Permutation Sensitivity in t-SVD-based Multi-view Clustering

Jintian Ji[1], He Zhang[1], Yanjun Zhang[1], Leo Yu Zhang[1], Cong Wang[2], Shirui Pan[1]*
[1]Griffith University, [2]City University of Hong Kong
jintian.ji@griffithuni.edu.au
{h.zhang,yanjun.zhang,leo.zhang,s.pan}@griffith.edu.au
congwang@cityu.edu.hk

Abstract

Tensor Singular Value Decomposition (t-SVD) has achieved strong empirical performance in multi-view clustering (MVC) and the incomplete setting (IMVC). However, the success of this framework heavily relies on the Fourier transform along the sample dimension, whose inherent dependence on sample ordering raises serious concerns about sample-permutation sensitivity. To address this issue, we first formalize permutation sensitivity in multi-view clustering by introducing two complementary notions: outcome-oriented permutation invariance and representation-oriented permutation equivariance. Second, by revealing the structural risks associated with the fixed Fourier basis, we propose two plug-and-play strategies: (i) an Order-Restoration strategy (K-OR), designed to recover a locally smooth sample ordering compatible with the Fourier transform; and (ii) a permutation-aware t-SVD framework ($\mathcal{T}$-SVD), instantiated as Cor-SVD, which addresses the structural source of Fourier-transform sensitivity by replacing the fixed basis with a data-dependent transform. Finally, we design a controlled permutation evaluation protocol to assess the robustness of tensorial MVC methods. Extensive experiments show substantial performance variability of existing t-SVD-based methods under sample permutations, while our approaches generally improve permutation robustness and maintain competitive clustering performance. These findings highlight sample-permutation sensitivity as a key robustness issue in tensorial MVC and motivate permutation-aware evaluation and design. Code is publicly available at https://github.com/jijintian/Cor-SVD.

1 Introduction

In real-world scenarios, a comprehensive understanding of an entity often requires observing it from multiple perspectives [45]. For instance, in precision medicine, researchers can group patients into distinct clinical subtypes by analyzing their MRI scans, genetic profiles, and electronic health records simultaneously [16]. Multi-view clustering (MVC) [29, 12, 32] is fundamentally designed for such scenarios, aiming to group instances by discovering consistent and complementary structures across different modalities [4]. However, sensor failures or privacy restrictions frequently plague practical data collection, leaving some views missing for certain instances [36]. This practical bottleneck gives rise to Incomplete Multi-view Clustering (IMVC) [37, 43, 9]. To address both complete and incomplete scenarios, tensor-based frameworks [44, 40] have emerged as a widely used paradigm. Rather than modeling multi-view data individually or in pairs [6, 26, 30, 27, 5], these approaches stack view-specific representations into a third-order tensor and leverage the Tensor Singular Value Decomposition (t-SVD) [17, 39, 3] to mine high-order correlations in the spectral domain. By

*Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

holistically modeling spatial and cross-view structures, t-SVD-based methods have demonstrated strong empirical performance and become an important component of multi-view learning.

Although t-SVD-based methods excel in handling the inherent challenges of MVC and IMVC, such as data noise [38, 28] and missing information [42, 35, 33], their representational stability under sample permutations remains a critical blind spot. Currently, existing evaluations focus on value-level perturbations [46], while overlooking the impact of the arbitrary ordering of samples. In unsupervised clustering [1], samples should be considered as an unordered set rather than an ordered sequence, as the order of the samples does not carry semantic meaning. Therefore, a reliable clustering framework[20] should preserve clustering outcomes under sample re-indexing while transforming sample-wise representations equivariantly.

To investigate sample-permutation robustness, two recent studies [19, 14] have begun to examine the permutation sensitivity of t-SVD-based tensorial clustering methods. Their empirical results show that methods applying FFT along the sample dimension of the tensor are highly sensitive to sample permutations. This sensitivity arises because the Fourier transform implicitly treats the transformed dimension as an ordered signal, making the learned tensor representations dependent on the sequential arrangement of samples. To improve stability, these studies avoid applying FFT along the sample dimension by moving the transform to other tensor dimensions, such as the view dimension [19] or the anchor dimension [14].

Despite these important observations, the permutation-sensitivity problem in t-SVD-based clustering remains insufficiently understood and lacks a principled solution. First, existing studies do not provide a clear formal definition of the desired permutation property for clustering, but mainly report empirical performance changes under different sample orders [13]. Second, their remedies improve stability by bypassing the sample dimension, but they do not explain why the resulting representations should be permutation-consistent or guarantee this property. In other words, the issue is mitigated empirically rather than resolved from a principled equivariance perspective. Third, existing evaluations lack a unified protocol for systematically measuring sample-order sensitivity, making it difficult to evaluate different tensorial frameworks under a common robustness criterion.

To fill these gaps, we provide a systematic study of sample-permutation sensitivity in t-SVD-based tensorial clustering. First, we formulate the sample-permutation sensitivity problem in multi-view clustering and introduce two complementary evaluation criteria: outcome-oriented permutation invariance and representation-oriented permutation equivariance. Second, guided by these criteria, we revisit existing sample-mode t-SVD frameworks and show that sample-mode FFT is not permutation equivariant, creating a structural risk for downstream representations and clustering outcomes. Third, we propose two targeted plug-and-play remedies. K-OR is a heuristic order-restoration strategy that rearranges samples into a more favorable order for FFT operation. Another remedy is an equivariant transform-based t-SVD framework, $\mathcal{T}$-SVD, which replaces the fixed Fourier basis with a permutation-aligned transform. We instantiate this framework as Cor-SVD by constructing the transform from sample correlations. Finally, we introduce a controlled permutation evaluation protocol to systematically measure sample-order sensitivity and assess permutation robustness. The main contributions are summarized as follows:

1. We formulate the sample-permutation invariance problem in multi-view clustering and introduce two complementary criteria: outcome-oriented permutation invariance and representation-oriented permutation equivariance.

2. We identify the structural risk of sample-mode t-SVD and propose two plug-and-play remedies: K-OR, a heuristic strategy that promotes local sample smoothness before FFT-based shrinkage, and $\mathcal{T}$-SVD, an equivariant framework instantiated as Cor-SVD.

3. We establish a controlled permutation evaluation protocol based on varying sample shuffle ratios. Extensive experiments show that existing sample-mode FFT frameworks are highly sensitive to sample permutations, while our methods effectively improve permutation robustness and maintain competitive clustering performance.

2 Preliminaries

In this paper, scalars are denoted by lowercase letters (e.g., x), matrices by boldface uppercase letters (e.g., $\mathbf{X}$), and third-order tensors by calligraphic letters (e.g., $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$). Given

a tensor $\mathcal{X}$, its k-th frontal slice is denoted as $\mathcal{X}^{(k)} = \mathcal{X}(:,:,k)$. The Fast Fourier Transform (FFT) [8] applied along the third dimension (mode-3) is denoted as $\mathcal{X}_f = \text{fft}(\mathcal{X}, 3) = \mathcal{X} \times_3 \mathbf{F}$, where $\mathbf{F} \in \mathbb{C}^{n_3 \times n_3}$ is the Fourier transform matrix, and $\times_3$ denotes the mode-3 tensor-matrix product (e.g., $\mathcal{X}_f^{(k)} = \sum_{j=1}^{n_3} \mathbf{F}_{kj} \mathcal{X}^{(j)}$). Correspondingly, the inverse FFT (iFFT) is denoted as $\mathcal{X} = \text{ifft}(\mathcal{X}_f, 3) = \mathcal{X}_f \times_3 \mathbf{F}^{-1}$. The following are the relevant definitions of the tensor.

Tensor-tensor product (t-product) [15]. Given $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the t-product $\mathcal{X} * \mathcal{Y}$ is a tensor $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$, where $\mathcal{C}_f^{(k)} = \mathcal{X}_f^{(k)} \mathcal{Y}_f^{(k)}, \forall k = 1, \ldots, n_3$.

Definition 2.1 (Tensor Singular Value Decomposition (t-SVD) [17])**.** Given a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, its t-SVD is defined as: $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, where $\mathcal{U} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors (i.e., $\mathcal{U}^\top * \mathcal{U} = \mathcal{I}$, where $\mathcal{I}$ is the identity tensor), and $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor whose frontal slices are diagonal matrices.

Definition 2.2 (Tensor Average Rank [24])**.** Given a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its t-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, its tensor average rank, denoted as $\text{rank}_a(\mathcal{X})$, is defined as, $\text{rank}_a(\mathcal{X}) = 1/n_3 \sum_{k=1}^{n_3} \text{rank}(\mathcal{S}_f^{(k)})$.

Minimizing the tensor average rank directly is NP-hard. Therefore, existing methods typically replace the rank function with a tractable surrogate:

Definition 2.3 (Approximate Tensor Average Rank [7])**.** Given a third-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and its t-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$, its *Approximate Tensor Average Rank*, denoted as $\Psi_t(\mathcal{X})$, is defined as

$$\Psi_t(\mathcal{X}) = \frac{1}{n_3} \sum_{k=1}^{n_3} \sum_{i=1}^{\min(n_1, n_2)} \psi\Big(\big(\mathcal{S}_f^{(k)}\big)_{i,i} \Big), \tag{1}$$

where $\psi(\cdot)$ penalizes singular values. Choosing $\psi(x) = |x|$ recovers the standard TNN [25], while non-convex choices such as the Schatten-p penalty [11] yield tighter average-rank surrogates.

The Generalized t-SVD Framework for MVC and IMVC. Let $\{\mathbf{X}^{(v)}\}_{v=1}^m$ be a multi-view dataset with n samples and m views, where $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d_v}$. For IMVC scenarios, we introduce an observation indicator matrix $\mathbf{W}^{(v)} \in \{0,1\}^{n \times d_v}$, where 1 denotes an observed entry and 0 denotes missing data. For complete MVC, $\mathbf{W}^{(v)}$ simply defaults to an all-ones matrix. The t-SVD-based paradigm of tensorial multi-view clustering [7, 21, 22, 34, 31, 18] aims to learn a view-specific representation matrix $\mathbf{Z}^{(v)} \in \mathbb{R}^{n \times d}$ (often representing a similarity graph or latent embedding) for each view. By stacking and rotating these matrices into a third-order tensor $\mathcal{Z} \in \mathbb{R}^{d \times m \times n}$, where the third dimension (mode-3) strictly corresponds to the sample dimension n, many sample-mode t-SVD methods considered in this work can be formulated as:

$$\min_{\mathcal{Z}, \{\mathbf{Z}^{(v)}\}} \Psi_t(\mathcal{Z}) + \lambda \sum_{v=1}^{m} \mathcal{L}_{data}\big(\mathbf{Z}^{(v)}; \mathbf{X}^{(v)}, \mathbf{W}^{(v)}\big) \quad \text{s.t. } \mathcal{Z} = \Phi(\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(m)}), \tag{2}$$

where $\Psi_t(\mathcal{Z})$ represents the continuous surrogate for the tensor average rank, and $\mathcal{L}_{data}(\cdot)$ indicates the view-specific fidelity or reconstruction loss. $\Phi(\cdot)$ denotes the tensorization mapping [40], which involves folding and lateral alignment operations to construct a 3D tensor from 2D matrices.

To solve the constrained optimization problem in Eq. (2), researchers usually adopt the Alternating Direction Method of Multipliers (ADMM). For the ADMM-based formulations considered here, the update of the low-rank tensor variable typically reduces to the following proximal subproblem:

$$\min_{\mathcal{Z}} \Psi_t(\mathcal{Z}) + \frac{\mu}{2} \|\mathcal{Z} - \mathcal{A}\|_F^2, \tag{3}$$

where $\mathcal{A}$ is the auxiliary tensor aggregated from the ADMM updates, $\mu > 0$ is the penalty scalar, and $\|\cdot\|_F$ denotes the Frobenius norm. The solution to Eq. (3) is obtained via the Generalized Singular Value Thresholding (GSVT) operator [23, 2] applied in the Fourier domain. The standard computational procedure is summarized in Algorithm 1.

Remark. This paper focuses on t-SVD-based tensorial MVC and IMVC frameworks that apply the transform along the sample mode, a widely adopted design with strong benchmark performance. Since sample order is arbitrary and semantically meaningless in clustering, it remains unclear whether the unusually strong results under fixed dataset orders reflect intrinsic data geometry or partly rely on

favorable sample arrangements. This motivates our study of sample-mode t-SVD frameworks and the development of a controlled permutation evaluation protocol for assessing order-induced instability.

Algorithm 1 t-SVD-based Low-rank Tensor Learning

Input: Auxiliary tensor $\mathcal{A} \in \mathbb{R}^{d\times m\times n}$, ADMM penalty μ, rank surrogate function $\psi(\cdot)$.
Output: Recovered low-rank tensor $\mathcal{Z}^*$.
$\mathcal{A}_f \leftarrow \text{fft}(\mathcal{A}, 3)$;
for $k = 1$ **to** n **do**
  $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] \leftarrow \text{SVD}\big(\mathcal{A}_f^{(k)}\big)$;
  $\mathbf{\Sigma}_{\text{new}} \leftarrow \text{Prox}_{\psi,1/\mu}(\mathbf{\Sigma})$, where $\text{Prox}(\cdot)$ denotes the proximal operator [23];
  $\mathcal{Z}_f^{(k)} \leftarrow \mathbf{U}\mathbf{\Sigma}_{\text{new}}\mathbf{V}^H$;
end for
$\mathcal{Z}^* \leftarrow \text{ifft}(\mathcal{Z}_f, 3)$.

