

AXIAL SPHERE LOSS: Encouraging open-space risk minimization in face identification tasks

Rafael Henrique Vareto^{*◇} and William Robson Schwartz^{*}

^{*}Universidade Federal de Minas Gerais (UFMG) | [◇]Samsung R&D Institute Brazil (SRBR)

{rafaelvareto, william}@dcc.ufmg.br

Abstract—Open-set face recognition challenges biometric systems by requiring them to identify registered subjects while rejecting unregistered individuals. This task is particularly demanding in watchlist scenarios, where biometric systems must focus on subjects of interest and disregard irrelevant faces. To address real-world face applications, this study associates quickly trainable adaptation networks with a logit-and-distance-based cost function that explores non-gallery samples in favor of minimizing the open-space risk. These background instances are either specified in dataset protocols or synthetically built at training time. The proposed Axial Sphere Loss (ASL) shifts each class into pre-defined regions in the latent space and mutually pushes non-gallery samples toward the space origin, forming spherical containers around each class template at inference time. We show that training an adapter network with ASL does not hinder closed-set recognition scores but significantly boosts open-set identification rates, achieving state-of-the-art performance on three well-known face benchmarks, namely, LFW, IJB-C, and UCFS datasets.

I. INTRODUCTION

Smart face recognition has become a prominent research area in recent decades. Nowadays, face identification systems are not only prevalent in social media and photo tagging applications but also serve as a crucial tool for governments, law enforcement, and private companies [8, 13]. Despite the ongoing progress, a serious limitation remains: the inability to handle images of never-seen individuals, a typical scenario in surveillance or security environments. Essentially, research addressing such a level of uncertainty, in which a biometric system may encounter entirely new distributions or data points not belonging to any known class, has received comparatively less attention [38].

Open-set face recognition characterizes the aforementioned setting where anonymous individuals, unseen during training and enrollment stages, only come into sight during evaluation time [1, 16, 41]. For a clear illustration, an open-set face application is expected to dismiss law-abiding passengers at airports, but warn the security personnel whenever law offenders turn up. Such biometric systems may also raise alerts when unauthorized people attempt to access an organization’s premises. Unfortunately, recent newspaper articles have shown that people being misidentified is not a hypothetical exercise but has actually occurred several times across the United States [19, 35]. To make matters worse, false alarms should be avoided by any means since a system identification error may bias the security approach and mistakenly hold up innocent people in custody [16].

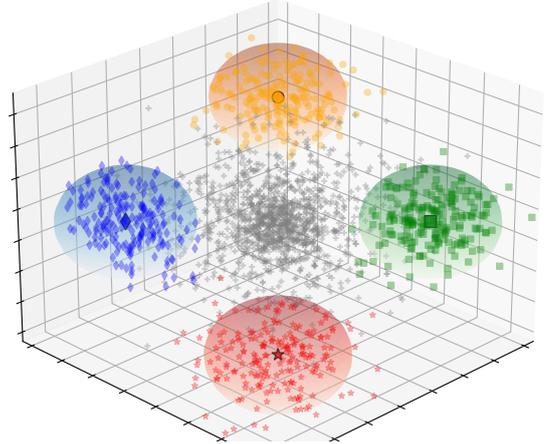


Fig. 1: OPEN-SET CLASSIFICATION WITH AXIAL SPHERE LOSS. Background instances (grey crosses +) are pushed to the space origin whereas gallery-enrolled classes, represented by a set of symbols ■ ★ ◆ ● are shifted to specific regions alongside each axis. Samples outside their corresponding spherical container hold low acceptance scores and, consequently, should be regarded as unknown categories.

The open-space risk is a concept concerning the chance of misclassifying a sample as belonging to a known class when it literally pertains to an unknown category [38]. For most machine learning algorithms learn patterns from finite data, they implicitly assume that probe samples are drawn from the same underlying data distribution. These methods often struggle to delineate clear decision boundaries between unknown and known classes, acting like nearest-neighbor classifiers and leading to further false-positive occurrences. The unreliable results arise from the lack of any reference point in the training set to support the classification of probe data originating from different distributions [39].

The majority of open-set studies evaluate approaches on non-face datasets containing a reduced number of classes but holding an abundant volume of samples per category [9, 22, 30]. These datasets usually contain training and test samples of known classes drawn from the same underlying data distribution [44]. Methods evaluated in such conditions are not hampered by the shortage of samples available for training and, in fact, provide better generalization capabilities since they effectively preserve the inherent data statistics [6]. Then, we believe that the good results attained by many methods in well-supplied datasets would not always extend to more demanding open-set tasks [11, 16].

Quick install: use `pip install openloss` for reproducibility.

The goal of this paper is to introduce a technique that can be successfully deployed to realistic biometric environments, like face recognition tasks. In fact, we propose the AXIAL SPHERE LOSS (ASL), a novel cost function that induces artificial neural networks to push each class of interest, that is, the gallery set, to fixed regions along each axis and also encourages non-gallery training samples to lie close to the logit space origin. More precisely, ASL has been designed to benefit from background samples, training instances meant to resemble unknown categories, that may provide information about decision boundaries between known classes and uncharted data. Seeing that most datasets do not contain readily available background samples, we find synthesizing new data within the latent space during training a feasible alternative. In consequence, ASL can take advantage of diversified data and assist neural networks in refining decision boundaries, identifying potential outliers, reducing false positives, and improving the overall accuracy of the model.

Figure 1 illustrates how ASL builds a sphere-like probabilistic container around each class template (centroid). At inference time, the closer a probe sample gets to a sphere center, the greater its chance of being contained in that class. The farther it lies from all gallery templates, the lower the prospects of being a class of interest, increasing its chance of rejection. In summary, the acceptance/rejection scheme provides a mechanism to quantify the open-space risk for each probe sample, assuming that known samples are close to their corresponding class centroid but unknown samples lie somewhere else in the open-space.

This work consists of neural adapter networks acting as a replacement of fully-connected layers from pre-trained deep architectures in pursuance of agile gallery specialization [43]. We conduct the analysis with three pretrained RESNET-based backbones [5, 10, 23], which generate feature embeddings that serve as input to the adapter network. The main experiments are carried out on three face benchmarks, namely LFW, IJB-C and UCCS [17, 20, 27]. We utilize LFW to find the best hyperparameters and then apply the very same configuration when evaluating other face datasets. Such criterion confirms the robustness of our proposed cost function across different domains. The investigation demonstrates the superiority of ASL, achieving superior performance on different open-set protocols, especially when combined with either genuine or synthesized background samples.

The major contributions of our work are:

- 1) ASL, a cost function that either pushes known classes further apart onto fixed reference points (coordinate axes) or drives non-gallery samples toward the origin.
- 2) A new inference metric designed to improve decision-making by more effectively distinguishing between known and unknown samples during inference.
- 3) An optimized method for interpolating hidden-layer feature maps to generate new training samples in the absence of non-gallery classes.
- 4) An analysis of how ASL distinguishes known from unknown identities, attenuates the open-space risk and achieves state-of-the-art results.

II. RELATED WORK

The majority of modern face recognition applications rely upon deep neural networks (DNNs) [10, 21, 28, 36]. DNNs used in biometric tasks are typically trained on large datasets of public figures before being applied to specific face populations [13]. This condition makes the identification task inherently domain-adaptive, as none of the individuals available during the network pre-training are included in the gallery set, which encompasses only the subjects of interest, also referred to as the watchlist.

