ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

When Curvature Counts: Hyperbolic Geometry in Prototype-Based Image Classification

Silvia Grosso[1,2], Samuele Fonio[1], Mirko Polato[1], Roberto Esposito[1], Sara Bouchenak[2] *

1. University of Turin – Computer Science Department, Turin, Italy
2. INSA Lyon – LIRIS, Lyon, France

Abstract.

Prototype Learning offers an interpretable and efficient classification framework by mapping data into an embedding space structured around class prototypes. Recent research has explored non-Euclidean geometries, such as hyperspherical and hyperbolic spaces, to more effectively model latent hierarchical structures and complex data relationships. While these geometries have shown potential, leveraging them within an image classification context is not trivial. To address this, we propose HypPNet, a hyperbolic prototypical model on the Poincaré ball that integrates Riemannian optimization and norm-based regularization to perform effectively without prior data knowledge. Experiments on three benchmark datasets and multiple embedding dimensions show that HypPNet outperforms its competitors across alternative geometries, improving classification performance over various metrics.

1 Introduction

Prototype Learning (PL) has recently emerged as a promising and interpretable alternative to conventional classification approaches. The key idea is to represent each target class through a prototype in the embedding space, enabling the model to leverage metric relations between samples and prototypes for classification. PL has shown promising results in various domains, such as image classification and few-shot learning [1].

Since PL directly operates on metric elements, a major concern is equipping the embedding space with an appropriate metric. Non-Euclidean manifolds have shown strong potential, particularly in scenarios where data exhibit latent hierarchies or tree-like structures [2]. In such cases, the negative curvature of hyperbolic spaces allows for modeling inter-sample relationships more effectively than Euclidean ones. In addition to choosing the appropriate metric, another key aspect in PL is how prototypes are positioned and updated within the embedding space. Some methods adopt fixed prototypes, while others employ dynamic ones [2,3], which rely on class centroids or learnable prototypes jointly optimized with the network. It is worth mentioning that non-Euclidean approaches mainly rely on fixed prototypes, while neglecting their update through backpropagation.

*This work was partly supported by the Agence Nationale de la Recherche under the France 2030 program (ANR-22-PESN-0013), as part of project M4DI of the national strategy program Digital Health, and by the Spoke "FutureHPC & BigData" of the ICSC - Centro Nazionale di Ricerca in "High Performance Computing, Big Data and Quantum Computing".

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

To this end, we propose HypPNet, a prototype learning framework that exploits hyperbolic geometry and jointly optimizes the prototypes with the network parameters. HypPNet further benefits from two key aspects: norm-based regularization and shrink initialization of the prototypes, which we thoroughly investigate, to the best of our knowledge, for the first time. All these aspects enable HypPNet to consistently outperform other geometries across various embedding dimensions and evaluation metrics, highlighting the potential of hyperbolic geometry for training robust and effective neural networks.

Our contribution can be summarized as follows: (*i*) we introduce HypPNet, a framework for dynamically updating prototypes on hyperbolic manifolds through Riemannian optimization; (*ii*) we show that the combination of shrinkage initialization for prototypes and norm regularization for embeddings significantly enhances the effectiveness of hyperbolic PL; (*iii*) we provide extensive empirical evidence that HypPNet outperforms euclidean and hyperspherical baselines across multiple metrics and regardless of the embedding dimension.

2 Method

In this section, we introduce HypPNet, which extends the idea of learning prototypes as network parameters to the non-Euclidean domain. This approach overcomes the computational challenge of calculating the hyperbolic centroid, which is particularly demanding, while offering a more adaptable alternative to methods based on fixed prototypes. Our methodology involves extracting the output from a backbone network, projecting it onto the Poincaré ball, and then optimizing a distance-based cross-entropy loss. Formally, the n-dimensional Poincaré ball and the corresponding distance are defined as:

$$\mathbb{B}^n = \{x \in \mathbb{R}^n : \|x\|^2 < 1\}$$
$$d_{\mathbb{B}}(x, y) = \operatorname{arcosh}\left(1 + 2\frac{\|x - y\|^2}{(1 - \|x\|^2)(1 - \|y\|^2)}\right), \quad x, y \in \mathbb{B}^n.$$