3 Problem Definition

Given a multi-view dataset $\mathbb{X} = \{\mathbf{X}^{(v)}\}_{v=1}^m$ with n samples and m views, where $\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}$ denotes the feature matrix of the v-th view, multi-view clustering aims to learn a clustering pipeline $\mathcal{F}(\cdot)$ that assigns each sample to one of K clusters. The pipeline is commonly decomposed as $\mathcal{F} = \mathcal{C} \circ f$, where $f(\cdot)$ learns a unified representation and $\mathcal{C}(\cdot)$ denotes a downstream clustering operator, such as k-means or spectral clustering. Let Π_n denote the set of $n \times n$ permutation matrices. For any $\mathbf{P}_n \in \Pi_n$, the permuted multi-view dataset is defined as $\mathbf{P}_n\mathbb{X} := \{\mathbf{P}_n\mathbf{X}^{(v)}\}_{v=1}^m$.

Sample-Permutation Sensitivity Problem.

Since multi-view clustering operates on unordered samples, a reliable clustering framework should not depend on the arbitrary order in which samples are stored or fed into the model. We refer to the phenomenon that sample re-indexing changes the learned representations or final clustering results as *sample-permutation sensitivity*. To characterize this problem, we consider two levels of robustness: outcome-oriented permutation invariance, which requires the final clustering performance to remain unchanged, and representation-oriented permutation equivariance, which further requires the learned representations to transform consistently with the input permutation.

Definition 3.1 (Outcome-oriented Permutation Invariance)**.** Given a multi-view clustering pipeline $\mathcal{F}(\mathbb{X}) = \hat{\mathbf{y}}$, we say that $\mathcal{F}$ satisfies outcome-oriented permutation invariance if its clustering performance remains unchanged under any sample permutation $\mathbf{P}_n \in \Pi_n$. Formally, for a permutation-invariant evaluation metric $\mathcal{M}$ and ground-truth labels $\mathbf{y}$,

$$\mathcal{M}\big(\mathcal{F}(\mathbf{P}_n\mathbb{X}), \mathbf{P}_n\mathbf{y}\big) = \mathcal{M}\big(\mathcal{F}(\mathbb{X}), \mathbf{y}\big). \tag{4}$$

Definition 3.2 (Representation-oriented Permutation Equivariance)**.** Given a multi-view clustering pipeline $\mathcal{F} = \mathcal{C} \circ f$, where $f(\cdot)$ denotes the representation learning module, we say that $f(\cdot)$ satisfies representation-oriented permutation equivariance if the learned representations are re-indexed consistently with the input samples under any sample permutation $\mathbf{P}_n \in \Pi_n$. Formally,

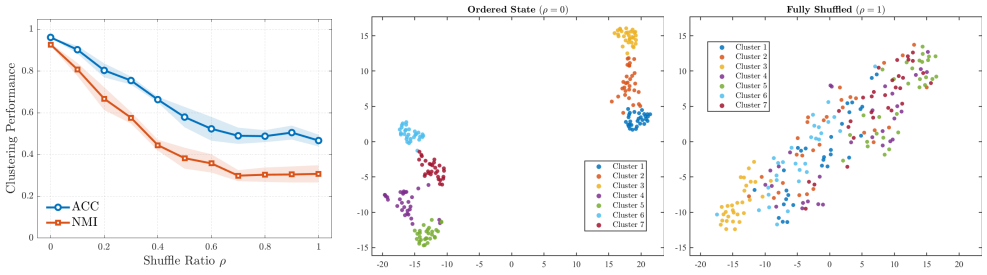
$$f(\mathbf{P}_n\mathbb{X}) = \mathbf{P}_n f(\mathbb{X}). \tag{5}$$

Remark. Representation-oriented permutation equivariance provides a stronger structural guarantee than outcome-oriented permutation invariance. A reliable clustering framework should learn order-stable representations, meaning that each sample representation is determined by its intrinsic features and semantic relations rather than by its arbitrary input index. In addition, under a permutation-equivariant downstream clustering operator $\mathcal{C}(\cdot)$, representation-oriented permutation equivariance is sufficient to ensure outcome-oriented permutation invariance. See Appendix B.1 for the proof.

4 Methodology

4.1 Structural Risk of Sample-mode t-SVD

From a representation-equivariance perspective, we now examine the t-SVD-based low-rank recovery module in Algorithm 1. Since this module produces the representation tensor used by downstream



(a) Performance variation (b) Representation variation ($\Delta_F = 9.19$)

Figure 1: (a) Clustering performance of TLSpNM under varying permutation ratios ρ on the MSRCv1 dataset. (b) t-SNE of unified representations. $\Delta_F = \|\mathbf{Z}_{\rho=0} - \mathbf{P}_\rho^\top \mathbf{Z}_{\rho=1}\|_F / \|\mathbf{Z}_{\rho=0}\|_F$ denotes the normalized distance between permutation-aligned representations.

clustering, it should also preserve sample-wise equivariance. However, we show that the standard sample-mode FFT breaks this property.

Proposition 4.1 (Non-Equivariance of Sample-mode FFT)**.** *Let $\mathcal{A} \in \mathbb{R}^{d\times m\times n}$ be a tensor whose third mode indexes samples. In general, the sample-mode FFT is not equivariant to arbitrary sample permutations. That is, there exist $\mathcal{A}$ and $\mathbf{P}_n \in \Pi_n$ such that*

$$\text{fft}(\mathcal{A} \times_3 \mathbf{P}_n, 3) \neq \text{fft}(\mathcal{A}, 3) \times_3 \mathbf{P}_n. \tag{6}$$

Proof. The result follows from the non-commutativity between the fixed DFT matrix and arbitrary sample permutation matrices, i.e., $\mathbf{F}\mathbf{P}_n \neq \mathbf{P}_n\mathbf{F}$ in general. Therefore, sample permutation and sample-mode FFT do not commute in general. See Appendix B.2 for details. □

Proposition 4.1 shows that sample-mode FFT does not preserve the re-indexing relation required by sample-wise representation equivariance. After a sample permutation, the Fourier-domain tensor is not merely a permuted version of the original Fourier-domain tensor; its spectral slices may be structurally changed. This discrepancy is further propagated by the slice-wise singular value thresholding step in Algorithm 1. Since SVT is applied independently to each Fourier-domain frontal slice $\mathcal{A}_f^{(k)}$, changes in the Fourier-domain decomposition can induce different shrinkage patterns. Consequently, the recovered tensor $\mathcal{Z}^*$ is not guaranteed to be a sample-wise permutation of the original output. Thus, sample-mode t-SVD lacks a structural guarantee for representation equivariance, and consequently lacks a reliable guarantee for satisfying the sample-wise permutation invariance requirement.

Empirical Verification of the Structural Risk. While Proposition 4.1 shows that sample-mode FFT lacks a structural guarantee for representation equivariance, this result alone does not establish that the full pipeline violates either permutation invariance or representation equivariance. We conduct two empirical tests on the TLSpNM baseline to examine whether this structural risk appears in practice.

First, we test outcome-oriented permutation invariance through a controlled permutation protocol. Starting from a class-contiguous ordering, we randomly permute a fraction $\rho \in [0, 1]$ of samples and record the performance change of TLSpNM with sample-mode FFT. As shown in Figure 1 (a) (more in Appendix A), increasing ρ leads to degraded performance and amplified variance, suggesting that this baseline does not reliably preserve permutation-invariant clustering behavior.

Second, we examine representation-oriented consistency by comparing the unified representations learned under a class-contiguous order and a fully random order after aligning sample indices. As shown in Figure 1(b), the representation geometry changes substantially after sample shuffling, as reflected by both the t-SNE visualization and the normalized Frobenius discrepancy. This indicates that the learned representations are not merely permuted versions of each other, but can be structurally altered by sample re-indexing. These observations provide empirical evidence that sample-mode FFT may induce order-dependent representations in practical tensorial clustering pipelines.

4.2 Mitigating the Structural Risk: Two Plug-and-Play Remedies

Based on the above analysis, we propose two strategies to mitigate the structural risk in the sample-mode t-SVD framework.

Order Restoration Strategy

Since a locally smooth sample ordering can yield a more concentrated Fourier representation, a natural remedy is to arrange similar samples close to each other before FFT-based shrinkage. Given an auxiliary tensor $\mathcal{A} \in \mathbb{R}^{d\times m\times n}$, we construct a data-dependent reorder matrix $\mathbf{P}_{\mathcal{A}} \in \Pi_n$ and apply the FFT-based shrinkage to the reordered tensor. The final result is restored to the original indexing:

$$\mathcal{Z}^* = \mathrm{Shrink}_{\mathrm{FFT}}\left(\mathcal{A} \times_3 \mathbf{P}_{\mathcal{A}}\right) \times_3 \mathbf{P}_{\mathcal{A}}^{\top}. \tag{7}$$

This strategy does not remove the structural limitation of FFT, but practically reduces its mismatch with arbitrary sample orders.

Instantiation: k-means Order Restoration (K-OR). We instantiate $\mathbf{P}_{\mathcal{A}}$ using a simple k-means-based ordering strategy. Specifically, we unfold $\mathcal{A}$ along the sample mode as $\mathbf{A}_{(3)} \in \mathbb{R}^{n\times(dm)}$, apply k-means to group samples by representation-level similarity, and then arrange samples according to their cluster assignments. This places similar samples locally along the sample mode, mitigating the spectral leakage induced by order disorder. The resulting K-OR procedure is plug-and-play and can be inserted into existing FFT-based t-SVD solvers. However, it remains heuristic and does not provide a formal invariance guarantee.

Algorithm 2 FFT-based Low-rank Tensor Learning with Order Restoration (K-OR)

Input: Auxiliary tensor $\mathcal{A} \in \mathbb{R}^{d\times m\times n}$, penalty μ, surrogate function $\psi(\cdot)$.
Output: Recovered low-rank tensor $\mathcal{Z}^*$.
Unfold $\mathcal{A}$ along the sample mode to obtain $\mathbf{A}_{(3)} \in \mathbb{R}^{n\times(dm)}$.
Obtain a data-dependent permutation matrix $\mathbf{P}_{\mathcal{A}}$ via k-means-based ordering on $\mathbf{A}_{(3)}$.
$\mathcal{A}_{\mathrm{ord}} \leftarrow \mathcal{A} \times_3 \mathbf{P}_{\mathcal{A}}$.
$\widehat{\mathcal{A}}_{\mathrm{ord}} \leftarrow \mathrm{fft}(\mathcal{A}_{\mathrm{ord}}, 3)$.
for $k = 1$ **to** n **do**
  $[\mathbf{U}, \mathbf{\Sigma}, \mathbf{V}] \leftarrow \mathrm{SVD}(\widehat{\mathcal{A}}^{(k)}_{\mathrm{ord}})$.
  $\widehat{\mathcal{Z}}^{(k)}_{\mathrm{ord}} \leftarrow \mathbf{U}\,\mathrm{Prox}_{\psi,1/\mu}(\mathbf{\Sigma})\mathbf{V}^H$.
end for
$\mathcal{Z}_{\mathrm{ord}} \leftarrow \mathrm{ifft}(\widehat{\mathcal{Z}}_{\mathrm{ord}}, 3)$.
$\mathcal{Z}^* \leftarrow \mathcal{Z}_{\mathrm{ord}} \times_3 \mathbf{P}_{\mathcal{A}}^{\top}$.

Equivariant Transform-based t-SVD ($\mathcal{T}$-SVD)

To address the structural source of order sensitivity in sample-mode t-SVD, we replace the fixed Fourier basis with a transform that is aligned with sample permutations. Specifically, we propose a generalized equivariant transform-based t-SVD ($\mathcal{T}$-SVD) framework based on a data-dependent transform basis that remains aligned under sample re-indexing.

Let $\mathbb{X} = \{\mathbf{X}^{(v)}\}_{v=1}^{m}$ denote a multi-view dataset with $\mathbf{X}^{(v)} \in \mathbb{R}^{n\times d_v}$. We construct a tensor representation $\mathcal{A} \in \mathbb{R}^{d\times m\times n}$ from $\mathbb{X}$, where the third mode corresponds to the sample dimension. We define a data-dependent orthogonal transform basis $\mathbf{T}(\mathbb{X}) \in \mathbb{R}^{n\times n}$ and the corresponding transform: $\mathcal{T}(\mathcal{A}, 3) = \mathcal{A} \times_3 \mathbf{T}(\mathbb{X})^{\top}$. To establish permutation consistency, we assume that the transform basis satisfies the following condition:

Assumption 4.2 (Permutation-Aligned Transform Basis)**.** The transform basis $\mathbf{T}(\mathbb{X})$ is permutation-aligned if, for any permutation matrix $\mathbf{P}_n \in \Pi_n$, $\mathbf{T}(\mathbf{P}_n\mathbb{X}) = \mathbf{P}_n\mathbf{T}(\mathbb{X})$, where $\mathbf{P}_n\mathbb{X} := \{\mathbf{P}_n\mathbf{X}^{(v)}\}_{v=1}^{m}$.

Under this condition, the transform remains consistently aligned with the permuted input, which leads to the following permutation equivariance of $\mathcal{T}$-SVD. The proof is provided in Appendix B.3.

