapes, materials, and wearing styles (e.g., loose or partially covering masks), which increases intra-class variability and reduces feature consistency. Moreover, ENSA-MFRD was collected under semi-controlled conditions with limited pose variation, while CASIA exhibits larger head rotations and non-frontal views; since the proposed approach relies strongly on discriminative cues from the upper facial region, strong pose changes degrade the stability of these cues. Finally, CASIA images often present lower resolution and higher compression artifacts, which hinder the extraction of fine-grained periocular and forehead features that are critical for masked face recognition.

We evaluated our proposed loss function against two established alternatives, the triplet loss (Schroff et al. [3]) and the quadruplet loss (Chen et al. [6]), to assess its effectiveness. The triplet loss seeks to minimize intra-class distance via margin-based separation between similar and dissimilar samples, while the quadruplet loss enhances inter-class discrimination by introducing an additional negative sample. We evaluated all three loss functions on our ENSA-MFRD dataset, using Euclidean, Manhattan, and an NN distance. The experimental results are summarized in Table 6.

The outcomes in Table 6 demonstrate the efficiency of all loss functions across various distance measurement metrics. The Improved Quadruplet Loss demonstrates a significantly better performance than both Triplet Loss and the

standard Quadruplet Loss, mainly when the Euclidean distance measure is employed. It achieves a validation accuracy of 99.27%, precision of 99.29%, recall of 99.26%, and F1-score of 99.28%, all of which are accompanied by a very low validation loss of only 0.0593. This confirms its strong ability to optimize feature embedding space for enhanced face verification. In contrast, the Triplet Loss, especially with the NN distance, shows significantly lower performance, with a validation accuracy of only 61.04%, indicating its limitations in handling complex variations in masked face recognition. The quadruplet loss offers an improvement over the triplet loss by introducing additional constraints. Yet, it does not achieve the robustness of the improved quadruplet loss, particularly when Manhattan and NN distances are used. These findings emphasize the efficiency of improved quadruplet loss, making it a promising approach for real-world masked face recognition applications.

5 Discussion

5.1 Quantitative comparison with the state of the art

To clearly position our approach with respect to existing work, we compare the best configurations of our model with the most recent masked face recognition methods presented in Table 1. FaceMaskNet-21 achieves an accuracy of 92.8%, PLFace reaches 96.4%, while ArcFace-based variants (CosFace, SubFace, etc.) generally lie between 97% and 98% on datasets such as RMFRD, SMFD, or their derivatives. In comparison, our model using the improved quadruplet loss achieves: 99.27% on our internal dataset (ENSA-MFRD), 97.93% on MFRD-80K, 73.19% on CASIA, 94.20% on LFW-SMFD, and 97.84% on COMASK20, which consistently outperforms the models from the state of the art reported in the literature, including deep pre-trained networks such as ResNet-100, VGGFace, or HRNet.

Unlike these massive approaches, our method relies on a more compact architecture and a discriminative loss function, which minimizes complexity while improving robustness to occlusions caused by face masks. The results reported in Tables 3, 4, and 5 clearly show that the choice of distance metric strongly influences the ability of the model to effectively separate facial representations under occlusion.

The main observations are as follows:

- The Euclidean distance provides the most stable and best results in terms of accuracy, F1-score, and loss, confirming that it remains the most suitable metric for high-dimensional normalized embeddings.
- The Manhattan distance yields slightly lower performance, suggesting a weaker suitability to the geometric structure of the learned encodings.
- The distance learned via a neural network (NN) exhibits competitive recall but higher variance, which reflects increased sensitivity to data fluctuations and a higher risk of overfitting.

These results confirm that the combination of Euclidean distance and the improved quadruplet loss represents the most robust trade-off for masked face recognition across different scenarios.

5.2 Impact of distance metric under different occlusion levels

In order to assess the impact of various distance metrics under a variety of occlusion conditions, we implemented ablation experiments. The following similarity measures were evaluated: Manhattan distance, Euclidean distance, and the distance module based on neural networks. The test images were subjected to three artificial levels of occlusion: low (25%), medium (40%), and high (70%) mask coverage.