Given a dataset $\{(x_i, y_i)\}_{i=1}^N$ and a backbone network $f_\theta(\cdot)$, we define class prototypes as parameters on the manifold $\Pi = \{\pi_j\}_{j=1}^C \in \mathbb{B}^n$. To embed the data onto the Poincaré we employ the exponential map centered at the origin $\mathbf{0}$, as done by [2], $z_i = \exp_{\mathbf{0}}(f_\theta(x_i)) \in \mathbb{B}^n$. Classification is performed by comparing each embedding z_i with the class prototypes π_j through the hyperbolic distance $d_{\mathbb{B}}(z_i, \pi_j)$. The network parameters θ and prototypes Π are jointly optimized by minimizing a distance-based cross-entropy loss:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{N}\sum_i \log \frac{\exp[-\,d_{\mathbb{B}}(z_i, \pi_{y_i})/\tau]}{\sum_{j=1}^C \exp[-\,d_{\mathbb{B}}(z_i, \pi_j)/\tau]}, \tag{1}$$

where τ is the *temperature* parameter [4], which controls the sharpness or smoothness of the distance distribution. Lower temperatures produce sharper, more confident distributions, while higher temperatures yield flatter ones, favoring greater exploration and uncertainty.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

Unlike previous works with fixed prototypes, we update prototypes directly on the manifold using Riemannian Stochastic Gradient Descent (RSGD) [5]:

$$\pi \leftarrow \exp_\pi(-\mu \cdot \nabla_\pi^R \mathcal{L}),$$

where the Riemannian gradient is defined as $\nabla_x^R = \nabla_x/(\lambda_x)^2$, with $\lambda_x = \frac{1}{1-\|x\|^2}$ the conformal factor of the Poincaré metric (with curvature equal to 1), which scales distances to account for the curvature of the manifold, and ∇_x is the standard Euclidean gradient. It has been observed that embeddings approaching the Poincaré boundary (*i.e.*, having a norm close to 1) exhibit a vanishing gradient behavior [6]. To mitigate this effect, we apply norm clipping before projection, as proposed by [6]. To further enhance this regularization, we incorporate two main elements: a norm-based regularization term [2], and a shrink initialization of the prototypes. The norm-based regularization is defined as:

$$\mathcal{L}_{\text{reg}}(z) = -\lambda \cdot d \log(1 - \|z\|^2), \tag{2}$$

where λ, denoted as the *slope*, controls the attraction toward the manifold center, and d denotes the embedding dimension. The shrink initialization simply consists of initializing the prototypes within a specific range, *i.e.*, $\pi \in [-\epsilon, \epsilon]$. As we will show, these two techniques are both crucial to achieving an advantage over other geometries and show very promising results across a wide range of embedding dimensions and evaluation metrics.

3 Experiments and Discussion

Experimental Setting. Experiments were conducted on three fine-grained image classification benchmarks with hierarchical label structures (CIFAR100, Aircraft, CUB [7–9]), across a variety of embedding dimensions, which depend on the number of classes. This choice is motivated by the fact that hyperbolic spaces are reported to perform particularly well in low-dimensional settings. However, we also investigate their effectiveness in higher-dimensional spaces and across a wider range of evaluation metrics. As baseline methods, we selected models operating under different geometries. ECL exploits a Euclidean geometry, using dynamic prototypes updated through backpropagation, and optimizing a distance-based cross-entropy loss [10]. Its variant ECL^p has a separate optimizer for the prototypes. NormFace (NF) is a hyperspherical method leveraging cosine similarity with cosine softmax and dynamic prototypes [11]. All the details about hyperparameters used for the training, as well as the code to reproduce the results, can be found at this link.

