MULTI-VIEW COLLABORATIVE LABEL LEARNING WITH BALANCED REGULARIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Anchor graph-based methods reduce computational complexity and improve scalability in multi-view clustering. However, existing approaches suffer from two major limitations: they treat anchor graphs merely as sample representations — leading to under-representation and thus losing fine-grained structural information — and they ignore the consistency between anchor distribution and sample distribution, which results in suboptimal and unstable performance. To address these issues, we propose a balanced multi-view collaborative label learning method based on anchor representation. Specifically, instead of using anchors to represent samples, we use samples to represent anchors, which enhances the representational power of anchors and effectively resolves the under-representation problem inherent in the former approach. In addition, we design an anchor balancing mechanism based on the Schatten p-norm, supported by theoretical analysis, to solve the distribution consistency problem between the anchor space and the sample space. Collaborative learning is achieved by alternately updating sample labels and anchor labels, while a low-rank tensor constraint is imposed on the sample labels to capture complementary information across views. Extensive experiments demonstrate the effectiveness and superiority of our method.

1 INTRODUCTION

Clustering is an unsupervised learning technique that groups similar data points to uncover underlying patterns in the data. With the advancement of multi-modal perception technologies, it has become very easy to describe the same object from multiple perspectives, leading to an increasing volume of unlabeled multi-view data in real applications. Compared to single-view data, multi-view data help to provide complementary information from multiple dimensions, which is crucial for data analysis. Multi-view clustering integrates information from different views, improving the reliability of representation and the stability of label learning Zhang et al. (2020); Yan et al. (2023); Xu et al. (2015); Xie et al. (2018); Zhao et al. (2024b); Du et al. (2026). As a significant branch of multi-view clustering, graph-based clustering models data as a graph, which effectively captures similarity between samples, and can efficiently handle data with arbitrary shape. Consequently, it has attracted considerable attention in recent years Nie et al. (2017); Xu et al. (2020); Qiang et al. (2023); Wu et al. (2019).

Graph based clustering methods typically require constructing an $n \times n$ graph to represent relationships between samples, followed by graph decomposition to obtain a discrete label matrix. n is the number of samples. As a result, these methods often exhibit high computational complexity, typically $\mathcal{O}(n^2)$ or $\mathcal{O}(n^3)$, making them impractical for large-scale datasets. To address this issue, anchor graph-based multi-view clustering is one of the most representative and successful strategies due to the impressive performance and good mathematical theoryLiu et al. (2010); Li et al. (2022); Yang et al. (2024a); Chen et al. (2025). Where, connected components based anchor graph clustering is one of the most representative techniques Yang et al. (2022b); Liu et al. (2022); Fang et al. (2023); Xia et al. (2023). It aims to learn anchor graphs of all views so that the learned view consistent bipartite graph, which are composed of anchor graphs of views, has c connected components, and then directly gets indicator matrix according to the connected components, where c denotes the number of clusters. In real applications, however, it is very demanding on the selection of hyper-parameters and may not find a suitable bipartite graph with c-connected components in some cases, reducing the flexibility of algorithms.

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

To mitigate these issues and improve the interpretability of clustering models, an effective and efficient strategy is to directly apply linear regression Zheng et al. (2015); Zhao et al. (2024a); Xia et al. (2021) or non-negative matrix factorization(NMF) Lee & Seung (1999); Ding et al. (2010); Yang et al. (2022a); Li et al. (2023) to anchor graphs of views, followed by decision-level fusion to obtain the final indicator matrix. While these methods achieve promising performance and their motivations are different, all of them solely consider anchor graphs as a representation of the data, which leads to under-representation and fails to exploit the richer structural information encoded in the original sample space. Moreover, they ignore the consistency between anchor distribution and sample distribution, which results in suboptimal and unstable performance.

To address the aforementioned issues, We treat the anchor graph as a representation of anchors, that is, using rich sample information to represent the anchors. This view significantly enhances the representational capacity of anchors and naturally links anchor labels to sample labels via the anchor graph. Using anchors instead of samples for clustering learning requires that the distribution of the learned anchors be consistent with that of the original samples. To ensure this consistency, we propose a novel cluster balance regularization that enforces a balanced anchor distribution by maximizing the Schatten p-norm of the anchor label matrix. A detailed theoretical analysis is then provided. Subsequently, we present an effective and efficient clustering model for multi-view label learning, which enables alternating learning of sample labels and anchor labels, allowing the two types of variables to mutually calibrate each other and thus improving clustering robustness. Finally, we impose a tensor rank constraint to better exploit complementary information across views.

Importantly, the proposed framework is not tied to any particular way of selecting anchors. It works with the standard shared-anchor strategy, where anchors are chosen jointly from all views, and also with the view-specific alternative, where anchors are picked independently per view. Our experiments show that the method performs well in both cases, with no noticeable drop when moving from shared to view-specific anchors. This robustness comes from the collaborative learning scheme, which separates how anchors are represented from how they are obtained, so the model adapts to different anchor construction choices.

The contributions of our paper are summarized as follows:

- ***Schatten p*-norm Balance Principle.** We theoretically demonstrate that maximizing the *Schatten p*-norm of the indicator matrix promotes balanced cluster assignments. This finding offers a new perspective on the *Schatten p*-norm and broadens its role beyond traditional norm-minimization paradigms.
- **Balanced Anchor Representation.** Unlike existing methods that use anchors to represent samples (which leads to under-representation because $m \ll n$), we instead use samples to represent anchors. By leveraging abundant sample information and encouraging anchors to be evenly distributed across clusters, the learned anchors become more representative of the underlying cluster structure, significantly enhancing their representational capacity and alleviating the under-representation problem inherent in conventional anchor-based approaches.
- **Collaborative Learning with Tensor Low-Rank Constraint.** We achieve collaborative learning by alternately updating sample labels and anchor labels. A low-rank tensor constraint is imposed on the sample labels to capture complementary information across multiple views, which enhances the robustness and effectiveness of multi-view clustering.