Proposition 4.3 (Permutation Equivariance of $\mathcal{T}$-SVD)**.** *Let $f_{\mathcal{T}}(\cdot)$ denote the proximal mapping induced by $\mathcal{T}$-SVD, defined as $f_{\mathcal{T}}(\mathcal{A}) = \mathcal{G}_\tau\big(\mathcal{A} \times_3 \mathbf{T}(\mathbb{X})^{\top}\big) \times_3 \mathbf{T}(\mathbb{X})$, where $\mathcal{G}_\tau(\cdot)$ denotes the slice-wise singular value shrinkage operator. If the transform basis satisfies Assumption 4.2, then*

$$f_{\mathcal{T}}(\mathcal{A} \times_3 \mathbf{P}_n) = f_{\mathcal{T}}(\mathcal{A}) \times_3 \mathbf{P}_n. \tag{8}$$

Instantiation via Sample Correlation Transform (Cor-SVD). To instantiate the $\mathcal{T}$-SVD framework, we construct the transform basis from the sample-wise correlation structure of the data and denote this basis by $\mathbf{T}_c(\mathbb{X})$. This design anchors the transform to the intrinsic geometry of the dataset rather than an arbitrary ordering. Let $\mathbf{X}^{(v)} \in \mathbb{R}^{n \times d_v}$ denote the normalized feature matrix of the v-th view. We define the sample correlation matrix $\mathbf{S} \in \mathbb{R}^{n \times n}$ as:

$$\mathbf{S} = \begin{cases} \dfrac{1}{m} \displaystyle\sum_{v=1}^{m} \mathbf{X}^{(v)} \big(\mathbf{X}^{(v)}\big)^\top, & \text{for complete MVC,} \\ \left(\displaystyle\sum_{v=1}^{m} (\mathbf{X}^{(v)} \odot \mathbf{W}^{(v)})(\mathbf{X}^{(v)} \odot \mathbf{W}^{(v)})^\top \right) \oslash \mathbf{O}, & \text{for IMVC,} \end{cases} \tag{9}$$

where $\odot$ and $\oslash$ denote the element-wise Hadamard product and division, respectively. $\mathbf{O} = \sum_{v=1}^{m} \mathbf{o}^{(v)} (\mathbf{o}^{(v)})^\top$ is the co-occurrence matrix, and $\mathbf{o}^{(v)} \in \{0,1\}^n$ indicates observed samples. $\mathbf{W}^{(v)} \in \{0,1\}^{n \times d_v}$ is an observation indicator matrix where 1 denotes an observed entry and 0 denotes missing data. In practice, the division by $\mathbf{O}$ is performed only on entries with $\mathbf{O}_{ij} > 0$. Then, the basis $\mathbf{T}_c(\mathbb{X})$ is obtained by eigen-decomposition: $\mathbf{S} = \mathbf{T}_c(\mathbb{X}) \mathbf{\Lambda} \mathbf{T}_c(\mathbb{X})^\top$.

Theorem 4.4 (Conditional Permutation Equivariance of Cor-SVD)**.** *For any permutation matrix $\mathbf{P}_n \in \Pi_n$, the transform bases constructed from $\mathbb{X}$ and $\mathbf{P}_n \mathbb{X}$ satisfy $\mathbf{T}_c(\mathbf{P}_n \mathbb{X}) = \mathbf{P}_n \mathbf{T}_c(\mathbb{X}) \mathbf{Q}_P$, where $\mathbf{Q}_P$ is block-orthogonal within eigenspaces associated with the same eigenvalue. Let $\bar{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{T}_c(\mathbb{X})^\top$. For the ℓ-th eigenspace, let $\mathcal{I}_\ell$ denote the indices of the transformed slices associated with this eigenspace, and define $\mathcal{R} = \{\ell : |\mathcal{I}_\ell| > 1\}$. Then the following hold:*

1. Exact equivariance. *If the sample correlation matrix has a simple spectrum, then*

$$f_{\mathrm{Cor}}(\mathcal{A} \times_3 \mathbf{P}_n) = f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n. \tag{10}$$

2. Bounded equivariance error. *If the sample correlation matrix has repeated eigenvalues, then*

$$\| f_{\mathrm{Cor}}(\mathcal{A} \times_3 \mathbf{P}_n) - f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n \|_F \leq 2 \left(\sum_{\ell \in \mathcal{R}} \big\| \bar{\mathcal{A}}_{\mathcal{I}_\ell} \big\|_F^2 \right)^{1/2}. \tag{11}$$

The proof is provided in Appendix B.4. This theorem clarifies how the spectral structure of the correlation matrix governs the permutation robustness of Cor-SVD and characterizes the source of any remaining equivariance error.

5 Experiments

5.1 Experimental Setup

Datasets. We conduct evaluations on four benchmark multi-view datasets: MSRCv1, NGs, BBCSport, and HW. For the incomplete setting, we follow the existing study [10] by randomly masking a portion of samples (missing rate = 10%, 50%) and ensuring each sample is observed in at least one view. Detailed statistical summaries are provided in Appendix C.1.

Baselines. To evaluate permutation sensitivity in both complete and incomplete multi-view clustering (MVC/IMVC), we adopt four state-of-the-art tensor-based methods: TLSpNM (TPAMI 2023) [11] and ESTMC (TPAMI 2025) [13] for MVC, and SEIC (TKDE 2025) [42] and ASCR (AAAI 2025) [41] for IMVC. For each baseline, we consider four variants: **FFT (sample mode)**, the original setting that applies FFT along the sample mode; **FFT (view mode)**, a comparison setting following [19] that applies FFT along the view mode; and two proposed variants, **K-OR** and **Cor-SVD**. Detailed information is provided in Appendix C.1.

Evaluation Metrics. The clustering performance is measured using five standard metrics: Accuracy (ACC), Normalized Mutual Information (NMI), Purity, F-score, and Adjusted Rand Index (ARI). Higher values indicate better clustering performance for all five metrics.

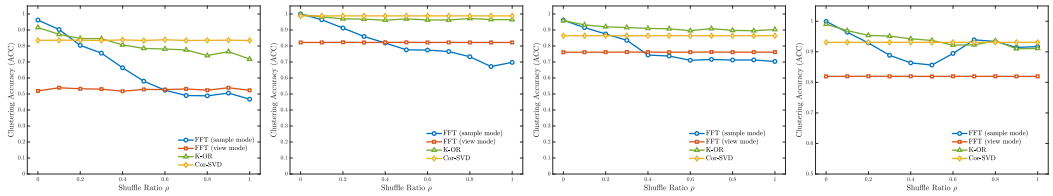
Controlled Permutation Evaluation. Beyond conventional evaluation under a fixed sample order, we introduce a *controlled permutation protocol* to assess robustness against order-induced instability. Specifically, we define a shuffle ratio $\rho \in [0,1]$ to interpolate between structured and random sample

Table 1: Clustering performance (mean (std)%) on the complete MSRCv1 and NGs datasets.

Baselines	Variants	MSRCv1					NGs				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
TLSpNM	FFT (sample mode)	64.94 (17.54)	48.94 (21.80)	65.45 (17.04)	50.16 (21.20)	42.02 (24.68)	81.56 (10.43)	59.75 (19.49)	81.56 (10.43)	68.82 (16.33)	61.31 (20.24)
	FFT (view mode)	52.83 (1.80)	38.49 (1.84)	53.13 (1.92)	37.44 (1.54)	27.18 (1.76)	82.23 (0.07)	61.17 (0.23)	82.23 (0.08)	68.47 (0.12)	60.60 (0.15)
	K-OR	80.46 (6.52)	69.22 (7.92)	80.53 (6.42)	67.93 (8.35)	62.66 (9.80)	97.02 (1.19)	90.99 (3.38)	97.02 (1.19)	94.09 (2.34)	92.63 (2.92)
	Cor-SVD	**84.29 (0)**	**73.38 (0.03)**	**84.29 (0)**	**72.76 (0.14)**	**68.33 (0.10)**	**98.20 (0)**	**94.26 (0)**	**98.20 (0)**	**96.43 (0)**	**95.55 (0)**
ESTMC	FFT (sample mode)	82.46 (5.66)	73.40 (7.27)	82.46 (5.66)	71.84 (8.08)	67.26 (9.41)	91.62 (2.88)	78.63 (6.88)	91.62 (2.88)	84.16 (5.34)	80.09 (6.67)
	FFT (view mode)	79.32 (1.27)	70.39 (1.49)	79.32 (1.27)	67.92 (1.36)	62.70 (1.57)	86.60 **(0.02)**	68.61 (0.13)	86.60 **(0.02)**	75.34 (0.09)	69.15 (0.30)
	K-OR	**85.58** (1.93)	**75.67** (2.40)	**85.58** (1.93)	**74.35** (2.86)	**70.10** (3.36)	92.93 (1.45)	81.12 (3.51)	92.93 (1.45)	86.44 (2.67)	83.08 (3.34)
	Cor-SVD	83.75 **(0.89)**	72.66 **(0.91)**	83.75 **(0.89)**	72.31 **(1.06)**	67.76 **(1.25)**	**95.80** (0.03)	**88.17 (0.07)**	**95.80** (0.03)	**91.80 (0.05)**	**89.78 (0.06)**

Table 2: Clustering performance (mean (std)%) on the incomplete MSRCv1 dataset under different missing rates (10% and 50%).

Baselines	Variants	MSRCv1 (Missing rate: 10%)					MSRCv1 (Missing rate: 50%)				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
SEIC	FFT (sample mode)	70.40 (8.28)	63.10 (12.22)	72.62 (8.22)	60.76 (11.40)	53.96 (13.34)	55.87 (13.03)	46.87 (17.82)	58.08 (13.31)	44.80 (15.91)	34.12 (18.28)
	FFT (view mode)	71.49 (3.48)	66.06 (2.42)	73.73 (2.92)	62.65 (2.93)	56.97 (3.55)	48.16 (3.25)	40.20 **(3.15)**	50.11 (3.20)	35.31 (2.56)	21.21 (3.48)
	K-OR	**74.79** (2.05)	**69.01** (2.47)	**76.74** (1.87)	**65.87** (2.28)	**59.89** (2.71)	**59.09** (5.95)	**50.99** (7.30)	**61.25** (6.00)	**45.39** (7.26)	**34.70** (9.05)
	Cor-SVD	74.12 **(1.93)**	67.66 **(1.93)**	75.75 **(1.68)**	64.27 **(1.97)**	58.08 **(2.34)**	52.32 **(2.53)**	43.11 (3.23)	54.31 **(2.50)**	39.13 **(2.54)**	27.21 **(3.13)**
ASCR	FFT (sample mode)	**73.59** (8.11)	**66.67** (10.46)	**75.68** (7.75)	**63.81** (10.80)	**57.60** (12.66)	58.42 (16.57)	46.61 (20.89)	60.29 (16.46)	**46.16** (19.84)	**36.78** (23.35)
	FFT (view mode)	69.21 **(2.82)**	62.24 **(2.36)**	70.83 (2.49)	58.61 **(2.58)**	51.50 **(3.02)**	48.37 **(2.61)**	39.42 **(2.48)**	50.51 **(2.34)**	36.88 (2.36)	26.02 (2.79)
	K-OR	72.05 (3.34)	65.94 (3.65)	74.48 (2.97)	61.67 (3.52)	54.93 (4.18)	**59.79** (4.59)	**50.46** (4.63)	**61.92** (4.23)	45.37 (4.20)	35.29 (5.13)
	Cor-SVD	72.21 (2.86)	64.52 (2.55)	74.23 **(2.34)**	61.15 (2.68)	54.47 (3.17)	31.51 (2.79)	18.33 (3.39)	32.70 (2.74)	23.37 **(2.28)**	9.47 **(2.77)**



(a) MSRCv1 (b) NGs (c) BBCSport (d) HW

Figure 2: Performance trajectories of four TLSpNM variants on four datasets.

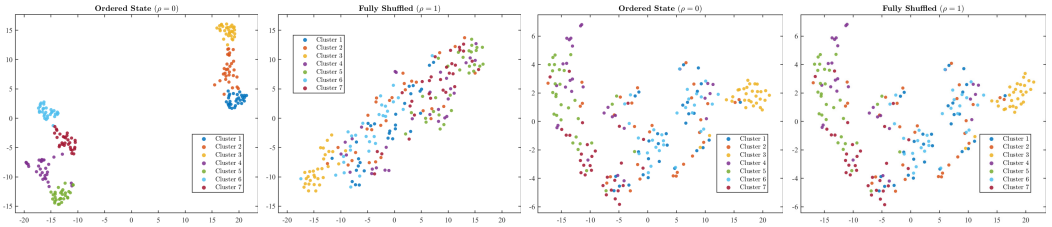
orderings. We vary ρ from 0 to 1 with step size 0.1, i.e., $\rho \in \{0, 0.1, \ldots, 1\}$. For each ρ, we randomly select $\lfloor \rho n \rfloor$ samples and permute their positions, while the remaining samples keep their original order. Thus, $\rho = 0$ corresponds to the original class-contiguous dataset ordering, which may exhibit strong local smoothness along the sample mode, whereas $\rho = 1$ corresponds to a uniformly random permutation of all samples. For each ρ, the entire pipeline is repeated $T = 10$ times to account for randomness from sample shuffling. We report the mean and standard deviation over all evaluated shuffle ratios and repeated trials. The mean summarizes overall clustering performance under different permutation levels, while the standard deviation reflects empirical variability across these settings. For IMVC, this variability may additionally include randomness from incomplete-data generation.

5.2 Evaluating Outcome-oriented Permutation Invariance

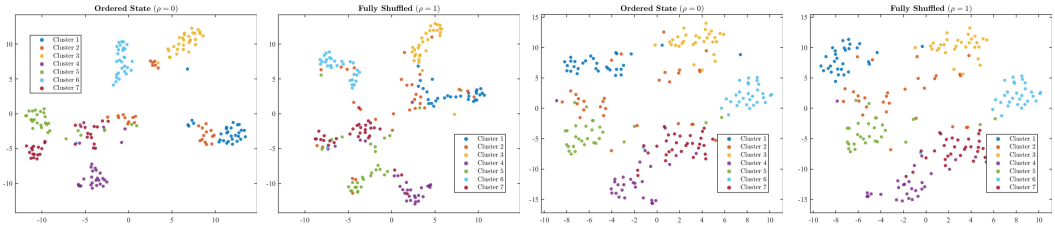
Table 1 and Table 2 (more in Appendix C.2) summarize clustering performance over all shuffle ratios $\rho \in \{0, 0.1, \ldots, 1\}$, whereas the trajectory plots report the performance at each ρ. The mean and standard deviation in the tables are computed over the results across all shuffle ratios and repeated trials, reflecting overall clustering quality and empirical variability, respectively.

FFT (sample mode) is sensitive to sample permutations. Across both complete and incomplete settings, FFT (sample mode) exhibits large performance variance. For example, in Table 1, TLSpNM with FFT (sample mode) achieves an ACC of 64.94% with a standard deviation of 17.54% on MSRCv1, and its ACC standard deviation reaches 10.43% on NGs. Under a 50% missing rate, ASCR with FFT (sample mode) also shows a large ACC standard deviation of 16.57%.

