

Tensorized Label Learning Based Fast Fuzzy Clustering

Xingyu Xue*, Jingjing Xue*, Quanxue Gao[†], Qianqian Wang

School of Telecommunication Engineering, Xidian University, Shaanxi 710071, China.
23011211150@stu.xidian.edu.cn, xuejingjing@xidian.edu.cn, qxgao@xidian.edu.cn, qqwang@xidian.edu.cn

Abstract

Multi-view graph clustering methods have been widely concerned due to the ability of dealing with arbitrarily shaped datasets. However, many methods with higher time and space complexity make them challenging to deal with large-scale datasets. Besides, many fuzzy clustering methods need additional regularization terms or hyper-parameters to obtain the membership matrix or avoid trivial solutions, which weakens the model generalization ability. Furthermore, inconsistent clustering labels can arise when there are significant discrepancies between views, making it challenging to effectively leverage the complementary information from different views. To this end, we propose Tensorized Label Learning based Fast Fuzzy Clustering (TLLFFC). Specifically, we design a novel balanced regularization term to reduce pressure of tuning regularization parameters for fuzzy clustering. The label transmission strategy with the anchor graph makes TLLFFC suitable for large-scale datasets. Moreover, incorporating the Schatten p -norm regularization on the label matrices can effectively unearth the complementary information distributed among views, thereby align the labels across views more consistently. Extensive experiments verify the superiority of TLLFFC.

Introduction

In recent decades, data mining and machine learning techniques have been increasingly applied in fields such as image processing (Gong, Yuan, and Bao 2021a), video annotation (Liu and Tsang 2015), and data analysis (Liu and Tsang 2017). Meanwhile, with the widespread use of various sensors and technologies, descriptions of the same object have become increasingly diverse and heterogeneous (Gong et al. 2022; Gong, Yuan, and Bao 2021b). Multi-view clustering methods (Cui et al. 2023) can identify these multi-view datasets with common features and integrate them into groups, ensuring that samples within the same cluster exhibit high similarity. Among them, multi-view graph clustering methods (Wan et al. 2023) become popular due to their abilities of dealing with arbitrarily shaped multimedia data better. Multi-view graph clustering methods (Wang, Yang, and Liu 2019; Wu, Lin, and Zha 2019) usually need to construct an

$n \times n$ large graph, and implement the eigenvalue decomposition step, or solve the optimal problem directly (Wang, Yang, and Liu 2019), whose time complexity are $\mathcal{O}(Vn^3)$ or $\mathcal{O}(Vn^2)$. Therefore, they are not suitable for large-scale multi-view datasets.

The anchor technique (Xia et al. 2023) has been utilized to reduce the size of the similarity matrix and expedite the optimization process, yielding impressive performance across various applications. Anchor points, which are representative data points, effectively cover the entire data point cloud and capture the internal structure of the data. Anchor-based methods first generate representative anchors from the original samples using a specific strategy and then construct a similarity graph to measure the similarity between samples and anchors. Following this, clustering processes are applied to the representation matrix.

Further, some researches use the two-step strategy to get final label matrices (Wang et al. 2021; Kang et al. 2020), whose solutions are suboptimal, as they are far away from the solutions obtained by directly solving the original problem. Therefore, (1) the rank constraint are introduced to ensure c connected components to avoid the post-process (Xia et al. 2023; Li et al. 2020; Fang et al. 2023; Xue et al. 2024); (2) some novel optimization strategies (Qiang et al. 2023; Nie et al. 2022) or Nonnegative Matrix Factorization (NMF) models (Yang et al. 2020) are proposed, which can obtain the final solution directly. The former has high requirements on parameters, and in some cases it may not be possible to obtain bipartite graphs with clear connected components. The latter employs the crisp partitioning to make each data point belong to a cluster. In contrast to crisp partitioning, fuzzy clustering methods (Zhang et al. 2024) can represent the degree of ambiguity or certainty with which a data point belongs to a cluster. In practical scenarios, the class attributes of most objects are inherently ambiguous, making fuzzy clustering techniques a more accurate reflection of the real world. Consequently, many researchers have chosen to explore and advance this technique.

However, in order to obtain the membership matrix or avoid trivial solutions, some regularization terms or hyper-parameters are introduced (Jiang and Gao 2022; Hu et al. 2024; Yang et al. 2024). Considering these parameters do not have practical interpretations, how to tactfully determine them is a severe challenge. A common way is to search these

*These authors contributed equally.

[†]Corresponding author.

Copyright © 2025, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

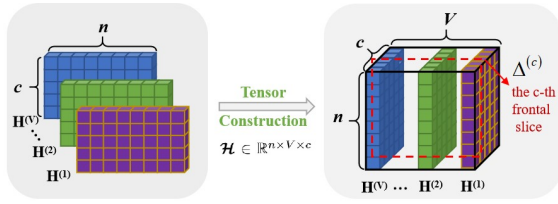


Figure 1: Construction of 3rd-order tensor

parameters in a wide range for each dataset, which weakens the model generalization ability. Therefore, designing a fuzzy clustering model with less parameters is necessary.

To this end, we design a Tensorized Label Learning based Fast Fuzzy Clustering (TLLFFC) model to solve the above problems simultaneously. First, in order to cope with large-scale multi-view datasets and avoid the post-process, we utilize anchor and label transmission strategies. Second, we design a novel balanced regularization term which, combined with prior anchor graph information, can avoid trivial solutions and achieve fuzzy clustering without additional regularization parameters. Finally, in order to explore the complementary information of different views better, inspired by the tensor Schatten p -norm (Li et al. 2023, 2024), we use the soft labels of each view to form a 3rd-order tensor as illustrated in Figure 1, and exploit tensor Schatten p -norm regularizer on this term. This helps to get the complementary information embedded in the soft labels of different views. The main contributions are summarized as follows

- A novel balanced regularization term is proposed to avoid trivial solutions without additional regularization terms. Thus, TLLFFC has no pressure of tuning regularization parameters w.r.t. trivial solutions or fuzziness.
- The label transmission strategy with the anchor graph transmits the label information from samples to anchors, which makes TLLFFC suitable for large-scale problems and can mine the intrinsic clustering structure between samples and anchors.
- We leverage the tensor Schatten p -norm regularization to fully exploit the structural and complementary information among different views. This ensures that labels of samples from different views are more likely to align, thereby improving clustering performance.
- We provide an efficient and effective optimization algorithm for the proposed model. Extensive experiments on different datasets show the superiority of proposed TLLFFC.

