Learning Topology-Driven Multi-Subspace Fusion for Grassmannian Deep Networks

Xuan Yu, Tianyang Xu*

School of Artificial Intelligence and Computer Science, Jiangnan University
Wuxi, Jiangsu 214122, China
xuan.yu@stu.jiangnan.edu.cn, tianyang.xu@jiangnan.edu.cn

Abstract

Grassmannian manifold offers a powerful carrier for geometric representation learning by modelling high-dimensional data as low-dimensional subspaces. However, existing approaches predominantly rely on static single-subspace representations, neglecting the dynamic interplay between multiple subspaces critical for capturing complex geometric structures. To address this limitation, we propose a topology-driven multi-subspace fusion framework that enables adaptive subspace collaboration on the Grassmannian. Our solution introduces two key innovations: (1) Inspired by the Kolmogorov-Arnold representation theorem, an adaptive multi-subspace modelling mechanism is proposed that dynamically selects and weights task-relevant subspaces via topological convergence analysis, and (2) a multi-subspace interaction block that fuses heterogeneous geometric representations through Fréchet mean optimisation on the manifold. Theoretically, we establish the convergence guarantees of adaptive subspaces under a projection metric topology, ensuring stable gradient-based optimisation. Practically, we integrate Riemannian batch normalisation and mutual information regularisation to enhance discriminability and robustness. Extensive experiments on 3D action recognition (HDM05, FPHA), EEG classification (MAMEM-SSVEPII), and graph tasks demonstrate state-of-the-art performance. Our work not only advances geometric deep learning but also successfully adapts the proven multi-channel interaction philosophy of Euclidean networks to non-Euclidean domains, achieving superior discriminability and interpretability.

Code — https://github.com/Xua-Yu/GMSF-Net
Extended version — https://arxiv.org/abs/2511.08628

Introduction

In recent years, unitary subspace modelling on the Grassmannian has achieved great success in tasks which require extended discriminative capacity, such as activity recognition (Cherian et al. 2017), emotion recognition (Liu et al. 2014), face verification (Huang et al. 2015), and classification of time-series data in brain-computer interfaces (Gao et al. 2022). In the above tasks, Grassmannian typically model the input data with a single and static orthogonal subspace (Huang, Wu, and Van Gool 2018; Wang et al. 2024; Xu, Zhu, and Wu 2023). However, a fixed single-subspace assumption fails to capture local geometric variations and multi-modal distribution, limiting the expressiveness and adaptability of the model in structurally diverse or heterogeneous task scenarios (Hamm and Lee 2008). To overcome this issue, we introduce an adaptive multi-subspace representation that dynamically adjusts multiple subspaces to accurately capture task-specific characteristics and the harmonious distribution among subspaces. The convergence of obtaining adaptive subspaces is analysed from a topological perspective (Munkres 2000), satisfying specific demands across different tasks.

In deep neural networks, different vectors in the Euclidean space reflect their specific feature identities. However, the existing Euclidean metric (Olver and Shakiban 2006) driven by the inner product, measures the semantic difference between two vectors, resulting in a distorted relevance in task-specific decisions (as shown in Figure 1 (a) and (d)). Based on non-Euclidean geometry, this distorted relevance can be effectively addressed on Riemannian manifolds. Intuitively, Riemannian manifolds provide compact spaces embedded in the dense Euclidean space, where the accessible path can preserve the geometric consistency (Boumal 2023). As a typical Riemannian manifold, the Grassmannian (Bendokat, Zimmermann, and Absil 2024) can model high-dimensional data as linear subspaces to effectively address semantic distortion. Despite this merit, existing solutions (Huang, Wu, and Van Gool 2018; Wang et al. 2024) are primarily driven by a single subspace, lacking consideration of the collaborative interaction of multi-subspaces, which hinders the characterisation of subspace geometry diversity in terms of extended semantics (as shown in Figure 1 (b) and (e)). In the Euclidean space, the promising progress of deep learning in the recent decade is attributed to multi-channel interactions and non-linear activations (Krizhevsky, Sutskever, and Hinton 2012). In particular, the seminal work LeNet-5 (LeCun et al. 2002) extracts multi-channel feature maps using stacked convolution blocks, thus obtaining extended representative capacity via channel interaction. Admitting the natural representation power of linear subspace (Xu, Wu, and Kittler 2018), the potential of multi-subspace interactions has been overlooked in the community. To reveal the power

*Corresponding author.

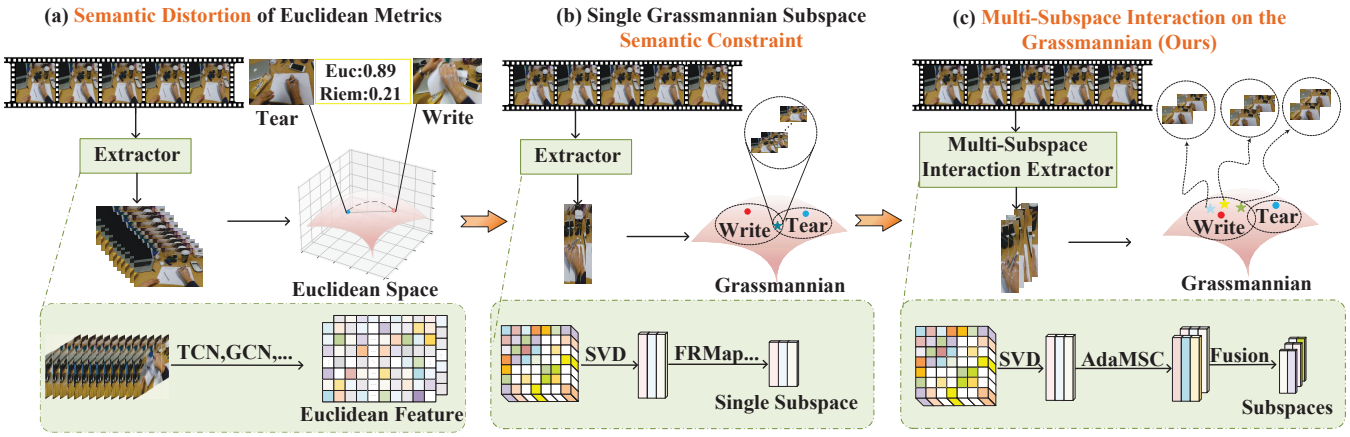

Figure 1: Illustration of the problem evolution. (a) shows semantic distortion of features in Euclidean space. (b) reveals semantic constraint using a single Grassmannian subspace. (c) presents the proposed framework incorporating multi-Grassmannian subspace fusion.

