Tensorized Label Learning via Balanced Tensor Regression

Guangyu Yang[1], Yuzhuo Feng[1], Qin Li[2*], Quanxue Gao[1†], Ming Yang[3], Rui Wang[1]

[1]School of Telecommunications Engineering, Xidian University, Shaanxi 710071, China
[2]School of Software Engineering, Shenzhen University of Information Technology, Shenzhen 518172, China
[3]School of Mathematical Sciences, Harbin Engineering University, Heilongjiang 150001, China.
24011211262@stu.xidian.edu.cn, 22049200389@stu.xidian.edu.cn, liqin@sziit.edu.cn, qxgao@xidian.edu.cn, yangmingmath@gmail.com, 23011210720@stu.xidian.edu.cn

Abstract

The multi-view clustering methods based on tensor regression can make full use of the potential structural information between views and achieve data-level fusion. However, existing tensor regression-based approaches for anchor graph often overlook the probabilistic nature of anchor graph, focusing solely on sample labels while ignoring the influence of anchor labels on clustering results. To overcome these limitations, we introduce Tensorized Label Learning via Balanced Tensor Regression (TLL-BTR). Our key idea is to exploit the probabilistic nature of the anchor graph by regarding the sample labels as a projection tensor that maps the anchor graph into the label space, thereby producing anchor labels. By enforcing constraints on these anchor labels, we guide the concurrent learning of sample labels and achieve co-label learning between anchors and samples. To prevent trivial solutions, we maximize the nuclear norm to promote an even distribution of samples across clusters. Extensive experiments on benchmark datasets demonstrate that TLL-BTR consistently outperforms state-of-the-art methods.

Introduction

Graph-based clustering methods have demonstrated strong performance by capturing the pairwise relationships among samples in each view(Zhan et al. 2019; Liang et al. 2024; Duan et al. 2024). These methods typically build an $n \times n$ full graph for each view, but incur at least $\mathcal{O}(n^2)$ computational complexity, which limits their scalability to large datasets.

Anchor graph-based methods(Li et al. 2025, 2023b) mitigate this bottleneck by selecting m representative anchors from n samples and construct an $n \times m$ anchor graph, significantly reducing computational cost while preserving clustering performance. However, many existing anchor graph-based approaches still depend on post-processing like k-means or spectral clustering, which affects overall efficiency. To address these limitations, recent works have embedded auxiliary mechanisms directly into the anchor graph framework, including k-connected component methods that detect clusters via graph connectivity (Xia et al. 2023; Yang et al. 2022b); Non-negative matrix factorization (NMF) models that factorize the anchor graph into low-rank, non-negative components (Li et al. 2023a, 2024b); and regression-based strategies that learn a projection from the anchor graph into the label space, thereby obviating external clustering steps (Zhao et al. 2024, 2025; Li et al. 2024a).

Regression-based anchor graph methods achieve competitive clustering performance and reduce reliance on separate post-processing, but they exhibit three main drawbacks: (1) they typically treat the anchor graph purely as a structural representation, overlooking its inherent probabilistic nature—particularly the probabilistic associations between samples and anchors; (2) they focus solely on the sample indicator matrix, overlooking the informative guidance of the anchor indicator matrix; and (3) their Frobenius-norm–based objective can yield trivial or unstable solutions when clusters are imbalanced.

To address the above problems, this paper proposes Tensorized Label Learning via Balanced Tensor Regression (TLL-BTR). This method leverages the probabilistic nature of the anchor graph, interpreting the regression model from a probabilistic perspective. Specifically, a probabilistic projection tensor is learned to project the anchor graph into the label space. The resulting anchor labels in turn facilitate the learning of sample labels, enabling co-clustering of anchors and samples. Given the assumption that anchors and samples share similar distributions, a well-distributed set of anchors naturally leads to a well-distributed sample clustering. The proposed method uses the maximum nuclear norm as the regularization constraint of the anchor label tensor, which can effectively ensure that the anchors are evenly distributed to each cluster, thereby enabling the samples to be distributed evenly. The main contributions of this paper are:

1. **Concept Regression:** We simplify traditional projection-based regression by treating the sample-label tensor itself as the projection variable, eliminating the need for joint optimization of multiple tensors. Anchor labels emerge implicitly from this formulation without requiring explicit optimization.
2. **Co-Label Learning:** By coupling anchor and sample labels via the probabilistic nature of the anchor graph, our method achieves simultaneous co-clustering of anchors and samples, leveraging the mutual reinforcement between the two.

*corresponding author
†corresponding author

3. **Balanced Clustering:** Instead of minimizing the Frobenius norm, we maximize the nuclear norm of the anchor label tensor to enforce an even distribution of anchors (and thus samples) across clusters, effectively preventing trivial solutions.

Related Work

Anchor graph–based multi-view clustering methods typically proceed by selecting m representative anchors from n samples and constructing a anchor graph for subsequent grouping. Early methods, such as LMVSC (Kang et al. 2020) and FPMVS-CAG (Wang et al. 2021b), achieve linear time complexity by leveraging efficient anchor selection and graph construction; however, they still require spectral clustering or similar post-processing to obtain final labels, which adds computational overhead.

To eliminate the need for post-processing, one line of work enforces that the anchor graph consists of exactly k connected components, each corresponding to a cluster. Representative examples include SFMC (Li et al. 2022), which integrates connected-component constraints into the optimization, and TBGL (Xia et al. 2023), which further incorporates tensor structure to capture high-order affinities among anchors.

A second paradigm applies non-negative matrix factorization directly on the anchor graph to learn cluster assignments in one step. EMVC-NTLC (Li et al. 2024b) factorizes the anchor adjacency matrix under non-negativity and orthogonality constraints, while Orth-NTF (Li et al. 2023a) extends this idea by imposing orthogonality on each factor to promote distinct cluster representations.

The third category employs regression models to map the anchor graph into the label space without additional spectral steps. AGFS-OMVC (Zhao et al. 2024) learns a separate projection for each view, enabling direct label inference but at the cost of ignoring dependencies across views. To jointly exploit multi-view structure, LLMTP (Li et al. 2024a) arranges the collection of view-specific anchor graphs into a third-order tensor and performs tensor regression, achieving effective data-level fusion and improved robustness against view heterogeneity.

Although these methods enhance efficiency and integration, they leave open challenges in balancing cluster sizes and fully leveraging both low-rankness and interview correlations within a unified framework. In this work, we address these gaps by introducing a balanced tensor regression approach that simultaneously ensures equitable cluster assignments and captures complementary multi-view information with linear computational complexity.

Notations and Preliminaries

We use uppercase bold letters for matrices (e.g., $\mathbf{W}$), lowercase bold letters for vectors (e.g., $\mathbf{w}$), bold calligraphic letters for third-order tensors (e.g., $\boldsymbol{\mathcal{W}}$), and scalars by regular letters with subscripts (e.g., q, W_{il}). The operator $\overline{\boldsymbol{\mathcal{W}}} = \mathrm{fft}\big(\boldsymbol{\mathcal{W}}, [\,], 3\big)$ computes the Fast Fourier Transform of $\boldsymbol{\mathcal{W}}$ along its third mode.

