Grassmann Neighborhood Preserving Autoencoder for Image Set Classification

Zeming Chen, Benchao Li, Weixiao Dai, Zicheng Hu, and Ruisheng Ran(✉)

School of Computer and Information Science, Chongqing Normal University,
Chongqing 401331, China
rshran@cqnu.edu.cn

Abstract. Grassmann manifolds have emerged as a powerful tool for high-dimensional data analysis tasks such as image set classification and video action recognition, owing to their ability to mathematically represent collections of linear subspaces. However, the computational complexity of Grassmann manifolds and the challenges faced by traditional dimensionality reduction methods in effectively handling complex nonlinear structures have hindered their widespread application. To address this issue, we propose a novel unsupervised shallow dimensionality reduction method: the Grassmann manifold-based Neighborhood-preserving Autoencoder (GAE-LLE), applied to the task of image set classification. This approach integrates the global dimensionality reduction representation power of the Grassmann autoencoder with the local topology-preserving principle of Neighborhood Embedding. It constructs a dual-constrained optimization framework in the manifold space: on one hand, a neighborhood similarity graph is constructed on the Grassmann manifold using geodesic distances to enforce local structural constraints in the low-dimensional manifold space; on the other hand, the Grassmann autoencoder is employed to preserve the global manifold characteristics of the data. Our comparative experiments demonstrate that the proposed unsupervised shallow dimensionality reduction method outperforms existing traditional dimensionality reduction techniques, significantly improving classification accuracy across multiple image set classification tasks.

Keywords: Manifold learning · Image set classification · Dimensionality reduction · Autoencoder · Neighborhood preservation

1 Introduction

In recent years, Riemannian manifold learning [1] has emerged as a research hotspot due to its effectiveness in handling data with complex geometric structures. Compared to traditional manifold learning methods, Riemannian manifolds naturally capture intrinsic geometric constraints and avoid the information loss that results from forcing data into Euclidean spaces. As a typical Riemannian manifold, the Grassmann manifold [2] has been widely applied in high-dimensional data analysis tasks such as image set classification and video recognition, as it can effectively preserve complex nonlinear

D.-S. Huang et al. (Eds.): ICIC 2025, CCIS 2567, pp. 398–409, 2025.
https://doi.org/10.1007/978-981-96-9952-0_34

relationships among data samples. To enhance manifold modeling capabilities, numerous dimensionality reduction methods based on the Grassmann manifold have been proposed. For example, Harandi et al. introduced Grassmannian discriminant analysis via graph embedding (GEDA) [3] and Grassmann kernels based on canonical correlation [4], both of which achieved promising results in image set classification. However, due to the high computational complexity on the Grassmann manifold, existing methods still struggle with limitations in both efficiency and robustness. To alleviate this problem, researchers have introduced some classical dimensionality reduction methods into the Grassmann manifold, such as multi-kernel clustering with tensor fusion on the Grassmann manifold (MKCTM) [5] and Grassmann discriminant analysis (GKDA) [6], achieving dimensionality reduction and classification through the reproducing kernel Hilbert space (RKHS).

In Euclidean spaces, autoencoders [7] are widely used tools for nonlinear dimensionality reduction. By employing an encoder-decoder structure to extract global features, autoencoders offer a new direction for dimensionality reduction on Grassmann manifold data. However, relying solely on global information is often insufficient in practical applications. The complex local variations and neighborhood relationships within manifold data also play a critical role during dimensionality reduction. To address this, some studies have introduced local structure-preserving strategies, proposing methods such as Grassmann Locality Preserving Projection (GLPP) [8], GNPE [9], and GALL [10]. Among them, GNPE is inspired by Locally Linear Embedding (LLE) [11] and Neighborhood Preserving Embedding (NPE) [12], and models local geometry through linear combination coefficients. Nevertheless, despite LLE achieving remarkable results on manifold data, no existing research has yet integrated the neighborhood-preserving principle of LLE with Grassmann autoencoders. Current Grassmann manifold-based dimensionality reduction methods often focus on preserving either global features or local topology, making it difficult to achieve a balance between the two.

