Robustness Meets Low-Rankness: Unified Entropy and Tensor Learning for Multi-view Subspace Clustering

Shuqin Wang, Yongyong Chen, *Member, IEEE,* Zhiping Lin, *Senior Member, IEEE,* Yigang Cen and Qi Cao

***Abstract*—In this paper, we develop the weighted error entropy-regularized tensor learning method for multi-view subspace clustering (WETMSC), which integrates the noise disturbance removal and subspace structure discovery into one unified framework. Unlike most existing methods which focus only on the affinity matrix learning for the subspace discovery by different optimization models and simply assume that the noise is independent and identically distributed (i.i.d.), our WETMSC method adopts the weighted error entropy to characterize the underlying noise by assuming that noise is independent and piecewise identically distributed (i.p.i.d.). Meanwhile, WETMSC constructs the self-representation tensor by storing all self-representation matrices from the view dimension, preserving high-order correlation of views based on the tensor nuclear norm. To solve the proposed nonconvex optimization method, we design a half-quadratic (HQ) additive optimization technology and iteratively solve all subproblems under the alternating direction method of multipliers framework. Extensive comparison studies with state-of-the-art clustering methods on real-world datasets and synthetic noisy datasets demonstrate the ascendancy of the proposed WETMSC method.**

***Index Terms*—Multi-view Subspace Clustering, Tensor Learning, Weighted Error Entropy**

I. Introduction

With the development of information technology, multimedia data can be described by multiple views. Resorting to the comprehensive information of multi-view features, multi-view learning has improved multimedia data (such as images, videos and documents) analysis performance and has been widely used in clustering, classification etc. [1], [2]. Multi-view clustering (MVC) [3]–[6] learns the correlation of multiple views in an unsupervised manner to divide multimedia data into different clusters [7]. However, the high-dimensional multimedia data in real-world applications increases the computational complexity. Currently, multi-view subspace clustering (MVSC) [8]–[10] is widely exploited by assuming that high-dimensional multimedia data is located in several low-dimensional latent subspaces. These low-dimensional subspaces are learned by using original multimedia data as a dictionary to preserve the inherent structure of self-representation matrices.

Subspace structure discovery and noise disturbance removal are the two sides of a coin and they help each other in a mutual reinforcement manner. However, most existing methods pay attention only to subspace structure discovery by different prior-based optimization models, such as sparse prior termed sparse subspace clustering (SSC) [11], low-rankness termed low-rank representation (LRR) [12] or block diagonal representation termed least squares regression (LSR) [13]. Specifically, SSC used l_1 norm to constrain self-representation matrix and error matrix by assuming that the noise is sparse and obeys the Laplacian distribution. LRR used nuclear norm to constrain self-representation matrix and $l_{2,1}$ norm on error matrix by assuming that the sample-specific noise is dense and obeys the Gaussian distribution. LSR used Frobenius norm to constrain self-representation matrix and error matrix by assuming that the noise obeys the Gaussian distribution. Along the above line, reference [14] based on LSR to learn self-representation matrices separately and then obtained the averaged affinity matrix. Zhang *et al*. [15] pointed out that consistent and complementary information among multiple views contributes differently to the described multimedia data. To merge the different information of all views, the tensorized MVSC methods [16]–[19] have been proposed to serve as the advanced variants of matrix-dimensional LRR methods. They usually constrain complementary and consistent information by third-order tensor in a holistic manner but not independently [20]. In addition, several technologies including hypergraph [17], anchor graph [14], kernel trick [21], [22], joint learning [23]–[25] have been integrated with MVSC. However, they focus on the graph construction and thus pay much less attention to the noise disturbance removal.

In practice, noise is often complex and may contain multiple distributions leading to multimedia data degradation [26], [27]. Most existing MVSC methods [16], [28] simply assume that the noise obeys a single distribution, such as Gaussian or Laplacian, without taking the corruption of complex noise into consideration. The l_1 norm imposes the Laplacian distribution

This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFE0110500, in part by the National Natural Science Foundation of China under Grant 62062021 and 62106063, in part by the Guangdong Natural Science Foundation under Grant 2022A1515010819, and in part by the Shenzhen Science and Technology Program under Grant RCBS20210609103708013. (Corresponding authors: Yigang Cen and Qi Cao.)

S. Wang and Y. Cen are with the Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China, and also with the Beijing Key Laboratory of Advanced Information Science and Network Technology, Beijing 100044, China (Email: ShuqinWang.cn@hotmail.com; ygcen@bjtu.edu.cn).

Y. Chen is with the School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China (Email: YongyongChen.cn@gmail.com).

Z. Lin is with the School of Electrical and Electronic Engineering, Nanyang Technological University, 639798, Singapore (Email: ezplin@ntu.edu.sg).

Q. Cao is with the School of Computing Science, University of Glasgow, 567739, Singapore (Email: qi.cao@glasgow.ac.uk).

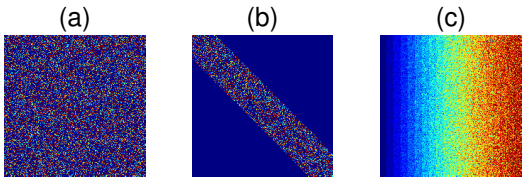

Fig. 1. (a) Gaussian random noise. (b) Diagonal structure noise. (c) Simulated illumination noise.

assumption to characterize sparse noise and is sensitive to dense noise. The mean squared error (MSE, l_2 norm) and $l_{2,1}$ norm aim to eliminate the Gaussian distribution noise but cannot handle complex dense noise and outliers. To deal with mixed noise, the combination of G_1 norm and $l_{2,1}$ norm [29], the mixture of Gaussian constraint [30] have been used. Although the disturbances of different noise combinations are suppressed to some extent, they still rely mostly on the Gaussian assumption and are invalid for non-Gaussian noise and outliers.

Different from the above research lines which either focus only on the affinity matrix learning for the subspace discovery or simply assume that noise obeys the Gaussian or Laplacian distribution, we consider the noise disturbance removal and subspace structure discovery in a holistic manner. In this paper, we propose the weighted error entropy-regularized tensor learning method for MVSC (WETMSC) to learn self-representation tensor based on weighted error entropy within the subspace clustering framework. Error entropy is an extension of the traditional metric method in information theoretic learning (ITL), which has been successfully applied to signal processing due to its robustness to non-Gaussian noise [31], [32]. For example, Erdogmus *et al*. [33] proposed the correntropy concept, in which the correntropy-induced metric (CIM) is applied to MVC in [34]. Correntropy is essentially equivalent to error entropy, both of which are derived from Rényi's quadratic entropy. From the perspective of information geometry, minimizing error entropy can be simply understood as minimizing the probability distribution between the desired and system outputs [33]. However, most existing entropy-based methods still assume that the noise is independent and identically distributed (i.i.d.), i.e. they treat the entries of error term independently, ignoring their structural information. Corresponding to ITL, the entropy of different distribution noise should be different and thus the i.i.d. assumption is not sufficient to fully describe the noise behavior. As shown in Fig. 1, three images are contaminated with Gaussian noise, but they have completely different spatial distributions. The traditional MSE or CIM describe the noise behavior under the i.i.d. assumption, and they will obey the same value. Obviously, the noise in Fig. 1(a) has higher randomness than that in Fig. 1(b) and (c), so the entropy is larger. In addition, Fig. 1(b) and (c) have non-i.i.d. structured noise. Hence, it is not realistic to be accurately described by a single noise distribution criterion. Specially, Fig. 1(c) simulates different illumination in face images, which is typical and common noise. Therefore, the weak ability of existing methods to describe complex noise in real scenes needs to be improved. To address these challenges, we utilize the weighted error entropy with the independent and piecewise identically distributed (i.p.i.d.) model to encode non-i.i.d. and i.i.d. noise in the low-rank tensor learning framework. An overview of the proposed WETMSC is shown in Fig. 2. The contributions of our work are summarized as follows:

- We propose the weighted error entropy-regularized tensor learning method for MVSC (WETMSC) without the single distribution assumption or i.i.d. assumption of the noise, in which the weighted error entropy and low-rank tensor learning are integrated into one unified model to remove noise with multiple distributions and discover subspace structure simultaneously.
- WETMSC constructs all self-representation matrices as a third-order tensor to explore the high-order correlation among multiple views. The weighted error entropy with the i.p.i.d. model is robust to non-i.i.d. and i.i.d. noise with different distributions. WETMSC transforms the weighted error entropy function into a convex optimization problem by using the additive form of the half-quadratic (HQ) optimization technology.
- We develop the iterative optimization algorithm to solve the proposed optimization method based on the alternating direction method of multipliers (ADMM). Experimental results on real and synthetic noisy datasets verify the superiority of the proposed method.

The rest of this paper is organized as follows. Section II briefly reviews some related MVSC methods and entropy-based methods from ITL. Section III shows some preliminaries involved in this paper. Section IV and V describe the proposed WETMSC model and its solution process. Section VI reports the clustering results of extensive experiments and model analysis. The conclusions of this paper are summarized in Section VII.