The recognition results derived on the LFW dataset are presented in Table 7. These results are based on low, medium, and high simulated occlusion levels. The negative impact of facial obstruction on identity recognition is confirmed by the progressive decrease in performance that is observed as the severity of occlusion increases. Across all occlusion levels, the Euclidean distance consistently obtains the highest accuracy and F1-score among the evaluated similarity metrics, suggesting more robustness and stability. Although Manhattan and NN-based methods offer competitive precision values, their accuracy and recall deteriorate more significantly in the presence of medium and high occlusion situations. The growing classification difficulty is further reflected in the increase in loss values as larger facial regions become concealed. In conclusion, the findings indicate that the recognition accuracy is inversely proportional to the severity of occlusion and that the selection of a similarity metric is crucial for performance in challenging environments.

5.3 Comparison of loss functions

The analysis presented in Table 5 clearly demonstrates that the proposed loss function, $\mathcal{L}_{\text{imQuad}}$, outperforms classical approaches, including the triplet loss and the original quadruplet loss. Our enhanced formulation encourages more discriminative learning by structuring the latent space in a more coherent manner, in contrast to these traditional loss functions, which do not always maintain an optimal inter-class separation under partial occlusion. This capability results in quantifiable improvements in the areas of precision, recall, accuracy, and F1-score. The enhancements, which vary from 1% to 3% depending on the dataset, illustrate that $\mathcal{L}_{\text{imQuad}}$ is notably effective in enhancing the margin between distinct classes and reinforcing intra-class compactness, which are two critical components of robust masked face recognition. Another substantial benefit of $\mathcal{L}_{\text{imQuad}}$ is its capacity to minimize the optimization error,

Table 7: Comparative analysis of distance metrics under low, medium, and high simulated occlusion levels on the LFW dataset.

Occ. Level	Metric	V_Acc (%)	V_Loss	V_Prec (%)	V_Rec (%)	V_F1 (%)
Low	Euclidean	95.02	0.1609	96.20	95.81	96.00
	Manhattan	89.42	1.7191	92.86	91.53	92.19
	NN	88.25	0.6421	91.40	90.28	90.84
Medium	Euclidean	95.58	0.1077	96.39	96.04	96.21
	Manhattan	86.78	0.2485	87.85	86.80	87.32
	NN	85.12	0.3927	88.96	87.92	88.44
High	Euclidean	94.77	0.0809	96.27	95.79	96.03
	Manhattan	90.28	1.1395	92.88	91.56	92.21
	NN	85.94	0.5236	94.37	93.37	93.87

as evidenced by the reduction in V_{Loss} in comparison to the other evaluated losses. This decrease implies a more rapid and consistent convergence, which is indicative of a more accurate modeling of the relationships between positive and negative samples during the training process. In other words, the loss function not only improves the separation between different individuals, but also optimizes the cohesion of samples belonga mask partially occludes themhey are partially occluded by a mask. This property gives our model enhanced resilience to the visual perturbations introduced by face masks, which some competing methods based on fixed margins or less adaptive penalties fail to provide.

Overall, the comparison of loss functions confirms that $\mathcal{L}_{\text{imQuad}}$ is better suited to the specific challenges of masked face recognition. Its formulation encourages a more discriminative representation, ensures better numerical stability, and, most importantly, improves the overall robustness of the system under various occlusion conditions. Thus, $\mathcal{L}_{\text{imQuad}}$ enables superior performance without requiring a more complex architecture, consolidating its central role in improving the proposed model.

5.4 Parallel between the adaptivity of the improved quadruplet loss and adaptive control strategies

Although our work lies in the field of facial recognition and not in control theory, our approach shares several fundamental principles with adaptive control techniques used to handle nonlinear, uncertain, or perturbed systems. In our model, the improvement of the quadruplet loss enables an implicit adaptation to variations induced by masks, partial occlusions, illumination changes, or noise present in the images. This adaptive effect is manifested through the dynamic reorganization of intra-class and inter-class distances in the latent space. the ability of the model to strengthen or relax the separation between classes depending on the observed level of occlusion, and a progressive reduction of sensitivity to perturbations that are not explicitly visible in the training data, guided by the structure of the loss. Conceptually, these mechanisms are related to the behavior of adaptive control systems, where a controller continuously adjusts its parameters to maintain stability and performance in the presence of environmental uncertainties.

