HyperSpaceX: Radial and Angular Exploration of HyperSpherical Dimensions Supplementary Material

Chiranjeev Chiranjeev, Muskan Dosi, Kartik Thakral Mayank Vatsa, and Richa Singh

Indian Institute of Technology Jodhpur, India
{chiranjeev.1,dosi.1,thakral.1,mvatsa,richa}@iitj.ac.in
https://github.com/IAB-IITJ/HyperSpaceX

1 Derivation of Decision Boundary and Predictive Measure for HyperSpaceX

The *HyperSpaceX* framework depends on three primary components: θ, ϕ, and δ for introducing more refined decision boundaries for separability among classes and more discriminative representation learning of features. We derive the formulations dependent on these three essential components of the HyperSpaceX framework for two purposes: (1) for the decision boundary and (2) predictive measure for the most favourable class distribution. The formulations are built using cosine angular-based triangle law properties. In which triangular formation is created between three vectors: scaled proxy vector ω_r, sample's feature vector x, and resultant vector R between ω_r and x, having angle θ between ω_r and x, and angle ϕ between ω_r and R.

(1) Decision Boundary: Assuming a scenario of 2-class classification problem. The decision boundary dependent on three essential loss factors is derived as follows:

$$(\cos(\theta_1 + m) + ||x_1||_2) - (\cos(\theta_2) + ||x_2||_2) = 0 \text{ for class 1}$$
$$(\cos(\theta_1) + ||x_1||_2) - (\cos(\theta_2 + m) + ||x_2||_2) = 0 \text{ for class 2}$$

where, according to cosine angular-based triangle law $||x_1||_2$ is defined as,

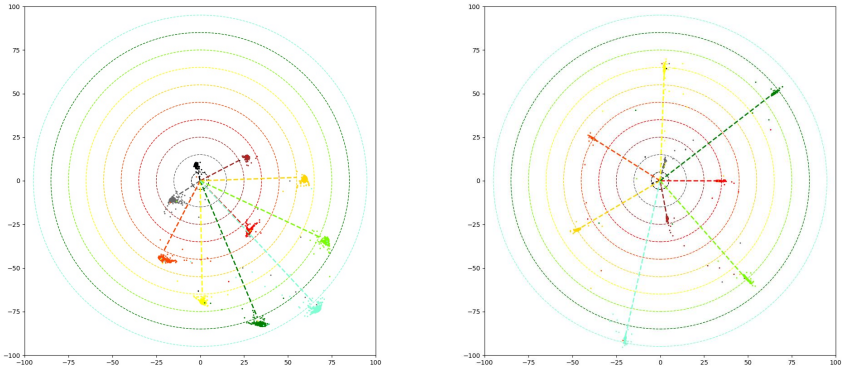
$$||R_1||_2 \cos(\theta_1 + m) + ||\omega_{r_1}||_2 \cos(\pi - ((\theta_1 + m) + \phi_1)) \tag{1}$$

and $||x_2||_2$ as,

$$||R_2||_2 \cos(\theta_2 + m) + ||\omega_{r_2}||_2 \cos(\pi - ((\theta_2 + m) + \phi_2)) \tag{2}$$

(2) Predictive Measure: The class prediction is made based on the smallest resultant vector being selected from a set of resultant vectors computed between scaled proxy and test sample embedding. The resultant vector (R_i) and its magnitude computation for a test sample feature (x) is calculated against each scaled proxy vector (ω_{r_i}) using the vector formulation,

$$||R_i||_2 = ||x - \omega_{r_i}||_2 \quad i \in 1, 2, 3, ... \tag{3}$$



(a) (b)

Fig. 1: Visualization of geometric ablation of feature representation learning via DistArc loss components. This image shows the significance of (a) $\cos\theta$ and δ, and (b) the incorporation of $\cos\phi$ in DistArc loss formulation. Part (a) illustrates the distribution and grouping of features in both angular and radial dimensions. Part (b) further refines this concept by displaying enhanced feature compactness and the scaling of features to specific radii-hyperspheres, each associated with their class proxies in radial-angular dimensions. The diverse color spectrum represents unique class distributions, each aligned with their scaled class proxies, depicted as colored lines in distinct angular orientations. The overlay of concentric circles symbolizes the 2D-hyperspherical manifolds, each with varying radii, underlining the multi-dimensional aspect of this learning model.

and, resultant vector's magnitude calculation using cosine angular-based triangle side length computation property, showing its dependence on the introduced angular and radial components of HyperspaceX framework *i.e.*, θ, ϕ and length of test sample feature vector $||x||_2$,

$$||R_i||_2 = ||x||_2 \cos\phi_i \; + \; ||\omega_{r_i}||_2 \cos(\pi - (\theta_i + \phi_i)) \tag{4}$$

where, $(\pi - (\theta_i + \phi_i))$, represents the value of the angle between x and R, which is calculated from all interior angles in a triangle sums to 180°or π in radians.

