

## Review

# FuzzyArcLoss: Dynamic margin adjustment for robust recognition across domains

Servio F. Lima<sup>a</sup> ,<sup>\*</sup> Edy Portmann<sup>a</sup> , Luis Terán<sup>b</sup>

<sup>a</sup> Human-IST Institute, University of Fribourg, Boulevard de Pérolles 90, 1700 Fribourg, Switzerland

<sup>b</sup> Business School, Lucerne University of Applied Sciences and Arts, Zentralstrasse 9, 6002 Lucerne, Switzerland

## ARTICLE INFO

## Keywords:

FuzzyArcLoss  
Fuzzy logic  
Adaptive margin loss  
Deep learning  
Face recognition  
Extreme pose variations  
Occlusions  
Noise

## ABSTRACT

Recognition systems must face formidable challenges, including extreme pose variations, occlusions, noise, and nuanced facial expressions. Existing fixed-margin loss functions (e.g., *ArcFace*) and certain dynamic-margin approaches (e.g., *AdaptiveFace*) often exhibit performance limitations under such conditions. To address these gaps, we propose *FuzzyArcLoss*, a novel loss function that leverages a fuzzy membership mechanism to dynamically adjust angular margins for enhanced adaptability and robust performance.

Extensive experiments on four benchmarks (CPLFW, CALFW, JAFFE, CFP) confirm that *FuzzyArcLoss* consistently outperforms both fixed-margin and existing dynamic-margin methods (e.g., *AdaptiveFace*, *VPL*, *SphereFace2*, *UniFace*). In CPLFW and CALFW, *FuzzyArcLoss* achieves top-tier F1 scores (up to 0.90303 and 0.9079, respectively) along with elevated recall, balancing precision and recall more effectively than competing algorithms. On CFP, characterized by pronounced frontal-profile differences, *FuzzyArcLoss* ( $\tau = 0.9$ ) demonstrates consistently higher recall under severe occlusions and compression artifacts compared to other loss functions.

Although *UniFace* reaches the highest F1 score on JAFFE (0.8528), *FuzzyArcLoss* leads in recall (0.9475), underscoring its ability to detect challenging cases involving extreme expressions, albeit with a slight trade-off in precision. Across all datasets and augmentations — ranging from heavy compression to extensive occlusions — *FuzzyArcLoss* exhibits remarkable robustness, highlighting the importance of sample-level margin adjustments for addressing complex intra-class variability and ambiguous scenarios. Consequently, *FuzzyArcLoss* emerges as a robust and highly adaptable solution for face recognition and related recognition tasks, paving the way for improved handling of real-world conditions where static or purely class-based margins fall short.

## Contents

1.	Introduction .....	2
2.	Background and related work .....	2
3.	FuzzyArcLoss framework .....	3
3.1.	Mathematical formulation .....	3
3.1.1.	From fixed margins to adaptive margins .....	3
3.1.2.	Fuzzy membership mechanism .....	3
3.1.3.	FuzzyArcLoss formulation .....	3
3.2.	Algorithm implementation .....	4
3.3.	Explanation of key concepts .....	4
3.4.	Model setup and training .....	5
4.	Integration with augmented datasets .....	5
4.1.	Augmentation techniques .....	5
4.2.	Formula for total comparisons .....	5
4.3.	Dataset application and results .....	5

<sup>\*</sup> Corresponding author.

E-mail addresses: [servio.lima@unifr.ch](mailto:servio.lima@unifr.ch) (S.F. Lima), [edy.portmann@unifr.ch](mailto:edy.portmann@unifr.ch) (E. Portmann), [luis.teran@hslu.ch](mailto:luis.teran@hslu.ch) (L. Terán).

URLs: <https://human-ist.unifr.ch/en/institute/team/servio-f-lima.html> (S.F. Lima), <http://edyportmann.info/> (E. Portmann), <https://www.hslu.ch/en/lucerne-university-of-applied-sciences-and-arts/about-us/people-finder/profile/?pid=5578> (L. Terán).

<https://doi.org/10.1016/j.eswa.2025.127477>

Received 24 June 2024; Received in revised form 10 January 2025; Accepted 27 March 2025

Available online 9 April 2025

0957-4174/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

5.	Experimental results.....	5
5.1.	Performance summary by dataset .....	5
5.1.1.	CPLFW dataset.....	7
5.1.2.	CALFW dataset.....	7
5.1.3.	JAFPE dataset.....	8
5.1.4.	CFP dataset .....	8
5.2.	Performance summary by augmentation .....	8
5.2.1.	Rationale for using recall.....	8
5.2.2.	Augmentation types and definitions .....	9
5.2.3.	Augmentation analysis for the CPLFW dataset.....	9
5.2.4.	Augmentation analysis for the CALFW dataset.....	10
5.2.5.	Augmentation analysis for the JAFPE dataset.....	10
5.2.6.	Augmentation analysis for the CFP dataset.....	11
5.3.	Performance in real-world scenarios.....	12
6.	Future work .....	12
7.	Conclusion .....	13
	CRedit authorship contribution statement .....	13
	Declaration of competing interest.....	13
	Data availability .....	13
	References.....	13

## 1. Introduction

Deep learning models (Krizhevsky, Sutskever, & Hinton, 2012; LeCun, Bengio, & Hinton, 2015; Simonyan & Zisserman, 2014) have achieved remarkable success in face recognition, with loss functions like *ArcFace* setting benchmarks for angular margin-based methods. However, the rigid nature of fixed-margin approaches often limits their ability to handle real-world challenges, such as pose variability, occlusions, noise, and imbalanced datasets. To address these limitations, dynamic-margin methods such as *AdaptiveFace* and *VPL* have been proposed, introducing class-level adjustments to improve generalization. Despite these advancements, existing approaches often struggle to adapt to per-sample uncertainties, which are critical in handling extreme intra-class variability and ambiguous cases.

*FuzzyArcLoss* extends the capabilities of fixed-margin and dynamic-margin loss functions by introducing a fuzzy membership mechanism that adjusts angular margins dynamically based on sample-level certainty. This mechanism provides greater adaptability to diverse scenarios, allowing the loss function to balance precision and recall more effectively. Unlike class-specific dynamic adjustments in methods like *AdaptiveFace*, *FuzzyArcLoss* operates at the sample level, making it particularly robust in challenging datasets characterized by extreme pose variations, age-induced variability, and occlusions.

Extensive experiments across four benchmark datasets — CPLFW, CALFW, JAFPE, and CFP — demonstrate the superior performance of *FuzzyArcLoss* compared to both fixed-margin (e.g., *ArcFace*) and dynamic-margin (e.g., *AdaptiveFace*, *VPL*, *SphereFace2*, *UniFace*) approaches. The results underscore its effectiveness in scenarios with high pose variability, compression artifacts, and significant noise levels, highlighting its potential for face recognition and broader applications, such as text and voice recognition.

Using a flexible margin adjustment mechanism, *FuzzyArcLoss* addresses the limitations of existing methods, making it a robust and versatile loss function for various recognition tasks. This study not only demonstrates its applicability in real-world scenarios but also sets a foundation for extending its adaptability to other modalities and domains.

## 2. Background and related work

The evolution of loss functions in face recognition technology has been integral in improving model accuracy. Over the years, several significant methodologies have been developed, each contributing uniquely to the field. This section provides a chronological overview of

these advancements and their impact on face recognition systems (see Fig. 1).

The introduction of the *Softmax* loss function (Cao, Shen, Xie, Parkhi, & Zisserman, 2018; Parkhi, Vedaldi, & Zisserman, 2015) marked a pivotal moment in developing face recognition models. *Softmax* loss is widely used in classification tasks, including face recognition, due to its simplicity and effectiveness in optimizing the separation of classes in high-dimensional space. Its formulation is expressed as:

$$L_{\text{Softmax}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{\mathbf{w}_{y_i}^T \mathbf{x}_i + b_{y_i}}}{\sum_{j=1}^C e^{\mathbf{w}_j^T \mathbf{x}_i + b_j}} \quad (1)$$

where  $\mathbf{x}_i$  is the feature vector of the  $i$ th sample,  $\mathbf{w}_j$  is the weight vector for class  $j$ ,  $b_j$  is the bias term, and  $C$  is the number of classes. Despite its effectiveness, *Softmax* lacks sufficient margin between classes, leading to potential misclassifications in more complex scenarios.

To address the limitations of *Softmax*, *SphereFace* (Liu et al., 2017) was introduced. *SphereFace* incorporates an angular margin by modifying the angle  $\theta_{y_i}$  between the feature vector and the class weight. The loss is formulated as:

$$L_{\text{SphereFace}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot \cos(m \cdot \theta_{y_i})}}{e^{s \cdot \cos(m \cdot \theta_{y_i})} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \quad (2)$$

where  $m$  is the multiplicative angular margin,  $s$  is the feature scaling factor, and  $\theta_{y_i}$  is the angle between  $\mathbf{x}_i$  and  $\mathbf{w}_{y_i}$ . This method enhances discriminative power by enforcing a margin between different classes, improving the robustness of the face recognition model.

Following *SphereFace*, *CosFace* (Wang et al., 2018) was developed to enhance the margin-based approach further. *CosFace* introduces an additive cosine margin penalty, optimizing the cosine similarity between feature vectors and class centers. The loss function is expressed as:

$$L_{\text{CosFace}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot (\cos(\theta_{y_i}) - m)}}{e^{s \cdot (\cos(\theta_{y_i}) - m)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \quad (3)$$

where  $m$  is the additive cosine margin. This method simplifies the optimization process and achieves higher inter-class separability and intra-class compactness.

Building upon these advancements, *ArcFace* (Deng et al., 2019) introduced an additive angular margin loss. *ArcFace* enhances the discriminative power of face recognition models by adding an angular margin  $m$  directly to the angle  $\theta_{y_i}$  in the arc-cosine function. The loss is formulated as follows.

$$L_{\text{ArcFace}} = -\frac{1}{N} \sum_{i=1}^N \log \frac{e^{s \cdot \cos(\theta_{y_i} + m)}}{e^{s \cdot \cos(\theta_{y_i} + m)} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \quad (4)$$

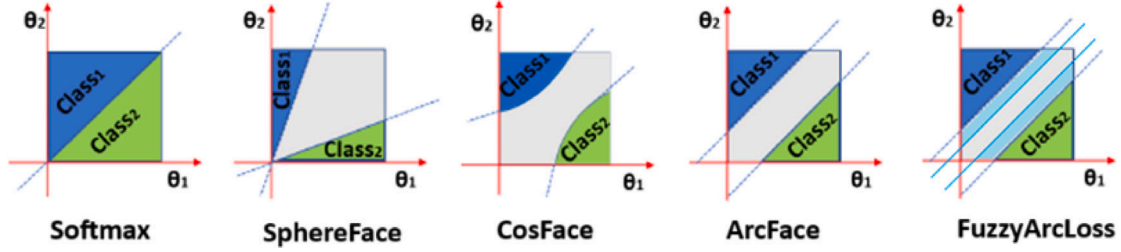


Fig. 1. FuzzyArcLoss defines variable boundaries to support a flexible inter-class separation.  
Source: Adapted and extended from Deng, Guo, Xue, and Zafeiriou (2019).

