LEARNING DISTRIBUTIONS OF GRASSMANN SUBSPACE COMBINATIONS FOR EFFECTIVE GEOMETRIC REPRESENTATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Grassmann manifolds provide a principled framework for modelling data as low-dimensional subspaces, capturing intrinsic geometric structure and achieving considerable success in geometric representation learning. Existing approaches, however, typically treat all inner subspace combinations within a Grassmann point as equally important, overlooking their non-uniform distribution. To address this issue, we theoretically characterise the distribution of subspace combinations within a Grassmann point based on spectral measure theory and quadratic utility theory, revealing that different combinations contribute non-uniformly to representation. This insight suggests that Grassmann representation learning should adaptively model such distributions and exploit compatibility among low-dimensional subspaces to select and construct effective combinations. To this end, we propose a Graph-Structured adaptive SubSpace (GraSS) modelling on Grassmann manifolds. GraSS builds a relational graph among constituent subspaces to jointly capture their importance and structural compatibility. Based on the constructed graph, we identify the maximum connected subgraph with high coherence and accordingly adaptively determine the mutually compatible constituent subspaces and their composition. Furthermore, we introduce an adaptive gating selection mechanism to dynamically filter the less powerful combinations, suppressing redundant combinations while emphasising those with higher representation utility. Finally, we apply GraSS to multiple downstream tasks. Extensive experiments show that GraSS consistently improves existing Grassmann representation learning methods across tasks and exhibits strong generalisation.

1 INTRODUCTION

In recent years, subspace modelling on the Grassmannian has achieved considerable success in tasks requiring strong discriminative capacity, such as activity recognition (Cherian et al., 2017), emotion recognition (Liu et al., 2014; Wang et al., 2023), face verification (Huang et al., 2015; Chen et al., 2021), and time-series classification in brain computer interfaces (Gao et al., 2022). In these tasks, existing Grassmannian methods typically start from a set of orthogonal representations and construct one or more fixed-dimensional subspaces to characterise the geometric structure of the input data (Huang et al., 2018; Wang et al., 2024; Xu et al., 2023; Chen et al., 2025; Yu & Xu, 2026), as illustrated in Fig. 1. However, such subspace modelling treats all subspace combinations within a Grassmann point as equally important (Hamm & Lee, 2008), thereby overlooking the non-uniform distribution of combinations formed by these constituent subspaces within the Grassmann point.

To address the above limitation, we adopt a distributional perspective and investigate the latent combinatorial distribution among the constituent subspaces within a Grassmann point. In recent years, studies in representation learning have increasingly shown that modelling the non-uniformity within representations can reveal structural information overlooked by homogeneous modelling. For example, Distribution-aware Knowledge Prototyping (Xu et al., 2024) and Continual-MAE (Liu et al., 2024) explicitly account for distributional variations within representations, ultimately leading to richer and more task-relevant representations. Inspired by these studies, we theoretically characterise the subspace combination distribution within a Grassmann point based on spectral measure

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

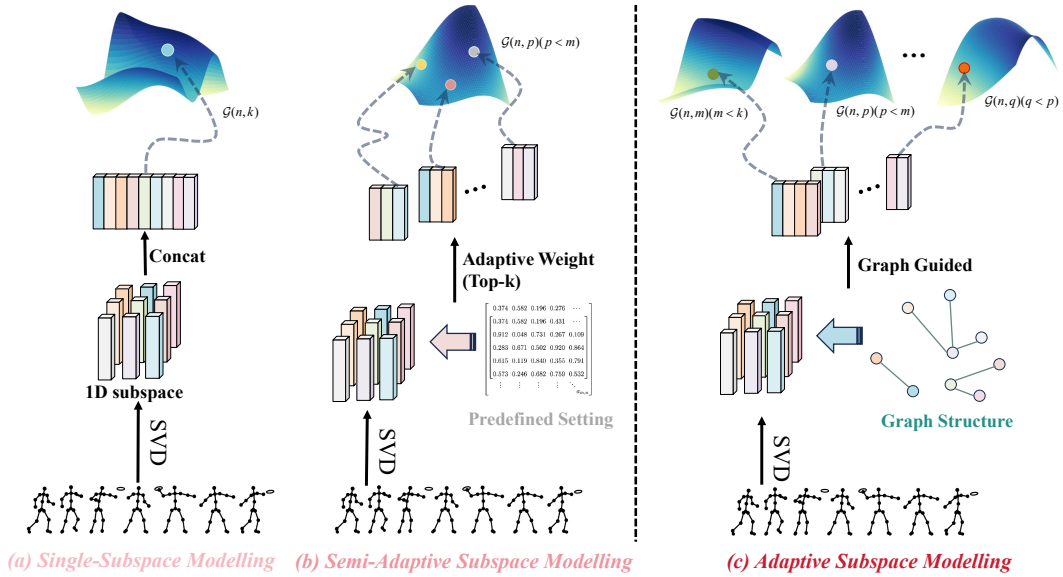


Figure 1: **Illustration of Grassmannian subspace modelling.** (a) illustrates the semantic conflicts arising from single-subspace modelling. (b) reveals the dependence of multi-subspace modelling on predefined subspace dimensions and numbers. (c) presents the proposed Graph-Structured adaptive SubSpace (GraSS) modelling framework.

theory (Reed & Simon, 1972) and quadratic utility theory (Chew et al., 1991). Furthermore, we demonstrate that different subspace combinations contribute non-uniformly to the representation. This theory provides a theoretical basis for understanding the non-uniform combinatorial relationships among constituent subspaces on the Grassmannian. Therefore, effective Grassmannian representations require adaptively modelling the subspace combination distribution.

Existing studies have also begun to explore the modelling and combination of constituent subspaces within a Grassmann point to obtain more flexible geometric representations. Early approaches typically model a Grassmann point as a unified holistic subspace, GrNet(Huang et al., 2018) and GDLNet (Wang et al., 2024) directly perform geometric transformations and representation learning on the complete subspace. More recently, GMSF-Net (Yu & Xu, 2026) introduces adaptive weights for different constituent subspaces and selects the Top-k subspaces with the highest weights to construct a more coherent multi-subspace representation. However, existing adaptive approaches typically still require the dimension of the subspaces and the number of selected subspaces to be specified in advance (Cai et al., 2024), thereby imposing manually defined priors on subspace selection and composition. Moreover, these methods primarily select constituent subspaces according to their individual importance, without fully exploiting the non-uniform distributional characteristics of subspace combinations within a Grassmann point. Therefore, a natural question arises: *How can we adaptively uncover the non-uniform distribution of subspace combinations within a Grassmann point, thereby obtaining more effective Grassmannian geometric representations?*

To answer this question, we propose a **Graph-Structured adaptive SubSpace (GraSS)** modelling framework for learning subspace combination distributions on the Grassmannian. Specifically, GraSS adaptively constructs a relational graph among constituent subspaces to characterise their geometric contributions and structural compatibility, thereby uncovering potentially effective subspace combinations within a Grassmann point. Based on the constructed graph, we further identify the maximum connected subgraph with high structural coherence, and accordingly adaptively determine mutually compatible constituent subspaces and their composition, thereby overcoming the fixed subspace dimensionality and cardinality constraints imposed by conventional Grassmannian subspace modelling. Furthermore, we also introduce an adaptive gating selection mechanism to dynamically adjust the contributions of different subspace combinations according to their representation utility, suppressing redundant combinations while retaining more discriminative ones, thereby yielding more effective geometric representations.

We summarise the main contributions as follows:

- We propose Graph-Structured adaptive SubSpace (GraSS), which adaptively constructs a relational graph among constituent subspaces to learn the subspace combination distribution within a Grassmann point.
- We provide the first theoretical characterisation of the non-uniform intrinsic distribution of subspace compositions within a Grassmann point based on spectral measure theory and quadratic utility theory.
- We conduct extensive experiments on three datasets across multiple downstream tasks, demonstrating that GraSS consistently improves existing Grassmannian representation learning approaches.

2 Related Works

The Grassmannian, denoted by $\mathcal{G}(n, p)$, is the set of all p-dimensional linear subspaces of $\mathbb{R}^n$, providing a natural geometric space for representing data with intrinsic subspace structures (Souza et al., 2020; Zhang et al., 2018; Baudat & Anouar, 2000; Von Luxburg, 2007; Dai et al., 2012). In recent years, deep Grassmannian models have been further developed to learn more discriminative nonlinear representations while preserving the underlying geometric structure of the manifold. GrNet (Huang et al., 2018) constructs a deep network on the Grassmannian by introducing geometry-aware operations, including FRMap, OrthMap, and ProjMap, to directly transform and learn subspace representations on the manifold. GDLNet (Wang et al., 2024) further incorporates a self-attention mechanism to model dependencies among different subspaces and enhance the representation of globally relevant geometric information (Vaswani et al., 2017). Despite their differences in network architectures and geometric operations, these methods generally follow a unified modelling paradigm, where a single and static subspace representation is adopted to model the original global features, treating each Grassmann point as a complete geometric object. Such unified modelling does not explicitly distinguish the contributions of different constituent subspaces within a Grassmann point, thereby overlooking the influence of constituent subspaces on the final representation.