Notations

$\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times n}$ denotes the sample data of v -views, $x_{ij}^{(v)}$ is (i, j) -element of $\mathbf{X}^{(v)}$, $\mathbf{x}_j^{(v)}$ and $\mathbf{x}^{(v)i}$ are the j th column and i th row of $\mathbf{X}^{(v)}$ respectively. $\mathbf{x}_j^{(v)\top}$ and $\mathbf{X}^{(v)\top}$ are the transpose of $\mathbf{x}_j^{(v)}$ and $\mathbf{X}^{(v)}$ respectively. $\mathbf{1}$ is a column vector of all ones. $\mathbf{H}^{(v)} \in \mathbb{R}^{n \times c}$ is a membership matrix, and $h_{ij}^{(v)}$ represents the membership of the i th sample to the j th class,

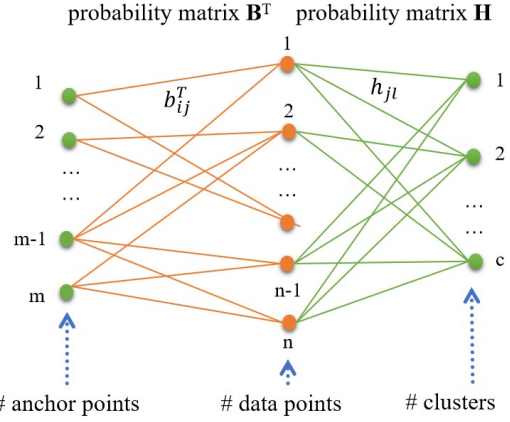


Figure 2: The illustration of probability matrix \mathbf{B} and \mathbf{H} .

satisfying $\sum_{j=1}^c h_{ij}^{(v)} = 1$. We use bold calligraphy letters for 3rd-order tensors, $\mathcal{H} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Moreover, the i th frontal slice of \mathcal{H} is $\mathbf{H}^{(i)}$. $\overline{\mathcal{H}}$ is the discrete Fourier transform of \mathcal{H} along the third dimension, $\overline{\mathcal{H}} = \text{fft}(\mathcal{H}, [], 3)$. Thus, $\mathcal{H} = \text{ifft}(\overline{\mathcal{H}}, [], 3)$. Besides, all boldface uppercase letters represent matrices, all boldface lowercase letters represent vectors.

Definition 1 (Tensor Schatten p -norm (Gao et al. 2021)) Given $\mathcal{H} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, $h = \min(n_1, n_2)$, the tensor Schatten p -norm of \mathcal{H} is

$$\|\mathcal{H}\|_{\mathfrak{S}} = \left(\sum_{i=1}^{n_3} \|\overline{\mathcal{H}}^{(i)}\|_{\mathfrak{S}}^p \right)^{\frac{1}{p}} = \left(\sum_{i=1}^{n_3} \sum_{j=1}^h \sigma_j^p(\overline{\mathcal{H}}^{(i)}) \right)^{\frac{1}{p}}, \quad (1)$$

where $\sigma_j(\overline{\mathcal{H}}^{(i)})$ denotes the j th singular value of $\overline{\mathcal{H}}^{(i)}$.

Methodology

Label Transmission Strategy

The anchor graph $\mathbf{B} \in \mathbb{R}^{n \times m}$ shows the prior probability knowledge between n samples and m anchors. Given the label of samples $\mathbf{H} \in \mathbb{R}^{n \times c}$, the label of anchors $\mathbf{F} \in \mathbb{R}^{m \times c}$ can be obtained by the Label Transmission strategy (Liu, He, and Chang 2010), i.e., $\mathbf{F} = \mathbf{B}^T \mathbf{H}$. The label of anchors can be decided by the label of its neighboring nodes, i.e., $f_{il} = \sum_{j=1}^n b_{ij}^T h_{jl}$ ($i = 1, 2, \dots, m$) ($l = 1, 2, \dots, c$). The greater the similarity to the node, the greater the influence weight of its neighboring nodes on its label. Figure 2 shows the illustration of \mathbf{B} and \mathbf{H} .

Balanced Clustering

If we only require that the elements of the anchor label matrix \mathbf{F} are non-negative, and each row sums to one. Then, the solution is not unique, and it cannot guarantee the balance of anchor across clusters. In order to avoid the trivial solution and obtain a clear clustering partition, our goal is to make the degree of each sample's membership to each class vary as much as possible. Therefore, some researchers propose the

following clustering model

$$\max_{\mathbf{F}\mathbf{1}=\mathbf{1}, f_{ij} \geq 0} \text{Tr}(\mathbf{F}^T \mathbf{F}) = \sum_{j=1}^c \sum_{i=1}^n f_{ij}^2. \quad (2)$$

Obviously, as $\mathbf{F}\mathbf{1} = \mathbf{1}$, the model (2) tends to make the maximum value of each row of \mathbf{F} be 1. That is, it tends to make the probability that each data point belongs to a certain category as large as possible, in short, it has a clear clustering partition. However, this problem has a trivial solution $\mathbf{F}^{(1)}$ in which all data points are divided into one category.

