Geometric Autoencoder with Manifold Learning

Mahboubeh Farahat and Ali Ahmadi

December 7, 2024

Geometric Autoencoder with Manifold Learning

Mahboubeh Farahat
Faculty of *Computer Engineering*
K. N. Toosi University of Technology
Tehran, Iran
mfarahat@email.kntu.ac.ir

Ali Ahmadi
Faculty of *Computer Engineering*
K. N. Toosi University of Technology
Tehran, Iran
ahmadi@ kntu.ac.ir

***Abstract*— Autoencoders are widely recognized as non-probabilistic learning models for extracting useful information from data. Most autoencoder models assume a Euclidean geometry for the underlying nature of the data. However, recent advancements in geometric learning suggest that incorporating curvature information of the intrinsic manifold of data may yield richer representations. In this work, we investigate the performance of a learning method that embeds data under curved geometric constraints. Our method assumes that the data manifold consists of both curved and Euclidean spaces. Experimental results demonstrate that our method achieves performance comparable to state-of-the-art techniques**

Keywords—Manifold Learning, Geometric Learning, Auto-Encoders, Representation Learning

I. Introduction

Autoencoders and manifold learning methods are crucial components of machine learning, designed to reduce the dimensionality of high-dimensional data points. Existing manifold learning methods can be broadly categorized into two approaches: global and local [1]. Global methods aim to preserve the global geometric structure of the manifold, while local methods focus on preserving its local geometric structure. Both approaches share the fundamental assumption that the input data points lie on a manifold that is either globally or locally isometric to a subset of Euclidean space [2-4].

Recent research suggests that incorporating curvature information can improve performance over Euclidean-based methods [4,5]. For example, employing spherical spaces enhances the discriminative ability of models in zero-shot learning for unknown classes. Similarly, hyperbolic spaces have demonstrated comparable benefits when used as embedding spaces [6]. Curved spaces can effectively encode complex data structures by leveraging their discriminative capabilities, which allow for more precise decision boundaries, especially in low-shot scenarios.

In this paper, we leverage the curvature information of the intrinsic data structure to learn a meaningful latent embedding space within an autoencoder framework. Our method utilizes two distinct autoencoders with different latent spaces: Hyperbolic and Euclidean. Each space is governed by a latent-specific manifold regularization term. The Riemannian metric enhances the model’s capacity to generate more meaningful latent representations. Experimental results show that our proposed method performs comparably to existing state-of-the-art approaches

The remainder of this paper is organized as follows: Section II provides a brief overview of related work. Section III introduces the proposed approach. Section IV presents experimental results comparing our method to other feature selection techniques. Finally, Section V concludes the paper and discusses future directions.

II. Related Works

An Autoencoder (AE) is a type of neural network consisting of two parts: an encoder and a decoder. A basic AE is an unsupervised model that learns a compact representation of data through the hidden layer of the encoder. Several modified versions of the basic AE have been introduced [7]. Regularized AEs, for example, include additional regularization terms to improve the model's generalization capability. Denoising AEs [8], a specific type of regularized AE, are trained with noise-corrupted input data, enabling the model to learn representations that ignore input noise. Other notable regularized AEs include HSAE (Hessian Sparse Auto Encoder) and LMAE (Large Magin Auto Encoder). HSAE focuses on dimensionality reduction while preserving the locality of the data, and MRAE [9] employs two regularization strategies—linearity and sparsity—to encourage sparsity and flexible structure learning.

In addition to AE models, manifold learning methods are often used to uncover the intrinsic structure of data. Manifold learning approaches can be categorized into several classes: local manifold learning methods, aim to obtain a low-dimensional representation by preserving the local geometric structure of the data [5]. For instance, LLE (Locally Linear Embedding) constructs a linear subspace within the neighborhood of each data point, while HLLE (Hessian (Locally Linear Embedding) assumes that the manifold is locally isometric to Euclidean space. LTSA applies PCA for dimensionality reduction within local neighborhoods and patches of the data.

Variational autoencoders (VAEs) [11] are generative models designed to produce new sample data. As probabilistic models, VAEs generate encoded representations of features from which new data samples can be created. Various extensions of VAEs have been introduced in recent research, including Conditional VAE (CVAE), which generates new images conditioned on specific features.