II. Related Works

MVSC methods learn representation coefficients with a certain structure through self-representation technology. Currently, most methods explore the underlying global low-rank property of the self-representation matrix based on LRR. For example, Tang *et al*. [35] learned a joint affinity graph based on LRR for MVSC. References [36], [37] pursued the latent low-rank representation matrix in latent subspace by projection matrix. To explore the correlation among multiple views, the tensorized self-representation matrices were utilized to preserve the high-order structure of views [15], [16], [38]. However, the LRR model follows the linear representation assumption and may fail to discover the local geometry structure of the views when the data is nonlinear. To address this issue, many multi-view subspace clustering methods that fuse LRR and other techniques have been proposed. For example, manifold learning-based methods [39], [40] have been proposed to explore the local structure of views by Laplacian regularization constraint. Xie *et al*. [17] and Wang *et al*. [41] applied the hypergraph idea in the low-rank tensor learning framework to explore the local structure of views through hyper-Laplace regularization. Another way to deal

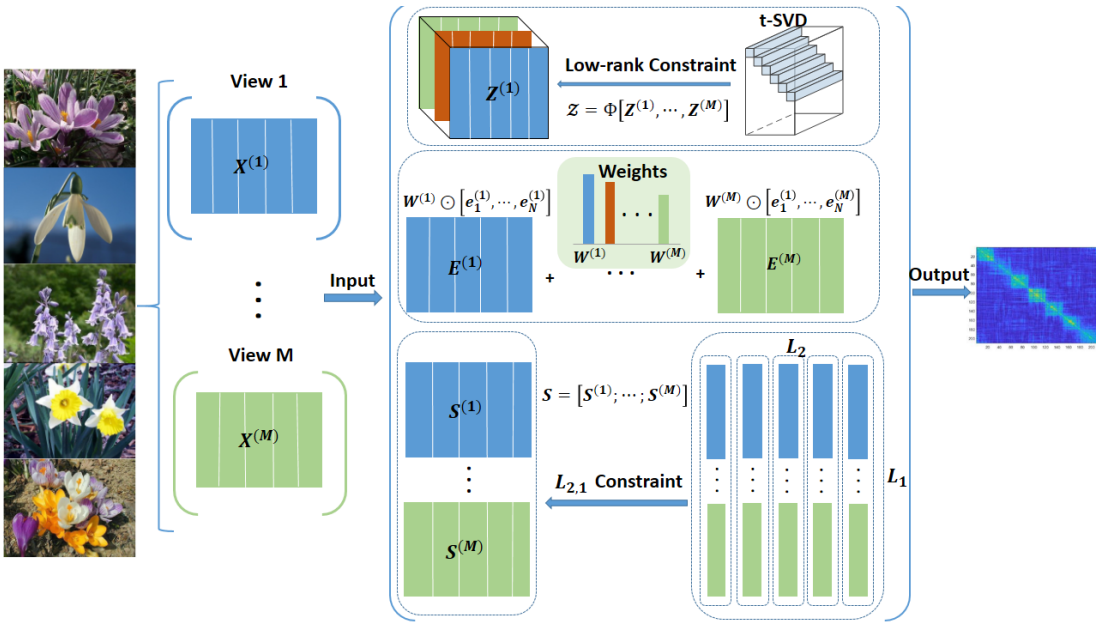

Fig. 2. Given the multimedia data with M views: $\{X^{(v)}\}_{v=1}^{M}$, WETMSC learns self-representation matrix $Z^{(v)}$, error term $E^{(v)}$ and complementary noise $S^{(v)}$ under the self-representation model. First, all $Z^{(v)}$ are constructed as tensor $\mathcal{Z}$ and constrained by the nuclear norm based on t-SVD. Then, WETMSC characterizes the error term $E^{(v)}$ by weighted error entropy and applies the $l_{2,1}$ norm to the separated complementary noise $S^{(v)}$ to extract the complex non-i.i.d. noise and the sample-specific noise. Finally, the affinity matrix is output for spectral clustering.

with nonlinear structured data is the kernel trick, which finds a new feature space for linearly separated multi-view data [21], [23]. However, the LRR model still faces the challenging task of reducing computational complexity and enhancing the block diagonal structure of the representation matrix. To address these problems, many matrix factorization-based MVC methods have been proposed [42]–[44]. Their aim is to factorize multi-view data into common representations with smaller dimension by minimizing the overall loss function for different views. The recent works in [14], [45]–[47] proposed anchor graphs to handle large-scale datasets. Their difference is that methods in [14], [45] directly obtained the anchors through the cluster center of k-means, while methods in [46], [47] learned the public graph through the projected unified anchors, avoiding the isolation of the fixed anchors and the graph construction.

Although the above methods have made significant progress in MVSC, they focus only on exploring the structure of learned self-representative matrix or tensor. For the design of the loss function, most clustering methods simply assume that the noise within one view is a single distribution (e.g., Laplacian distribution, Gaussian distribution or the sample-specific noise obeys Gaussian distribution). They encode noise by using several methods [29], [36], [48]–[50], such as MSE (l_2 norm), l_1 norm, $l_{2,1}$ norm, G_1 norm and mixed Gaussian noise model. Fan *et al*. [51] fused the noiseless structure among views and samples through l_2 norm. However, the above assumptions are limited to complex datasets with non-Gaussian noise or outliers. In addition, the MSE cannot fully exploit the high-order information of the noise. Fortunately, the concept of entropy was first introduced in [33] from ITL. Its followers have proven that entropy is effective in removing non-Gaussian noise and outliers through information metric. For example, Wang *et al*. [32] introduced the minimum error entropy for face recognition. He *et al*. [52] proposed the rotational-invariant principal component analysis algorithm based on entropy metric. The works in [53], [54] introduced correntropy into subspace clustering models with nonlinear kernels. The effectiveness of the above methods lies in that when the error is relatively large, the value of CIM is close to 1 with the ability to suppress large outliers. Although entropy-based methods are robust to complex noise, they assume that the noise is i.i.d. and cannot be applied to datasets with different noise distributions. For example, the face dataset is continuously occluded, and these noises may come from multiple distributions. Thus the i.i.d. assumption is unreasonable. Recently, Li *et al*. [26] proposed a weighted error entropy minimization to model non-i.i.d. noise, which was successfully used for face denoising and background modeling. Weighted error entropy is an extension of correntropy, which constructs a weight matrix based on the information between samples to strengthen the connection of error samples and encodes noise through i.p.i.d. model. Therefore, we leverage the weighted error entropy technique to adequately capture the information of multiple views and characterize different noise distributions to improve clustering performance on different datasets. Specifically, compared with matrix-based methods, WETMSC preserves more category information by exploiting high-order correlation among views by tensorizing self-representing matrices. Compared with tensor-based methods, WETMSC not only focuses on representing the structural information of tensor, but also considers the complex noise. Obviously, the WETMSC method builds on the above methods in a simple and robust way: tensorization and weighted error entropy minimization, while considering

the inherent low-rank property of the self-representation matrix or tensor and the complex noise distribution in views.

TABLE I
SYMBOLS AND EXPLANATIONS

Symbols	Explanations
$\mathcal{X}$, X, $\mathbf{x}$, x	the tensor, matrix, sequence, vector
$\mathcal{X}_{ijk}$	the element (i,j,k) of tensor $\mathcal{X}$
$\mathcal{X}^{(i)}$	the i-th frontal slice of tensor $\mathcal{X}$
$\bar{\mathcal{X}} = \mathbf{fft}(\mathcal{X}, [], 3)$	the Fourier transform alone tube fiber
$d^{(v)}$	feature dimension of the v-th view
N, M	the number of samples, views
$X^{(v)} \in \mathbb{R}^{d^{(v)} \times N}$	feature matrix of the v-th view
$Z^{(v)} \in \mathbb{R}^{N \times N}$	the self-representation matrix
$\mathcal{Z} \in \mathbb{R}^{N \times N \times M}$	the self-representation tensor
$E^{(v)} \in \mathbb{R}^{d^{(v)} \times N}$	the complex non-i.i.d. noise matrix
$S^{(v)} \in \mathbb{R}^{d^{(v)} \times N}$	the sample-specific complementary noise
$S \in \mathbb{R}^{d^{(v)} \times N*M}$	noise matrix constructed by $S^{(v)}$
$\|X\|_{\circledast}$	t-SVD-nuclear norm
$\|X\|_F = \left(\sum_{i,j} \|x_{ij}\|^2\right)^{\frac{1}{2}}$	Frobenius norm
$\|X\|_{2,1} = \sum_i \sqrt{\sum_j x_{ij}^2}$	$l_{2,1}$ norm
$\|X\|_\infty = \max_{0 \le i,j \le n} \|x_{ij}\|$	infinity norm

III. PRELIMINARIES

In this section, we introduce two core contents: tensor operations and weighted error entropy. The tensor operations involve the definition of the tensor nuclear norm. The minimum weighted error entropy model is an optimization for modeling non-i.i.d and i.i.d. noise.

A. Tensor Operations

For easy understanding, we explain the meanings of common symbols in Table I. For tensor $\mathcal{X} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$, the tensor nuclear norm based on tensor singular value decomposition (t-SVD) is the sum of tensor singular values in the Fourier domain, defined as:

$$\|\mathcal{X}\|_{\circledast} = \sum_{i=1}^{min(n_1,n_2)} \sum_{k=1}^{n_3} \bar{\mathcal{S}}(i,i,k), \tag{1}$$

where the tensor singular value $\bar{\mathcal{S}}(i,i,k)$ is obtained by the following correlation definitions.

