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

Rafael Henrique Vareto William Robson Schwartz

DCC DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO — UFMG

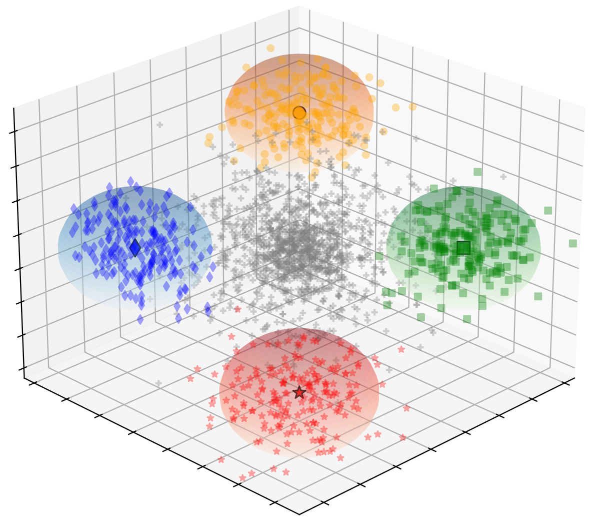
Open-set Recognition

Modern face recognition systems excel in **closed-set** scenarios but struggle with **open-set** tasks where unknown individuals (classes not seen during training) must be rejected. This is critical for real-world applications like surveillance, watchlist screening, and access control. Some of the key challenges are:

- **False positives:** High-confidence errors can lead to wrongful identifications with serious societal consequences.
- **Lack of unknown references:** Most loss functions ignore non-gallery samples during training, limiting discrimination.
- **Open-space risk:** Traditional methods misclassify unknown individuals due to unconstrained feature distributions.

Axial Sphere Loss (ASL)

ASL is a cost function that restructures the latent space to explicitly separate known and unknown identities. It encourages embeddings of known classes to cluster around predefined anchor points, while guiding unknown samples toward the origin.



ADVANTAGES:

- ☑ Minimizes open-space risk via geometric constraints.
- ☑ Enhances discrimination: Intra-class compactness + inter-class separation.

CORE MECHANISM:

- **Fixed axial centers:** Gallery (known) classes are mapped to orthogonal, pre-defined positions along feature axes.
- **Origin-bound unknowns:** Non-gallery samples are pushed toward the origin, creating a clear rejection boundary.
- **Spherical decision:** At inference, samples are accepted only if they fall near their class centroid (low open-space risk).

HOW ASL WORKS:

- **Fixed class centers:** Each anchor corresponds to a subject of interest in the latent space.

$$\mathcal{P} = \left[\mathbf{p}_1 = (\alpha, 0, \cdots, 0, 0),\ \mathbf{p}_2 = (0, \alpha, \cdots, 0, 0), \ldots,\ \mathbf{p}_{|G|} = (0, 0, \cdots, 0, \alpha)\right]$$

- **Intra-class term:** Logits of known samples align with their fixed axial centers.

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

- **Inter-class term:** Maximizes angular separation between different classes.

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

- **Magnitude regularization:** Known samples have larger logit magnitudes; unknowns cluster near zero.

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

- **Loss aggregation:** Associates the aforementioned functions to generate a penalty score.

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

- **Inference:** Determines whether a probe sample belongs to a known subject.

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

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

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

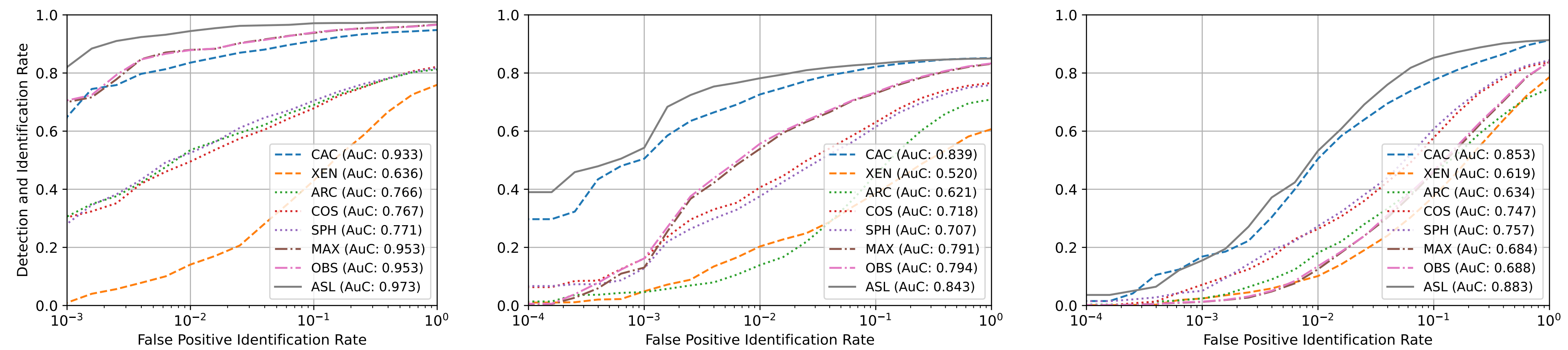
Further Contributions

- **Adaptation Networks:** ASL integrates with pre-trained ResNet models (e.g., AFFFE, ArcFace, VGGFace2) via lightweight and plug-and-play adapter networks, enabling quick deployment without retraining the entire model [1].
- **Optimized MixUp Manifold:** We introduce a strategy for generating *synthetic background samples* in the latent space, allowing the model to simulate unknown classes during training when labeled out-of-distribution data is unavailable [2].