The combination of manifold learning methods and VAEs represents a novel extension of VAEs. For example, the Flat Manifold VAE learns a smooth latent manifold using a Euclidean metric to compute similarity [11].

Most computer vision tasks traditionally rely on Euclidean or spherical embeddings and linear hyperplanes. However, the application of curved spaces to improve model efficiency is an emerging area of research. Khrulkov [6] [citation] proposed hyperbolic image embeddings for learning representations, particularly in hierarchical datasets. In this work, the network operates in Euclidean space in all layers except the final layer, where it transitions from Euclidean to hyperbolic space using the exponential map. Experiments show that this hyperbolic network outperforms other methods by better conforming to the intrinsic geometry of image manifolds. Further research has explored the use of hyperbolic spaces for various tasks, including visual anomaly detection, natural language processing, and identifying out-of-distribution objects [12].

Gulcehre [13] proposed a hyperbolic attention network, which modifies the geometry of the embedding space for object representations. The results indicate that altering the embedding space leads to a more efficient neural network without increasing the model's parameter count.

Few-shot learning is another application that benefits from curved geometries, such as spherical spaces. While most methods for learning in curved spaces utilize Euclidean-based optimization methods, there is growing evidence that non-Euclidean optimization methods are necessary for effectively training models in curved spaces. Several studies have shown improvements in model performance when using specialized optimization methods. Tabealhojeh [14] introduced a new optimization technique for Riemannian manifolds based on ADAM, incorporating orthogonality constraints into bi-level optimization problems. This method uses Riemannian operations such as retraction and parallel transport to optimize models while preserving the underlying Riemannian geometry.

III. PROPOSED METHOD

Inspired by the concept of product spaces [12,15,16], our proposed method for representation learning employs two individual networks: an autoencoder in hyperbolic space and a simple autoencoder in Euclidean space. Each autoencoder produces a reduced representation of the input that captures specific characteristics of the data. For instance, the hyperbolic space effectively uncovers the intrinsic hierarchical structure of the input.

Unlike traditional product models, which combine features directly, our method constructs a dynamic product input space by training different autoencoders under distinct Riemannian manifolds.

Before training the model, a k-nearest neighbors (KNN) graph is used to calculate the similarity between samples and the structural relationships within the input data. To transfer intrinsic information into the latent embedding space, a weight parameter W is introduced. This parameter ensures that similar samples are positioned closer together, while dissimilar samples are pushed farther apart in the latent space

To ensure consistency between the two spaces, we introduce a regularization term in the loss function that minimizes the divergence between the similarity matrices of the two spaces, promoting a coherent alignment of their latent representations.

Fig. 1. Shows the overall structure of proposed method. z_i' and z_j' are latent representation of inputs. The overall representation is concatenated representation of two autoencoders.

Before go further, we provide the a few required definitions.

Exponential and logarithmic maps: to perform operations in Riemannian spaces, an exponential map, maps a tangent vector in Euclidean space to a manifold. The logarithmic map, projects a point on manifold to its tangent vector representation.

A. Hyperbolic Autoencoder (HAE)

Hyperbolic space, characterized by constant negative curvature, provides an efficient framework for modeling hierarchical and tree-like structures, as well as graph neural networks. It offers significant advantages for encoding hierarchical relationships and efficiently modeling data with exponential growth patterns.

Several common models exist for hyperbolic space, such as the Poincaré Disk Model [6]. In our method, we apply the Poincaré model, which represents hyperbolic space within a unique disk. The n-dimensional Poincaré ball with curvature k, is defined with Eq.1 and Eq.2 is applied to transform an embedded $x \in R^n$ to the Poincare Ball $x \in D^n$. To calculate the geodesic distance between two points on a hyperbolic space, the Eq. 3 and Eq. 4 are used.