2 Motivation and Objective

We construct an anchor graph $\mathbf{G} \in \mathbb{R}^{n\times m}$ to capture the relationships between the original data matrix $\mathbf{X} \in \mathbb{R}^{n\times d}$ and the anchor matrix, where n, m, and d denote the number of samples, anchors, and feature dimensions, respectively. Existing regression-based clustering methods typically treat $\mathbf{G}$ as a compact representation of $\mathbf{X}$ and learn a sample label matrix $\mathbf{H} \in \mathbb{R}^{n\times c}$ (with c the number of clusters) via a projection matrix $\mathbf{Q}$, i.e., $\mathbf{GQ} \approx \mathbf{H}$. The objective is often formulated as

$$\min_{\mathbf{Q}} \|\mathbf{GQ} - \mathbf{H}\|_F^2 . \tag{1}$$

However, this formulation only focuses on obtaining sample labels and completely ignores the relationship between anchor assignments and sample labels. The anchor graph describes the relationship

between samples and anchors, while they treat the anchor graph merely as a representation of samples, i.e., using m anchors to represent n sample points. Since the number of anchors is much smaller than the number of samples ($m \ll n$), this leads to under-representation, and thus it will miss the fine-grained structural information of the original space. Moreover, they ignore the consistency between anchor distribution and sample distribution, which results in suboptimal and unstable performance.

To address this issue, we treat the anchor graph as a representation of the anchors. That is, we use n samples to represent m anchors. Since n is much larger than m ($n \gg m$), this greatly enhances the representational capacity of the anchors. Using anchors instead of samples for clustering learning requires a key prerequisite: the distribution of the learned anchors must be consistent with the spatial distribution of the samples. Unfortunately, such consistency is often difficult to measure directly. Instead, from a label perspective, this consistency in spatial distribution naturally translates into the following requirement: assuming the samples belong to c clusters, the learned anchors should also be distributed across these c clusters, and should be as evenly distributed as possible among the clusters. Thus, the problem of consistency is transformed into the problem of achieving a balanced distribution of anchor labels.

Given the anchor graph $\mathbf{G}$ and the sample label matrix $\mathbf{H}$, the anchor label matrix $\mathbf{Q} \in \mathbb{R}^{m \times c}$ can be naturally obtained as $\mathbf{Q} = \mathbf{G}^\top \mathbf{H}$. The anchor label balancing problem essentially seeks a function $f(\cdot)$ such that when $f(\mathbf{Q})$ achieves its optimal value, the anchor labels in $\mathbf{Q}$ are distributed as evenly as possible across all classes. This leads to the general formulation: $f(\mathbf{Q}), \quad \text{s.t.} \quad \mathbf{Q} \geq 0,\ \mathbf{Q}\mathbf{1} = \mathbf{1}$.

Balanced Clustering: To ensure the balance of anchors within clusters, researchers have proposed some effective balance methods in machine learning and pattern recognition area, three of the most representative balance methods are Liu et al. (2018); Mi et al. (2024); Sun et al. (2023), formulated as follows:

$$\min_{\boldsymbol{Q}\geq 0, \boldsymbol{Q}\cdot\mathbf{1}=\mathbf{1}} \|\boldsymbol{Q}\|_F^2, \quad \min_{\boldsymbol{Q}\geq 0, \boldsymbol{Q}\cdot\mathbf{1}=\mathbf{1}} \left\|\boldsymbol{Q}^\top\right\|_{2,1}, \max_{\boldsymbol{Q}\geq 0, \boldsymbol{Q}\cdot\mathbf{1}=\mathbf{1}} \|\boldsymbol{Q}\|_F^2. \tag{2}$$

However, the aforementioned methods fail to prevent the occurrence of trivial solutions. Specifically, when the probability of each data point being assigned to every class is uniform, the first two terms in (2) attain their respective minima, i.e., $\mathbf{Q}^{(1)} = \frac{1}{c}\mathbf{1}_n\mathbf{1}_c^{\mathrm{T}} \in \mathbb{R}^{n\times c}$. At this point, the clustering results become indistinguishable.

Moreover, it is easy to prove that, when one column of $\mathbf{Q}$ has all 1 elements and the other columns have 0 elements, the last term in (2) attains its maximum value. This scenario implies that all data points are grouped into one cluster, leaving multiple empty clusters. This constitutes the second trivial solution.

To circumvent these two trivial solutions, we find that maximizing the Schatten p-norm of the label matrix not only help ensure a balance of samples across clusters, but also obtain discrete labels. Thus, we replace $f(\cdot)$ with the maximum Schatten p-norm and obtain a balanced clustering model as follows:

$$\max_{\mathbf{Q1=1}, \mathbf{Q}\geqslant\mathbf{0}} \|\mathbf{Q}\|_{Sp}^p \ (0 < p \leq 2) \tag{3}$$

For model (3), we have the following theorems.

Theorem 1. *Denote by $\beta_i = \tau_i(\mathbf{Q}^T\mathbf{Q})$ the i-th largest eigenvalue of $\mathbf{Q}^T\mathbf{Q}$, then the optimal solution of model (3) is equivalent to the optimal solution of Model (4).*

$$\max_{\boldsymbol{Q1=1}, \boldsymbol{Q}\geqslant\boldsymbol{0}} \|\boldsymbol{Q}\|_F^2, \ \ s.t. \beta_1 = \beta_2 = \cdots = \beta_c \tag{4}$$

Proof. Let $\sigma_i(\mathbf{Q})$ denote the i-th largest singular value of $\mathbf{Q}$, and $\tau_i(\mathbf{Q}^T\mathbf{Q})$ denote the i-th largest eigenvalue of $\mathbf{Q}^T\mathbf{Q}$. We have:

$$\|\mathbf{Q}\|_{Sp}^p = \sum_{i=1}^{c} \sigma_i^p(\mathbf{Q}) = \sum_{i=1}^{c} \tau_i^{\frac{p}{2}}(\mathbf{Q}^T\mathbf{Q}) = \sum_{i=1}^{c} \beta_i^{\frac{p}{2}} \tag{5}$$

where $\beta_i = \tau_i(\mathbf{Q}^T\mathbf{Q})$.

Let $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_c]^T \in \mathbb{R}^{c\times 1}$, $\gamma_1 = \cdots = \gamma_c = \frac{1}{c}$. For $0 < p \leq 2$, the function $f(\beta_i) = {\beta_i}^{\frac{p}{2}}$ is concave function with respect to β_j. Therefore, by applying Jensen's inequality, we obtain

$$f\left(\sum_{i=1}^{c}\gamma_i\beta_i\right) \geq \sum_{i=1}^{c}\gamma_i f(\beta_i) \tag{6}$$

Equality holds if and only if $\beta_1 = \beta_2 = \cdots = \beta_c$.