Similarly, our improved quadruplet loss implicitly adjusts the structure of the representation space to reduce the impact of occlusions and maximize separability, thereby ensuring reliable recognition despite the uncertainty introduced by masks. To further strengthen this conceptual link between adaptive face recognition and adaptive control, we draw inspiration from the robustness and disturbance-compensation properties highlighted in the following works: [27] [28] [29] [30] [31] [32]. These studies demonstrate how adaptive mechanisms allow a system to maintain its performance despite variations in input conditions, a principle that parallels the adaptivity of our model to masks and occlusions.

5.5 Real-world applications and deployment considerations

The proposed method is particularly well-suited for real-world security and biometric applications, where facial occlusion is a common occurrence. The robustness of the proposed framework to masked or partially visible faces can directly benefit access control systems, intelligent surveillance platforms, and automated identification solutions. The method is also applicable to industrial and robotic environments, where the reliable identification of operators donning protective equipment (e.g., masks, helmets, or visors) is crucial for ensuring operational continuity and safety. The proposed framework demonstrates behavior that is comparable to adaptive control systems utilized in industrial and autonomous platforms from a conceptual perspective. In these systems, the controller parameters are perpetually adjusted to ensure stability in the presence of external disturbances, noise, or uncertainties. In the same vein, the enhanced quadruplet loss facilitates a structured reorganization of the embedding space in response to

fluctuations in occlusion patterns or acquisition conditions. Stable performance under genuine operational constraints is facilitated by this adaptive representation learning mechanism.

Several technical considerations must be addressed in order to deploy in practical environments. Despite the computational demands of quadruplet generation and model complexity during training, inference is still lightweight and compatible with embedded or edge-based systems. Standard optimization techniques, including pruning, quantization, and knowledge distillation, can be implemented to satisfy real-time performance requirements. Adaptive thresholding or auto-calibration strategies may be necessary to maintain reliable decision boundaries in scenarios with severe occlusion, extreme pose variations, or poor image quality.

5.6 Limitations and future work

Despite its strong performance across the evaluated datasets, the proposed approach presents several limitations. First, it is important to consider the demographic characteristics of the ENSA dataset, which mainly consists of university students. This relatively homogeneous age distribution may limit the assessment of the model's behavior across broader populations, particularly with respect to age- and gender-related variations in facial morphology. Consequently, further evaluation on more diverse and heterogeneous datasets is necessary to thoroughly assess the generalization capability and fairness of the proposed method.

In addition, recognition performance decreases under extreme occlusion levels, suggesting that improved exploitation of the remaining visible facial regions is still required. Although the neural network–based learned distance is expressive, additional regularization strategies may help reduce sensitivity to noise and cross-dataset variability. The hyperparameter optimization process involving the α and β coefficients proved effective; however, more efficient or guided search strategies could further reduce the associated computational overhead. Furthermore, the generation of quadruplets and the dimensionality of the latent representation introduce computational complexity, indicating opportunities for architectural simplification and execution-time optimization.

Future work will therefore focus on several directions. First, the integration of attention mechanisms targeting unoccluded facial regions could further improve recognition robustness under severe masking conditions. Second, optimizing the model architecture and inference pipeline may facilitate real-time deployment in practical applications such as intelligent surveillance or access control systems. Finally, extending the framework toward multimodal face recognition by combining visible and thermal imagery could enhance robustness in challenging acquisition environments and improve system reliability under varying illumination and environmental conditions.

6 Conclusion

This paper presents a deep learning model for masked face identification, tackling the issues of obscured facial characteristics. Our project was divided into two primary phases: the creation of a specific masked face dataset and the development of a recognition model using a specialized loss function to enhance performance. The system utilizes a convolutional neural network to extract facial features, subsequently measuring similarity by Euclidean distance, Manhattan distance, and an alternative CNN-based method. Experimental assessments confirmed the model's efficacy, with a maximum precision of 99.27% with the application of Euclidean distance. These findings highlight the potential of our approach for robust masked face recognition.