Definition 1: (t-Product): For tensor $\mathcal{X} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$, and $\mathcal{Y} \in \mathcal{R}^{n_2 \times n_4 \times n_3}$, the t-product $\mathcal{X} * \mathcal{Y}$ is a tensor with size $n_1 \times n_4 \times n_3$.

$$\mathcal{M} = \mathcal{X} * \mathcal{Y} =: bvfold\{bcirc(\mathcal{X}) bvec(\mathcal{Y})\}, \tag{2}$$

where the the block circular matrix (bcirc), block vectorization (bvec), block diagonal matrix (bdiag) and the corresponding opposite operations (bvfold(bvec), bdfold(bdiag)) are as follows.

$$bcirc(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & \mathcal{X}^{(n_3)} & \dots & \mathcal{X}^{(2)} \\ \mathcal{X}^{(2)} & \mathcal{X}^{(1)} & \dots & \mathcal{X}^{(3)} \\ \vdots & \ddots & \ddots & \vdots \\ \mathcal{X}^{(n_3)} & \mathcal{X}^{(n_3-1)} & \dots & \mathcal{X}^{(1)} \end{bmatrix},$$

$$bvec(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} \\ \mathcal{X}^{(2)} \\ \vdots \\ \mathcal{X}^{(n_3)} \end{bmatrix}, \ bvfold(bvec(\mathcal{X})) = \mathcal{X},$$

$$bdiag(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^{(1)} & & \\ & \ddots & \\ & & \mathcal{X}^{(n_3)} \end{bmatrix}, \ bdfold(bdiag(\mathcal{X})) = \mathcal{X}.$$

Definition 2: (Tensor Transpose): For the tensor $\mathcal{X}$ with size $n_1 \times n_2 \times n_3$, the transpose tensor $\mathcal{X}^T \in \mathcal{R}^{n_2 \times n_1 \times n_3}$ is obtained by transposing each frontal slices and then reversing the order of frontal slices 2 to n_3. For example, let $\mathcal{X}$ and its frontal slices be $\mathcal{X}^{(1)}$, $\mathcal{X}^{(2)}$, $\mathcal{X}^{(3)}$ and $\mathcal{X}^{(4)}$, then

$$\mathcal{X}^T = fold\left(\begin{bmatrix} \mathcal{X}^{(1)T} \\ \mathcal{X}^{(4)T} \\ \mathcal{X}^{(3)T} \\ \mathcal{X}^{(2)T} \end{bmatrix}\right).$$

Definition 3: (Identity Tensor): A tensor $\mathcal{I}$ with size $n_1 \times n_1 \times n_3$ is an identity tensor if its first frontal slice is the $n_1 \times n_1$ identity matrix and all other frontal slices are zero.

Definition 4: (Orthogonal Tensor): A tensor $\mathcal{Q} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$ is orthogonal if it satisfies

$$\mathcal{Q}^T * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^T = \mathcal{I}, \tag{3}$$

where $*$ is the t-product. $\mathcal{Q}^T \in \mathcal{R}^{n_2 \times n_1 \times n_3}$ is the transpose of tensor $\mathcal{Q}$. $\mathcal{I} \in \mathcal{R}^{n_1 \times n_1 \times n_3}$ is an identity tensor.

Definition 5: (f-Diagonal Tensor): A tensor is called f-diagonal if each of its frontal slices is diagonal matrix.

Theorem 1: (t-SVD): Let $\mathcal{X} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$, it can be factored as

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T, \tag{4}$$

where $\mathcal{U} \in \mathcal{R}^{n_1 \times n_1 \times n_3}$ and $\mathcal{V} \in \mathcal{R}^{n_2 \times n_2 \times n_3}$ are orthogonal tensors, $\mathcal{S} \in \mathcal{R}^{n_1 \times n_2 \times n_3}$ denotes an f-diagonal tensor.

B. Weighted Error Entropy

The weighted error entropy is used to remove non-i.i.d. noise. Non-i.i.d. noise distribution means that there is different noise in different parts of the data, which can describe the signal using multiple density functions, *i.e.*, i.p.i.d. model. The one dimension i.p.i.d. model has the following definition.

Definition 6: Given the sequence $\mathbf{e} = [e_1, e_2, \cdots, e_N] \in R^N$, N is the number of independent samples and $\{\mathcal{P}_q\}_{q=1}^L$ is a nonoverlapping, sequential partition of the index vector $[1, 2, \cdots, N]$, defined as

$$\mathcal{P}_q = [p_{q-1} + 1, p_{q-1} + 2, \cdots, p_q], q \in \{1, \cdots, L\}, \tag{5}$$

where $p_q < p_{q+1}, p_L = N$ and $p_0 = 0$. The sequence $\mathbf{e}$ is generated by an i.p.i.d. source if the subsequence of $\mathbf{e}$ determined by $\mathcal{P}_q$ is independently generated according to a probability density function $\{f_q\}_{q=1}^L$, which is given by

$$e_{p_{q-1}+1}, e_{p_{q-1}+2}, \cdots, e_{p_q} \overset{i.i.d.}{\sim} f_q. \tag{6}$$

The one dimension i.p.i.d. model can be extended to multi-dimensional signals such as images and videos. The i.p.i.d.

model is different from the traditional i.i.d. noise: 1) i.p.i.d. utilizes multiple density functions instead of a single distribution to describe the noise of non-overlapping partitions, enhancing the ability to characterize complex signals and describe local structural information. 2) When $L = 1$, i.p.i.d. degenerates to traditional i.i.d., which indicates that i.p.i.d. as an extension of i.i.d. still retains the ability to describe purely random signals.

The traditional error entropy comes from Rényi's quadratic entropy in ITL. The second-order Rényi's entropy of random sequence $\mathbf{e}$ generated by traditional i.i.d. is defined as:

$$H_2(\mathbf{e}) = -\log\big(\int\big(\frac{1}{N}\sum_{i=1}^{N}k_\sigma(e-e_i)\big)^2\big), \tag{7}$$

where $\frac{1}{N}\sum_{i=1}^{N}k_\sigma(e-e_i)$ is the probability density function estimated directly from the sample by the Parzen window method. $k_\sigma(e-e_i) = exp(-\frac{(e-e_i)^2}{2\sigma^2})$ denotes the Gaussian kernel function with the kernel size σ.

For the random sequence $\mathbf{e}$ generated by i.p.i.d., the second-order Rényi's entropy is defined as:

$$\hat{H}_2(\mathbf{e}) = -\log\big(\sum_q\frac{|\mathcal{P}_q|}{N}\int\big(\frac{1}{|\mathcal{P}_q|}\sum_{i\in\mathcal{P}_q}k_\sigma(e-e_i)\big)^2\big). \tag{8}$$

Specifically, when $L = 1$, $\hat{H}_2(\mathbf{e})$ is equivalent to $H_2(\mathbf{e})$. $\hat{H}_2(\mathbf{e})$ applies the weighted average information potential over different partitions. As $\{\mathcal{P}_q\}_{q=1}^{L}$ is usually hard to obtain, it is difficult to compute $\hat{H}_2(\mathbf{e})$. According to [26], assuming that the local region of $\mathbf{e}$ with i.p.i.d. model satisfies i.i.d., the position of the sample sequence e_q in each partition can be expressed as I_q, and the corresponding local region is Ω_{I_q}. When Ω_{I_q} coincides with $\{\mathcal{P}_q\}_{q=1}^{L}$, $\hat{H}_2(\mathbf{e})$ is completely estimated. However, the sample probability density function in a very small region is difficult to estimate. To solve $\hat{H}_2(\mathbf{e})$, the weighting function $Dis(\cdot)$ (l_2 norm) is introduced to calculate the weight of each sample. The weight is determined by the distance $D_{q,i} = Dis(I_q, I_i)$ between the samples I_i from I_q in the region Ω_{I_q}. The probability density function for Ω_{I_q} is approximated as $\sum_{i=1}^{N}c(D_{q,i})k_\sigma(e-e_i)$, where $c(\cdot) = \frac{1}{Q}e^{\frac{(D_{q,i})^2}{2\sigma^2}}$ is the weighting function. Q is a normalization term to satisfy $\sum_{I_i}c(D_{q,i}) = 1$, $\sigma^2 = \frac{N}{1000}$. Then $\hat{H}_2(\mathbf{e})$ is transformed into the following form:

$$\begin{aligned}\bar{H}_2(\mathbf{e}) &= -\log\frac{1}{N}\big(\sum_{I_q}\int\big(\sum_{i=1}^{N}c(D_{q,i})k_\sigma(e-e_i)\big)^2\big)\\ &= -\log\frac{1}{N}\sum_{I_q}\sum_{i,j=1}^{N}c(D_{q,i})c(D_{q,j})k_{\sqrt{2}\sigma}(e_i-e_j)\\ &= -\log\sum_{i,j=1}^{N}\omega_{i,j}k_\sigma(e_i-e_j)+\log N\\ &= -\log S(\mathbf{e})+\log N.\end{aligned} \tag{9}$$

Note that $\bar{H}_2(\mathbf{e})$ is an approximation of $\hat{H}_2(\mathbf{e})$ without using the partitions prior $\{\mathcal{P}_q\}_{q=1}^{L}$. $S(\mathbf{e}) = \sum_{i,j=1}^{N}\omega_{i,j}k_\sigma(e_i-e_j)$, $\omega_{i,j} = \sum_{I_q}c(D_{q,i})c(D_{q,j})$. For the convenience of calculation, in this paper, we set $w_{i,j} = \frac{e^{-\frac{(D_{i,j})^2}{2\sigma^2}}}{\sum_i e^{-\frac{(D_{i,j})^2}{2\sigma^2}}}$. According to [26], double-summation function $S(\mathbf{e})$ can be relaxed as follows:

$$\hat{S}(\mathbf{e}) = \sum_{i,j=1}^{N}k_\sigma(\mathbf{w}_i\mathbf{e}), \tag{10}$$

where $\mathbf{w}_i = [w_{i,1}, \cdots, w_{i,N}]$. Since $\bar{H}_2(\mathbf{e})$ is a monotonic decreasing function with respect to $S(\mathbf{e})$, minimizing $\bar{H}_2(\mathbf{e})$ is equivalent to minimizing $-\hat{S}(\mathbf{e})$.