Furthermore, in order to avoid this trivial solution, some researchers always add additional regularization terms which increase the pressure of tuning parameters. On the contrary, we design a novel non-parameter clustering model (3) that can avoid the above trivial solution

$$\max_{\mathbf{F}\mathbf{1}=\mathbf{1}, \mathbf{F} \geq \mathbf{0}} \|\mathbf{F}^T\|_{2,1} = \max_{\mathbf{F}\mathbf{1}=\mathbf{1}, \mathbf{F} \geq \mathbf{0}} \sum_{j=1}^c \sqrt{\mathbf{f}_j^T \mathbf{f}_j}. \quad (3)$$

Obviously, Model (3) can avoid **trivial solution** $\mathbf{F}^{(1)}$, its optimal solution is $\mathbf{F}^{(2)}$, a discrete indicator matrix, where, each sample belongs to only one category and the number of samples in each class is the same. The objective values of model (3) about two label matrices are as follows

$$\begin{aligned} \text{obj}(\mathbf{F}^{(1)}) &= \sqrt{n}; \\ \text{obj}(\mathbf{F}^{(2)}) &= \sum_{i=1}^c \sqrt{n_i}. \end{aligned} \quad (4)$$

It is clear that the inequality $\sum_{i=1}^c \sqrt{n_i} \geq \sqrt{\sum_{i=1}^c n_i}$ holds when $n_i \geq 0$. Thus, $\mathbf{F}^{(1)}$ is not optimal solution of model (3), which can avoid trivial solution without additional regularization terms. Next, Theorem 1 gives a further analysis of model (3).

Theorem 1 *The following problem arrives its maximum value when \mathbf{F} is the discrete indicator matrix and $n_j = \frac{n}{c}$.*

$$\max_{\mathbf{F}\mathbf{1}=\mathbf{1}, \mathbf{F} \geq \mathbf{0}} \|\mathbf{F}^T\|_{2,1} = \max_{\mathbf{F}\mathbf{1}=\mathbf{1}, \mathbf{F} \geq \mathbf{0}} \sum_{j=1}^c \sqrt{\mathbf{f}_j^T \mathbf{f}_j}.$$

Proof 1 *Obviously, when \mathbf{F} is the discrete indicator matrix, we have $\sum_{j=1}^c \sqrt{\mathbf{f}_j^T \mathbf{f}_j} = \sum_{j=1}^c \sqrt{n_j}$, where n_j is the number of samples belonging to the j th category, $\sum_{j=1}^c n_j = n$. According to Cauchy–Schwarz inequality (Bhatia and Davis 1995), we have*

$$\|\mathbf{F}^T\|_{2,1}^2 = \left(\sum_{j=1}^c \sqrt{\mathbf{f}_j^T \mathbf{f}_j} \right)^2 = \left(\sum_{j=1}^c \sqrt{n_j} \right)^2$$

$$\leq \sum_{j=1}^c n_j \sum_{j=1}^c 1 = cn,$$

the two sides are equal if and only if $\sqrt{n_1} = \sqrt{n_2} = \dots = \sqrt{n_c}$, i.e., $n_j = \frac{n}{c}$. At this time, $\sum_{j=1}^c \sqrt{\mathbf{f}_j^T \mathbf{f}_j}$ reaches its maximum value.

The Proposed Model

Therefore, motivated by the above, we can tend this idea to the multi-view field and propose a parameter-free model (the first term of model (6)) to avoid trivial solution, $\mathbf{H}^{(v)} \in \mathbb{R}^{n \times c}$ is the label matrix of samples in the v th view. Furthermore, in order to harness the complementary information and spatial structure across views, we employ the tensor Schatten p -norm constraint on the tensorial form of $\mathbf{H}^{(v)}$ to ensure that the labels for samples within each view remain consistent. The objective function is

$$\max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} \sum_{v=1}^V \|\mathbf{H}^{(v)T} \mathbf{B}^{(v)}\|_{2,1} - \lambda \|\mathcal{H}\|_{\mathfrak{S}}^p, \quad (5)$$

According to the definition of the $\ell_{2,1}$ -norm, the objective function can be written as:

$$\max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} \sum_{v=1}^V \sum_{j=1}^c \sqrt{\mathbf{h}_j^{(v)T} \mathbf{B}^{(v)} \mathbf{B}^{(v)T} \mathbf{h}_j^{(v)}} - \lambda \|\mathcal{H}\|_{\mathfrak{S}}^p, \quad (6)$$

where, $0 < p \leq 1$, λ is the hyper-parameter of the Schatten p -norm term. $\mathbf{H}^{(v)}$ represents the v th lateral slice of tensor $\mathcal{H} \in \mathbb{R}^{n \times V \times c}$. The tensor construction process is illustrated in Figure 1.

Remark 1 *The regularizer in the model (6) is used to explore the complementary information embedded in inter-views cluster assignment matrices $\mathbf{H}^{(v)}$ ($v = 1, 2, \dots, V$). The c -th frontal slice $\Delta^{(c)}$ in Figure 1 describes the similarity between n sample points and the c clusters in different views. The ideal label matrix $\mathbf{H}^{(v)}$ is consistent in different views. Since different views usually show different cluster structures, imposing tensor Schatten p -norm minimization (Gao et al. 2021; Xia et al. 2023) constraint on \mathcal{H} can make sure each $\Delta^{(c)}$ has spatial low-rank structure. Thus, $\Delta^{(c)}$ can well characterize the complementary information embedded in inter-views.*

Optimization

We employ the Augmented Lagrange Multiplier (ALM) method to address problem (6). Introduce the auxiliary variable \mathcal{J} and set $\mathcal{H} = \mathcal{J}$, we have

$$\begin{aligned} \max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}, \mathcal{J}} \sum_{v=1}^V \sum_{j=1}^c \sqrt{\mathbf{h}_j^{(v)T} \mathbf{B}^{(v)} \mathbf{B}^{(v)T} \mathbf{h}_j^{(v)}} - \lambda \|\mathcal{J}\|_{\mathfrak{S}}^p \\ - \frac{\mu}{2} \|\mathcal{H} - \mathcal{J} + \frac{\mathcal{Q}}{\mu}\|_F^2, \end{aligned} \quad (7)$$

where \mathcal{Q} represents the Lagrange multiplier, μ is the penalty parameter. The optimization process can be separated into two steps:

(1) **Solve \mathcal{H} with fixed \mathcal{J} .** (7) becomes

$$\begin{aligned} \max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} \sum_{v=1}^V \sum_{j=1}^c \sqrt{\mathbf{h}_j^{(v)T} \mathbf{B}^{(v)} \mathbf{B}^{(v)T} \mathbf{h}_j^{(v)}} \\ - \frac{\mu}{2} \|\mathcal{H} - \mathcal{J} + \frac{\mathcal{Q}}{\mu}\|_F^2. \end{aligned} \quad (8)$$

Algorithm 1: Solving problem (6)

Input: Data matrices $\{\mathbf{X}^{(v)}\}_{v=1}^V \in \mathbb{R}^{d_v \times n}$, the number of anchors m , the number of classes c .