$$H_k^n = \left\{ x \in R^n : ||x|| < {}^{-1}/_{k} \right\} \quad (1)$$

$$X_H = \begin{cases} X, & if\ ||x|| < {}^{1}/_{k} \\ \frac{1-\varepsilon}{|k|}\frac{x}{||x||}, & else \end{cases} \quad (2)$$

$$d_{Geo}(x, y) = \frac{2}{\sqrt{|k|}} tanh^{-1}(\sqrt{|k|}|| - x \oplus y) \quad (3)$$

$$x \oplus y = \frac{(1 + 2|k| < x, y > +|k|||y||^2)x + (1 - |k|||x||^2)y}{1 + 2|k| < x, y > +|k|^2||x||^2||y||^2} \quad (4)$$

The Hyperbolic Autoencoder (HAE) is designed to handle non-Euclidean manifolds for the input data. In this framework, the Euclidean input is projected into a non-Euclidean space to serve as the input to the encoder. Both the encoder and decoder in our model are constructed using hyperbolic layers. Specifically, the encoder receives non-Euclidean input and uses Eq. (1) to project this input onto a point in the Poincaré ball. The output of the encoder is a non-Euclidean latent representation, which is then processed by the decoder, also operating in a non-Euclidean space.

To compute the reconstruction error, the decoder's output is projected back to the original input space, and the mean squared error (MSE) is used to measure the discrepancy. To preserve the local geometry of the input KNN graph in latent space of HAE, the geodesic distance, as defined in Eq. (6), is calculated within a neighborhood of a central point.

Eq. (5) presents the overall loss function for the two autoencoders. The first term represents the reconstruction error, while the second term is a manifold regularization term that preserves the local intrinsic geometry of the manifold. The loss function is composed of both Euclidean and non-Euclidean terms. As a result, gradient descent optimization is employed, rather than Riemannian optimization methods, to efficiently minimize the loss.

B. Euclidean Autoencoder(EAE)

Another component of our proposed method is a Euclidean autoencoder. The structures of the two networks are similar, with the key distinction being that the mapping between Euclidean and non-Euclidean spaces is not required in the Euclidean network. Similar to the Hyperbolic Autoencoder (HAE), a regularization term is incorporated into the Euclidean network to preserve the local intrinsic geometry. This regularization term is shown in Eq. (7)

$$loss = \sum_{i=1}^{N} (\left\|x_i - z_i\right\| + \sum_{j=1}^{N_{mi}} W_{ij}(\left\|z_i' - z_j'\right\|)) + \alpha\left\|A_{Eucluidean} - A_{Hyperbolic}\right\| \tag{5}$$

$$\left\|z_i' - z_j'\right\| = d_{geo}(z_i', z_j') \tag{6}$$

$$\left\|z_i' - z_j'\right\| = d_{euclidean}(z_i', z_j') \tag{7}$$

C. Experiments

We investigate the efficiency of our proposed methods in comparison with traditional manifold learning techniques and state-of-the-art autoencoders. To assess the performance of our model, we conduct experiments in two phases. In the first phase, we compare the proposed method with traditional manifold learning approaches, and in the second phase, we evaluate the performance of our model against other autoencoders. To ensure a fair comparison, we use datasets based on the results reported in [7]. Since no performance data for autoencoders on the USPS dataset is available, we compare our method with traditional manifold learning methods using the results reported in [5].

In the second phase of the experiments, to achieve a fair comparison with other autoencoders, we configure our model following the setup described in [7,17].

The datasets used in each phase are as follows:

- **YaleFace B**: 2,414 source images resized to 32x32.
- **USPS**: 9,298 images resized to 16x16 grayscale.
- **MNIST**: 70,000 28x28 grayscale handwritten digits (0–9).
- **Fashion-MNIST**: A dataset similar to MNIST, containing 70,000 28x28 grayscale images across 10 fashion product categories.
- **CIFAR-10**: 60,000 32x32 color images categorized into 10 classes.