IV. Motivation and The Proposed WETMSC

For MVC, most existing methods pursue the structure information of self-representation matrix or tensor under the single noise distribution assumption or i.i.d. assumption. Although the global correlation of inter-view and intra-view have been well studied, simple constraint on noise limits the scalability of MVSC. It can be concluded from the diversity of existing views that the impact of non-i.i.d. noise is inevitable in clustering tasks. Therefore, we use the weighted error entropy with i.p.i.d. model to measure the error according to the correlation among samples instead of a single Gaussian or Laplacian distribution assumption. In this section, we propose the unified entropy and tensor learning for MVSC (WETMSC) using the tensor nuclear norm and weighted error entropy. The proposed WETMSC mainly clusters multimedia data with M views and N samples: $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \cdots, x_N^{(v)}] \in R^{d^{(v)}\times N}$. $d^{(v)}$ is the dimension of the v-th view. Specifically, WETMSC expresses $X^{(v)}$ as $X^{(v)} = X^{(v)}Z^{(v)} + S^{(v)} + E^{(v)}$. Accordingly, the proposed WETMSC model is formulated as follows:

$$\begin{aligned}&\min_{\mathcal{Z},S,E} -\sum_{v=1}^{M}\sum_{i,j=1}^{N}k_\sigma(\mathbf{w}_i^{(v)}\mathbf{e}_j^{(v)}) + \lambda\|S\|_{2,1} + \beta\|\mathcal{Z}\|_{\circledast}\\ &s.t.\ X^{(v)} = X^{(v)}Z^{(v)} + S^{(v)} + E^{(v)}, v = 1, \cdots, M,\\ &\quad S = [S^{(1)}; S^{(2)}; \cdots; S^{(M)}],\\ &\quad \mathcal{Z} = \Phi(Z^{(1)}, Z^{(2)}, \cdots, Z^{(M)}).\end{aligned} \tag{11}$$

Different from the traditional methods to directly optimize self-representation matrix $Z^{(v)}$, WETMSC constructs all self-representation matrix $Z^{(v)}$ as a third-order tensor $\mathcal{Z}$ for overall optimization, which could strengthen the high-order correlation among all views. On the other hand, WETMSC characterizes the error term $E^{(v)}$ by the weighted error entropy rather than a single noise distribution assumption to effectively remove complex non-i.i.d. or i.i.d. noise and their combination. Considering the inherently dense property of noise in each sample, WETMSC applies the $l_{2,1}$ norm to the separated complementary noise $S^{(v)}$ to extract the sample-specific noise. λ and β are the trade-off parameters for low-rank and error terms. $\mathbf{e}_j^{(v)}$ is the j-th column of $E^{(v)}$. $\sum_{i,j=1}^{N}k_\sigma(\mathbf{w}_i^{(v)}\mathbf{e}_j^{(v)}) = \sum_{i,j=1}^{N}exp(-\frac{(\mathbf{w}_i^{(v)}\mathbf{e}_j^{(v)})^2}{2\sigma^2})$ is a nonconvex function, which can been solved by HQ technology. HQ as an optimization method for ITL-based methods can transform nonconvex function

into quadratic terms that are tractable. Let $\phi(x)$ be a loss function, the HQ technology based on the additive form has the following expression:

$$\phi(x) = \min_{u}\{(x-u)^2 + \psi(u)\}, \tag{12}$$

where $\psi(u)$ denotes the dual function, and u as an auxiliary variable is determined by the minimizer function $\delta(x) = x - x * exp(-\frac{x^2}{\sigma^2})$.

We first construct the v-th weight matrix $W^{(v)} = [\mathbf{w}_1^{(v)}, \cdots, \mathbf{w}_N^{(v)}]^T$, which i,j-th element is $w_{i,j}^{(v)}$. Then, we transform the proposed objective function (11) into a convex optimization problem with the HQ technology:

$$\begin{aligned} \min_{\mathcal{Z},S,E} & \sum_{v=1}^{M}\sum_{j=1}^{N}\|W^{(v)}\mathbf{e}_j^{(v)} - \mathbf{u}_j^{(v)}\|_F^2 + \psi(\mathbf{u}_j^{(v)}) \\ & + \lambda\|S\|_{2,1} + \beta\|\mathcal{Z}\|_{\circledast} \\ s.t.\ & X^{(v)} = X^{(v)}Z^{(v)} + S^{(v)} + E^{(v)}, v = 1, \cdots, M, \\ & S = [S^{(1)}; S^{(2)}; \cdots; S^{(M)}], \\ & \mathcal{Z} = \Phi(Z^{(1)}, Z^{(2)}, \cdots, Z^{(M)}). \end{aligned} \tag{13}$$

Since problem (13) is a constrained optimization problem, and both the objective function and the constraint condition contain $Z^{(v)}$. We use ADMM to solve the objective function (13).

V. Optimization of WETMSC

ADMM is uesd to transform objective function (13) into an unconstrained optimization problem. To separate the original complex problem into several easy subproblems, we introduce one auxiliary variable $\mathcal{H}$ to separate variable $\mathcal{Z}$ and utilize alternating minimization scheme to solve each subproblem. The augmented Lagrangian function of Eq. (13) is:

$$\begin{aligned} \mathcal{L}(\mathcal{Z},\mathcal{H},E^{(v)},S) &= \lambda\|S\|_{2,1} + \frac{\mu}{2}\|\mathcal{Z} - \mathcal{H} + \frac{\mathcal{Y}_2}{\mu}\|_F^2 \\ &+ \sum_{v=1}^{M}\sum_{j=1}^{N}\|W^{(v)}\mathbf{e}_j^{(v)} - \mathbf{u}_j^{(v)}\|_F^2 + \beta\|\mathcal{H}\|_{\circledast} \\ &+ \sum_{v=1}^{M}\big(\frac{\mu}{2}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)} - S^{(v)} + \frac{Y_1^{(v)}}{\mu}\|_F^2\big), \end{aligned} \tag{14}$$

where $\{Y_1^{(v)}\}_{v=1}^M$ and $\mathcal{Y}_2$ represent the Lagrange multipliers and μ denotes the penalty parameter.

Solving auxiliary variable u: According to the HQ technology, when $E^{(v)}$ is fixed, the auxiliary variable $\mathbf{u}$ can be updated by

$$\mathbf{u}_{jk+1}^{(v)} = \sum_{i=1}^{N}(\mathbf{w}_i^{(v)}\mathbf{e}_{jk}^{(v)} - \mathbf{w}_i^{(v)}\mathbf{e}_{jk}^{(v)} * exp(-\frac{(\mathbf{w}_i^{(v)}\mathbf{e}_{jk}^{(v)})^2}{\sigma^2})). \tag{15}$$

Solving auxiliary variable $\mathcal{H}$: We fix variable $\mathcal{Z}$. The optimization problem $\mathcal{H}$ is equivalent to

$$\mathcal{H}_{k+1} = \arg\min_{\mathcal{H}}\|\mathcal{H}\|_{\circledast} + \frac{\mu_k}{2}\|\mathcal{Z}_k - \mathcal{H} + \frac{\mathcal{Y}_{2k}}{\mu_k}\|_F^2. \tag{16}$$

Eq.(16) is a typical low-rank tensor approximation problem which has a closed-form solution and can be easily solved by the following tensor tubal-shrinkage operator [16].

$$\mathcal{H}_{k+1} = \mathcal{C}_{\frac{M}{2\mu_k}}(\mathcal{F}) = \mathcal{U} * \mathcal{C}_{\frac{M}{2\mu_k}}(\mathcal{S}) * \mathcal{V}^T, \tag{17}$$

where $\mathcal{F} = \mathcal{Z}_k + \frac{\mathcal{Y}_{2k}}{\mu_k} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, $\mathcal{C}_{\frac{M}{2\mu_k}}(\mathcal{S}) = \mathcal{S} * \mathbf{diag}\{\max(\bar{\mathcal{S}}^{(j)}(i,i) - \frac{M}{2\mu_k}, 0)\}$, $\bar{\mathcal{S}} = fft(\mathcal{S}, [], 3)$ and $j = 1, 2, \cdots, N$.

Solving regularization term $\mathcal{Z}$: We first fix variabl $\mathcal{H}$, then solve the regularization term $\mathcal{Z}$ view-by-view since the following two terms are regularized Frobenius-norm. The subproblem of the v-th view of $\mathcal{Z}$ is shown as follows:

$$\begin{aligned} Z_{k+1}^{(v)} = \arg\min_{Z^{(v)}} & \frac{\mu_k}{2}\|Z^{(v)} - H_{k+1}^{(v)} + \frac{Y_{2k}^{(v)}}{\mu_k}\|_F^2 \\ & + \frac{\mu_k}{2}\|X^{(v)} - X^{(v)}Z^{(v)} - E_k^{(v)} - S_k^{(v)} + \frac{Y_{1k}^{(v)}}{\mu_k}\|_F^2. \end{aligned} \tag{18}$$

Since $Z^{(v)}$ is constrained by two Frobenius norms, we obtain the optimal solution of $Z^{(v)}$ by solving the derivative with respect to $Z^{(v)}$ as follows

$$\begin{aligned} Z_{k+1}^{(v)} = & \left(I + X^{(v)^T}X^{(v)}\right)^{-1} \times \big(X^{(v)^T}X^{(v)} - X^{(v)^T}E_k^{(v)} \\ & - X^{(v)^T}S_k^{(v)} + X^{(v)^T}\frac{Y_{1k}^{(v)}}{\mu_k} + H_{k+1}^{(v)} - \frac{Y_{2k}^{(v)}}{\mu_k}\big), \end{aligned} \tag{19}$$

where I denotes the identity matrix with proper size. The term $\left(I + X^{(v)^T}X^{(v)}\right)^{-1}$ could be precomputed before iteration to reduce the computational cost of the proposed algorithm to some extent.