Parameter: $\mu = 10^{-3}$, $\eta = 1.1$, $max\text{-}\mu = 10^9$, λ .

Output: Cluster assignment matrix $\mathbf{H} \in \mathbb{R}^{n \times c}$ of each data points.

- 1: Initialize $\mathbf{Q} = \mathcal{J} = \mathbf{0}$, $\mathbf{H}^{(v)}$ is an identity matrix. Select anchor points and construct anchor graphs $\mathbf{B}^{(v)}$ ($v = 1, 2, \dots, V$) by (Xia et al. 2023).
 - 2: **while** not converge **do**
 - 3: Update \mathcal{J} by (18);
 - 4: Update $\mathbf{H}^{(v)}$ by solving problem (9);
 - 5: Update \mathbf{Q} and μ : $\mathbf{Q} = \mathbf{Q} + \mu(\mathcal{H} - \mathcal{J})$, $\mu = \eta\mu$;
 - 6: **end while**
 - 7: Directly achieve c clusters based on the cluster assignment matrix $\mathbf{H} = \sum_{v=1}^V \mathbf{H}^{(v)}$.
 - 8: **return** Clustering results.
-

Since all $\mathbf{H}^{(v)}$ ($v = 1, 2, \dots, V$) are independent, then (8) can be decomposed into V independent sub-optimization problems. Then, the optimization problem of solving $\mathbf{H}^{(v)}$ is

$$\max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} \sum_{j=1}^c \sqrt{\mathbf{h}_j^{(v)\top} \mathbf{B}^{(v)} \mathbf{B}^{(v)\top} \mathbf{h}_j^{(v)}} - \frac{\mu}{2} \|\mathbf{H}^{(v)} - \mathbf{G}^{(v)}\|_F^2, \quad (9)$$

where, $\mathbf{G}^{(v)} = \mathbf{J}^{(v)} - \frac{\mathbf{Q}^{(v)}}{\mu}$. As $\sqrt{\mathbf{h}_j^{(v)\top} \mathbf{B}^{(v)} \mathbf{B}^{(v)\top} \mathbf{h}_j^{(v)}}$ is convex w.r.t. $\mathbf{h}_j^{(v)}$, we use Iteratively Re-Weighted (IRW) (Nie et al. 2016) algorithm to solve problem (9) iteratively, where, the intermediate variable $\mathbf{D}^{(v)}$ is

$$\mathbf{d}_j^{(v)} = \frac{\mathbf{B}^{(v)} \mathbf{B}^{(v)\top} \mathbf{h}_j^{(v)}}{\sqrt{\mathbf{h}_j^{(v)\top} \mathbf{B}^{(v)} \mathbf{B}^{(v)\top} \mathbf{h}_j^{(v)}}}. \quad (10)$$

Therefore, $\mathbf{H}^{(v)}$ is updated by solving the following problem

$$\max_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} Tr(\mathbf{D}^{(v)\top} \mathbf{H}^{(v)}) - \frac{\mu}{2} \|\mathbf{H}^{(v)} - \mathbf{G}^{(v)}\|_F^2. \quad (11)$$

Let $\mathbf{A}^{(v)} = \mathbf{D}^{(v)} + \mu \mathbf{G}^{(v)}$, the above problem is

$$\min_{\mathbf{H}^{(v)} \geq \mathbf{0}, \mathbf{H}^{(v)} \mathbf{1} = \mathbf{1}} Tr(\mathbf{H}^{(v)\top} \mathbf{H}^{(v)}) - \frac{2}{\mu} Tr(\mathbf{A}^{(v)\top} \mathbf{H}^{(v)}). \quad (12)$$

As each row of $\mathbf{H}^{(v)}$ is independent, solving (12) is equivalent to solving the following n problems

$$\min_{\mathbf{h}^{(v) i} \geq 0, \mathbf{h}^{(v) i} \mathbf{1} = 1} \|\mathbf{h}^{(v) i} - \frac{1}{\mu} \mathbf{a}^{(v) i}\|_2^2, \quad (13)$$

where, $\mathbf{h}^{(v) i}$ denotes the i th row of $\mathbf{H}^{(v)}$. Problem (13) can be solved by an efficient iterative method (Huang, Nie, and Huang 2015). Therefore, problem (9) can be solved by updating $\mathbf{D}^{(v)}$ by Eq. (10) and updating $\mathbf{H}^{(v)}$ by solving problem (13) iteratively.

(2) Solve \mathcal{J} with fixed \mathcal{H} . (7) becomes

$$\min_{\mathcal{J}} \frac{\lambda}{\mu} \|\mathcal{J}\|_{\mathcal{S}}^p + \frac{1}{2} \|\mathcal{H} - \mathcal{J} + \frac{\mathbf{Q}}{\mu}\|_F^2. \quad (14)$$

After completing the square regarding \mathcal{J} , we can deduce

$$\mathcal{J}^* = \arg \min_{\mathcal{J}} \frac{\lambda}{\mu} \|\mathcal{J}\|_{\mathcal{S}}^p + \frac{1}{2} \|\mathcal{H} - \mathcal{J} + \frac{\mathbf{Q}}{\mu}\|_F^2, \quad (15)$$

which has a closed-form solution as Theorem 2 (Gao et al. 2021):

Theorem 2 Let $\mathcal{S} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ have the t -SVD $\mathcal{S} = \mathbf{U} * \mathcal{A} * \mathbf{V}^T$. For the problem

$$\min_{\mathcal{X}} \frac{1}{2} \|\mathcal{X} - \mathcal{S}\|_F^2 + \tau \|\mathcal{X}\|_{\mathcal{S}}^p, \quad (16)$$

the optimal solution is

$$\mathcal{X}^* = \Gamma_{\tau}(\mathcal{S}) = \mathbf{U} * \text{ifft}(\mathbf{P}_{\tau}(\overline{\mathcal{S}})) * \mathbf{V}^T, \quad (17)$$

where $\mathbf{P}_{\tau}(\overline{\mathcal{S}}) \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is a f -diagonal tensor, whose diagonal elements can be obtained by the General Shrinkage Thresholding algorithm (Gao et al. 2021).