Code availability

The implementation of the proposed method is publicly available at: https://github.com/sihamahmam/Improved-Quadruplet..

References

[1] Organization WH et al. (2020). Advice on the use of masks in the context of covid-19: interim guidance.*5 June 2020. Tech. rep.,World Health Organization.*

[2] Ngan M, Grother P, Hanaoka K (2020). Ongoing Face Recognition Vendor Test (FRVT) Part 6B: Face recognition accuracy with face masks using post-COVID-19 algorithms, *(NISTIR), Gaithersburg, MD.* https://doi.org/10.6028/NIST.IR.8331

[3] Schroff F, Kalenichenko D, Philbin J (2015). Facenet: A unified embedding for face recognition and clustering. *Proc. CVPR. (pp. 815–823).* https://doi.org/10.1109/CVPR.2015.7298682

[4] Goel R, Mehmood I, Ugail H (2021). A Study of Deep Learning-Based Face Recognition Models for Sibling Identification. *Sensors 21, 5068.* https://doi.org/10.3390/s21155068

[5] Bromley J, Guyon I, LeCun Y, Sackinger E, Shah R (1993). Signature Verification Using a Siamese Time Delay Neural Network *Morgan Kaufmann Publishers Inc. San Francisco, CA, USA.* https://doi.org/10.1142/S0218001493000339

[6] Chen W, Chen X, Zhang J, Huang K (2017) Beyond triplet loss: a deep quadruplet network for person re-identification. *In: Proc. IEEE CVPR (pp. 403-412).* https://doi.org/10.1109/CVPR.2017.145

[7] Boutros F, Damer N, Kirchbuchner F, Kuijper A (2021) Self-restrained triplet loss for accurate masked

face recognition. *Comput. Vis. Pattern Recognit.* https://doi.org/10.1016/j.patcog.2021.108473

[8] Huang B, Wang Z, Wang G, Jiang K, Han Z, Lu T, Liang C (2023) PLFace: Progressive learning for face recognition with mask bias. *Pattern Recognition, 135, 109142.* https://doi.org/10.1016/j.patcog.2022.109142

[9] Salim RJ, Surantha N (2023) Masked face recognition by zeroing the masked region without model retraining. *Int. J. Innov. Comput. Inf. Control, 19(4), 1087-1101.* https://doi.org/10.24507/ijicic.19.04.1087

[10] Golwalkar R, Mehendale N (2022) Masked-face recognition using deep metric learning and FaceMaskNet-21. *Appl Intell 52, 13268-13279.* https://doi.org/10.1007/s10489-021-03150-3

[11] Sikha OK, Bharath B (2022) VGG16-random Fourier hybrid model for masked face recognition. *Soft Comput 26, 12795–12810.* https://doi.org/10.1007/s00500-022-07289-0

[12] Du H, Shi H, Liu Y, Zeng D, Mei T (2021) Towards NIR-VIS masked face recognition. *IEEE Signal Process. Lett., 28, 768-772.* https://doi.org/10.1109/LSP.2021.3076335

[13] Omar M, Rashedul M, Touhid M (2024) Advanced Masked Face Recognition using Robust and Light Weight Deep Learning Model. *In IJCA (Vol. 186, No. 2, pp. 42-51).* https://doi.org/10.5120/ijca2024923351

[14] Zhang J, An D, Zhang Y, Wang X, Wang X, Wang Q, Pan Z, Yue Y (2025) A Review on Face Mask Recognition. *Sensors, 25(2), 387.* https://doi.org/10.3390/s25020387

[15] Mahmoud M, Kasem MS, Kang HS (2024) A Comprehensive Survey of Masked Faces: Recognition, Detection, and Unmasking. *Applied Sciences, 14(19), 8781.* https://doi.org/10.3390/app14198781

[16] Huang YC, Rahardjo DAB, Shiue RH, Chen HH (2024) Masked face recognition using domain adaptation. *Pattern Recognition, 153, 110574.* https://doi.org/10.1016/j.patcog.2024.110574

[17] Ge Y, Liu H, Du J, Li Z, Wei Y (2023) Masked face recognition with a convolutional visual self-attention network. *Neurocomputing, 518, 496-506.* https://doi.org/10.1016/j.neucom.2022.10.025