Solving the error matrix $E^{(v)}$: When fixing variables $\mathbf{u}^{(v)}$ and $\mathcal{Z}$, the error matrix $E^{(v)}$ from Eq. (14) has the following formula:

$$\begin{aligned} E_{k+1}^{(v)} = \arg\min_{E^{(v)}} & \sum_{v=1}^{M}\sum_{i=1}^{N}\|W^{(v)}\mathbf{e}_{jk+1}^{(v)} - \mathbf{u}_{jk+1}^{(v)}\|_F^2 \\ & + \frac{\mu_k}{2}\|E^{(v)} - B_k^{(v)}\|_F^2, \end{aligned} \tag{20}$$

where $B_k^{(v)} = X^{(v)} - X^{(v)}Z_{k+1}^{(v)} - S_k^{(v)} + \frac{Y_{1k}^{(v)}}{\mu_k}$. Similar to Eq. (18), the problem (20) can be solved by setting the derivative of Eq. (20) to zero and the closed-form solution is

$$\mathbf{e}_{jk+1}^{(v)} = (2W^{(v)T}W^{(v)} + \mu I)^{-1}(2W^{(v)T}\mathbf{u}_{jk+1}^{(v)} + \mu b_{jk}^{(v)}). \tag{21}$$

Solving error matrix S: Similar to the strategy of the above subproblem, the variables $\mathcal{Z}$ and $E^{(v)}$ are first fixed, and the formula for S is as follows:

$$S_{k+1} = \arg\min_{S}\frac{\lambda}{\mu_k}\|S\|_{2,1} + \frac{1}{2}\|S - D_k\|_F^2, \tag{22}$$

where $D_k = [D_k^{(1)}; D_k^{(2)}; \cdots; D_k^{(M)}]$ is constructed by $D_k^{(v)} = X^{(v)} - X^{(v)}Z_{k+1}^{(v)} - E_{k+1}^{(v)} + \frac{Y_{1k}^{(v)}}{\mu_{1k}}$. The j-th column of optimal solution S^* can be obtained by

$$S_{k+1}(:,j) = \begin{cases} \frac{\|D_k(:,j)\|_2 - \frac{\lambda}{\mu_k}}{\|D_k(:,j)\|_2}D_k(:,j), & \frac{\lambda}{\mu_k} < \|D_k(:,j)\|_2; \\ 0, & otherwise. \end{cases} \tag{23}$$

At last, we update the Lagrange multipliers $Y_1, \mathcal{Y}_2$ and penalty parameter μ by

$$Y_{1k+1}^{(v)} = Y_{1k}^{(v)} + \mu_k(X^{(v)} - X^{(v)}Z_{k+1}^{(v)} - E_{k+1}^{(v)} - S_{k+1}^{(v)}); \quad (24)$$

$$\mathcal{Y}_{2k+1} = \mathcal{Y}_{2k} + \mu_k(\mathcal{Z}_{k+1} - \mathcal{H}_{k+1}); \quad (25)$$

$$\mu_{k+1} = \min\{\rho * \mu_k, \mu_{max}\}, \quad (26)$$

where ρ is used to accelerate the convergence speed. Algorithm 1 shows the computational process of the proposed WETMSC.

Algorithm 1: WETMSC for MVSC

Require: multi-view features $\{X^{(v)}, v = 1, 2, \cdots, M\}$; parameter λ, β, k; u, $\mathcal{H}$, $\mathcal{Z}$, E, S, Y_1, $\mathcal{Y}_2$ initialized to $\mathbf{0}$; $\mu = 10^{-5}$, $\rho = 2$, $tol = 10^{-7}$;

while not converged **do**
  Update $\mathcal{H}_{t+1}$ by Eq. (17);
  for $v = 1$ to M **do**
    Update $\mathbf{u}_{k+1}^{(v)}$ by Eq. (15);
    Update $Z_{k+1}^{(v)}$ by Eq. (19);
    Update $E_{k+1}^{(v)}$ by Eq. (21);
    Update $S_{k+1}^{(v)}$ by Eq. (23);
    Update $Y_{1k+1}^{(v)}$ by Eq. (24);
  end for
  Update $\mathcal{Y}_{2k+1}$ and ρ_{k+1} by Eqs. (25) and (26);
  Check the convergence conditions:
  $$\max\left\{ \begin{array}{l} \|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)} - S^{(v)}\|_\infty, \\ \|\mathcal{Z} - \mathcal{H}\|_\infty \end{array} \right\} \leq tol$$
end while

Ensure: $Z_{k+1}^{(v)}$.

Complexity analysis: The weight W in this paper can be pre-calculated, so the complexity of the proposed WETMSC algorithm comes from the calculation of variables $\mathbf{u}^{(v)}$, $\mathcal{H}$, $\mathcal{Z}$, $E^{(v)}$ and S. Each step of solving $\mathbf{u}_j^{(v)}$, $Z^{(v)}$, $E^{(v)}$ and $S^{(v)}$ involves matrix multiplication and matrix inverse calculation, and the total computational complexity is about $\mathcal{O}(N^3)$. The most important computation is singular value decomposition, fast Fourier transform (FFT) and inverse FFT of the tensor $\mathcal{H}$. To reduce the computational complexity, we rotate the $N \times N \times M$ tensor $\mathcal{H}$ into an $N \times M \times N$ tensor during the computation. Then, the computational complexity of tensor $\mathcal{H}$ is about $\mathcal{O}(2N^2Mlog(N)) + \mathcal{O}(N^2M^2)$. In addition, after obtaining the coefficient matrix $Z^{(v)}$, we can calculate the affinity matrix $C = \frac{1}{M}\sum_{v=1}^{M}(|Z^{(v)}| + |Z^{(v)T}|)$ and input it into the spectral clustering algorithm to obtain the final clustering result whose computational complexity is usually $\mathcal{O}(N^3)$. In summary, the whole computational complexity of our WETMSC algorithm is about $\mathcal{O}(2kN^2Mlog(N)) + \mathcal{O}(N^3)$, where k is the number of iterations.

VI. Experiment

A. Dataset and Experimental Setup

We investigate the clustering performance of the proposed WETMSC on six real-world datasets including text and image datasets. **BBCSport**[1] is a News article dataset from BBC website with 2 views. **UCI-3views**[2] is a handwritten digit dataset from the UC Irvine machine learning repository, including 3 types of features: $240d$ Fourier coefficients, $76d$ pixel averages and $6d$ morphological. **MSRC-V1** dataset contains images of trees, cars, buildings, faces, etc.. There are 5 types of features including $24d$ colour moment, $576d$ histogram of oriented gradients, $512d$ GIST, $254d$ CENTRIST feature and $256d$ local binary pattern. **MITIndoor** is a scene dataset, including 4 types of features ($4096d$ PHOW, $3600d$ LBP, $1770d$ CENTRIST and $1240d$ deep features). **Scene-15**[3] dataset contains scene images of outdoor and indoor, including 3 types of features ($1800d$ PHOW, $1180d$ PRI-CoLBP and 1240 CENTRIST). **Caltech101** dataset contains object images with 4 types of features. The details of the six datasets are summarized in Table II and some samples are shown in Fig. 3.

TABLE II
STATISTICS OF SIX REAL-WORLD MULTI-VIEW DATASETS

Datasets	View	Dimension (d)	Samples	Categories
BBCSport	2	3183/3203	544	5
UCI-3views	3	240/76/6	2000	10
MSRC-V1	5	24/576/512/256/254	210	7
MITIndoor	4	4096/3600/1770/1240	5360	67
Scene-15	3	1800/1180/1240	4485	15
Caltech101	4	2048/4800/3540/1240	8677	101

Fig. 3. Some samples of (a) UCI-3views, (b) MSRC-V1, (c) Scene15 and (d) MITIndoor datasets.