To address this, we propose a new unsupervised shallow dimensionality reduction method: the Grassmann Neighborhood Preserving Autoencoder (GAE-LLE). This method combines the global feature-preserving capability of autoencoders with the local neighborhood-preserving idea of LLE, and constructs a manifold optimization framework with dual constraints. Specifically, we design a neighborhood similarity graph based on geodesic distance to preserve the local topological relationships between samples on top of constructing the Grassmannian autoencoder. We then establish a dual optimization function that combines reconstruction loss and neighborhood-preserving loss, where the reconstruction term employs projection mapping to achieve isometric embedding from the Grassmannian manifold to its tangent space. Through this framework, we achieve an organic integration of global and local structures in Grassmannian manifold data. The main contributions of this work are as follows:

- This paper constructs an autoencoder framework based on the Grassmann manifold and designs a manifold-adaptive reconstruction loss based on the projected Frobenius norm, effectively enhancing the preservation of the global geometric structure of manifold data.
- For the first time, the neighborhood-preserving idea of LLE is introduced into the Grassmann autoencoder, proposing the Grassmann Neighborhood Preserving

Autoencoder (GAE-LLE) method. This method effectively integrates both local and global information, making the reduced-dimensional representation more stable and discriminative.

- The proposed GAE-LLE method performs excellently in image set classification tasks. Compared with various traditional dimensionality reduction methods and deep Grassmann networks, it achieves higher classification accuracy on multiple datasets, demonstrating strong adaptability and stability.

2 Related Work

In this section, we introduce the relevant theoretical foundations, including the definition of the Grassmann manifold and the projection distance.

Grassmann Manifold. The Grassmann manifold is a Riemannian manifold with a non-Euclidean space structure, denoted as $\mathcal{G}(p, D)$, and is generally represented using the orthogonal basis of subspaces. The mathematical definition of the Grassmann manifold is as follows.

Definition. The Grassmann manifold $\mathcal{G}(p, D)$ consists of all p-dimensional linear subspaces of $\mathbb{R}^D$ (where $0 \leq p \leq D$). It is represented by all $D \times p$ orthogonal matrices, where the p orthonormal columns form a quotient space under the orthogonal group $O(p)$:

$$\mathcal{G}(p, D) \triangleq \left\{ \boldsymbol{X} \in \mathbb{R}^{D \times p} : \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{I}_p \right\} / O(p) \tag{1}$$

Any point on the Grassmann manifold $\mathcal{G}(p, D)$ can be represented as a standard orthonormal matrix $\boldsymbol{X}$ of size $D \times p$, which corresponds to a linear subspace span($\boldsymbol{X}$) spanned by an orthonormal basis. It is rigorously defined as: $\mathrm{span}(\boldsymbol{X}) = \left\{ \boldsymbol{X} | \boldsymbol{X}^{\mathrm{T}}\boldsymbol{X} = \boldsymbol{I}_p \right\}$, where $\boldsymbol{I}_p$ represents the $p \times p$ identity matrix.

It is possible to represent the Grassmann manifold within the space of symmetric matrices, as demonstrated below.

$$\mathcal{F} : \boldsymbol{X} \in \mathcal{G}(p, D) \mapsto \boldsymbol{X}\boldsymbol{X}^T \in \mathrm{Sym}(D) \tag{2}$$

Given Grassmann points $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, the projection distance on the Grassmann manifold can be defined as:

$$\mathrm{dist}_g(\boldsymbol{X}_1, \boldsymbol{X}_2) = \frac{1}{\sqrt{2}} \|\mathcal{F}(\boldsymbol{X}_1) - \mathcal{F}(\boldsymbol{X}_2)\|_F \tag{3}$$

3 Proposed Method

In this section, we first introduce a method for constructing a neighborhood similarity graph on the Grassmann manifold based on the projection metric. On this basis, we present the building blocks of a Grassmann manifold-based autoencoder. To address the mismatch between traditional Euclidean loss functions and the geometry of manifold spaces, we derive a manifold reconstruction loss based on the projection Frobenius norm, which better aligns with the geometric structure of Grassmann manifold data. Finally, by integrating these key components, we propose a Grassmann Neighborhood Preserving Autoencoder (GAE-LLE) that effectively captures both global and local structures, thereby improving the performance of manifold-based dimensionality reduction and classification.