of multi-subspace interaction for the Grassmannian, there are two major issues, *i.e.*, *how to perform subspace interaction and how to formulate its deep architecture.*

For the first issue, we design a multi-subspace interaction block in a geometric manner, *i.e.*, several intermediate representations are generated by a unified mapping and then fused through alignment, facilitating effective interaction among different Grassmannian subspaces. For the second issue, we leverage an intrinsic geometric approach to ensure that the fusion process occurs at accurate positions on the manifold (Tao et al. 2024), which enables the design of our multi-subspace interaction block in a stackable form, allowing for flexible capacity expansion (Song et al. 2025). To the best of our knowledge, this is the first work that introduces deep interactions of Grassmannian subspaces in Riemannian neural networks, as shown in Figure 1 (c) and (f), these deep interactions capture the dynamic relationships between multiple subspaces, effectively modelling their geometric structures, semantic features, and mutual influences. Euclidean deep networks naturally implement backpropagation, given that the gradient superposition property satisfies the axioms of linear spaces (Fei et al. 2025). However, parameter optimisation on a Riemannian manifold must be strictly restricted to its tangent space (Smith 2014). To this end, we develop a topology-driven framework for adaptive multi-subspace construction on the Grassmannian. In summary, our main contributions are the following:

- We propose an adaptive multi-subspace modelling approach tailored to diverse recognition tasks.
- We propose a novel Grassmannian Multi-Subspace Fusion network (GMSF-Net) to fuse heterogeneous subspace representations.
- We propose a topology-driven framework with theoretical guarantees for adaptive Grassmannian subspaces construction.

Related Work

The Grassmannian is the set of all linear subspaces of a fixed dimension, showing notable algebraic and geometric merits, reflecting the stability and invariance of subspace geometric structures (Souza et al. 2020). In related research, deep learning applications on the Grassmannian have evolved from shallow methods to deep models.

In the early research stage, shallow Grassmannian learning (Zhang et al. 2018) is achieved by projecting subspaces onto the manifold, such as discriminant analysis (Baudat and Anouar 2000), high-dimensional data clustering (Von Luxburg 2007), and low-rank matrix completion (Dai, Kerman, and Milenkovic 2012). However, shallow learning on the manifold cannot mine deep geometric characteristics from the data, thus only suitable for simple tasks. The development of GrNet breaks this limitation by formulating a deep network version of the Grassmannian (Huang, Wu, and Van Gool 2018). In particular, GrNet projects data onto the Grassmannian and follows the manifold's geometric properties to design network layers such as FRMap, OrthMap, and ProjMap. A brief summary of GrNet is provided in Appendix B. Typically, GrNet considers the manifold's constraints and the low-dimensional subspace representation of the data, thereby enhancing the model capacity to learn delicate geometric structures and effectively solve nonlinear relationships and high-dimensional problems.

In recent studies, equipped with the self-attention mechanism, models can focus on the global relevance of the input and build long-range dependencies among them, greatly enhancing the model capacity (Vaswani et al. 2017). To this end, to explore the potential of self-attention on the Grassmannian, GDLNet (Wang et al. 2024) integrates the mechanism to build dependencies between different subspaces. It helps the model effectively focus more on important subspace information. However, GrNet and GDLNet typically adopt a single and static subspace to model the original global features, which makes it difficult to extract diverse geometric subspaces, thus delivering internal disturbances

and weaker discriminative power. To address this issue, we introduce an adaptive multi-subspace construction mechanism to capture coherent geometric features. Additionally, we design a multi-subspace interaction block to fuse different subspaces, facilitating intra-subspace transformation and inter-subspace complement.

Preliminaries

This section provides a brief overview of the basic geometries of Grassmannian and Topology. More detailed review and relevant notations can be found in Appendix A.

Grassmannian

Grassmannian geometries: The Grassmannian $\mathcal{G}(n, p)$ is the set of all p-dimensional linear subspaces of $\mathbb{R}^n$. Each point represents a p-dimensional subspace, typically represented by an orthonormal basis matrix U. The tangent space, exponential map, and logarithmic map are key tools to understand its geometric structure, and these concepts will serve as prerequisites for subsequent operations (a detailed explanation is provided in Appendix A.2).

Grassmannian Metric: Points on the Grassmannian represent different subspaces, and their distances are typically measured by metrics such as the Projection metric (Huang et al. 2015) or Principal Angles metric (Qiu, Zhang, and Li 2005). Considering computational efficiency, this paper adopts the Projection metrics to analyse points on the manifold. For two subspaces X_1 and X_2, the Projection metric is defined as:

$$d_p(X_1, X_2) = 2^{-1/2}\|X_1X_1^T - X_2X_2^T\|_F, \qquad (1)$$

where $\|\cdot\|_F$ is the Frobenius norm (Harandi et al. 2013). This metric serves as an effective tool for quantifying the distance between subspaces and enabling geometric analysis on the Grassmannian (Fei et al. 2025).

Topology

Topological spaces underpin Riemannian geometry by abstracting spatial properties beyond specific shapes (Lee 2000). On the Grassmannian, each subspace inherently constitutes a topological space. This topological framework rigorously characterises subspace structure and metric, ensuring the critical convergence property for sequences of subspaces. This theoretical guarantee of convergence motivates the design of adaptive multi-subspace modelling, ensuring stable convergence towards limiting subspaces during optimisation and enabling efficient geometric feature extraction and fusion. This section provides a brief introduction to the topological structure, with detailed definitions and proofs presented in Appendices A.3 and F.

Approach

In this section, we will systematically introduce our method. We first introduce the adaptive multi-subspace encoder on the Grassmannian, analysing its design principles in depth and proving its convergence based on topological theory. Subsequently, we explain the design of the multi-subspace interaction block based on a geometric approach in detail. Figure 2 (a) presents an overview of GMSF-Net.