In all experimental settings, the parameters of the comparison methods are set according to the corresponding literature. The regularization parameters of SSC and LRR are chosen from $[0.01, 10]$. The regularization parameter of RMSC is set from the interval $[0.005, 0.01, 0.05, 0.1]$. For DiMSC, two parameters are set in $[0.01, 0.03]$ and $[20 : 20 : 180]$, respectively. For LT-MSC, the trade-off parameter is selected from 0.01 to 100. The parameter of GMC is set to 1 as the initial value, which is tuned automatically in the clustering process. For LMSC, the parameter is tuned from $[0.001, 0.01, 0.1, 1, 10, 100, 1000]$. For MLAN, one parameter is set to a random number between 1 and 30. The trade-off parameter of tSVDMSC is set within the range $[0.1, 2]$. The parameter of ETLMSC is selected from the range of $[0.0008, 0.01]$. For LRTG, two parameters are empirically selected from the sets of $[0.001, 0.01, 0.1, 0.5, 1, 2, 5, 10, 50, 100]$ and $[5, 15]$, respec-

[1] http://mlg.ucd.ie/datasets/segment.html
[2] http://archive.ics.uci.edu/ml/datasets/Multiple+Features
[3] http://www-cvr.ai.uiuc.edu/ponce_grp/data/

TABLE III
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *BBCSport*, *UCI-3views*, *MSRC-V1* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
BBCSport	SSC [11]	0.627±0.003	0.534±0.008	0.364±0.007	0.565±0.005	0.427±0.004	0.834±0.004
	LRR [12]	0.836±0.001	0.698±0.002	0.705±0.001	0.776±0.001	0.768±0.001	0.784±0.001
	RMSC [48]	0.826±0.001	0.666±0.001	0.637±0.001	0.719±0.001	0.766±0.001	0.677±0.001
	DiMSC [49]	0.922±0.000	0.785±0.000	0.813±0.000	0.858±0.000	0.846±0.000	0.872±0.000
	LT-MSC [15]	0.460±0.046	0.222±0.028	0.167±0.043	0.428±0.014	0.328±0.028	0.629±0.053
	GMC [55]	0.807±0.000	0.760±0.000	0.722±0.000	0.794±0.000	0.727±0.000	0.875±0.000
	LMSC [36]	0.847±0.003	0.739±0.001	0.749±0.001	0.810±0.001	0.799±0.001	0.822±0.001
	MLAN [56]	0.721±0.000	0.779±0.000	0.591±0.000	0.714±0.000	0.567±0.000	0.962±0.000
	tSVDMSC [16]	0.879±0.000	0.765±0.000	0.784±0.000	0.834±0.000	0.863±0.000	0.807±0.000
	ETLMSC [28]	0.959±0.086	0.972±0.058	0.949±0.107	0.961±0.081	0.963±0.078	0.960±0.085
	LRTG [57]	0.943±0.005	0.869±0.009	0.840±0.012	0.879±0.010	0.866±0.006	0.892±0.014
	GNLTA [18]	0.980±0.064	0.986±0.043	0.973±0.086	0.979±0.065	0.979±0.067	0.980±0.064
	ERMC-AGR [45]	0.576±0.072	0.382±0.129	0.300±0.136	0.465±0.101	0.475±0.116	0.457±0.091
	RLMSC [51]	0.822±0.000	0.723±0.000	0.687±0.000	0.758±0.000	0.796±0.000	0.724±0.000
	WETMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
UCI-3views	SSC [11]	0.815±0.011	0.840±0.001	0.770±0.005	0.794±0.004	0.747±0.010	0.848±0.004
	LRR [12]	0.871±0.001	0.768±0.002	0.736±0.002	0.763±0.002	0.759±0.002	0.767±0.002
	RMSC [48]	0.915±0.024	0.822±0.008	0.789±0.014	0.811±0.012	0.797±0.017	0.826±0.006
	DiMSC [49]	0.703±0.010	0.772±0.006	0.652±0.006	0.695±0.006	0.673±0.005	0.718±0.007
	LT-MSC [15]	0.803±0.001	0.775±0.001	0.725±0.001	0.753±0.001	0.739±0.001	0.767±0.001
	GMC [55]	0.736±0.000	0.815±0.000	0.678±0.000	0.713±0.000	0.644±0.000	0.799±0.000
	LMSC [36]	0.893±0.000	0.815±0.000	0.783±0.000	0.805±0.000	0.798±0.000	0.812±0.000
	MLAN [56]	0.874±0.000	0.910±0.000	0.847±0.000	0.864±0.000	0.797±0.000	0.943±0.000
	tSVDMSC [16]	0.955±0.000	0.932±0.000	0.924±0.000	0.932±0.000	0.930±0.000	0.934±0.000
	ETLMSC [28]	0.958±0.078	0.977±0.028	0.953±0.069	0.958±0.062	0.940±0.088	0.980±0.029
	LRTG [57]	0.981±0.000	0.953±0.000	0.957±0.000	0.961±0.000	0.961±0.000	0.962±0.000
	GNLTA [18]	0.981±0.036	0.979±0.012	0.972±0.036	0.975±0.032	0.968±0.052	0.983±0.008
	ERMC-AGR [45]	0.746±0.063	0.675±0.035	0.582±0.060	0.625±0.053	0.607±0.065	0.645±0.040
	RLMSC [51]	0.754±0.000	0.759±0.000	0.608±0.000	0.648±0.000	0.637±0.000	0.659±0.000
	WETMSC	**0.996±0.000**	**0.989±0.000**	**0.991±0.000**	**0.992±0.000**	**0.992±0.000**	**0.992±0.000**
MSRC-V1	SSC [11]	0.791±0.007	0.750±0.005	0.651±0.006	0.701±0.005	0.670±0.008	0.736±0.003
	LRR [12]	0.695±0.000	0.590±0.000	0.491±0.000	0.562±0.000	0.560±0.000	0.564±0.000
	RMSC [48]	0.761±0.054	0.673±0.032	0.587±0.041	0.646±0.035	0.633±0.041	0.660±0.031
	DiMSC [49]	0.759±0.009	0.622±0.015	0.548±0.015	0.611±0.013	0.606±0.013	0.616±0.012
	LT-MSC [15]	0.831±0.003	0.743±0.004	0.665±0.004	0.712±0.004	0.699±0.004	0.725±0.003
	GMC [55]	0.748±0.000	0.771±0.000	0.640±0.000	0.697±0.000	0.612±0.000	0.809±0.000
	LMSC [36]	0.770±0.022	0.679±0.025	0.596±0.028	0.654±0.024	0.635±0.025	0.673±0.023
	MLAN [56]	0.859±0.003	0.751±0.003	0.709±0.004	0.750±0.003	0.727±0.004	0.776±0.002
	tSVDMSC [16]	0.991±0.000	0.982±0.000	0.978±0.000	0.981 ±0.000	0.980±0.000	0.982±0.000
	ETLMSC [28]	0.872±0.082	0.805±0.053	0.764±0.083	0.797±0.071	0.784±0.082	0.812±0.059
	LRTG [57]	0.895±0.000	0.829±0.000	0.775±0.000	0.807±0.000	0.794±0.000	0.821±0.000
	GNLTA [18]	0.894±0.089	0.880±0.053	0.829±0.092	0.854±0.078	0.836±0.095	0.873±0.060
	ERMC-AGR [45]	0.663±0.073	0.578±0.065	0.472±0.084	0.549±0.072	0.523±0.065	0.577±0.080
	RLMSC [51]	0.719±0.000	0.660±0.000	0.508±0.000	0.578±0.000	0.567±0.000	0.590±0.000
	WETMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**

tively. For GNLTA, two parameters are tuned from interval $[0.001, 0.1]$ and $[1.1, 2]$. The ranges of the two parameters in RLMSC are $[1, 100]$ and $[0.01, 0.5]$, respectively. For the proposed WETMSC method, the value range of parameters λ and β are $[0.01 : 0.01 : 0.1], [0.1 : 0.1 : 1]$, respectively. All comparison methods in this paper use the k-means clustering algorithm on the affinity matrix, where the clusters number is the categories number of datasets. Considering the effect of randomness of k-means, we repeat 10 times k-means clustering for each dataset and report the average clustering results with. Clustering performance is quantitatively evaluated under the six general clustering metrics, including Accuracy (ACC), normalized mutual information (NMI), adjusted rand index (AR), F-score, Precision, and Recall. All experiments are performed on desktop computer with 3.19 GHz CPU, 32 GB RAM and MATLAB 2018b (64-bit).

B. Comparison with State-of-the-art Methods

To verify the effectiveness of the proposed WETMSC method, we compare it with 2 single-view clustering methods and 12 MVC methods, including 5 tensor-based clustering methods. Specifically, sparse representation clustering (SSC) [11] is used for each view independently to generate a sparse self-representation matrix and reports the best performance; Low-rank representation clustering (LRR) [12] is used for each view independently to generate a low-rank self-representation matrix and reports the best performance; Robust multi-view spectral clustering (RMSC) [48] recovers a shared low-rank transition probability matrix by Markov chain method. Diversity-induced clustering (DiMSC) [49] explores the complementarity of views by diversity term; Graph learning clustering (GMC) [55] fuses data graph matrices of all views to generate a unified graph matrix; Latent subspace clustering (LMSC) [36] learns latent low-rank representation by mapping the original space into a low-dimensional latent

TABLE IV
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON *MITIndoor-67*, *Scene-15*, *Caltech101* DATASETS.