FFT (view mode) improves sample-order stability but can introduce view-order sensitivity. With TLSpNM, it reduces the ACC standard deviation on NGs from 10.43% to 0.07%, yet achieves only 52.83% ACC on MSRCv1, below K-OR (80.46%) and Cor-SVD (84.29%). It also shows view-order sensitivity on MSRCv1 and HW (Appendix D: Table 10), indicating that moving FFT can transfer rather than eliminate order dependence.



(a) FFT (sample mode) ($\Delta_F = 9.19$) (b) FFT (view mode) ($\Delta_F = 1.75 \times 10^{-11}$)



(c) K-OR ($\Delta_F = 5.53$) (d) Cor-SVD ($\Delta_F = 1.1 \times 10^{-11}$)

Figure 3: t-SNE visualization for different strategies with the TLSpNM baseline on MSRCv1. Δ_F denotes the normalized distance between permutation-aligned representations.

K-OR and Cor-SVD achieve better accuracy–stability trade-offs. K-OR improves performance by restoring a more favorable sample order before FFT-based shrinkage, while Cor-SVD replaces the fixed Fourier basis with a permutation-aligned transform. Overall, both methods reduce order sensitivity while maintaining competitive clustering accuracy, with Cor-SVD showing particularly low variance in complete MVC. In IMVC, however, its performance can be limited by less reliable correlation estimation under incomplete observations.

Performance trajectories highlight the robustness–accuracy trade-off. As shown in Figure 2, the clustering performance of FFT (sample mode) degrades progressively as ρ increases, while FFT (view mode) remains stable but less accurate. K-OR alleviates the degradation, whereas Cor-SVD achieves a more favorable balance between stability and clustering performance across shuffle ratios.

5.3 Evaluating Representation-oriented Permutation Equivariance

We further evaluate representation-oriented permutation robustness on complete datasets using t-SNE and the normalized discrepancy Δ_F between representations learned before and after full sample shuffling. We focus on complete datasets because varying missing patterns in IMVC introduce additional representation discrepancies that cannot be attributed solely to sample permutations. Results are shown in Figure 3 and Appendix C.3, leading to the following observations:

FFT (sample mode) is strongly order-dependent. As shown in Figure 3(a), FFT (sample mode) produces substantially different representation geometries after permutation, with the originally separable structure becoming more entangled and $\Delta_F = 9.19$.

FFT (view mode) is representation-stable. As shown in Figure 3(b), it yields an almost zero discrepancy ($\Delta_F = 1.75 \times 10^{-11}$) under sample re-indexing. This stability is expected because the Fourier transform is applied along the view dimension rather than the permuted sample dimension.

K-OR mitigates but does not eliminate distortion. As shown in Figure 3(c), K-OR reduces representation distortion, but its discrepancy remains non-negligible ($\Delta_F = 5.53$), indicating that it is a heuristic mitigation rather than a strict equivariance-guaranteed solution.

Cor-SVD exhibits near-zero empirical equivariance error. As shown in Figure 3(d), Cor-SVD yields a near-zero empirical discrepancy on MSRCv1. This observation is consistent with the theoretical analysis, which guarantees exact equivariance under a simple spectrum and bounds the error contributed by repeated eigenspaces otherwise.

6 Ablation Study

Table 3: Ablation and control study for transform design with TLSpNM.

Variants	MSRCv1		NGs		BBCSport		HW	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Original FFT	64.94 (17.54)	48.94 (21.80)	81.56 (10.43)	59.75 (19.49)	78.38 (9.77)	56.87 (14.59)	91.82 (4.08)	82.67 (8.70)
Spectral clustering	66.28 (0.38)	60.43 (0.53)	97.60(0)	92.86(0)	79.07(0.06)	62.15(0.23)	76.88(0.69)	70.13(0.21)
Random orthogonal transform	79.26 (4.29)	66.72 (4.52)	49.13(3.99)	22.70(4.43)	55.82(3.57)	29.98(2.54)	82.84(0.91)	73.12(0.96)
DCT basis	57.69 (15.32)	41.69 (17.01)	66.13(12.71)	44.03(18.80)	70.84(14.04)	50.81(17.45)	87.86(5.34)	76.94(9.34)
Ground-truth basis (Oracle)	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**	**100.00 (0.00)**
Cor-SVD	84.29 (0)	73.38 (0.03)	98.20 (0)	94.26 (0)	86.40 (0)	75.33 (0)	93.26 (0.01)	87.37 (0.04)

Table 4: Ablation study of different sample-ordering strategies with TLSpNM.

Variants	MSRCv1		NGs		BBCSport		HW	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Original FFT	64.94 (17.54)	48.94 (21.80)	81.56 (10.43)	59.75 (19.49)	78.38 (9.77)	56.87 (14.59)	91.82 (4.08)	82.67 (8.70)
Random ordering	63.61 (1.96)	47.49 (2.46)	74.34(1.24)	47.71(1.53)	76.27(0.99)	61.77(1.17)	91.55(0.57)	84.94(0.59)
Hash-key ordering	63.96 (1.70)	46.92 (1.52)	77.04(0.24)	54.51(0.27)	61.51(0.22)	50.56(2.04)	91.44(0.09)	85.00(0.18)
K-OR	80.46 (6.52)	69.22 (7.92)	97.02 (1.19)	90.99 (3.38)	91.29 (2.16)	77.31 (4.59)	94.05 (2.64)	88.06 (4.20)
Ground-truth ordering (Oracle)	**97.14 (0.00)**	**93.89 (0.00)**	**99.80 (0.00)**	**99.30 (0.00)**	**96.32 (0.00)**	**89.68 (0.00)**	**99.80 (0.00)**	**99.30 (0.00)**

Table 5: Effect of transform-basis quality under a 50% missing rate with SEIC.

Variants	MSRCv1		NGs		BBCSport		HW	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
Cor-SVD	52.32 (2.53)	43.11 (3.23)	33.74 (2.57)	15.93 (3.11)	37.11 **(1.79)**	15.37 (6.02)	66.18 **(1.95)**	61.30 **(1.17)**
+KNN completion	66.30 (3.00)	57.65 (3.12)	32.32 **(2.32)**	13.04 **(2.88)**	36.96 (1.93)	15.96 (6.27)	78.43 (2.02)	73.54 (2.09)
+Source completion	73.53 **(2.43)**	68.17 **(2.05)**	45.44 (5.52)	**35.08** (7.24)	55.18 (6.97)	38.68 (8.95)	79.21 (2.50)	77.66 (1.32)
+Ground-truth basis (Oracle)	**78.48** (4.08)	**89.93** (2.34)	**46.03** (4.36)	34.89 (6.53)	**82.36** (4.34)	**82.04 (5.59)**	**79.55** (2.94)	**92.28** (1.28)

Transform-basis ablation. To separate the effect of replacing FFT from that of the data-dependent basis, we compare Cor-SVD with spectral clustering (on the sample correlation matrix $\mathbf{S}$), random orthogonal and DCT transforms, and an oracle ground-truth basis in Table 3. These results indicate that simply replacing FFT with another fixed transform is insufficient, while exploiting data-dependent sample structure yields substantially stronger performance in the evaluated settings.

Sample-ordering ablation. To verify that K-OR benefits from structure-aware ordering rather than arbitrary reordering, we compare it with random, hash-key (a fixed deterministic ordering), and oracle ground-truth ordering in Table 4. K-OR consistently outperforms non-oracle ordering strategies, supporting the benefit of restoring local sample structure for permutation-robust clustering.

Basis-quality ablation under incomplete observations. To examine the effect of basis quality under severe missingness, Table 5 compares Cor-SVD with several basis variants at a 50% missing rate, where only the basis construction is modified while all other components remain unchanged. Specifically, we consider bases obtained from KNN-completed data, source-data completion (using the original complete data), and an oracle construction. More informative bases improve performance in several cases, highlighting reliable basis estimation as a key challenge under severe missingness.

7 Conclusion

This paper studied sample-permutation sensitivity in t-SVD-based tensorial frameworks for MVC and IMVC. We formulated this problem through two complementary criteria: outcome-oriented permutation invariance and representation-oriented permutation equivariance. Guided by these criteria, we revealed a structural risk of sample-mode FFT and proposed two plug-and-play remedies: K-OR, a heuristic order-restoration strategy that restores a smoother sample arrangement before FFT-based shrinkage, and $\mathcal{T}$-SVD, a permutation-aware transform framework instantiated as Cor-SVD with spectrum-dependent equivariance guarantees. We further introduced a controlled permutation evaluation protocol to assess robustness from both outcome-oriented and representation-oriented perspectives. Experiments show that existing sample-mode FFT frameworks exhibit strong outcome-level and representation-level sensitivity, whereas our methods improve permutation robustness while maintaining competitive clustering performance. Future work will explore more reliable transform construction under severe missingness and more efficient equivariant tensor optimization frameworks.

Acknowledgments

We thank the anonymous reviewers for their constructive feedback. This work was supported by the Australian Research Council through the Linkage Projects scheme (LP240100315) and the Discovery Projects scheme (DP250102634), and by the Queensland State Department of Environment and Science under the Quantum Challenges 2032 Program (Project No. Q2032001). The authors declare no competing interests.

References

[1] Horace B Barlow. Unsupervised learning. *Neural Computation*, 1(3):295–311, 1989.

[2] Jian-Feng Cai, Emmanuel J Candès, and Zuowei Shen. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization*, 20(4):1956–1982, 2010.

[3] Wei Chang, Huimin Chen, Feiping Nie, Rong Wang, and Xuelong Li. Tensorized and compressed multi-view subspace clustering via structured constraint. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10434–10451, 2024.

[4] Guoqing Chao, Shiliang Sun, and Jinbo Bi. A survey on multiview clustering. *IEEE Transactions on Artificial Intelligence*, 2(2):146–168, 2021.

[5] Guoqing Chao, Kaixin Xu, Xijiong Xie, and Yongyong Chen. Global graph propagation with hierarchical information transfer for incomplete contrastive multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 15713–15721, 2025.

[6] Xinyue Chen, Yazhou Ren, Jie Xu, Fangfei Lin, Xiaorong Pu, and Yang Yang. Bridging gaps: Federated multi-view clustering in heterogeneous hybrid views. In *Advances in Neural Information Processing Systems*, pages 1–23, 2024.

[7] Yongyong Chen, Shuqin Wang, Chong Peng, Zhongyun Hua, and Yicong Zhou. Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering. *IEEE Transactions on Image Processing*, 30:4022–4035, 2021.

[8] James W Cooley and John W Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, 1965.

[9] Xiaojian Ding, Lin Zhao, Xian Li, and Xiaoying Zhu. Incomplete multi-view clustering via hierarchical semantic alignment and cooperative completion. In *Advances in Neural Information Processing Systems*, pages 1–21, 2025.

[10] Zhibin Gu, Zhendong Li, and Songhe Feng. Edison: Enhanced dictionary-induced tensorized incomplete multi-view clustering with gaussian error rank minimization. In *Proceedings of the International Conference on Machine Learning*, 2024.

[11] Jipeng Guo, Yanfeng Sun, Junbin Gao, Yongli Hu, and Baocai Yin. Logarithmic Schatten-p norm minimization for tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3396–3410, 2023.

[12] Shudong Huang, Hongjie Wu, Yazhou Ren, Ivor Tsang, Zenglin Xu, Wentao Feng, and Jiancheng Lv. Multi-view subspace clustering on topological manifold. In *Advances in Neural Information Processing Systems*, volume 35, pages 25883–25894, 2022.

[13] Jintian Ji and Songhe Feng. Anchors crash tensor: Efficient and scalable tensorial multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(4): 2660–2675, 2025.

[14] Jintian Ji and Songhe Feng. Anchors bring stability and efficiency: Fast tensorial multi-view clustering on shuffled datasets. In *Proceedings of the ACM International Conference on Multimedia*, pages 1676–1685, 2025.

[15] Eric Kernfeld, Misha Kilmer, and Shuchin Aeron. Tensor–tensor products with invertible linear transforms. *Linear Algebra and its Applications*, 485:545–570, 2015. ISSN 0024-3795.

[16] Shahper Nazeer Khan, Danishuddin, Mohd Wajid Ali Khan, Luca Guarnera, and Syed Mohammad Fauzan Akhtar. Multi-modal ai in precision medicine: integrating genomics, imaging, and ehr data for clinical insights. *Frontiers in Artificial Intelligence*, 8:1743921, 2026.

[17] Misha E Kilmer and Carla D Martin. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3):641–658, 2011.

[18] Zisen Kong, Dongxia Chang, Yiming Wang, Pengyuan Li, and Yao Zhao. Tensorial multi-view clustering via alternative rank minimization and inter-view alignment. *IEEE Transactions on Knowledge and Data Engineering*, 38(5):3196–3209, 2026.

[19] Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian Qu, Yu Feng, Renxiang Guan, et al. Large-scale multi-view tensor clustering with implicit linear kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20727–20736, 2025.

[20] Tianmou Liu, Han Yu, and Rachael Hageman Blair. Stability estimation for unsupervised clustering: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 14(6):e1575, 2022.

[21] Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, and Ce Zhu. S2MVTC: A simple yet efficient scalable multi-view tensor clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26213–26222, 2024.

[22] Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, and Ce Zhu. TLRLF4MVC: Tensor low-rank and low-frequency for scalable multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(8):6900–6911, 2025.

[23] Canyi Lu, Changbo Zhu, Chunyan Xu, Shuicheng Yan, and Zhouchen Lin. Generalized singular value thresholding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.

[24] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5249–5257, 2016.

[25] Canyi Lu, Jiashi Feng, Yudong Chen, Wei Liu, Zhouchen Lin, and Shuicheng Yan. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(4):925–938, 2020.