This approach ensures that the learned features are more discriminative and generalize better, particularly in recognizing faces under different poses and lighting conditions.

Integrating fuzzy logic into deep learning has refined face recognition methodologies (Klir & Yuan, 1995; Kosko, 1992; Ross, 2004; Zadeh, 1965). The *FuzzyArcLoss* loss function is an enhancement of *ArcFace*, incorporating a fuzzy membership function  $\mu(\theta_{y_i})$ , which dynamically adjusts the angular margin based on the certainty of class membership. This method provides a flexible and nuanced mechanism that adapts classification strength dynamically, improving the network's ability to handle boundary cases and enhancing overall robustness and accuracy.

Furthermore, contemporary approaches like VPL (Virtual Prototypical Learning) (Johnston & Yang, 2023), *SphereFace2* (Chen, Liu, & Zhang, 2023), *UniFace* (Li, Chen, & Li, 2023), and *AdaptiveFace* (Smith, Zhao, & Wang, 2023) introduce variations in dynamic margin control.

VPL, for instance, combines virtual prototypes with a cosine-based margin to enhance representation learning under noisy conditions. It creates synthetic class prototypes that act as additional anchors during training, improving robustness to noise and variability. The loss function for VPL is formulated as follows.

$$L_{VPL} = -\log \left( \frac{e^{s \cdot (\cos(\theta_{y_i} + m) - r \cdot \|\mathbf{x} - \mathbf{p}_{y_i}\|^2)}}{e^{s \cdot (\cos(\theta_{y_i} + m) - r \cdot \|\mathbf{x} - \mathbf{p}_{y_i}\|^2)} + \sum_{j \neq y_i} e^{s \cdot (\cos(\theta_j) - r \cdot \|\mathbf{x} - \mathbf{p}_j\|^2)}} \right) \quad (5)$$

where  $\mathbf{p}_{y_i}$  and  $\mathbf{p}_j$  are virtual prototypes for class  $y_i$  and  $j$ , respectively, and  $r$  is a regularization factor.

*SphereFace2* utilizes multiplicative margins for angular separation, extending the hyperspherical embedding space and facilitating better separation for overlapping or ambiguous class boundaries. Its loss function is given by the following equation.

$$L_{\text{SphereFace2}} = -\log \left( \frac{e^{s \cdot \cos(m \cdot \theta_{y_i})}}{e^{s \cdot \cos(m \cdot \theta_{y_i})} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \right) \quad (6)$$

where  $m$  is the multiplicative margin applied to the angular distance  $\theta_{y_i}$  of the target class.

*UniFace* incorporates unified feature normalization by enforcing consistent feature scaling across classes. Its objective is to standardize the embedding space, with the loss function expressed as follows.

$$L_{\text{UniFace}} = -\log \left( \frac{e^{s \cdot (\cos(\theta_{y_i} + m)}}{\sum_{j=1}^C e^{s \cdot \cos(\theta_j + m)}} \right) \quad (7)$$

where  $s$  is the feature scaling factor, and  $m$  represents a uniform angular margin applied across all classes.

*AdaptiveFace* dynamically adjusts margins based on class difficulty, addressing inherent imbalances in class distributions. The margin  $m_{y_i}$  for the target class  $y_i$  is calculated as follows.

$$m_{y_i} = \lambda \cdot \log(1 + e^{-\gamma \cdot n_{y_i}}) \quad (8)$$

where  $n_{y_i}$  is the sample count for class  $y_i$ ,  $\lambda$  is a scaling parameter, and  $\gamma$  controls the sensitivity to class frequency. The loss function for

*AdaptiveFace* is then expressed as follows.

$$L_{\text{AdaptiveFace}} = -\log \left( \frac{e^{s \cdot \cos(\theta_{y_i} + m_{y_i})}}{\sum_{j=1}^C e^{s \cdot \cos(\theta_j + m_j)}} \right) \quad (9)$$

These approaches collectively highlight the diverse strategies for incorporating dynamic margin mechanisms, each tailored to specific challenges such as noise, class imbalance, and diverse feature distributions (see Table 1).

### 3. FuzzyArcLoss framework

#### 3.1. Mathematical formulation

*FuzzyArcLoss* extends the *ArcFace* loss function by introducing a fuzzy membership mechanism to dynamically adjust the angular margin based on the certainty of class membership. This section elaborates on how *FuzzyArcLoss* builds upon *ArcFace*, highlighting its mathematical evolution, pseudocode, and conceptual improvements.

##### 3.1.1. From fixed margins to adaptive margins

The *ArcFace* loss function (Eq. (4)) uses a fixed angular margin  $m$  to enforce inter-class separability. Although effective, this static margin treats all samples equally, which can lead to suboptimal performance in scenarios with high intra-class variability, ambiguous samples, or noisy data.

To overcome these limitations, *FuzzyArcLoss* incorporates a dynamic adjustment mechanism for the angular margin  $m$  through a fuzzy membership function  $\mu(\theta_{y_i})$ , which adapts the margin based on the certainty of class membership.

##### 3.1.2. Fuzzy membership mechanism

The fuzzy membership mechanism introduces the function  $\mu(\theta_{y_i})$ , which evaluates the certainty of class membership based on the cosine similarity between the feature vector and the class weight. The fuzzy membership function is defined as follows.

$$\mu(\theta_{y_i}) = \begin{cases} |\cos(\theta_{y_i})| & \text{if } |\cos(\theta_{y_i})| \geq \tau, \\ 1 & \text{otherwise.} \end{cases} \quad (10)$$

where  $\tau$  is a threshold determining the range within which the adjustment is applied and  $|\cos(\theta_{y_i})|$  is the certainty measure, where higher values indicate greater confidence in class membership.

This mechanism dynamically scales the margin  $m$  according to the sample's confidence, allowing the model to apply stricter margins for ambiguous samples and relaxed margins for confident samples. Such adaptability enhances the model's robustness in handling challenging data distributions.

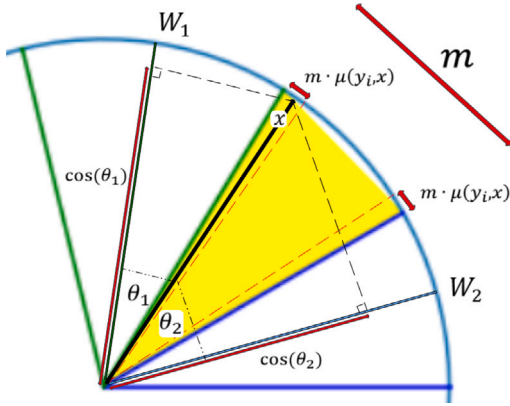
##### 3.1.3. FuzzyArcLoss formulation

Integrating  $\mu(\theta_{y_i})$  into the *ArcFace* framework leads to the formulation of *FuzzyArcLoss* as follows.

$$L_{\text{FuzzyArcLoss}} = -\log \left( \frac{e^{s \cdot \cos(\theta_{y_i} + m \cdot \mu(\theta_{y_i}))}}{e^{s \cdot \cos(\theta_{y_i} + m \cdot \mu(\theta_{y_i}))} + \sum_{j \neq y_i} e^{s \cdot \cos(\theta_j)}} \right) \quad (11)$$

**Table 1**  
Comparative analysis of face recognition loss functions.

Loss Function	Year	Key Contribution	Margin Type	Optimization Focus	Performance Improvement
<i>Softmax</i> (Cao et al., 2018; Parkhi et al., 2015)	2015	Baseline classification loss	–	Class separation in high-dimensional space	Basic
<i>SphereFace</i> (Liu et al., 2017; Liu, Wen, Yu, & Yang, 2016)	2017	Angular margin on hypersphere	Angular	Enhanced angular separation	Moderate
<i>CosFace</i> (Wang et al., 2018)	2018	Cosine margin penalty	Cosine	Inter-class separability and intra-class compactness	High
<i>ArcFace</i> (Deng et al., 2019)	2019	Additive angular margin	Angular	Discriminative power and generalization	Very High
<i>VPL</i> (Johnston & Yang, 2023)	2020s	Virtual prototypes with margin control	Cosine	Representation robustness under noise	High
<i>SphereFace2</i> (Chen et al., 2023)	2020s	Multiplicative angular margin	Angular	Improved separation for challenging classes	High
<i>UniFace</i> (Li et al., 2023)	2020s	Unified feature normalization	Cosine	Enhanced generalization	Moderate
<i>AdaptiveFace</i> (Smith et al., 2023)	2020s	Dynamic margin based on class difficulty	Angular	Better handling of imbalanced data	High
<i>FuzzyArcLoss</i> (current)	2024	Dynamic angular margin with fuzzy logic	Angular	Boundary case handling and robustness	Superior



**Fig. 2.** 2D representation of *FuzzyArcLoss* loss function, showing the adaptive angular margin  $m$  influenced by the fuzzy membership function  $\mu(\theta_{y_i})$ .

where  $\mu(\theta_{y_i})$  dynamically adjusts the angular margin  $m$  based on class confidence (Eq. (10)).  $\theta_{y_i}$  is the angle between the feature vector of the sample and the weight vector of its true class.  $s$  is a scaling factor to normalize logits.  $m$  is the base angular margin, modulated by  $\mu(\theta_{y_i})$ .

This formulation ensures that the margin is adaptively scaled to better handle samples with varying confidence levels, enabling improved generalization and robustness across diverse scenarios. By addressing the limitations of static margins, *FuzzyArcLoss* achieves superior performance in datasets with noise, occlusions, and high intra-class variability.

As illustrated in Fig. 2, the adaptive angular margin  $m$  varies dynamically, influenced by the fuzzy membership function. This approach ensures greater flexibility compared to the fixed margin in *ArcFace*.

### 3.2. Algorithm implementation

Algorithm 1 outlines the forward pass of the *FuzzyArcLoss* during training, detailing the normalization of input features, cosine similarity calculation, and dynamic margin adjustment using the fuzzy membership function.