In the past decade, several works have detailed strategies that adjust pretrained networks to watchlists. Some explore transfer learning techniques or consist of traditional machine learning algorithms fitted on deep feature representations [3, 15, 31]. Hashing functions have also been used to solve open-set face recognition tasks [12, 41, 46]. Other approaches rely on clustering methods that act as a filtering barrier to unknown samples [18, 40]. Some researchers have attempted to optimize the speed-accuracy trade-off of DNNs by designing lighter architectures [16, 26, 50]. Despite all contributions, the aforementioned methods present unbounded open-space risk [38] and are not very well suited for rejecting unknown individuals as generally required in the watchlist context.

Vaze *et al.* [44] argued that achieving high accuracy in closed-set classification is strongly correlated with open-set performance. The findings suggest that thresholding techniques applied to the maximum logit score of a strong closed-set classifier can achieve competitive open-set results. However, they only considered non-face datasets containing at least 500 instances per class¹ and, as a result, there is little guarantee the obtained results would propagate to datasets holding scarce samples per subject, as in biometric applications. For face recognition, researchers have chased “tailor-made” approaches to better handle unknown classes. In fact, many of them have been looking for cost functions that are able to model the differences between known and unknown data distributions during the training process [14, 49], as reported in the subsequent paragraphs.

Liu *et al.* [25] proposed *SphereFace Loss* (SPH), a cost function designed to support neural networks in learning discriminative embeddings by imposing angular constraints to the decision boundary on a hypersphere manifold. Later, Wang *et al.* [48] disclosed *CosFace Loss* (COS), a loss function that adds penalties to the angular space, separating feature representations of different identities but maintaining intra-class compactness. Similarly, Deng *et al.* [10] introduced *ArcFace Loss* (ARC), another cost function that incorporates angular margins into the decision boundary in exchange for larger gaps between embeddings of different identities. Even though these angular-based functions are intuitive, they are sensitive to their hyper-parameters and, eventually, prone to overfitting. They assume embeddings are angularly separable, which may not align with other domains where data distributions lack a natural angular structure, such as geological formations and medical imaging.

¹Datasets: MNIST, SVHN, CIFAR10, CIFAR+50 and TinyIMAGENET.

The vast majority of open-set approaches fail to quantify how confident models are about their predictions. In contrast, Miller *et al.* [29] introduced *Class Anchor Clustering* (CAC), a loss function that encourages data points from the same class to cluster tightly around fixed “anchors”. CAC provides a mechanism for computing a rejection score considering the distance between a probe instance and all class centers. As a limitation, CAC only explores samples from gallery-enrolled identities, neglecting any information that could enhance a model’s ability to discriminate known from unknown classes.

Dhamija *et al.* [11] coined “agnostophobia” the fact that neural networks are generally overconfident when dealing with instances from unknown classes. The authors came up with *Entropic Open-Set* and *ObjectoSphere* (OBS) losses. The former maximizes the uncertainty for background samples by inducing the network responses to lie uniformly distributed whereas the latter also modifies the magnitude of training instances based on a gallery/non-gallery principle. Both methods were later revisited by Günther *et al.* [16] as they coupled both losses with adaptation layers to quickly readjust pre-trained deep networks to specific face watchlists. More recently, Vareto *et al.* [42] disclosed *Maximal Entropy Loss* (MAX), a function that adds penalty margins to known identities and increases the entropy for background samples. Even though the described methods require training data that includes non-gallery samples from the same underlying distribution of the gallery set, they do not provide probe samples with indicative values of open-space risk minimization.

In the interest of modifying data in the latent feature space, Verma *et al.* [45] came up with an interpolation strategy to generate new feature representations. Li *et al.* [24] proposed a stochastic feature augmentation procedure to perturb Gaussian noise embeddings, whereas Volpi *et al.* [47] employed adversarial samples as background identities. Even though these works haven’t been evaluated on face benchmarks, they may provide meaningful insights into how additional background samples can be incorporated into the training stage for strong domain generalization.

Labeled Faces in the Wild (LFW) [20] consists of 13,233 images unevenly distributed among nearly six thousand identities. As LFW was originally designed for verification, researchers have proposed protocols in an attempt to leverage open-set identification tasks [15, 26]. IJB-C [27] contains two disjoint gallery partitions with known subjects: *Gallery A* contains 1772 identities alongside 5832 samples whereas *Gallery B* holds 1759 individuals and a total of 6024 instances, an average of 3.35 samples per subject. The UCCS dataset [37] consists of surveillance images taken from more than 1,700 identities collected over a period of 19 months. UCCS’s gallery set includes 1,085 subjects of interest holding nearly 20 instances per class as well as a myriad of samples belonging to unknown identities. Despite LFW, IJB-C, and UCCS being well-established datasets, their open-set protocols highlight three persistent challenges in real-world biometrics: the large number of classes, the limited number of samples per category, and the inherent complexities of working with unconstrained data environments.

III. PROPOSED APPROACH

AXIAL SPHERE LOSS (ASL) is a cost function designed to enhance open-set classification by leveraging both geometric and probabilistic perspectives in the latent feature space. ASL incorporates three key components: intra-class compactness, inter-class separation, and magnitude regularization. More precisely, it impels neural networks to minimize the intra-class spread and maximize the inter-class distance by handling not only gallery-enrolled identities but also maintaining background training samples² with contained magnitude. In addition, the proposed approach is capable of handling both known and unknown identities by synthesizing representative background samples when unavailable, ensuring robust generalization in challenging biometric scenarios.

A. Feature Extraction

Typical neural network pipelines can be characterized as

$$\hat{y} = f(x) = C_{\psi} \circ F_{\Theta}(x) = C_{\psi} \circ F_{\Theta^L} \circ \dots \circ F_{\Theta^1}(x)$$

where $F_{\Theta}(x)$ contains L convolutional layers connected to an eventual classification layer C_{ψ} . Given a face image x , the feature extraction and classification modules can be represented, respectively, as two modular parts: $z = F_{\Theta}(x)$ and $\hat{y} = C_{\psi}(z)$. In our pipeline, $F_{\Theta}(x)$ consists of RESNET architectures used for getting discriminative representations whereas C_{ψ} comprises a compact adaptation network in charge of watchlist/gallery specialization (see Section IV-4).

B. Background training embeddings

Not having any indicative attribute of unknown sample distribution is one of the critical drawbacks in deploying open-set systems. Therefore, we incorporate background face data into the training stage as we hypothesize that non-gallery samples may equip neural networks with tighter class boundaries and enhanced generalization power. Since background samples can boost classification performance when sharing equivalent statistics with the gallery set [43], we synthesize new data using samples from gallery-enrolled subjects, according to the following equations:

$$\bar{z} = \lambda \cdot z_i + (1 - \lambda) \cdot z_j \quad (1a)$$

$$\begin{aligned} \text{st.}: z_j &= \arg \max_{(z_{i'}, g_{i'}) \in G} \cos(z_i, z_{i'}) \\ g_i &\neq g_j \wedge 0 \leq \lambda \leq 1 \end{aligned} \quad (1b)$$

Equation (1a) defines the feature augmentation technique that takes place in the latent space and \bar{z} is the fabricated embedding. Observe that \bar{z} is expressed as a linear combination of feature vectors z_i and z_j coming from distinct identities g_i and g_j of the gallery set G , respectively. The first constraint in Equation (1b) guarantees that z_j is selected in favor of maximizing the cosine similarity with z_i among all feature vectors $z_{i'}$ available in set G . The second constraint states that z_i and z_j come from different classes whereas the weight factor λ sets the contribution of each vector.