By simplifying the right-hand side of the inequality equation 6, we derive

$$\sum_{i=1}^{c}\gamma_i f(\beta_i) = \frac{1}{c}\sum_{i=1}^{c}\beta_i^{\frac{p}{2}} = \frac{1}{c}\|\mathbf{Q}\|_{Sp}^{p} \tag{7}$$

Similarly, by simplifying the left-hand side of the inequality equation 6, we have

$$\sum_{i=1}^{c}\gamma_i\beta_i = \sum_{i=1}^{c}\gamma_i\tau_i(\mathbf{Q}^T\mathbf{Q}) = \frac{1}{c}\sum_{i=1}^{c}\tau_i(\mathbf{Q}^T\mathbf{Q}) = \frac{1}{c}\|\mathbf{Q}\|_F^2 \tag{8}$$

Thus,

$$f(\sum_{i=1}^{c}\gamma_i\beta_i) = f(\frac{1}{c}\|\mathbf{Q}\|_F^2) = (\frac{1}{c}\|\mathbf{Q}\|_F^2)^{\frac{p}{2}} \tag{9}$$

Substituting equation 7 and equation 9 into equation 6, we have

$$c\left(\frac{1}{c}\|\mathbf{Q}\|_F^2\right)^{\frac{p}{2}} \geq \|\mathbf{Q}\|_{Sp}^{p} \tag{10}$$

According to inequality equation 6, we have that equality in equation 10 holds if and only if $\beta_1 = \cdots = \beta_c$. Consequently, maximizing the right-hand side of equation 10 is equal to maximize the left-hand side with constraint $\beta_1 = \cdots = \beta_c$. □

□

Theorem 2. *Let n_j be the number of samples in the j-th cluster. Then,*

$$\max_{\boldsymbol{Q}} \|\boldsymbol{Q}\|_{Sp}^{p} \quad s.t. \boldsymbol{Q} \geq 0, \boldsymbol{Q1} = \boldsymbol{1} \tag{11}$$

can ensure a balanced cluster distribution, i.e, $n_1 = n_2 = \ldots = n_c$.

Proof. According to Theorem 1, we have that model (3) is equal to model (12), i.e.,

$$\max_{\forall i, z_{ij}\geq 0,\ \sum_j z_{ij}=1} \|\mathbf{Q}\|_F^2, \text{s.t. } \beta_1 = \cdots = \beta_c \tag{12}$$

Temporarily ignoring the constraint $\beta_1 = \cdots = \beta_c$, Eq.equation 12 can be written to:

$$\max_{\forall i, z_{ij}\geq 0,\ \sum_j z_{ij}=1} \|\mathbf{Q}\|_F^2 = \max_{\forall i, z_{ij}\geq 0,\ \sum_j z_{ij}=1} \sum_{i,j} z_{ij}^2 \tag{13}$$

Since all rows of $\mathbf{Q}$ are independent, model (13) becomes a problem of optimizing each row of $\mathbf{Q}$ separately. The optimal solution to the problem equation 13 is that $\mathbf{Q}_i$ has only one element equal to 1 with all other elements being 0. In this case, $\mathbf{Q}^T\mathbf{Q} \in \mathbb{R}^{c\times c}$ is a diagonal matrix whose the i-th diagonal element is n_i, i.e., the number of samples in the i-th cluster. As is well known, the diagonal elements of a diagonal matrix are also the eigenvalues of the diagonal matrix. Thus, $\beta_i = n_i$.

However, theoretical analysis shows that Eq.equation 13 leads to a trivial solution: when one column of $\mathbf{Z}$ is all ones and the remaining columns are all zeros, the Frobenius norm reaches its maximum value. At this point, the clustering results become indistinguishable.

Therefore, we introduce the constraint $\beta_1 = \cdots = \beta_c$ to avoid this trivial solution. Enforcing it in Eq.equation 13 ensures $n_1 = n_2 = \cdots = n_c$, meaning the samples are evenly distributed across all clusters.

According to Theorem 1 and combining the aforementioned analysis, we have that maximizing $\|\mathbf{Q}\|_{Sp}^{p}$, i.e., model (11), helps achieve a balanced cluster distribution, i.e.,$n_1 = n_2 = \ldots = n_c$. □

Objective Function: Applying model equation 3 to multi-view clustering which is one of the most active topic in machine learning and pattern recognition area, we present a new multi-view label learning model as follows:

$$\max_{\mathbf{H}^{(v)}\geqslant\mathbf{0},\,\mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}}\sum_{v=1}^{V}\|{\mathbf{G}^{(v)}}^{\mathrm{T}}\mathbf{H}^{(v)}\|_{Sp}^{p}+\lambda\sum_{v=1}^{V}\mathcal{R}(\mathbf{H}^{(\nu)}),\tag{14}$$

where $\mathbf{G}^{(v)}\in\mathbb{R}^{n\times m}$ represents the anchor graph in the v-th view, and $\mathbf{H}^{(v)}\in\mathbb{R}^{n\times c}$ denotes the label matrix of samples in the v-th view. The second regularization term is to eliminate differences between the view-specific indicator matrix.

To better exploit the complementary information across different views, inspired by the work of Xia et al. (2023); Gao et al. (2025); Li et al. (2023), we minimize the *Schatten p*-norm of the sample label tensor, ensuring that the labels of samples from different views are more likely to remain consistent. Thus, the final objective function is

$$\max_{\mathbf{H}^{(v)}}\sum_{v=1}^{V}\|{\mathbf{G}^{(v)}}^{\mathrm{T}}\mathbf{H}^{(v)}\|_{Sp}^{p_1}-\lambda\|\boldsymbol{\mathcal{H}}\|_{\mathcal{S}p}^{p_2},\ \text{s.t.}\mathbf{H}^{(v)}\geqslant\mathbf{0},\ \mathbf{H}^{(v)}\cdot\mathbf{1}=\mathbf{1}\tag{15}$$

where $0<p_1,p_2\leqslant 1$, $\boldsymbol{\mathcal{H}}\in\mathbb{R}^{n\times V\times c}$, and λ is a hyperparameter.