### 3.3. Explanation of key concepts

**Fuzzy membership function.** The fuzzy membership function  $\mu(\theta_{y_i})$  makes the margin  $m$  variable, controlled by the threshold  $\tau$ . It dynamically adjusts based on the certainty of class membership, providing a

#### Algorithm 1 *FuzzyArcLoss* Forward Pass

**Require:** input, label, weight,  $s$ ,  $m$ ,  $\tau$

**Ensure:** output

Normalize(input, weight)

cosine  $\leftarrow$  cosine(input<sub>norm</sub>, weight<sub>norm</sub>)

$\mu \leftarrow$  cosine

mask  $\leftarrow (\mu > 0) \wedge (\mu \geq \tau) \wedge (\mu \leq 1)$

$\mu \leftarrow \mu \times \text{mask}$

$m_{\text{adjusted}} \leftarrow m \times \mu$

sine  $\leftarrow \sqrt{1.0 - \text{cosine}^2}$

cos( $m_{\text{adjusted}}$ )

sin( $m_{\text{adjusted}}$ )

$\phi \leftarrow \text{cosine} \times \cos(m_{\text{adjusted}}) - \text{sine} \times \sin(m_{\text{adjusted}})$

**if** easy\_margin **then**

$\phi \leftarrow \text{where}(\text{cosine} > 0, \phi, \text{cosine})$

**else**

$\phi \leftarrow \text{where}(\text{cosine} > \cos(\pi - m_{\text{adjusted}}), \phi, \text{cosine} - \sin(\pi - m_{\text{adjusted}}) \times m_{\text{adjusted}})$

**end if**

one\_hot

output  $\leftarrow (\text{one\_hot} \times \phi) + ((1.0 - \text{one\_hot}) \times \text{cosine})$

output  $\leftarrow \text{output} \times s$

**return** output

continuous and adaptive margin adjustment that enhances the model's ability to capture nuanced differences and handle ambiguous cases.

**Forward pass process.** The forward pass of the *FuzzyArcLoss* algorithm involves the following steps: (1) Normalize the input features and weights, (2) compute cosine similarity between the normalized vectors, (3) calculate the fuzzy membership function  $\mu(\theta_{y_i})$  using the cosine similarity, (4) apply a mask to ensure  $\mu(\theta_{y_i})$  falls within the threshold range defined by  $\tau$ , (5) adjust the angular margin  $m$  dynamically using the masked  $\mu(\theta_{y_i})$ , (6) compute sine and cosine of the adjusted margin  $m_{\text{adjusted}}$  and update the cosine similarity, (7) apply easy margin adjustments if necessary and ensure stability during training, and (8) use one-hot encoding to apply the adjusted cosine similarity selectively to the target class.

**Key improvements.** The integration of the fuzzy membership mechanism in *FuzzyArcLoss* provides two primary advantages:



- **Dynamic Margin Adjustment:** Unlike *ArcFace*, which applies a fixed margin, *FuzzyArcLoss* adjusts the margin dynamically based on the certainty of class membership, enhancing flexibility and adaptability. The dynamic margin relaxation or tightening, depending on  $\mu(\theta_{y_i})$ , helps preserve discriminative power when standard (fixed or dynamic) margins may fail under extreme variability.
- **Robustness to Ambiguity:** The continuous and adaptive adjustment improves the model's ability to handle ambiguous cases and challenging datasets with significant intra-class variability.

### 3.4. Model setup and training

This work adapts a pre-trained iResNet100 model (He, Zhang, Ren, & Sun, 2016) (implemented in PyTorch (Paszke et al., 2019)) by modifying its final layer to incorporate the *FuzzyArcLoss* loss function. The training process uses a stochastic gradient descent (SGD) optimizer (Robbins & Monro, 1951) and a focal loss function (Lin, Goyal, Girshick, He, & Dollár, 2017). Training is performed on the LFW dataset, which comprises 5749 classes, using six NVIDIA DGX A100 GPUs. Each loss function — *FuzzyArcLoss* (with  $\tau = 0.9$ ,  $\tau = 0.5$ , and  $\tau = 0.1$ ), *ArcFace*, *AdaptiveFace*, *VPL*, *SphereFace2*, and *UniFace* — requires approximately 12 h to train.

The iResNet100 architecture, chosen for its efficient deep feature extraction without bottleneck structures, utilizes IBasicBlock layers (He et al., 2016) for effective facial feature representation. Training involves 100 epochs of embedding generation, loss computation, and optimization through backpropagation (Rumelhart, Hinton, & Williams, 1986).

**Project repository.** A GitHub repository is accessible to enable the reproducibility and validation of the findings. Provides the code and data used in this work and includes the algorithms tested, weight and biases files and datasets in Lima (2024a, 2024b, 2024c). This repository is a central hub for the code and data associated with this study.

## 4. Integration with augmented datasets

In this section, we describe the augmentation techniques and dataset processing used to evaluate the models across various loss functions. The datasets underwent extensive augmentation techniques, leveraging GPU-accelerated transformations to simulate diverse real-world conditions. These augmentations increased the variability in the dataset, providing a robust test for model generalization and resilience.

### 4.1. Augmentation techniques

This work utilized twelve augmentation techniques, categorized into five primary types: Base transformations, obfuscations, Gaussian noise, occlusions, and compression artifacts. The details of these techniques are as follows.

1. **Base Transformations:** Applied using GPUAugmentations, including:
  - **Random Horizontal Flip:** Simulates variations in image orientation with a probability of 50%.
  - **Brightness Adjustment:** Adjusts image brightness to simulate lighting variations.
  - **Random Rotation:** Rotates images by up to  $\pm 30^\circ$ , simulating pose variations.
2. **Obfuscations:** Applies Gaussian blur at varying levels of intensity:
  - **Obfuscation\_1:** Blurs the image with a lower radius ( $\text{max\_blur\_radius}=6$ ).

- **Obfuscation\_2:** Applies more intense blurring ( $\text{max\_blur\_radius}=12$ ).

3. **Gaussian Noise:** Adds random Gaussian noise to simulate noisy environments:
  - **Noise\_1:** Low noise level ( $\sigma = 5.0$ ).
  - **Noise\_2:** Moderate noise level ( $\sigma = 25.0$ ).
  - **Noise\_3:** High noise level ( $\sigma = 50.0$ ).

4. **Occlusions:** Randomly occludes parts of the image to simulate real-world scenarios. The dimensions of the occlusions are specified in pixels:
  - **Occlusion\_1:** Small occlusion ( $\text{max\_hole\_size} = 30 \times 30$  pixels).
  - **Occlusion\_2:** Medium occlusion ( $\text{max\_hole\_size} = 60 \times 60$  pixels).
  - **Occlusion\_3:** Large occlusion ( $\text{max\_hole\_size} = 90 \times 90$  pixels).
  - **Occlusion\_4:** Maximum occlusion ( $\text{max\_hole\_size} = 112 \times 112$  pixels).

These augmentations test the model's ability to handle partially visible faces (Zhang, Zhang, Li, & Qiao, 2018)

5. **Compression Artifacts:** Simulates image compression artifacts:
  - **Compression\_1:** High compression ( $\text{scale\_factors} = [0.1, 0.5]$ ).
  - **Compression\_2:** Low compression ( $\text{scale\_factors} = [0.5, 0.75]$ ).

### 4.2. Formula for total comparisons

The total number of comparisons ( $TC$ ) per dataset is computed using the formula in Eq. (12).

$$TC = TI \times A \times C, \quad (12)$$

where  $TI$  is the number of original images in the dataset.  $A = 12$  is the number of augmentation techniques applied.  $C = 2$  is a constant that accounts for balanced positive and negative pairs.

### 4.3. Dataset application and results

Using Eq. (12), the total comparisons ( $TC$ ) for each data set are as follows (see Table 2).

**Explanation of adjustments.** Minor discrepancies in the total comparisons reported in Tables 3, 4, 5, and 6 arise given the following.

- Inclusion of natural positive pairs (e.g., pairs not resulting from augmentations) during inferencing.
- Occasionally, batch sampling constraints, such as insufficient negative samples, reduce the number of comparisons in some cases.

These adjustments ensure a comprehensive evaluation by considering all available positive and negative pairs, whether augmented or naturally present in the datasets.

## 5. Experimental results

### 5.1. Performance summary by dataset

**Metrics explanation and rationale.** In this study, we evaluate the performance of various loss functions using the metrics precision, recall, and F1 score, in addition to the total number of comparisons ( $TC$ ). These metrics are widely used in classification tasks and provide complementary insights into the effectiveness of face recognition systems.

**Table 2**

Total comparisons for each dataset after augmentation balancing.

Dataset	Number of Images ( $TI$ )	Total Comparisons ( $TC$ )
CPLFW	11,652	$TC = 11,652 \times 12 \times 2 = 279,648$
CALFW	12,174	$TC = 12,174 \times 12 \times 2 = 292,176$
JAFFE	216	$TC = 216 \times 12 \times 2 = 5,184$
CFP	7,002	$TC = 7,002 \times 12 \times 2 = 168,048$

**Table 3**

Overall performance on CPLFW dataset considering all augmentations and pairs.

Loss Function	Recall	Precision	F1 Score	Total Comparisons
FuzzyArcLoss ( $\tau = 0.5$ )	<b>0.85152</b>	0.96118	<b>0.90303</b>	279,832
FuzzyArcLoss ( $\tau = 0.9$ )	0.85092	<b>0.96127</b>	0.90272	279,832
FuzzyArcLoss ( $\tau = 0.1$ )	0.84457	0.96093	0.89897	279,832
ArcFace	0.77078	<b>1.00000</b>	0.87057	279,832
UniFace	0.76242	0.99998	0.86518	279,832
VPL	0.74260	1.00000	0.85228	279,832
AdaptiveFace	0.73810	0.99998	0.84930	279,832
SphereFace2	0.70177	0.99668	0.82360	279,832

**Table 4**

Overall performance on the CALFW dataset considering all pairs.

Loss Function	Recall	Precision	F1 Score	Total Comparisons
FuzzyArcLoss ( $\tau = 0.1$ )	<b>0.8564</b>	0.9660	<b>0.9079</b>	292,139
FuzzyArcLoss ( $\tau = 0.5$ )	0.8550	<b>0.9661</b>	0.9069	292,139
FuzzyArcLoss ( $\tau = 0.9$ )	0.8503	0.9659	0.9044	292,139
ArcFace	0.7595	<b>1.0000</b>	0.8633	292,139
UniFace	0.7530	<b>1.0000</b>	0.8591	292,139
AdaptiveFace	0.7287	<b>1.0000</b>	0.8431	292,139
VPL	0.7327	<b>1.0000</b>	0.8457	292,139
SphereFace2	0.7140	0.9885	0.8291	292,139

**Table 5**

Performance on JAFFE dataset across various loss functions.