Adaptive Multi-Subspace Encoder (AMSE)

This subsection introduces the Adaptive Multi-Subspace Encoder (AMSE) on the Grassmannian, which adaptively constructs multiple coherent subspaces to preserve the diverse geometric characteristics of data, as illustrated in Figure 2 (b). This subsection introduces the construction of adaptive multiple subspaces and the analysis of their convergence in detail.

Adaptive Multi-Subspace Construction (AdaMSC) The construction of adaptive multi-subspaces is the core of AMSE, as shown in Figure 2 (c), which builds multiple stable and harmonious low-dimensional subspace representations on the high-dimensional Grassmannian space $\mathcal{G}(n, k)$, and adaptively adjusts these subspaces based on different tasks. In particular, we first extract frame-level features and compute a covariance matrix $X \in \mathbb{R}^{n\times n}$ to model the statistical dependencies of features along the temporal dimension (Nguyen et al. 2019). Subsequently, we apply the Schmidt orthogonalisation process to map X into a set of low-dimensional orthogonal subspaces (Yaghooti, Raviv, and Sinopoli 2024):

$$\mathcal{S} = \{S_1, S_2, \ldots, S_k\}, \quad S_i^\top S_j = 0 \quad (\forall\, i \neq j), \qquad (2)$$

where each subspace $S_j \in \mathcal{G}(n, 1)$. This design aims to decouple statistical information from physical information in the features, transforming them into a set of localised geometric encodings that facilitate subsequent subspace modeling (Edelman, Arias, and Smith 1998).

We define the index set of subspaces as $\mathcal{I} = \{i_1, \ldots, i_k\}$, where each index i_j corresponds to a subspace $S_j \in \mathcal{G}(n, 1)$. For each new subspace $S'_{m'}$, we initialise a set of learnable weight parameters $\mathcal{W}^{(m')} = \{w_1^{(m')}, \ldots, w_k^{(m')}\}$, which are normalised via the Softmax function to obtain the weight distribution $\widetilde{\mathcal{W}}^{(m')} = \{\widetilde{w}_1^{(m')}, \ldots, \widetilde{w}_k^{(m')}\}$. Each $\widetilde{w}_j^{(m')}$ reflects the relative importance of the atomic subspace S_j during the construction of S'_m. We then select the top p indices with the highest values in $\widetilde{\mathcal{W}}^{(m')}$, obtaining the index set $\{i_{j_1}, i_{j_2}, \ldots, i_{j_p}\}$ and the corresponding weights $\{\widetilde{w}_{j_1}^{(m')}, \ldots, \widetilde{w}_{j_p}^{(m')}\}$. The m'-th new subspace is constructed from these weighted subspaces, with the construction defined as above based on Theorem F.1 in the Appendix:

$$S'_{m'} = [\widetilde{w}_{j_1}^{(m')} S_{j_1},\ \widetilde{w}_{j_2}^{(m')} S_{j_2},\ \ldots,\ \widetilde{w}_{j_p}^{(m')} S_{j_p}]. \qquad (3)$$

In this way, the new subspace $S'_{m'} \in \mathcal{G}(n, p)$ is composed of distinct key atomics which enable efficient modelling and adaptive adjustment of diverse features and lay a solid foundation for subsequent multi-subspace fusion. The detailed process is shown in algorithm 1.

Next, we introduce the mutual information function $f(S')$ to measure the complementarity between different new subspaces S'_i and S'_j (Liu et al. 2009). This function aims to enhance the discriminative ability of the overall representation by maximising the information complementarity between subspaces:

$$f(S') = I(S'_i, S'_j), \quad i \neq j, \quad i, j \in \{1, \ldots, m'\}. \qquad (4)$$

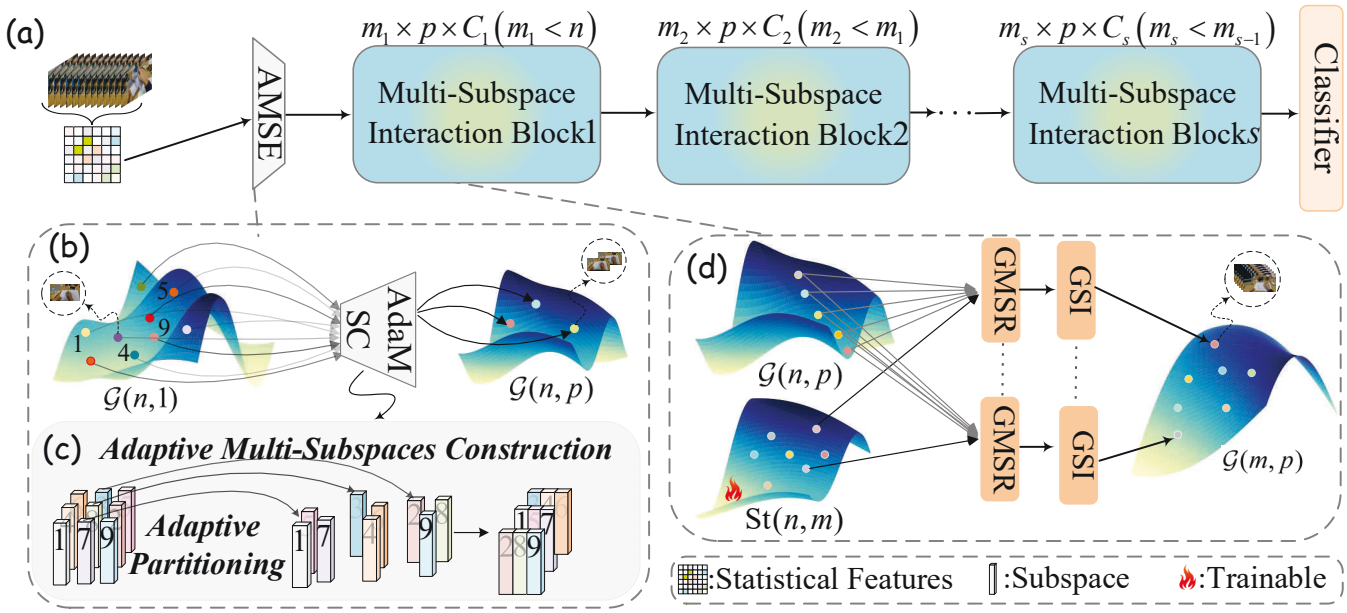

Figure 2: *Overview of GMSF-Net.* (a) Architecture of GMSF-Net. (b) Structure of the Adaptive Multi-Subspace Encoder (AMSE). (c) *Fine-Grained Design.* The proposed efficient Adaptive Multi-Subspace Construction (AdaMSC). (d) *Subspace Interaction Design.* The proposed discriminative Multi-Subspace Interaction Block.