Definition 1 (t-product (Kilmer and Martin 2011))**.** *Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{n_1\times m\times n_3}$ and $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{m\times n_2\times n_3}$. Their t-product $\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is defined by*

$$\boldsymbol{\mathcal{X}} * \boldsymbol{\mathcal{Y}} = \mathrm{ifft}\big(\mathrm{bdiag}(\overline{\mathbf{X}}\,\overline{\mathbf{Y}}), [\,], 3\big),$$

where $\overline{\mathbf{X}} = \mathrm{bdiag}(\overline{\boldsymbol{\mathcal{X}}})$ and $\overline{\mathbf{Y}} = \mathrm{bdiag}(\overline{\boldsymbol{\mathcal{Y}}})$ form block-diagonal matrices of their frontal slices.

Definition 2 (t-SVD (Kilmer and Martin 2011))**.** *The tensor singular value decomposition of $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is*

$$\boldsymbol{\mathcal{W}} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top,$$

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, and $\mathcal{S} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is f-diagonal. Here, $*$ denotes the t-product.

Definition 3 (Tensor Schatten p-norm (Gao et al. 2021))**.** *For $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, let $m = \min(n_1, n_2)$. Its tensor Schatten p-norm is*

$$\|\boldsymbol{\mathcal{W}}\|_{\circledS p} = \Big(\sum_{i=1}^{n_3} \big\|\overline{\boldsymbol{\mathcal{W}}}^{(i)}\big\|_{\circledS p}^{p}\Big)^{1/p} = \Big(\sum_{i=1}^{n_3}\sum_{j=1}^{m} \sigma_j\big(\overline{\boldsymbol{\mathcal{W}}}^{(i)}\big)^{p}\Big)^{1/p} \tag{1}$$

where $\sigma_j(\overline{\boldsymbol{\mathcal{W}}}^{(i)})$ is the j-th singular value of the i-th frontal slice $\overline{\boldsymbol{\mathcal{W}}}^{(i)}$ and $0 < p \le 1$.

Methodology

Motivation and Proposed Model

Anchor graphs avoid the need to construct a full similarity graph, making them well suited for large-scale datasets. It is therefore natural to extend anchor-graph techniques to regression. In classical (matrix) regression, one seeks a projection matrix that maps the anchor graph into the label space to obtain anchor labels (Zhao et al. 2024, 2025). To exploit structural information across multiple views, LLMTP (Li et al. 2024a) generalizes this idea to tensor regression. Given the tensorized anchor graph $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n\times m\times V}$, the projection tensor $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n\times c\times V}$, and the anchor-label tensor $\boldsymbol{\mathcal{Q}} \in \mathbb{R}^{m\times c\times V}$, the tensor regression objective is:

$$\min_{\boldsymbol{\mathcal{W}},\,\boldsymbol{\mathcal{Q}}} \ \big\|\boldsymbol{\mathcal{B}} * \boldsymbol{\mathcal{Q}} - \boldsymbol{\mathcal{W}}\big\|_F^2 \ + \ \lambda\,\big\|\boldsymbol{\mathcal{W}}\big\|_{\circledS p,r}^{p}, \tag{2}$$

where $\|\boldsymbol{\mathcal{W}}\|_{\circledS p,r}^{p} = \|\boldsymbol{\mathcal{W}}_\circ\|_{\circledS p}^{p}$, and $\boldsymbol{\mathcal{W}}_\circ \in \mathbb{R}^{n\times V\times c}$ denotes the rotated version of $\boldsymbol{\mathcal{W}}$ (see Fig. 2). Here n is the number of samples, m is the number of anchors, V the number of views, and λ the regularization weight for the tensor Schatten-p term.

Existing methods overlook the fact that the anchor graph itself induces a probability distribution encoding associations between anchors and samples. Since anchors and samples share similar distributions, anchor labels should be equally informative for clustering—an insight we incorporate via probability-induced concept regression.

Theorem 1. *Given an anchor-graph matrix $\boldsymbol{B} \in \mathbb{R}^{n\times m}$ with nonnegative entries and row sums equal to one, and a sample-label matrix $\boldsymbol{W} \in \mathbb{R}^{n\times c}$, the corresponding anchor-label matrix*

$$\boldsymbol{Q} = \boldsymbol{B}^\top \boldsymbol{W}. \tag{3}$$

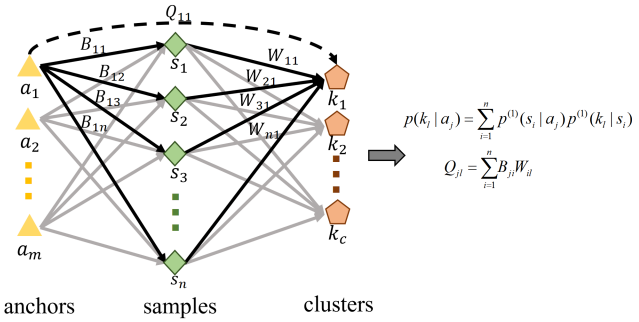


Figure 1: The process of probability transition.

Proof. Under the stationary Markov-random-walk framework of (Liu, He, and Chang 2010), the one-step transition probability from the i-th sample to the j-th anchor is:

$$p^{(1)}(a_j|s_i) = \frac{B_{ij}}{\sum_{j'}^{m} B_{ij'}} = B_{ij}. \tag{4}$$

Let $W_{il} = p^{(1)}(k_l \mid s_i)$ and $Q_{jl} = p(k_l \mid a_j)$. Assuming independence between steps,

$$Q_{jl} = p(k_l|a_j) = \sum_{i=1}^{n} p^{(1)}(s_i|a_j)p^{(1)}(k_l|s_i) = \sum_{i=1}^{n} B_{ji}W_{il} \tag{5}$$

In matrix form, this yields

$$\boldsymbol{Q} = \boldsymbol{B}^\top \boldsymbol{W}. \tag{6}$$

□

Figure 1 illustrates that $\boldsymbol{B}$ encodes connections between n samples and m anchors, and the transition to clusters follows a first-order Markov process determined solely by the current state.

By combining Theorem 1 with (2), we arrive at the probability-induced concept regression:

$$f\left(\boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}\right). \tag{7}$$

Theorem 2. *Suppose $m_1 + \cdots + m_c = m$, where m_j is the number of anchors assigned to cluster j. If each $\overline{\boldsymbol{\mathcal{Q}}}^{(v)}$ is a discrete matrix, and satisfies $m_j = m/c$, then the balanced allocation*

$$\max_{\overline{\boldsymbol{\mathcal{Q}}}^{(v)} \geq 0,\ \overline{\boldsymbol{\mathcal{Q}}}^{(v)} \cdot \mathbf{1} = \mathbf{1}} \|\boldsymbol{\mathcal{Q}}\|_{\circledast} \tag{8}$$

achieves its optimum under these conditions.