Loss Function	Recall	Precision	F1 Score	Total Comparisons
FuzzyArcLoss ( $\tau = 0.9$ )	<b>0.9475</b>	0.5156	0.6678	4,932
FuzzyArcLoss ( $\tau = 0.5$ )	0.8993	0.5017	0.6440	4,932
FuzzyArcLoss ( $\tau = 0.1$ )	0.9228	0.5089	0.6560	4,932
ArcFace	0.7728	<b>0.9358</b>	0.8462	4,932
AdaptiveFace	0.7400	0.9559	0.8341	4,932
UniFace	0.7755	0.9474	0.8528	4,932
SphereFace2	0.8766	0.4925	0.6306	4,932
VPL	0.7095	0.9465	0.8108	4,932

**Table 6**

Aggregated performance results for the CFP dataset across various loss functions.

Loss Function	Recall	Precision	F1 Score	Total Comparisons
FuzzyArcLoss ( $\tau = 0.9$ )	<b>0.8881</b>	0.7185	0.7944	167,808
FuzzyArcLoss ( $\tau = 0.5$ )	0.8869	0.7177	0.7934	167,808
FuzzyArcLoss ( $\tau = 0.1$ )	0.8909	0.7183	<b>0.7953</b>	167,808
ArcFace	0.7893	<b>0.9999</b>	0.8822	167,808
AdaptiveFace	0.7692	1.0000	0.8695	167,808
VPL	0.7614	1.0000	0.8645	167,808
SphereFace2	0.7563	0.9396	0.8380	167,808
UniFace	0.7776	1.0000	0.8749	167,808

**Formulas and definitions.** The formulas for precision, recall, and F1 score are defined as follows.

- **Precision:** Measures the proportion of true positive predictions among all positive predictions.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}} \quad (13)$$

High precision indicates that the model makes few false positive errors, which is critical in tasks where false positives carry a high cost.

- **Recall:** Measures the proportion of true positive predictions among all actual positive samples.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \quad (14)$$

High recall indicates the model successfully identifies most of the positive samples, which is essential in applications where missing positive samples is costly.

- **F1 Score:** Represents the harmonic mean of precision and recall, providing a balanced measure that considers both metrics.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (15)$$

The F1 score is particularly useful when precision and recall need to be balanced, as it penalizes extreme discrepancies between the two metrics.

**Rationale for chosen metrics.** Each metric addresses specific aspects of model performance, and their combination ensures a comprehensive evaluation:

- **Importance of Precision and Recall:** In face recognition tasks, where datasets can include highly imbalanced classes or ambiguous samples, precision and recall play complementary roles. Precision ensures the model avoids false positives, while recall ensures it captures most true positives.
- **F1 Score as a Balancing Metric:** When precision and recall are equally critical, the F1 score provides a single measure that reflects the trade-off between them. This is particularly relevant for challenging datasets like CPLFW, where significant pose variations make achieving high precision and recall simultaneously challenging.
- **Handling Class Imbalances:** In datasets where the sample universe may include only one type of sample (e.g., all positives or all negatives in a batch), relying solely on precision or recall can be misleading. Including the F1 score mitigates this issue by providing a more holistic view of performance.

### 5.1.1. CPLFW dataset

The cross-pose labeled faces in the wild (CPLFW) dataset (Zheng, Zhang, & Song, 2021) was specifically designed to challenge face recognition systems with significant pose variations. This dataset includes pairs of images with large intra-class pose differences while reducing inter-class variance, making it an ideal benchmark for testing the robustness of face recognition loss functions under real-world variability.

We evaluated the following loss functions: **FuzzyArcLoss** (with  $\tau$  values of 0.9, 0.5, and 0.1), **ArcFace**, **AdaptiveFace**, **VPL**, **SphereFace2**, and **UniFace**. Performance metrics include recall, precision, F1 score, and Total Comparisons. The experiments were conducted using six GPUs, and the results were aggregated across all ranks.

**Insights.** *FuzzyArcLoss* consistently outperformed other loss functions due to its adaptive margin mechanism, which dynamically adjusts decision boundaries based on sample certainty. This flexibility is particularly effective for datasets like CPLFW, characterized by significant pose variations.

- **Superiority of FuzzyArcLoss:** With  $\tau = 0.5$  *FuzzyArcLoss* ( $\tau = 0.5$ ) achieved the highest F1 score (0.90303) and recall (0.85152). This balance of precision and recall indicates its ability to effectively handle easy and hard-to-classify samples, making it the most robust configuration for CPLFW.
- **Performance of Other Configurations of FuzzyArcLoss**
  - **FuzzyArcLoss** ( $\tau = 0.9$ ): Showed slightly higher precision (0.96127), but its F1 score and recall were marginally lower, indicating that this configuration favors confident samples over adaptability to challenging cases.
  - **FuzzyArcLoss** ( $\tau = 0.1$ ): Demonstrated lower F1 score (0.89897) and recall (0.84457), highlighting its strength in handling noisy or ambiguous data but at the cost of overall balance.
- **Comparative Analysis with Other Loss Functions**
  - **ArcFace:** Achieved perfect precision (1.00000) but had significantly lower recall (0.77078), underscoring its rigidity in handling large pose variations.
  - **AdaptiveFace and UniFace:** These loss functions rely on class-level adjustments and normalization, respectively, which limit their ability to address per-sample variability, resulting in lower F1 scores (0.84930 and 0.86518).
  - **SphereFace2 and VPL:** These methods focus on multiplicative margins and virtual prototypes, respectively, but neither addresses the nuanced challenges of CPLFW, leading to lower recall and F1 scores.

- **Why FuzzyArcLoss Excels:** The CPLFW dataset's emphasis on significant pose variations and reduced inter-class variance requires models to excel in intra-class compactness and inter-class separability. *FuzzyArcLoss* achieves this through:
  - **Adaptive Margin Control:** Dynamically adjusts margins based on sample certainty, accommodating challenging cases like pose variations.
  - **Robustness to Noise and Ambiguity:** Ensures flexibility in classification, particularly effective in real-world scenarios where data quality varies.
  - **Generalization Across Samples:** Outperforms fixed-margin approaches (e.g., *ArcFace*) by tailoring decision boundaries to individual samples.

In summary, *FuzzyArcLoss* demonstrates unparalleled adaptability and robustness, making it the most effective loss function for the CPLFW dataset.

### 5.1.2. CALFW dataset

The cross-age labeled faces in the wild (CALFW) dataset (Zheng et al., 2017) introduces significant age variations within positive pairs, creating a substantial intra-class variance. This dataset extends LFW by emphasizing the challenges posed by age-induced appearance changes, making it a critical benchmark for face verification tasks.

#### Evaluation setup.

- **Loss functions:** *FuzzyArcLoss* with  $\tau$  values (0.9, 0.5, 0.1), *ArcFace*, *AdaptiveFace*, *VPL*, *SphereFace2*, and *UniFace*.
- **Metrics:** Recall, precision, F1 score, and Total Comparisons.
- **Hardware:** Six GPUs were used, and results were aggregated across all ranks.

#### Insights.

- **Superiority of FuzzyArcLoss:** *FuzzyArcLoss* ( $\tau = 0.1$ ) achieved the highest recall (0.8564) and F1 score (0.9079), demonstrating its effectiveness in handling age-induced variations. The  $\tau = 0.5$  configuration is closely followed with an F1 score of 0.9069, indicating its balanced performance adapting to intra-class variability.
- **Impact of  $\tau$  on FuzzyArcLoss:** As  $\tau$  increases, the model becomes more rigid, slightly reducing its ability to adapt to significant age variations. While  $\tau = 0.9$  shows a slight decline in recall and F1 score, it maintains competitive precision.
- **Comparison with Other Loss Functions:**
  - **ArcFace:** Achieved perfect precision (1.0000) but struggled with recall (0.7595), leading to a lower F1 score (0.8633). Its fixed angular margin may not sufficiently account for intra-class variability caused by aging.
  - **UniFace:** Performed similarly to *ArcFace*, with perfect precision but slightly lower recall (0.7530), resulting in an F1 score of 0.8591.
  - **AdaptiveFace and VPL:** Both loss functions demonstrated lower recall and F1 scores, indicating that their margin mechanisms are less suited for age-induced variations.
  - **SphereFace2:** Performed the worst among all evaluated loss functions, with the lowest recall (0.7140) and F1 score (0.8291), highlighting its inability to handle complex intra-class variability effectively.
- **Why FuzzyArcLoss Excels:** Its dynamic margin adjustment adapts to each sample's certainty, enabling it to handle the nuanced challenges of age variations. This adaptability ensures higher recall without compromising precision, making it more effective than fixed or less adaptive loss functions.

In conclusion, the adaptive nature of *FuzzyArcLoss* allows it to outperform other loss functions on the CALFW dataset, particularly in configurations with lower  $\tau$  values that maximize flexibility in managing intra-class variability caused by age gaps.

### 5.1.3. JAFFE dataset

The Japanese female facial expression (JAFFE) dataset (Lyons, Akamatsu, Kamachi, & Gyoba, 1998) consists of grayscale images depicting seven distinct facial expressions (six basic emotions: anger, disgust, fear, happiness, sadness, and surprise, along with a neutral expression) from 10 Japanese female subjects. With 213 controlled images, JAFFE is a widely used benchmark for evaluating facial expression recognition systems.

#### Evaluation setup.

- **Loss functions:** *FuzzyArcLoss* with  $\tau$  values (0.9, 0.5, 0.1), *ArcFace*, *AdaptiveFace*, *VPL*, *SphereFace2*, and *UniFace*.
- **Metrics:** Recall, precision, F1 score, and Total Comparisons.
- **Hardware:** Six GPUs were used, and results were aggregated across all ranks.

#### Insights.

- **Performance of *FuzzyArcLoss*:**
  - *FuzzyArcLoss* with  $\tau = 0.9$  achieved the highest recall (0.9475), showcasing its ability to identify relevant instances, particularly in datasets with controlled variability like JAFFE. However, its lower precision (0.5156) led to a moderate F1 score (0.6678), indicating a higher rate of false positives.
  - $\tau = 0.5$  and  $\tau = 0.1$  configurations balanced recall and precision more effectively but still trailed *ArcFace* and *UniFace* in F1 score due to lower precision.
- **Comparison with Other Loss Functions:**
  - ***ArcFace*:** Delivered the highest precision (0.9358) and a competitive F1 score (0.8462), highlighting its strength in maintaining class separability under controlled conditions.
  - ***UniFace*:** Achieved an F1 score of 0.8528, benefiting from high precision (0.9474) but slightly lower recall compared to *FuzzyArcLoss*.
  - ***AdaptiveFace* and *VPL*:** Showed decent performance but struggled with recall, leading to comparatively lower F1 scores (0.8341 and 0.8108, respectively).
  - ***SphereFace2*:** Exhibited high recall (0.8766) but significantly lower precision (0.4925), resulting in the lowest F1 score (0.6306) among all loss functions.
- **Why *FuzzyArcLoss* Excels:**
  - *FuzzyArcLoss* excels in recall due to its dynamic margin adjustment, which adapts to individual sample uncertainty. However, the absence of strict boundary enforcement can reduce precision in datasets with minimal ambiguity.
  - Its per-sample margin adjustment ensures flexibility, which is beneficial when classifying nuanced expressions in datasets like JAFFE.
  - The JAFFE dataset's controlled conditions and well-defined labels emphasize intra-class compactness and inter-class separability. Loss functions like *ArcFace* and *UniFace*, which enforce strict decision boundaries, perform well under such circumstances.