2 Geometric Visual and Experimental Ablation of DistArc Loss in HyperSpaceX Methodology

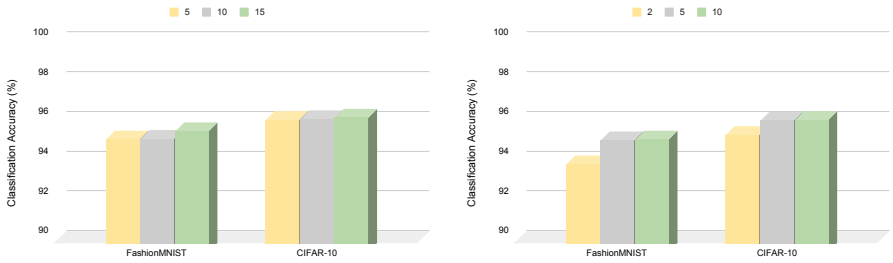
This section provides a geometric and experimental interpretation of each component of the proposed *DistArc* loss by learning in both combined angular and radial dimensions. Fig. 1 represents the overall geometric visualization of all parts of the proposed loss function *i.e.*, $\cos\theta$, $\cos\phi$ and δ.

Visual Geometric Ablation: The geometric interpretation of both components, $\cos(\theta_{y_i} + m)$, δ_{y_i} and δ_j in the proposed *DistArc* loss function is visually represented through the depiction of features in a 2-dimensional latent hyperspherical space, as illustrated in Fig. 1a. This representation demonstrates that these terms help position each class feature to the particular radial hypersphere along with the minimization of angles θ. These δ-based loss components help increase separability among different classes in radial dimensions, and discriminability is also achieved by further minimizing intra-class distance among similar class features and maximizing inter-class distance among distinct classes.

Fig. 1b shows the significance of all components along with $\cos\phi_{y_i}$ term in the DistArc loss. Feature representation in Fig. 1b interprets that these terms help create a more compact distribution of each class cluster along with keeping the dissimilar class features apart by distance/radial factor. We can observe that the minimization of angles θ and ϕ leads to a concentrated angular distribution of features and clustering among features belonging to the same class. Furthermore, it prevents features from extending beyond their respective hypersphere of radii $||\omega_{r_{y_i}}||_2$. The angular separability and discriminability among classes are achieved through these cosine-based terms in the loss.

The *HyperSpaceX* framework, as observed through the geometric visualization of feature representation learned using DistArc loss in Fig. 1, leads to the enlargement and alignment of feature vectors to their respective angular and radial positions within the hyperspheres space. This results in a more separable and discriminative arrangement of data points across various hyperspherical dimensions. The overall combination of angular loss minimization due to θ and ϕ, and angular-radial loss minimization due to δ and ϕ helps to increase the intra-class similarity among same class features at point $\omega_{r_{y_i}}$ by shifting the features up to the $||\omega_{r_{y_i}}||_2$ radial hypersphere and compacting it closer to $\omega_{r_{y_i}}$ scaled proxy vector. Therefore, the DistArc loss ensures that features x_i maintain specific angular orientations and prevent them from extending beyond their respective radial hyperspheres. Additionally, it elevates inter-class distance among features x_i and the dissimilar class proxies ω_j*'s*. This approach, incorporated within the HyperSpaceX methodology, contributes to enhancing feature discriminability and separability.

Experimental Ablation: Experimentally, we highlight the importance of each field introduced in the proposed *HyperSpaceX* framework. We reported extensive results showcasing the classification accuracy on various image classification datasets while utilizing ViT-B architecture with an embedding layer of size 512 dimensions. From Fig. 2a, we can analyze the performance increment as we increase the radial gap between adjacent hyperspheres. It illustrates as the gap increases, the separability among the different classes increases. It leads to reduced miss-classifications among class feature points of distinct classes in the latent space. Fig. 2b represents the classification performance on varying the number of hyperspheres, for distributing the class features over different hyperspherical regions. It can be analyzed as we increase the number of hyperspheres; it leads to more enhanced separability and discriminability among class fea-



(a) **(b)**

Fig. 2: Classification performance (in %) (a) by varying radial gap between adjacent hyperspheres where each color denotes gap between the hyperspheres; (b) by varying number of hyperspheres where each color denotes the different number of hyperspheres chosen.