More recently, GMSF-Net (Yu & Xu, 2026) has begun to investigate the diversity among different low-dimensional subspaces and explicitly distinguish their contributions to the final representation. Specifically, GMSF-Net assigns an individual importance weight to each low-dimensional subspace and selects the Top-K low-dimensional subspaces according to their weights to construct multiple more coherent Grassmannian representations. This strategy enables the model to adaptively focus on more important subspaces according to the input data. However, its selection process still primarily relies on the individual importance of constituent subspaces, thereby overlooking the non-uniform contributions of different combinations formed by these constituent subspaces to the representation. To address the above limitations, we propose GraSS, which adaptively constructs a relational graph among constituent subspaces, thereby enabling the discovery of their non-uniform distribution on the Grassmannian.

3 Preliminaries

3.1 Grassmannian

The Grassmann manifold $\mathcal{G}(n, p)$ is the set of all p-dimensional linear subspaces of $\mathbb{R}^n$:

$$\mathcal{G}(n, p) = \{\mathcal{S} \subset \mathbb{R}^n \mid \dim(\mathcal{S}) = p\}. \tag{1}$$

A Grassmann point $\mathcal{S} \in \mathcal{G}(n, p)$ can be represented by an orthonormal basis $U \in \mathbb{R}^{n\times p}$ satisfying $U^\top U = I_p$, with $\mathcal{S} = \mathrm{span}(U)$. For two Grassmann points represented by $U_1, U_2 \in \mathbb{R}^{n\times p}$, we adopt the projection metric (Huang et al., 2015):

$$d_{\mathrm{p}}(U_1, U_2) = \frac{1}{\sqrt{2}} \left\| U_1U_1^\top - U_2U_2^\top \right\|_F. \tag{2}$$

Since $(UR)(UR)^\top = UU^\top$, this metric is invariant to the choice of basis and therefore measures the intrinsic geometric discrepancy between the underlying subspaces.

3.2 Spectral Measure

Let $X \in \mathbb{R}^{n\times d}$ be a matrix with singular value decomposition $X = U\Sigma V^\top$, where $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$ are the singular values. For the leading K singular components, the normalised spectral weight of the i-th component is defined as

$$p_i = \frac{\sigma_i^2}{\sum_{j=1}^K \sigma_j^2}, \qquad i = 1, \ldots, K. \tag{3}$$

Let $P_i = u_i u_i^\top$ denote the corresponding rank-one projection. The spectral measure of X is defined:

$$\mu_X = \sum_{i=1}^K p_i \delta_{P_i}, \tag{4}$$

where δ_{P_i} denotes the Dirac measure concentrated at P_i. Thus, μ_X characterises the distribution of spectral energy across the leading singular components.

3.3 Quadratic Utility

Let $\phi_i = \phi(P_i) \in \mathbb{R}^m$ denote the representation of the i-th constituent component, and define its weighted representation as $g_i = \sqrt{p_i}\phi_i$. For any subset $S \subseteq \{1, \ldots, K\}$, let $z_S = \sum_{i\in S} g_i$. Given a positive semidefinite matrix $M \succeq 0$, the quadratic utility of S is defined by

$$\mathcal{R}(S) = z_S^\top M z_S. \tag{5}$$

Equivalently, defining $a_i = \phi_i^\top M \phi_i$ and $b_{ij} = \phi_i^\top M \phi_j$, we have

$$\mathcal{R}(S) = \sum_{i\in S} p_i a_i + 2 \sum_{\substack{i,j\in S \\ i<j}} \sqrt{p_i p_j}\, b_{ij}. \tag{6}$$

4 Theoretical Foundations of Grassmannian Subspace Distributions

Proposition 1 (Spectral Distribution)**.** *For each constituent direction u_i, define its sign-invariant projection $P_i = u_i u_i^\top$ and spectral weight*

$$p_i = \frac{\sigma_i^2}{\sum_{j=1}^K \sigma_j^2}, \qquad i = 1, \ldots, K. \tag{7}$$

Then the SVD induces the discrete spectral distribution $\mu_{\mathcal{U}} = \sum_{i=1}^K p_i \delta_{P_i}$, where δ_{P_i} is the Dirac measure centred at P_i.

Theorem 1 (Utility Decomposition)**.** *Let $\phi_i = \phi(P_i) \in \mathbb{R}^m$ and $M \succeq 0$. For $S \subseteq \{1, \ldots, K\}$, define*

$$\mathcal{R}(S) = \left(\sum_{i\in S} \sqrt{p_i}\phi_i\right)^\top M \left(\sum_{i\in S} \sqrt{p_i}\phi_i\right). \tag{8}$$

Then

$$\mathcal{R}(S) = \sum_{i\in S} p_i a_i + 2 \sum_{\substack{i,j\in S \\ i<j}} \sqrt{p_i p_j}\, b_{ij}, \tag{9}$$

where

$$a_i = \phi_i^\top M \phi_i, \qquad b_{ij} = \phi_i^\top M \phi_j. \tag{10}$$

Proposition 2 (Importance and Compatibility)**.** *For two constituent directions i and j, the marginal utility of adding j to i is*

$$\Delta(j \mid \{i\}) = p_j a_j + 2\sqrt{p_i p_j}\, b_{ij}. \tag{11}$$

Thus, even if $p_j a_j > 0$, we have

$$\Delta(j \mid \{i\}) < 0 \quad \textit{if} \quad b_{ij} < -\frac{1}{2}\sqrt{\frac{p_j}{p_i}}\, a_j. \tag{12}$$

Hence, individual importance alone does not guarantee the utility of joint composition, which is jointly determined by spectral importance and pairwise compatibility.

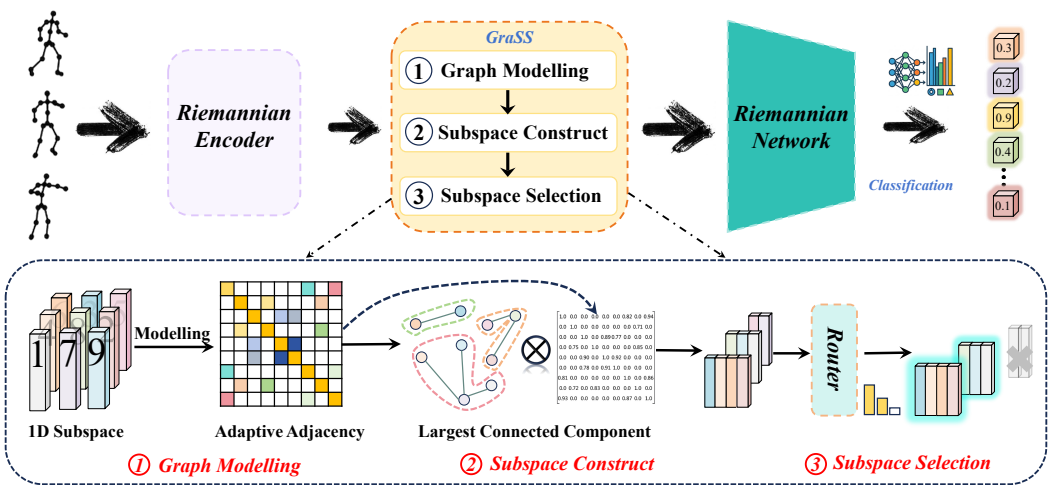


Figure 2: **The GraSS pipeline for adaptive subspace modelling on Grassmann manifolds.** GraSS consists of three steps: (1) **Graph Modelling** adaptively learns relationships among constituent subspaces to characterise their combinatorial distribution structure. (2) **Subspace Construction** aggregates highly coherent subspaces based on maximal connected subgraphs to form effective composite geometric representations. (3) **Adaptive Subspace Selection** evaluates the importance of composite subspaces through a learnable gating mechanism, retaining discriminative structures while suppressing redundancy.

Theorem 2 (Joint Contribution)**.** *Let $\beta > 0$ and define*

$$\mathcal{F}(S) = \mathcal{R}(S) - \beta|S|. \tag{13}$$

For $t \notin S$, its marginal gain is

$$\Delta(t \mid S) = p_t a_t - \beta + 2 \sum_{i \in S} \sqrt{p_i p_t}\, b_{it}. \tag{14}$$

Hence, the selection of t depends jointly on its individual importance $p_t a_t$ and its compatibility with the selected set S.

Theorem 3 (Bounded Omission)**.** *Let $\widehat{S}$ be the selected set and $R = \{1, \ldots, K\} \setminus \widehat{S}$. Then*

$$\Gamma(R \mid \widehat{S}) = \sum_{t \in R} \Delta(t \mid \widehat{S}) + 2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t}\, b_{st}. \tag{15}$$

If

$$\Delta(t \mid \widehat{S}) \le 0, \quad \forall t \in R, \qquad 2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} [b_{st}]_+ \le \epsilon, \tag{16}$$

then

$$\Gamma(R \mid \widehat{S}) \le \epsilon. \tag{17}$$

5 GraSS: Graph-Structured Adaptive Subspace Modelling

In this section, we further propose Graph-Structured Adaptive Subspace Modelling (GraSS), which learns a structural relationship graph among different constituent subspaces to adaptively uncover their underlying combinatorial distribution structure, thereby determining mutually compatible constituent subspaces and their composition. Furthermore, we introduce an adaptive gating mechanism to suppress redundant composite subspaces, thereby obtaining more discriminative geometric representations. The overall pipeline of GraSS is illustrated in Fig. 2.