Therefore, the solution of (15) is

$$\mathcal{J}^* = \Gamma_{\frac{\lambda}{\mu}}(\mathcal{H} + \frac{\mathbf{Q}}{\mu}). \quad (18)$$

Algorithm 1 summaries the details.

Time and Space Complexity Analysis

Time Complexity: TLLFFC consists of two stages: 1) Construction of $\{\mathbf{B}^{(v)}\}_{v=1}^V$, needs $\mathcal{O}(Vnmd + Vnm \log(m))$, where $d = \sum_{v=1}^V d_v$; 2) Iterative updating \mathcal{J} and $\{\mathbf{H}^{(v)}\}_{v=1}^V$, needs $\mathcal{O}(Vnc \log(Vn) + V^2nc)$ and $\mathcal{O}(Vnc + Vm^2c)$ respectively. V, n, m, d_v, c are the number of views, samples, anchors, features and classes, respectively. Given that m, c, V are relatively small, the overall time complexity of our proposed method is $\mathcal{O}(Vnmd + Vm^2c)$.

Space Complexity: The storage memories for $\{\mathbf{B}^{(v)}\}_{v=1}^V$, \mathcal{J} , \mathcal{H} and \mathbf{Q} needs $\mathcal{O}(Vnm)$, $\mathcal{O}(Vnc)$, $\mathcal{O}(Vnc)$ and $\mathcal{O}(Vnc)$, respectively. The overall space complexity of TLLFFC is $\mathcal{O}(Vnm + 3Vnc)$.

Experiments

Datasets, Compared Methods and Metrics

Datasets: Our experiments are executed on five widely-recognized datasets to validate our model's effectiveness. Table 1 gives a brief description of these datasets. All experiments are implemented on a standard Windows 10 Server with two Intel (R) Xeon (R) Gold 6230 CPUs 2.1 GHz and 128 GB RAM, MATLAB R2020a.

- MSRC-v5(MSRC) (Winn and Jovic 2005) includes 7 types of objects with a total of 210 images. We selected CM, HOG, GIST, LBP, and CENT features as five different views.
- HandWritten (Asuncion, Newman et al. 2007) includes 10 digits with a total of 2,000 images. We selected FOU, FAC, ZER, and MOR features as four different views.

Datasets	# of samples	# of views	# of classes	# of features
MSRC	210	5	7	24,576,512,256,254
HandWritten	2000	4	10	76,216,47,6
MNIST	4000	3	4	30,9,30
Scene	4485	3	15	1800,1180,1240
Reuters	18758	5	6	21531,24892,34251,15506,11547

Table 1: Multi-view Datasets

Datasets	MSRC			HandWritten			MNIST			Scene		
	m $0.7n$	p 0.5	λ 50	m n	p 0.3	λ 1000	m $0.3n$	p 0.9	λ 100	m $0.9n$	p 0.4	λ 500
Methods	ACC	NMI	Purity	ACC	NMI	Purity	ACC	NMI	Purity	ACC	NMI	Purity
CSMSC	0.862	0.767	0.862	0.806	0.793	0.867	0.643	0.601	0.728	0.576	0.574	0.629
GMC	0.895	0.809	0.895	0.879	0.882	0.879	0.921	0.807	0.921	0.409	0.430	0.417
ETLMSC	0.962	0.937	0.962	0.938	0.894	0.938	0.934	0.847	0.934	0.218	0.166	0.221
LMVSC	0.814	0.717	0.814	0.904	0.831	0.904	0.892	0.726	0.892	0.561	0.512	0.581
FMCNOF	0.713	0.648	0.714	0.541	0.484	0.54	0.686	0.513	0.695	0.263	0.249	0.258
SFMC	0.809	0.721	0.781	0.853	0.871	0.873	0.917	0.801	0.917	0.343	0.522	0.346
FPMVS-CAG	0.843	0.738	0.843	0.850	0.787	0.850	0.887	0.719	0.887	0.618	0.566	0.651
OMVFC-LICAG	0.728	0.590	0.738	0.837	0.798	0.838	0.888	0.725	0.888	0.484	0.517	0.537
Orth-NTF	0.990	0.978	0.990	0.985	0.969	0.985	0.977	0.926	0.977	0.758	0.804	0.759
TLL-AG	0.986	0.968	0.986	0.982	0.962	0.982	0.981	0.935	0.981	0.719	0.785	0.752
TLLFFC	1	1	1	0.999	0.996	0.999	1	1	1	0.840	0.874	0.870

Table 2: Clustering performance

Datasets	Reuters		
	m $0.001n$	p 0.4	λ 10
Methods	ACC	NMI	Purity
CSMSC	OM	OM	OM
GMC	-	-	-
ETLMSC	OM	OM	OM
LMVSC	0.589	0.335	0.615
FMCNOF	0.343	0.125	0.358
SFMC	0.602	0.354	0.604
FPMVS-CAG	0.526	0.323	0.603
OMVFC-LICAG	0.426	0.087	0.431
Orth-NTF	0.694	0.686	0.809
TLL-AG	0.686	0.554	0.740
TLLFFC	0.808	0.697	0.810

Table 3: Clustering performance. 'OM' means out of memory. '-' means it takes more than 3 hours

- Mnist4 (Deng 2012) includes four categories of handwritten digits from 0 to 3, totaling 4,000 images. We used ISO, LDA, and NPE features as three different views.
- Scene15 (Oliva and Torralba 2001) comprises 15 categories of natural scenes encompassing indoor and outdoor environments. We selected PHOW, LBP, and CENTRIST as three different views.
- Reuters (Apté, Damerou, and Weiss 1994) consists of document data with six categories, totaling 18,758 samples. We selected five different views for each category,