²Background/non-gallery training data consist of face instances that do not correspond to any identity registered in the gallery of known subjects.

Each new representation receives a $\bar{y} = -1$ label, ensuring that the tuple (\bar{z}, \bar{y}) contains information about two different subjects of interest. Adding supplementary instances makes the training process more challenging, while simultaneously enhancing the model’s ability to distinguish between classes due to the increased diversity in the training data.

C. Axial Sphere Loss

The AXIAL SPHERE LOSS (ASL) is an open-set cost function that handles training samples in three distinct ways: (i) ASL induces neural networks to push each class of interest (gallery set) to fixed regions along axes in the logit space; (ii) it also pushes non-gallery training samples to lie close to the logit space origin; in addition, (iii) the proposed loss focuses on enhancing the intra-class compactness as well as inter-class separability among known subjects by penalizing target classes during training time.

1) *Fixed class centers*: Class centroids correspond to multi-dimensional points or vectors that summarize the characteristics of classes in a dataset. In the proposed method, however, we establish them as reference basis vectors that represent each category in the latent space. They behave as some sort of static anchoring and play a crucial role in guiding neural networks when defining decision boundaries and improving their ability to distinguish between classes. Equation (2a) and (2b) specify a fixed point \mathbf{p}_g for each training class g registered in the gallery set G , as below:

$$\mathcal{P} = (\alpha \cdot \mathbf{p}_1, \alpha \cdot \mathbf{p}_2, \dots, \alpha \cdot \mathbf{p}_{|G|-1}, \alpha \cdot \mathbf{p}_{|G|}) \quad (2a)$$

$$\begin{aligned} \mathbf{p}_1 &= (1, 0, \dots, 0, 0), \mathbf{p}_2 = (0, 1, \dots, 0, 0), \\ \mathbf{p}_{|G|-1} &= (0, 0, \dots, 1, 0), \mathbf{p}_{|G|} = (0, 0, \dots, 0, 1) \end{aligned} \quad (2b)$$

Essentially, \mathbf{p}_g can be pictured as $|G|$ -dimensional points, coordinate axes or one-hot encodings where all indices are nil except for a single element filled with 1. Each reference vector \mathbf{p}_g is multiplied by a scaling factor α that modifies the magnitude while maintaining their direction and orthogonality. The fixed class centres are mutually orthogonal in such a way that $\mathbf{p}_a \cdot \mathbf{p}_b = 0$ for $a \neq b$. Such nature results in an equal separation among all classes as the Euclidean distance between any pair of one-hot vectors is $\alpha\sqrt{2}$, as follows:

$$d(\mathbf{p}_a, \mathbf{p}_b) = \sqrt{\sum_{i=0}^{|G|} (\alpha \mathbf{p}_{a_{[i]}} - \alpha \mathbf{p}_{b_{[i]}})^2} = \sqrt{\alpha^2 + \alpha^2} = \alpha\sqrt{2}$$

Unlike the traditional use of one-hot target vectors and posterior probability scores in categorical classification tasks, our approach assumes that each reference point represents a fixed coordinate/point in the Euclidean space. This shift in perspective implicitly encourages neural networks to treat logit scores geometrically, that is, the relationship between feature vectors and class representations is measured through spatial distances instead of probabilistic distributions. As a result, the adoption of fixed class centers promotes a more organized and interpretable structure within the feature space, improving the model’s capacity to represent class-specific information in a spatially meaningful manner.

2) *Intra-class distance*: The intra-class distance refers to a statistical metric that quantifies the distance among feature vectors and anchors, commonly associated with class means or predefined class prototypes. In the proposed approach, we compute the intra-class distance between a logit vector \hat{y} predicted by a neural network and its corresponding reference point $\mathbf{p}_y \in \mathcal{P}$. Equation (3) designates the Euclidean distance to estimate the length of the straight-line segment between \hat{y} and \mathbf{p}_y , as indicated below:

$$\mathcal{L}_{intra}(\hat{y}, y) = d(\hat{y}, \mathbf{p}_y) = \|\hat{y} - \alpha \cdot \mathbf{p}_y\|_2 \quad (3)$$

Basically, \mathcal{L}_{intra} encourages any generic neural network $f(x)$ to produce logits positioned closer to their respective fixed-class centers. The loss score obtained through \mathcal{L}_{intra} indicates how well the network has successfully learned to align the predictions with their respective class centers. Note that the Euclidean distance is minimized when \hat{y} perfectly matches $\alpha \cdot \mathbf{p}_y$. Therefore, it helps establish clearer decision boundaries in classification tasks by implicitly improving the feature space organization and reducing the overlap among gallery-enrolled classes.

3) *Inter-class distance*: The inter-class distance is a measure of the dissimilarity between different classes or groups of data points, widely used to quantify how far apart distinct classes are from each other in a feature space. Unlike the previous step, now we estimate the Euclidean distance between a logit vector \hat{y} and all fixed class centers $\mathbf{p}_g \in \mathcal{P}$ associated with subjects enrolled in the gallery set G .

$$\mathcal{L}_{inter}(\hat{y}, y) = \log \left(\sum_g e^{d(\hat{y}, \mathbf{p}_y) - d(\hat{y}, \mathbf{p}_g)} \right) \quad (4)$$

Function \mathcal{L}_{inter} focuses on maximizing the separation between different classes by penalizing the model when the distance to the target category \mathbf{p}_y is larger than the distance to any other class center \mathbf{p}_g . Exponentiating the differences among $d(\hat{y}, \mathbf{p}_y)$ and $d(\hat{y}, \mathbf{p}_g)$ for all $g \in G$ is a common tactic to eliminate negative values and ease the interpretation. Given the asymptotic behavior of the exponential function e^x , condition $d(\hat{y}, \mathbf{p}_y) > d(\hat{y}, \mathbf{p}_g)$ implies that logit vector \hat{y} is distant to its true class, which results in positive or higher error scores; otherwise, the penalty remains lower. The summation over all classes $g \in G$ aggregates penalties across all centroids during training, pushing feature vectors away from their incorrect reference points. Ultimately, the logarithmic function compresses the scale of large penalties, ensuring that the contributions of smaller penalties are not entirely suppressed by the prioritization of larger errors.