The proof of the Theorem 2 is in the appendix.

To prevent trivial solutions and encourage informative cluster assignments, we leverage Theorem 2 to maximize the nuclear norm. Since the anchor distribution should mirror that of the data points, we introduce a regularization on the anchor labels. This yields the final optimization problem:

$$\begin{aligned} &\max_{\boldsymbol{\mathcal{W}}} \left\|\boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}\right\|_{\circledast} - \lambda \left\|\boldsymbol{\mathcal{W}}\right\|_{\circledS_p,r}^p, \\ &\text{s.t.} \quad \boldsymbol{W}^{(v)} \geq 0, \quad \boldsymbol{W}^{(v)}\mathbf{1} = \mathbf{1}, \end{aligned} \tag{9}$$

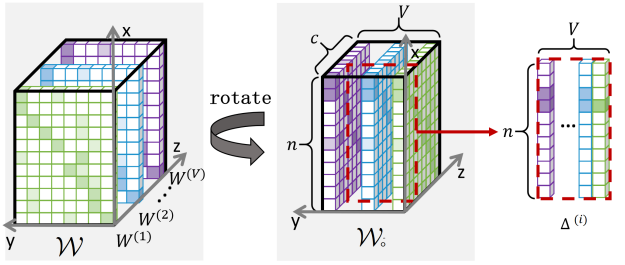


Figure 2: The process of constructing the rotated 3rd-order tensor. $\boldsymbol{\mathcal{W}} \in \mathbb{R}^{n\times c\times V}$, $\boldsymbol{\mathcal{W}}_\circ \in \mathbb{R}^{n\times V\times c}$ is the rotated form of $\boldsymbol{\mathcal{W}}$.

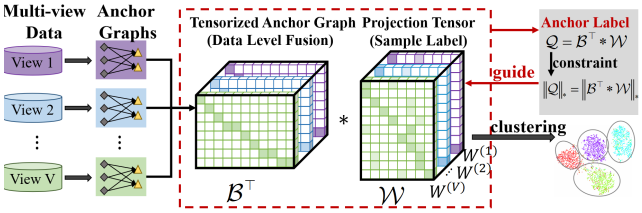


Figure 3: The framework of TLL-BTR.

where $\boldsymbol{W}^{(v)}$ is the v-th frontal slice of the tensor $\boldsymbol{\mathcal{W}}$. Anchor selection and graph construction details are given in (Xia et al. 2023).

Maximizing $\|\boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}\|_{\circledast}$ enforces balanced anchor assignments across clusters, avoiding trivial solutions. Only when anchors are evenly spread across different clusters can they better characterize the data distribution. At this point, we can learn accurate sample labels from the anchor graph.

Note that we do not optimize the anchor labels $\boldsymbol{\mathcal{Q}}$ directly. Instead, the nuclear-norm regularizer $\|\boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}\|_{\circledast}$ implicitly aligns $\boldsymbol{\mathcal{Q}} = \boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}$ with the sample labels $\boldsymbol{\mathcal{W}}$, achieving co-label learning while explicitly optimizing only one variable.

Furthermore, the tensor Schatten p-norm term $\|\boldsymbol{\mathcal{W}}\|_{\circledS_p,r}^p$ captures complementary information across different views. The overall algorithmic framework is depicted in Figure 3.

Remark 1. *For the rotated tensor $\boldsymbol{\mathcal{W}}_\circ$ in Fig. 2, each lateral slice $\Delta^{(i)}$ encodes the relationships between the n samples and the i-th cluster across v views. Multi-view clustering seeks to align the sample-cluster affinities $\mathbf{W}_{:,i}^{(1)}, \ldots, \mathbf{W}_{:,i}^{(v)}$, yet view-specific structures often diverge. By applying the tensor Schatten p-norm, we enforce a low-rank structure on each $\Delta^{(i)}$, integrating complementary view information and promoting consistency in the cluster indicators.*

Optimization

To solve (9), we employ the Augmented Lagrangian method. Introducing an auxiliary tensor $\boldsymbol{\mathcal{J}}$ and enforcing $\boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{W}}$, we obtain the following problem:

$$\begin{aligned} &\max_{\boldsymbol{\mathcal{W}},\boldsymbol{\mathcal{J}}} \left\|\boldsymbol{\mathcal{B}}^\top * \boldsymbol{\mathcal{W}}\right\|_{\circledast} - \lambda \left\|\boldsymbol{\mathcal{J}}\right\|_{\circledS_p,r}^p - \frac{\mu}{2}\left\|\boldsymbol{\mathcal{W}} - \boldsymbol{\mathcal{J}} + \tfrac{1}{\mu}\boldsymbol{\mathcal{Y}}\right\|_F^2, \\ &\text{s.t.} \quad \boldsymbol{W}^{(v)} \geq 0, \quad \boldsymbol{W}^{(v)}\mathbf{1} = \mathbf{1} \quad (v = 1, \ldots, V), \end{aligned} \tag{10}$$

where $\mu > 0$ is the penalty parameter and $\mathcal{Y}$ denotes the Lagrange multiplier.

Since (10) is non-convex in all variables jointly, we apply the Alternating Direction Method of Multipliers (ADMM) to split it into subproblems.

• **Update $\mathcal{J}$**: When $\mathcal{W}$ and $\mathcal{Y}$ are fixed, the subproblem for $\mathcal{J}$ is:

$$\max_{\mathcal{J}} \ -\lambda_1\|\mathcal{J}\|^p_{\circledast,r} - \frac{\mu}{2}\left\|\mathcal{W} - \mathcal{J} + \frac{\mathcal{Y}}{\mu}\right\|_F^2. \tag{11}$$

which is equivalent to:

$$\min_{\mathcal{J}} \frac{1}{2}\left\|\mathcal{J}_\circ - \left(\mathcal{W}_\circ + \frac{\mathcal{Y}_\circ}{\mu}\right)\right\|_F^2 + \frac{\lambda_1}{\mu}\|\mathcal{J}_\circ\|^p_{\circledast}. \tag{12}$$

where $\mathcal{J}_\circ \in \mathbb{R}^{n\times V\times c}$, $\mathcal{W}_\circ \in \mathbb{R}^{n\times V\times c}$ and $\mathcal{Y}_\circ \in \mathbb{R}^{n\times V\times c}$ are the rotated forms of $\mathcal{J}$, $\mathcal{W}$ and $\mathcal{Y}$ respectively.

Theorem 3. *(Gao et al. 2020) Given $\mathcal{A} \in \mathbb{R}^{n_1\times n_2\times n_3}$, it has t-SVD as $\mathcal{A} = \mathcal{U} * \mathcal{S} * \mathcal{V}^\top$. The solution of problem*

$$\arg\min_{\mathcal{X}} \ \frac{1}{2}\|\mathcal{X} - \mathcal{A}\|_F^2 + \tau\|\mathcal{X}\|^p_{\circledast}. \tag{13}$$

is given by:

$$\mathcal{X}^* = \Gamma_\tau(\mathcal{A}) = \mathcal{U} * \mathit{ifft}\left(P_\tau(\overline{\mathcal{A}})\right) * \mathcal{V}^\top, \tag{14}$$

where $P_\tau(\overline{\mathcal{A}})$ is the third-order diagonal tensor, whose diagonal elements are computed by using the Generalized Soft-Thresholding (GST) algorithm (Gao et al. 2021).