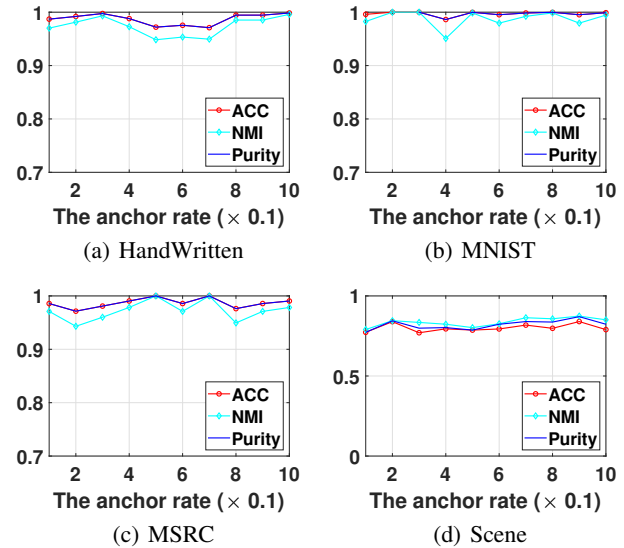


Figure 3: Clustering performance w.r.t. m .

described by different languages and their corresponding translations.

Compared Methods: We choose the following 10 state-of-art multi-view clustering algorithms to compare with our proposed methods: CSMSC (Luo et al. 2018), GMC (Wang, Yang, and Liu 2019), ETLMSC (Wu, Lin, and Zha 2019), LMVSC (Kang et al. 2020), FMCNOF (Yang et al. 2020),

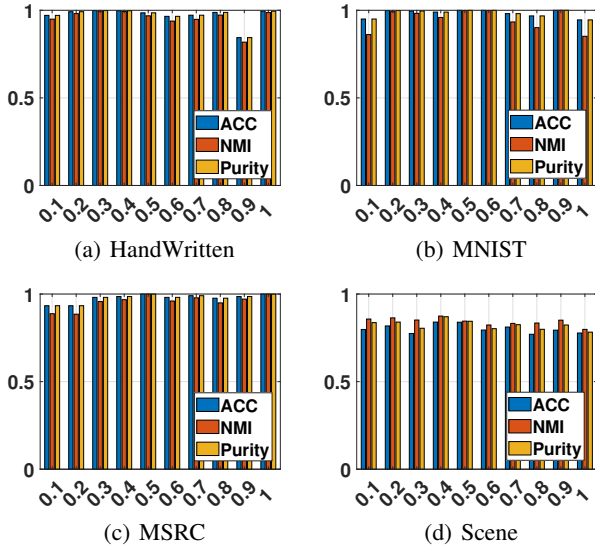


Figure 4: Clustering performance w.r.t. p .

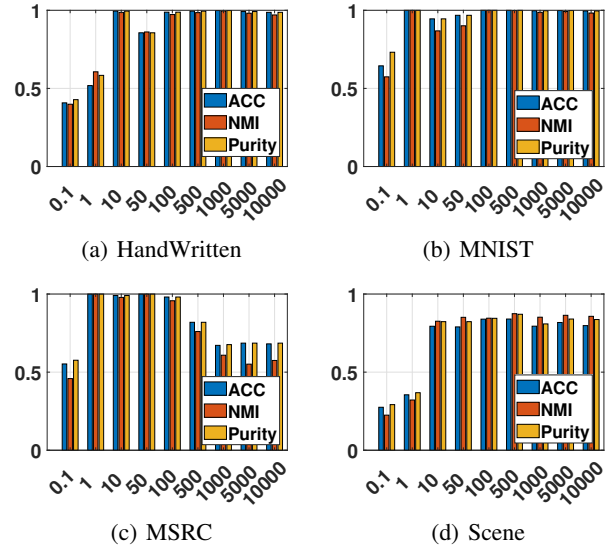


Figure 5: Clustering performance w.r.t. λ .

SFMC (Li et al. 2020), FPMVS-CAG (Wang et al. 2021), Orth-NTF (Li et al. 2023), OMVFC-LICAG (Zhang et al. 2024), TLL-AG (Li et al. 2024).

Metrics: We employ three criteria to comprehensively measure the clustering quality including Accuracy (ACC), Normalized Mutual Information (NMI) and Purity. These criteria measure the consistency between real labels and predicted labels from different viewpoints. They are between 0 and 1, and the larger they are, the better the performance is.

Experimental Results

The experimental results are delineated in Tables 2 and 3 along with the hyper-parameters setting on five datasets, the best results are highlighted in boldface. From these results, we found that our TLLFFC significantly competing methods across all datasets, i.e.,

(1) Compared with CSMSC, GMC, and ETLMSC, which have higher time and space complexity leading to time and memory overload when processing large-scale datasets, TLLFFC utilizes an anchor strategy to reduce computational demands. This enables it to efficiently manage and produce impressive results on large-scale datasets.

(2) Compared with discrete and fuzzy anchor graph methods, LMVSC, FMCNOF, SFMC, FPMVS, OMVFC-LICAG, where, LMVSC and FPMVS are two-step methods, TLLFFC performs better. As TLLFFC with the tensor constraint can extract complementary information from different perspectives, ensuring that the cluster indicator matrices across various views align as closely as possible. Besides, TLLFFC can obtain fuzzy clustering results with one-step.

(3) For Orth-NTF, TLL-AG and our method TLLFFC, which have tensor constraints, it is clear that these are the first three best methods, which demonstrate the necessity of the tensor constraint further. However, TLLFFC is better among them. As Orth-NTF and TLL-AG are NMF-based methods, our proposed fuzzy method TLLFFC with label transmission

strategy can mine the intrinsic clustering structure between samples and anchors.