3 Optimization and Complexity Analysis

We employ the Augmented Lagrange Multiplier (ALM) method to solve problem equation 15, introducing the auxiliary variable $\boldsymbol{\mathcal{D}}$ and setting $\mathcal{H}=\mathcal{D}$. Consequently, the objective function is reformulated as:

$$\begin{aligned}\max_{\mathbf{H}^{(v)},\mathbf{D}^{(v)}}&\sum_{v=1}^{V}\|{\mathbf{G}^{(v)}}^{\mathrm{T}}\mathbf{H}^{(v)}\|_{Sp}^{p_1}-\lambda\|\boldsymbol{\mathcal{D}}\|_{\mathcal{S}p}^{p_2}-\frac{\mu}{2}\|\boldsymbol{\mathcal{H}}-\boldsymbol{\mathcal{D}}+\frac{\boldsymbol{\mathcal{Y}}}{\mu}\|_F^2,\\&\text{s.t.}\mathbf{H}^{(v)}\geqslant\mathbf{0},\ \mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}\end{aligned}\tag{16}$$

where $\boldsymbol{\mathcal{Y}}$ represents the Lagrange multipliers, and μ is a penalty parameter. The problem is decomposed into three steps.

• **Updating $\mathcal{H}$ while keeping $\mathcal{D}$ fixed.** Under this condition, equation 16 simplifies to:

$$\max_{\mathbf{H}^{(v)}\geqslant\mathbf{0},\,\mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}}\sum_{v=1}^{V}\|{\mathbf{G}^{(v)}}^{\mathrm{T}}\mathbf{H}^{(v)}\|_{Sp}^{p_1}-\frac{\mu}{2}\|\boldsymbol{\mathcal{H}}-\boldsymbol{\mathcal{D}}+\frac{\boldsymbol{\mathcal{Y}}}{\mu}\|_F^2.\tag{17}$$

Since all $\mathbf{H}^{(v)}(v=1,2,\ldots,V)$ are independent, problem equation 17 can be decomposed into V independent sub-optimization problems. Consequently, solving equation 17 reduces to solving:

$$\max_{\mathbf{H}^{(v)}\geqslant\mathbf{0},\,\mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}}\|{\mathbf{G}^{(v)}}^{T}\mathbf{H}^{(v)}\|_{Sp}^{p_1}-\frac{\mu}{2}\|\mathbf{H}^{(v)}-\mathbf{M}^{(v)}\|_F^2,\tag{18}$$

where $\mathbf{M}^{(v)}=\mathbf{D}^{(v)}-\frac{\mathbf{y}^{(v)}}{\mu}$.

To solve equation 18, we replace the first term of the model equation 18 with a first-order Taylor expansion as the lower bound of the first term, and then iteratively solve it. To be specific,

Defining $\Phi(\mathbf{Y}^{(v)})=\|\mathbf{Y}^{(v)}\|_{Sp}^{p_1}$, where $\mathbf{Y}^{(v)}={\mathbf{G}^{(v)}}^{T}\mathbf{H}^{(v)}$, we perform a first-order Taylor expansion at $\mathbf{Y}^{(v)}$ as follows:

$$\Phi(\mathbf{Y}^{(v)})=\Phi({\mathbf{Y}^{(v)}}^{(t)})+\langle\nabla\Phi({\mathbf{Y}^{(v)}}^{(t)}),\mathbf{Y}^{(v)}-{\mathbf{Y}^{(v)}}^{(t)}\rangle,\tag{19}$$

where ${\mathbf{Y}^{(v)}}^{(t)}$ represents the solution at the t-th iteration. Clearly, we can obtain:

$$\mathbf{W}^{(v)}=\nabla\Phi(\mathbf{Y}^{(v)})=\frac{\partial\|\mathbf{Y}^{(v)}\|_{Sp}^{p_1}}{\partial\mathbf{Y}^{(v)}}=p_1\mathbf{U}\boldsymbol{\Sigma}^{-1}|\boldsymbol{\Sigma}|^{p_1}\mathbf{V}^{\mathrm{T}},\tag{20}$$

where the singular value decomposition (SVD) of $\mathbf{Y}^{(v)}$ is given by $\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, and $\mathbf{\Sigma}^{-1}$ denotes the Moore-Penrose pseudo-inverse of $\mathbf{\Sigma}$.

Ignoring the constant term in (19) and substituting it into the model equation 18, we have

$$\min_{\mathbf{H}^{(v)}\geqslant\mathbf{0},\ \mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}} \left(\frac{\mu}{2}\|\mathbf{H}^{(v)}-\mathbf{M}^{(v)}\|_F^2 - Tr(\mathbf{W}^{(v)^T}\mathbf{G}^{(v)^T}\mathbf{H}^{(v)})\right). \tag{21}$$

Defining $\mathbf{S}^{(v)} = \mathbf{G}^{(v)}\mathbf{W}^{(v)} + \mu\mathbf{M}^{(v)}$, model equation 21 becomes

$$\min_{\mathbf{H}^{(v)}\geqslant\mathbf{0},\ \mathbf{H}^{(v)}\mathbf{1}=\mathbf{1}} \left(Tr(\mathbf{H}^{(v)^T}\mathbf{H}^{(v)}) - \frac{2}{\mu}Tr(\mathbf{S}^{(v)^T}\mathbf{H}^{(v)})\right). \tag{22}$$

Since each row of $\mathbf{H}^{(v)}$ is independent, equation 22 can be decomposed into the following n sub-problems:

$$\min_{\mathbf{H}^{(v)^i}\mathbf{1}=1,\ \mathbf{H}^{(v)^i}\geqslant\mathbf{0}} \sum_{i=1}^{n} \|\mathbf{H}^{(v)^i} - \frac{1}{\mu}\mathbf{s}^{(v)^i}\|_2^2, \tag{23}$$

where $\mathbf{H}^{(v)^i}$ represents the i-th row of $\mathbf{H}^{(v)}$. The optimization problem equation 23 can be efficiently solved using an iterative method proposed in Huang et al. (2015).

• **Solving $\mathcal{D}$ while keeping $\mathcal{H}$ fixed.** When $\mathcal{H}$ is fixed, equation 16 simplifies to:

$$\min_{\mathcal{D}} \frac{\lambda}{\mu}\|\mathcal{D}\|_{\circledS}^{p_2} + \frac{1}{2}\left\|\mathcal{D} - (\mathcal{H} + \frac{\mathcal{Y}}{\mu})\right\|_F^2. \tag{24}$$

The solution for $\mathcal{D}$ follows directly from Theorem 3.