5.1 CONSTITUENT SUBSPACE GRAPH MODELLING

Given an input sample, the Riemannian encoder first maps it to a geometric representation $X \in \mathbb{R}^{M\times M}$. We then perform singular value decomposition (SVD) on X and obtain M one-dimensional subspaces from its left singular vectors:$X = U\Sigma V^\top$, where each singular vector corresponds to a constituent subspace $S_i \in \mathcal{G}(M, 1)$, yielding the constituent subspace set $\mathcal{S} = \{S_1, S_2, \ldots, S_M\}$. Since each S_i is a one-dimensional orthonormal subspace, its basis satisfies $S_i^\top S_i = 1$.

Based on these constituent subspaces, we first construct an initial constituent subspace graph, which provides geometric priors for subsequent graph structure learning. Specifically, each constituent subspace is regarded as a node, yielding

$$\mathcal{G}^{(0)} = (\mathcal{V}, \mathcal{E}^{(0)}, A^{(0)}), \tag{18}$$

where $\mathcal{V} = \{S_1, S_2, \ldots, S_M\}$, and $A^{(0)} \in \mathbb{R}^{M\times M}$ denotes the initial adjacency matrix (Jin et al., 2020).

We then characterise the initial geometric relationship between two constituent subspaces through the distance between their projection matrices:$d_{\mathrm{p}}(S_i, S_j)$. Based on this geometric relationship, we construct the initial edge weights (Chen et al., 2020b) as

$$A_{ij}^{(0)} = f\left(d_{\mathrm{p}}(S_i, S_j)\right), \tag{19}$$

where $f(\cdot)$ denotes a mapping. The resulting $A^{(0)}$ serves as an initial structural prior and is used to initialise the learnable adjacency matrix A, which is subsequently optimised end-to-end during training.

5.2 SUBSPACE CONSTRUCTION

After obtaining the final learned constituent subspace graph, GraSS further exploits its topological structure to adaptively uncover the distribution of subspace combinations. Specifically, we perform connected-component analysis (Bondy & Murty, 2008) on the learned adjacency matrix A and denote the resulting maximal connected subgraphs by $\mathcal{C} = \{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K\}$. Each $\mathcal{C}_k \subseteq \mathcal{V}$ represents a maximal connected subgraph composed of mutually connected constituent subspaces. Since the maximal connected subgraphs form a partition of the graph nodes, they are mutually disjoint and collectively cover all constituent subspaces, i.e., $\mathcal{C}_k \cap \mathcal{C}_r = \varnothing, \quad k \neq r$, and $\bigcup_{k=1}^{K} \mathcal{C}_k = \mathcal{V}$.

The learned graph adaptively partitions the M constituent subspaces into K connected components $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$, where both the number and composition of the components are determined by graph connectivity. For each component $\mathcal{C}_k$, we assign a connectivity-based weight to each constituent subspace:

$$\alpha_{ik} = \frac{\sum_{j\in\mathcal{C}_k} A_{ij}}{\sum_{u\in\mathcal{C}_k}\sum_{j\in\mathcal{C}_k} A_{uj}}, \qquad S_i \in \mathcal{C}_k, \tag{20}$$

This enables the construction of composite subspaces to consider the structural compatibility among constituent subspaces while preserving their individual contributions to the final representation. For each constituent subspace $S_i \in \mathcal{C}_k$ within a connected component $\mathcal{C}_k$, let Q_i denote its orthonormal basis. We use the connectivity weights within the maximal connected subgraph to obtain the composite subspace representation:

$$\widetilde{Q}_k = \sum_{i\in\mathcal{C}_k} \alpha_{ik} Q_i. \tag{21}$$

Since $\widetilde{Q}_k$ is generally not orthonormal, we further perform QR decomposition:

$$\widetilde{Q}_k = Q_k R_k, \tag{22}$$

and take the orthonormal factor Q_k as the final composite subspace representation. Through this construction, GraSS can adaptively discover constituent subspaces with high structural coherence, thereby forming effective composite subspace representations.

5.3 Adaptive Subspace Selecting

Although graph connectivity enables the discovery of multiple composite subspaces, their contributions to the downstream task may differ. GraSS therefore introduces an adaptive gating mechanism to selectively emphasise discriminative structures and eliminate redundant ones. For the k-th composite subspace, we first obtain its gate through a learnable mapping:

$$g_k = \sigma\left(\phi(\widetilde{Q}_k)\right), \tag{23}$$

where $\phi(\cdot)$ is a learnable mapping and $\sigma(\cdot)$ denotes the Sigmoid function. The gating weight g_k measures the importance of the k-th composite subspace and is used to adaptively select effective composite subspaces, denoted by

$$\widetilde{\mathcal{Q}} = \left\{\widetilde{Q}_1, \widetilde{Q}_2, \ldots, \widetilde{Q}_{K'}\right\}, \qquad K' \leq K, \tag{24}$$

and fed into the subsequent network for representation learning and prediction. In this way, GraSS adaptively preserves informative composite structures while reducing the influence of redundant ones.

Algorithm 1 Graph-Structured Adaptive Subspace Modelling (GraSS)

Input: Constituent subspaces $\mathcal{S} = \{S_i\}_{i=1}^{M}$, $S_i \in \mathcal{G}(M, 1)$
Parameter: Gating parameters ϕ,Learnable Graph A.
Output: Gated composite subspaces $\{\widetilde{Q}_k\}_{k=1}^{K'}$
Initialisation: Construct the initial graph $A^{(0)}$ according to Eq. 19. Initialise the learnable graph as $A \leftarrow A^{(0)}$ and randomly initialise the gating parameters ϕ.
Identify the connected components of A: $\mathcal{C} = \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$, where K denotes their number.
for $k = 1$ **to** K **do**
    Extract the connectivity weights from A: $\alpha_{ik} \leftarrow \frac{\sum_{j \in \mathcal{C}_k} A_{ij}}{\sum_{u \in \mathcal{C}_k} \sum_{j \in \mathcal{C}_k} A_{uj}}$
    Subspace fusion: $\widetilde{Q}_k \leftarrow \sum_{i \in \mathcal{C}_k} \alpha_{ik} Q_i$
    Orthogonalise: $\widetilde{Q}_k \leftarrow \mathrm{QR}(\widetilde{Q}_k)$
end for
Apply gating and obtain $K' < K$ composite subspaces: $\{\widetilde{Q}_{k'}\}_{k'=1}^{K'} \leftarrow \{\phi(\widetilde{Q}_k)\}_{k=1}^{K}$
Return: $\{\widetilde{Q}_{k'}\}_{k'=1}^{K'}$

6 Experiments

To validate the effectiveness of the proposed Graph-Structured Adaptive Subspace Modelling (GraSS), we conduct experiments on action recognition tasks. Specifically, we investigate whether GraSS can adaptively construct discriminative composite subspaces, thereby learning more compact and discriminative geometric representations.

6.1 Experimental Setup

Datasets. We evaluate GraSS on three skeleton-based human action recognition datasets: FPHA (Garcia-Hernando et al., 2018), HDM05 (Müller et al., 2007), and NTU60 (Shahroudy et al., 2016). All experiments are conducted with 10 different random seeds, and the mean and standard deviation of the results are reported. The best-performing results are highlighted in bold in the tables. More experimental settings and implementation details are provided in Appendices G.1 and G.2.

6.2 Effectivene Analysis

As shown in Table 1, GraSS achieves competitive recognition performance across all three datasets while maintaining a low computational cost. Compared with existing Riemannian methods, GraSS not only demonstrates strong recognition performance but also exhibits a more pronounced performance advantage on NTU60. Specifically, on the FPHA dataset, GraSS achieves an accuracy of

Table 1: Comparison with Existing Riemannian Methods on the HDM05, FPHA, and NTU60 Datasets. FLOPs are reported on the NTU60 dataset.

Method	HDM05	FPHA	NTU60	
	Acc.	***Acc.***	***Acc.***	***FLOPs(M)***
SPDNet (Huang & Van Gool, 2017)	60.45%±1.12	87.65%±1.02	57.92%±0.36	5.0760
MATT (Pan et al., 2022)	62.25%±1.68	87.70%±0.68	58.63%±0.65	29.7180
GrNet (Huang et al., 2018)	59.23%±1.78	78.79%±1.82	57.45%±0.29	**0.6953**
GDLNet (Wang et al., 2024)	60.08%±1.78	87.60%±0.69	56.29%±0.78	20.4600
GMSF-Net (Yu & Xu, 2026)	63.64%±1.24	90.43%±0.74	58.52%±0.99	21.7389
GraSS	**63.72%±1.43**	**90.44%±0.38**	**61.04%±0.29**	**0.6953**

Table 2: Ablation study of GraSS on the HDM05 and FPHA datasets.