By Theorem 3, the solution of Eq. (12) can be written as:

$$\mathcal{J}^* = \Gamma_{\frac{\lambda}{\mu}}\left(\mathcal{W} + \frac{\mathcal{Y}}{\mu}\right). \tag{15}$$

• **Update $\mathcal{W}$**: Holding all other variables fixed, $\mathcal{W}$ is updated by solving:

$$\begin{aligned} &\max \left\|\mathcal{B}^\top * \mathcal{W}\right\|_{\circledast} - \frac{\mu}{2}\left\|\mathcal{W} - \mathcal{J} + \frac{\mathcal{Y}}{\mu}\right\|_F^2, \\ &s.t.\ \boldsymbol{W}^{(v)} \geq 0,\ \boldsymbol{W}^{(v)}\cdot\mathbf{1} = \mathbf{1}. \end{aligned} \tag{16}$$

Applying the discrete Fourier transform along the third mode converts (16) into

$$\begin{aligned} &\max \sum_{v=1}^{V}\left\|(\overline{\boldsymbol{B}}^{(v)})^\top\overline{\boldsymbol{W}}^{(v)}\right\|_* - \frac{\mu}{2}\|\overline{\boldsymbol{W}}^{(v)} - \overline{\boldsymbol{J}}^{(v)} + \frac{\overline{\boldsymbol{Y}}^{(v)}}{\mu}\|_F^2, \\ &s.t.\ \boldsymbol{W}^{(v)} \geq 0,\ \boldsymbol{W}^{(v)}\cdot\mathbf{1} = \mathbf{1}. \end{aligned} \tag{17}$$

where $\|\cdot\|_*$ is the nuclear norm of the matrix.

Let $\overline{\boldsymbol{Q}}^{(v)} = (\overline{\boldsymbol{B}}^{(v)})^\top\overline{\boldsymbol{W}}^{(v)}$, we have $f(\overline{\boldsymbol{Q}}^{(v)}) = \|\overline{\boldsymbol{Q}}^{(v)}\|_* = \|(\overline{\boldsymbol{B}}^{(v)})^\top\overline{\boldsymbol{W}}^{(v)}\|_*$. To simplify the calculation, we use the first-order Taylor expansion to convert the nuclear norm term $f(\overline{\boldsymbol{Q}}^{(v)})$ into the trace form. Specifically, we obtain the following first-order Taylor expansion at the i-th iteration:

$$f(\overline{\boldsymbol{Q}}^{(v)}) = f(\overline{\boldsymbol{Q}}_i^{(v)}) + \langle\nabla f(\overline{\boldsymbol{Q}}_i^{(v)}), \overline{\boldsymbol{Q}}^{(v)} - \overline{\boldsymbol{Q}}_i^{(v)}\rangle, \tag{18}$$

where $\nabla f(\overline{\boldsymbol{Q}}_i^{(v)})$ is the gradient of the $f(\overline{\boldsymbol{Q}}_i^{(v)})$.

Discarding the constant term in Eq. (18) gives:

$$f(\overline{\boldsymbol{Q}}^{(v)}) = \langle\nabla f(\overline{\boldsymbol{Q}}_i^{(v)}), \overline{\boldsymbol{Q}}^{(v)}\rangle = Tr(\nabla f(\overline{\boldsymbol{Q}}_i^{(v)}), \overline{\boldsymbol{Q}}^{(v)}) \tag{19}$$

Substituting (19) to (17) resulting in:

$$\begin{aligned} &\max \sum_{v=1}^{V} Tr\left[(\overline{\boldsymbol{K}}^{(v)})^\top\overline{\boldsymbol{Q}}^{(v)}\right] - \frac{\mu}{2}\|\overline{\boldsymbol{W}}^{(v)} - \overline{\boldsymbol{J}}^{(v)} + \frac{\overline{\boldsymbol{Y}}^{(v)}}{\mu}\|_F^2, \\ &s.t.\ \boldsymbol{W}^{(v)} \geq 0,\ \boldsymbol{W}^{(v)}\cdot\mathbf{1} = \mathbf{1}. \end{aligned} \tag{20}$$

where $\overline{\boldsymbol{K}}^{(v)} = \nabla f(\overline{\boldsymbol{Q}}_i^{(v)})$.

Theorem 4. *(Zhen et al. 2018) Given the matrix $\boldsymbol{E}$, and define $f(\boldsymbol{E}) = \|\boldsymbol{E}\|_*$, the derivative of $\|\boldsymbol{E}\|_*$ can be derived as:*

$$\nabla f(\boldsymbol{E}) = \frac{\partial\|\boldsymbol{E}\|_*}{\partial\boldsymbol{E}} = \boldsymbol{\Lambda}\boldsymbol{\Sigma}^{-1}|\boldsymbol{\Sigma}|\boldsymbol{V}^\top, \tag{21}$$

where $\boldsymbol{\Lambda}\boldsymbol{\Sigma}\boldsymbol{V}^\top = svd(\boldsymbol{E})$. $\boldsymbol{\Sigma}^{-1}$ denotes the Moore-Penrose pseudo-inverse of $\boldsymbol{\Sigma}$.

By Theorem 4, the update for the frequency-domain tensor slice $\overline{\boldsymbol{K}}^{(v)}$ in Eq. (20) is:

$$\overline{\boldsymbol{K}}^{(v)} = \overline{\boldsymbol{\Lambda}}^{(v)}(\overline{\boldsymbol{\Sigma}}^{(v)})^{-1}|\overline{\boldsymbol{\Sigma}}^{(v)}|(\overline{\boldsymbol{V}}^{(v)})^\top, \tag{22}$$

where $\overline{\boldsymbol{\Lambda}}^{(v)}\overline{\boldsymbol{\Sigma}}^{(v)}(\overline{\boldsymbol{V}}^{(v)})^\top = svd(\overline{\boldsymbol{Q}}^{(v)}) = svd[(\overline{\boldsymbol{B}}^{(v)})^\top\overline{\boldsymbol{W}}^{(v)}]$.