3.1 Constructing the Grassmann Neighborhood Similarity Graph

We utilize the Grassmann projection metric Eq. (3) to adopt a local linear representation on the Grassmann manifold. Specifically, given a set of Grassmann samples $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$, where $\boldsymbol{X}_i \in \mathcal{G}(p, D)$, we can express the cost function as follows:

$$\begin{aligned} &\min_{\boldsymbol{C}} \sum_i \left\| \boldsymbol{X}_i \boldsymbol{X}_i^T - \sum_{j \in N_k(\boldsymbol{X}_i)} C_{ij} \boldsymbol{X}_j \boldsymbol{X}_j^T \right\|_F^2 \\ &\text{s.t.} \quad \boldsymbol{C}^T \mathbf{1} = \mathbf{1} \end{aligned} \tag{4}$$

where $\mathbf{1} \in \mathbb{R}^{n \times 1}$ is a column vector with all entries equal to 1, $N_k(\boldsymbol{X}_i)$ denotes the set of k nearest neighbors of $\boldsymbol{X}_i$ based on the embedding distance Eq. (3) and the coefficient C_{ij} represents the contribution of the j-th Grassmann sample in reconstructing the i-th Grassmann sample.

The main steps for constructing the similarity graph $\boldsymbol{C}$ are given below. We define $\boldsymbol{Z}_i = \boldsymbol{X}_i \boldsymbol{X}_i^T$, $\boldsymbol{Z}_j = \boldsymbol{X}_j \boldsymbol{X}_j^T$, thus we have:

$$\min_{\boldsymbol{C}} \sum_i \left\| \boldsymbol{Z}_i - \sum_{j \in N_k(\boldsymbol{X}_i)} C_{ij} \boldsymbol{Z}_j \right\|_F^2 \tag{5}$$

For a given i, and $\sum_{j=1}^{k} C_j = 1$, we have:

$$\begin{aligned} &\min \left\| \boldsymbol{Z} - \sum_{j=1}^{k} C_j \boldsymbol{Z}_j \right\|_F^2 \\ &= \min \sum_{p=1}^{k} \sum_{q=1}^{k} C_p C_q tr((\boldsymbol{Z} - \boldsymbol{Z}_p)(\boldsymbol{Z} - \boldsymbol{Z}_q)^T) \end{aligned} \tag{6}$$

Let $S_{pq} = tr((\boldsymbol{Z} - \boldsymbol{Z}_p)(\boldsymbol{Z} - \boldsymbol{Z}_q)^T)$, then the above expression becomes $\min \sum_{p=1}^{k} \sum_{q=1}^{k} C_p C_q S_{pq}$.

Using the method of Lagrange multipliers, we have:

$$L(C, \lambda) = \sum_{p=1}^{k} \sum_{q=1}^{k} C_p C_q S_{pq} - \lambda(\sum_{j=1}^{k} C_j - 1) \tag{7}$$

From this, we obtain $\boldsymbol{C}$. In general, $\boldsymbol{C}$ is a singular matrix, but after regularization, it is constructed to ensure that it is non-singular. By introducing a parameter τ, we get $\boldsymbol{C} = \boldsymbol{C} + \tau \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix.

3.2 Grassmann Autoencoder

This section introduces the constructed shallow Grassmann manifold autoencoder. The core idea is to model image sets as Grassmann manifold data, achieving feature compression and reconstruction through an encoding-decoding architecture under manifold geometric constraints. Finally, a Grassmann KNN classifier is employed to classify the more discriminative low-dimensional manifold data.

Encoding Phase. Given a set of Grassmann samples $\boldsymbol{X} = \{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_n\}$, where $\boldsymbol{X}_i \in \mathcal{G}(p, D)$, we need to learn a mapping $\mathcal{G}(p, D) \rightarrow \mathcal{G}(p, d)$, with $D > d$, as shown below:

$$\boldsymbol{Z}_i = \boldsymbol{A}^T \boldsymbol{X}_i \tag{8}$$

let $\boldsymbol{A} \in \mathbb{R}^{D \times d}$. Since the latent space of this weight matrix lies on a non-compact Stiefel manifold, the geodesic distance is unbounded and thus cannot be directly optimized. According to [13], we impose an orthogonality constraint on it, transforming its space into a compact Stiefel manifold. Meanwhile, to ensure that $\boldsymbol{Z}_i \in \mathbb{R}^{d \times p}$ can serve as a representative point on the low-dimensional Grassmann manifold, certain conditions need to be satisfied. To address this issue, we employ QR decomposition, which to some extent plays the role of a nonlinear activation.

$$\begin{aligned} \boldsymbol{Z}_i &= \boldsymbol{A}^T \boldsymbol{X}_i = \boldsymbol{Q}_i \boldsymbol{R}_i \\ \boldsymbol{Y}_i &= \boldsymbol{Q}_i = \boldsymbol{A}^T (\boldsymbol{X}_i \boldsymbol{R}_i^{-1}) \end{aligned} \tag{9}$$

here, $\boldsymbol{Q}_i \in \mathbb{R}^{d \times p}$ is an orthogonal matrix, and $\boldsymbol{R}_i \in \mathbb{R}^{p \times p}$ is an upper triangular matrix. The resulting Grassmannian low-dimensional embedded sample point is denoted as $\boldsymbol{Y}_i$. We refer to Eq. (8) as the Full-Rank Mapping Layer (FRMap Layer), and Eq. (9) as the Re-orthogonalization Layer (ReOrth Layer) [13].