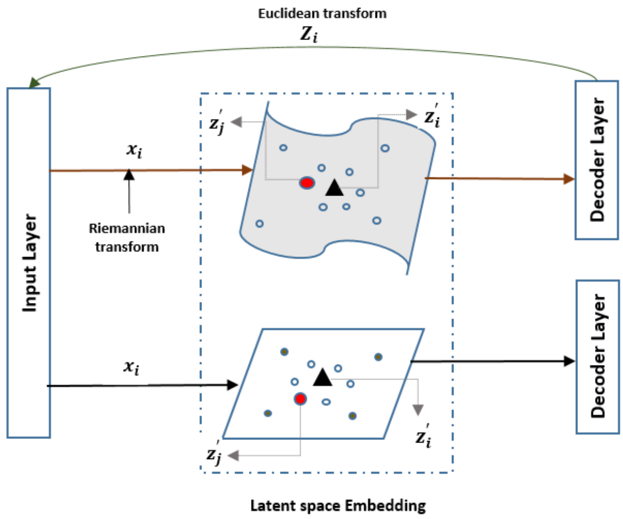

Fig. 1. Architecture of Proposed Method

As outlined in [7], the number of hidden neurons in each layer of the compared methods is typically selected to be between 100% and 200% of the input dimension. However, our experiments demonstrate that the hyperbolic space outperforms the Euclidean network in lower-dimensional spaces. While using two autoencoders may initially appear more complex than employing a single autoencoder, a single autoencoder with a higher number of neurons per layer will have more parameters than two parallel autoencoders. To get close to fair comparison, in our method, both the EAE and HAE models consist of three fully connected layers with [256, 64, 32] hidden neurons, and an output layer with 16 neurons for the MNIST and Fashion-MNIST datasets. For the CIFAR-10 dataset, the autoencoder architectures are [512, 256, 64]. The activation function used in each hidden layer of both autoencoders is ReLU. We used fewer neurons in each layer compared to other methods mentioned in [7]. Moreover, we include a regularization term to maintain consistency between the two spaces. However, this may lead to performance trade-offs due to the heterogeneity between the spaces. To mitigate this, we set a small value for α=0.001 to regulate the information transfer between the spaces.

Table 1 presents the results on the USPS and Yale datasets. For USPS, our method outperforms the others in terms of classification accuracy, demonstrating its superior ability to uncover and preserve the local structure of the input data. While the YaleFace dataset is a flat Riemannian space that is locally isometric to Euclidean space, our model still achieves better performance than CA-LLE, a curvature-aware manifold learning method. This highlights the advantage of employing a model that considers a mixed intrinsic structure, especially when prior knowledge of the underlying data structure is unavailable.

As confirmed by the results in Tables 2 and 3, the proposed method achieves comparable classification performance. Since MNIST is not characterized by a curved space, only marginal improvements are observed compared to a simple autoencoder. However, Fashion-MNIST, with its more curved structure, shows a significant performance gain from the proposed method, outperforming all other approaches. For the CIFAR-10 dataset, which exhibits an

Table 1- Classification Accuracy on YaleFace and USPS datase*ts*

	USPS	YaleFace
LLP	90.14	70.29
LEP	92.81	68.62
LLE	92.48	60.42
CA-LLE	94.52	73.87
AE	**95.68**	**83.56**
Proposed Method	**97.14**	**84.48**

Table 2-Classification Accuracy comparison with other Autoencoders

	AE	HSAE	RAE	DMRAE	Proposed Method
MNIST	98.22	98.95	98.96	**98.99**	98.55
CIFAR-10	44.53	53.57	54.06	54.06	**55.34**
MNIST-Fashion	89.16	90.04	90.06	90.24	**90.76**

Table 3-AUC comparison with state-of-the-art Autoencoders

	AE	HSAE	RAE	DMRAE	Proposed Method
MNIST	98.22	98.95	98.21	98.98	**99.43**
CIFAR-10	78.92	83.76	84.67	84.63	**87.45**
MNIST-Fashion	99.42	99.65	99.63	**99.70**	99.50

even more pronounced curvature, our proposed method delivers the best performance.

Given the use of a mixed model, we utilized the Adam optimization method. While Adam is efficient for input spaces with lower curvature, for higher curvature spaces, a Riemannian optimization method or a hybrid approach may be more effective.

IV. CONCLUSION

In this paper, we introduced a mixed autoencoder consisting of two independent components: a Hyperbolic Autoencoder (HAE) and a Euclidean Autoencoder (EAE). The HAE is designed to capture the hierarchical structure of data within a non-zero curvature space, while the EAE represents a conventional model assuming zero curvature for the input space. To construct the HAE, a Riemannian metric is employed to map the input data to a Riemannian space. In the HAE, the similarity between points is computed using geodesic distance, whereas in the EAE, Euclidean distance is used. Experimental results demonstrate that the mixed model, incorporating different Riemannian manifolds, performs comparably to other existing methods. Future work will focus on developing deeper mixed networks that combine the advantages of convolutional networks with the benefits of mixed manifold models.