Parameter Analysis

- **parameter m :** We investigate the clustering performance w.r.t. the number of anchors m , the results are illustrated in Figure 3. The abscissa is the anchor rate, the ratio of anchors to the total number of samples, varies from 0.1 to 1. From it, we know that the clustering performance does not exhibit a consistent upward trend with increasing anchor rates. The quality of the anchor graph influences the clustering performance. Therefore, the construction of the anchor graph is critical.
- **parameter p :** p varies from 0.1 to 1, and the corresponding results are depicted in Figure 4. From it, it is clear that TLLFFC is insensitive to p . And the model performs better when $p < 1$ than $p = 1$. As when $p < 1$, the Schatten p -norm ensures that the tangent plane of the tensor has a spatially low-rank structure, which can mine of complementary information among multiple views better.
- **parameter λ :** We tune λ over the set $\{0.1, 1, 10, 50, 100, 500, 1000, 5000, 10000\}$. Figure 5 shows that λ significantly affects the clustering results. Selecting extremely high or low values can degrade the clustering quality.

Convergence Analysis

Figure 6 gives the convergence curves with the clustering performance of TLLFFC. We monitor the changes in reconstruction errors (i.e., $\|\mathcal{H} - \mathcal{J}\|_\infty$) over increasing iterations. Notably, the errors exhibit significant fluctuations during the initial 80 iterations but stabilize. Our model typically achieves convergence after 100 iterations. At the same time, within the first 80 iterations, ACC fluctuates at lower levels owing to the fact that the model has not reached convergence at this time.

Datasets	MSRC			HandWritten			MNIST			scene		
Methods	ACC	NMI	Purity	ACC	NMI	Purity	ACC	NMI	Purity	ACC	NMI	Purity
w.o. \mathcal{H}	0.742	0.654	0.776	0.487	0.578	0.584	0.612	0.552	0.702	0.470	0.502	0.526
TLLFFC	1	1	1	0.999	0.996	0.999	1	1	1	0.840	0.874	0.870

Table 4: The result of ablation study

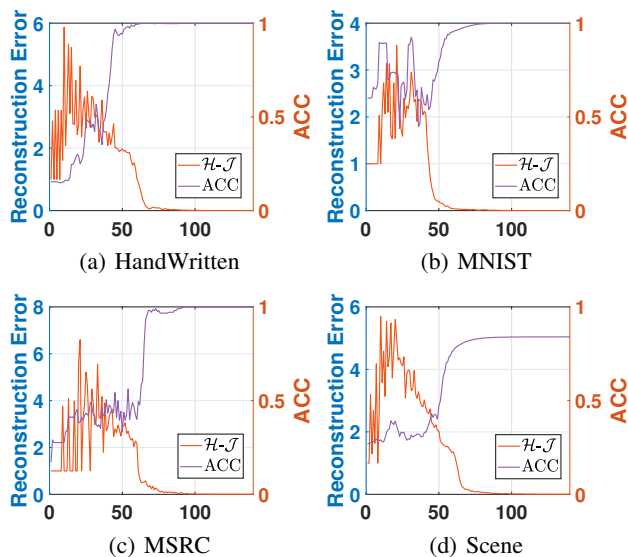


Figure 6: Clustering performance and convergence curves.

At about 100 iterations, ACC peaks and remains relatively stable.

Ablation Study

To examine the significance of the tensor regularization term, we perform ablation studies by omitting the Schatten p -norms. The outcomes are presented in Table 4. Observations reveal that relying solely on fusion of independent view results in significantly poorer clustering performance, with ACC over 25% below that of our proposed TLLFFC. Incorporating Schatten p -norms enhances the clustering outcome, with the term having a more pronounced impact. As utilizing tensors uncovers complementary information across views, yielding more precise clustering labels.

T-SNE Visualization

We show the T-SNE visualization results on the Handwritten dataset in Figure 7, with the number of iterations varies from 10 to 120. When the number of iterations is small, the majority of the samples belong to the same cluster, indicating poor clustering performance. However, as the number of iterations increases, the distinction between clusters gradually becomes more apparent. Samples belonging to the same cluster gather together, and the clustering accuracy steadily improves. At the 120th iteration, the model has converged, dividing the pieces into 10 distinct clusters. The clustering

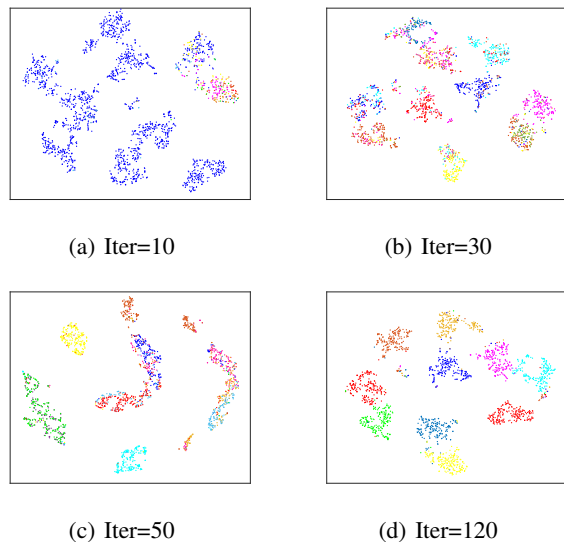


Figure 7: T-SNE visualization results on the Handwritten dataset with different numbers of iterations.

accuracy reaches 0.999, which provides further evidence of the model’s effectiveness.

Conclusions

This study proposes a novel fuzzy label learning model with multi-view graph clustering. To be specific, the proposed novel balanced regularization term can reduce the pressure of tuning more regularization parameters for fuzzy clustering. Introducing the label transmission and anchor graph strategies can deal with large-scale datasets. Introducing the tensor Schatten p -norm minimization, the complementary information and spatial structure embedded in data with different types can be well utilized to make the clustering results more reliable. Comprehensive experiments verify the superiority of TLLFFC.

Acknowledgements

This work is supported by the National Natural Science Foundation of China under Grants 62176203, the Postdoctoral Fellowship Program of CPSF under Grant Number GZC20241318, and the Open Projects Program of the State Key Laboratory of Multimodal Artificial Intelligence Systems under Grant Number MAIS2024114.