Decoding Phase. The weights in the decoding phase are transposed with those in the encoding phase, i.e., $(\boldsymbol{A}^T)^T = \boldsymbol{A}$. The decoding phase first passes through the FRMap Layer, which remaps the low-dimensional manifold data back to the high-dimensional manifold space: $\mathcal{G}(p, d) \rightarrow \mathcal{G}(p, D)$, where $D > d$. The specific representation is as follows:

$$\hat{\boldsymbol{Z}}_i = \boldsymbol{A} \boldsymbol{Y}_i \tag{10}$$

here, $\boldsymbol{A} \in \mathbb{R}^{D \times d}$. Similarly, since $\hat{\boldsymbol{Z}}_i$ is not an orthogonal matrix, it cannot serve as a representative point on the Grassmann manifold. We still need to apply the ReOrth Layer to obtain the final reconstructed data, which is expressed as follows:

$$\begin{aligned} \hat{\boldsymbol{Z}}_i &= \boldsymbol{A} \boldsymbol{Y}_i = \hat{\boldsymbol{Q}}_i \hat{\boldsymbol{R}}_i \\ \hat{\boldsymbol{X}}_i &= \hat{\boldsymbol{Q}}_i = \boldsymbol{A} (\boldsymbol{Y}_i \hat{\boldsymbol{R}}_i^{-1}) \end{aligned} \tag{11}$$

In order to make the loss function adapt to the geometric properties of the Grassmann manifold data, we generalize the traditional mean squared error loss function to the Grassmann manifold space based on the projection metric of the Grassmann manifold in Eq. (3). We derive the Grassmann manifold reconstruction loss function with differential geometric properties, which is mathematically expressed as:

$$\mathcal{L}_{GAE} = \frac{1}{2N} \sum_{i=1}^{N} \left\| \boldsymbol{X}_i \boldsymbol{X}_i^T - \hat{\boldsymbol{X}}_i \hat{\boldsymbol{X}}_i^T \right\|_F^2 \tag{12}$$

where $\boldsymbol{X}_i$ represents the original sample point, and $\hat{\boldsymbol{X}}_i$ represents the reconstructed sample point obtained through the Grassmann autoencoder. This loss directly measures the difference between the subspace projection matrices on the manifold, and its geometric nature is equivalent to calculating the tangent space distance between two points, thus avoiding the metric bias caused by manifold curvature in traditional loss.

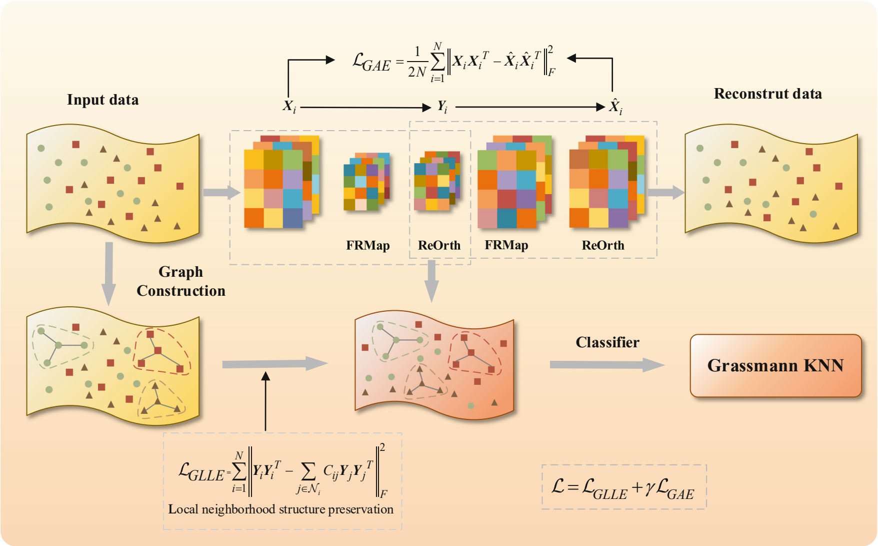
$\mathcal{L}_{GAE} = \frac{1}{2N}\sum_{i=1}^{N}\left\|\boldsymbol{X}_i\boldsymbol{X}_i^T - \hat{\boldsymbol{X}}_i\hat{\boldsymbol{X}}_i^T\right\|_F^2$
Input data
$\boldsymbol{X}_i$
$\boldsymbol{Y}_i$
$\hat{\boldsymbol{X}}_i$
Reconstrut data
FRMap
ReOrth
FRMap
ReOrth
Graph Construction
Classifier
Grassmann KNN
$\mathcal{L}_{GLLE} = \sum_{i=1}^{N}\left\|\boldsymbol{Y}_i\boldsymbol{Y}_i^T - \sum_{j\in\mathcal{N}_i} C_{ij}\boldsymbol{Y}_j\boldsymbol{Y}_j^T\right\|_F^2$
Local neighborhood structure preservation
$\mathcal{L} = \mathcal{L}_{GLLE} + \gamma\mathcal{L}_{GAE}$