4) *Magnitude Estimation*: The magnitude (or norm) is a non-negative function that assigns a measure of length/size to feature vectors. In this work, we use the Euclidean L_2 -norm to impose certain constraints on the predicted logit vector \hat{y} . Function \mathcal{L}_{mag} behaves differently depending on whether the true class y belongs to the gallery set G , as follows:

$$\mathcal{L}_{mag}(\hat{y}, y) = \begin{cases} \max(\alpha - \|\hat{y}\|_2, 0) & \text{if } y \in G \\ \|\hat{y}\|_2 & \text{if } y \notin G \end{cases} \quad (5)$$

The Euclidean norm is mathematically expressed as $\|\cdot\|_2$, inducing a model to produce a logit vector with magnitude closer or greater than α when its label y is enrolled in G . Then, $\|\hat{y}\|_2 \geq \alpha$ indicates a good estimate for y as the loss reaches zero; otherwise, the error increases as the norm of \hat{y} decreases. For samples belonging to background classes, the error merely comprehends the magnitude of \hat{y} , which drives the neural network toward producing logit predictions with smaller norms. Essentially, logit vectors with low magnitudes are less likely to align with any reference point, making it easier to distinguish them as out-of-distribution samples.

5) *Loss aggregation*: Associating all the aforementioned distance metrics, as shown below, is expected to encourage class separability, minimize ambiguity and improve the discriminative power on open-set tasks:

$$\mathcal{L}_{ASL}(\hat{y}, y) = \mathcal{L}_{inter}(\hat{y}, y) + \lambda[\mathcal{L}_{intra}(\hat{y}, y) + \mathcal{L}_{mag}(\hat{y}, y)]$$

By construction, function \mathcal{L}_{inter} is designed to prevent overlaps among different classes by penalizing the embeddings when they fail to maintain sufficient distance to mismatched reference points. Higher λ emphasizes pulling known samples closer to their respective class center as well as keeping background instances near the origin. In addition, it provides flexibility in emphasizing certain aspects of the loss function, making it adaptable to various tasks and datasets.

D. Enrollment and Inference

Following the training process, we construct a gallery of templates \mathcal{T} by iterating over all subjects of interest present in the training set. Specifically, we substitute the fixed class centers $p_g \in \mathcal{P}$ for a template $t_g \in \mathcal{T}$ that captures the central tendency through the mean logit vector of all samples available for subject g . It ensures that reference points are not merely predefined anchors but, instead, dynamically adapted to reflect the model’s learned feature distribution.

The following equations demonstrate how our approach addresses the open-space risk at inference time by assigning an acceptance scores to a prediction $\hat{y} = f(x)$, as follows:

$$\delta = \mathbf{d} \circ [1 - \text{softmin}(\mathbf{d})] \quad (6a)$$

$$\rho = [\max(\delta) - \delta] \circ \|\hat{y}\|_2 \quad (6b)$$

Equation (6a) establishes an element-wise (\circ) relationship between absolute distance vector \mathbf{d} and its relative *softmin* counterpart, which holds the probability of not belonging to any known class. Each item $d_g \in \mathbf{d}$ represents the distance between logit vector \hat{y} and a specific class template $t_g \in \mathcal{T}$. Then, δ_g is expected to hold a small rejection score when \hat{y} belongs to subject g . Equation (6b) converts rejection vector δ into an acceptance vector ρ by subtracting each element from its maximum value. Term $\|\hat{y}\|_2$ incorporates magnitude information assuming that unknown probe samples are close to the logit space origin and have smaller norm/length. Finally, a probe sample is regarded as unknown when all values $\rho_g \in \rho$ fail to surpass a threshold θ , as shown below:

$$\hat{g} = \begin{cases} \text{argmax}(\rho) & \text{if } \max(\rho) > \theta, \\ \text{'unknown'} & \text{otherwise.} \end{cases}$$

IV. EXPERIMENTS

We implemented the proposed approach using the PyTorch framework [32] along with the Bob library [2, 7] for feature extraction. Our application operates on deep representations extracted from the long-established RESNET architectures, specifically utilizing three pre-trained models: VGGFACE2 [5], AFFFE [23] and ARCFACE [10].

1) *Evaluation Metric*: The Open-set Receiver Operating Characteristics (O-ROC) curve plots the Detection and Identification Rate (DIR) against the False Positive Identification Rate (FPIR) by a varying decision threshold θ [34]. DIR represents the proportion of known probe samples from a target g that are correctly detected and identified, while FPIR quantifies the proportion of unknown probe samples mistakenly identified. An ideal open-set face identification system achieves a DIR of 1.0 while maintaining an FPIR close to 0, ensuring high recognition accuracy for known subjects and high rejection of unknowns. The area under the O-ROC curve (AUC) provides a single scalar value that summarizes the overall performance of a recognition system across all threshold values θ . We also report the closed-set *Rank-1* recognition rate, derived from the Cumulative Match Characteristic (CMC) curve, which measures the proportion of probe samples whose correct identity appears within the top- k ranked matches, providing insight into the system’s ranking performance across different values of k .

2) *Datasets and Protocols*: We evaluate the proposed cost function using LFW, IJB-C, and UCCS datasets [20, 27, 37]. The LFW dataset is employed for parameter selection, picked to determine optimal settings that are subsequently applied to other domains without modification. The IJB-C dataset is evaluated under the *test-4 protocol*, which does not provide background samples, requiring us to dynamically synthesize them during training, as detailed in Section III-B. We train the model using only *gallery A*, ensuring that probe samples corresponding to identities exclusive to *gallery B* are treated as unknown subjects. The UCCS dataset comprises a real-world surveillance scenario, containing an official watchlist protocol that includes bounding box coordinates and identity labels for both known and unknown faces. Additionally, we conduct a qualitative analysis on the PUBFIG83 dataset [33] to illustrate how ASL reshapes the latent feature space.

3) *Evaluated Methods*: We conducted experiments in the interest of providing a comprehensive comparison on the performance of AXIAL SPHERE LOSS (ASL). Specifically, we benchmarked ASL against several loss functions, such as *Cross-Entropy* (XEN) [44], *ArcFace* (ARC) [10], *CosFace* (COS) [48], *SphereFace* (SPH) [25], *Class Anchor Clustering* (CAC) [29], *Maximal Entropy* (MAX) [43], and *Object-toSphere* (OBS) [11]. These cost functions have been chosen as they represent a diverse range of methodologies, including margin-based classification, angular-based metric learning, and open-set recognition approaches. By comparing ASL to these widely recognized baselines, we aim to highlight its advantages in terms of improving class separability, reducing open-space risk, and enhancing recognition accuracy for both closed-set and open-set scenarios.