Algorithm 1: Adaptive Multi-Subspace Construction

Input: Low-dimensional subspaces $\{S_i\}_{i=1}^{k}$, $S_i \in \mathcal{G}(n,1)$, subspace dimension p, number of output subspaces m'
Parameter: Importance matrix $\mathcal{W} \in \mathbb{R}^{m' \times k}$
Output: Weighted multi-subspaces $\{S_i'\}_{i=1}^{m'}$
$\mathcal{W}' \leftarrow \text{Softmax}(\mathcal{W}, \text{dim} = 1)$ *# row-wise softmax*
for $i = 1$ **to** m' **do**
  Sort indices: $I_i \leftarrow argsort_{descending}(\mathcal{W}'[i,:])$
  Select top-p components: $\widetilde{S}_i \leftarrow S_i[:, I_i[1:p]]$
  Get corresponding weights: $\mathbf{w}_i \leftarrow \mathcal{W}'[i, I_i[1:p]]$
  $T_i \leftarrow \widetilde{S}_i \odot (\mathbf{1}_n \cdot \mathbf{w}_i^\top)$ {$\mathbf{1}_n$: all-ones vector}
end for
Concatenate subspaces: $\{S_i'\}_{i=1}^{m'} \leftarrow \bigcup_{k=1}^{m'} \{T_k\}$

The specific formula for the mutual information $I(S_i', S_j')$ is:

$$I(S_i', S_j') = \int \int p(S_i', S_j') \log \frac{p(S_i', S_j')}{p(S_i')p(S_j')} \, dS_i' \, dS_j'. \quad (5)$$

In practical modelling, since it is difficult to directly obtain the probability distributions $p(S_i')$ and $p(S_j')$ for the subspaces, we typically use embedding or feature-based estimation methods to approximate the mutual information (for detailed information see the *Optimisation Strategies*) (Maeda, Kawaguchi, and Tezuka 2025).

Analysis of Adaptive Subspace Convergence We further analyse multi-subspace construction on the Grassmannian from a topological perspective to ensure model convergence. By introducing a metric topology and projection metric, we ensure the subspaces converge iteratively. In the metric topology, the distance on the Grassmannian is defined by the projection metric, as shown in Equation (1). This metric satisfies non-negativity, symmetry, and the triangle inequality, ensuring that the Grassmannian is a metric topology. The proof is in Appendix F.2.

In this metric topology, the convergence of a sequence of subspaces is naturally defined as detailed in Proposition F.3 in the Appendix. During training (Ionescu, Vantzos, and Sminchisescu 2015), the subspace is iteratively updated through Riemannian gradient descent, with the update rule as follows:

$$S'(t+1) = \text{Exp}_{S'(t)} \left(-\eta \, \text{Proj}_{T_{S'(t)}} \nabla L(S') \right), \quad (6)$$

where $\text{Exp}_{S'(t)}$ is the exponential map, $\text{Proj}_{T_{S'(t)}}$ is the tangent space projection, $\nabla L(S')$ is the gradient,and the step size η controls the learning rate. The proof is in Appendices F.3 to F.5. Due to the continuity and differentiability of the loss function with a lower bound, along with the completeness of the manifold and the continuity of the exponential map (Eschenauer and Olhoff 2001), this subspace eventually satisfies (see Appendix F.6):

$$d(S'(t), S^*) \to 0 \quad \text{as} \quad t \to \infty. \quad (7)$$

This means that in the topology induced by the projection metric, the sequence converges to the stable subspace S^*, ensuring the convergence and consistency of the model's geometric modelling process.

Multi-Subspace Interaction Block

The multi-subspace interaction block fuses multiple Grassmannian subspaces into a single integrated subspace, as

shown in Figure 2, effectively capturing geometric features of data and enhancing the ability to represent complex geometric cues. This block consists of two main components: Grassmannian Multi-Representations and Subspace Interaction.

This work introduces a semantic perspective for multi-subspace interaction design, addressing the often-overlooked intuition behind such interactions. We posit that distinct subspaces, governed by different geometric frameworks, capture complementary task-relevant information through diverse local geometric features. Their interaction transcends mere feature fusion, embodying semantic co-aggregation within the manifold. By incorporating alignment constraints and weighted aggregation, our model dynamically integrates multi-subspace representations, enhancing both discriminability and geometric diversity.

Grassmannian Multi-Subspace Representations (GMSR) Grassmannian Multi-Subspace Representations (GMSR) generates subspace representations with distinct geometric features under a unified geometric framework by applying the same learnable matrix to a set of input subspaces $\{S'_{m'}\}$, producing a set of output representations $\{X^{c,m'}_{GMSR}\}$. The formulation is as follows:

$$X^{c,m'}_{GMSR} = f^{(c)}_{GMSR}(S'_{m'}, W_c) = W_c^T S'_{m'}, \tag{8}$$

where $S'_{m'} \in \mathcal{G}(n,p)$ is the m-th output subspace from the AMSE module, $W_c \in \mathbb{R}^{n\times m}, (m < n)$ is the mapping matrix that is basically required to be a full rank matrix (Huang and Van Gool 2017), $X^{c,m'}_{GMSR} \in \mathcal{G}(m,p)$ is the new Grassmannian representations. Generally, $W_c^T S'$ is not an orthonormal basis matrix, we employ QR or SVD decomposition to preserve the intrinsic geometric structure of the data. To obtain a closed-form expression for the mapping matrix W_c during the optimisation process, we constrain it to the Stiefel manifold $St(n,m)$ (Huang and Van Gool 2017). We define c geometric frameworks as mapping matrices $\{W_1, \ldots, W_c\}$ capturing different subspace representations, where c denotes the number of geometric frameworks. The learnability of the mapping matrix enables the model to adaptively adjust the geometric framework, capture the geometric information of multiple subspaces, and improve modelling capacity for geometric feature extraction and subspace fusion.