Fig. 1. The proposed Grassmann Neighborhood Preserving Autoencoder consists of FRMap and ReOrth layers. Its loss function comprises two components: (1) the GAE loss, which measures the reconstruction error on the Grassmann manifold; and (2) the GLLE loss, which constructs a similarity graph on the original manifold to constrain the low-dimensional embedding to preserve the original neighborhood relationships. The model learns discriminative low-dimensional manifold features and performs classification using a Grassmann KNN classifier.

3.3 Grassmann Neighborhood Preserving Autoencoder

The model diagram of the GAE-LLE method is shown in Fig. 1. We model the image set data as Grassmann matrix manifold data and use this shallow dimensionality reduction method to achieve neighborhood-preserving dimensionality reduction on the Grassmann manifold data. Finally, we use the low-dimensional data from the hidden layer for classification (the classifier uses GKNN).

To ensure that the low-dimensional manifold data preserves the original high-dimensional manifold data's neighborhood relationships, we introduce manifold local geometric constraints during the encoding phase, and refer to this part as GLLE:

$$\mathcal{L}_{GLLE}=\sum_{i=1}^{N}\left\|\boldsymbol{Y}_i\boldsymbol{Y}_i^T-\sum_{j\in\mathcal{N}_i}C_{ij}\boldsymbol{Y}_j\boldsymbol{Y}_j^T\right\|_F^2 \tag{13}$$

where $\boldsymbol{Y}_i$ represents the i-th sample mapped to the low-dimensional space. By embedding the sample into the symmetric matrix space using Eq. (3) the geodesic distance can be effectively approximated.

Finally, the loss function of the model consists of two parts, as shown in the following formula, where γ is the balancing parameter:

$$\mathcal{L}=\mathcal{L}_{GLLE}+\gamma\mathcal{L}_{GAE} \tag{14}$$

By optimizing the total loss, an optimal projection matrix $\boldsymbol{A}$ is obtained, enabling the original Grassmann manifold data to be mapped onto a more discriminative low-dimensional manifold. Both the global geometric structure consistency and the local adjacency relationships are well preserved.

4 Experiments

This section provides an overview of the Grassmann datasets employed in our experiments, which encompass both image-based and video-based collections. We then compare several commonly used dimensionality reduction methods on the Grassmann manifold, including Grassmannian K-Nearest Neighbors (GKNN), Support Vector Machine-based methods (GSVM [14], GKDA), and various manifold learning approaches such as GLPP, GNPE, GALL, and Nested Grassmann (NG) [15]. In addition, our proposed unsupervised shallow dimensionality reduction method is compared with the representative deep Grassmann network, GrNet. Among these methods, the implementations of GSVM, NG, and GrNet are publicly available. For the rest of the methods, since official source codes are not publicly available, we implemented them ourselves by closely following the pseudo-code and algorithm descriptions from their original publications.

4.1 Dataset Description

- **Ballet Dataset** [16]: It contains 44 instructional DVD videos, covering 8 complex racket actions demonstrated by 3 performers. The intra-class variations (including speed, spatiotemporal scale, clothing, and motion amplitude) pose challenges to the classification task.
- **ETH-80 Dataset** [17]: This dataset contains 8 categories of objects, such as cows, cups, and horses. Each category consists of 10 subcategories (forming image set samples), and each sample is composed of 41 multi-view images.