REFERENCES

[1] V. D. Silva and J. B. Tenenbaum, "Global versus local methods in nonlinear dimensionality reduction," *Neural Information Processing Systems*, vol. 15, pp. 721–728, Jan. 2002.

[2] D. L. Donoho and C. Grimes, "Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data," *Proceedings of the National Academy of Sciences*, vol. 100, no. 10, pp. 5591–5596, Apr. 2003, doi: 10.1073/pnas.1031596100.

[3] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM Journal on Scientific Computing*, vol. 26, no. 1, pp. 313–338, Jan. 2004, doi: 10.1137/s1064827502419154

[4] Y. Goldberg, A. Zakai, D. Kushnir, and Y. Ritov, "Manifold learning: the price of normalization," *Journal of Machine Learning Research*, vol. 9, no. 63, pp. 1909–1939, Jun. 200

[5] Y. Li, "Curvature-aware manifold learning," *Pattern Recognition*, vol. 83, pp. 273–286, Jun. 2018, doi: 10.1016/j.patcog.2018.06.007

[6] V. Khrulkov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempitsky, "Hyperbolic image embeddings," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6417–6427, Jun. 2020, doi: 10.1109/cvpr42600.2020.00645.

[7] N. Farajian and P. Adibi, "DMRAE: discriminative manifold regularized auto-encoder for sparse and robust feature learning," *Progress in Artificial Intelligence*, vol. 9, no. 3, pp. 263–274, Jul. 2020, doi: 10.1007/s13748-020-00211-5.

[8] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion," *Journal of Machine Learning Research*, vol. 11, no. 110, pp. 3371–3408, Mar. 2010

[9] W. Liu, T. Ma, D. Tao, and J. You, "HSAE: A Hessian regularized sparse auto-encoders," *Neurocomputing*, vol. 187, pp. 59–65, Nov. 2015, doi: 10.1016/j.neucom.2015.07.119

[10] Y. Shi, M. Lei, R. Ma, and L. Niu, "Learning Robust Auto-Encoders with regularizer for linearity and sparsity," *IEEE Access*, vol. 7, pp. 17195–17206, Jan. 2019, doi: 10.1109/access.2019.2895884.

[11] P. Shamsolmoali, M. Zareapoor, H. Zhou, D. Tao, and X. Li, "VTAE: Variational Transformer Autoencoder with Manifolds learning," *IEEE Transactions on Image Processing*, vol. 32, pp. 4486–4500, Jan. 2023, doi: 10.1109/tip.2023.3299495.

[12] Z. Huang, R. Wang, S. Shan, L. Van Gool, and X. Chen, "Cross Euclidean-to-Riemannian Metric Learning with Application to Face Recognition from Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2827–2840, Nov. 2017, doi: 10.1109/tpami.2017.2776154.

[13] Ç. Gülçehre *et al.*, "Hyperbolic attention networks.," *International Conference on Learning Representations*, May 2018.

[14] H. Tabealhojeh, P. Adibi, H. Karshenas, S. K. Roy, and M. Harandi, "RMAML: Riemannian meta-learning with orthogonality constraints," *Pattern Recognition*, vol. 140, p. 109563, Mar. 2023, doi: 10.1016/j.patcog.2023.109563.

[15] J. Hong, Z. Hayder, J. Han, P. Fang, M. Harandi, and L. Petersson, "Hyperbolic audio-visual zero-shot learning," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2023, doi: 10.1109/iccv51070.2023.00724.

[16] L. Li, Y. Zhang, and S. Wang, "The Euclidean Space is Evil: Hyperbolic Attribute Editing for Few-shot Image Generation," *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22657–22667, Oct. 2023, doi: 10.1109/iccv51070.2023.02076.

[17] M. Salehi *et al.*, "ARAE: Adversarially robust training of autoencoders improves novelty detection," *Neural Networks*, vol. 144, pp. 726–736, Sep. 2021, doi: 10.1016/.