Grassmannian Subspace Interaction (GSI) In the GMSR layer, we construct subspace representations under various geometric frameworks c. The input is a set of m' Grassmannian subspace points denoted as $\{P_i\}^{m'}_{i=1}$, where each $P_i = X^{c,i}_{GMSR} \in \mathcal{G}(m,p)$. To enable effective fusion of the set $\{P_i\}^{m'}_{i=1}$, we introduce the Fréchet mean to preserve the intrinsic geometric structure (Pennec, Fillard, and Ayache 2006). Specifically, Fréchet mean Γ is computed as the point that minimises the sum of squared distances on the manifold (Yang, Arnaudon, and Barbaresco 2010):

$$\Gamma = \mathcal{F}_w(\{P_i\}_{i\le m'}) := arg \min_{\Gamma\in\mathcal{G}(m,p)} \sum_{i=1}^{m'} w_i \delta_R^2(\Gamma, P_i), \tag{9}$$

where, w_i denotes the weight of each point, satisfying $w_i \geq 0$ and $\sum_{i=1}^{m'} w_i = 1$. The function $\delta_R(\cdot,\cdot)$ represents the geodesic distance on the Grassmannian.

When $m' = 2$, *i.e.*, the weights are given by $\{w, 1-w\}$, a closed-form solution exists, which exactly corresponds to the geodesic between the two points P_1 and P_2, parameterised by $w \in [0,1]$:

$$\Gamma = V\left[P_1\cos(w\Theta) + U\sin(w\Theta)\right]V^\top, \tag{10}$$

where $P_1^\top P_2 = V\cos(\Theta)Q^\top$ is part of its singular value decomposition (SVD), U, Θ, V come from the SVD of $(I - P_1P_2^\top)P_2V$, and the parameter $w \in [0,1]$ controls the position of the fused point (Beik-Mohammadi et al. 2021).

When $m' > 2$, the optimisation problem does not have a closed-form solution. Therefore, the Karcher flow algorithm is employed to perform optimisation in the tangent space iteratively, ultimately converging to the Fréchet mean through exponential mapping. The update formula of the Karcher flow (Karcher 1977; Yang, Arnaudon, and Barbaresco 2010) algorithm is as follows (Appendix C provides detailed steps):

$$\Gamma_{new} = \exp_\Gamma\left(\Gamma, \alpha\cdot\frac{1}{m'}\sum_{i=1}^{m'}\log_\Gamma(\Gamma, P_i)\right), \tag{11}$$

where α is the step size, Γ is the current mean, and P_i represents each subspace within the same geometric framework. $\log_G(\cdot,\cdot)$ and $\exp_G(\cdot,\cdot)$ are provided in the Appendix A.

Deep Subspace Interaction and Optimisation Principles

Deep Subspace Interaction Principles: We propose a stackable GMSF-Net by applying multiple blocks to meet the flexible and diverse requirements of practical applications. The core principle is to constrain the feature manifold distribution through Riemannian batch normalisation, thereby enhancing geometric discriminability while stabilising the feature space topology. The feature construction based on Riemannian BN provides navigation signals with geometric interpretability, enabling the stacking structure to systematically enhance and refine the features.

For the input of the AMSE block, *i.e.* statistical features, we introduce a Riemannian normalisation strategy (Brooks et al. 2019) based on manifold geometry, mapping statistical features to the SPD manifold (Nguyen et al. 2019) and normalising them to extract more discriminative key Riemannian features (see Appendix E for the detailed implementation of SPDBN):

$$\hat{X} = \exp_{\hat{B}}\left(\hat{B}^{1/2}\left((\hat{B}B)^{-1/2}X(\hat{B}B)^{-1/2}\right)\hat{B}^{1/2}\right), \tag{12}$$

where $\hat{X}$ is the refined feature, X is the original input, B is the Riemannian barycenter, and $\hat{B}$ is the learnable common feature, which is iteratively updated using the Riemannian geodesic formula (Yair, Ben-Chen, and Talmon 2019).

Optimisation Strategies: In order to effectively train the proposed GMSF-Net, we restructure the loss function by combining classification cross-entropy loss (Hinton et al.

Method	Acc.(%)	Size(MB)	FLOPs(M)
SPDNet	87.65%±1.02	13.60	1595.50
GrNet	78.79%±1.82	6.73	38.60
MATT	87.70%±0.68	1.83	142.07
SPDNetBN	89.33%±0.49	13.63	1902.97
GDLNet	87.60%±0.69	1.83	**33.69**
GMSF-Net-1Block	90.43%±0.74	**1.20**	48.42
GMSF-Net-2Blocks	90.70%±0.70	1.26	63.44
GMSF-Net-3Blocks	**91.22%±0.53**	1.30	81.07

Table 1: Comparison Between GMSF-Net and Other Riemannian Solutions on the FPHA Dataset.

Method	Acc.(%)	Size(MB)	FLOPs(M)
SPDNet	60.45%±1.12	11.71	2050.49
GrNet	59.23%±1.78	6.88	**81.96**
MATT	62.25%±1.68	3.99	442.45
GDLNet	60.08%±1.78	3.95	89.11
GMSF-Net-1Block	63.64%±1.24	**3.33**	96.08
GMSF-Net-2Blocks	63.98%±1.07	3.34	100.07
GMSF-Net-3Blocks	**64.19%±0.88**	3.39	114.12

Table 2: Comparison Between GMSF-Net and Other Riemannian Solutions on the HDM05 Dataset.

1995) and MI-inspired regularizer loss, effectively capturing diverse geometric subspaces. The total loss function for the proposed method is as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{CE} + \lambda \cdot \mathcal{L}_C, \tag{13}$$

where, $\mathcal{L}_{total}$ is the total loss, $\mathcal{L}_{CE}$ is the cross-entropy loss, $\mathcal{L}_C$ is the MI-inspired regularizer loss, and λ is a hyperparameter that balances classification accuracy with structural preservation. The details of these two loss functions (Appendix D.1) and their backpropagation derivations (Appendix D.2) are provided in Appendix D.