[26] Feiping Nie, Jingjing Xue, Weizhong Yu, and Xuelong Li. Fast clustering with anchor guidance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(4):1898–1912, 2024.

[27] ErLin Pan and Zhao Kang. Multi-view contrastive graph clustering. In *Advances in Neural Information Processing Systems*, volume 34, pages 2148–2159, 2021.

[28] Xinyu Pu, Hangjun Che, Baicheng Pan, Man-Fai Leung, and Shiping Wen. Robust weighted low-rank tensor approximation for multiview clustering with mixed noise. *IEEE Transactions on Computational Social Systems*, 11(3):3268–3285, 2024.

[29] Yalan Qin, Xinpeng Zhang, Shui Yu, and Guorui Feng. A survey on representation learning for multi-view data. *Neural Networks*, 181:106842, 2025.

[30] Hao Wang, Yan Yang, and Bing Liu. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6):1116–1129, 2020.

[31] Huibing Wang, Yawei Chen, Mingze Yao, Wenzhe Liu, Jinjia Peng, and Xianping Fu. Tensor completion framework by graph refinement for incomplete multi-view clustering. *IEEE Transactions on Multimedia*, 27:9385–9398, 2025.

[32] Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, and En Zhu. Align then fusion: Generalized large-scale multi-view clustering with anchor matching correspondences. In *Advances in Neural Information Processing Systems*, volume 35, pages 5882–5895, 2022.

[33] Xinxin Wang, Yongshan Zhang, and Yicong Zhou. Learn multi-task anchor: Joint view imputation and label generation for incomplete multi-view clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 6451–6459, 2025.

[34] Yapeng Wang, Quanxue Gao, Fangfang Li, Yu Yun, and Ming Yang. Unified view extraction with low-rankness and smoothness fusion for multi-view subspace clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pages 26634–26642, 2026.

[35] Youqing Wang, Yu Cao, Jinlu Wang, Xiang Xu, Jiapu Wang, Tengfei Liu, Junbin Gao, and Jipeng Guo. Aligning collaborative view recovery and tensorial subspace learning via latent representation for incomplete multi-view clustering. In *Proceedings of the International Conference on Learning Representations*, 2026.

[36] Jie Wen, Zheng Zhang, Zhao Zhang, Lei Zhu, Lunke Fei, Bob Zhang, and Yong Xu. Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10273–10281, 2021.

[37] Jie Wen, Zheng Zhang, Lunke Fei, Bob Zhang, Yong Xu, Zhao Zhang, and Jinxing Li. A survey on incomplete multiview clustering. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(2):1136–1149, 2023.

[38] Jianlong Wu, Zhouchen Lin, and Hongbin Zha. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12):5910–5922, 2019.

[39] Wei Xia, Xiangdong Zhang, Quanxue Gao, Xiaochuang Shu, Jungong Han, and Xinbo Gao. Multiview subspace clustering by an enhanced tensor nuclear norm. *IEEE Transactions on Cybernetics*, 52(9):8962–8975, 2022.

[40] Yuan Xie, Dacheng Tao, Wensheng Zhang, Yan Liu, Lei Zhang, and Yanyun Qu. On unifying multi-view self-representations for clustering by tensor multi-rank minimization. *International Journal of Computer Vision*, 126(11):1157–1179, 2018.

[41] Deng Xu, Chao Zhang, Cong Guo, Chunlin Chen, and Huaxiong Li. Fast incomplete multi-view clustering with adaptive similarity completion and reconstruction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21734–21742, 2025.

[42] Tianchuan Yang, Haiqiang Chen, Haoyan Yang, Man-Sheng Chen, Xiangcheng Li, Youming Sun, and Chang-Dong Wang. Smoothness-induced efficient incomplete multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 37(10):6173–6188, 2025.

[43] Shengju Yu, Zhibing Dong, Siwei Wang, Xinhang Wan, Yue Liu, Weixuan Liang, Pei Zhang, Wenxuan Tu, and Xinwang Liu. Towards resource-friendly, extensible and stable incomplete multi-view clustering. In *Proceedings of the International Conference on Machine Learning*, 2024.

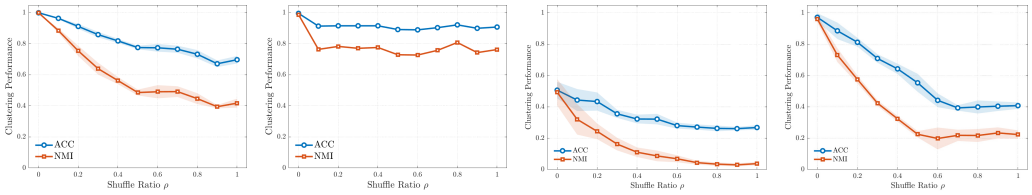
[44] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1582–1590, 2015.

[45] Jing Zhao, Xijiong Xie, Xin Xu, and Shiliang Sun. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, 38:43–54, 2017.

[46] Lihua Zhou, Guowang Du, Kevin Lü, Lizheng Wang, and Jingwei Du. A survey and an empirical evaluation of multi-view clustering approaches. *ACM Computing Surveys*, 56(7): 1–38, 2024.

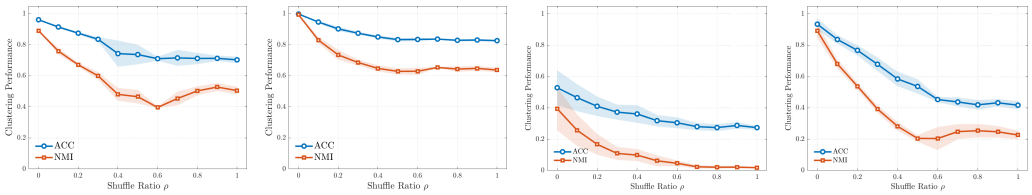
A Permutation Sensitivity Across Diverse Datasets

In the main text, we illustrated the performance degradation of sample-mode FFT-based methods on MSRCv1. We further examine whether similar behavior appears on NGs, BBCSport, and HW. As shown in Figures 4 to 6, these additional datasets exhibit similar sensitivity trends, indicating that the observed order dependence is not specific to MSRCv1.



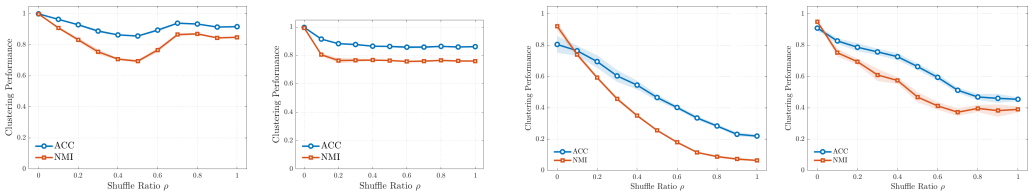
(a) TLSpNM (b) ESTMC (c) SEIC (missing rate:50%) (d) ASCR(missing rate:50%)

Figure 4: Clustering performance of FFT-based t-SVD methods under varying permutation ratios ρ on the NGs dataset.



(a) TLSpNM (b) ESTMC (c) SEIC (missing rate:50%) (d) ASCR(missing rate:50%)

Figure 5: Clustering performance of FFT-based t-SVD methods under varying permutation ratios ρ on the BBCSport dataset.



(a) TLSpNM (b) ESTMC (c) SEIC (missing rate:50%) (d) ASCR(missing rate:50%)

Figure 6: Clustering performance of FFT-based t-SVD methods under varying permutation ratios ρ on the HW dataset.

B Proofs of Theoretical Results

In this section, we provide the detailed proofs for the propositions and theorems presented in the main text.

B.1 Sufficient Condition for Outcome-oriented Invariance

Proof. Let $\mathbb{X}$ be a multi-view dataset and let $\mathbf{Z} = f(\mathbb{X})$ be its learned representation. For any permutation matrix $\mathbf{P}_n \in \Pi_n$, representation-oriented permutation equivariance gives

$$f(\mathbf{P}_n\mathbb{X}) = \mathbf{P}_n f(\mathbb{X}) = \mathbf{P}_n\mathbf{Z}. \tag{12}$$

Assume that the downstream clustering operator $\mathcal{C}$ is permutation equivariant, up to an arbitrary relabeling of cluster indices. Then

$$\mathcal{C}\big(f(\mathbf{P}_n\mathbb{X})\big) = \mathcal{C}(\mathbf{P}_n\mathbf{Z}) \equiv \mathbf{P}_n\mathcal{C}(\mathbf{Z}), \tag{13}$$

where $\equiv$ denotes equality up to a permutation of cluster labels.

Let $\hat{\mathbf{y}} = \mathcal{C}(\mathbf{Z})$. Since standard clustering metrics $\mathcal{M}$ are invariant to synchronized sample re-indexing and cluster-label permutation,

$$\mathcal{M}\big(\mathcal{C}(f(\mathbf{P}_n\mathbb{X})), \mathbf{P}_n\mathbf{y}\big) = \mathcal{M}(\hat{\mathbf{y}}, \mathbf{y}). \tag{14}$$

Therefore, representation-oriented permutation equivariance, together with a permutation-equivariant downstream clustering operator, is sufficient for outcome-oriented permutation invariance. □

B.2 Proof of Proposition 4.1: Non-Equivariance of FFT

Proof. Following the definition of the mode-3 FFT transform:

$$\begin{aligned} \text{fft}(\mathcal{A} \times_3 \mathbf{P}_n, 3) &= (\mathcal{A} \times_3 \mathbf{P}_n) \times_3 \mathbf{F} = \mathcal{A} \times_3 (\mathbf{F}\mathbf{P}_n), \\ \text{fft}(\mathcal{A}, 3) \times_3 \mathbf{P}_n &= (\mathcal{A} \times_3 \mathbf{F}) \times_3 \mathbf{P}_n = \mathcal{A} \times_3 (\mathbf{P}_n\mathbf{F}), \end{aligned} \tag{15}$$

where $\mathbf{F}$ is the fixed Discrete Fourier Transform matrix. The non-commutativity of the DFT matrix and an arbitrary permutation matrix, i.e., $\mathbf{F}\mathbf{P}_n \neq \mathbf{P}_n\mathbf{F}$ for $n > 2$ in general, implies that the transform does not preserve sample-wise correspondence. Consequently, the standard FFT-based t-SVD does not guarantee permutation equivariance.

Note that the DFT matrix diagonalizes cyclic permutations and may commute with a restricted subset of structured permutations. However, for generic permutations, the non-commutativity $\mathbf{F}\mathbf{P}_n \neq \mathbf{P}_n\mathbf{F}$ holds. □

B.3 Proof of Proposition 4.3: Permutation Equivariance of $\mathcal{T}$-SVD mapping

Proof. The $\mathcal{T}$-SVD proximal mapping $f_{\mathcal{T}}(\mathcal{A})$ is defined as:

$$f_{\mathcal{T}}(\mathcal{A}) = \mathcal{G}_\tau\left(\mathcal{A} \times_3 \mathbf{T}(\mathbb{X})^\top\right) \times_3 \mathbf{T}(\mathbb{X}), \tag{16}$$

where $\mathcal{G}_\tau(\cdot)$ denotes the slice-wise singular value shrinkage operator. Under sample permutation $\mathbf{P}_n$, the transform basis satisfies $\mathbf{T}(\mathbf{P}_n\mathbb{X}) = \mathbf{P}_n\mathbf{T}(\mathbb{X})$ (Assumption 4.2). For the permuted input $\tilde{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{P}_n$, the transform-domain representation is:

$$\begin{aligned} \tilde{\mathcal{A}}_{\mathcal{T}} &= (\mathcal{A} \times_3 \mathbf{P}_n) \times_3 (\mathbf{P}_n\mathbf{T}(\mathbb{X}))^\top \\ &= \mathcal{A} \times_3 \left(\mathbf{T}(\mathbb{X})^\top\mathbf{P}_n^\top\mathbf{P}_n\right) \\ &= \mathcal{A} \times_3 \mathbf{T}(\mathbb{X})^\top = \mathcal{A}_{\mathcal{T}}. \end{aligned} \tag{17}$$

Since $\mathbf{P}_n$ is orthogonal ($\mathbf{P}_n^\top\mathbf{P}_n = \mathbf{I}$), the spectral coefficients in the transform domain remain identical to those obtained from the original input. Applying the inverse transform with the permuted basis yields:

$$\begin{aligned} f_{\mathcal{T}}(\tilde{\mathcal{A}}) &= \mathcal{G}_\tau(\mathcal{A}_{\mathcal{T}}) \times_3 (\mathbf{P}_n\mathbf{T}(\mathbb{X})) \\ &= (\mathcal{G}_\tau(\mathcal{A}_{\mathcal{T}}) \times_3 \mathbf{T}(\mathbb{X})) \times_3 \mathbf{P}_n = f_{\mathcal{T}}(\mathcal{A}) \times_3 \mathbf{P}_n. \end{aligned} \tag{18}$$

This confirms that the $\mathcal{T}$-SVD proximal mapping is permutation equivariant. □

B.4 Proof of Theorem 4.4: Conditional Permutation Equivariance of Cor-SVD

We first establish two properties of the slice-wise shrinkage operator.