Eq. (20) can be simplified to:

$$\begin{aligned} &\min \sum_{v=1}^{V}\frac{\mu}{2}\left\|\overline{\boldsymbol{W}}^{(v)} - \left(\overline{\boldsymbol{J}}^{(v)} - \frac{1}{\mu}\overline{\boldsymbol{Y}}^{(v)} + \frac{1}{\mu}\overline{\boldsymbol{B}}^{(v)}\overline{\boldsymbol{K}}^{(v)}\right)\right\|_F^2, \\ &\text{s.t. } \boldsymbol{W}^{(v)} \geq 0,\ \boldsymbol{W}^{(v)}\cdot\mathbf{1} = \mathbf{1}. \end{aligned} \tag{23}$$

Convert Eq. (23) to the time domain through inverse Fourier transform and obtain:

$$\begin{aligned} &\min \frac{\mu}{2}\|\mathcal{W} - \mathcal{C}\|_F^2, \\ &\text{s.t. } \boldsymbol{W}^{(v)} \geq 0,\ \boldsymbol{W}^{(v)}\cdot\mathbf{1} = \mathbf{1}. \end{aligned} \tag{24}$$

where $\overline{\boldsymbol{C}}^{(v)} = \overline{\boldsymbol{J}}^{(v)} - \frac{1}{\mu}\overline{\boldsymbol{Y}}^{(v)} + \frac{1}{\mu}\overline{\boldsymbol{B}}^{(v)}\overline{\boldsymbol{K}}^{(v)}$.

Following (Nie et al. 2016), the solution of Eq. (24) is $\boldsymbol{w}^{(v)} = (\boldsymbol{c}^{(v)} + \gamma\mathbf{1})$, where $\boldsymbol{w}^{(v)}$ denotes pulling the matrix $\boldsymbol{W}^{(v)}$ into a vector, $\boldsymbol{c}^{(v)}$ denotes pulling the matrix $\boldsymbol{C}^{(v)}$ into a vector, and γ is a Lagrange multiplier.

The complete optimization procedure is summarized in Algorithm 1.

Convergence Analysis Sketch

We briefly outline why the ADMM iterations for (10) converge to a stationary (KKT) point. Let $\mathcal{L}_\mu(\mathcal{W},\mathcal{J},\mathcal{Y}) = -\|\mathcal{B}^\top * \mathcal{W}\|_{\circledast} + \lambda\|\mathcal{J}\|^p_{\circledast,r} + \langle\mathcal{Y}, \mathcal{W} - \mathcal{J}\rangle + \frac{\mu}{2}\|\mathcal{W} - \mathcal{J}\|_F^2$. Denote the iterate at step i by $(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_i)$.

Because each block subproblem (for $\mathcal{J}$ in (12) and for $\mathcal{W}$ in (24)) is solved exactly in closed form, there exist constants $\alpha, \beta > 0$ such that

Algorithm 1: Tensorized Label Learning via Balanced Tensor Regression

Input: Initial anchor graph $\boldsymbol{B}^{(v)}$
Parameter: Anchor rate, p, λ
Output: Clustering label

Initialize: Let $\boldsymbol{W}^{(v)} = \boldsymbol{I}$, $\boldsymbol{\mathcal{J}} = \boldsymbol{\mathcal{Y}} = 0$, $\mu = 10^{-5}$, $\mu_{\max} = 10^{12}$, $\eta = 1.1$.
while not converged **do**
  Update $\overline{\boldsymbol{\mathcal{K}}}^{(v)}$ by Eq. (22).
  Update $\boldsymbol{\mathcal{W}}$ by Eq. (24).
  Update $\boldsymbol{\mathcal{J}}$ by Eq. (15).
  Update $\boldsymbol{\mathcal{Y}}, \mu$: $\boldsymbol{\mathcal{Y}} = \boldsymbol{\mathcal{Y}} + \mu(\boldsymbol{\mathcal{W}} - \boldsymbol{\mathcal{J}})$, $\mu = \min(\eta\mu, \mu_{\max})$.
end while
Compute the shared indicator matrix $\boldsymbol{W}$ by $\boldsymbol{W} = \frac{\sum_{v=1}^{V} \boldsymbol{W}^{(v)}}{V}$ and obtain the clustering label by maximizing $\boldsymbol{W}$ by rows.
return Clustering label

$\mathcal{L}_{\mu_i}(\mathcal{W}_{i-1}, \mathcal{J}_{i-1}, \mathcal{Y}_{i-1}) - \mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_{i-1}, \mathcal{Y}_{i-1}) \geq \alpha \|\mathcal{W}_i - \mathcal{W}_{i-1}\|_F^2, \mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_{i-1}, \mathcal{Y}_{i-1}) - \mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_{i-1}) \geq \beta \|\mathcal{J}_i - \mathcal{J}_{i-1}\|_F^2$. The dual update $\mathcal{Y}_i = \mathcal{Y}_{i-1} + \mu_i(\mathcal{W}_i - \mathcal{J}_i)$ modifies $\mathcal{L}$ by

$$\mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_i) - \mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_{i-1}) = \tfrac{1}{\mu_i} \|\mathcal{Y}_i - \mathcal{Y}_{i-1}\|_F^2.$$

Combining these shows that $\{\mathcal{L}_{\mu_i}(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_i)\}$ is non-increasing and bounded below.

Summing the above decreases inequalities over i yields

$$\sum_{i=1}^{\infty} \|\mathcal{W}_i - \mathcal{W}_{i-1}\|_F^2 + \|\mathcal{J}_i - \mathcal{J}_{i-1}\|_F^2 < \infty,$$

so both sequences $\{\mathcal{W}_i\}$ and $\{\mathcal{J}_i\}$ are Cauchy and converge to limits $\mathcal{W}^*, \mathcal{J}^*$. The multiplier updates satisfy $\|\mathcal{W}_i - \mathcal{J}_i\| \to 0$, enforcing primal feasibility.

At the limit $(\mathcal{W}^*, \mathcal{J}^*, \mathcal{Y}^*)$, first-order optimality of each block gives

$$0 \in -\partial_{\mathcal{W}} \|\mathcal{B}^\top * \mathcal{W}^*\|_{\circledast} + \mathcal{Y}^*, \quad 0 \in \partial_{\mathcal{J}} \lambda \|\mathcal{J}^*\|_{\mathfrak{S}p,r}^p - \mathcal{Y}^*,$$

together with $\mathcal{W}^* = \mathcal{J}^*$. These coincide with the KKT system of (10).

Since $\mathcal{L}_\mu$ is a semi-algebraic function, it satisfies the Kurdyka–Łojasiewicz property. Standard arguments then imply that the entire sequence $\{(\mathcal{W}_i, \mathcal{J}_i, \mathcal{Y}_i)\}$ converges to the single stationary point $(\mathcal{W}^*, \mathcal{J}^*, \mathcal{Y}^*)$.

Complexity Analysis

The overall computational complexity of our method decomposes into two primary stages: (1) constructing the initial anchor graph $\boldsymbol{B}^{(v)}$ and (2) solving the optimization problem in (9).

In stage (1), building each view's anchor graph requires forming pairwise similarities and sorting, yielding $\mathcal{O}\big(V\,n\,m\,d + V\,n\,m\,\log m\big)$, where V is the number of views, n the number of samples, m the number of anchors per view, and d the total feature dimension across all views.