Evaluation

In this section, we evaluate the performance of the proposed GMSF-Net on two challenging classification tasks: HDM05 (Müller et al. 2007) and FPHA (Garcia-Hernando et al. 2018) datasets for video-based 3D action recognition, and the MAMEM-SSVEP-II (Pan, Chou, and Wei 2022) dataset for EEG signal classification. Additionally, we perform node classification (NC) and link prediction (LP) tasks on graph datasets, where the underlying structure of the data can naturally be represented as Grassmannian, possessing multi-subspace geometric properties. More experimental details are presented in Appendix G.

Experimental Evaluation on the Classification Datasets

We evaluate the proposed GMSF-Net on video-based 3D action recognition and EEG signal classification tasks. For action recognition, we use the FPHA and HDM05 datasets, following the experimental settings of SymNet (Wang, Wu,

Method	Acc.(%)	Size(MB)	FLOPs (M)
EEGNet	53.72%±7.23	**0.075**	60.13
ShallowConvNet	56.93%±6.97	0.18	127.00
EEG-TCNet	55.45%±7.66	0.016	60.80
SCCNet	62.11%±7.70	0.55	108.56
MBEEGSE	56.45%±7.27	1.64	52.40
FBCNet	53.09%±5.67	0.24	**51.80**
SPDNet	62.30%±3.12	2.81	2274.56
GrNet	61.23%±3.56	1.95	2020.78
MATT	65.19%±3.14	1.97	2068.44
SPDNetBN	62.76%±3.01	2.81	2312.78
GDLNet	65.52%±2.86	1.95	2028.34
GMSF-Net-1Block	66.74%±1.79	1.94	2020.14
GMSF-Net-2Blocks	66.32%±1.84	1.94	2021.78
GMSF-Net-3Blocks	**66.87%±1.46**	1.94	2032.55

Table 3: Performance comparison between GMSF-Net and the baseline on SSVEP.

and Kittler 2021) and GrNet (Huang, Wu, and Van Gool 2018). For EEG classification, we follow the preprocessing pipeline of GDLNet (Wang et al. 2024) and conduct experiments on the MAMEM-SSVEP-II dataset. In the model evaluation, we compare performance involving different numbers of multi-subspace interaction blocks. As performance saturates with more blocks, we report results with up to three blocks in this section. For each task, we highlight the best performance results in bold.

FPHA: Following (Wang, Wu, and Kittler 2021), we evaluate GMSF-Net using three network configurations: 1Block, 2Blocks, and 3Blocks. The results from 10 random experiments (mean ± std) are shown in Table 1. Notably, GMSF-Net consistently outperforms most Riemannian deep models, with a **12.43%** improvement over GrNet. As the number of blocks increases, performance improvements plateau, yet the effectiveness of the adaptive subspace interaction method remains evident. Additionally, our model demonstrates significant advantages in terms of model size and computational cost.

HDM05: Following (Huang, Wu, and Van Gool 2018), we also adopt three architectures to evaluate GMSF-Net: 1Block, 2Blocks, and 3Blocks configurations. The 10-fold results (mean ± std) are presented in Table 2. It should be noted that GMSF-Net achieves a performance improvement of up to **4.96%** over GrNet. This highlights the effectiveness of our approach and significantly reduces memory and computational costs compared to SPD manifold-based neural networks. Results on HDM05 and FPHA demonstrate that GMSF-Net consistently outperforms other methods in 3D action recognition.

MAMEM-SSVEP-II: In Table 3, we follow the protocol of GDLNet and report the results of ten random cross-validation runs (mean ± std). It can be observed that GMSF-Net outperforms existing Riemannian models as well as EEG deep learning models, including EEGNet (Lawhern et al. 2018), ShallowConvNet (Schirrmeister et al. 2017), EEG-TCNet (Ingolfsson et al. 2020), SCCNet (Wei, Koike-

Dataset	Disease		Airport		PubMed		CoRA	
Task	**LP**	**NC**	**LP**	**NC**	**LP**	**NC**	**LP**	**NC**
Euc	59.8%±2.0	32.5%±1.1	92.0%±0.0	60.9%±3.4	83.3%±0.1	48.2%±0.7	82.5%±0.3	23.8%±0.7
Hyp	63.5%±0.6	45.5%±3.3	94.5%±0.0	70.2%±0.1	87.5%±0.1	68.5%±0.3	87.6%±0.2	22.0%±1.5
Euc-Mixed	49.6%±1.1	35.2%±3.4	91.5%±0.1	68.3%±2.3	86.0%±1.3	63.0%±0.3	84.4%±0.2	46.1%±0.4
Hyp-Mixed	55.1%±1.3	56.9%±1.5	93.3%±0.0	69.6%±0.1	83.8%±0.3	**73.9%**±0.2	85.6%±0.5	45.9%±0.3
MLP	72.6%±0.6	28.8%±2.5	89.8%±0.5	68.6%±0.6	84.1%±0.9	72.4%±0.2	83.1%±0.5	51.5%±1.0
HNN	75.1%±0.3	41.0%±1.8	90.8%±0.2	80.5%±0.5	94.9%±0.1	69.8%±0.4	89.0%±0.1	54.6%±0.4
GMSF-Net-1Block	95.0%±1.0	79.7%±2.9	**93.4%**±0.2	**82.5%**±0.8	94.6%±0.1	72.7%±0.5	88.6%±0.6	**57.8%**±1.2
GMSF-Net-2Blocks	94.6%±0.7	81.1%±1.9	92.7%±0.6	80.7%±1.0	**95.0%**±0.1	72.8%±0.4	**89.3%**±0.6	55.8%±1.5
GMSF-Net-3Blocks	**95.5%**±0.4	**82.7%**±0.8	91.2%±0.2	78.3%±0.6	93.6%±0.3	72.9%±0.3	88.9%±0.3	56.2%±1.1

Table 4: Performance comparison between GMSF-Net and the baselines on Graph Tasks.