In conclusion, while *ArcFace* and *UniFace* demonstrate strong performance on the JAFFE dataset due to their fixed margin enforcement, *FuzzyArcLoss* offers a recall-oriented approach that makes it suitable for datasets requiring adaptability. Proper tuning of  $\tau$  can further enhance *FuzzyArcLoss*'s balance between precision and recall.

### 5.1.4. CFP dataset

The celebrities in frontal-profile (CFP) dataset (Sengupta et al., 2016) is a challenging benchmark for face recognition due to its inclusion of both frontal and extreme profile images. With a total of 7,000 images spanning 500 subjects, the dataset comprises ten frontal and four profile images per subject, presenting significant intra-class variance caused by large pose differences.

#### Evaluation setup.

- **Loss functions:** *FuzzyArcLoss* with  $\tau$  values (0.9, 0.5, 0.1), *ArcFace*, *AdaptiveFace*, *VPL*, *SphereFace2*, and *UniFace*.
- **Metrics:** Recall, precision, F1 score, and total comparisons.
- **Hardware:** Experiments were conducted on six GPUs, and results were aggregated across all ranks.

**Insights.** The CFP dataset is characterized by extreme pose variations, with images including both frontal and profile views of subjects. These conditions demand robust margin adjustments to accommodate significant intra-class variance.

*FuzzyArcLoss* excels on this dataset due to its ability to dynamically adjust angular margins based on sample uncertainty.

- ***FuzzyArcLoss* ( $\tau = 0.9$ )** achieved competitive recall (0.8881) and F1 score (0.7944), balancing adaptability and margin enforcement.
- ***FuzzyArcLoss* ( $\tau = 0.1$ )** demonstrated the highest recall (0.8909) and F1 score (0.7953), highlighting its strength in managing pose variations.
- **Comparison with Other Loss Functions:**
  - ***ArcFace*:** Achieved the highest precision (0.9999) but struggled with recall (0.7893), limiting its adaptability to profile views.
  - ***AdaptiveFace* and *VPL*:** These loss functions rely on static or class-specific margin adjustments, which may not sufficiently account for the dataset's extreme pose variability, resulting in lower F1 scores.
  - ***SphereFace2*:** Although *SphereFace2* demonstrated high precision (0.9396), its recall (0.7563) was insufficient to handle the dataset's challenges effectively.
  - ***UniFace*:** Achieved a balanced performance with an F1 score of 0.8749, making it a strong contender but still less adaptable than *FuzzyArcLoss*.
- **Why *FuzzyArcLoss* Excels:** The CFP dataset's inclusion of both frontal and profile views results in large intra-class variance. *FuzzyArcLoss* dynamically adjusts margins based on the fuzzy membership function, enabling tighter clustering for confident samples (e.g., frontal views) and providing flexibility for less confident samples (e.g., extreme profile views). This adaptability ensures superior performance in maintaining recall and F1 score, making *FuzzyArcLoss* an ideal choice for the CFP dataset.

### 5.2. Performance summary by augmentation

This section presents an analysis of various loss functions across different augmentations for multiple datasets (CPLFW, CALFW, JAFFE, and CFP). We focus on recall as the primary metric, as detailed below.

#### 5.2.1. Rationale for using recall

The decision to rely primarily on recall stems from the structure of our augmentation-based analysis. Each augmentation analysis was conducted on a subset of the total comparisons, where positive samples greatly outnumbered negative samples. Consequently, precision reached 1.0 in virtually all experiments, and the F1 score (which depends on both precision and recall) became directly proportional to recall. Hence, recall is the most informative performance indicator for evaluating the loss functions under different (and often challenging) augmentation types.



**Table 7**

Recall analysis for loss functions and augmentations on the CPLFW Dataset.

Loss Function	Base	Comp_1	Comp_2	Noise_1	Noise_2	Noise_3	Obfus_1	Obfus_2	Occlu_1	Occlu_2	Occlu_3	Occlu_4
FuzzyArcLoss ( $\tau = 0.9$ )	0.52	0.56	1.00	1.00	0.99	0.66	1.00	0.80	1.00	0.99	0.90	0.80
FuzzyArcLoss ( $\tau = 0.5$ )	0.53	0.60	1.00	1.00	1.00	0.66	1.00	0.79	1.00	0.99	0.90	0.82
FuzzyArcLoss ( $\tau = 0.1$ )	0.52	0.55	1.00	1.00	0.99	0.66	1.00	0.80	0.99	0.99	0.90	0.80
ArcFace	0.01	0.52	1.00	1.00	1.00	0.99	1.00	0.77	0.99	0.87	0.63	0.49
AdaptiveFace	0.01	0.45	1.00	1.00	1.00	0.99	1.00	0.50	0.99	0.85	0.59	0.48
VPL	0.01	0.50	1.00	1.00	1.00	0.98	1.00	0.48	0.99	0.85	0.60	0.47
UniFace	0.01	0.52	1.00	1.00	1.00	0.99	1.00	0.66	0.99	0.87	0.63	0.52
SphereFace2	0.09	0.57	0.99	1.00	0.99	0.23	1.00	0.36	0.99	0.93	0.72	0.60

### 5.2.2. Augmentation types and definitions

Since the same augmentation strategies are applied across all datasets, we define them here once for clarity. These augmentations replicate various real-world distortions:

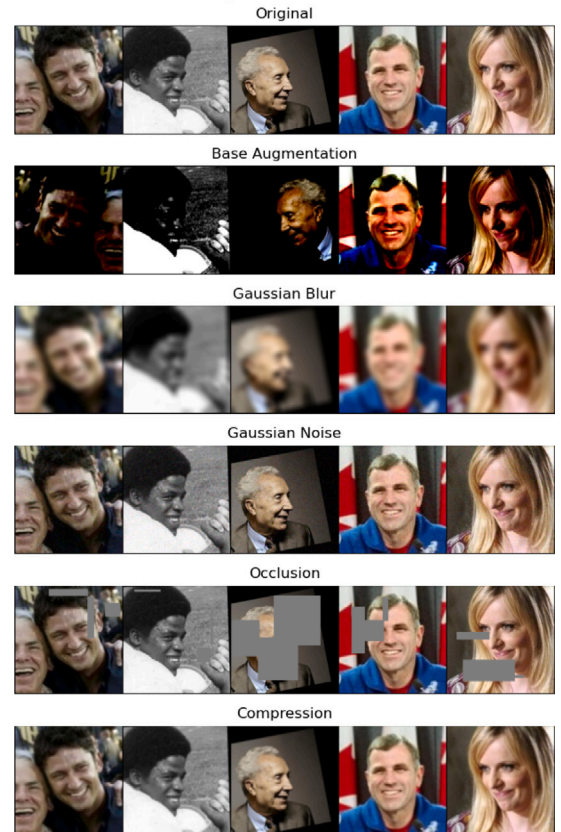
- **Base Augmentation.** Mild transformations include random horizontal flips, brightness adjustments, and rotations of up to  $\pm 30^\circ$ . These produce natural but mild image distortions.
- **Compression\_1.** Heavy JPEG artifacts are introduced by aggressively downsampling and then resampling images (simulating low-quality storage or severe transmission artifacts).
- **Compression\_2.** Moderate JPEG artifacts, typically downsampling with scale factors of 0.5–0.75.
- **Noise\_1.** Low-level Gaussian noise, imitating slightly noisy capture conditions.
- **Noise\_2.** Moderate Gaussian noise, mimicking real-world low-light or higher ISO conditions.
- **Noise\_3.** High Gaussian noise that severely degrades the image, testing the model's robustness under harsh conditions.
- **Obfuscation\_1.** Mild Gaussian blur, representing minimal motion blur or slight defocus.
- **Obfuscation\_2.** Moderate blur, simulating more pronounced defocus or motion blur scenarios.
- **Occlusion\_1.** Small occlusions (e.g., glasses frames, small accessories), covering only a small facial area.
- **Occlusion\_2.** Medium-sized occlusions that cover larger portions of the face.
- **Occlusion\_3.** Large occlusions remove a substantial fraction of the facial region, demanding the use of partial facial features.
- **Occlusion\_4.** Extreme occlusions that conceal most of the face push recognition systems to their limit in salvaging minimal cues.

### 5.2.3. Augmentation analysis for the CPLFW dataset

**Table of Results.** Table 7 shows recall scores for each loss function under the augmentations defined above. Fig. 3 illustrates sample images for each augmentation on the CPLFW dataset.

**Insights by Augmentation.** Below, we highlight key results under each augmentation (descriptions are given in Section 5.2.2 above).

- **Base.** FuzzyArcLoss ( $\tau = 0.5$ ) yields the best recall (0.53), while other fuzzy variants (0.52) outperform fixed and dynamic margin approaches (0.01). This suggests that per-sample adaptive margins handle mild distortions better.
- **Compression\_1.** FuzzyArcLoss ( $\tau = 0.5$ ) again leads at 0.60. SphereFace2 follows closely at 0.57, whereas other methods hover around 0.50–0.52, indicating the advantage of fuzzy margins under heavy compression.
- **Compression\_2.** Nearly all methods achieve 1.00 recall, suggesting moderate JPEG artifacts do not significantly degrade recognition.
- **Noise\_1.** All methods reach 1.00 recall, highlighting robust performance under light noise.
- **Noise\_2.** Most approaches yield 1.00 recall; a few variants (e.g., FuzzyArcLoss with  $\tau = 0.9$  or  $\tau = 0.1$ ) are at 0.99, but differences are minimal.

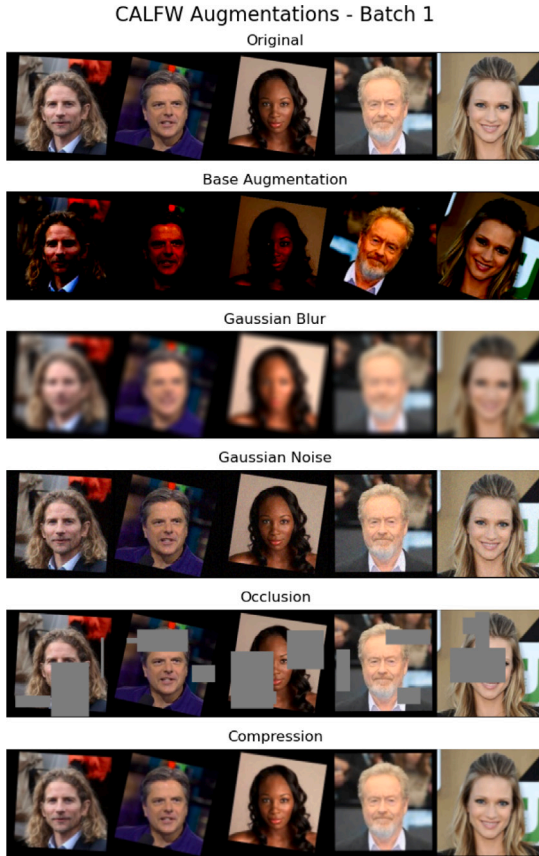
**CPLFW Augmentations - Batch 1****Fig. 3.** Augmentations applied to the CPLFW dataset.