Variant	Constituent Subspace	Graph	Gate	$\mathcal{L}_{\text{graph}}$	$\mathcal{L}_{\text{gate}}$	HDM05	FPHA
GrNet	×	×	×	×	×	57.36%±1.68	81.53%±0.81
Fixed Partition	✓	×	×	×	×	62.11%±1.67	81.99%±0.83
Graph	✓	✓	×	×	×	62.44%±1.47	88.08%±0.48
Graph + Gate	✓	✓	✓	×	×	62.95%±1.73	88.99%±0.62
GraSS	✓	✓	✓	✓	✓	**63.72%±1.43**	**90.44%±0.38**

$90.44\% \pm 0.38$, slightly outperforming GMSF-Net with $90.43\% \pm 0.74$. Although the mean accuracies of the two methods are comparable, the standard deviation of GraSS is reduced from 0.74 to 0.38, indicating that it learns more stable Grassmannian subspace representations. Notably, on the NTU60 dataset, GraSS achieves an accuracy of $61.04\% \pm 0.29$, improving upon GMSF-Net by 2.52 percentage points from $58.52\% \pm 0.99$, while also exhibiting a substantially lower standard deviation. This demonstrates that GraSS can obtain more stable Grassmannian representations by adaptively modelling the relationships among constituent subspaces. In terms of computational efficiency, GraSS requires only 0.6953 *MFLOPs*, which is substantially lower than those of the other baseline methods. Notably, with the same computational cost as GrNet, GraSS improves the recognition accuracy by 4.49, 11.65, and 3.59 percentage points on the three datasets, respectively. This demonstrates that GraSS does not rely on substantially increased computational costs to achieve performance gains, but instead improves the utilisation of Grassmannian representations through more effective modelling of subspace relationships.

6.3 Ablation Experiments

We conduct ablation experiments on the HDM05 and FPHA datasets to investigate the contribution of each component in GraSS. The results are reported in Table 2. First, introducing the subspace composition mechanism leads to a substantial improvement over the baseline. Specifically, the Fixed Partition variant improves the accuracy on HDM05 from 57.36% to 62.11%, while achieving an accuracy of 81.99% on FPHA. This result demonstrates that explicitly composing the constituent subspaces of a Grassmann subspace into structured groups can provide more effective geometric representations than treating the original subspace as a whole. It also indicates that exploiting the intrinsic structure within Grassmann subspaces is beneficial for downstream action recognition. Building upon this, the introduction of graph-based structural modelling enables the model to further capture the relationships among constituent subspaces. Compared with the fixed partition strategy, the Graph variant employs graph-guided adaptive subspace composition, allowing the model to discover more appropriate structural relationships from the data rather than relying on a predefined partition. We then introduce the gating mechanism on top of the Graph variant. This enables the model to adaptively modulate the contribution of different composed subspaces according to their importance, thereby providing greater flexibility in representation learning.

Finally, the complete GraSS incorporates both the graph regularisation $\mathcal{L}_{\text{graph}}$ and the gate regularisation $\mathcal{L}_{\text{gate}}$. GraSS achieves the best performance on both datasets, reaching $63.12\% \pm 1.43\%$ on HDM05 and $90.34\% \pm 0.38\%$ on FPHA. Compared with the GrNet, GraSS improves the accuracy by 5.76 percentage points on HDM05 and 8.91 percentage points on FPHA. These results demonstrate that the performance gains are not attributable to a single component, but rather arise from

Table 3: Downstream classification performance under different Grassmannian dimensions across complexity levels on the Salinas dataset.

Complexity	$p = 2$	$p = 3$	$p = 5$	$p = 8$	$p = 10$	$p = 15$	p^*_{cls}
Q1 (Very Low)	92.74	93.66	95.12	95.20	95.20	92.66	5
Q2 (Low)	93.61	93.89	95.39	96.49	96.05	92.99	8
Q3 (Medium)	94.13	95.01	95.39	95.94	96.68	93.68	8
Q4 (High)	91.77	94.32	94.59	96.34	95.61	92.83	8
Q5 (Very High)	90.75	93.80	96.75	97.11	97.76	95.42	8

the complementary effects of subspace composition, adaptive graph modelling, gating, and their corresponding regularisation objectives. Overall, the ablation results verify that the individual components of GraSS play complementary roles in constructing effective geometric representations. In particular, the substantial improvement of the complete model over the baseline demonstrates that explicitly modelling the structural organisation and relative importance of constituent subspaces within Grassmann subspaces provides an effective strategy for improving Grassmannian geometric representation learning.

6.4 Further Analysis

To validate the preceding theoretical analysis that the capacity of Grassmann subspace representations should adapt dynamically to sample characteristics, we conduct a sample dimensionality experiment on the Salinas dataset. We divide the samples into five levels, Q1–Q5, according to their local spectral complexity, and examine their classification performance under different Grassmann representation dimensions to determine whether their representation capacity requirements differ. As shown in Table 3, samples with different complexity levels exhibit different responses to the representation dimension. For the low-complexity Q1 samples, a dimension of $p = 5$ already achieves 95.12%, close to the best performance of 95.20%; in contrast, the high-complexity Q5 samples improve from 90.75% at $p = 2$ to 97.76% at $p = 10$.

Furthermore, the minimum representation dimensions p^*_{cls} for Q1–Q5 are $5, 8, 8, 8, 8$, respectively, indicating that samples with different complexity levels have different requirements for Grassmann representation capacity. Therefore, the key finding of this experiment is that the representation capacity requirements vary across samples. A fixed Grassmann subspace dimension cannot simultaneously accommodate the structural and information requirements of different samples, suggesting that the representation capacity should be dynamically adjusted according to sample characteristics. These results provide empirical support for the adaptive Grassmann subspace representation capacity proposed in this work. Further analyses of the consistency of dimensionality requirements and the robustness of the complexity dimensionality relationship under different patch sizes are provided in Appendix G.3.

7 Conclusion

This paper proposes a Graph-Structured adaptive SubSpace (GraSS) modelling framework, providing a new perspective for understanding and exploiting the distribution of composite subspaces within a Grassmann point. Through theoretical analysis based on spectral measure theory and quadratic utility theory, we reveal that different subspace combinations within a Grassmann point contribute non-uniformly to the representation. Specifically, GraSS constructs a relational graph among constituent subspaces and adaptively discovers highly structurally coherent constituent subspaces and their composition based on maximal connected subgraphs. Furthermore, we introduce an adaptive gating mechanism to dynamically regulate the contributions of different subspaces to the final representation, thereby obtaining more compact and discriminative geometric representations on the Grassmannian. Overall, this work extends the subspace modelling paradigm on the Grassmannian from the perspectives of intrinsic distribution and structural compatibility, providing a new approach to adaptive geometric representation learning on Grassmann manifolds.

ETHICS STATEMENT

This work uses only publicly available benchmark datasets, including NTU60, HDM05, and FPHA, which contain no personally identifiable or sensitive information. We do not identify any ethical concerns associated with this work.

REPRODUCIBILITY STATEMENT

This work aims to ensure the reproducibility of the experimental results. For experimental evaluation, we provide detailed descriptions of the all datasets, experimental protocols, and evaluation metrics, with further information on model implementation, and training parameters provided in the supplementary materials. If the paper is accepted, we will release the source code and necessary configuration files to facilitate the reproduction of our experiments and support future research.

REFERENCES

Gaston Baudat and Fatiha Anouar. Generalized discriminant analysis using a kernel approach. *Neural computation*, 12(10):2385–2404, 2000.

Thomas Bendokat, Ralf Zimmermann, and P-A Absil. A grassmann manifold handbook: Basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(1):6, 2024.

John Adrian Bondy and Uppaluri Siva Ramachandra Murty. *Graph theory*. Springer Publishing Company, Incorporated, 2008.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for spd neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Lipeng Cai, Jun Shi, Shaoyi Du, Yue Gao, and Shihui Ying. Self-adaptive subspace representation from a geometric intuition. *Pattern Recognition*, 149:110228, 2024.

Ming Chen, Zhewei Wei, Zengfeng Huang, Bolin Ding, and Yaliang Li. Simple and deep graph convolutional networks. In *International conference on machine learning*, pp. 1725–1735. PMLR, 2020a.

Yu Chen, Lingfei Wu, and Mohammed Zaki. Iterative deep graph learning for graph neural networks: Better and robust node embeddings. *Advances in neural information processing systems*, 33:19314–19326, 2020b.

Ziheng Chen, Tianyang Xu, Xiao-Jun Wu, Rui Wang, and Josef Kittler. Hybrid riemannian graph-embedding metric learning for image set classification. *IEEE transactions on big data*, 9(1): 75–92, 2021.

Ziheng Chen, Yue Song, Xiaojun Wu, and Nicu Sebe. Gyrogroup batch normalization. In *International Conference on Learning Representations*, volume 2025, pp. 101200–101237, 2025.

Anoop Cherian, Basura Fernando, Mehrtash Harandi, and Stephen Gould. Generalized rank pooling for activity recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3222–3231, 2017.

Soo Hong Chew, Larry G Epstein, and Uzi Segal. Mixture symmetry and quadratic utility. *Econometrica: Journal of the Econometric Society*, pp. 139–163, 1991.