- **Traffic Video Database** [18]: This dataset includes 254 traffic video sequences recorded by stationary cameras, representing different traffic conditions and weather scenarios. Each video is captured at 10 frames per second, with each frame transformed into a 20 × 20 grayscale image. The videos are treated as samples of image sets.
- **UCF-S Dataset** [19]: The dataset consists of 150 video sequences containing 13 action categories, covering a variety of scenes and viewpoints. Each video contains between 22 and 144 frames, which are converted into grayscale images.
- **UT-Kinect Datase**t [20]: The dataset includes video, depth sequences, and skeleton data of 10 action categories, collected from 10 subjects using Kinect, with a total of 200 sequences.

Table 1. Datasets and Samples Information.

Datasets	Samples	Training Set	Test Set	Categories	$\mathcal{G}(p, D)$	$\mathcal{G}(p, d)$
Ballet	59	30	29	8	$\mathcal{G}(8, 400)$	$\mathcal{G}(8, 50)$
ETH-80	80	40	40	8	$\mathcal{G}(11, 400)$	$\mathcal{G}(11, 50)$
Traffic	254	127	127	3	$\mathcal{G}(15, 400)$	$\mathcal{G}(15, 50)$
UCF-S	150	75	75	13	$\mathcal{G}(9, 400)$	$\mathcal{G}(9, 50)$
UT-Kinect	199	100	99	10	$\mathcal{G}(8, 400)$	$\mathcal{G}(8, 50)$

Table 1 presents the key information of the datasets we used, where $\mathcal{G}(p, D)$ represents the original high-dimensional space. In both the proposed method and the comparison methods, the reduced dimensionality d is uniformly set to 50.

4.2 Experimental Comparison

Comparison with Other Methods. We performed a comparative evaluation of the classification performance between the proposed GAE-LLE method and other techniques across five datasets, with the corresponding accuracy results presented in Table 2. The results indicate that GAE-LLE consistently offers strong classification performance. On the Ballet dataset, GAE-LLE achieves an accuracy of 78.57%, notably surpassing other methods, which suggests its superior capability in modeling intra-class variations within complex actions. In the ETH-80 multi-view classification task, GAE-LLE achieves a perfect accuracy of 100%, improving by 12.5% over the best-performing existing methods (GrNet and GNPE, both at 87.50% accuracy), illustrating its proficiency in capturing the underlying manifold structure of multi-view features. On the Traffic dataset, GAE-LLE leads with an accuracy of 94.44%. For the UCF-S and UT-Kinect datasets, GAE-LLE attains accuracies of 49.32% and 52.53%, respectively. The former shows a 2.74% improvement over the second-best method, and although the latter is slightly below GrNet, it outperforms most other Grassmannian methods, indicating its effectiveness in dynamic action recognition. Overall, GAE-LLE excels at balancing both local geometric structure and global discriminative features in complex datasets, resulting in more accurate and reliable classification outcomes.

Table 2. Accuracy Comparison.

Method	Ballet	ETH-80	Traffic	UCF-S	UT-Kinect
GKNN	0.6429	0.8500	0.8889	0.4521	0.4343
GSVM	0.4286	0.8500	0.7778	0.1781	0.4444
GKDA	0.5714	0.8500	0.8968	0.4110	0.4848
GLPP	0.5714	0.8000	0.8968	0.3288	0.4646
GNPE-I	0.5357	0.8750	0.9206	0.4521	0.4848
GNPE-II	0.5357	0.8750	0.8571	0.4521	0.4747
GALL	0.6071	0.7500	0.8810	0.3973	0.2424
NG	0.5357	0.8250	0.9127	0.4658	0.3939
GrNet	0.6429	0.8750	0.8493	0.4658	**0.5353**
GAE-LLE	**0.7857**	**1.0000**	**0.9444**	**0.4932**	0.5253

To provide a comprehensive assessment of each method, we compared three evaluation metrics in addition to accuracy: Precision (PRE), F1 score, and Recall (REC). Figure 2 illustrates the results of these metrics on the Ballet, UT-Kinect, and ETH-80 datasets. Each subplot corresponds to one metric, where the X-axis indicates the method, the Y-axis denotes the dataset, and the Z-axis shows the corresponding score. Different bar colors represent different datasets, and the bar heights clearly reflect the performance variations across methods.