Datasets	n	V	c	$d_v (v = 1, ..., V)$
MSRC	210	5	7	24,576,512,256,254
ORL	400	3	40	4096,3304,6750
HW4	2000	4	10	76,216,47,6
Scene15	4485	3	15	1800,1180,1240
NUS	30000	5	31	64,225,144,73,128
AWA	30475	6	50	2688,2000,252,2000,2000,2000

Table 1: Multi-view datasets used in the experiments. n, V and c are the numbers of samples, views and classes respectively, and d_v is the feature dimension in the v-th view.

Stage (2) alternately updates the core variables, including $\boldsymbol{\mathcal{W}}$ and $\boldsymbol{\mathcal{J}}$. The per-iteration costs are $\mathcal{O}\big(n\,m\,c\,V\big)$ and $\mathcal{O}\big(2\,n\,c\,V\log(cV) + n\,c\,V^2\big)$, respectively, where c denotes the number of clusters. Noting that $c, V \ll n$, the dominant term scales as $\mathcal{O}(n)$, so the algorithm is effectively linear in the sample size.

Experiments

Experimental Setup

All experiments were performed on a Windows10 server equipped with dual IntelXeonGold6230 CPUs (2.10GHz) and 128GB of RAM, using MATLAB R2020b.

We tune three hyperparameters: the anchor rate, the Schatten-p exponent p, and the regularization weight λ. A two-stage grid search is employed: first, we explore a coarse range to locate promising regions; then, we refine each parameter within a narrower interval. Specifically, On small- to medium-scale datasets (MSRC, ORL, HW4, Scene15), the anchor rate is varied over $\{0.1, 0.2, \ldots, 1.0\}$. On large-scale datasets (NUS, AWA), the anchor rate is swept over $\{0.001,\ 0.005,\ 0.01,\ 0.05,\ 0.1\}$. The exponent p of the tensor Schatten-p norm is chosen from $\{0.1, 0.2, \ldots, 1.0\}$. The regularization parameter λ is drawn from $\{5,\ 10,\ 50,\ 100,\ 500,\ 1000\}$.

When adjusting λ, we first ensure that each term in the objective function has comparable magnitude, and then fine-tune λ within the selected range.

Datasets and Compared Methods

We conducted experiments on seven widely used benchmark datasets: MSRC (Winn and Jojic 2005), ORL (Samaria and Harter 1994), HW4 (van Breukelen et al. 1998), Scene15 (Oliva and Torralba 2001), NUS (Chua et al. 2009), and AWA (Lampert, Nickisch, and Harmeling 2009). Their key statistics are summarized in Table 1. To evaluate clustering performance, we adopt three standard metrics: Accuracy (ACC), Normalized Mutual Information (NMI), and Purity.

We compare our method against a range of state-of-the-art algorithms, including MultiNMF (Liu et al. 2013), NMF-CC (Liang et al. 2020), LMVSC (Kang et al. 2020), PLCMF (Wang et al. 2021a), FPMVS-CAG (Wang et al. 2021b), ERMC-AGR (Yang et al. 2022a), SFMC (Li et al. 2022), AWMVC (Wan et al. 2023), FSMSC (Chen et al. 2023), Orth-NTF (Li et al. 2023a),

Datasets / Methods	MSRC ACC	MSRC NMI	MSRC Purity	ORL ACC	ORL NMI	ORL Purity	HW4 ACC	HW4 NMI	HW4 Purity	Scene15 ACC	Scene15 NMI	Scene15 Purity
MultiNMF	0.757	0.662	0.790	0.710	0.854	0.803	0.803	0.678	0.803	0.369	0.380	0.396
NMF-CC	0.577	0.469	0.606	0.793	0.901	0.817	0.624	0.567	0.587	0.368	0.364	0.390
LMVSC	0.814	0.717	0.814	0.653	0.806	0.710	0.904	0.831	0.904	0.355	0.331	0.399
PLCMF	0.634	0.567	0.660	0.546	0.751	0.587	0.679	0.648	0.696	0.480	0.476	0.520
FPMVS-CAG	0.786	0.686	0.786	0.785	0.893	0.810	0.744	0.753	0.744	0.463	0.486	0.481
ERMC-AGR	0.440	0.345	0.449	0.346	0.535	0.371	0.385	0.370	0.386	0.218	0.166	0.221
SFMC	0.810	0.721	0.810	0.760	0.914	0.815	0.853	0.871	0.873	0.188	0.135	0.202
AWMVC	0.843	0.727	0.843	0.788	0.897	0.808	0.876	0.790	0.876	0.613	0.577	0.653
FSMSC	0.876	0.753	0.876	0.770	0.884	0.798	0.878	0.787	0.878	0.440	0.414	0.450
Orth-NTF	0.990	0.978	0.990	0.785	0.879	0.798	0.985	0.969	0.985	0.758	0.804	0.759
LMVC-MFCG	0.852	0.738	0.852	0.653	0.815	0.688	0.856	0.832	0.856	0.505	0.566	0.522
LLMTP	0.986	0.971	0.986	0.755	0.855	0.772	0.963	0.937	0.963	0.800	0.835	0.808
EMVC-NTLC	0.857	0.811	0.857	0.690	0.821	0.708	0.767	0.774	0.774	0.674	0.724	0.680
AGFS-OMVC	1.000	1.000	1.000	0.998	0.998	0.998	0.995	0.987	0.995	0.796	0.834	0.806
Ours	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.996**	**0.989**	**0.996**	**0.977**	**0.962**	**0.977**

Table 2: Experimental performance of the compared multi-view clustering methods on MSRC, ORL, HW4, and Scene15.

Datasets / Methods	NUS ACC	NUS NMI	NUS Purity	AWA ACC	AWA NMI	AWA Purity
MultiNMF	0.113	0.092	0.191	0.069	0.087	0.092
NMF-CC	0.107	0.073	0.180	0.079	0.078	0.092
LMVSC	0.133	0.110	0.167	0.075	0.086	0.089
PLCMF	0.120	0.098	0.208	0.082	0.085	0.097
FPMVS-CAG	0.174	0.128	0.210	0.091	0.109	0.096
ERMC-AGR	0.133	0.040	0.155	0.050	0.022	0.051
SFMC	0.179	0.191	0.255	0.039	0.003	0.040
AWMVC	0.126	0.122	0.229	0.094	0.113	0.115
FSMSC	0.143	0.131	0.237	0.097	0.113	0.118
Orth-NTF	0.355	0.544	0.465	0.109	0.115	0.134
LMVC-MFCG	0.183	0.135	0.219	0.097	0.118	0.104
LLMTP	0.574	0.719	0.681	0.250	0.340	0.266
EMVC-NTLC	0.453	0.519	0.483	0.234	0.375	0.236
AGFS-OMVC	0.300	0.467	0.390	0.240	**0.520**	0.278
Ours	**0.633**	**0.762**	**0.702**	**0.315**	0.460	**0.315**

Table 3: Experimental performance of the compared multi-view clustering methods on NUS and AWA.

LMVC-MFCG (Yang, Tan, and Yang 2024), LLMTP (Li et al. 2024a), EMVC-NTLC (Li et al. 2024b), and AGFS-OMVC (Zhao et al. 2024).