Wei Dai, Ely Kerman, and Olgica Milenkovic. A geometric approach to low-rank matrix completion. *IEEE Transactions on Information Theory*, 58(1):237–247, 2012.

Yanhong Fei, Yingjie Liu, Chentao Jia, Zhengyu Li, Xian Wei, and Mingsong Chen. A survey of geometric optimization for deep learning: from euclidean space to riemannian manifold. *ACM Computing Surveys*, 57(5):1–37, 2025.

Yunyuan Gao, Yici Liu, Qingshan She, and Jianhai Zhang. Domain adaptive algorithm based on multi-manifold embedded distributed alignment for brain-computer interfaces. *IEEE Journal of Biomedical and Health Informatics*, 27(1):296–307, 2022.

Guillermo Garcia-Hernando, Shanxin Yuan, Seungryul Baek, and Tae-Kyun Kim. First-person hand action benchmark with rgb-d videos and 3d hand pose annotations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 409–419, 2018.

Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

Jihun Hamm and Daniel D Lee. Grassmann discriminant analysis: a unifying view on subspace-based learning. In *Proceedings of the 25th international conference on Machine learning*, pp. 376–383, 2008.

Mehrtash T Harandi, Conrad Sanderson, Sareh Shirazi, and Brian C Lovell. Kernel analysis on grassmann manifolds for action recognition. *Pattern Recognition Letters*, 34(15):1906–1915, 2013.

Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, and Xilin Chen. Projection metric learning on grassmann manifold with application to video based face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 140–149, 2015.

Zhiwu Huang, Jiqing Wu, and Luc Van Gool. Building deep networks on grassmann manifolds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Wei Jin, Yao Ma, Xiaorui Liu, Xianfeng Tang, Suhang Wang, and Jiliang Tang. Graph structure learning for robust graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 66–74, 2020.

Jiaming Liu, Ran Xu, Senqiao Yang, Renrui Zhang, Qizhe Zhang, Zehui Chen, Yandong Guo, and Shanghang Zhang. Continual-mae: Adaptive distribution masked autoencoders for continual test-time adaptation. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28653–28663. IEEE, 2024.

Mengyi Liu, Ruiping Wang, Shaoxin Li, Shiguang Shan, Zhiwu Huang, and Xilin Chen. Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. In *Proceedings of the 16th International Conference on multimodal interaction*, pp. 494–501, 2014.

Christos Louizos, Max Welling, and Diederik P Kingma. Learning sparse neural networks through l_0 regularization. *arXiv preprint arXiv:1712.01312*, 2017.

Meinard Müller, Tido Röder, Michael Clausen, Bernhard Eberhardt, Björn Krüger, and Andreas Weber. Mocap database hdm05. *Institut für Informatik II, Universität Bonn*, 2(7), 2007.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. *Advances in Neural Information Processing Systems*, 35:31116–31129, 2022.

Chunmei Qing, Jiawei Ruan, Xiangmin Xu, Jinchang Ren, and Jaime Zabalza. Spatial-spectral classification of hyperspectral images: a deep learning framework with markov random fields based modelling. *IET Image Processing*, 13(2):235–245, 2019.

Li Qiu, Yanxia Zhang, and Chi-Kwong Li. Unitarily invariant metrics on the grassmann space. *SIAM journal on matrix analysis and applications*, 27(2):507–531, 2005.

M Reed and B Simon. Methods of modern mathematical physics. volume i: Functional analysis (1972), 1972.

Amir Shahroudy, Jun Liu, Tian-Tsong Ng, and Gang Wang. Ntu rgb+ d: A large scale dataset for 3d human activity analysis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1010–1019, 2016.

Lincon S Souza, Naoya Sogi, Bernardo B Gatto, Takumi Kobayashi, and Kazuhiro Fukui. An interface between grassmann manifolds and vector spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 846–847, 2020.

Gilbert W Stewart. On the early history of the singular value decomposition. *SIAM review*, 35(4): 551–566, 1993.

Edric Tam and David Dunson. Fiedler regularization: Learning neural networks with graph sparsity. In *International Conference on Machine Learning*, pp. 9346–9355. PMLR, 2020.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Ulrike Von Luxburg. A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416, 2007.

Rui Wang, Xiao-Jun Wu, and Josef Kittler. Symnet: A simple symmetric positive definite manifold deep learning method for image set classification. *IEEE Transactions on Neural Networks and Learning Systems*, 33(5):2208–2222, 2021.

Rui Wang, Xiao-Jun Wu, Tianyang Xu, Cong Hu, and Josef Kittler. U-spdnet: An spd manifold learning-based neural network for visual classification. *Neural networks*, 161:382–396, 2023.

Rui Wang, Chen Hu, Ziheng Chen, Xiao-Jun Wu, and Xiaoning Song. A grassmannian manifold self-attention network for signal classification. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pp. 5099–5107, 2024.

Rui Wang, Shaocheng Jin, Ziheng Chen, Xiaoqing Luo, and Xiao-Jun Wu. Learning to normalize on the spd manifold under bures-wasserstein geometry. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8289–8298. IEEE, 2025.

Kunlun Xu, Xu Zou, Yuxin Peng, and Jiahuan Zhou. Distribution-aware knowledge prototyping for non-exemplar lifelong person re-identification. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16604–16613. IEEE, 2024.

Tianyang Xu, Xue-Feng Zhu, and Xiao-Jun Wu. Learning spatio-temporal discriminative model for affine subspace based visual object tracking. *Visual Intelligence*, 1(1):4, 2023.

Xuan Yu and Tianyang Xu. Learning topology-driven multi-subspace fusion for grassmannian deep networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 27987–27995, 2026.

Jiayao Zhang, Guangxu Zhu, Robert W Heath Jr, and Kaibin Huang. Grassmannian learning: Embedding geometry awareness in shallow and deep learning, 2018.

APPENDIX

A Default Notation **14**

B Limitations **14**

C AI use statement **14**

D Preliminaries **14**

D.1 Grassmannian 14

D.2 Spectral Measure 16

D.3 Quadratic Utility 16

E Proof **17**

E.1 Proof of Pro. 1 17

E.2 Proof of The. 1 18

E.3 Proof of Pro. 2 18

E.4 Proof of The. 2 19

E.5 Proof of The. 3 20

F Loss **20**

F.1 Graph Sparsity Loss 20

F.2 Gate Sparsity Loss 21

F.3 Overall Objective 21

G Experiments **21**

G.1 Datasets 21

G.2 Implementation Details 22

G.3 Further Analysis 23

A Default Notation

Table 4: Summary of notation.

Notation	Explanation
$\mathcal{G}(n,p)$	Grassmann manifold of p-dimensional subspaces
$T_U\mathcal{G}(n,p)$	Tangent space of the Grassmann manifold at U
$\mathrm{St}(n,p)$	Stiefel manifold
U	Orthonormal Basis Representation of the Grassmann Manifold
u	singular vector
σ	singular value
$\mathcal{S}$	Set of constituent low-dimensional subspaces
$\mathcal{C}_k$	k-th connected component of the learned subspace graph
$UU^\top$	Projection representation of a Grassmann point
Exp_U	Riemannian exponential map at U
Log_U	Riemannian logarithmic map at U
$d_{\mathrm{p}}(U_1,U_2)$	Projection metric between two Grassmann points
$\mathcal{V}$	Node set of the constituent-subspace graph

B Limitations

Although GraSS achieves consistent performance improvements across multiple datasets and downstream tasks, it still has certain limitations. GraSS relies on a relational graph among constituent subspaces for adaptive composition, and the quality of the learned graph directly affects the discovery of connected subgraphs and the construction of the resulting geometric representations. Therefore, how to characterise relationships among subspaces more accurately and robustly in complex or high-dimensional scenarios remains an open direction for further research.

C AI use statement

Generative AI tools were used to assist with the organisation and language of the manuscript, primarily to improve the clarity, fluency, and readability of the text. Beyond these language-related tasks, the research ideas, theoretical derivations, methodological design, experimental validation, and associated figures and tables were developed and completed independently by the authors. The authors take full responsibility for the content and final outcomes of this work.

D Preliminaries

D.1 Grassmannian

The Grassmann manifold, denoted by $\mathcal{G}(n,p)$, is the set of all p-dimensional linear subspaces of $\mathbb{R}^n$, given by

$$\mathcal{G}(n,p) = \{\mathcal{S} \subset \mathbb{R}^n \mid \dim(\mathcal{S}) = p\}. \tag{25}$$

Each point $\mathcal{S} \in \mathcal{G}(n,p)$ represents a p-dimensional linear subspace (Bendokat et al., 2024). In practice, a Grassmann point can be represented by an orthonormal basis matrix $U \in \mathbb{R}^{n\times p}$ satisfying

$$U^\top U = I_p, \tag{26}$$

such that

$$\mathcal{S} = \mathrm{span}(U). \tag{27}$$

Importantly, the orthonormal basis is not unique. For any orthogonal matrix $R \in \mathbb{R}^{p\times p}$ satisfying $R^\top R = I_p$, U and UR represent the same Grassmann point:

$$\mathrm{span}(U) = \mathrm{span}(UR). \tag{28}$$

Accordingly, the Grassmann manifold can be viewed as the quotient of the Stiefel manifold $\mathrm{St}(n,p)$ by the orthogonal group $O(p)$, i.e.,

$$\mathcal{G}(n,p) = \mathrm{St}(n,p)/O(p), \tag{29}$$

and has dimension $p(n-p)$. This basis invariance is essential for characterising intrinsic geometric structures on the Grassmannian.