Results. Our evaluation included accuracy, Out-Of-Distribution (OOD) detection, and robustness. The results shown are the average over three independent runs. OOD detection was evaluated using the AUROC metric, computed from the maximum softmax confidence. Robustness was evaluated with the Projected Gradient Descent (PGD) attack, a standard and strong benchmark for adversarial testing.

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

CUB								
	Accuracy				OOD			
d	16	64	128	200	16	64	128	200
ECL	$42.07_{\pm 0.54}$	$44.70_{\pm 0.43}$	$45.21_{\pm 0.30}$	$44.99_{\pm 0.71}$	$74.17_{\pm 1.54}$	$75.47_{\pm 1.77}$	$76.71_{\pm 0.63}$	$75.41_{\pm 0.86}$
ECL^p	$22.31_{\pm 1.00}$	$36.75_{\pm 0.63}$	$39.79_{\pm 0.47}$	$39.81_{\pm 0.60}$	$57.05_{\pm 4.77}$	$65.64_{\pm 1.91}$	$67.05_{\pm 0.57}$	$67.83_{\pm 1.78}$
NF	$42.92_{\pm 0.24}$	$43.73_{\pm 0.31}$	$42.43_{\pm 0.66}$	$42.49_{\pm 0.73}$	$69.14_{\pm 0.56}$	$70.32_{\pm 0.57}$	$70.13_{\pm 0.55}$	$69.48_{\pm 1.85}$
Ours	**43.41**$_{\pm 0.46}$	**47.66**$_{\pm 0.69}$	**49.71**$_{\pm 0.34}$	**49.45**$_{\pm 0.73}$	**80.14**$_{\pm 1.42}$	**82.99**$_{\pm 0.80}$	**82.48**$_{\pm 0.40}$	**82.50**$_{\pm 0.33}$
Aircraft								
	Accuracy				OOD			
d	8	32	64	100	8	32	64	100
ECL	$74.89_{\pm 0.28}$	**76.58**$_{\pm 0.35}$	$76.11_{\pm 0.44}$	$76.74_{\pm 0.33}$	$88.49_{\pm 1.28}$	$90.59_{\pm 0.75}$	$90.33_{\pm 0.79}$	$91.63_{\pm 0.32}$
ECL^p	$56.31_{\pm 0.77}$	$70.16_{\pm 0.81}$	$70.42_{\pm 0.70}$	$69.80_{\pm 0.40}$	$76.15_{\pm 0.73}$	$85.14_{\pm 0.75}$	$86.93_{\pm 0.81}$	$85.31_{\pm 0.45}$
NF	**75.38**$_{\pm 0.42}$	$76.19_{\pm 0.60}$	$76.11_{\pm 0.54}$	$76.83_{\pm 0.60}$	$83.45_{\pm 0.46}$	$85.30_{\pm 0.36}$	$85.14_{\pm 1.23}$	$86.50_{\pm 0.48}$
Ours	$74.87_{\pm 0.78}$	$76.19_{\pm 0.52}$	**78.21**$_{\pm 0.09}$	**77.70**$_{\pm 0.12}$	**93.06**$_{\pm 1.30}$	**93.39**$_{\pm 0.07}$	**92.94**$_{\pm 0.35}$	**93.98**$_{\pm 0.50}$
CIFAR100								
	Accuracy				OOD			
d	8	32	64	100	8	32	64	100
ECL	$73.47_{\pm 0.53}$	$75.51_{\pm 0.26}$	$75.2_{\pm 0.38}$	$75.79_{\pm 0.12}$	$75.25_{\pm 1.08}$	$78.68_{\pm 0.35}$	**78.7**$_{\pm 0.6}$	**78.73**$_{\pm 0.22}$
ECL^p	$72.35_{\pm 0.1}$	$74.47_{\pm 0.25}$	$75.18_{\pm 0.63}$	$75.08_{\pm 0.15}$	**75.96**$_{\pm 0.55}$	$77.6_{\pm 0.23}$	$77.73_{\pm 0.17}$	$78.02_{\pm 0.54}$
NF	**73.98**$_{\pm 0.41}$	$75.46_{\pm 0.33}$	$74.90_{\pm 0.53}$	$75.30_{\pm 0.25}$	$75.25_{\pm 0.12}$	$76.56_{\pm 0.26}$	$76.84_{\pm 0.17}$	$76.56_{\pm 0.08}$
Ours	$73.74_{\pm 0.30}$	**75.84**$_{\pm 0.24}$	**76.89**$_{\pm 0.28}$	**77.03**$_{\pm 0.21}$	$74.47_{\pm 0.55}$	**78.83**$_{\pm 0.24}$	$78.53_{\pm 0.37}$	$78.36_{\pm 0.15}$