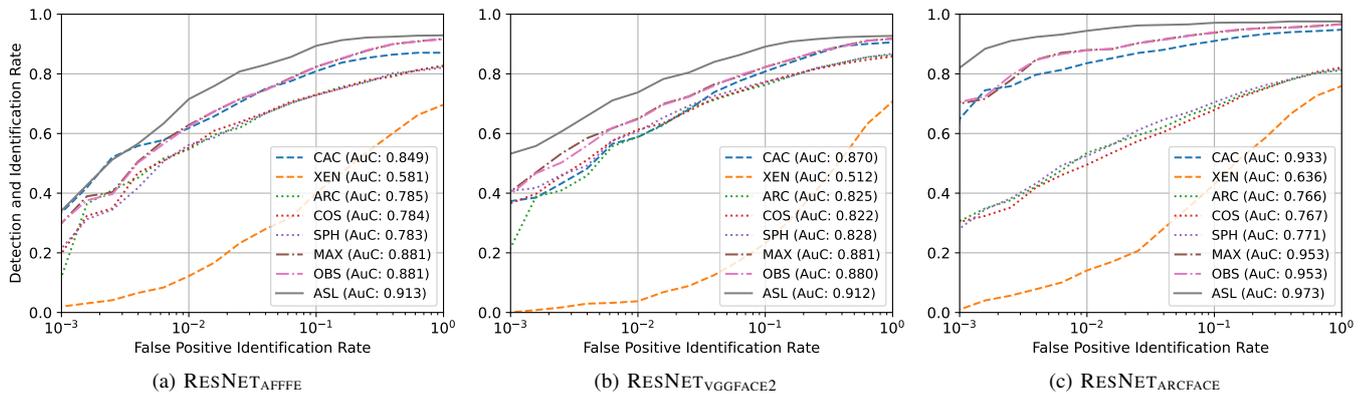


Fig. 2: LITERATURE COMPARISON - LFW. The charts demonstrate open-set recognition results obtained when training an adapter network with either the proposed ASL or seven other cost functions (CAC, XEN, ARC, COS, SPH, MAX, and OBS). Feature embeddings are extracted using three RESNET-based architecture and then fed into the adapter network. The performance is summarized in the form of AUC scores and indicated in the legend.

4) *Watchlist Adaptation*: We incorporate the concept of adapter networks [16] to enable rapid updates to the gallery set without requiring a complete retraining of face DNNs. Our compact adapter model is implemented as a multi-layer perceptron network with two hidden fully-connected layers and 20% of the nodes dropped out. Both intermediate layers enclose 128 neurons for experiments conducted on the LFW dataset. In contrast, the first hidden layer is configured with 256 neurons while the second comprises 512 neurons for evaluations involving IJB-C and UCCS (larger gallery sets). Since the adapter network benefits from deep representations extracted with RESNET-based architectures; namely AFFFE, ARCFACE and VGGFACE2; the input layer size ranges from 1000, 512 and 2018 dimensions, respectively.

During watchlist specialization, the adapter network and all assessed cost functions are trained for a maximum of 100 epochs. The learning rate has been set to $3e^{-4}$ and an early stopping mechanism is triggered whenever the training accuracy reaches 0.995. However, we observe that both CAC and ASL algorithms exhibit a tendency to make the adapter network overfit to the gallery set due to their strong influence on feature space optimization. We impose a stricter training limit for these two cost functions to mitigate this issue, restraining the maximum number of epochs at 50. This adjustment ended up preventing overfitting while maintaining the integrity of the comparison with other loss functions.

A. Parameter Selection

Table I presents the AUC results of a grid search conducted to identify the optimal hyperparameters for the ASL function on the LFW dataset. Results are divided into three sections, each focusing on a different hyperparameter (non-linearity functions along with scaling and weighting factors) in which the best-performing configurations are highlighted in bold. The first section compares different activation functions, revealing that *Tanh* achieves the highest AUC of 0.909. The second one shows that $\alpha = 10$ yields the best AUC with performance gradually decreasing as α increases to 30. Lastly, the optimal value is found to be $\lambda = 0.10$, achieving an AUC of 0.912, while other values such as $\lambda = 0.30$ and $\lambda = 0.05$ deliver comparable results.

Parameter Selection - LFW dataset					
Activation	CeLU	Leaky	ReLU	SiLU	Tanh
AUC	0.904	0.844	0.807	0.817	0.909
Parameter α	5.	10.	15.	20.	30.
AUC	0.909	0.909	0.878	0.861	0.819
Parameter λ	.05	.10	.20	.30	.50
AUC	0.909	0.912	0.911	0.910	0.896

TABLE I: Seeking optimal parameters for AXIAL SPHERE LOSS following a grid search mechanism. Note that best results highlighted in bold font.

We conjecture that *Tanh* outperforms the other activation functions by virtue of computing bounded values in the range $[-1, +1]$, smoothly normalizing the feature space. As a result, it makes the optimization process more stable and prevents the adapter network from learning inconsistent representations. We noticed that setting α to large values can cause the class anchors to become overly dispersed, disrupting the network’s ability to maintain a well-structured feature space and hindering the effective distinction between known and unknown samples. Since λ represents a trade-off among ASL components, larger values may have a negative impact on closed-set classification, whereas small digits may reduce the open-set performance. In favor of contrasting ASL with other literature works, we set $\alpha = 10.$, $\lambda = .10$, and adopt *Tanh* as the activation function for the adapter network.

B. Literature Comparison

To showcase the superiority of AXIAL SPHERE LOSS over other cost functions commonly utilized in face recognition tasks, we conduct a comprehensive evaluation across multiple datasets, including LFW, IJB-C, and UCCS. Figures 2 and 3 disclose O-ROC curves alongside their corresponding AUC scores, generated from experiments carried out on open-set protocols. Additionally, Table II also presents a comparison of closed-set performance indicating the results taken from the standard CMC curve. The table provides a two-level ranking performance for each dataset: *Rank-1* representing the top-1 match accuracy, and *Rank-10*, which measures accuracy when considering the top-10 candidates.

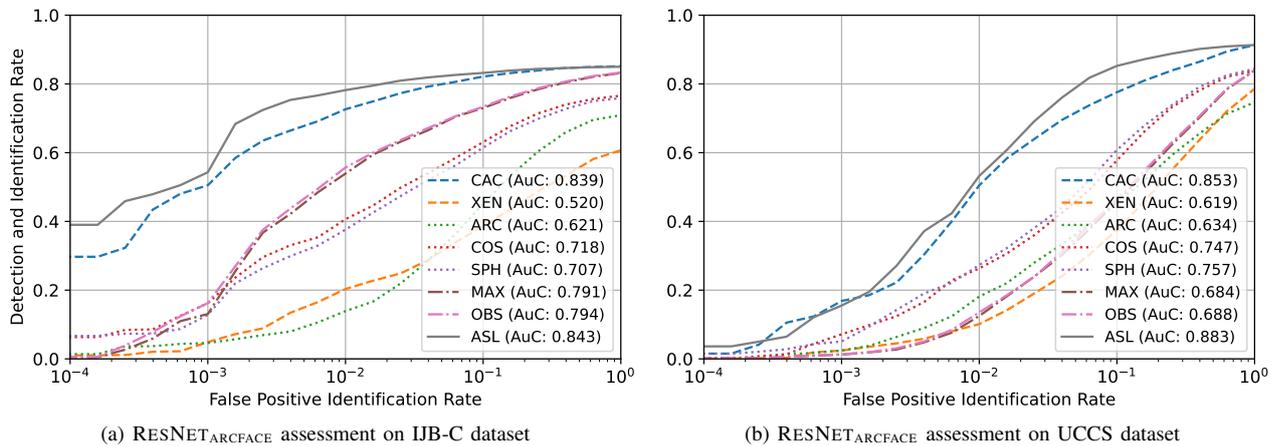


Fig. 3: LITERATURE COMPARISON - IJB-C AND UCCS. Two charts demonstrating open-set recognition results obtained when training an adapter network with either the proposed ASL or seven other cost functions (CAC, XEN, ARC, COS, SPH, MAX, and OBS). Feature embeddings are extracted using $\text{RESNET}_{\text{ARCFACE}}$ architecture and then fed into the adapter network. The performance is summarized in the form of AUC scores and indicated in the legend.