Tangent Space. As a Riemannian manifold, the Grassmannian admits a tangent space at each point. Given a Grassmann point represented by U, its tangent space in the orthonormal-basis viewpoint is

$$T_U\mathcal{G}(n,p) = \left\{\Delta \in \mathbb{R}^{n\times p} \;\middle|\; U^\top \Delta = 0\right\}. \tag{30}$$

A tangent vector $\Delta \in T_U\mathcal{G}(n,p)$ therefore represents an infinitesimal direction along which the subspace can move from U while remaining orthogonal to the current subspace.

Exponential and Logarithmic Maps. The exponential map establishes a correspondence between tangent vectors and points on the Grassmann manifold. For $\Delta \in T_U\mathcal{G}(n,p)$, let

$$\Delta = Q\Sigma R^\top \tag{31}$$

be its compact singular value decomposition, where $Q^\top Q = I_p$, $R^\top R = I_p$, and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$. The exponential map is given by

$$\mathrm{Exp}_U(\Delta) = UR\cos(\Sigma) + Q\sin(\Sigma), \tag{32}$$

where the trigonometric functions are applied element-wise to the diagonal entries of Σ.

Conversely, for two Grassmann points represented by orthonormal basis matrices $U, Y \in \mathbb{R}^{n\times p}$, the logarithmic map $\mathrm{Log}_U(Y)$ gives the tangent vector at U pointing towards Y. Specifically, define

$$M = \left(I_n - UU^\top\right) Y \left(U^\top Y\right)^{-1}, \tag{33}$$

and let its compact singular value decomposition be

$$M = O\Sigma R^\top. \tag{34}$$

Then,

$$\mathrm{Log}_U(Y) = O\Theta R^\top, \qquad \Theta = \arctan(\Sigma), \tag{35}$$

where $\arctan(\cdot)$ is applied element-wise. Equivalently, if $M = O\tan(\Theta)R^\top$, then

$$\mathrm{Log}_U(Y) = O\Theta R^\top. \tag{36}$$

The exponential and logarithmic maps provide the basic tools for transferring between the nonlinear Grassmann manifold and its local tangent spaces.

Grassmannian Metric. The geometric distance between two Grassmann points can be measured using different Riemannian (Qiu et al., 2005) or projection-based metrics (Huang et al., 2015). In this work, we adopt the projection metric for its computational simplicity. For two Grassmann points represented by orthonormal basis matrices $U_1, U_2 \in \mathbb{R}^{n\times p}$, the projection metric is defined as

$$d_{\mathrm{p}}(U_1, U_2) = \frac{1}{\sqrt{2}} \left\| U_1U_1^\top - U_2U_2^\top \right\|_F, \tag{37}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. Since $UU^\top$ is invariant to the choice of orthonormal basis U (Fei et al., 2025), *i.e.*,

$$(UR)(UR)^\top = UU^\top, \tag{38}$$

the projection metric is well-defined on the Grassmann manifold and depends only on the underlying subspaces rather than on their particular basis representations. This metric provides an effective measure of geometric discrepancy between Grassmann points and will be used for subsequent geometric analysis.

D.2 Spectral Measure

Let $X \in \mathbb{R}^{d\times n}$ have the singular value decomposition

$$X = U\Sigma V^\top, \tag{39}$$

where $\sigma_1 \geq \cdots \geq \sigma_r \geq 0$ are the singular values. For the leading K singular components, the normalised spectral weights are defined as

$$p_i = \frac{\sigma_i^2}{\sum_{j=1}^K \sigma_j^2}, \qquad i = 1, \ldots, K, \tag{40}$$

which satisfy

$$p_i \geq 0, \qquad \sum_{i=1}^K p_i = 1. \tag{41}$$

Let u_i denote the corresponding left singular vector. To remove the sign ambiguity of singular vectors, we represent each component by its rank-one projection

$$P_i = u_i u_i^\top. \tag{42}$$

The spectral measure is then defined as

$$\mu_X = \sum_{i=1}^K p_i \delta_{P_i}, \tag{43}$$

where δ_{P_i} denotes the Dirac measure at P_i. Since the weights form a probability distribution, μ_X is a discrete probability measure over the leading spectral components.

The projection representation is invariant to the sign of the singular vectors, since

$$(-u_i)(-u_i)^\top = u_i u_i^\top = P_i. \tag{44}$$

Therefore, μ_X is independent of the arbitrary sign choices of the singular vectors.

D.3 Quadratic Utility

Let $\phi_i = \phi(P_i) \in \mathbb{R}^m$ denote the representation of the i-th constituent component. To incorporate its spectral importance p_i, we define the weighted directional atom as

$$g_i = \sqrt{p_i}\phi_i. \tag{45}$$

For a subset $S \subseteq \{1, \ldots, K\}$, the corresponding aggregated representation is

$$z_S = \sum_{i\in S} g_i = \sum_{i\in S} \sqrt{p_i}\phi_i. \tag{46}$$

Let $M \succeq 0$ be a positive semidefinite matrix that characterises a task-related metric or objective. The quadratic utility of the composition S is defined as

$$\mathcal{R}(S) = z_S^\top M z_S. \tag{47}$$

Substituting the definition of z_S gives

$$\mathcal{R}(S) = \left(\sum_{i\in S} \sqrt{p_i}\phi_i\right)^\top M \left(\sum_{j\in S} \sqrt{p_j}\phi_j\right). \tag{48}$$

Expanding the quadratic form yields

$$\mathcal{R}(S) = \sum_{i\in S} p_i \phi_i^\top M \phi_i + \sum_{\substack{i,j\in S \\ i\neq j}} \sqrt{p_i p_j}\, \phi_i^\top M \phi_j. \tag{49}$$

Since M is symmetric, i.e., $M = M^\top$, we have

$$\phi_i^\top M \phi_j = \phi_j^\top M \phi_i. \tag{50}$$

Therefore, the off-diagonal terms can be grouped into pairs:

$$\mathcal{R}(S) = \sum_{i \in S} p_i \phi_i^\top M \phi_i + 2 \sum_{\substack{i,j \in S \\ i<j}} \sqrt{p_i p_j}\, \phi_i^\top M \phi_j. \tag{51}$$

Define the individual utility of component i as

$$a_i = \phi_i^\top M \phi_i, \tag{52}$$

and the pairwise interaction between components i and j as

$$b_{ij} = \phi_i^\top M \phi_j. \tag{53}$$

The quadratic utility can then be expressed as

$$\mathcal{R}(S) = \sum_{i \in S} p_i a_i + 2 \sum_{\substack{i,j \in S \\ i<j}} \sqrt{p_i p_j}\, b_{ij}. \tag{54}$$

The first term represents the individual contribution of each component, weighted by its spectral importance, while the second term captures pairwise interactions induced by joint composition. This decomposition shows that the utility of a composition is determined jointly by component importance and their interactions.

E PROOF

E.1 PROOF OF PRO. 1

Proof. By the SVD, the contribution of the i-th left singular direction to the energy of X is quantified by σ_i^2. Hence, the total spectral energy contained in the first K directions is

$$E_K = \sum_{j=1}^{K} \sigma_j^2. \tag{55}$$

Therefore,

$$p_i = \frac{\sigma_i^2}{E_K} \geq 0, \qquad i = 1, \ldots, K. \tag{56}$$

Moreover,

$$\sum_{i=1}^{K} p_i = \frac{\sum_{i=1}^{K} \sigma_i^2}{\sum_{j=1}^{K} \sigma_j^2} = 1. \tag{57}$$

Thus, $\{p_i\}_{i=1}^K$ forms a valid probability distribution over the K constituent directions.

Furthermore, the Grassmann representation is invariant to the sign ambiguity of individual singular vectors. In particular, if u_i is replaced by $-u_i$, then

$$(-u_i)(-u_i)^\top = u_i u_i^\top = P_i. \tag{58}$$

Hence, defining the distribution over the projection matrices P_i, rather than over the signed vectors u_i, is consistent with the sign-invariant geometry of the Grassmannian.