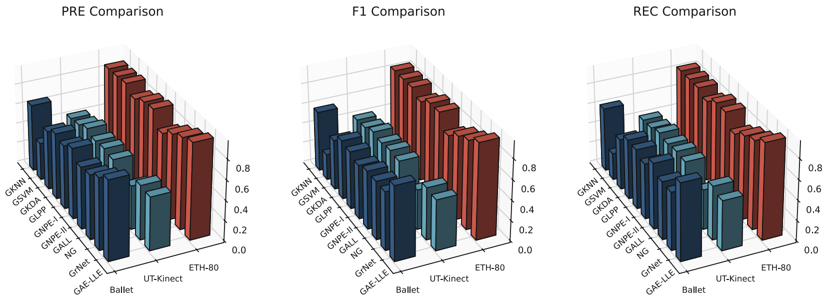

Fig. 2. Comparison of Precision, F1 Score, and Recall Across Different Methods.

As can be seen from the figure, GAE-LLE outperforms in all three metrics. Particularly in ETH-80, GAE-LLE leads with 95.42% (PRE), 94.97% (F1), and 95.00% (REC), demonstrating its outstanding performance in multi-view classification tasks.

In the Ballet dataset, GAE-LLE also significantly outperforms other methods in all metrics, reflecting its strong ability to model complex actions. In UT-Kinect, although GAE-LLE is slightly lower than GrNet, it outperforms most Grassmann methods in all three metrics, showing good competitiveness. Overall, GAE-LLE achieves leading or near-optimal performance in multiple tasks, validating its stability and generalization ability.

Ablation Study. To further analyze the effectiveness of each component in the proposed GAE-LLE method, we conducted ablation studies. The GAE-LLE framework consists of two modules: GAE and GLLE. We separately compare the performance of three configurations on the aforementioned datasets: using only the GAE module, using only the GLLE module, and the complete GAE-LLE method that combines both. This comparison allows us to clearly identify the contribution of each module to the final performance, thereby providing a better understanding of their roles and importance within the GAE-LLE method.

Table 3. Ablation Study: Accuracy Comparison.

Method	Ballet	ETH-80	Traffic	UCF-S	UT-Kinect
GAE	0.7143	0.8750	0.9206	0.4795	0.4444
GLLE	0.7857	0.9000	0.9127	0.4521	0.4949
GAE-LLE	**0.7857**	**1.0000**	**0.9444**	**0.4932**	**0.5253**

As shown in Table 3, GAE-LLE generally outperforms using GAE or GLLE alone on all datasets, with the best performance observed on ETH-80, surpassing both individual modules. Although the improvements on other datasets are relatively small, GAE-LLE still shows a clear advantage on the Traffic and UT-Kinect datasets. GAE helps capture the global structure, while GLLE enhances local geometric features, and the combination of both modules improves classification performance. Notably, the image set classification accuracy of the proposed modules (GAE and GLLE) also demonstrates a clear advantage over other traditional dimensionality reduction methods.

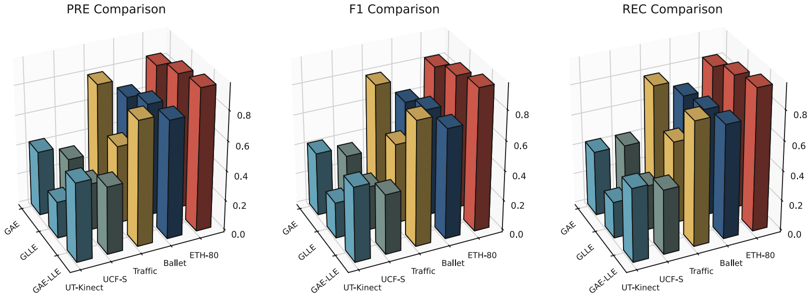

Fig. 3. Ablation Study: Comparison of Precision, F1 Score, and Recall.

In addition to accuracy, we further compared precision (PRE), F1 score, and recall (REC) to verify the effectiveness of the proposed method. Figure 3 shows the performance of three methods (GAE, GLLE, GAE-LLE) on five datasets (UT-Kinect, UCF-S, Traffic, Ballet, ETH-80). From the perspective of the three metrics, the GAE method performs better than GLLE on the Traffic, UCF-S, and UT-Kinect datasets, but worse on the Ballet and ETH-80 datasets. The GLLE method shows better performance on the Ballet and ETH-80 datasets, but performs worse on the other three datasets. Finally, GAE-LLE outperforms both GAE and GLLE across all three metrics. It demonstrates a clear advantage on all datasets, indicating that GAE-LLE combines the benefits of both the autoencoder and neighborhood-preserving constraints, making it suitable for diverse datasets. This validates its effectiveness and advantages in image set classification tasks.