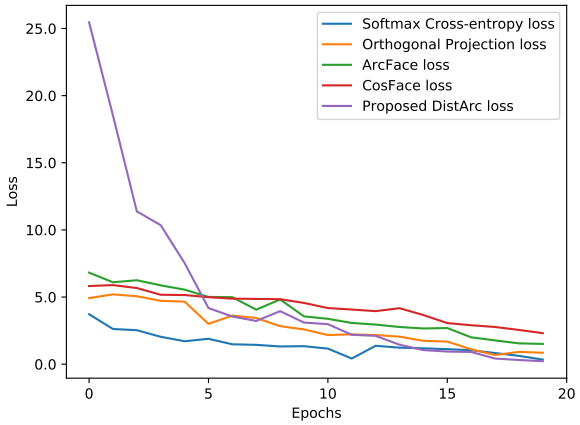


Fig. 3: Comparative training convergence graph across different loss formulations.

tures. This causes performance improvement as we increase the angular and radial search space exploration through multi-hyperspherical regions.

Consequently, from both Visual and Experimental Ablation study, we can highlight the key significance of *HyperSpaceX* framework for learning highly separable and discriminative representation of features in latent space using *DistArc* loss. In the proposed DistArc loss function, each introduced component plays a crucial role in enhancing performance across multiple datasets by representing feature points in various angular and radial dimensions. All components are carefully designed to optimize the function's effectiveness, demonstrating their individual significance in diverse data environments.

Table 1: Quantitative results (in %) of face recognition model trained using different loss functions on MS1Mv2 dataset. The best and second best performances are **bolded** and underlined.

Loss	LFW	AgeDB-30	CA-LFW	CP-LFW
Center loss	98.75	-	85.48	77.48
SphereFace	99.42	92.88	90.30	81.40
CosFace	99.73	92.98	92.83	91.03
ArcFace	99.82	98.15	**95.45**	**92.08**
SphereFace2	99.50	93.68	93.47	91.07
DistArc	**99.82**	**98.21**	94.87	91.10

3 Graphical Representation of Training Convergence

The results of various loss formulations and training convergence are depicted through the graph in Fig. 3. It shows the loss convergence over epochs. It can be visualized that the DistArc loss leads to higher loss values in initial epochs, which decrease gradually due to the diverging of each identity to different radial hyperspheres. It can be observed that the DistArc loss curve converges faster in comparison to ArcFace, CosFace, and Orthogonal Projection losses, which further finally converges to the lowest minimum value of 0.22 in comparison to other loss functions even including cross-entropy loss, which attains a minimum of 0.34 loss value.

4 More Empirical Results

Face Recognition: We conducted another experiment by training the model on one of the largest face datasets, MS1Mv2. This experiment is conducted to ensure a fair comparison with state-of-the-art (SoTA) methods that have also been trained on this dataset. Table 1 presents the performances of several face recognition techniques on testing face databases. It can be observed that the model when trained with the proposed DistArc loss on the MS1Mv2 dataset, achieves superior performance on most of the testing face datasets. This demonstrates the efficacy of the proposed DistArc Loss in large-scale face datasets as well.

Image Classification: The classification results with 1024 embedding size on CIFAR-100 and TinyImageNet datasets are shown in **Table 2**. We can observe superior performance on backbones trained with DistArc loss. Further, an improvement of up to 4.36% on CIFAR-100 and about 2.0% on TinyImageNet datasets for models trained with the proposed DistArc loss and other existing loss functions. We have also conducted **McNemar's test** for iResNet50 backbone on TinyImageNet dataset with 512 embedding size, which yields $\chi^2 = 4.848$ ($>\chi^2_{0.05}$), stating a significant difference between models trained using cross-entropy loss and proposed DistArc loss.

Table 2: Image classification accuracy (in %) of models trained on 1024 dimensions with various loss functions. The best and second best performances are **bolded** and underlined.

CIFAR-100				
Backbone	**Cross entropy**	**ArcFace**	**CosFace**	**DistArc**
iResNet50	85.16	82.67	83.11	**89.52**
ViT-B	82.41	81.00	80.84	**85.74**
TinyImageNet				
Backbone	**Cross entropy**	**ArcFace**	**CosFace**	**DistArc**
iResNet50	60.24	61.47	62.01	**64.01**
ViT-B	**78.26**	77.96	78.07	78.22

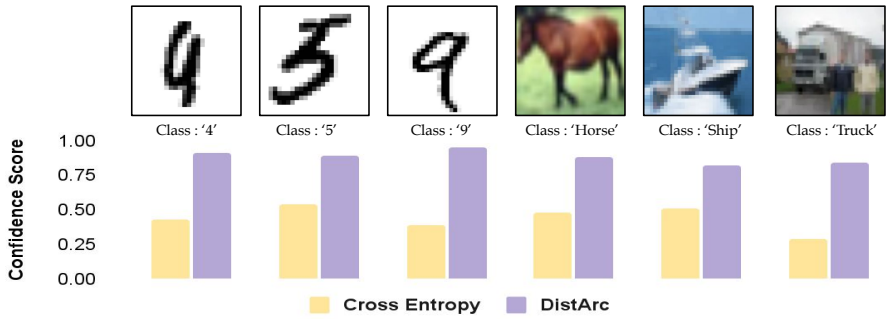


Fig. 4: Analytical results demonstrating the comparative study of confidence scores in correct class predictions. The first row shows the images from two datasets, MNIST and CIFAR-10, along with their ground truth class. The second row illustrates the corresponding confidence scores of class predictions made by models trained with Cross entropy loss and DistArc loss.