Lemma B.1 (Properties of Slice-wise Singular Value Shrinkage)**.** *For a matrix with SVD* $\mathbf{B} = \mathbf{U}\,\mathrm{diag}(\sigma_i)\mathbf{V}^\top$*, define*

$$G_\tau(\mathbf{B}) = \mathbf{U}\,\mathrm{diag}\big(s_\tau(\sigma_i)\big)\mathbf{V}^\top,$$

where s_τ *is a fixed single-valued rule satisfying* $0 \le s_\tau(\sigma) \le \sigma$ *for every* $\sigma \ge 0$*. Let* $\mathcal{G}_\tau$ *apply* G_τ *independently to each frontal slice, that is,*

$$[\mathcal{G}_\tau(\mathcal{B})]^{(k)} = G_\tau(\mathcal{B}^{(k)}).$$

Then, for any tensor $\mathcal{B} \in \mathbb{R}^{d\times m\times n}$ *and diagonal sign matrix* $\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_n)$ *with* $d_k \in \{-1, 1\}$*,*

$$\mathcal{G}_\tau(\mathcal{B} \times_3 \mathbf{D}) = \mathcal{G}_\tau(\mathcal{B}) \times_3 \mathbf{D}, \tag{19}$$

$$\|\mathcal{G}_\tau(\mathcal{B})\|_F \le \|\mathcal{B}\|_F. \tag{20}$$

Proof. If $\mathbf{B} = \mathbf{U}\,\mathrm{diag}(\sigma_i)\mathbf{V}^\top$ is an SVD of $\mathbf{B}$, then $-\mathbf{B} = (-\mathbf{U})\,\mathrm{diag}(\sigma_i)\mathbf{V}^\top$ is an SVD of $-\mathbf{B}$. Applying the same scalar shrinkage rule therefore gives

$$G_\tau(-\mathbf{B}) = -G_\tau(\mathbf{B}).$$

Since $(\mathcal{B} \times_3 \mathbf{D})^{(k)} = d_k \mathcal{B}^{(k)}$, applying this identity to each frontal slice proves Eq. (19).

Let $\sigma_i^{(k)}$ denote the singular values of $\mathcal{B}^{(k)}$. By orthogonal invariance of the Frobenius norm and $0 \le s_\tau(\sigma) \le \sigma$,

$$\|\mathcal{G}_\tau(\mathcal{B})\|_F^2 = \sum_{k=1}^{n} \sum_{i=1}^{\min(d,m)} s_\tau(\sigma_i^{(k)})^2 \le \sum_{k=1}^{n} \sum_{i=1}^{\min(d,m)} (\sigma_i^{(k)})^2 = \|\mathcal{B}\|_F^2.$$

Taking square roots proves Eq. (20). □

We now apply Lemma B.1 to prove Theorem 4.4.

Proof. Let $\mathcal{G}_\tau(\cdot)$ be the slice-wise shrinkage operator defined in Lemma B.1. For $\mathbf{T}_c = \mathbf{T}_c(\mathbb{X})$, the Cor-SVD mapping can be written as

$$f_{\mathrm{Cor}}(\mathcal{A}) = \mathcal{G}_\tau\left(\mathcal{A} \times_3 \mathbf{T}_c^\top\right) \times_3 \mathbf{T}_c. \tag{21}$$

We suppress the dependence of f_{Cor} on $\mathbb{X}$. For permuted inputs, the basis is constructed from $\mathbf{P}_n\mathbb{X}$. For simplicity, denote $\bar{\mathcal{A}} = \mathcal{A} \times_3 \mathbf{T}_c^\top$.

Let $\mathbf{S}'$ be the sample correlation matrix constructed from $\mathbf{P}_n\mathbb{X}$. For complete MVC,

$$\mathbf{S}' = \frac{1}{m}\sum_{v=1}^{m}(\mathbf{P}_n\mathbf{X}^{(v)})(\mathbf{P}_n\mathbf{X}^{(v)})^\top = \mathbf{P}_n\mathbf{S}\mathbf{P}_n^\top.$$

For IMVC, the observation masks are permuted jointly with the data. Both the numerator and the co-occurrence matrix in Eq. (9) then transform by conjugation with $\mathbf{P}_n$. Entry-wise division preserves this relation, with $S_{ij} = 0$ whenever $O_{ij} = 0$. Thus, $\mathbf{S}' = \mathbf{P}_n\mathbf{S}\mathbf{P}_n^\top$ also holds for IMVC.

Since $\mathbf{S} = \mathbf{T}_c\mathbf{\Lambda}\mathbf{T}_c^\top$,

$$\mathbf{S}' = (\mathbf{P}_n\mathbf{T}_c)\mathbf{\Lambda}(\mathbf{P}_n\mathbf{T}_c)^\top.$$

Using the same eigenvalue ordering for both orthonormal eigenbases, we obtain

$$\mathbf{T}_c' = \mathbf{T}_c(\mathbf{P}_n\mathbb{X}) = \mathbf{P}_n\mathbf{T}_c\mathbf{Q}_P, \tag{22}$$

where $\mathbf{Q}_P$ is block-orthogonal within eigenspaces associated with the same eigenvalue. It follows that

$$(\mathbf{T}_c')^\top = \mathbf{Q}_P^\top\mathbf{T}_c^\top\mathbf{P}_n^\top. \tag{23}$$

Under the adopted mode-3 product convention, $(\mathcal{X} \times_3 \mathbf{A}) \times_3 \mathbf{B} = \mathcal{X} \times_3 (\mathbf{B}\mathbf{A})$. Therefore, the forward transform of the permuted tensor satisfies

$$\begin{aligned}(\mathcal{A} \times_3 \mathbf{P}_n) \times_3 (\mathbf{T}_c')^\top &= (\mathcal{A} \times_3 \mathbf{P}_n) \times_3 \left(\mathbf{Q}_P^\top\mathbf{T}_c^\top\mathbf{P}_n^\top\right)\\ &= \mathcal{A} \times_3 \left(\mathbf{Q}_P^\top\mathbf{T}_c^\top\mathbf{P}_n^\top\mathbf{P}_n\right)\\ &= \mathcal{A} \times_3 \left(\mathbf{Q}_P^\top\mathbf{T}_c^\top\right)\\ &= \bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top.\end{aligned} \tag{24}$$

Applying the corresponding inverse transform gives

$$\begin{aligned}f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) &= \mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \mathbf{T}_c'\\ &= \mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \left(\mathbf{P}_n\mathbf{T}_c\mathbf{Q}_P\right).\end{aligned} \tag{25}$$

Using the associativity of the mode-3 product again, this can be rewritten as

$$f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) = \left[\mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \mathbf{Q}_P\right] \times_3 \left(\mathbf{P}_n\mathbf{T}_c\right). \tag{26}$$

On the other hand, from Eq. (21),

$$\begin{aligned} f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n &= \left[\mathcal{G}_\tau(\bar{\mathcal{A}}) \times_3 \mathbf{T}_c\right] \times_3 \mathbf{P}_n \\ &= \mathcal{G}_\tau(\bar{\mathcal{A}}) \times_3 (\mathbf{P}_n \mathbf{T}_c)\,. \end{aligned} \tag{27}$$

Subtracting Eq. (27) from Eq. (26), we obtain

$$\begin{aligned} &f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) - f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n \\ &= \left[\mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \mathbf{Q}_P - \mathcal{G}_\tau(\bar{\mathcal{A}})\right] \times_3 (\mathbf{P}_n \mathbf{T}_c)\,. \end{aligned} \tag{28}$$

Since both $\mathbf{P}_n$ and $\mathbf{T}_c$ are orthogonal, $\mathbf{P}_n\mathbf{T}_c$ is also orthogonal. Hence, multiplication along mode-3 by $\mathbf{P}_n\mathbf{T}_c$ preserves the Frobenius norm. Therefore,

$$\begin{aligned} &\left\| f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) - f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n \right\|_F \\ &= \left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \mathbf{Q}_P - \mathcal{G}_\tau(\bar{\mathcal{A}}) \right\|_F . \end{aligned} \tag{29}$$

Simple-spectrum case. If $\mathbf{S}$ has a simple spectrum, every eigenspace is one-dimensional. Therefore, $\mathbf{Q}_P$ reduces to a diagonal sign matrix $\mathbf{D}_P$ satisfying $\mathbf{D}_P^\top = \mathbf{D}_P$ and $\mathbf{D}_P^2 = \mathbf{I}$.

Multiplication by $\mathbf{D}_P$ only changes the sign of individual transformed frontal slices. Applying Lemma B.1 with $\mathbf{D} = \mathbf{D}_P$ gives

$$\mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{D}_P\right) = \mathcal{G}_\tau(\bar{\mathcal{A}}) \times_3 \mathbf{D}_P. \tag{30}$$

Consequently,

$$\begin{aligned} \mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{D}_P\right) \times_3 \mathbf{D}_P &= \mathcal{G}_\tau(\bar{\mathcal{A}}) \times_3 \mathbf{D}_P^2 \\ &= \mathcal{G}_\tau(\bar{\mathcal{A}}). \end{aligned} \tag{31}$$

Substituting this identity into Eq. (29) yields

$$f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) = f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n. \tag{32}$$

Repeated-eigenvalue case. If $\mathbf{S}$ has repeated eigenvalues, $\mathbf{Q}_P$ is block diagonal and can be written as

$$\mathbf{Q}_P = \mathrm{blkdiag}\left(\mathbf{Q}_{P,1}, \ldots, \mathbf{Q}_{P,L}\right), \tag{33}$$

where L is the number of distinct eigenvalues and $\mathbf{Q}_{P,\ell}$ is the orthogonal block acting within the ℓ-th eigenspace.

For each one-dimensional eigenspace, $\mathbf{Q}_{P,\ell} \in \{-1, 1\}$, and its contribution cancels exactly as in the simple-spectrum case. Thus, only repeated eigenspaces indexed by $\mathcal{R}$ may contribute to the equivariance error.

For each $\ell \in \mathcal{R}$, let $\mathcal{I}_\ell$ denote the transformed-slice indices associated with the ℓ-th repeated eigenspace, and define the corresponding block error as

$$\mathcal{E}_\ell = \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell} \times_3 \mathbf{Q}_{P,\ell}^\top\right) \times_3 \mathbf{Q}_{P,\ell} - \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell}\right). \tag{34}$$

Since the eigenspace blocks correspond to disjoint sets of transformed frontal slices, their squared Frobenius norms add. Hence,

$$\left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}} \times_3 \mathbf{Q}_P^\top\right) \times_3 \mathbf{Q}_P - \mathcal{G}_\tau(\bar{\mathcal{A}}) \right\|_F^2 = \sum_{\ell \in \mathcal{R}} \|\mathcal{E}_\ell\|_F^2. \tag{35}$$

We next bound each $\mathcal{E}_\ell$. By Lemma B.1, the shrinkage operator satisfies

$$\|\mathcal{G}_\tau(\mathcal{B})\|_F \le \|\mathcal{B}\|_F. \tag{36}$$

Using the triangle inequality, orthogonal invariance of the Frobenius norm, and Eq. (36), we obtain

$$\begin{aligned} \|\mathcal{E}_\ell\|_F &\le \left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell} \times_3 \mathbf{Q}_{P,\ell}^\top\right) \times_3 \mathbf{Q}_{P,\ell} \right\|_F + \left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell}\right) \right\|_F \\ &= \left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell} \times_3 \mathbf{Q}_{P,\ell}^\top\right) \right\|_F + \left\| \mathcal{G}_\tau\left(\bar{\mathcal{A}}_{\mathcal{I}_\ell}\right) \right\|_F \\ &\le \left\| \bar{\mathcal{A}}_{\mathcal{I}_\ell} \times_3 \mathbf{Q}_{P,\ell}^\top \right\|_F + \left\| \bar{\mathcal{A}}_{\mathcal{I}_\ell} \right\|_F \\ &= 2 \left\| \bar{\mathcal{A}}_{\mathcal{I}_\ell} \right\|_F . \end{aligned} \tag{37}$$

Combining Eqs. (29), (35) and (37), we obtain

$$\begin{aligned} &\left\|f_{\mathrm{Cor}}\left(\mathcal{A} \times_3 \mathbf{P}_n\right) - f_{\mathrm{Cor}}(\mathcal{A}) \times_3 \mathbf{P}_n\right\|_F \\ &= \left(\sum_{\ell \in \mathcal{R}} \|\mathcal{E}_\ell\|_F^2\right)^{1/2} \\ &\leq 2\left(\sum_{\ell \in \mathcal{R}} \left\|\bar{\mathcal{A}}_{\mathcal{I}_\ell}\right\|_F^2\right)^{1/2}. \end{aligned} \tag{38}$$

□

C Extended Experiments Section

All algorithms are implemented in MATLAB R2020b and executed on a workstation equipped with an Intel (R) Core (TM) Ultra 5 225H (1.70 GHz) and 32GB of RAM.

C.1 Experimental Setup

Datasets: To verify the effectiveness of our proposed methods, four challenging datasets with different properties are used and summarized in Table 6. More detailed information about these datasets can be found below. To construct the shuffled datasets, we apply the same random sample permutation to all views and the corresponding labels.

Table 6: Summary of the benchmark datasets.

Dataset	Sam./Clu.	Dimensions
MSRCv1	210 / 7	24/576/512/256/254
NGs	500 / 5	2000/2000/2000
BBCSport	544/5	3183/3203
HW	2000 /10	216/76/64/6/240/47

- **MSRCv1**[2]. It consists of 210 scene recognition images and is divided into 7 categories. Each image is represented by five different feature sets.
- **NGs**[3]. It is a subset of the 20 Newsgroup dataset and consists of 500 newsgroup documents. Each raw document is pre-processed with three different methods (three views) and is assigned one of five class labels.
- **BBCSport**[4]. It consists of 544 sports news articles in five topical areas (athletics, cricket, football, rugby, and tennis). Each sample is described by two views, and their dimensions are 3183 and 3203, respectively.
- **HandWritten (HW)**. It comes from the UCI repository and consists of handwritten numbers, divided into 10 categories from 0 to 9. Each image is represented by six different feature sets.

Incomplete Data Settings: Following the experimental setting in [10], we simulate incomplete multi-view datasets by randomly masking instances to achieve a target missing rate $\eta \in \{0.1, 0.5\}$. Crucially, to ensure no instance is entirely excluded from the learning process, we strictly enforce the constraint that every sample must be observed in at least one view (i.e., $\sum_{v=1}^{m} o_i^{(v)} \geq 1$). If the initial random masking results in a sample being missing across all views, we dynamically restore its observation in a randomly selected view to satisfy this condition while maintaining the overall missing rate η.