Theorem 3. *Let $\mathcal{T} \in \mathbb{R}^{n_1\times n_2\times n_3}$ have a t-SVD decomposition given by $\mathcal{T} = \mathcal{G} * \mathcal{Q} * \mathcal{V}^{\mathrm{T}}$. Then, for the optimization problem:*

$$\min_{\mathcal{X}} \frac{1}{2}\|\mathcal{X}-\mathcal{T}\|_F^2 + \tau\|\mathcal{X}\|_{\circledS}^p, \tag{25}$$

the optimal solution is given by:

$$\mathcal{X}^* = \Gamma_\tau(\mathcal{T}) = \mathcal{G} * \mathit{ifft}(P_\tau(\overline{\mathcal{T}})) * \mathcal{V}^{\mathrm{T}}, \tag{26}$$

where $P_\tau(\overline{\mathcal{T}})$ is a third-order tensor whose diagonal elements are computed via the generalized soft-thresholding algorithm Gao et al. (2020).

By applying Theorem 3, we obtain the solution for $\mathcal{D}$:

$$\mathcal{D}^* = \Gamma_{\frac{\lambda}{\mu}}(\mathcal{H} + \frac{\mathcal{Y}}{\mu}). \tag{27}$$

• **Updating the Lagrange multipliers and penalty parameters.**

The Lagrange multiplier $\mathcal{Y}$ and the penalty parameter μ are updated as follows:

$$\mathcal{Y} = \mathcal{Y} + \mu(\mathcal{H}-\mathcal{D}), \mu = \min(\mu\times 1.1, \max_\mu),. \tag{28}$$

where max_μ is a predefined upper bound.

Our algorithm alternately updates $\mathcal{H}$, $\mathbf{W}^{(v)}$, $\mathcal{D}$, and $\mathcal{Y}$ (with μ) until convergence after anchor graph construction, and then obtains the clustering label matrix $\mathbf{H}$ by averaging the view-specific label matrices.

The computational complexity of the proposed model primarily arises from the following two stages: 1) Construction of the anchor graph $\mathbf{G}$, which has a computational complexity of $\mathcal{O}(nmdV + nmV\log(m))$. 2) Optimization of $\mathcal{D}$ and $\mathcal{H}$, where the respective computational complexities are $\mathcal{O}(2ncV\log(cV) + ncV^2)$ and $\mathcal{O}(nmcV)$. Since m, c, and V are relatively small, the overall computational complexity of the model can be approximated as $\mathcal{O}(nmV\log(m) + nmctV)$, where t denotes the number of iterations.

Table 1: The results on the MSRC, Handwritten4, Mnist4 and Scene15 datasets.

Datasets	MSRC			Handwritten4			Mnist4			Scene15		
Methods	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
ETLMSC (2019)	0.962	0.937	0.962	0.938	0.893	0.938	0.934	0.847	0.934	0.709	0.774	0.887
LMVSC(2020)	0.814	0.717	0.814	0.904	0.831	0.904	0.892	0.726	0.892	0.561	0.512	0.581
FPMVS-CAG(2021)	0.843	0.738	0.843	0.850	0.787	0.850	0.887	0.719	0.887	0.618	0.566	0.651
FMCNOF(2022)	0.713	0.648	0.714	0.541	0.484	0.54	0.686	0.513	0.695	0.263	0.249	0.258
MSC-BG(2022)	0.981	0.960	0.981	0.889	0.922	0.889	0.938	0.861	0.938	0.690	0.759	0.734
SFMC(2022)	0.809	0.721	0.781	0.853	0.871	0.873	0.917	0.801	0.917	0.343	0.522	0.346
FSMSC(2023)	0.876	0.753	0.876	0.878	0.787	0.878	0.791	0.579	0.791	0.440	0.414	0.450
Orth-NTF(2023)	0.990	0.978	0.990	0.985	0.969	0.985	0.977	0.926	0.977	0.758	0.804	0.759
OMVFC-LICAG(2024)	0.728	0.590	0.738	0.837	0.798	0.838	0.888	0.725	0.888	0.484	0.517	0.537
TLL-AG(2024)	0.986	0.968	0.986	0.982	0.962	0.982	0.981	0.935	0.981	0.719	0.785	0.752
LMVC-MFCG(2024)	0.852	0.738	0.852	0.856	0.832	0.856	0.886	0.717	0.886	0.505	0.566	0.522
AGFS-OMVC(2024)	**1.000**	**1.000**	**1.000**	0.995	0.987	0.995	0.979	0.928	0.979	0.796	0.834	0.806
ALPC(2025)	0.881	0.775	0.881	0.710	0.662	0.730	0.783	0.580	0.783	0.363	0.336	0.393
LMTC(2025)	0.881	0.808	0.881	0.710	0.662	0.730	0.783	0.580	0.783	0.363	0.336	0.393
MVC-TPL(2025)	0.995	0.989	0.995	0.987	0.971	0.987	0.995	0.979	0.995	0.845	0.848	0.845
Dual MoE MVKM(2026)	0.714	0.661	0.581	0.708	0.668	0.710	0.893	0.734	0.750	0.372	0.373	0.195
Our work(Shared Anchors)	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.839	0.853	0.840
Our work(View-specific Anchors)	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**0.898**	**0.890**	**0.898**

4 EXPERIMENT

4.1 DATASETS AND COMPETITORS

Our experiments are conducted on four small-scale datasets and three large-scale datasets: **MSRC-v5 (MSRC)** Winn & Jojic (2005), **Handwritten4 (HW4)** Asuncion & Newman (2007), **Mnist4** Deng (2012), **Scene15** Kilmer et al. (2013), **Reuters** Apté et al. (1994), **NUS-WIDE (NUS)** Chua et al. (2009), and **AWA** Fu et al.. All experiments are conducted a standard Windows 11 with AMD Ryzen 5 9600X 6-Core Processor and 32 GB RAM, using MATLAB R2025b. We assess the effectiveness of models by using three indicators: Accuracy(ACC), normalized mutual information(NMI) and Purity.