Consequently, the SVD provides not only the location of the Grassmann point

$$\mathcal{U} = \mathrm{span}(U_K), \tag{59}$$

but also an internal spectral distribution over its constituent directions:

$$\mu_{\mathcal{U}} = \sum_{i=1}^{K} p_i \delta_{P_i}, \tag{60}$$

where p_i quantifies the relative spectral energy contributed by the i-th constituent direction. This completes the proof. □

E.2 Proof of The. 1

Proof. By definition, the aggregated representation is

$$z_S = \sum_{i \in S} \sqrt{p_i} \phi_i. \tag{61}$$

Therefore, the representation utility can be written as

$$\mathcal{R}(S) = z_S^\top M z_S. \tag{62}$$

Substituting the expression for z_S gives

$$\mathcal{R}(S) = \left(\sum_{i \in S} \sqrt{p_i} \phi_i \right)^\top M \left(\sum_{j \in S} \sqrt{p_j} \phi_j \right). \tag{63}$$

Expanding the bilinear form yields

$$\mathcal{R}(S) = \sum_{i \in S} \sum_{j \in S} \sqrt{p_i p_j}\, \phi_i^\top M \phi_j. \tag{64}$$

Separating the diagonal terms corresponding to $i = j$ from the off-diagonal terms gives

$$\mathcal{R}(S) = \sum_{i \in S} p_i \phi_i^\top M \phi_i + 2 \sum_{\substack{i,j \in S \\ i<j}} \sqrt{p_i p_j}\, \phi_i^\top M \phi_j. \tag{65}$$

By defining

$$a_i = \phi_i^\top M \phi_i, \qquad b_{ij} = \phi_i^\top M \phi_j, \tag{66}$$

we obtain

$$\mathcal{R}(S) = \sum_{i \in S} p_i a_i + 2 \sum_{\substack{i,j \in S \\ i<j}} \sqrt{p_i p_j}\, b_{ij}. \tag{67}$$

This completes the proof. □

E.3 Proof of Pro. 2

Proof. By Theorem 1, the utility of the individual direction i is

$$\mathcal{R}(\{i\}) = p_i a_i, \tag{68}$$

while the utility of the pair $\{i, j\}$ is

$$\mathcal{R}(\{i, j\}) = p_i a_i + p_j a_j + 2\sqrt{p_i p_j}\, b_{ij}. \tag{69}$$

Therefore, the marginal utility of adding direction j to direction i is

$$\Delta(j \mid \{i\}) = \mathcal{R}(\{i, j\}) - \mathcal{R}(\{i\}) = p_j a_j + 2\sqrt{p_i p_j}\, b_{ij}. \tag{70}$$

Suppose that

$$b_{ij} < -\frac{1}{2} \sqrt{\frac{p_j}{p_i}}\, a_j. \tag{71}$$

Then

$$2\sqrt{p_i p_j}\, b_{ij} < 2\sqrt{p_i p_j} \left(-\frac{1}{2} \sqrt{\frac{p_j}{p_i}}\, a_j \right). \tag{72}$$

The right-hand side simplifies to

$$2\sqrt{p_i p_j} \left(-\frac{1}{2} \sqrt{\frac{p_j}{p_i}}\, a_j \right) = -p_j a_j. \tag{73}$$

Consequently,

$$2\sqrt{p_i p_j}\, b_{ij} < -p_j a_j. \tag{74}$$

It follows that

$$p_j a_j + 2\sqrt{p_i p_j}\, b_{ij} < 0. \tag{75}$$

Hence,

$$\Delta(j \mid \{i\}) < 0. \tag{76}$$

Therefore, although direction j has a positive individual contribution $p_j a_j$, its inclusion can decrease the overall representation utility when its compatibility with direction i is sufficiently negative. This completes the proof. □

E.4 Proof of The. 2

Proof. By definition,

$$\mathcal{F}(S) = \sum_{i \in S} p_i a_i + 2 \sum_{\substack{i,j \in S \\ i<j}} \sqrt{p_i p_j} b_{ij} - \beta |S|. \tag{77}$$

After adding a new direction $t \notin S$, the newly introduced individual contribution is

$$p_t a_t, \tag{78}$$

while the complexity penalty increases by

$$-\beta. \tag{79}$$

Moreover, the new pairwise interaction terms are those between t and every previously selected direction $i \in S$, yielding

$$2 \sum_{i \in S} \sqrt{p_i p_t} b_{it}. \tag{80}$$

Hence,

$$\mathcal{F}(S \cup \{t\}) = \mathcal{F}(S) + p_t a_t - \beta + 2 \sum_{i \in S} \sqrt{p_i p_t} b_{it}. \tag{81}$$

Subtracting $\mathcal{F}(S)$ from both sides gives

$$\Delta(t \mid S) = p_t a_t - \beta + 2 \sum_{i \in S} \sqrt{p_i p_t} b_{it}. \tag{82}$$

This proves the stated marginal-gain expression. □

Remark 1. The expression

$$p_t a_t \tag{83}$$

measures the individual task-related importance of direction t, whereas

$$2 \sum_{i \in S} \sqrt{p_i p_t} b_{it} \tag{84}$$

measures its aggregate compatibility with the currently selected set S.

Consequently, if

$$\Delta(t \mid S) > 0, \tag{85}$$

then adding direction t increases the compatibility-aware objective, and thus t is beneficial to the current effective representation. Conversely, if

$$\Delta(t \mid S) \leq 0, \tag{86}$$

then direction t provides no positive marginal contribution to the current set and can therefore be regarded as a weak or incompatible direction.

E.5 Proof of The. 3

Proof. By expanding $\mathcal{F}(\widehat{S} \cup R)$, we obtain

$$\mathcal{F}(\widehat{S} \cup R) = \mathcal{F}(\widehat{S}) + \sum_{t \in R} \left(p_t a_t - \beta + 2 \sum_{i \in \widehat{S}} \sqrt{p_i p_t} b_{it} \right) + 2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} b_{st}. \tag{87}$$

The term inside the first summation is precisely the marginal gain of adding direction t to $\widehat{S}$:

$$\Delta(t \mid \widehat{S}) = p_t a_t - \beta + 2 \sum_{i \in \widehat{S}} \sqrt{p_i p_t} b_{it}. \tag{88}$$

Therefore,

$$\Gamma(R \mid \widehat{S}) = \sum_{t \in R} \Delta(t \mid \widehat{S}) + 2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} b_{st}. \tag{89}$$

Since

$$\Delta(t \mid \widehat{S}) \leq 0, \qquad \forall t \in R, \tag{90}$$

we have

$$\sum_{t \in R} \Delta(t \mid \widehat{S}) \leq 0. \tag{91}$$

Moreover,

$$b_{st} \leq [b_{st}]_+, \tag{92}$$

and since

$$\sqrt{p_s p_t} \geq 0, \tag{93}$$

it follows that

$$2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} b_{st} \leq 2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} [b_{st}]_+. \tag{94}$$

By assumption,

$$2 \sum_{\substack{s,t \in R \\ s<t}} \sqrt{p_s p_t} [b_{st}]_+ \leq \epsilon. \tag{95}$$

Combining the above inequalities yields

$$\Gamma(R \mid \widehat{S}) \leq 0 + \epsilon = \epsilon. \tag{96}$$

Thus, the total improvement that could be obtained by reintroducing the omitted directions is at most ϵ. In particular, when ϵ is small, discarding the weak or incompatible directions incurs only a bounded loss in the compatibility-aware objective. □

F Loss

F.1 Graph Sparsity Loss

Although the low-dimensional subspace graph adaptively captures the structural relationships between different low-dimensional subspaces, the learned graph may become overly dense without appropriate constraints (Chen et al., 2020a). Excessive connections can introduce redundant relationships between low-dimensional subspaces and weaken the ability of graph connectivity to reveal distinct local structures. Therefore, we impose a sparsity constraint on the learned adjacency matrix $\mathbf{A}$.

Specifically, the graph sparsity loss is defined as

$$\mathcal{L}_{\text{graph}} = \frac{1}{M(M-1)} \sum_{i \neq j} |A_{ij}| \,. \tag{97}$$

This term penalises the non-diagonal edge weights in the low-dimensional subspace graph, encouraging unnecessary connections to approach zero (Tam & Dunson, 2020). Consequently, only structurally meaningful relationships between low-dimensional subspaces tend to be preserved, resulting in a more compact and interpretable graph structure.

F.2 Gate Sparsity Loss

After discovering the low-dimensional subspaces according to the connectivity of the learned graph, different low-dimensional subspaces may contribute unequally to the final recognition task. In particular, some low-dimensional subspaces may contain informative action structures, whereas others may provide redundant or weakly discriminative information (Harandi et al., 2013). To encourage the model to suppress such redundant composite subspaces, we further impose a sparsity constraint on the adaptive gates. Let

$$\mathbf{g} = [g_1, g_2, \ldots, g_K] \tag{98}$$

denote the gating coefficients of the K composite subspaces, where $g_k \in (0, 1)$. The gate sparsity loss is defined as

$$\mathcal{L}_{\text{gate}} = \frac{1}{K} \sum_{k=1}^{K} g_k. \tag{99}$$

Minimising $\mathcal{L}_{\text{gate}}$ encourages the gating coefficients to become small, thereby promoting the suppression of uninformative composite subspaces. However, the gate sparsity loss does not independently determine which composite subspaces should be removed (Louizos et al., 2017). For an informative composite subspace, excessively reducing its gate would impair the classification performance, and the classification loss consequently encourages its gate to remain sufficiently large. The interaction between the classification objective and the gate sparsity constraint therefore allows GraSS to retain informative composite subspaces while suppressing redundant ones.