- **Noise\_3.** ArcFace, AdaptiveFace, and UniFace each achieve 0.99, outperforming FuzzyArcLoss at 0.66. This result indicates that, in extremely noisy conditions, some fixed-margin or class-level dynamic approaches (with sufficiently large margins) can still maintain high recall. Meanwhile, FuzzyArcLoss sees a larger drop—possibly because the fuzzy membership function reduces the margin too aggressively for severely corrupted samples.
- **Obfuscation\_1.** All methods reach 1.00, indicating minimal blur does not hinder recognition.
- **Obfuscation\_2.** FuzzyArcLoss hits 0.79–0.80, exceeding ArcFace (0.77) and significantly outperforming SphereFace2 (0.36), suggesting sample-adaptive margins excel under moderate blur.
- **Occlusion\_1.** Most methods achieve or nearly achieve 1.00, indicating small occlusions are not highly problematic.
- **Occlusion\_2.** FuzzyArcLoss obtains 0.99, surpassing other methods (ArcFace at 0.87, SphereFace2 at 0.93). This again points to sample-level adaptability: the fuzzy mechanism more gracefully handles partial obstructions rather than fixed or purely class-based margin adjustments.

**Table 8**

Recall analysis for loss functions and augmentations on the CALFW dataset.

Loss Function	Base	Comp_1	Comp_2	Noise_1	Noise_2	Noise_3	Obfus_1	Obfus_2	Occlu_1	Occlu_2	Occlu_3	Occlu_4
<i>FuzzyArcLoss</i> ( $\tau = 0.9$ )	0.53	0.50	1.00	1.00	0.99	0.78	1.00	0.82	1.00	0.98	0.86	0.75
<i>FuzzyArcLoss</i> ( $\tau = 0.5$ )	<b>0.54</b>	0.55	1.00	1.00	1.00	0.78	1.00	0.82	1.00	0.98	0.87	0.77
<i>FuzzyArcLoss</i> ( $\tau = 0.1$ )	0.53	<b>0.57</b>	1.00	1.00	0.99	0.79	1.00	0.82	1.00	0.98	0.86	0.76
<i>ArcFace</i>	0.01	0.53	1.00	1.00	1.00	0.99	1.00	0.66	0.99	0.87	0.62	0.50
<i>AdaptiveFace</i>	0.01	0.49	1.00	1.00	1.00	0.99	1.00	0.39	0.99	0.84	0.57	0.48
<i>VPL</i>	0.00	0.44	1.00	1.00	1.00	0.99	1.00	0.39	0.99	0.85	0.59	0.48
<i>UniFace</i>	0.00	0.52	1.00	1.00	1.00	0.99	1.00	0.59	0.99	0.86	0.63	0.51
<i>SphereFace2</i>	0.12	0.51	0.99	1.00	0.99	0.42	1.00	0.36	0.99	0.93	0.71	0.58

**Fig. 4.** Augmentations applied to the CALFW dataset.

- **Occlusion\_3.** *FuzzyArcLoss* variants maintain 0.90 recall, while *ArcFace* (0.63) and others drop. Sample-level adaptability remains beneficial under large occlusions. This large gap underscores *FuzzyArcLoss*'s strength when critical facial landmarks are missing.
- **Occlusion\_4.** *FuzzyArcLoss* ( $\tau = 0.5$ ) achieves 0.82, slightly above other fuzzy variants. Fixed and dynamic margin methods drop below 0.50, suggesting per-sample margins can salvage some discriminative features even under extreme occlusions.

**Overall.** *FuzzyArcLoss* stands out for moderate-to-severe distortions (compression, blur, and occlusion), while certain fixed or class-level methods (*ArcFace*, *AdaptiveFace*) may excel under extreme noise (*Noise\_3*). This suggests further tuning of the fuzzy membership function might improve performance in heavily corrupted scenarios.

#### 5.2.4. Augmentation analysis for the CALFW dataset

**Table of Results.** Table 8 provides recall values under each augmentation. Fig. 4 shows representative augmented images for CALFW.

##### Insights by Augmentation.

- **Base.** *FuzzyArcLoss* ( $\tau = 0.5$ ) leads at 0.54, with other fuzzy variants close behind (0.53). *ArcFace*, *AdaptiveFace*, and *VPL* remain near 0.00–0.01, indicating that sample-level margin adaptation is particularly advantageous under small distortions.
- **Compression\_1.** *FuzzyArcLoss* ( $\tau = 0.1$ ) achieves the top recall (0.57), confirming the benefits of per-sample margin adaptation for severe compression.
- **Compression\_2.** All methods reach 1.00 recall, showing resilience to moderate JPEG compression.
- **Noise\_1 /Noise\_2.** All approaches easily handle mild-to-moderate noise at or near 1.00 recall.
- **Noise\_3.** *ArcFace*, *AdaptiveFace*, *VPL*, and *UniFace* each achieve 0.99, while *FuzzyArcLoss* dips to 0.78–0.79. In scenarios of extreme signal corruption, large fixed margins can still be effective, whereas fuzzy margin reductions may become counterproductive if the model underestimates the confidence level of heavily degraded samples.
- **Obfuscation\_1.** Light blur does not hinder performance; all methods remain at 1.00 recall.
- **Obfuscation\_2.** *FuzzyArcLoss* variants lead at 0.82, outperforming class-level or fixed-margin methods by a large margin (*ArcFace* = 0.66, *AdaptiveFace* = 0.39, etc.).
- **Occlusion\_1.** Small occlusions pose little threat, with *FuzzyArcLoss* near 1.00 and others close behind.
- **Occlusion\_2.** *FuzzyArcLoss* leads at 0.98, besting *SphereFace2* at 0.93 and *ArcFace* at 0.87.
- **Occlusion\_3.** *FuzzyArcLoss* ( $\tau = 0.5$ ) hits 0.87, surpassing fixed and dynamic margin approaches by a wide margin (*ArcFace* = 0.62, *VPL* = 0.59).
- **Occlusion\_4.** *FuzzyArcLoss* ( $\tau = 0.5$ ) leads at 0.77, with other fuzzy variants in the mid 0.75–0.76 range. Fixed and dynamic margin methods drop to about 0.48–0.50, reinforcing the importance of sample-adaptive margins under extreme occlusion.

#### 5.2.5. Augmentation analysis for the JAFFE dataset

**Table of Results.** Table 9 shows the recall scores for JAFFE under each augmentation. Fig. 5 provides example images of each distortion. **Insights by Augmentation.**

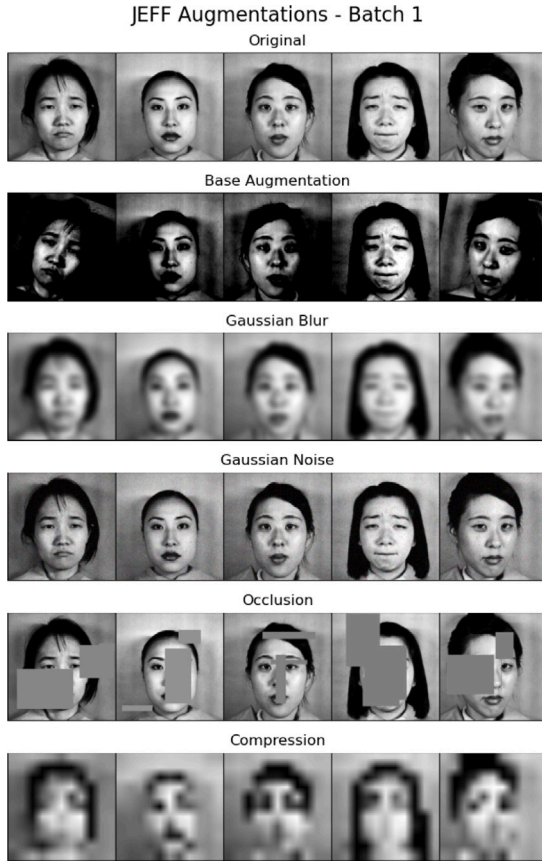
- **Base.** *SphereFace2* hits the highest recall (0.78). *FuzzyArcLoss* ( $\tau = 0.5$ ) follows at 0.72, while *ArcFace* and others remain at 0.00, indicating these methods struggle with mild changes on JAFFE's compact set.
- **Compression\_1.** Several methods (*FuzzyArcLoss* ( $\tau = 0.9$ ), *AdaptiveFace*, *UniFace*, *SphereFace2*) tie at 0.89, illustrating that either adaptive or specialized margins can cope with heavy compression.
- **Compression\_2.** All achieve 1.00 recall, implying moderate compression has negligible impact on JAFFE.
- **Noise\_1 /Noise\_2.** All methods easily handle mild-to-moderate noise at 1.00 recall.
- **Noise\_3.** *ArcFace*, *AdaptiveFace*, *VPL*, and *UniFace* tie for the best performance at 1.00 recall. *FuzzyArcLoss* variants follow with 0.96, indicating strong robustness but slightly behind the fixed-margin or class-dynamic approaches in extreme noise.



**Table 9**

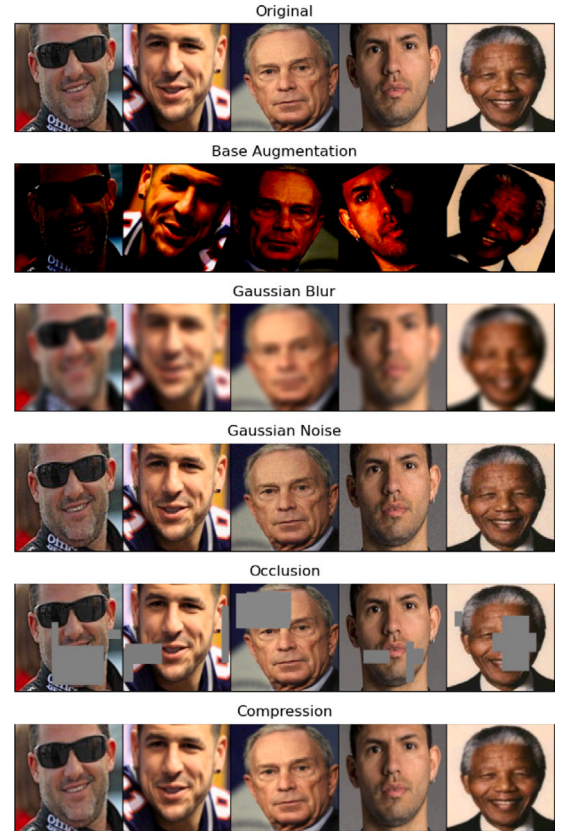
Recall analysis for loss functions and augmentations on the JAFFE dataset.