	With Interaction			Without Interaction		
Configuration	**HDM05**	**FPHA**	**SSVEP**	**HDM05**	**FPHA**	**SSVEP**
Adaptive-Subspace	**63.64%**±1.24	**90.43%**±0.74	**66.74%**±1.79	**56.49%**±1.37	**80.68%**±2.15	**59.83%**±2.72
Random-Subspace	50.29%±2.10	72.47%±3.19	56.01%±2.46	52.50%±1.82	73.10%±2.19	57.43%±1.10
Fixed-Subspace	53.04%±0.88	83.06%±0.61	66.05%±1.95	51.12%±3.55	76.67%±0.16	58.85%±1.19

Table 5: Performance Comparison of Subspace Construction and Interaction Mechanisms.

Akino, and Wang 2019), MBEEGSE (Altuwaijri et al. 2022), and FBCNet (Mane et al. 2021), in terms of overall performance. The accuracy is improved by **5.64%** over GrNet, and by **1.68%** and **1.35%** over the Riemannian self-attention models MATT (Pan, Chou, and Wei 2022) and GDLNet, with model stability also enhanced. EEG visualisation results are shown in Appendix G.4. These results further highlight the versatility and effectiveness of our framework.

Experimental Evaluation on Graph Datasets

To evaluate the broad applicability of GMSF-Net, we perform node classification (NC) and link prediction (LP) on graph datasets with prominent Grassmannian subspace structures. The implementation details is provided in Appendix G.2.2. The experimental results show that increased deviation from the ideal Grassmannian structure leads to performance degradation due to modelling limits on non-natural manifold geometries. In contrast, GMSF-Net exhibits greater performance gains on datasets more aligned with the Grassmannian assumption, validating the soundness and adaptability of its design. Specifically, GMSF-Net achieves notable gains on Cora (Sen et al. 2008) and Disease (Chami et al. 2019) datasets, which exhibit clear subspace structures, while improvements are more modest on structurally complex PubMed (Sen et al. 2008) and Airport (Chami et al. 2019) datasets that deviate from ideal subspace assumptions. GMSF-Net achieves notable gains on all graph datasets except PubMed, as shown in Table 4.

Ablation Experiments and Analysis

As shown in Table 5, we conducted six ablation studies across three datasets to evaluate the effectiveness of the adaptive subspace and interaction mechanisms. The results show that the framework combining adaptive subspace construction with the interaction mechanism consistently achieves the best performance (HDM05: 63.64%, FPHA: 90.43%, SSVEP: 66.74%), significantly outperforming other configurations. This validates the necessity of adaptive mechanisms for optimising the feature space and highlights the synergistic enhancement provided by the interaction mechanism within high-confidence subspaces. As shown in Table 5, the random subspace achieves the lowest performance (*e.g.*, HDM05: 52.50% $\rightarrow$ 50.29%), indicating that low-quality interactions may introduce noise. While the fixed subspace reaches 83.06% on FPHA, its static structure limits its ability to adapt to diverse geometric characteristics. The adaptive subspace is essential for effectively collaborating with the interaction mechanism to capture feature correlations and enhance generalisation performance. Additional experiments on model hyperparameters are provided in Appendix G.3.

Conclusion

This study innovatively proposes a topology-driven multi-subspace fusion framework on the Grassmannian, unifying adaptive geometric representation learning with differentiable subspace interaction. By dynamically optimising subspaces to generate task-oriented geometric representation primitives and incorporating a geometry-aware interaction network for subspace topology correlation modelling, we successfully achieve simultaneous improvements in model accuracy and noise robustness in cross-dataset experiments. This work provides a new differentiable manifold learning paradigm for fields such as computer vision and natural language processing.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (62576152, 62336004, 62020106012), the Basic Research Program of Jiangsu (BK20250104), and the Fundamental Research Funds for the Central Universities (JUSRP202504007)

References

Altuwaijri, G. A.; Muhammad, G.; Altaheri, H.; and Alsulaiman, M. 2022. A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for EEG-based motor imagery signals classification. *Diagnostics*, 12(4): 995.

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

Beik-Mohammadi, H.; Hauberg, S.; Arvanitidis, G.; Neumann, G.; and Rozo, L. 2021. Learning Riemannian Manifolds for Geodesic Motion Skills. arXiv:2106.04315.

Bendokat, T.; Zimmermann, R.; and Absil, P.-A. 2024. A Grassmann manifold handbook: Basic geometry and computational aspects. *Advances in Computational Mathematics*, 50(1): 6.

Boumal, N. 2023. *An introduction to optimization on smooth manifolds*. Cambridge University Press.

Brooks, D.; Schwander, O.; Barbaresco, F.; Schneider, J.-Y.; and Cord, M. 2019. Riemannian batch normalization for SPD neural networks. *Advances in Neural Information Processing Systems*, 32.

Chami, I.; Ying, Z.; Ré, C.; and Leskovec, J. 2019. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32.

Cherian, A.; Fernando, B.; Harandi, M.; and Gould, S. 2017. Generalized rank pooling for activity recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3222–3231.

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

Edelman, A.; Arias, T. A.; and Smith, S. T. 1998. The geometry of algorithms with orthogonality constraints. *SIAM journal on Matrix Analysis and Applications*, 20(2): 303–353.

Eschenauer, H. A.; and Olhoff, N. 2001. Topology optimization of continuum structures: a review. *Appl. Mech. Rev.*, 54(4): 331–390.

Fei, Y.; Liu, Y.; Jia, C.; Li, Z.; Wei, X.; and Chen, M. 2025. A survey of geometric optimization for deep learning: from Euclidean space to Riemannian manifold. *ACM Computing Surveys*, 57(5): 1–37.

Gao, Y.; Liu, Y.; She, Q.; and Zhang, J. 2022. 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.

Garcia-Hernando, G.; Yuan, S.; Baek, S.; and Kim, T.-K. 2018. 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*, 409–419.

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

Harandi, M.; Sanderson, C.; Shen, C.; and Lovell, B. C. 2013. Dictionary learning and sparse coding on Grassmann manifolds: An extrinsic solution. In *Proceedings of the IEEE international conference on computer vision*, 3120–3127.

Hinton, G. E.; Dayan, P.; Frey, B. J.; and Neal, R. M. 1995. The" wake-sleep" algorithm for unsupervised neural networks. *Science*, 268(5214): 1158–1161.

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

Huang, Z.; Wang, R.; Shan, S.; and Chen, X. 2015. 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*, 140–149.