1) *Labeled Faces in the Wild*: Figure 2 presents the results of experiments conducted on the LFW dataset under an open-set protocol introduced by Günther *et al.* [15]. This protocol strategically sets aside subject classes holding a single sample to serve as supplementary identities during the training phase. In fact, evaluating ASL and other background-handling loss functions on Günther’s protocol does not require synthesizing additional feature embeddings to leverage their performance on open-set scenarios.

Results demonstrate that approaches capable of handling background samples, such as MAX and OBS, tend to perform better than other long-established cost functions, known for being employed to train face recognition systems [49]. This suggests that the inclusion of additional background instances equips the adapter network with enhanced capabilities to distinguish between LFW subjects and effectively reject unknown identities. Notably, AXIAL SPHERE LOSS achieves a significant leap in performance, surpassing all baseline methods with outstanding AUC scores of 0.913, 0.912, and 0.973 across the different RESNET architectures. These results highlight ASL’s robust ability to maintain superior detection and identification rates across the entire false-positive range, spanning from 10^{-3} to 10^0 . Among the three evaluated backbones, $\text{RESNET}_{\text{ARCFACE}}$ delivers the best overall performance, making it the most suitable choice for subsequent analyses on IJB-C and UCCS datasets.

2) *IARPA Janus Benchmark series C*: IJB-C dataset lacks background samples available for training in its official open-set protocol. We also observed that the selected additional data did not improve performance, likely due to their statistical mismatch with the gallery set [43]. As picking the right background data plays a crucial role in open-set classification, Figure 3a discloses the performance of experiments in which synthetic additional instances are generated for background-handling cost functions (ASL, MAX and OBS) in accordance with the formulation specified in Section III-B. Note that these embeddings are generated during training and follow the same underlying data distribution of identities enrolled in the gallery set.

Figure 3a converts into numbers how the evaluated cost functions handle open-set face recognition in a more demanding scenario. The proposed ASL achieves the highest AUC score of 0.843 and the curves demonstrate its superiority in maintaining higher detection and identification rates on the full false-positive spectrum. In contrast, traditional loss functions show relatively weaker performance, culminating in XEN reaching the lowest AUC score of 0.520. All methods undergo a drop below $\text{DIR@FPIR} = 2 \times 10^{-3}$, suggesting they struggle to maintain accuracy under more strict conditions. IJB-C’s open-set protocol consists of nearly 20 thousand probe templates when $\text{FPIR} = 10^0$ but decreases significantly when FPIR ranges from 10^{-3} to 10^{-4} where DIR is computed on fewer than 20 samples. This scarcity of test samples makes the curve more susceptible to abrupt changes, potentially leading to misleading interpretations.

3) *UnConstrained College Students*: Unlike previously assessed benchmarks, the UCCS dataset encloses background samples alongside face misdetections. For a more realistic analysis, we get rid of all non-face detections as they end up easing the recognition process, eventually pushing O-ROC curves upwards. As for LFW, UCCS’s open-set protocol eliminates the need for synthesizing additional feature embeddings, enabling a direct assessment of loss functions without the need for artificial data augmentation.

As demonstrated in Figure 3b, ASL persists as the most effective approach with an AUC score of 0.883 on the UCCS dataset, attaining better generalization and discrimination degrees than all seven challengers. CAC also achieved a strong performance with an AUC of 0.853, performing surprisingly well without supplementary background data. However, the results show that it was not able to preserve the fine-grained feature separation learned with ASL. The faster performance degradation on UCCS may be explained by its surveillance-style imagery, which amplifies open-set recognition challenges. UCCS’s unconstrained characteristics imply higher data complexity, introducing noise and data variability, a condition that makes it harder for any model to distinguish between known and unknown classes.

Method	IJB-C			LFW			UCCS			
	RESNET_ARCFACE	$D@F=1\%$	Rank-1	Rank-10	$D@F=1\%$	Rank-1	Rank-10	$D@F=1\%$	Rank-1	Rank-10
XEN		0.20	0.61	0.79	0.14	0.76	0.94	0.10	0.79	0.93
ARC		0.14	0.71	0.85	0.54	0.81	0.94	0.18	0.75	0.88
COS		0.41	0.77	0.86	0.50	0.82	0.94	0.26	<i>0.84</i>	0.93
SPH		0.38	0.76	0.86	0.53	0.82	0.94	0.27	<i>0.84</i>	0.93
CAC		<i>0.73</i>	0.85	<i>0.89</i>	0.84	0.95	<i>0.98</i>	<i>0.53</i>	0.91	0.96
MAX		0.59	<i>0.83</i>	0.90	0.87	<i>0.97</i>	0.99	0.12	<i>0.84</i>	<i>0.94</i>
OBS		0.59	<i>0.83</i>	0.90	<i>0.88</i>	<i>0.97</i>	0.99	0.14	<i>0.84</i>	<i>0.94</i>
ASL		0.78	0.85	0.90	0.94	0.98	0.99	0.55	0.91	0.96

TABLE II: We report $DIR@FPIR=1\%$, showing the detection and identification rate at 10^{-2} false positives, using RESNET_ARCFACE as the feature extractor. Additionally, Rank-1 and Rank-10 scores summarize the closed-set CMC curve. The best and second-best results are highlighted in bold and italics.

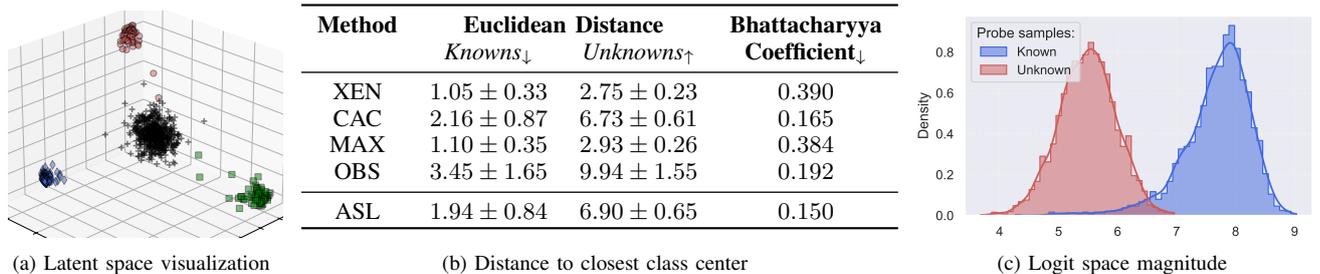


Fig. 4: ABLATION STUDY - PUBFIG83. Scatter plot (a) displays the position of three subjects of interest in the latent space (color markers) and unknown samples close to the origin (black/gray crosses). Table (b) outlines the distance’s $mean \pm std.dev.$ of each probe sample to the nearest class center and a distribution similarity coefficient. Histogram (c) exhibits the different magnitude distribution of known and unknown probe samples.

C. Ablation Study

Hereafter, we adopt the PUBFIG83 dataset to check how ASL influences the weights of the adapter network. Apart from the space visualization, all experiments hold 40 subjects of interest and 40 unknown identities from a total of 83 individuals. Given the known set, we assign 60% of the samples for training and the other 40% for testing. Background samples are generated during training by applying the feature augmentation technique to the gallery set samples.