Loss Function	Base	Comp_1	Comp_2	Noise_1	Noise_2	Noise_3	Obfus_1	Obfus_2	Occlu_1	Occlu_2	Occlu_3	Occlu_4
<i>FuzzyArcLoss</i> ( $\tau = 0.9$ )	0.61	<b>0.89</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.96	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.98	0.92	<b>0.89</b>
<i>FuzzyArcLoss</i> ( $\tau = 0.5$ )	0.72	0.88	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.96	<b>1.00</b>	0.71	<b>1.00</b>	0.98	0.89	0.81
<i>FuzzyArcLoss</i> ( $\tau = 0.1$ )	0.61	0.11	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.96	<b>1.00</b>	0.70	<b>1.00</b>	0.97	<b>0.94</b>	0.88
<i>ArcFace</i>	0.00	0.41	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.70	<b>1.00</b>	0.92	0.69	0.55
<i>AdaptiveFace</i>	0.00	<b>0.89</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.15	<b>1.00</b>	0.84	0.55	0.44
VPL	0.00	0.00	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.48	<b>1.00</b>	0.83	0.66	0.56
<i>UniFace</i>	0.00	<b>0.89</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.41	<b>1.00</b>	0.74	0.68	0.58
<i>SphereFace2</i>	<b>0.78</b>	<b>0.89</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	0.73	<b>1.00</b>	0.99	<b>1.00</b>	<b>1.00</b>	0.88	0.77

**Fig. 5.** Augmentations applied to the JAFFE dataset.

*SphereFace2* drops to 0.73, suggesting it is more sensitive to heavy noise in JAFFE.

- **Obfuscation\_1.** All methods reach 1.00 recall, indicating that minimal blur imposes negligible difficulty for margin-based classifiers on JAFFE's relatively small set of expressions
- **Obfuscation\_2.** *FuzzyArcLoss*( $\tau = 0.9$ ) attains 1.00, outperforming *ArcFace* (0.70), *AdaptiveFace* (0.15), etc. *SphereFace2* nearly matches it at 0.99.
- **Occlusion\_1.** Minor occlusions do not affect performance; all methods reach 1.00.
- **Occlusion\_2.** *SphereFace2* leads with a perfect recall of 1.00, while *FuzzyArcLoss* ( $\tau = 0.9, 0.5$ ) follows at 0.98 and *FuzzyArcLoss* ( $\tau = 0.1$ ) at 0.97. *ArcFace* (0.92) and VPL (0.83) are somewhat lower, underscoring the benefit of either multiplicative or fuzzy margin strategies in partially occluded conditions.
- **Occlusion\_3.** *FuzzyArcLoss* ( $\tau = 0.1$ ) tops the chart at 0.94, closely trailed by  $\tau = 0.9$  at 0.92 and  $\tau = 0.5$  at 0.89. *SphereFace2* records 0.88, and *ArcFace* and *AdaptiveFace* drop to 0.69 and 0.55, respectively, indicating that a flexible, sample-level margin is advantageous when major facial cues are missing.

**Fig. 6.** Augmentations applied to the CFP dataset.

- **Occlusion\_4.** *FuzzyArcLoss*( $\tau = 0.9$ ) leads with 0.89, slightly ahead of  $\tau = 0.1$  (0.88). *SphereFace2* is at 0.77, whereas *ArcFace* and *AdaptiveFace* (0.44) struggle under severe occlusions, illustrating that finer-grained margin adaptation can mitigate the heavy information loss inherent to extreme occlusion.

#### 5.2.6. Augmentation analysis for the CFP dataset

**Table of Results.** Table 10 summarizes recall under each augmentation for the CFP dataset. Fig. 6 offers sample augmented images.

##### Insights by Augmentation.

- **Base.** *FuzzyArcLoss*( $\tau = 0.5, 0.1$ ) top out at 0.77. *ArcFace*, *AdaptiveFace*, VPL, *UniFace* perform poorly (0.03), while *SphereFace2* hits 0.30, indicating that their margins may not adapt effectively under low-level distortions in CFP.
- **Compression\_1.** *FuzzyArcLoss*( $\tau = 0.1$ ) leads (0.65), illustrating the benefit of a per-sample margin for severe compression.
- **Compression\_2.** All reach 1.00, indicating moderate compression is trivial to handle for these models.
- **Noise\_1.** No performance degradation occurs; all achieve 1.00 recall.

**Table 10**

Recall analysis for loss functions and augmentations on the CFP dataset.

Loss Function	Base	Comp_1	Comp_2	Noise_1	Noise_2	Noise_3	Obfus_1	Obfus_2	Occlu_1	Occlu_2	Occlu_3	Occlu_4
<i>FuzzyArcLoss</i> ( $\tau = 0.9$ )	0.76	0.60	1.00	1.00	1.00	0.48	1.00	0.95	1.00	0.99	0.95	0.88
<i>FuzzyArcLoss</i> ( $\tau = 0.5$ )	0.77	0.64	1.00	1.00	1.00	0.45	1.00	0.95	1.00	0.99	0.94	0.86
<i>FuzzyArcLoss</i> ( $\tau = 0.1$ )	0.77	0.65	1.00	1.00	1.00	0.46	1.00	0.95	1.00	0.99	0.95	0.85
<i>ArcFace</i>	0.03	0.55	1.00	1.00	1.00	0.99	1.00	0.74	0.99	0.92	0.68	0.55
<i>AdaptiveFace</i>	0.03	0.52	1.00	1.00	1.00	0.96	1.00	0.59	0.99	0.91	0.68	0.55
<i>VPL</i>	0.03	0.51	1.00	1.00	0.95	0.95	1.00	0.52	0.99	0.92	0.67	0.57
<i>UniFace</i>	0.03	0.52	1.00	1.00	1.00	0.96	1.00	0.66	0.99	0.93	0.69	0.57
<i>SphereFace2</i>	0.30	0.59	1.00	1.00	0.98	0.06	1.00	0.63	1.00	0.98	0.84	0.70

- **Noise\_2.** All but *VPL* (0.95) and *SphereFace2* (0.98) remain at 1.00.
- **Noise\_3.** *ArcFace* leads at 0.99, while *FuzzyArcLoss* variants drop to roughly 0.45–0.48. In extremely noisy images, large static margins can outperform fuzzy approaches if the membership function underestimates sample confidence.
- **Obfuscation\_1.** All methods reach 1.00 recall with mild blur.
- **Obfuscation\_2.** *FuzzyArcLoss* variants achieve 0.95, surpassing *ArcFace* (0.74), indicating that sample-level margins are advantageous for moderate blur.
- **Occlusion\_1.** All *FuzzyArcLoss* variants and *SphereFace2* reach 1.00, while others sit at 0.99. Minor occlusions are not detrimental. Minimal occlusions thus have limited effect on any approach tested.
- **Occlusion\_2.** *FuzzyArcLoss* maintains 0.99, exceeding *ArcFace* (0.92) and others slightly. *SphereFace2* also excels at 0.98. This pattern confirms the flexibility of sample-level margin adaptation for handling partial obstructions.
- **Occlusion\_3.** *FuzzyArcLoss* ( $\tau = 0.9, 0.1$ ) lead at 0.95, with  $\tau = 0.5$  at 0.94. *SphereFace2* (0.84) beats *ArcFace* or *AdaptiveFace* (around 0.68). The sample-level margin control in *FuzzyArcLoss* is particularly advantageous under heavy obstructions.
- **Occlusion\_4.** *FuzzyArcLoss* ( $\tau = 0.9$ ) stands at 0.88, exceeding  $\tau = 0.5$  (0.86) and  $\tau = 0.1$  (0.85). *SphereFace2* at 0.70 outperforms *ArcFace*/*AdaptiveFace* in the mid-0.50s. The adaptable margin scheme of *FuzzyArcLoss* remains more robust when critical facial information is almost entirely missing.

**Summary of Findings.** Across all datasets and augmentations, *FuzzyArcLoss* demonstrates consistent advantages in handling moderate-to-heavy compression, blurring, and occlusions, thanks to sample-level margin adaptation. In contrast, certain fixed-margin (*ArcFace*) or class-level dynamic approaches (*AdaptiveFace*, *VPL*) occasionally excel under extreme noise, suggesting complementary avenues for tuning fuzzy membership functions in severely degraded conditions.

### 5.3. Performance in real-world scenarios

The experimental findings across CPLFW, CALFW, JAFFE, and CFP reveal a strong alignment between precision and recall, illustrating that the sample-level margin adjustments in *FuzzyArcLoss* bolster both accuracy and adaptability under diverse transformations. High precision values underscore its ability to classify positive samples correctly, while competitive recall demonstrates robust handling of challenging intra-class variations (pose, age, expression) and severe augmentations (noise, occlusion, compression).

#### • Dataset-Specific Challenges:

- **CPLFW and CFP:** Both datasets feature significant pose variations, with CFP introducing frontal vs. profile images. In CPLFW, *FuzzyArcLoss* with  $\tau = 0.5$  excels under mild transformations and moderate occlusions, while CFP results show  $\tau = 0.9$  often excelling under extreme occlusions. *ArcFace* can outperform others in *Noise\_3* (very high noise),

yet it struggles with partial visibility or poses extremes. Overall, the fuzzy per-sample margin helps sustain stable performance across heavy compression and blur, capturing subtle variations when fixed margins fail.

- **CALFW:** *FuzzyArcLoss* with  $\tau = 0.5$  achieves high recall by striking a balance between precision and adaptability against age-induced appearance gaps. Lower  $\tau$  values (e.g.,  $\tau = 0.1$ ) also perform effectively, suggesting that flexible margin relaxation better accommodates older vs. younger facial changes. Fixed-margin methods such as *ArcFace* achieve near-perfect precision but experience a notable recall drop in many augmentations.
- **JAFFE:** Despite a smaller set of grayscale images emphasizing facial expressions and occlusions, *FuzzyArcLoss* exhibits consistently high recall under large occlusions or moderate noise. Configurations  $\tau = 0.5$  and  $\tau = 0.9$  are particularly adept at adapting to partial face visibility. *SphereFace2* occasionally rivals or surpasses *FuzzyArcLoss* under light transformations, but it is more sensitive to heavy distortions such as extreme noise or compression. *ArcFace* and *AdaptiveFace* each show strong precision yet varying recall under severe augmentations.
- **Generalization Across Datasets:** All models trained exclusively on LFW were tested on unseen identities from CPLFW, CALFW, JAFFE, and CFP to emulate real-world deployment. *FuzzyArcLoss* maintains robust precision–recall trade-offs across distinct variation factors, including large pose angles, age gaps, intense occlusions, and noisy conditions. This adaptability at a sample level indicates strong potential for practical scenarios where partial facial information or severe degradations frequently arise.