Table 1: Average test accuracy and OOD detection on three datasets for our proposed method HypPNet and the competing approaches. The results are the mean over three runs with their standard deviation, and the best results are in bold.

Table 1 shows the accuracy and Out-Of-Distribution performances on the 3 selected datasets. Overall, HypPNet achieves the highest accuracy across all embedding dimensions in CUB, and in the higher embedding dimensions in Aircraft and CIFAR100, highlighting the effectiveness of our method. NF performs slightly better in the lowest-dimensional setting, likely due to its lower temperature ($\tau = 0.08$), which sharpens class boundaries and helps compensate for the limited representational capacity in low dimensions. Furthermore, the unit-norm constraint imposed by hyperspherical geometry may act as an implicit regularizer, preventing feature collapse and improving class discrimination.

For OOD detection, greater separability between ID and OOD samples leads to higher AUROC values. We restricted OOD evaluation to compatible architectures: CIFAR10 was used as the OOD dataset for CIFAR100, and CUB/Aircraft were used reciprocally. As observed in Table 1, our method often achieves the best OOD performance. This suggests that our method is intrinsically more robust to OOD samples, probably due to the exponential growth of the hyperbolic distance, which amplifies the discrepancy between ID and OOD samples. The weaker performance of HypPNet on CIFAR100 is likely due to the semantic overlap between ID and OOD samples (CIFAR10 and CIFAR100), which may make OOD detection more difficult and narrow the performance gap among methods.

For the robustness, we adopted a PGD attack with perturbation levels $\varepsilon \in \{0, 0.8/255, 1.6/255, 3.2/255\}$. The higher a model's accuracy remains under stronger perturbations, the more robust it is. As shown in Figure 1, HypPNet consistently outperforms the baselines across all datasets, showing superior robustness to adversarial perturbations. This behavior is particularly evident

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

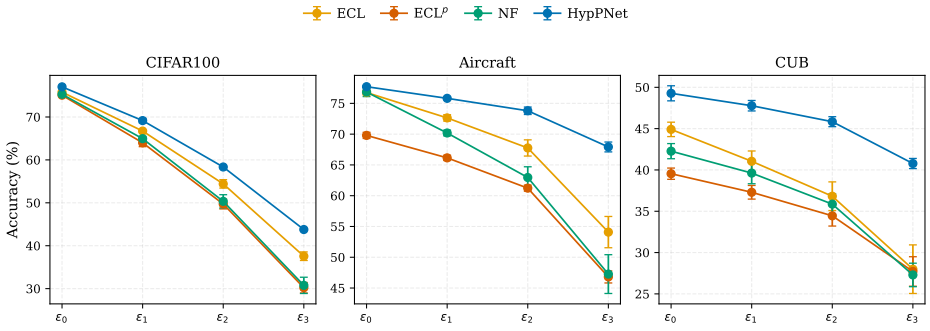

Fig. 1: Accuracy under increasing perturbation levels ε_0 to ε_3.

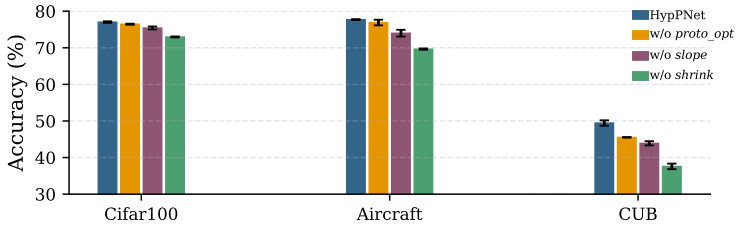

Fig. 2: Ablation study. Accuracy after individually removing each key component of HypPNet: the separate prototype optimizer, slope regularization, and shrink initialization.

on Aircraft and CUB, where our method exhibits the smallest drop in accuracy as the perturbation level increases. The superiority of hyperbolic methods in adversarial robustness has been well established, which is likely an intrinsic consequence of the exponentially inflated distances.