1) *Latent Space Visualization*: We choose nine random identities to visualize the latent space in a low dimensionality. Three subjects are picked to compose the gallery set, background set, and unknown set, respectively. Figure 4a demonstrates how ASL enforces a structured separation between known and unknown samples. Known identities (color markers) hold simultaneous intra-class compactness and inter-class separation, while unknown samples (black crosses) are pushed toward the origin. The plot shows ASL’s effectiveness in balancing feature compactness and openness, making it well-suited for open-set face recognition tasks.

2) *Distance to Closest Class*: Table 4b presents a quantitative comparison based on Euclidean distance metrics and the *Bhattacharyya* coefficient [4]. ASL ensures that known probe samples remain closer to their respective class centers (1.94 ± 0.84) but pushes unknown probe samples significantly farther apart (6.90 ± 0.65). Additionally, ASL demonstrates the lowest *Bhattacharyya* coefficient (0.150), indicating minimal overlap between the distribution of known and unknown identities. This performance suggests a superior ability to distinguish unseen identities, reducing misclassification risks in open-set scenarios and face recognition tasks.

3) *Feature Magnitudes*: Figure 4c shows the distribution of magnitudes of known and unknown samples in the logit space. Through the norm-based term (Equation 5), ASL leads networks to produce larger magnitude logits for known samples and smaller magnitudes for unknown samples. This distinction in logit magnitudes provides a valuable cue for the model to discriminate between known and unknown identities, enhancing the performance of open-set recognition.

V. CONCLUSION

In this work, we introduced AXIAL SPHERE LOSS (ASL), a cost function designed to enhance open-set recognition by restructuring the latent space. Experiments across face recognition benchmarks demonstrate that ASL consistently outperforms state-of-the-art loss functions in both closed-set and open-set scenarios. By enforcing an axial constraint on feature representations, ASL achieves superior separation between known and unknown identities, mitigating the open-space risk – a critical challenge in open-set recognition. The empirical study confirms that ASL maintains compact intra-class distributions for known identities while increasing the distance of unknown samples, as evidenced by higher Euclidean separation and lower *Bhattacharyya* coefficient. The improved open-set ROC curves and AUC scores reinforce its capability to reject unseen classes more effectively.

These findings establish ASL as a powerful alternative to traditional softmax-based and angular margin losses, especially in open-set recognition applications. Future work may explore further refinements to ASL, including adaptive hyperparameter tuning and extensions to other biometric and object recognition tasks, solidifying its impact in real-world scenarios where open-set generalization is paramount.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the Brazilian National Council for Scientific and Technological Development (CNPq) – Grant 312565/2023-2; the Samsung R&D Institute Brazil (SRBR); and the Universidade Federal de Minas Gerais (UFMG).

ETHICAL IMPACT STATEMENT

The development of Axial Sphere Loss (ASL) aims to improve open-set face recognition by enhancing the separability between known and unknown identities while mitigating the open-space risk. While this advancement contributes to the robustness of biometric systems, it is crucial to assess its ethical implications.

Review and Oversight

This study does not involve direct interaction with human subjects or the collection of personally identifiable information. The datasets used in our experiments, including LFW, IJB-C, PUBFIG83 and UCCS, are publicly available and widely used in face recognition research. We strictly adhere to dataset usage policies and avoid any re-identification attempts beyond what is necessary for our research objectives. Considering the nature of our study, formal oversight from an Institutional Review Board (IRB) was not required.

Potential Risks and Societal Impact

The Axial Sphere Loss (ASL) proposed in this paper improves open-set face recognition by enhancing feature separability and mitigating open-space risk. While this contributes to advancements in security and identity verification, it also raises ethical concerns related to privacy, surveillance, and potential misuse in applications such as mass monitoring and law enforcement. Face recognition technology has historically exhibited bias across demographic groups, leading to disparities in performance. If not properly managed, ASL could unintentionally reinforce these biases, resulting in unfair treatment of underrepresented populations. Additionally, ASL's ability to reject unknown identities more effectively could be exploited in restrictive environments, leading to concerns about exclusion or misuse in authoritarian surveillance systems.

Risk Mitigation Strategies

To reduce these risks, we implement several mitigation strategies:

- **Fairness and Bias Awareness:** We recommend that ASL be evaluated on diverse and balanced datasets before deployment in real-world applications. Future work should explore fairness-aware training and demographic bias analysis.
- **Privacy and Data Ethics:** We do not introduce any new datasets and only use publicly available resources, ensuring compliance with dataset policies. Organizations adopting ASL should implement data protection measures and comply with legal frameworks such as the General Data Protection Regulation (GDPR) and California Consumer Privacy Act (CCPA).
- **Responsible Deployment:** ASL should be used in ethical and regulated environments, with safeguards preventing misuse in mass surveillance and discrimination-based applications. We encourage the development of ethical AI policies when deploying open-set recognition models.

Balancing Risks and Benefits

The potential benefits of ASL, including improved biometric security, fraud prevention, and robust identity verification, outweigh its risks when deployed with fairness considerations and regulatory oversight. By improving open-set recognition, ASL enhances model reliability and real-world applicability, making it a valuable contribution to biometric and security applications. However, we strongly advocate for ethical AI practices, continued bias assessments, and the responsible use of face recognition technology to prevent unintended societal harm.

CODE AVAILABILITY

The source code containing the proposed AXIAL SPHERE LOSS has been developed by Mr. Rafael Henrique Vareto, currently pursuing a Ph.D. degree at the Federal University of Minas Gerais, Brazil. The approach is publicly available as a Python package through the PIP installer program (`pip install openloss`), an effortless way to integrate our method into new projects. The code can also be accessed from the following GitHub repository: <https://github.com/rafaelvareto/open-loss>.

REFERENCES

- [1] Wasseem N Ibrahim Al-Obaydy and Shahrel Azmin Suandi. Open-set face recognition in video surveillance: a survey. In *International Conference on Electrical, Control and Computer Engineering*. Springer, 2020.
- [2] André Anjos, Laurent El-Shafey, Roy Wallace, Manuel Günther, Christopher McCool, and Sébastien Marcel. Bob: a free signal processing and machine learning toolbox for researchers. In *ACM Multimedia*, 2012.
- [3] Jianmin Bao, Dong Chen, Fang Wen, Houqiang Li, and Gang Hua. Towards open-set identity preserving face synthesis. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.
- [4] Anil Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distribution. *Bulletin of the Calcutta Mathematical Society*, 35, 1943.
- [5] Qiong Cao, Li Shen, Weidi Xie, Omkar M. Parkhi, and Andrew Zisserman. Vggface2: A dataset for recognising faces across pose and age. In *International Conference on Automatic Face and Gesture Recognition*, 2018.
- [6] Paul R Cohen and David Jensen. Overfitting explained. In *WAIS*. PMLR, 1997.
- [7] Tiago de Freitas Pereira, Dominic Schmidli, Yu Linghu, Xinyi Zhang, Sébastien Marcel, and Manuel Günther. Eight years of face recognition research: Reproducibility, achievements and open issues. *arXiv*, 2022.
- [8] Jose Sanchez del Rio, Daniela Moctezuma, Cristina Conde, Isaac Martin de Diego, and Enrique Cabello. Automated border control e-gates and facial recognition systems. *Computers & Security*, 2016.
- [9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009.
- [10] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. ArcFace: Additive angular margin loss for deep face recognition. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019.
- [11] Akshay Raj Dhamija, Manuel Günther, and Terrance E. Boult. Reducing network agnostophobia. In *Conference on Neural Information Processing Systems*, 2018.
- [12] Cassio Elias Dos Santos and William Robson Schwartz. Extending face identification to open-set face recognition. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2014.
- [13] Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, and Tao Mei. The elements of end-to-end deep face recognition: A survey of recent advances. *ACM Computing Surveys*, 2022.
- [14] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. Recent advances in open set recognition: A survey. *Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [15] Manuel Günther, Steve Cruz, Ethan M. Rudd, and Terrance E. Boult. Toward open-set face recognition. In *Computer Vision and Pattern Recognition Workshops*. IEEE, 2017.
- [16] Manuel Günther, Akshay Raj Dhamija, and Terrance E. Boult. Watchlist adaptation: Protecting the innocent. In *BIO SIG*, 2020.