5 Conclusion

This paper proposes an unsupervised shallow dimensionality reduction method, GAE-LLE, which constructs an autoencoder architecture tailored to the Grassmann manifold and introduces a manifold reconstruction loss based on the projection metric. Additionally, a method for constructing a neighborhood similarity graph on the Grassmann manifold using the projection metric is presented. For the first time, the neighborhood preservation concept of LLE is organically integrated with the Grassmann autoencoder, resulting in a dual-constrained optimization framework that achieves coordinated optimization of global geometric structure and local topological relationships. However, this method is not yet applicable to large-scale datasets, primarily because the similarity graph construction based on the projection metric incurs significant computational overhead when the sample size is large, limiting the model's scalability. In the future, the graph construction efficiency and the applicability of the method could be improved by incorporating graph approximation algorithms or index-based fast neighborhood search techniques.

References

1. Lin, T., Zha, H.: Riemannian manifold learning. IEEE Trans. Pattern Anal. Mach. Intell. **30**(5), 796–809 (2008)
2. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM J. Matrix Anal. Appl. **20**(2), 303–353 (1998)
3. Harandi, M.T., Sanderson, C., Shirazi, S., Lovell, B.C.: Graph embedding discriminant anaysis on Grassmannian manifolds for improved image set matching. In: CVPR 2011, pp. 2705–2712. IEEE (2011)
4. Harandi, M.T., Salzmann, M., Hartley, R.: From manifold to manifold: geometry- aware dimensionality reduction for SPD matrices. In: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6–12, 2014, Proceedings, Part II 13, pp. 17–32. Springer, Cham (2014)
5. Qi, F., et al.: Multi-kernel clustering with tensor fusion on Grassmann manifold for high-dimensional genomic data. Methods **231**, 215–225 (2024)
6. Hamm, J., Lee, D.D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: Proceedings of the 25th International Conference on Machine Learning, pp. 376–383 (2008)

7. Hinton, G.E., Zemel, R.: Autoencoders, minimum description length and Helmholtz free energy. In: Advances in Neural Information Processing Systems, vol. 6 (1993)
8. Wang, B., Hu, Y., Gao, J., Sun, Y., Chen, H., Yin, B.: Locality preserving projections for Grassmann manifold. arXiv preprint arXiv:1704.08458 (2017)
9. Wei, D., Shen, X., Sun, Q., Gao, X., Ren, Z.: Neighborhood preserving embedding on Grassmann manifold for image-set analysis. Pattern Recogn. **122**, 108335 (2022)
10. Wei, D., Shen, X., Sun, Q., Gao, X., Ren, Z.: Learning adaptive Grassmann neighbors for image-set analysis. Expert Syst. Appl. **247**, 123316 (2024)
11. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
12. He, X., Cai, D., Yan, S., Zhang, H.J.: Neighborhood preserving embedding. In: Tenth IEEE International Conference on Computer Vision (ICCV 2005), vol. 2, pp. 1208–1213. IEEE (2005)
13. Huang, Z., Wu, J., Van Gool, L.: Building deep networks on Grassmann manifolds. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32 (2018)
14. Al-Samhi, W., Al-Soswa, M., Al-Dhabi, Y.: Time series data classification on Grassmann manifold. In: Journal of Physics: Conference Series, vol. 1848, p. 012037. IOP Publishing (2021)
15. Yang, C.H., Vemuri, B.C.: Nested Grassmanns for dimensionality reduction with applications to shape analysis. In: International Conference on Information Processing in Medical Imaging, pp. 136–149. Springer, Cham (2021)
16. Wang, Y., Mori, G.: Human action recognition by semilatent topic models. IEEE Trans. Pattern Anal. Mach. Intell. **31**(10), 1762–1774 (2009)
17. Leibe, B., Schiele, B.: Analyzing appearance and contour based methods for object categorization. In: Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. II–409. IEEE (2003)
18. Chan, A.B., Vasconcelos, N.: Probabilistic kernels for the classification of auto- regressive visual processes. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005), vol. 1, pp. 846–851. IEEE (2005)
19. Rodriguez, M.D., Ahmed, J., Shah, M.: Action mach a spatio-temporal maximum average correlation height filter for action recognition. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8. IEEE (2008)
20. Xia, L., Chen, C.C., Aggarwal, J.K.: View invariant human action recognition using histograms of 3D joints. In: 2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 20–27. IEEE (2012)