References

- Apté, C.; Damerau, F.; and Weiss, S. M. 1994. Automated Learning of Decision Rules for Text Categorization. *ACM Trans. Inf. Syst.*, 12(3): 233–251.
- Asuncion, A.; Newman, D.; et al. 2007. UCI machine learning repository.
- Bhatia, R.; and Davis, C. 1995. A Cauchy-Schwarz inequality for operators with applications. *Linear Algebra and its Applications*, 223-224: 119–129.
- Cui, C.; Ren, Y.; Pu, J.; Li, J.; Pu, X.; Wu, T.; Shi, Y.; and He, L. 2023. A Novel Approach for Effective Multi-View Clustering with Information-Theoretic Perspective. In Oh, A.; Naumann, T.; Globerson, A.; Saenko, K.; Hardt, M.; and Levine, S., eds., *Advances in Neural Information Processing Systems*, volume 36, 44847–44859. Curran Associates, Inc.
- Deng, L. 2012. The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]. *IEEE Signal Process. Mag.*, 29(6): 141–142.
- Fang, S.-G.; Huang, D.; Cai, X.-S.; Wang, C.-D.; He, C.; and Tang, Y. 2023. Efficient Multi-View Clustering via Unified and Discrete Bipartite Graph Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 1–12.
- Gao, Q.; Zhang, P.; Xia, W.; Xie, D.; Gao, X.; and Tao, D. 2021. Enhanced Tensor RPCA and its Application. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(6): 2133–2140.
- Gong, X.; Yuan, D.; and Bao, W. 2021a. Discriminative metric learning for partial label learning. *IEEE Transactions on Neural Networks and Learning Systems*, 34(8): 4428–4439.
- Gong, X.; Yuan, D.; and Bao, W. 2021b. Understanding partial multi-label learning via mutual information. *Advances in Neural Information Processing Systems*, 34: 4147–4156.
- Gong, X.; Yuan, D.; Bao, W.; and Luo, F. 2022. A unifying probabilistic framework for partially labeled data learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(7): 8036–8048.
- Hu, X.; Qin, J.; Shen, Y.; Pedrycz, W.; Liu, X.; and Liu, J. 2024. An Efficient Federated Multiview Fuzzy C-Means Clustering Method. *IEEE Transactions on Fuzzy Systems*, 32(4): 1886–1899.
- Huang, J.; Nie, F.; and Huang, H. 2015. A new simplex sparse learning model to measure data similarity for clustering. In *Proceedings of the International Joint Conference on Artificial Intelligence*, 3569–3575.
- Jiang, T.; and Gao, Q. 2022. Fast multiple graphs learning for multi-view clustering. *Neural Networks*, 155: 348–359.
- Kang, Z.; Zhou, W.; Zhao, Z.; Shao, J.; Han, M.; and Xu, Z. 2020. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 4412–4419.
- Li, J.; Gao, Q.; Wang, Q.; and Xia, W. 2024. Tensorized Label Learning on Anchor Graph. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 13537–13544.
- Li, J.; Gao, Q.; WANG, Q.; Yang, M.; and Xia, W. 2023. Orthogonal Non-negative Tensor Factorization based Multi-view Clustering. In *Advances in Neural Information Processing Systems*, volume 36, 18186–18202.
- Li, X.; Zhang, H.; Wang, R.; and Nie, F. 2020. Multiview clustering: A scalable and parameter-free bipartite graph fusion method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(1): 330–344.
- 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-10)*, 679–686. Citeseer.
- Liu, W.; and Tsang, I. 2015. On the optimality of classifier chain for multi-label classification. *Advances in Neural Information Processing Systems*, 28.
- Liu, W.; and Tsang, I. W. 2017. Making decision trees feasible in ultrahigh feature and label dimensions. *Journal of Machine Learning Research*, 18(81): 1–36.
- Luo, S.; Zhang, C.; Zhang, W.; and Cao, X. 2018. Consistent and specific multi-view subspace clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- Nie, F.; Wang, X.; Jordan, M.; and Huang, H. 2016. The constrained laplacian rank algorithm for graph-based clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.
- Nie, F.; Xue, J.; Wang, R.; Zhang, L.; and Li, X. 2022. Fast clustering by directly solving bipartite graph clustering problem. *IEEE Transactions on Neural Networks and Learning Systems*.
- 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.
- Qiang, Q.; Zhang, B.; Wang, F.; and Nie, F. 2023. Multi-View Discrete Clustering: A Concise Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(12): 15154–15170.
- 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 *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 10078–10086.
- Wang, H.; Yang, Y.; and Liu, B. 2019. GMC: Graph-based multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 32(6): 1116–1129.
- Wang, S.; Liu, X.; Zhu, X.; Zhang, P.; Zhang, Y.; Gao, F.; and Zhu, E. 2021. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Transactions on Image Processing*, 31: 556–568.
- Winn, J. M.; and Jojic, N. 2005. LOCUS: Learning Object Classes with Unsupervised Segmentation. In *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*, 756–763. IEEE Computer Society.
- Wu, J.; Lin, Z.; and Zha, H. 2019. Essential tensor learning for multi-view spectral clustering. *IEEE Transactions on Image Processing*, 28(12): 5910–5922.

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.

Xue, J.; Nie, F.; Liu, C.; Wang, R.; and Li, X. 2024. Co-Clustering by Directly Solving Bipartite Spectral Graph Partitioning. *IEEE Transactions on Cybernetics*.

Yang, B.; Zhang, X.; Nie, F.; Wang, F.; Yu, W.; and Wang, R. 2020. Fast multi-view clustering via nonnegative and orthogonal factorization. *IEEE Transactions on Image Processing*, 30: 2575–2586.

Yang, H.; Deng, Z.; Zhang, W.; Wu, Q.; Choi, K.-S.; and Wang, S. 2024. End-to-End Multiview Fuzzy Clustering With Double Representation Learning and Visible-Hidden View Cooperation. *IEEE Transactions on Fuzzy Systems*, 32(2): 483–497.

Zhang, C.; Chen, L.; Shi, Z.; and Ding, W. 2024. Latent information-guided one-step multi-view fuzzy clustering based on cross-view anchor graph. *Information Fusion*, 102: 102025.