Experimental Results and Analysis

Table 2 and Table 3 reports the clustering accuracy of our proposed method versus several state-of-the-art baselines on six benchmark datasets. Optimal results are in bold and sub-optimal results are underlined.

Compared to anchor graph–based algorithms (LMVSC, FPMVS-CAG, ERMC-AGR, SFMC, Orth-NTF, EMVC-NTLC and AGFS-OMVC)—in particular AGFS-OMVC, which also employs a regression framework—our approach consistently achieves higher clustering scores. This gain arises because we model the data via third-order anchor graph tensors, which capture inter-view correlations and perform fusion directly at the data level.

In comparison with another tensor-based method, Orth-NTF, our model also demonstrates better performance. A possible explanation is that our framework explicitly considers the influence of anchor labels on clustering, where the collaborative label learning between anchors and samples contributes to more accurate cluster assignments.

Parameter Sensitivity Analysis

Our algorithm depends on three principal hyperparameters: the anchor rate, the penalty weight λ, and the exponent p in the Schatten–p norm.

Figure 4 illustrates the effect of the anchor rate on clustering performance. The best results occur at an anchor rate of 0.4 for HW4 and 0.9 for Scene15, indicating that increasing the anchor rate beyond an optimal point does not necessarily yield better clusters. Moreover, a higher anchor rate increases memory usage and computational cost, as evidenced by the approximately linear growth in running time shown in Fig. 5.

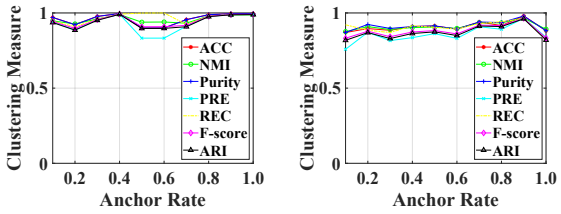
Figure 6 shows how λ and p influence clustering accuracy. When $p = 0.3$, both MSRC and HW4 attain their highest scores, suggesting $p \approx 0.3$ is a robust choice. For the penalty parameter, MSRC performs best at $\lambda = 100$ and HW4 at $\lambda = 10$. In practice, we first choose λ to balance the magnitudes of competing terms in the objective, and then fine-tune it within that range.

Convergence Analysis

We consider the algorithm to have converged when the reconstruction error $\|\mathcal{W} - \mathcal{J}\|_\infty < 10^{-3}$. As illustrated in Fig. 7, the error decreases sharply during the early iterations and then levels off at approximately the 100th iteration. This plateau indicates that the objective function has reached a (local) minimum, confirming the convergence of our method.

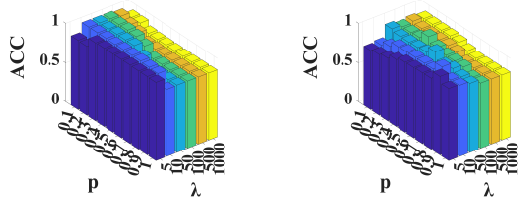
	Situations			MSRC			Scene15		
	concept	tensor-based	$\|\mathcal{W}\|^p_{\circledS p},r$	ACC	NMI	Purity	ACC	NMI	Purity
①	✔	✘	✘	0.733	0.582	0.733	0.324	0.322	0.381
②	✔	✘	✔	1.000	1.000	1.000	0.889	0.888	0.889
③	✔	✔	✘	0.857	0.748	0.857	0.509	0.465	0.548
④	✘	✔	✔	0.986	0.971	0.986	0.800	0.835	0.808
⑤	✔	✔	✔	1.000	1.000	1.000	0.977	0.962	0.977

Table 4: Ablation experiments on MSRC and Scene15



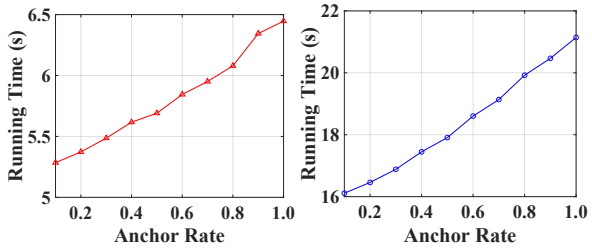
(a) HW4 (b) Scene15

Figure 4: Clustering performance with different anchor rate on HW4 and Scene15.



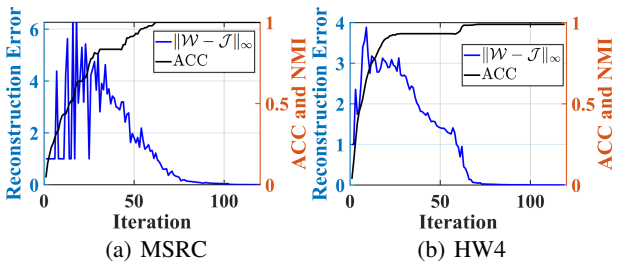
(a) MSRC (b) HW4

Figure 6: Clustering performance with different λ and p on MSRC and HW4.



(a) HW4 (b) Scene15

Figure 5: Running time(second) with different anchor rate on HW4 and Scene15.



(a) MSRC (b) HW4

Figure 7: Convergence experiments on MSRC and HW4.

Ablation Experiments

We perform controlled ablation studies on the MSRC and Scene15 datasets to assess the contributions of three core components: (1) traditional regression versus concept regression; (2) matrix regression versus tensor regression; and (3) Frobenius-norm regularization versus tensor Schatten p-norm regularization. The results are summarized in Table 4.

By comparing configurations ④ and ⑤, we find that concept regression yields superior clustering results. This gain arises because the probabilistic anchor-graph information is projected into the label space to generate anchor labels, which then guide sample-label learning and enhance the discriminative power.

Comparisons between ① versus ③ and ② versus ⑤ highlight the advantage of tensor regression over matrix regression. The tensor formulation captures inter-view structural correlations, enabling richer data-level fusion and improved clustering accuracy.

Finally, examining ① versus ② and ③ versus ⑤ demonstrates that the tensor Schatten p-norm regularizer outperforms the Frobenius norm. The Schatten p-norm more effectively leverages complementary information across views, resulting in more robust clustering.

Conclusion

In this work, we introduce a novel co-label learning framework, termed Tensorized Label Learning via Balanced Tensor Regression (TLL-BTR). By exploiting the probabilistic nature of the anchor graph, our method formulates a unified regression model that jointly infers both anchor and sample labels. A nuclear-norm maximization term is incorporated to promote balanced cluster assignments. Extensive experiments on multiple benchmark datasets demonstrate that TLL-BTR consistently outperforms existing state-of-the-art approaches.

Acknowledgments

This work was supported by the National Natural Science Foundation of China, Grant No.62176203, 62576263; the Natural Science Basic Research Program of Shaanxi Province, Grant No.2025JC-QYCX-051,and the Fundamental Research Funds for the Central Universities and the Innovation Fund of Xidian University, Grant No. YJSJ25007.