F.3 Overall Objective

Based on the above two structural constraints, the overall training objective of GraSS is formulated as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{ce}} + \lambda_{\text{graph}} \mathcal{L}_{\text{graph}} + \lambda_{\text{gate}} \mathcal{L}_{\text{gate}}, \tag{100}$$

where $\mathcal{L}_{\text{ce}}$ denotes the classification loss (Goodfellow et al., 2016), $\mathcal{L}_{\text{graph}}$ denotes the graph sparsity loss, and $\mathcal{L}_{\text{gate}}$ denotes the gate sparsity loss. λ_{graph} and λ_{gate} control the strengths of the graph and composite subspace regularisation, respectively.

The three objectives operate at different levels of the proposed framework. The classification loss determines which structures are useful for the recognition task, the graph sparsity loss controls which low-dimensional subspaces should maintain structural connections, and the gate sparsity loss determines which discovered composite subspaces should be retained. Through their joint optimisation, GraSS learns a compact low-dimensional subspace graph and selectively preserves discriminative composite subspaces, leading to a more structured and effective geometric representation.

G Experiments

G.1 Datasets

FPHA (Garcia-Hernando et al., 2018): The First-Person Hand Action Benchmark (FPHA) is a skeleton-based dataset designed for first-person hand action recognition. It contains 1,175 video sequences covering 45 hand action categories, with each frame providing the 3D spatial coordinates of 21 hand joints. Thus, each frame can be represented by a 63-dimensional vector obtained by concatenating the corresponding (x, y, z) coordinates. For each action sequence, we compute a 63×63 covariance or correlation matrix over the frame-level features, which serves as its structured

representation. Following the experimental protocol in Wang et al. (2021), we consider either the standard split with 600 sequences for training and 575 for testing, or the subject-wise protocol in which subjects 1–3 are used for training and subjects 4–6 for validation. This covariance-based representation provides a compact description of the joint-level motion correlations and facilitates subsequent action classification.

HDM05 (Müller et al., 2007): The HDM05 dataset is a motion-capture-based benchmark for human action recognition, consisting of 2,273 skeleton sequences collected from multiple subjects. Each skeleton frame contains the 3D coordinates of 31 body joints, resulting in a 93-dimensional frame-level representation. We subsequently construct a 93×93 covariance matrix from each action sequence to encode its joint-level motion statistics. Following the preprocessing procedure of brooks2019riemannian, we discard classes with insufficient samples and obtain 2,086 sequences belonging to 117 action categories. For evaluation, we follow the protocol introduced by Huang et al. (2018), where each sequence is assigned to one of the 117 action classes. The resulting covariance representations capture correlations among body-joint motions and provide a suitable geometric representation for structure-aware action recognition.

NTU60 (Shahroudy et al., 2016): NTU RGB+D 60 (NTU60) is a large-scale 3D skeleton-based action recognition dataset containing 56,880 sequences from 60 action categories. Each frame provides the 3D coordinates of 25 body joints. Following the joint-stream setting (Wang et al., 2025), we use only the joint information to represent each action sequence. Specifically, each sequence is transformed into a 75×10 Grassmannian matrix, where the 75-dimensional representation is obtained by concatenating the 3D coordinates of the 25 joints. For evaluation, we focus on the mutual-action classes and adopt the cross-view protocol following shahroudy2016ntu.

G.2 Implementation Details

All experiments are implemented in PyTorch and conducted on a single NVIDIA RTX 3090 GPU with 24 GB of GPU memory. We evaluate the proposed method on three sequential action recognition datasets, namely FPHA, HDM05, and NTU60. Unless otherwise specified, we follow the standard experimental protocols for each dataset. For clarity, we describe the implementation details using FPHA as an example.

For the FPHA dataset, we first transform the original sequential features into SPD covariance representations. Specifically, the input sequential features are first scaled by a factor of 10 and centred along the temporal dimension, followed by covariance estimation to obtain a 63×63 SPD matrix. To ensure numerical positive definiteness, a diagonal regularisation term proportional to the trace of the covariance matrix is added, with the regularisation coefficient set to 8×10^{-6}. The FPHA experiments follow the official training and validation split and contain 46 action categories.

Regarding the model architecture, we adopt a unified Grassmannian manifold-based feature learning framework. The input SPD representation is first processed by an SPD batch normalisation (Brooks et al., 2019) operation to reduce distribution shifts among different samples on the manifold. Subsequently, the top 10 orthogonal principal directions are extracted via singular value decomposition (SVD) (Stewart, 1993) and represented as 10 one-dimensional constituent subspaces. Based on these constituent subspaces, an initial constituent-subspace graph is constructed, and the structural relationships among different constituent subspaces are learned. The learned graph is then used to adaptively discover multiple composite subspaces according to its connectivity, where the constituent subspaces within the same structure are weighted according to their graph-based connectivity. After QR decomposition is applied to restore the orthogonal structure, an adaptive gating mechanism is further introduced to suppress redundant or weakly discriminative composite subspaces, thereby obtaining the final geometric representation.

The resulting geometric representation is subsequently fed into GrNet, where FRMap projects it onto a lower-dimensional Grassmannian space. The output subspace dimensions are set to 30 and 60 for FPHA and HDM05, respectively. The projected representations are then QR-reorthogonalised and transformed into matrix representations through the projection mapping $XX^\top$. The resulting representations are finally fed into the subsequent network modules for further feature learning and prediction. The model outputs class logits and is directly optimised using the cross-entropy loss.

G.3 Further Analysis

To verify whether the dimensionality of a subspace is sample-dependent, we divide the samples in the Salinas (Qing et al., 2019) dataset into five levels (Q1–Q5) according to their local spectral complexity, and examine both their intrinsic dimensionality and the dimensionality requirements of downstream tasks. As shown in Table 5, as the spectral complexity increases from Q1 to Q5, the effective spectral dimension D_{eff} increases from 1.74 to 7.42, while the dimension D_{req} required to preserve 99% of the local spectral energy increases from 7.28 ± 1.75 to 16.90 ± 1.54, indicating substantially different intrinsic dimensionality requirements across samples with different complexity levels. Furthermore, as shown in Table 3, the minimum dimension p^*_{cls} required to achieve stable downstream classification performance also varies with complexity: Q1 requires only 5 dimensions, whereas Q2–Q5 require at least 8 dimensions. These results indicate that the subspace dimensionality is not a universally suitable fixed quantity for all samples, but is closely related to the intrinsic structural complexity of the data.

This observation remains consistent under different analysis settings. As shown in Table 6, D_{req} and p^*_{cls} exhibit a consistent increasing trend, with a Spearman correlation coefficient of 0.707, indicating that variations in intrinsic structural complexity are reflected in the dimensionality requirements of downstream tasks. As shown in Table 7, D_{req} increases monotonically with complexity across different patch sizes, with a Spearman correlation coefficient of 1.000 in all cases, further confirming the robustness of the complexity–dimensionality relationship. Overall, these results demonstrate that a fixed subspace dimensionality is insufficient to accommodate the heterogeneous structural requirements of different Grassmann points. Therefore, a more appropriate Grassmannian modelling strategy should adaptively determine the effective subspace dimensionality according to the intrinsic structure of each Grassmann point.

Table 5: Sample-wise intrinsic dimensionality across different levels of spectral complexity on the Salinas dataset. D_{eff} measures the effective spectral dimension, while k_{95} and D_{req} denote the minimum spectral dimensions required to preserve 95% and 99% of the local spectral energy, respectively.

Complexity Level	# Patches	D_{eff}	k_{95}	D_{req}
Q1 (Very Low)	10826	1.74	2.67	7.28 ± 1.75
Q2 (Low)	10826	2.74	4.80	10.30 ± 2.66
Q3 (Medium)	10826	4.31	6.98	13.69 ± 1.10
Q4 (High)	10826	5.60	8.20	15.19 ± 0.82
Q5 (Very High)	10825	7.42	9.87	16.90 ± 1.54

Table 6: Comparison between the intrinsic required dimension and the dimension required for downstream recognition. D_{req} is determined from representation fidelity, while p^*_{cls} is determined from downstream classification performance.

Complexity Level	D_{req}	p^*_{cls}	$\lvert D_{\text{req}} - p^*_{\text{cls}} \rvert$	OA at p^*_{cls} (%)
Q1 (Very Low)	7.28 ± 1.75	5	2.28	95.12
Q2 (Low)	10.30 ± 2.66	8	2.30	96.49
Q3 (Medium)	13.69 ± 1.10	8	5.69	95.94
Q4 (High)	15.19 ± 0.82	8	7.19	96.34
Q5 (Very High)	16.90 ± 1.54	8	8.90	97.11
Spearman ρ		0.707		

Table 7: Robustness of the complexity–dimensionality relationship under different patch sizes.

Patch Size	Q1 D_{req}	Q2 D_{req}	Q3 D_{req}	Q4 D_{req}	Q5 D_{req}	Spearman ρ
5×5	7.34	10.26	11.26	12.02	13.01	1.000
7×7	7.22	11.43	13.28	14.44	15.82	1.000
9×9	7.28	10.30	13.69	15.19	16.90	1.000
11×11	7.54	9.23	12.81	15.29	17.24	1.000