[2]http://research.microsoft.com/en-us/projects/objectclassrecognition

[3]https://lig-membres.imag.fr/grimal/data.html

[4]http://mlg.ucd.ie/datasets/segment.html

Baselines: To provide a comprehensive evaluation, we select four representative state-of-the-art (SOTA), open-source FFT-based low-rank tensor learning algorithms, encompassing both complete and incomplete data scenarios. The algorithmic details and hyperparameter configurations for these baselines are outlined as follows:

- **TLSpNM (TPAMI 2023)** [11] is a subspace-based tensorial multi-view clustering method that uses the non-convex Schatten-p norm to approximate the tensor rank. The range of the hyperparameter λ is $\{1\mathrm{e}{-4}, 2\mathrm{e}{-4}, \cdots, 9\mathrm{e}{-4}\}$, consistent with the original article.
- **ESTMC (TPAMI 2025)** [13] is an anchor-based fast tensorial multi-view clustering method, which develops a general framework for non-convex functions to perform low-rank approximation. Consistent with the baseline literature, the hyperparameters are selected from $\lambda \in \{10^{-6}, 10^{-5}, \ldots, 1\}$ and $\delta \in \{10^{-2}, 10^{-1}, 1\}$.
- **SEIC (TKDE 2025)** [42] combines low-rank tensor completion, anchor-based acceleration, and smoothness constraints for incomplete multi-view clustering. To accommodate our variants (FFT-view, K-OR, and Cor-SVD), we remove the original difference-operator terms while keeping the remaining framework unchanged. For the adapted variants, we re-select β from $\{0.05, 0.5, 5, 50, 500\}$.
- **ASCR (AAAI 2025)** [41] constructs a graph representation tensor between samples and anchors, and utilizes the low-rank property of the tensor for missing data recovery and clustering. We set $k = K$ and select α from $\{2^2, 2^3, \ldots, 2^7\}$.

To ensure rigorous and fair comparisons, all experiments are conducted using the official implementations provided by the respective authors. Since K-OR adopts standard k-means as an intermediate partitioning step, it may also inherit sensitivity to k-means initialization. In practice, this effect can be reduced using multiple k-means restarts. Here, we focus specifically on the K-OR-specific hyperparameter k_{or} and evaluate its effect under both TLSpNM and ESTMC frameworks. As shown

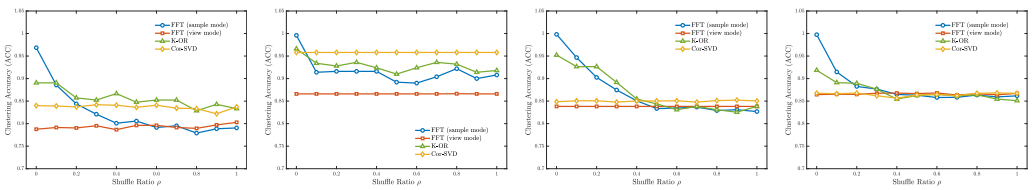
Table 7: Sensitivity of K-OR to the restoration cluster number k_{or} on MSRCv1.

Framework	**Metric**	$k_{or} = K-2$	$k_{or} = K-1$	$k_{or} = K$	$k_{or} = K+1$	$k_{or} = K+2$	$k_{or} = 2K$
TLSpNM	ACC (%)	73.94	78.92	80.46	79.83	**81.17**	73.68
	NMI (%)	64.30	66.52	69.22	68.49	**69.75**	60.25
ESTMC	ACC (%)	83.03	84.26	85.58	**85.65**	85.32	84.33
	NMI (%)	72.79	74.63	75.67	**75.78**	75.34	73.74

in Table 7, K-OR maintains stable clustering performance when k_{or} varies within a reasonable neighborhood of the true cluster number K under both TLSpNM and ESTMC. Based on this observation, we set $k_{or} = K$ in all other experiments.

Table 8: Clustering performance (mean (std)%) on complete datasets BBCSport and HW.

Baselines	**Variants**	**BBCSport**					**HW**				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
TLSpNM	FFT (sample mode)	78.38 (9.77)	56.87 (14.59)	80.54 (7.84)	67.15 (12.38)	57.76 (15.78)	91.82 (4.08)	82.67 (8.70)	91.82 (4.08)	84.66 (7.49)	82.96 (8.31)
	FFT (view mode)	76.14 (0.08)	62.04 (0.19)	81.29 (0.09)	67.28 (0.12)	58.07 (0.12)	81.97 (0.04)	74.67 (0.06)	81.97 (0.08)	71.96 (0.05)	68.81 (0.06)
	K-OR	**91.29** (2.16)	**77.31** (4.59)	**91.29** (2.16)	**84.35** (3.59)	**79.53** (4.67)	**94.05** (2.64)	**88.06** (4.20)	**94.05** (2.64)	**88.76** (4.66)	**87.51** (5.18)
	Cor-SVD	86.40 **(0)**	75.33 **(0)**	86.40 **(0)**	83.52 **(0)**	78.27 **(0)**	93.26 **(0.01)**	87.37 **(0.04)**	93.26 **(0.02)**	87.24 **(0.03)**	85.82 **(0.03)**
ESTMC	FFT (sample mode)	**86.95** (5.55)	70.33 (11.06)	**86.95** (5.55)	**77.46** (9.39)	**70.88** (12.08)	**88.18** (4.06)	78.74 (6.74)	**88.18** (4.06)	**78.80** (7.07)	**76.45** (7.86)
	FFT (view mode)	83.82 **(0)**	64.08 **(0)**	83.82 **(0)**	71.72 **(0)**	63.56 **(0)**	86.60 **(0.57)**	76.81 **(0.59)**	86.60 **(0.57)**	76.10 **(0.78)**	73.44 **(0.87)**
	K-OR	86.90 (4.47)	**70.60** (7.97)	86.90 (4.47)	76.99 (7.68)	70.12 (9.84)	87.17 (1.98)	77.90 (3.02)	87.17 (1.98)	77.05 (3.07)	74.50 (3.41)
	Cor-SVD	86.40 **(0)**	68.66 **(0)**	86.40 **(0)**	75.47 **(0)**	68.38 **(0)**	87.94 (0.58)	**78.94** (0.76)	87.94 (0.58)	78.08 (0.81)	75.64 (0.90)

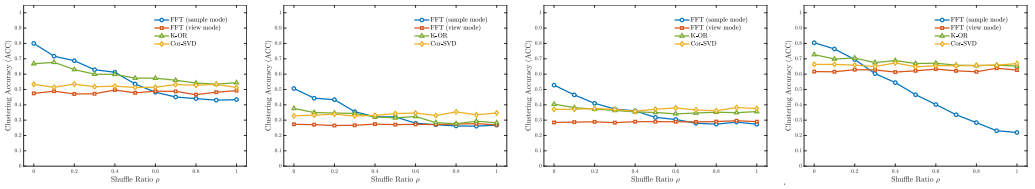


(a) MSRCv1 (b) NGs (c) BBCSport (d) HW

Figure 7: Performance trajectories of four ESTMC variants on MSRCv1, NGs, BBCSport, and HW.

Table 9: Clustering performance (mean (std)%) on incomplete datasets NGs, BBCSport, and HW with different missing rates (10% and 50%).

Baselines	Variants	NGs (Missing rate: 10%)					NGs (Missing rate: 50%)				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
SEIC	FFT (sample mode)	42.60 (15.13)	23.84 (21.90)	45.66 (15.98)	37.94 (14.56)	18.17 (19.78)	**33.84** (8.96)	14.78 (14.86)	**36.45** (10.42)	**31.82** (7.96)	**8.61** (10.70)
	FFT (view mode)	31.49 (**1.41**)	14.15 (**2.51**)	34.34 (**0.85**)	32.34 (**1.54**)	4.38 (**0.95**)	27.13 (**1.39**)	9.48 (**2.13**)	28.97 (**1.56**)	31.65 (**0.71**)	1.46 (**0.50**)
	K-OR	36.55 (5.05)	15.20 (7.37)	38.26 (5.27)	31.24 (4.47)	9.89 (6.02)	31.93 (4.78)	10.30 (6.05)	33.25 (5.08)	29.42 (3.91)	5.85 (5.07)
	Cor-SVD	**46.10** (3.65)	**37.46** (4.69)	**51.73** (3.40)	**40.77** (2.61)	**20.33** (3.90)	33.74 (2.57)	**15.93** (3.11)	36.10 (2.54)	30.44 (1.42)	5.10 (1.46)
ASCR	FFT (sample mode)	77.09 (9.71)	59.22 (15.18)	78.16 (8.95)	65.74 (13.04)	56.42 (16.79)	**60.16** (20.92)	**39.31** (24.79)	**61.03** (20.41)	**49.65** (22.04)	**33.88** (30.08)
	FFT (view mode)	**88.47** (4.58)	**80.24** (**3.23**)	**89.68** (**3.59**)	**83.39** (**4.09**)	**79.01** (**5.30**)	31.33 (14.49)	12.74 (14.85)	32.21 (15.10)	34.55 (**2.68**)	7.13 (9.59)
	K-OR	78.79 (4.70)	61.69 (4.85)	79.92 (4.03)	68.00 (4.69)	59.42 (6.21)	49.47 (5.39)	30.74 (5.95)	50.99 (5.06)	38.24 (3.48)	17.25 (5.26)
	Cor-SVD	78.93 (**4.14**)	68.46 (4.07)	80.39 (3.72)	70.46 (4.52)	62.07 (5.99)	53.43 (**4.78**)	35.04 (**5.55**)	54.48 (**4.63**)	39.32 (3.16)	20.09 (**4.37**)
Baselines	**Variants**	**BBCSport (Missing rate: 10%)**					**BBCSport (Missing rate: 50%)**				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
SEIC	FFT (sample mode)	44.21 (17.16)	21.85 (23.18)	51.52 (16.16)	38.60 (16.75)	18.35 (22.11)	34.05 (10.92)	10.14 (13.30)	41.98 (9.39)	29.63 (8.82)	6.73 (11.44)
	FFT (view mode)	41.60 (**1.23**)	15.90 (**2.13**)	43.69 (**1.06**)	38.31 (**0.76**)	3.59 (**1.54**)	34.74 (1.81)	5.74 (**1.46**)	38.50 (**0.97**)	**34.43** (**2.18**)	1.10 (**0.08**)
	K-OR	44.30 (4.86)	21.89 (9.31)	47.84 (5.70)	39.70 (4.33)	8.77 (8.30)	36.43 (2.84)	10.85 (7.49)	42.00 (5.06)	34.39 (2.38)	4.16 (4.64)
	Cor-SVD	**58.53** (6.07)	**44.94** (6.52)	**65.66** (4.56)	**49.33** (5.83)	**32.88** (7.98)	**37.11** (**1.79**)	**15.37** (6.02)	**45.62** (4.75)	33.52 (2.38)	**6.87** (3.97)
ASCR	FFT (sample mode)	**78.23** (8.50)	64.52 (12.25)	**81.81** (6.84)	**71.15** (10.96)	**62.47** (14.13)	**61.52** (17.87)	**39.39** (23.74)	**66.63** (16.03)	**53.11** (19.80)	**39.03** (25.72)
	FFT (view mode)	69.11 (4.65)	53.70 (4.32)	72.78 (3.37)	59.42 (4.08)	47.35 (5.28)	35.66 (**0.20**)	1.72 (**0.28**)	36.08 (**0.15**)	38.55 (**0.14**)	0.24 (**0.22**)
	K-OR	76.87 (3.45)	60.68 (4.65)	79.28 (2.40)	67.89 (4.11)	57.94 (5.19)	46.59 (2.93)	27.82 (4.03)	54.30 (3.11)	39.63 (2.97)	18.38 (3.91)
	Cor-SVD	78.16 (**2.52**)	**64.97** (**2.04**)	80.68 (**1.41**)	69.59 (**2.73**)	60.39 (**3.44**)	42.77 (2.07)	25.19 (3.42)	52.37 (2.91)	36.19 (2.18)	15.66 (2.92)
Baselines	**Variants**	**HW (Missing rate: 10%)**					**HW (Missing rate: 50%)**				
		ACC	NMI	Purity	F-score	ARI	ACC	NMI	Purity	F-score	ARI
SEIC	FFT (sample mode)	49.55 (19.48)	35.42 (27.64)	51.89 (20.80)	37.58 (23.28)	30.19 (25.87)	48.64 (20.05)	34.86 (28.06)	50.95 (21.41)	37.30 (23.53)	29.87 (26.17)
	FFT (view mode)	**82.29** (2.49)	79.68 (1.88)	**83.98** (2.05)	**76.86** (2.28)	**74.17** (2.56)	65.57 (2.01)	59.79 (1.63)	67.09 (1.86)	46.54 (2.16)	39.39 (2.53)
	K-OR	80.60 (2.27)	**80.88** (1.61)	83.18 (1.93)	76.83 (2.04)	74.07 (2.31)	**67.50** (2.66)	**63.45** (3.19)	**69.66** (2.64)	**50.75** (4.21)	**44.14** (4.90)
	Cor-SVD	80.78 (**2.06**)	78.10 (**1.55**)	82.93 (**1.73**)	75.09 (**1.99**)	72.21 (**2.23**)	66.18 (**1.95**)	61.30 (**1.17**)	68.03 (**1.66**)	49.87 (**1.88**)	43.47 (**2.20**)
ASCR	FFT (sample mode)	**80.17** (5.72)	**77.53** (6.95)	**82.13** (5.23)	**73.75** (7.78)	**70.73** (8.66)	**65.09** (15.95)	**54.92** (18.27)	**66.19** (16.01)	**51.61** (20.97)	**45.71** (23.71)
	FFT (view mode)	77.06 (**2.26**)	73.49 (**1.82**)	79.07 (**1.87**)	69.01 (**2.23**)	65.47 (**2.50**)	42.52 (**2.35**)	35.76 (**1.88**)	43.41 (**2.27**)	29.74 (**1.45**)	21.28 (**1.64**)
	K-OR	77.45 (2.39)	76.04 (2.04)	79.90 (1.99)	70.66 (2.41)	67.24 (2.69)	53.08 (2.85)	50.37 (2.42)	54.82 (2.63)	34.60 (2.25)	25.42 (2.78)
	Cor-SVD	77.49 (2.50)	74.16 (1.87)	79.33 (2.18)	69.78 (2.29)	66.31 (2.56)	56.16 (3.78)	48.13 (3.53)	57.36 (3.80)	41.93 (3.79)	35.31 (4.25)



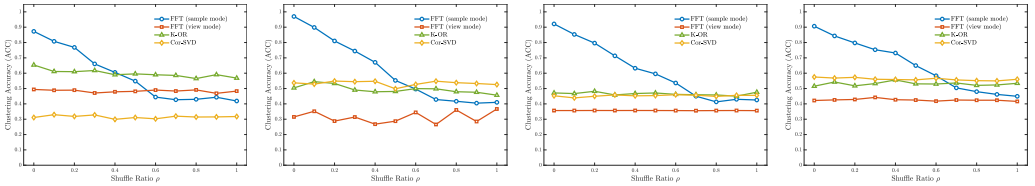
(a) MSRCv1 (b) NGs (c) BBCSport (d) HW

Figure 8: Performance trajectories of four SEIC variants on MSRCv1, NGs, BBCSport, and HW (missing rate = 50%).