Key Results & Performance

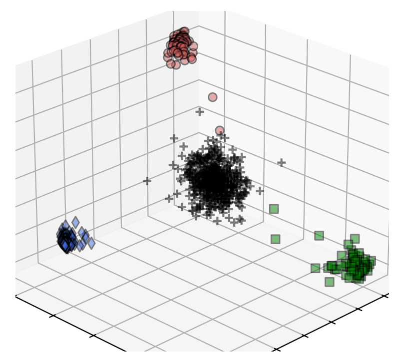
Table 1. We report DIR@FPIR=1%, showing the detection and identification rate at 10^{-2} false positives, using $\text{ResNet}_{\text{ArcFace}}$ as the feature extractor. Additionally, *Rank-1* and *Rank-10* scores summarize the closed-set CMC curve. The best and second-best results are highlighted in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

Method	IJB-C			LFW			UCCS		
$\text{ResNet}_{\text{ArcFace}}$	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	*0.84*	0.93
SPH	0.38	0.76	0.86	0.53	0.82	0.94	0.27	*0.84*	0.93
CAC	*0.73*	**0.85**	*0.89*	0.84	0.95	*0.98*	*0.53*	**0.91**	**0.96**
MAX	0.59	*0.83*	**0.90**	0.87	*0.97*	**0.99**	0.12	*0.84*	*0.94*
OBS	0.59	*0.83*	**0.90**	*0.88*	*0.97*	**0.99**	0.14	*0.84*	*0.94*
ASL	**0.78**	**0.85**	**0.90**	**0.94**	**0.98**	**0.99**	**0.55**	**0.91**	**0.96**



(a) $\text{ResNet}_{\text{ArcFace}}$ assessment on LFW dataset (b) $\text{ResNet}_{\text{ArcFace}}$ assessment on IJB-C dataset (c) $\text{ResNet}_{\text{ArcFace}}$ assessment on UCCS dataset

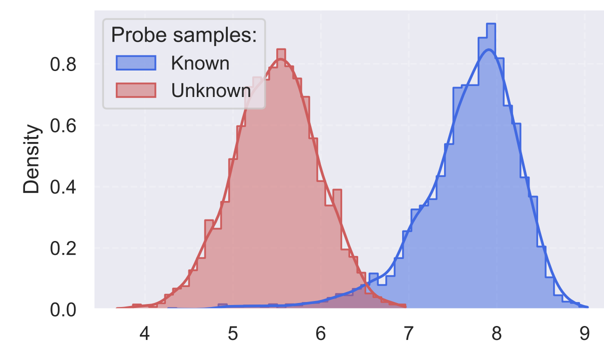
Figure 1. LITERATURE COMPARISON - LFW, IJB-C & UCCS – Three charts demonstrating open-set recognition results obtained when training an adapter network with either the proposed ASL or seven other cost functions (XEN, ARC, COS, SPH, CAC, MAX, and OBS). Feature embeddings are extracted using $\text{Resnet}_{\text{ArcFace}}$ architecture and then fed into the adapter network. The overall performance is summarized in the form of AUC scores as indicated in each plot legend.



(a) Latent space visualization

Method	Euclidean Distance Knowns$_\downarrow$	Euclidean Distance Unknowns$_\uparrow$	Bhattacharyya Coefficient$_\downarrow$
XEN	1.05 ± 0.33	2.75 ± 0.23	0.390
CAC	2.16 ± 0.87	6.73 ± 0.61	0.165
MAX	1.10 ± 0.35	2.93 ± 0.26	0.384
OBS	3.45 ± 1.65	9.94 ± 1.55	0.192
ASL	1.94 ± 0.84	6.90 ± 0.65	0.150

(b) Distance to closest class center



(c) Logit space magnitude

Figure 2. 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* ± *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.

Baselines: Cross-Entropy (XEN), ArcFace (ARC), CosFace (COS), SphereFace (SPH), Class Anchor Clustering (CAC), Maximal Entropy (MAX) and ObjectoSphere (OBS).

[1] Vareto, R., Linghu, Y., Boult, T., Schwartz, W., & Günther, M. (2024). Open-set face recognition with maximal entropy and objectosphere loss. IMAVIS.
[2] Vareto, R., Günther, M., & Schwartz, W. R. (2023). Open-set face recognition with neural ensemble, maximal entropy loss and feature augmentation. SIBGRAPI.