Confidence in Class Predictions: The quality of predictions should be evaluated not only based on their correctness but also on the level of confidence associated with them. In Fig. 4, we compare the confidence levels of predictions made using a model trained with cross-entropy loss and DistArc loss. For each sample depicted in Fig. 4, we consistently observe that predictions made using DistArc loss exhibit significantly higher confidence for both the MNIST and CIFAR-10 datasets. These experiments show that feature points near the decision boundary have lower prediction confidence with cross-entropy loss. With the proposed DistArc loss, features are compact within the same class and separative between distinct classes, leading to high predictive scores. This demonstrates the effectiveness of introducing the HyperSpaceX framework that primarily focused on separating the distinct classes with significantly larger distances and compacting the same class features, which is achieved through the proposed DistArc loss, hence resulting in producing confident classifications. We also applied a temperature-based calibration [1] technique for better confidence score understanding. After calibrating the iResNet50 backbone trained on the CIFAR-100

Table 3: Classification results across different radial factors (r_{y_i}) on CIFAR-10 dataset.

radius (r_{y_i})	**1,2,3,. . . ,10**	**5,10,15,. . . ,50**	**10,20,30,. . . ,100**
Accruacy (in %)	94.18	95.48	95.82

Table 4: Ablative experiment for λ on CIFAR-10 dataset with iResNet50 backbone.

$\boldsymbol{\lambda}$	**0.001**	**0.002**	**0.005**	**0.01**
Accruacy (in %)	94.55	95.17	95.82	94.07

Table 5: Statistics of the used datasets for the Image classification task.

Dataset	**Total # of images**	**# of training images**	**# of test images**
MNIST	60K	50K	10K
FashionMNIST	70K	60K	10K
CIFAR-10	60K	50K	10K
CIFAR-100	60K	50K	10K
TinyImageNet	0.11M	0.1M	10K
CUB-200	11,788	5,994	5,794
ImageNet	1.38M	1.28M	0.1M

dataset using Cross-entropy and DistArc loss, we achieve an Expected Calibration Error (ECE) of 0.467 and **0.337**, respectively (lower is better).

Sensitivity analysis based on r_{y_i}: Since r_{y_i} refers to a scaling radial factor defining the radials of multi-hyperspheres. We have performed its sensitivity analysis on the CIFAR-10 dataset with iResNet50 backbone through Table 3. It can be analyzed that larger radial values and gaps lead to more multi-spherical feature space exploration. It results in more separated distinct class features in the latent space, causing reduced inter-mixing of feature points. Hence, it leads to more enhanced discriminative features, causing improvement in classification performance as the radial gap increases.

Effect of λ: Through Table 4, we study the effect of λ as we vary the values of λ by training on CIFAR-10 dataset. It can be observed that HyperSpaceX performs optimally with λ as a multiplier of 1e$-$3, achieving a balanced DistArc loss and highly discriminative feature learning.

5 Limitations and Future Work

We note certain limitations of the HyperSpaceX framework that will be further addressed in future work. A flexible learnable margin can be introduced in the proposed HyperSpaceX framework to further reduce the inter-mixing of feature points of distinct classes over consecutive hyperspheres, *i.e.* enhancing more discriminability among features. Further, as the number of radials

Table 6: Statistics for the used datasets for Face Recognition task.

Dataset	**# of identities**	**# of images**	**split**
CASIA-WebFace	10K	0.5M	train
MS1Mv2	87K	5.8M	train
D-LORD	1,500	0.22M	train
LFW	5,749	13,233	test
CFP-FP	500	7,000	test
AgeDB-30	568	16,488	test
CA-LFW	5,749	11,652	test
CP-LFW	5,749	12,174	test
D-LORD	600	36,420	test

increases extremely, our approach necessitates the adoption of a learnable, unnormalized proxy expansion technique to ensure smoother learning and reduced feature spreading while expanding the radials to a greater extent. Meanwhile, our method excels in distinctive class learning and classification enhancement, especially in the visual domain by improving classification and recognition performance, its adoption to other modalities like text and audio currently remains unexplored.

6 Statistics for the Used Datasets

Table 5 and Table 6 represent the dataset statistics of the image classification and face recognition datasets, respectively. The face identification performance on D-LORD dataset is evaluated for high-resolution face images up to a distance of 5m. Overall these datasets are utilized to conduct training and testing experiments for the proposed and existing methodologies.

References

1. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: Int. Conf. Mach. Learn. pp. 1321–1330. PMLR (2017)