C.2 More Results for Outcome-oriented Permutation Invariance

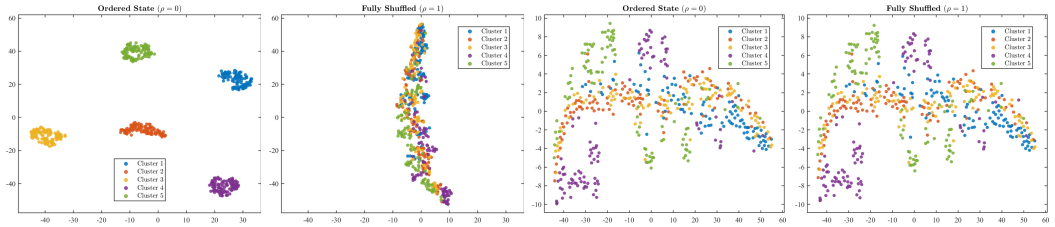
While the main text presents representative results across four datasets, we provide further detailed evaluations in Tables 8 and 9 and Figures 7 to 9. These additional results are intended to examine whether the observations in the main text generalize across different datasets, baselines, and missing rates.

Overall, the additional results are consistent with the main findings. FFT (sample mode) frequently exhibits large performance variability under sample permutations, indicating clear outcome-level order sensitivity. For IMVC, the observed variability may additionally contain the effect of random

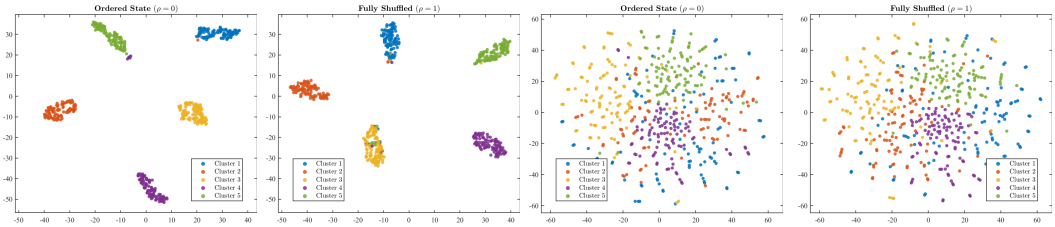


(a) MSRCv1 (b) NGs (c) BBCSport (d) HW

Figure 9: Performance trajectories of four ASCR variants on MSRCv1, NGs, BBCSport, and HW (missing rate = 50%).



(a) FFT (sample mode) ($\Delta_F = 3.55$) (b) FFT (view mode) ($\Delta_F = 1.4 \times 10^{-13}$)



(c) K-OR ($\Delta_F = 3.41$) (d) Cor-SVD ($\Delta_F = 3.11 \times 10^{-12}$)

Figure 10: t-SNE visualization for different strategies with the TLSpNM baseline on NGs. Δ_F denotes the normalized Frobenius discrepancy between the two representations after aligning samples back to the same indexing.

missing patterns. In contrast, FFT (view mode) is generally more stable because it avoids transforming along the sample dimension, but this often comes at the cost of reduced clustering accuracy.

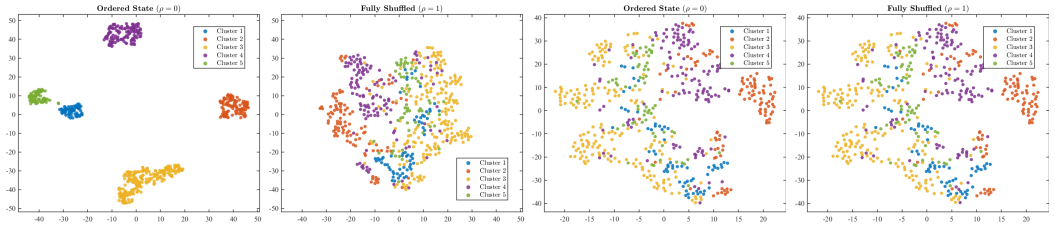
K-OR and Cor-SVD provide more favorable accuracy–stability trade-offs in most settings. K-OR improves robustness by restoring a more favorable sample order before FFT-based shrinkage, while Cor-SVD reduces permutation-induced variation through a data-dependent, permutation-aligned transform. The trajectory plots in Figures 7 to 9 further show that both variants are generally less sensitive to increasing shuffle ratios than FFT (sample mode).

For IMVC, however, Cor-SVD can still experience reduced clustering accuracy under severe missingness despite lower performance variability. Since its transform basis is estimated directly from the observed data, heavily incomplete observations may distort the estimated sample relationships and compromise basis quality. For example, on HW with ASCR at a 50% missing rate, Cor-SVD achieves 56.16% ACC with a standard deviation of 3.78%, whereas FFT (sample mode) achieves a higher ACC of 65.09% but a much larger standard deviation of 15.95%. This illustrates that improved empirical stability does not necessarily translate into higher clustering accuracy. Thus, while Cor-SVD mitigates permutation sensitivity, its effectiveness in IMVC remains dependent on the quality of the sample relationships estimated from incomplete observations.

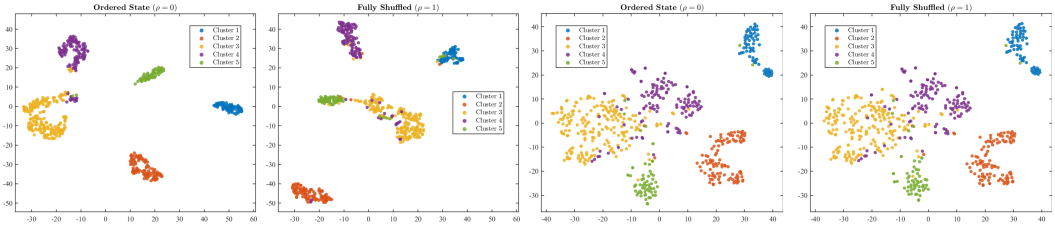
C.3 More Results for Representation-oriented Permutation Equivariance

While the main text reports representation-level results on MSRCv1, we provide additional t-SNE visualizations on NGs (Figure 10), BBCSport (Figure 11), and HW (Figure 12) to further evaluate representation-oriented permutation equivariance. For each dataset, we compare the representations learned from the original order and a fully shuffled order after aligning sample indices back to the same reference order. The normalized discrepancy Δ_F is reported to quantify representation-level consistency.

Overall, the additional visualizations are consistent with the main observations. FFT (sample mode) produces clear representation discrepancies after sample shuffling, with $\Delta_F = 3.55$ on NGs, 3.08 on BBCSport, and 22.94 on HW. K-OR slightly reduces the representation discrepancy on NGs and HW, but not on BBCSport, and substantial discrepancies remain across all three datasets. In contrast, FFT (view mode) and Cor-SVD exhibit near-zero representation discrepancies in these experiments. Combined with the outcome-oriented results, these visualizations further show that Cor-SVD better balances representation-level consistency and clustering performance.

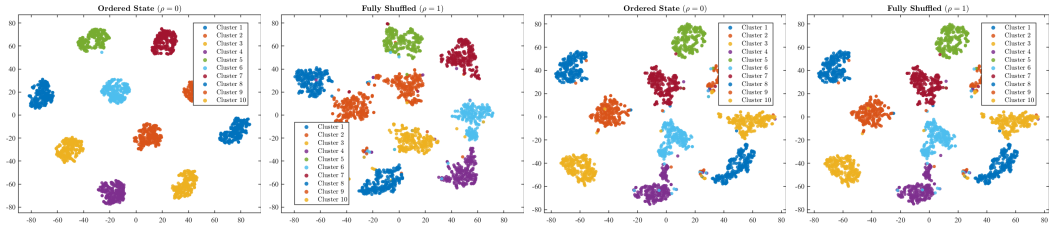


(a) FFT (sample mode) ($\Delta_F = 3.08$) (b) FFT (view mode) ($\Delta_F = 2.44 \times 10^{-13}$)

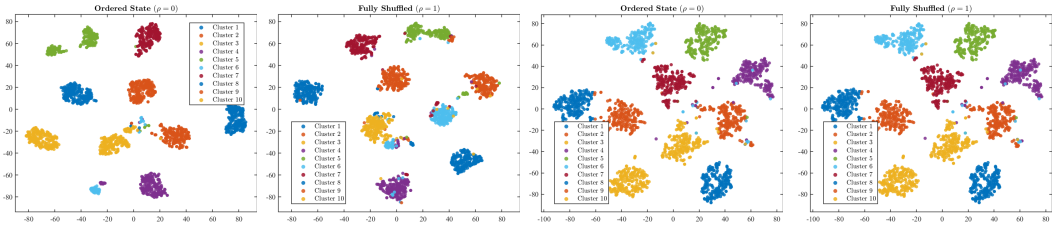


(c) K-OR ($\Delta_F = 3.11$) (d) Cor-SVD ($\Delta_F = 3.35 \times 10^{-13}$)

Figure 11: t-SNE visualization for different strategies with the TLSpNM baseline on BBCSport. Δ_F denotes the normalized Frobenius discrepancy between the two representations after aligning samples back to the same indexing.



(a) FFT (sample mode) ($\Delta_F = 22.94$) (b) FFT (view mode) ($\Delta_F = 1.46 \times 10^{-10}$)



(c) K-OR ($\Delta_F = 22.49$) (d) Cor-SVD ($\Delta_F = 1.21 \times 10^{-10}$)

Figure 12: t-SNE visualization for different strategies with the TLSpNM baseline on the HW dataset. Δ_F denotes the normalized Frobenius discrepancy between the two representations after aligning samples back to the same indexing.

Table 10: ACC (%) under view permutations with a fixed fully shuffled sample order ($\rho = 1$).

Baselines	Variants	MSRCv1	NGs	BBCSport	HW
Tested view orders		5!	3!	2!	6!
TLSpNM	FFT (sample mode)	53.33 (0.00)	67.00 (0.00)	69.85 (0.00)	92.00 (0.00)
	FFT (view mode)	46.76 (3.21)	82.20 (0.00)	76.10 (0.00)	86.12 (4.97)
	Cor-SVD	**84.29 (0.00)**	**98.20 (0.00)**	**86.40 (0.00)**	**93.26 (0.00)**
ESTMC	FFT (sample mode)	80.48 (0.00)	89.80 (0.00)	84.01 (0.00)	85.95 (0.00)
	FFT (view mode)	75.29 (6.01)	86.60 (0.00)	83.82 (0.00)	83.29 (3.71)
	Cor-SVD	**83.75 (0.00)**	**95.80 (0.00)**	**86.40 (0.00)**	**87.60 (0.00)**

D Additional Experiments

D.1 View Permutation Sensitivity

To examine whether moving FFT to the view mode [19] transfers permutation sensitivity from samples to views, we fix one fully shuffled sample order ($\rho = 1$) and report the mean and standard deviation of ACC over all view permutations. As shown in Table 10, FFT (sample mode) and Cor-SVD yield an empirical ACC standard deviation of 0.00 in all evaluated settings, whereas FFT (view mode) exhibits substantial variation on MSRCv1 and HW under both baselines. The absence of variation on NGs and BBCSport is consistent with their small numbers of views: with two or three views, every permutation can be expressed through circular shifts and reversals, which preserve the Fourier-slice singular values up to reordering and remain compatible with identical slice-wise shrinkage. With more than three views, general permutations can alter these singular values and consequently affect low-rank learning, as observed on MSRCv1 and HW. These results demonstrate that moving FFT to the view mode can transfer order dependence rather than eliminate it, so robustness to sample permutations alone does not ensure robustness to view permutations.

E Limitations & Future Work

Despite the promising empirical and theoretical findings, several limitations of this work should be acknowledged.

First, the proposed Cor-SVD framework relies on the estimation of a sample correlation matrix, whose quality may degrade under incomplete multi-view settings. As the missing rate increases, the correlation structure becomes less reliable, which can negatively affect both representation quality and clustering performance. This limitation is intrinsic to data-driven transform learning under insufficient observations.

Second, Cor-SVD requires constructing an $n \times n$ sample correlation matrix and computing its eigenbasis. This introduces additional computational overhead during basis construction and may limit scalability to large datasets. Developing more efficient transform construction strategies while preserving permutation robustness remains an important direction for future work.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: See the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: See Appendix E

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: See main text and Appendices B.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: See experimental section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Code is publicly available at `https://github.com/jijintian/Cor-SVD`.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: See Experimental Setup and Appendix C.1.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: See Experimental section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: See Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We have reviewed the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: Foundational research with no societal impact.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: No such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All code and data with their papers are cited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: This work falls within the field of traditional machine learning and does not involve large language models (LLMs).

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.