[18] Alqaralleh E, Afaneh A, Toygar Ö (2023) Masked face recognition using frontal and profile faces with multiple fusion levels. *Signal Image Video Process, 17(4), 1375-1382.* https://doi.org/10.1007/s11760-022-02345-6

[19] Cabani A, Hammoudi K, Benhabiles H, Melkemi M (2021) MaskedFace-Net–A dataset of correctly/incorrectly masked face images in the context of COVID-19. *Smart Health, 19, 100144.* https://doi.org/10.1016/j.smhl.2020.100144

[20] Zhang H, Tang J, Wu P, Li H, Zeng N (2023) A novel attention-based enhancement framework for face mask detection in complicated scenarios. *Signal Process. Image Commun., 116, 116985.* https://doi.org/10.1016/j.image.2023.116985

[21] Oulad-Kaddour M, Haddadou H, Palacios-Alonso D, Conde C, Cabello E (2024) Facial mask-wearing prediction and adaptive gender classification using convolutional neural networks. *EAI Endorsed Transactions on Industrial Networks and Intelligent Systems 11(2).* http://dx.doi.org/10.4108/eetinis.v11i2.4318

[22] Aly M (2025) Revolutionizing online education: Advanced facial expression recognition for real-time student progress tracking via deep learning model. *Multimedia Tools and Applications, 84, 12575–12614.* http://doi.org/10.1007/s11042-024-19392-5

[23] Vu HN, Nguyen MH, Pham C (2022) Masked face recognition with convolutional neural networks and local binary patterns. *Applied Intelligence 52, 5497–5512.* http://doi.org/10.1007/s10489-021-02728-1

[24] T, A., V, M. Explainable masked face recognition. Multimed Tools Appl 83, 31123–31138 (2024). https://doi.org/10.1007/s11042-023-16571-8

[25] Chong, W.-J. L., Chong, S.-C., Ong, T.-S. (2023). Masked face recognition using histogram-based recurrent neural network. J. Imaging, 9(2), 38. https://doi.org/10.3390/jimaging9020038

[26] Lee CP, Lim KM (2021) Mfrd-80k: A dataset and benchmark for masked face recognition. *Engineering Letters, 29 (4).*

[27] A. Boulkroune, F. Zouari, and A. Boubellouta, "Adaptive fuzzy control for practical fixed-time synchronization of fractional-order chaotic systems," *Journal of Vibration and Control*, 2025, doi:10.1177/10775463251320258.

[28] A. Boulkroune, F. Zouari, and A. Ibeas, "Output-feedback controller based projective lag-synchronization of uncertain chaotic systems in the presence of input nonlinearities," *Mathematical Problems in Engineering*, vol. 2017, Article ID 8045803, 12 pp., 2017, doi:10.1155/2017/8045803.

[29] F. Zouari, K. Ben Saad, and M. Benrejeb, "Robust neural adaptive control for a class of uncertain nonlinear complex dynamical multivariable systems," *International Review on Modelling and Simulations*, vol. 5, no. 5, pp. 2075–2103, 2012.

[30] F. Zouari, K. Ben Saad, and M. Benrejeb, "Adaptive backstepping control for a class of uncertain single input single output nonlinear systems," in *Proceedings of the 10th International Multi-Conference on Systems, Signals & Devices (SSD)*, 2013, pp. 1–6, doi:10.1109/SSD.2013.6564134.

[31] G. Rigatos, M. Abbaszadeh, B. Sari, P. Siano, G. Cuccurullo, and F. Zouari, "Nonlinear optimal control for a gas compressor driven by an induction motor," *Results in Control and Optimization*, vol. 11, p. 100226, 2023, doi:10.1016/j.rico.2023.100226.

[32] F. Zouari, K. Ben Saad, and M. Benrejeb, "Adaptive backstepping control for a single-link flexible robot manipulator driven DC motor," in *Proceedings of the 2013 International Conference on Control, Decision and Information Technologies (CoDIT)*, Hammamet, Tunisia, 2013, pp. 864–871, doi:10.1109/CODIT.2013.6689656.