Huang, Z.; Wu, J.; and Van Gool, L. 2018. Building deep networks on grassmann manifolds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Ingolfsson, T. M.; Hersche, M.; Wang, X.; Kobayashi, N.; Cavigelli, L.; and Benini, L. 2020. EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2958–2965. IEEE.

Ionescu, C.; Vantzos, O.; and Sminchisescu, C. 2015. Training Deep Networks with Structured Layers by Matrix Backpropagation. arXiv:1509.07838.

Karcher, H. 1977. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5): 509–541.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.

Lawhern, V. J.; Solon, A. J.; Waytowich, N. R.; Gordon, S. M.; Hung, C. P.; and Lance, B. J. 2018. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces. *Journal of neural engineering*, 15(5): 056013.

LeCun, Y.; Bottou, L.; Bengio, Y.; and Haffner, P. 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11): 2278–2324.

Lee, J. M. 2000. *Introduction to topological manifolds*. Springer.

Liu, H.; Sun, J.; Liu, L.; and Zhang, H. 2009. Feature selection with dynamic mutual information. *Pattern Recognition*, 42(7): 1330–1339.

Liu, M.; Wang, R.; Li, S.; Shan, S.; Huang, Z.; and Chen, X. 2014. Combining multiple kernel methods on riemannian manifold for emotion recognition in the wild. In *Proceedings of the 16th International Conference on multimodal interaction*, 494–501.

Maeda, Y.; Kawaguchi, H.; and Tezuka, H. 2025. Estimation of mutual information via quantum kernel methods. *Quantum Machine Intelligence*, 7(1): 1–20.

Mane, R.; Chew, E.; Chua, K.; Ang, K. K.; Robinson, N.; Vinod, A. P.; Lee, S.-W.; and Guan, C. 2021. FBCNet: A Multi-View Convolutional Neural Network for Brain-Computer Interface. arXiv:2104.01233.

Müller, M.; Röder, T.; Clausen, M.; Eberhardt, B.; Krüger, B.; and Weber, A. 2007. Mocap database hdm05. *Institut für Informatik II, Universität Bonn*, 2(7).

Munkres, J. R. 2000. *Topology*. Upper Saddle River, NJ: Prentice Hall, 2nd edition.

Nguyen, X. S.; Brun, L.; Lézoray, O.; and Bougleux, S. 2019. A neural network based on SPD manifold learning for skeleton-based hand gesture recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12036–12045.

Olver, P. J.; and Shakiban, C. 2006. *Applied linear algebra*, volume 1. Springer.

Pan, Y.-T.; Chou, J.-L.; and Wei, C.-S. 2022. MAtt: A manifold attention network for EEG decoding. *Advances in Neural Information Processing Systems*, 35: 31116–31129.

Pennec, X.; Fillard, P.; and Ayache, N. 2006. A Riemannian framework for tensor computing. *International Journal of computer vision*, 66(1): 41–66.

Qiu, L.; Zhang, Y.; and Li, C.-K. 2005. Unitarily invariant metrics on the Grassmann space. *SIAM journal on matrix analysis and applications*, 27(2): 507–531.

Schirrmeister, R. T.; Springenberg, J. T.; Fiederer, L. D. J.; Glasstetter, M.; Eggensperger, K.; Tangermann, M.; Hutter, F.; Burgard, W.; and Ball, T. 2017. Deep learning with convolutional neural networks for EEG decoding and visualization. *Human brain mapping*, 38(11): 5391–5420.

Sen, P.; Namata, G.; Bilgic, M.; Getoor, L.; Galligher, B.; and Eliassi-Rad, T. 2008. Collective classification in network data. *AI magazine*, 29(3): 93–93.

Smith, S. T. 2014. Optimization Techniques on Riemannian Manifolds. arXiv:1407.5965.

Song, C.; Li, H.; Xu, T.; Wu, X.-J.; and Kittler, J. 2025. RefineFuse: an end-to-end network for multi-scale refinement fusion of multi-modality images. *Visual Intelligence*, 3(1): 16.

Souza, L. S.; Sogi, N.; Gatto, B. B.; Kobayashi, T.; and Fukui, K. 2020. An interface between grassmann manifolds and vector spaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 846–847.

Tao, T.; Wang, G.; Lao, Y.; Chen, P.; Liu, J.; Lin, L.; Yu, K.; and Liang, X. 2024. Alignmif: Geometry-aligned multimodal implicit field for lidar-camera joint synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 21230–21240.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

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

Wang, R.; Hu, C.; Chen, Z.; Wu, X.-J.; and Song, X. 2024. A Grassmannian manifold self-attention network for signal classification. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, 5099–5107.

Wang, R.; Wu, X.-J.; and Kittler, J. 2021. 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.

Wei, C.-S.; Koike-Akino, T.; and Wang, Y. 2019. Spatial component-wise convolutional network (SCCNet) for motor-imagery EEG classification. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, 328–331. IEEE.

Xu, T.; Wu, X.-J.; and Kittler, J. 2018. Non-negative subspace representation learning scheme for correlation filter based tracking. In *2018 24th international conference on pattern recognition (ICPR)*, 1888–1893. IEEE.

Xu, T.; Zhu, X.-F.; and Wu, X.-J. 2023. Learning spatio-temporal discriminative model for affine subspace based visual object tracking. *Visual Intelligence*, 1(1): 4.

Yaghooti, B.; Raviv, N.; and Sinopoli, B. 2024. Gram-Schmidt Methods for Unsupervised Feature Selection. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, 7700–7707. IEEE.

Yair, O.; Ben-Chen, M.; and Talmon, R. 2019. Parallel transport on the cone manifold of SPD matrices for domain adaptation. *IEEE Transactions on Signal Processing*, 67(7): 1797–1811.

Yang, L.; Arnaudon, M.; and Barbaresco, F. 2010. Riemannian median, geometry of covariance matrices and radar target detection. In *The 7th European Radar Conference*, 415–418. IEEE.

Zhang, J.; Zhu, G.; Heath Jr, R. W.; and Huang, K. 2018. Grassmannian Learning: Embedding Geometry Awareness in Shallow and Deep Learning. arXiv:1808.02229.