To evaluate the superiority of our model, it is compared with: **ETLMSC** Wu et al. (2019), **LMVSC** Kang et al. (2020), **FMCNOF** Yang et al. (2020), **FPMVS-CAG** Wang et al. (2021), **MSC-BG** Yang et al. (2022b), **SFMC** Li et al. (2022), **FSMSC** Chen et al. (2023), **Orth-NTF** Li et al. (2023), **OMVFC-LICAG** Zhang et al. (2024), **TLL-AG** Li et al. (2024b), **LMVC-MFCG** Yang et al. (2024b), **AGFS-OMVC** Zhao et al. (2024a), **ALPC** Chen et al. (2025), **LMTC** Liu et al. (2025), **MVC-TPL** Xue et al. (2025), and **Dual MoE MVKM** Du et al. (2026).

4.2 CLUSTERING PERFORMANCE

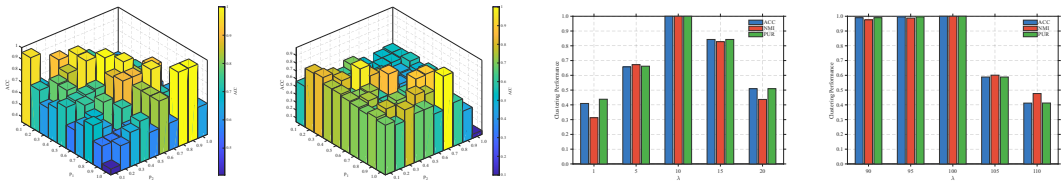
The experimental results are systematically presented in Tables 1 and 2, with optimal performance highlighted in bold and secondary results $\underline{underlined}$. In the table, both "Our work (Shared Anchors)" and "Our work (View-specific Anchors)" denote the model proposed in this paper. The former builds shared anchors after concatenating all views, while the latter selects anchors independently for each view using the same anchor rate. From the tables, we can observe the following results:

Compared with anchor graph-based methods (e.g., LMVSC, FPMVS-CAG, FMCNOF) and tensor constraint methods (e.g., Orth-NTF, TLL-AG), our algorithm achieves superior performance. Anchor graph-based methods often rely on element-wise error measurement and thus fail to capture cross-view complementary information, whereas our method (like TLL-AG) leverages the tensor Schatten-p norm to capture spatial structure and complementarity. Unlike tensor constraint methods, we directly model the relationships among anchor graph, anchor assignments, and sample labels, and enforce a balanced anchor distribution. Although AGFS-OMVC and MVC-TPL also build anchor graph connections, our collaborative learning of anchor and sample labels with balanced regularization better captures their interdependence, leading to improved clustering.

We also compare the two variants. On MSRC, Handwritten4, and Mnist4, both achieve perfect scores. On the four larger datasets, however, the view-specific variant consistently matches or exceeds the shared one. For example, on Reuters, it lifts ACC from 0.857 to 0.882 and NMI from 0.810 to 0.884; on NUS, gains are also clear (ACC: 0.686 vs. 0.648; NMI: 0.736 vs. 0.719). Overall, independent anchor selection performs better on most datasets, especially as data scale grows and views become more diverse.

Table 2: The results on the Reuters, NUS and AWA datasets, 'OM' means out of memory.

Datasets / Methods	Reuters ACC	Reuters NMI	Reuters PUR	NUS ACC	NUS NMI	NUS PUR	AWA ACC	AWA NMI	AWA PUR
ETLMSC (2019)	OM	OM	OM	OM	OM	OM	OM	OM	OM
LMVSC(2020)	0.589	0.335	0.615	0.133	0.110	0.166	0.075	0.086	0.089
FPMVS-CAG(2021)	0.526	0.323	0.603	0.173	0.127	0.210	0.091	0.109	0.096
FMCNOF(2022)	0.343	0.125	0.358	0.117	0.101	0.210	0.050	0.022	0.051
SFMC(2022)	0.602	0.354	0.604	0.168	0.060	0.190	0.039	0.003	0.040
FSMSC(2023)	0.517	0.307	0.558	0.143	0.131	0.237	0.097	0.113	0.118
Orth-NTF(2023)	0.694	0.686	0.809	0.235	0.372	0.333	0.109	0.115	0.134
OMVFC-LICAG(2024)	0.426	0.087	0.431	0.158	0.128	0.236	0.079	0.097	0.094
TLL-AG(2024)	0.686	0.554	0.740	0.115	0.083	0.185	0.243	0.279	0.279
LMVC-MFCG(2024)	0.579	0.313	0.586	0.183	0.135	0.219	0.097	0.118	0.104
AGFS-OMVC(2024)	0.646	0.593	0.771	0.300	0.467	0.390	0.240	0.520	0.287
ALPC(2025)	0.481	0.268	0.564	0.143	0.149	0.256	0.081	0.086	0.103
LMTC(2025)	0.459	0.268	0.528	0.122	0.101	0.212	0.109	0.120	0.126
MVC-TPL(2025)	0.817	0.783	0.841	0.394	0.583	0.524	0.430	0.665	0.481
Dual MoE MVKM(2026)	0.501	0.302	0.560	0.117	0.103	0.210	0.070	0.071	0.084
Our work(Shared Anchors)	0.857	0.810	0.857	0.648	0.719	0.669	**0.587**	0.757	0.626
Our work(View-specific Anchors)	**0.882**	**0.884**	**0.922**	**0.686**	**0.736**	**0.709**	0.585	**0.762**	**0.633**

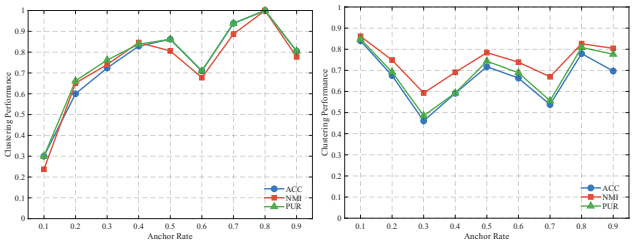


(a) MSRC (b) HandWritten4 (c) MSRC (d) HandWritten4

Figure 2: Effects of p_1, p_2, and λ on clustering results.

4.3 Parameter Analysis

- **Effect of the anchor rate parameter**: We vary the anchor point rate from 0.1 to 0.9 in increments of 0.1, and the corresponding experimental results are presented in Fig. 1. It can be observed that increasing the anchor point rate does not necessarily lead to performance improvement. Therefore, selecting an appropriate anchor rate is crucial for balancing efficiency and effectiveness.