References

Chen, Z.; Wu, X.; Xu, T.; and Kittler, J. 2023. Fast Self-Guided Multi-View Subspace Clustering. *IEEE Trans. Image Process.*, 32: 6514–6525.

Chua, T.-S.; Tang, J.; Hong, R.; Li, H.; Luo, Z.; and Zheng, Y. 2009. NUS-WIDE: a real-world web image database from National University of Singapore. In *ACM International Conference on Image and Video Retrieval*.

Duan, Y.; Nie, F.; Wang, R.; and Li, X. 2024. Harmonic cut: An efficient and directly solved balanced graph clustering. *Neurocomputing*, 578: 127381.

Gao, Q.; Xia, W.; Wan, Z.; Xie, D.; and Zhang, P. 2020. Tensor-SVD Based Graph Learning for Multi-View Subspace Clustering. In *AAAI Conference on Artificial Intelligence*, 3930–3937.

Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2021. Enhanced Tensor RPCA and its Application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6): 2133–2140.

Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale Multi-view Subspace Clustering in Linear Time. In *AAAI Conference on Artificial Intelligence*, 4412–4419.

Kilmer, M. E.; and Martin, C. D. 2011. Factorization strategies for third-order tensors. *Linear Algebra and its Applications*, 435(3): 641–658.

Lampert, C. H.; Nickisch, H.; and Harmeling, S. 2009. Learning to detect unseen object classes by between-class attribute transfer. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 951–958.

Li, J.; Gao, Q.; Wang, Q.; Deng, C.; and Xie, D. 2024a. Label Learning Method Based on Tensor Projection. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD '24, 1599–1609. New York, NY, USA: Association for Computing Machinery. ISBN 9798400704901.

Li, J.; Gao, Q.; Wang, Q.; Yang, M.; and Xia, W. 2023a. Orthogonal Non-negative Tensor Factorization based Multi-view Clustering. In *Neural Information Processing Systems*.

Li, J.; Wang, Q.; Yang, M.; Gao, Q.; and Gao, X. 2024b. Efficient Anchor Graph Factorization for Multi-View Clustering. *IEEE Trans. Multim.*, 26: 5834–5845.

Li, X.; Pan, Y. P.; Sun, Y.; Sun, Q.; Sun, Y.; W. Tsang, I.; and Ren, Z. 2025. Incomplete Multi-view Clustering with Paired and Balanced Dynamic Anchor Learning. *IEEE Transactions on Multimedia*, 7087–7098.

Li, X.; Sun, Y.; Sun, Q.; Ren, Z.; and Sun, Y. 2023b. Cross-view graph matching guided anchor alignment for incomplete multi-view clustering. *Information Fusion*, 100: 101941.

Li, X.; Zhang, H.; Wang, R.; and Nie, F. 2022. Multiview Clustering: A Scalable and Parameter-Free Bipartite Graph Fusion Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 330–344.

Liang, N.; Yang, Z.; Li, Z.; Sun, W.; and Xie, S. 2020. Multi-view clustering by non-negative matrix factorization with co-orthogonal constraints. *Knowl. Based Syst.*, 194(105582).

Liang, Y.; Huang, D.; Wang, C.; and Yu, P. S. 2024. Multi-View Graph Learning by Joint Modeling of Consistency and Inconsistency. *IEEE Trans. Neural Networks Learn. Syst.*, 35(2): 2848–2862.

Liu, J.; Wang, C.; Gao, J.; and Han, J. 2013. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM international conference on data mining*, 252–260. SIAM.

Liu, W.; He, J.; and Chang, S.-F. 2010. Large graph construction for scalable semi-supervised learning. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, 679–686. Citeseer.

Nie, F.; Wang, X.; Jordan, M. I.; and Huang, H. 2016. The Constrained Laplacian Rank Algorithm for Graph-Based Clustering. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, 1969–1976. AAAI Press.

Oliva, A.; and Torralba, A. 2001. Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *Int. J. Comput. Vis.*, 42(3): 145–175.

Samaria, F.; and Harter, A. 1994. Parameterisation of a stochastic model for human face identification. In *Proceedings of 1994 IEEE Workshop on Applications of Computer Vision*, 138–142.

van Breukelen, M.; Duin, R. P. W.; Tax, D. M. J.; and den Hartog, J. 1998. Handwritten digit recognition by combined classifiers. *Kybernetika*, 34: 381–386.

Wan, X.; Liu, X.; Liu, J.; Wang, S.; Wen, Y.; Liang, W.; Zhu, E.; Liu, Z.; and Zhou, L. 2023. Auto-Weighted Multi-View Clustering for Large-Scale Data. In *AAAI Conference on Artificial Intelligence*, 10078–10086.

Wang, D.; Han, S.; Wang, Q.; He, L.; min Tian, Y.; and Gao, X. 2021a. Pseudo-Label Guided Collective Matrix Factorization for Multiview Clustering. *IEEE Transactions on Cybernetics*, 52: 8681–8691.

Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2021b. Fast Parameter-Free Multi-View Subspace Clustering With Consensus Anchor Guidance. *IEEE Transactions on Image Processing*, 31: 556–568.

Winn, J.; and Jojic, N. 2005. LOCUS: learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV)*, volume 1, 756–763.

Xia, W.; Gao, Q.; Wang, Q.; Gao, X.; Ding, C.; and Tao, D. 2023. Tensorized Bipartite Graph Learning for Multi-View Clustering. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(4): 5187–5202.

Yang, B.; Zhang, X.; Lin, Z.; Nie, F.; Chen, B.; and Wang, F. 2022a. Efficient and Robust MultiView Clustering With Anchor Graph Regularization. *IEEE Trans. Circuits Syst. Video Technol.*, 32(9): 6200–6213.

Yang, H.; Gao, Q.; Xia, W.; Yang, M.; and Gao, X. 2022b. Multiview Spectral Clustering With Bipartite Graph. *IEEE Trans. Image Process.*, 31: 3591–3605.

Yang, Z.; Tan, Y.; and Yang, T. 2024. Large-scale multi-view clustering via matrix factorization of consensus graph. *Pattern Recognition*, 155: 110716.

Zhan, K.; Nie, F.; Wang, J.; and Yang, Y. 2019. Multiview Consensus Graph Clustering. *IEEE Trans. Image Process.*, 28(3): 1261–1270.

Zhao, W.; Li, Q.; Xu, H.; Gao, Q.; Wang, Q.; and Gao, X. 2024. Anchor Graph-Based Feature Selection for One-Step Multi-View Clustering. *IEEE Trans. Multim.*, 26: 7413–7425.

Zhao, X.; Fan, J.; Chang, X.; Nie, F.; Zhang, Q.; and Guo, J. 2025. Scalable Multi-view Regression Clustering for Large-scale Data. *IEEE Transactions on Circuits and Systems for Video Technology*, 1–1.

Zhen, X.; Yu, M.; He, X.; and Li, S. 2018. Multi-Target Regression via Robust Low-Rank Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(2): 497–504.