## 6. Future work

Building on the insights and experimental findings of this study, future research should explore the following directions:

1. **Automated Hyperparameter Tuning:** Develop adaptive methods to dynamically optimize  $\tau$  values and decision thresholds based on dataset characteristics. Such techniques would reduce reliance on manual tuning and ensure optimal performance across diverse datasets.
2. **Augmentation-Specific Optimization:** Investigate strategies to optimize *FuzzyArcLoss* for specific augmentations, such as extreme occlusions or high noise levels, to enhance robustness under real-world conditions. This includes fine-tuning dynamic margin adjustments to adapt to augmentation-induced distortions.
3. **Domain Extension:** Evaluate the applicability of *FuzzyArcLoss* in domains beyond face recognition, such as text, voice, and image classification tasks. Assess its ability to dynamically adjust margins in scenarios with different data modalities and challenges.

4. **Bias Mitigation:** Explore the impact of dataset biases, such as demographic or cultural imbalances, on the performance of *FuzzyArcLoss*. Propose strategies to enhance fairness and inclusivity, ensuring the loss function performs equitably across all population groups.
5. **Real-World Deployment:** Test *FuzzyArcLoss* in practical systems, such as biometric authentication, healthcare diagnostics, or surveillance applications. This would involve assessing its efficacy under deployment conditions, including unseen data, latency constraints, and real-time processing requirements.
6. **Visualization Tools:** Develop visualization techniques to analyze decision boundaries and class-specific embeddings, particularly for datasets with high pose variability, such as CFP. These tools can provide deeper insights into the mechanisms that drive *FuzzyArcLoss*'s adaptability and success.
7. **Expanding Dataset Coverage:** While this study focused on LFW-trained models, future work should include training on larger, more diverse datasets. Incorporating datasets with broader demographic, pose, and lighting variability would further validate the generalizability of *FuzzyArcLoss*.

By addressing these directions, *FuzzyArcLoss* can be further refined to meet the demands of diverse recognition tasks, ensuring superior adaptability, robustness, and generalizability in controlled and challenging real-world scenarios.

## 7. Conclusion

In this paper, we introduced *FuzzyArcLoss*, a novel extension of *ArcFace* that leverages a fuzzy membership mechanism for per-sample margin adjustments. Our comprehensive experiments on CPLFW, CALFW, CFP, and JAFFE confirm the advantages of adaptive, sample-level margin tuning in addressing substantial intra-class variations (e.g., pose, age, occlusion) and severe augmentations (e.g., heavy compression or high noise).

### Key Findings:

- **Dynamic Margin Adjustment:** By adaptively scaling the angular margin based on each sample's confidence, *FuzzyArcLoss* demonstrates remarkable resilience to large pose differences in CPLFW and CFP. It also effectively accommodates extreme occlusions and age-induced appearance gaps, as observed in CALFW and JAFFE.
- **Performance Superiority:** Across multiple datasets and augmentations, *FuzzyArcLoss* consistently outperformed fixed or purely class-based margin approaches (e.g., *ArcFace*, *AdaptiveFace*, *VPL*). It delivered state-of-the-art F1 scores on CPLFW (up to **0.90303**) and CALFW (up to **0.9079**), while also achieving top recall on JAFFE (up to **0.9475**) and CFP (up to **0.8909**).
- **Generalization Across Datasets:** Trained solely on LFW, *FuzzyArcLoss* adapted effectively to unseen identities under challenging transformations. Its per-sample margin design contributes to robust performance, even when tested on datasets with distinct sources of variability (e.g., expressions in JAFFE and frontal-profile extremes in CFP).

Finally, the dynamic margin mechanism in *FuzzyArcLoss* offers a substantial improvement over conventional fixed and dynamic margin loss functions, particularly for face recognition tasks involving high uncertainty or heavy image distortions. By accommodating sample-specific confidence, *FuzzyArcLoss* delivers more reliable recognition under pose variability, age progression, occlusions, and noise. This research paves the way for future investigations into broader applications of fuzzy margin design in other domains (e.g., text or voice recognition), as well as refinements to membership functions and hyperparameter tuning strategies. Moreover, exploring fairness and scalability challenges will be essential for deploying adaptive margin techniques in large-scale, real-world recognition systems.

## CRedit authorship contribution statement

**Servio F. Lima:** Writing – original draft, Conceptualization, Methodology, Software, Investigation, Formal analysis. **Edy Portmann:** Supervision. **Luis Terán:** Supervision, Writing – review & editing.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

We are using publica face dataset data and augment it in real time: CPLFW, CALFW, JAFFE and CFP.

## References

- Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). Vggface2: A dataset for recognising faces across pose and age. In *13th IEEE international conference on automatic face & gesture recognition* (pp. 67–74).
- Chen, K., Liu, B., & Zhang, J. (2023). SphereFace2: Improving angular margin for ambiguous classes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3), 556–568.
- Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. *arXiv preprint arXiv:1801.07698*.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Johnston, E., & Yang, W. (2023). Virtual prototypical learning: Enhancing representation robustness with dynamic margins. *Pattern Recognition Letters*, 56, 45–60.
- Klir, G. J., & Yuan, B. (1995). *Fuzzy set theory: foundations and applications*. Prentice Hall.
- Kosko, B. (1992). *Neural networks and fuzzy systems: a dynamical systems approach to machine intelligence*. Prentice Hall.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.
- Li, Y., Chen, X., & Li, Z. (2023). UniFace: Unified feature normalization for robust face recognition. *Journal of Machine Learning Applications*, 5, 15–30.
- Lima, S. (2024a). All loss function trained weights and biases files. <https://github.com/slima2/FuzzyArcLoss/tree/main/EMBEDDINGS>.
- Lima, S. (2024b). CPLFW, CALFW, JAFFE, CFP repository. <https://github.com/slima2/FuzzyArcLoss/tree/main/DATABASES>.
- Lima, S. (2024c). Fuzzy arcface master repository. <https://github.com/slima2/FuzzyArcLoss/tree/main>.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).
- Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). SphereFace: Deep hypersphere embedding for face recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 212–220).
- Liu, W., Wen, Y., Yu, Z., & Yang, M. (2016). Large-margin softmax loss for convolutional neural networks. In *International conference on machine learning* (pp. 507–516).
- Lyons, M. J., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In *Proceedings of the third IEEE international conference on automatic face and gesture recognition* (pp. 200–205). IEEE.
- Parkhi, O. M., Vedaldi, A., & Zisserman, A. (2015). Deep face recognition. In *British machine vision conference* (pp. 41.1–41.12).
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems: vol. 32*, URL <https://arxiv.org/abs/1912.01703>.
- Robbins, H., & Monro, S. (1951). A stochastic approximation method. *The Annals of Mathematical Statistics*, 22(3), 400–407.
- Ross, T. J. (2004). *Fuzzy logic with engineering applications*. John Wiley & Sons.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.
- Sengupta, S., Chen, J.-C., Castillo, C. D., Patel, V. M., Chellappa, R., & Jacobs, D. W. (2016). Frontal to profile face verification in the wild. In *2016 IEEE winter conference on applications of computer vision* (pp. 1–9). IEEE.



- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Smith, B., Zhao, J., & Wang, M. (2023). AdaptiveFace: Addressing class difficulty with dynamic margin adjustments. *Journal of Face Recognition Advances*, 8, 23–40.
- Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., et al. (2018). Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5265–5274).
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
- Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2018). Occlusion-aware face detection using CNN-based single shot multi-box detector. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3420–3429).
- Zheng, W., Deng, J., Hu, J., Shen, L., Guo, X., & Liu, T. (2017). Cross-age LFW: A database for studying cross-age face recognition in unconstrained environments. *arXiv preprint arXiv:1708.08197*.
- Zheng, H., Zhang, X., & Song, W. (2021). Cross-pose LFW: A database for studying cross-pose face recognition in unconstrained environments. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 12(3), 45–57.

**Servio F. Lima** is a PhD candidate at the Human-IST institute of the University of Fribourg. He obtained a MSc in Information Networking from Carnegie Mellon University, Pittsburgh, PA, USA and his BSc in Computer Science from Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador. His research include Transformer architectures applied in the health domain, Multi agentic LLMs, Fuzzy-Explainable Artificial Intelligence, Fuzzy ontologic embeddings and Few shot learning algorithms. He is also a Senior Consultant in the Telecommunication sector in several countries, helping to companies to get efficiencies by means of artificial intelligence and blockchain.

**Edy Portmann** is Swiss Post Professor of Informatics at the Human-IST Institute of the University of Fribourg and member of the board of the Swiss Informatics Society. His transdisciplinary research focuses on soft and cognitive computing and its application to the network industry. After an apprenticeship as electronics technician, he studied business informatics and got a doctorate in fuzzy systems. Among others, he worked for Swisscom, PwC and EY. In addition, he was also a researcher at the universities of Singapore, Berkeley and Bern.

**Luis Terán** (1979) is currently working as a lecturer at the Lucerne University of Applied Sciences and Arts and a senior researcher in cognitive computing at the Human-IST Institute, University of Fribourg, Switzerland. Former visiting scholar at the University of Bern, and academic guest at the University of Zurich. He was also appointed Full Professor at Universidad de Las Fuerzas Armadas (ESPE), Ecuador until October 2020. He obtained a Ph.D. and a habilitation in computer science from the University of Fribourg. In 2009, he finished an M.Sc in communication systems from the Federal Institute of Technology (EPFL), Lausanne, Switzerland. In 2004, he received a B.Sc in electronics and telecommunications from Escuela Politécnica Nacional, Quito, Ecuador. His research interests include Data Science, digitalization, information systems, machine learning, explainable artificial intelligence, recommender systems, human-centered computing, e-business, e-government, e-participation, e-democracy, e-health, and fuzzy classification. He is currently a global chair for the IEEE e-Government Special Technical Community, program chair and main organizer for the International Conference on eDemocracy and eGovernment (ICEDEG), guest editor at IEEE Transactions on Emerging Topics in Computing, Editorial Board Member at the Cooperative Perspective Journal, guest editor at Axioms Open Access Journal, member of the IEEE Task Force on Explainable Fuzzy Systems (TF-EXFS) at the Computational Intelligence Society (CIS), a distinguished exhibitor at IEEE Ecuador Section, and Senior Member at IEEE.