(a) MSRC (b) Scene15

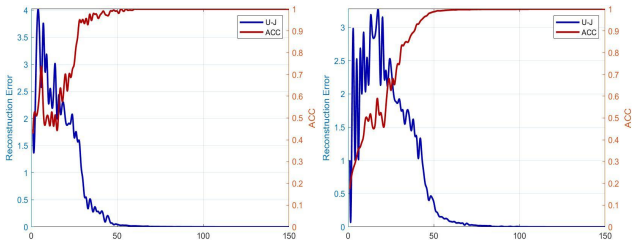
Figure 1: Effect of parameter anchor rate on clustering results

- **Effect of parameters p_1 and p_2**: Fig. 2 (a) and (b) illustrate the impact of p_1 and p_2 on clustering performance. It can be observed that clustering results are generally better when $p_1, p_2 < 1$ compared to when $p_1, p_2 = 1$. Clearly, the optimal parameter combination varies across datasets, and there is no fixed optimal combination of p_1 and p_2.
- **Effect of λ**: To investigate the influence of λ, we vary it within $\{1, 5, 10, 15, 20\}$ on MSRC and $\{90, 95, 100, 105, 110\}$ on HandWritten4. As shown in Fig. 2 (c) and (d), the performance first increases and then decreases with λ on both datasets. The best results are achieved at $\lambda = 10$ for MSRC and $\lambda = 100$ for HandWritten4, where all metrics approach 1.0, while larger values cause noticeable drops. Hence, only appropriately tuned λ can

effectively exploit cross-view complementary information, and the optimal value varies across datasets.

4.4 Convergence Analysis

When solving the objective function using the ALM method, we introduce an auxiliary variable. The model is considered to have converged when the reconstruction error (i.e., $\|\mathcal{H} - \mathcal{D}\|_\infty$) falls below a predefined threshold. The relationship between the reconstruction error, ACC, and the number of iterations is depicted in Fig. 3. As shown, the model converges after approximately 70 iterations.



(a) MSRC (b) HandWritten4

Figure 3: Convergence and clustering results on the two datasets.

5 Conclusion

In this paper, we propose a balanced multi-view collaborative label learning method based on anchor representation. By using samples to represent anchors, we solve the under-representation problem inherent in methods that use anchors to represent samples. We also introduce a label balancing mechanism based on the Schatten p-norm, which, together with collaborative learning of anchor labels and sample labels, solves the consistency problem between anchor distribution and sample distribution. Furthermore, we impose a low-rank constraint on the sample label tensor to exploit complementary information across multiple views, thereby enhancing the robustness and effectiveness of multi-view clustering. Extensive experiments demonstrate the effectiveness and superiority of the proposed method.

AI use statement

(This section is **required** and does not count toward the page limit.)

In this work, we did not use generative AI tools for any tasks. We take responsibility for the final content of this work.

Ethics statement

(This section is **recommended** and does not count toward the page limit.)

This work focuses on multi-view clustering using publicly available benchmark datasets. It does not involve human subjects, personal data, or any sensitive information. All datasets are publicly released and properly cited, and their licenses are respected. We do not foresee any harmful insights, discrimination, or fairness concerns arising from this work. The authors declare no conflicts of interest. No IRB approval was required because no human subjects were involved. We take responsibility for the research integrity of this work.

References

Chidanand Apté, Fred Damerau, and Sholom M Weiss. Automated learning of decision rules for text categorization. *ACM Transactions on Information Systems (TOIS)*, 12(3):233–251, 1994.

Arthur Asuncion and David Newman. Uci machine learning repository, 2007.

Yoshua Bengio and Yann LeCun. Scaling learning algorithms towards AI. In *Large Scale Kernel Machines*. MIT Press, 2007.

Hongmin Cai, Yu Wang, Fei Qi, Zhuoyao Wang, and Yiu Ming Cheung. Multiview tensor spectral clustering via co-regularization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (10):46, 2024.

Yawei Chen, Huibing Wang, Jinjia Peng, and Yang Wang. Anchor learning with potential cluster constraints for multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 15939–15947, 2025.

Zhe Chen, Xiao-Jun Wu, Tianyang Xu, and Josef Kittler. Fast self-guided multi-view subspace clustering. *IEEE Transactions on Image Processing*, 32:6514–6525, 2023.

Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. Nus-wide: a real-world web image database from national university of singapore. In *Proceedings of the ACM international conference on image and video retrieval*, pp. 1–9, 2009.

Li Deng. The mnist database of handwritten digit images for machine learning research. *IEEE signal processing magazine*, 29(6):141–142, 2012.

Chris H. Q. Ding, Tao Li, and Michael I. Jordan. Convex and semi-nonnegative matrix factorizations. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(1):45–55, 2010.

Liang Du, Yukai Shi, Yan Chen, Feijiang Li, Peng Zhou, Lei Duan, and Yuhua Qian. A dual mixture-of-experts framework for multi-view k-means clustering with view balance and regional sparsity. *Information Fusion*, pp. 103449, 2026.

Si-Guo Fang, Dong Huang, Xiao-Sha Cai, Chang-Dong Wang, Chaobo He, and Yong Tang. Efficient multi-view clustering via unified and discrete bipartite graph learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

Yanwei Fu, Timothy M Hospedales, Tao Xiang, and Shaogang Gong. Transductive multi-view zero-shot learning.

Quanxue Gao, Pu Zhang, Wei Xia, Deyan Xie, Xinbo Gao, and Dacheng Tao. Enhanced tensor rpca and its application. *IEEE transactions on pattern analysis and machine intelligence*, 43(6): 2133–2140, 2020.

Quanxue Gao, Fangfang Li, Qianqian Wang, Xinbo Gao, and Dacheng Tao. Manifold based multi-view k-means. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(4):3175–3182, 2025.

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

Geoffrey E. Hinton, Simon Osindero, and Yee Whye Teh. A fast learning algorithm for deep belief nets. *Neural Computation*, 18:1527–1554, 2006.

Jin Huang, Feiping Nie, and Heng Huang. A new simplex sparse learning model to measure data similarity for clustering. In *Twenty-fourth international joint conference on artificial intelligence*, 2015.

Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pp. 4412–4419, 2020.

Misha E Kilmer, Karen Braman, Ning Hao, and Randy C Hoover. Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging. *SIAM Journal on Matrix Analysis and Applications*, 34(1):148–172, 2013.