Ablation Study. To assess the contribution of each component of our model, we conducted an ablation study analyzing the effect of using a separate optimizer for prototypes, the slope regularization term, and the shrinkage initialization. Figure 2 reports the results obtained for the highest embedding dimension of each dataset, showing the drop in accuracy when removing each component individually. In all cases, accuracy consistently decreases regardless of the component removed or the dataset considered. Interestingly, using a separate optimizer for prototypes is beneficial in HypPNet, unlike in the Euclidean case, where ECL^p consistently performs worse than ECL.

A key factor influencing performance is the slope parameter, which controls the embedding norms. Although we only report the results obtained for the highest embedding dimensions, we evaluated its effect across all embedding dimensions for each dataset. HypPNet consistently outperformed the same method without the regularization ($\lambda = 0$) in all settings, highlighting the crucial stabilizing effect of the slope. Across our experiments, $\lambda = 0.1$ proved to be the best choice for low embedding dimensions, while $\lambda = 0.01$ worked better for higher

ESANN 2026 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium) and online event, 22-24 April 2026, i6doc.com publ., ISBN 9782875870964. Available from http://www.i6doc.com/en/.

ones, as expected [2].

Finally, the shrinkage initialization is another essential component of our HypPNet. Without it, prototypes could be initialized too close to the boundary of the Poincaré ball, leading to vanishing gradients and preventing the model from learning effectively. We fine-tuned the prototype initialization within the interval $[-\epsilon, \epsilon]$, testing $\epsilon \in \{0.01, 0.025, 0.05, 0.075, 0.1\}$. The value $\epsilon = 0.05$ consistently yielded the best performance across all datasets, regardless of the embedding dimension.

4 Conclusions

In this paper, we introduced HypPNet, a hyperbolic prototype learning model that leverages Riemannian optimization and norm-based regularization to perform classification without requiring prior data knowledge. Through extensive experiments on benchmark datasets and across multiple embedding dimensions, HypPNet consistently outperformed Euclidean and hyperspherical baselines in terms of accuracy, robustness, and out-of-distribution detection. Our findings highlight the advantages of dynamically learning prototypes on hyperbolic manifolds and the stabilizing effect of the proposed regularization.

References

[1] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

[2] Mina Ghadimi Atigh, Martin Keller-Ressel, and Pascal Mettes. Hyperbolic busemann learning with ideal prototypes. *Advances in neural information processing systems*, 34:103–115, 2021.

[3] Pascal Mettes, Elise Van der Pol, and Cees Snoek. Hyperspherical prototype networks. *Advances in neural information processing systems*, 32, 2019.

[4] Hong-Ming Yang, Xu-Yao Zhang, Fei Yin, and Cheng-Lin Liu. Robust classification with convolutional prototype learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3474–3482, 2018.

[5] Silvere Bonnabel. Stochastic gradient descent on riemannian manifolds. *IEEE Transactions on Automatic Control*, 58(9):2217–2229, 2013.

[6] Yunhui Guo, Xudong Wang, Yubei Chen, and Stella X Yu. Clipped hyperbolic classifiers are super-hyperbolic classifiers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11–20, 2022.

[7] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[8] Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *arXiv preprint arXiv:1306.5151*, 2013.

[9] Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

[10] Loic Landrieu and Vivien Sainte Fare Garnot. Leveraging class hierarchies with metric-guided prototype learning. In *British Machine Vision Conference (BMVC)*, 2021.

[11] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille. Normface: L2 hypersphere embedding for face verification. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1041–1049, 2017.