Dataset	Method	ACC	NMI	AR	F-score	Precision	Recall
MITIndoor	SSC [11]	0.475±0.008	0.615±0.003	0.322±0.006	0.343±0.006	0.314±0.007	0.377±0.007
	LRR [12]	0.120±0.004	0.226±0.006	0.031±0.007	0.045±0.004	0.044±0.006	0.047±0.004
	RMSC [48]	0.232±0.009	0.342±0.004	0.110±0.003	0.123±0.002	0.121±0.003	0.125±0.003
	DiMSC [49]	0.246±0.000	0.383±0.000	0.128±0.000	0.141±0.000	0.138±0.000	0.144±0.000
	LT-MSC [15]	0.431±0.002	0.546±0.004	0.280±0.008	0.290±0.002	0.279±0.006	0.306±0.005
	GMC [55]	0.099±0.000	0.204±0.000	0.003±0.000	0.032±0.000	0.016±0.000	0.838±0.000
	LMSC [36]	0.384±0.006	0.506±0.005	0.243±0.005	0.254±0.004	0.245±0.005	0.264±0.004
	MLAN [56]	0.232±0.010	0.408±0.012	0.012±0.009	0.041±0.003	0.021±0.001	0.662±0.015
	tSVDMSC [16]	0.684±0.005	0.750±0.007	0.555±0.005	0.562±0.008	0.543±0.005	0.582±0.004
	ETLMSC [28]	0.775±0.000	0.899±0.000	0.729±0.000	0.733±0.000	0.709±0.000	0.758±0.000
	LRTG [57]	0.404±0.008	0.522±0.005	0.244±0.006	0.255±0.005	0.243±0.007	0.269±0.004
	GNLTA [18]	0.791±0.047	0.907±0.016	0.740±0.050	0.744±0.050	0.727±0.050	0.762±0.049
	ERMC-AGR [45]	0.090±0.003	0.208±0.003	0.022±0.002	0.043±0.002	0.030±0.001	0.072±0.006
	RLMSC [51]	0.610±0.000	0.848±0.000	0.539±0.000	0.546±0.000	0.520±0.000	0.576±0.000
	WETMSC	**0.881±0.018**	**0.915±0.015**	**0.835±0.018**	**0.837±0.018**	**0.814±0.022**	**0.862±0.014**
Scene-15	SSC [11]	0.444±0.003	0.470±0.002	0.279±0.001	0.337±0.002	0.292±0.001	0.397±0.001
	LRR [12]	0.445±0.013	0.426±0.018	0.272±0.015	0.324±0.010	0.316±0.011	0.333±0.015
	RMSC [48]	0.503±0.000	0.495±0.000	0.325±0.000	0.371±0.000	0.374±0.000	0.368±0.000
	DiMSC [49]	0.300±0.010	0.269±0.009	0.117±0.012	0.181±0.010	0.173±0.016	0.190±0.010
	LT-MSC [15]	0.574±0.009	0.571±0.011	0.424±0.010	0.465±0.007	0.452±0.003	0.479±0.008
	GMC [55]	0.381±0.000	0.519±0.000	0.191±0.000	0.281±0.000	0.174±0.000	0.732±0.000
	LMSC [36]	0.563±0.000	0.525±0.000	0.397±0.000	0.440 ±0.000	0.430±0.000	0.450±0.000
	MLAN [56]	0.331±0.000	0.475±0.000	0.151±0.000	0.248±0.000	0.150±0.000	0.731±0.000
	tSVDMSC [16]	0.812±0.007	0.858±0.007	0.771±0.003	0.788±0.001	0.743±0.006	0.839±0.003
	ETLMSC [28]	0.878±0.000	0.902±0.000	0.851±0.000	0.862±0.000	0.848±0.000	0.877±0.000
	LRTG [57]	0.615±0.016	0.657±0.005	0.486±0.016	0.525±0.014	0.485±0.023	0.572±0.005
	GNLTA [18]	0.881±0.043	0.895±0.013	0.850±0.032	0.861±0.030	0.846±0.041	0.876±0.020
	ERMC-AGR [45]	0.430±0.023	0.430±0.012	0.259±0.007	0.315±0.007	0.289±0.005	0.347±0.013
	RLMSC [51]	0.551±0.000	0.730±0.000	0.477±0.000	0.513±0.000	0.516±0.000	0.511±0.000
	WETMSC	**0.904±0.011**	**0.929±0.007**	**0.891±0.009**	**0.899±0.008**	**0.887±0.011**	**0.911±0.006**
Caltech101	SSC [11]	0.420±0.015	0.723±0.005	0.303±0.011	0.317±0.012	0.441±0.025	0.248±0.010
	LRR [12]	0.510±0.009	0.728±0.014	0.304±0.017	0.339±0.008	0.627±0.012	0.231±0.010
	RMSC [48]	0.346±0.036	0.573±0.047	0.246±0.031	0.258±0.027	0.457±0.033	0.182±0.031
	DiMSC [49]	0.351±0.000	0.589±0.000	0.226±0.000	0.253±0.000	0.362±0.000	0.191±0.000
	LT-MSC [15]	0.559±0.012	0.788±0.005	0.393±0.007	0.403±0.003	0.670±0.009	0.288±0.012
	GMC [55]	0.331±0.000	0.544±0.000	0.031±0.000	0.081±0.000	0.044±0.000	0.470±0.000
	LMSC [36]	0.566±0.012	0.818±0.004	0.383±0.010	0.392±0.010	0.710±0.014	0.271±0.008
	MLAN [56]	0.579±0.024	0.748±0.020	0.222±0.015	0.265±0.015	0.173±0.009	**0.560±0.016**
	tSVDMSC [16]	0.607±0.005	0.858±0.003	0.430±0.005	0.440±0.010	0.742±0.007	0.323±0.009
	ETLMSC [28]	0.639±0.019	0.899±0.007	0.456±0.017	0.465±0.017	**0.825±0.029**	0.324±0.012
	LRTG [57]	0.490±0.000	0.750±0.000	0.340±0.000	0.350±0.000	0.547±0.000	0.260±0.000
	GNLTA [18]	0.604±0.016	0.875±0.005	0.444±0.017	0.453±0.016	0.776±0.018	0.320±0.015
	ERMC-AGR [45]	0.169±0.000	0.307±0.000	0.153±0.000	0.179±0.000	0.166±0.000	0.194±0.000
	RLMSC [51]	0.512±0.000	0.837±0.000	0.419±0.000	0.429±0.000	0.669±0.000	0.316±0.000
	WETMSC	**0.673±0.028**	**0.902±0.018**	**0.497±0.033**	**0.500±0.025**	0.817±0.029	0.360±0.029

space. Adaptive neighbors clustering (MLAN) [56] adaptively updates the graph for MVC; Anchor graph regularization (ERMC-AGR) [45] learns the embedded anchor graph under matrix factorization framework and introduces CIM to encode noise. Robust localized multi-view subspace clustering (RLMSC) [51] learns the robust consensus representation by fusing the noiseless structures of views. Furthermore, since matrix-based clustering methods may not fully capture the effective information in each view and across views, the proposed method differs from traditional methods by performing tensor optimization through subspace learning. Therefore, 5 tensor learning methods: LT-MSC [15], tSVDMSC [16], ETLMSC [28], LRTG [57] and GNLTA [18] are also used as contrasting methods for more comprehensive comparison. For noise consideration, SSC and RMSC assume that the noise obeys Laplacian distribution. DiMSC and RLMSC assume that the noise obeys Gaussian distribution. ERMC-AGR only assumes that the noise is i.i.d.. The remaining comparison methods assume that the sample-specific noise obeys Gaussian distribution.

ACC, NMI, AR, F-score, Precision, and Recall results of all methods are reported in Tables III and IV. The bolded results represent the best clustering results, and the underlined results represent the second best results. The results show that the proposed WETMSC outperforms all comparison methods on six datasets in most cases. Specifically, for the ACC metrics, WETMSC has an average improvement of $26.85\%, 15.3\%, 25.7\%, 58.35\%, 45.95\%, 20.8\%$ over single-view clustering methods. Compared with the matrix-based clustering methods, WETMSC stacks self-representation matrices into tensor for overall optimization, which avoids the loss of relevant information among views. Obviously, the most competitive methods tSVDMSC, ETLMSC, LRTG and GNLTA are tensor-based MVC ones. Especially, tSVDMSC

TABLE V
CLUSTERING RESULTS (MEAN±STANDARD DEVIATION) ON SIX DATASETS.