Abhishek Kumar, Piyush Rai, and Hal Daumé. Co-regularized multi-view spectral clustering. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 1–9, 2011.

Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.

Fangfang Li, Quanxue Gao, Qianqian Wang, Ming Yang, and Cheng Deng. Tensorized soft label learning based on orthogonal nmf. *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2024a. doi: 10.1109/TNNLS.2024.3442435.

Jing Li, Quanxue Gao, Qianqian Wang, Ming Yang, and Wei Xia. Orthogonal non-negative tensor factorization based multi-view clustering. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, pp. 18186–18202, 2023.

Jing Li, Quanxue Gao, Qianqian Wang, and Wei Xia. Tensorized label learning on anchor graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 13537–13544, 2024b.

Xuelong Li, Han Zhang, Rong Wang, and Feiping Nie. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1):330–344, 2022.

Yeqing Li, Feiping Nie, Heng Huang, and Junzhou Huang. Large-scale multi-view spectral clustering via bipartite graph. In *Proceedings of the AAAI conference on artificial intelligence*, volume 29, 2015.

Hongfu Liu, Ziming Huang, Qi Chen, Mingqin Li, Yun Fu, and Lintao Zhang. Fast clustering with flexible balance constraints. In *2018 IEEE International Conference on Big Data (Big Data)*, pp. 743–750. IEEE, 2018.

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 Computer Vision and Pattern Recognition Conference*, pp. 20727–20736, 2025.

Suyuan Liu, Siwei Wang, Pei Zhang, Kai Xu, Xinwang Liu, Changwang Zhang, and Feng Gao. Efficient one-pass multi-view subspace clustering with consensus anchors. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 7576–7584, 2022.

Wei Liu, Junfeng He, and Shih-Fu Chang. Large graph construction for scalable semi-supervised learning. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pp. 679–686. Citeseer, 2010.

Yong Mi, Hongmei Chen, Zhong Yuan, Chuan Luo, Shi-Jinn Horng, and Tianrui Li. Fast multi-view subspace clustering with balance anchors guidance. *Pattern Recognition*, 145:109895, 2024.

Feiping Nie, Jing Li, Xuelong Li, et al. Self-weighted multiview clustering with multiple graphs. In *IJCAI*, pp. 2564–2570, 2017.

Qianyao Qiang, Bin Zhang, Fei Wang, and Feiping Nie. Multi-view discrete clustering: A concise model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Lilei Sun, Jie Wen, Chengliang Liu, Lunke Fei, and Lusi Li. Balance guided incomplete multi-view spectral clustering. *Neural Networks*, 166:260–272, 2023.

Rui Wang, Quanxue Gao, Ming Yang, and Qianqian Wang. Tensorized tri-factor decomposition for multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2025. doi: 10.1109/TCSVT.2025.3536629.

Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31:556–568, 2021.

John Winn and Nebojsa Jojic. Locus: Learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pp. 756–763. IEEE, 2005.

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.

Wei Xia, Quanxue Gao, Qianqian Wang, and Xinbo Gao. Regression-based clustering network via combining prior information. *Neurocomputing*, 448:324–332, 2021.

Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5187–5202, 2023.

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. *Int. J. Comput. Vis.*, 126(11):1157–1179, 2018.

Chang Xu, Dacheng Tao, and Chao Xu. Multi-view intact space learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(12):2531–2544, 2015.

Huiling Xu, Xiangdong Zhang, Wei Xia, Quanxue Gao, and Xinbo Gao. Low-rank tensor constrained co-regularized multi-view spectral clustering. *Neural Networks*, 132:245–252, 2020.

Xingyu Xue, Wenhui Zhao, Quanxue Gao, Ming Yang, and Cheng Deng. Image clustering with transition probabilities learning. *IEEE Trans. Image Process.*, 34:1441–1453, 2025.

Weiqing Yan, Yuanyang Zhang, Chenlei Lv, Chang Tang, Guanghui Yue, Liang Liao, and Weisi Lin. Gcfagg: Global and cross-view feature aggregation for multi-view clustering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19863–19872, 2023.

Ben Yang, Xuetao Zhang, Feiping Nie, Fei Wang, Weizhong Yu, and Rong Wang. Fast multi-view clustering via nonnegative and orthogonal factorization. *IEEE transactions on image processing*, 30:2575–2586, 2020.

Ben Yang, Xuetao Zhang, Zhiping Lin, Feiping Nie, Badong Chen, and Fei Wang. Efficient and robust multiview clustering with anchor graph regularization. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(9):6200–6213, 2022a.

Ben Yang, Jinghan Wu, Xuetao Zhang, Xinhu Zheng, Feiping Nie, and Badong Chen. Discrete correntropy-based multi-view anchor-graph clustering. *Information Fusion*, 103:102097, 2024a.

Haizhou Yang, Quanxue Gao, Wei Xia, Ming Yang, and Xinbo Gao. Multiview spectral clustering with bipartite graph. *IEEE Transactions on Image Processing*, 31:3591–3605, 2022b.

Zengbiao Yang, Yihua Tan, and Tao Yang. Large-scale multi-view clustering via matrix factorization of consensus graph. *Pattern Recognition*, 155:110716, 2024b.

Changqing Zhang, Huazhu Fu, Qinghua Hu, Xiaochun Cao, Yuan Xie, Dacheng Tao, and Dong Xu. Generalized latent multi-view subspace clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(1):86–99, 2020.

Chuanbin Zhang, Long Chen, Zhaoyin Shi, and Weiping Ding. Latent information-guided one-step multi-view fuzzy clustering based on cross-view anchor graph. *Information Fusion*, 102:102025, 2024.

Wenhui Zhao, Qin Li, Huafu Xu, Quanxue Gao, Qianqian Wang, and Xinbo Gao. Anchor graph-based feature selection for one-step multi-view clustering. *IEEE Transactions on Multimedia*, 2024a.

Zihua Zhao, Feiping Nie, Rong Wang, Zheng Wang, and Xuelong Li. An balanced, and scalable graph-based multiview clustering method. *IEEE Transactions on Knowledge and Data Engineering*, 2024b.

Shuai Zheng, Xiao Cai, Chris Ding, Feiping Nie, and Heng Huang. A closed form solution to multi-view low-rank regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29, 2015.