- [17] Manuel Günther et al. Unconstrained face detection and open-set face recognition challenge. In *International Joint Conference on Biometrics (IJCB)*, 2017.
- [18] James Henrydoss, Steve Cruz, Chunchun Li, Manuel Günther, and Terrance E Boulton. Enhancing open-set recognition using clustering-based extreme value machine (C-EVM). In *Big Data*, 2020.
- [19] Kashmir Hill. Wrongfully accused by an algorithm. *New York Times*, June 2020.
- [20] Gary B Huang, Marwan Mattar, Tamara Berg, and Eric Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in Real-Life Images: detection, alignment, and recognition*, 2008.
- [21] Furkan Kasim, Terrance E Boulton, Rensso Mora, Bernardo Biesseck, Rafael Ribeiro, Jan Schlueter, Tomáš Repák, Rafael H. Vareto, David Menotti, William R. Schwartz, and Manuel Günther. Watchlist challenge: 3rd open-set face detection and identification. In *International Joint Conference on Biometrics (IJCB)*. IEEE, 2024.
- [22] Yann LeCun, Corinna Cortes, and Christopher J.C. Burges. Mnist handwritten digits dataset, 1998.
- [23] Chunchun Li, Manuel Günther, and Terrance E. Boulton. ECLIPSE: Ensembles of centroids leveraging iteratively processed spatial eclipse clustering. In *Winter Conference on Applications of Computer Vision (WACV)*, 2018.
- [24] Pan Li, Da Li, Wei Li, Shaogang Gong, Yanwei Fu, and Timothy M Hospedales. A simple feature augmentation for domain generalization. In *International Conference on Computer Vision (ICCV)*, 2021.
- [25] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. SpheroFace: Deep hypersphere embedding for face recognition. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017.
- [26] Yoanna Martindiez-Diaz, Luis S Luevano, Heydi Mendez-Vazquez, Miguel Nicolas-Diaz, Leonardo Chang, and Miguel Gonzalez-Mendoza. ShuffleFaceNet: A lightweight face architecture for efficient and highly-accurate face recognition. In *International Conference on Computer Vision Workshops*, 2019.
- [27] Brianna Maze et al. IARPA Janus Benchmark - C: Face dataset and protocol. In *International Conference on Biometrics*, 2018.
- [28] Qiang Meng, Shichao Zhao, Zhida Huang, and Feng Zhou. MagFace: A universal representation for face recognition and quality assessment. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2021.
- [29] Dimity Miller, Niko Sunderhauf, Michael Milford, and Feras Dayoub. Class anchor clustering: A loss for distance-based open set recognition. In *Winter Conference on Applications of Computer Vision (WACV)*, 2021.
- [30] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. Reading digits in natural images with unsupervised feature learning. In *Conference on Neural Information Processing Systems Workshop*, 2011.
- [31] Pau Panareda Busto and Juergen Gall. Open set domain adaptation. In *International Conference on Computer Vision (ICCV)*, 2017.
- [32] Adam Paszke et al. Pytorch: An imperative style, high-performance deep learning library. *Neural Information Processing Systems*, 2019.
- [33] Nicolas Pinto, Zak Stone, Todd Zickler, and David Cox. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In *Computer Vision and Pattern Recognition Workshops*. IEEE, 2011.
- [34] N Poh, C Chan, J Kittler, J Fierrez, and J Galbally. BEAT-biometrics evaluation and testing: Description of metrics for the evaluation of biometric performance. Technical report, Idiap, 2012.
- [35] Tony Romm. Amazon’s facial-recognition tool misidentified 28 lawmakers as people arrested for a crime. *Washington Post*, July 2017.
- [36] Swami Sankaranarayanan, Azadeh Alavi, Carlos D. Castillo, and Rama Chellappa. Triplet probabilistic embedding for face verification and clustering. In *International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2016.
- [37] Archana Sapkota and Terrance E. Boulton. Large scale unconstrained open set face database. In *International Conference on Biometrics: Theory, Applications, and Systems (BTAS)*, 2013.
- [38] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boulton. Toward open set recognition. *Pattern Analysis and Machine Intelligence*, 2012.
- [39] Walter J. Scheirer, Lalit P. Jain, and Terrance E. Boulton. Probability models for open set recognition. *Pattern Analysis and Machine Intelligence*, 36(11), 2014.
- [40] Rafael Henrique Vareto and William Robson Schwartz. Unconstrained face identification using ensembles trained on clustered data. In *International Joint Conference on Biometrics (IJCB)*, 2020.
- [41] Rafael Henrique Vareto, Samira Silva, Filipe Costa, and William Robson Schwartz. Towards open-set face recognition using hashing functions. In *International Joint Conference on Biometrics (IJCB)*, 2017.
- [42] Rafael Henrique Vareto, Manuel Günther, and William Robson Schwartz. Open-set face recognition with neural ensemble, maximal entropy loss and feature augmentation. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*. IEEE, 2023.
- [43] Rafael Henrique Vareto, Yu Linghu, Terrance Edward Boulton, William Robson Schwartz, and Manuel Günther. Open-set face recognition with maximal entropy and objectsphere loss. *Image and Vision Computing*, 141:104862, 2024.
- [44] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zissermann. Open-set recognition: A good closed-set classifier is all you need? In *International Conference on Learning Representations*, 2022.
- [45] Vikas Verma, Alex Lamb, Christopher Beckham, Amir Najafi, Ioannis Mitliagkas, David Lopez-Paz, and Yoshua Bengio. Manifold mixup: Better representations by interpolating hidden states. In *International Conference on Machine Learning*. PMLR, 2019.
- [46] V Esposito Vinzi, Wynne W Chin, Jörg Henseler, Huiwen Wang, et al. *Handbook of partial least squares*. Springer, 2010.
- [47] Riccardo Volpi, Pietro Morerio, Silvio Savarese, and Vittorio Murino. Adversarial feature augmentation for unsupervised domain adaptation. In *Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2018.
- [48] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. CosFace: Large margin cosine loss for deep face recognition. In *Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [49] Mei Wang and Weihong Deng. Deep face recognition: A survey. *Neurocomputing*, 2021.
- [50] Xiang Wu, Ran He, Zhenan Sun, and Tieniu Tan. A light CNN for deep face representation with noisy labels. *Transactions on Information Forensics and Security*, 2018.