Method	ACC	NMI	AR	F-score	Precision	Recall
	BBCSport					
TMSC	0.879±0.000	0.765±0.000	0.784±0.000	0.834±0.000	0.863±0.000	0.807±0.000
WETMSC-$l_{2,1}$	0.993±0.000	0.973±0.000	0.983±0.000	0.987±0.000	0.991±0.000	0.983±0.000
WETMSC+l_1	0.991±0.000	0.965±0.000	0.977±0.000	0.983±0.000	0.988±0.000	0.977±0.000
WETMSC+l_2	0.998±0.000	0.993±0.000	0.997±0.000	0.997±0.000	0.998±0.000	0.997±0.000
WETMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
	UCI-3views					
TMSC	0.955±0.000	0.932±0.000	0.924±0.000	0.932±0.000	0.930±0.000	0.934±0.000
WETMSC-$l_{2,1}$	0.991±0.000	0.979±0.000	0.981±0.000	0.982±0.000	0.982±0.000	0.983±0.000
WETMSC+l_1	0.993±0.000	0.981±0.000	0.983±0.000	0.985±0.000	0.985±0.000	0.985±0.000
WETMSC+l_2	**0.996±0.000**	**0.989±0.000**	**0.991±0.000**	**0.992±0.000**	**0.992±0.000**	**0.992±0.000**
WETMSC	**0.996±0.000**	**0.989±0.000**	**0.991±0.000**	**0.992±0.000**	**0.992±0.000**	**0.992±0.000**
	MSRC-V1					
TMSC	0.991±0.000	0.982±0.000	0.978±0.000	0.981±0.000	0.980±0.000	0.982±0.000
WETMSC-$l_{2,1}$	0.992±0.000	0.982±0.000	0.979±0.000	0.981±0.000	0.980±0.000	0.982±0.000
WETMSC+l_1	0.991±0.000	0.982±0.000	0.978±0.000	0.981±0.000	0.980±0.000	0.982±0.000
WETMSC+l_2	0.995±0.000	0.989±0.000	0.988±0.000	0.990±0.000	0.990±0.000	0.990±0.000
WETMSC	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**	**1.000±0.000**
	MITIndoor					
TMSC	0.684±0.005	0.750±0.007	0.555±0.005	0.562±0.008	0.543±0.005	0.582±0.004
WETMSC-$l_{2,1}$	0.828±0.000	0.884±0.000	0.755±0.000	0.759±0.000	0.735±0.000	0.812±0.000
WETMSC+l_1	0.824±0.013	0.876±0.006	0.754±0.013	0.757±0.012	0.733±0.015	0.785±0.010
WETMSC+l_2	0.862±0.015	0.908±0.005	0.827±0.013	0.829±0.012	0.803±0.016	0.858±0.008
WETMSC	**0.881±0.018**	**0.915±0.015**	**0.835±0.018**	**0.837±0.018**	**0.814±0.022**	**0.862±0.014**
	Scene-15					
TMSC	0.812±0.007	0.858±0.007	0.771±0.003	0.788±0.001	0.743±0.006	0.839±0.003
WETMSC-$l_{2,1}$	0.842±0.029	0.872±0.008	0.808±0.028	0.822±0.025	0.788±0.047	0.860±0.004
WETMSC+l_1	0.842±0.030	0.885±0.007	0.812±0.030	0.825±0.027	0.787±0.048	0.870±0.008
WETMSC+l_2	0.886±0.003	0.909±0.002	0.867±0.005	0.876±0.005	0.866±0.002	0.887±0.008
WETMSC	**0.904±0.011**	**0.929±0.007**	**0.891±0.009**	**0.899±0.008**	**0.887±0.011**	**0.911±0.006**
	Caltech101					
TMSC	0.607±0.005	0.858±0.003	0.430±0.005	0.440±0.010	0.742±0.007	0.323±0.009
WETMSC-$l_{2,1}$	0.644±0.023	0.890±0.012	0.467±0.019	0.476±0.020	0.783±0.024	0.342±0.016
WETMSC+l_1	0.645±0.018	0.890±0.006	0.469±0.019	0.478±0.018	0.784±0.022	0.344±0.016
WETMSC+l_2	0.662±0.011	0.904±0.003	0.471±0.011	0.480±0.011	0.822±0.016	0.352±0.009
WETMSC	**0.673±0.028**	**0.902±0.018**	**0.497±0.033**	**0.500±0.025**	**0.817±0.029**	**0.360±0.029**

and ETLMSC are the extensions of LRR and RMSC from matrix to tensor, respectively, which show better performance than their matrix-based versions on all datasets. The above results further confirm the absolute advantage of tensorizing the self-representation matrices to explore its global low-rank structure in the high-dimensional space, which not only preserves the pairwise correlation of views, but also explores the high-order correlation of multiple views. Compared with tensor-based methods, WETMSC considers different noise distributions and their combinations, and introduces the weighted error entropy to improve the noise robustness. In addition, WETMSC also preserves the inherent structure of sample-specific complementary noise.

To intuitively demonstrate the superior clustering performance of the proposed WETMSC method with i.p.i.d. model, we compare the affinity matrices of WETMSC with the 5 tensor-based methods on MSRC-V1 dataset, as shown in Fig. 4. The results show that the affinity matrix obtained by WETMSC has a distinct block diagonal structure (i.e., high intra-class similarity and low inter-class similarity). The reason is that the proposed WETMSC robustly learns a clean representation tensor, so that the affinity matrix can better represent the similarity between samples. Overall, the proposed WETMSC method can better eliminate the perturbation of noise and improve the accuracy of clustering.

C. Ablation Study and Noise Simulation

We also design 5 variants for the ablation experiments to verify the effectiveness of the weighted error entropy and $l_{2,1}$ norm in the proposed WETMSC method. The baseline is the proposed model shown in Eq. (13) without weighted error entropy, denoted as TMSC. The comparison methods are set as WETMSC-$l_{2,1}$ (WETMSC without $l_{2,1}$ norm), WETMSC+l_1 (The $l_{2,1}$ norm in the WETMSC model is replaced by l_1 norm), WETMSC+l_2 (The $l_{2,1}$ norm in the WETMSC model is replaced by l_2 norm) and WETMSC. Their corresponding clustering results on the six datasets are recorded in Table V. Specifically, the ACC of WETMSC-$l_{2,1}$ is improved by $11.4\%, 3.6\%, 0.1\%, 14.4\%, 3\%, 3.7\%$ compared to the baseline TMSC algorithm. The results of replacing $l_{2,1}$ norm with l_1 norm show that the difference between WETMSC-$l_{2,1}$ and WETMSC+l_1 is very small, which indicates that the weighted error entropy is very robust to Laplacian noise.

TABLE VI
CLUSTERING RESULTS (ACC/NMI) ON NOISY UCI-3VIEWS DATASET

	Method	RMSC	DiMSC	LT-MSC	GMC	LMSC	MLAN	tSVDMSC	ETLMSC	LRTG	GNLTA	WETMSC
ACC	UCI-3views+5%GM	0.810	0.446	0.921	0.749	0.775	0.867	0.956	0.939	0.976	0.952	**0.995**
	UCI-3views+10%GM	0.750	0.480	0.904	0.689	0.691	0.841	0.952	0.869	0.971	0.880	**0.990**
	UCI-3views+15%GM	0.729	0.465	0.893	0.676	0.717	0.830	0.948	0.855	0.962	0.864	**0.985**
	UCI-3views+20%GM	0.711	0.453	0.872	0.656	0.665	0.818	0.943	0.854	0.958	0.845	**0.983**
	UCI-3views+5%GS	0.808	0.494	0.800	0.952	0.851	0.874	0.956	0.913	0.981	0.965	**0.996**
	UCI-3views+10%GS	0.816	0.510	0.792	0.952	0.849	0.873	0.954	0.913	0.980	0.928	**0.995**
	UCI-3views+15%GS	0.806	0.529	0.787	0.748	0.824	0.877	0.954	0.895	0.979	0.930	**0.994**
	UCI-3views+20%GS	0.788	0.486	0.784	0.748	0.843	0.873	0.954	0.885	0.978	0.935	**0.993**
	UCI-3views+5%PS	0.832	0.480	0.921	0.749	0.777	0.874	0.957	0.924	0.977	0.928	**0.994**
	UCI-3views+10%PS	0.788	0.433	0.909	0.702	0.760	0.839	0.952	0.923	0.967	0.864	**0.992**
	UCI-3views+15%PS	0.772	0.464	0.898	0.670	0.712	0.837	0.951	0.885	0.960	0.874	**0.989**
	UCI-3views+20%PS	0.719	0.399	0.880	0.623	0.661	0.835	0.945	0.870	0.953	0.900	**0.982**
NMI	UCI-3views+5%GM	0.727	0.354	0.850	0.827	0.730	0.899	0.901	0.964	0.942	0.968	**0.986**
	UCI-3views+10%GM	0.680	0.376	0.817	0.766	0.651	0.892	0.892	0.925	0.931	0.941	**0.974**
	UCI-3views+15%GM	0.650	0.348	0.793	0.749	0.636	0.892	0.887	0.920	0.912	0.934	**0.964**
	UCI-3views+20%GM	0.626	0.343	0.766	0.720	0.631	0.875	0.878	0.941	0.905	0.926	**0.958**
	UCI-3views+5%GS	0.733	0.381	0.772	0.830	0.809	0.910	0.903	0.958	0.953	0.976	**0.988**
	UCI-3views+10%GS	0.739	0.427	0.770	0.830	0.801	0.909	0.902	0.958	0.951	0.965	**0.987**
	UCI-3views+15%GS	0.731	0.422	0.769	0.824	0.800	0.915	0.902	0.951	0.949	0.964	**0.984**
	UCI-3views+20%GS	0.722	0.415	0.769	0.824	0.801	0.910	0.902	0.948	0.947	0.963	**0.983**
	UCI-3views+5%PS	0.737	0.390	0.855	0.831	0.721	0.910	0.889	0.963	0.945	0.958	**0.983**
	UCI-3views+10%PS	0.696	0.359	0.829	0.778	0.699	0.890	0.889	0.954	0.926	0.945	**0.977**
	UCI-3views+15%PS	0.657	0.369	0.803	0.735	0.645	0.888	0.887	0.935	0.911	0.938	**0.970**
	UCI-3views+20%PS	0.617	0.319	0.776	0.716	0.576	0.892	0.879	0.929	0.897	0.930	**0.956**

TABLE VII
CLUSTERING RESULTS (ACC/NMI) ON NOISY MSRC-V1 DATASET

	Method	RMSC	DiMSC	LT-MSC	GMC	LMSC	MLAN	tSVDMSC	ETLMSC	LRTG	GNLTA	WETMSC
ACC	MSRC-V1+5%MN	0.689	0.600	0.857	0.700	0.640	0.843	0.981	0.737	0.652	0.872	**0.995**
	MSRC-V1+10%MN	0.636	0.576	0.691	0.660	0.455	0.716	0.919	0.730	0.593	0.852	**0.933**
NMI	MSRC-V1+5%MN	0.567	0.529	0.741	0.599	0.522	0.731	0.965	0.717	0.653	0.862	**0.989**
	MSRC-V1+10%MN	0.546	0.509	0.594	0.638	0.317	0.587	0.870	0.702	0.541	0.847	**0.892**
ACC	MSRC-V1+50%SI	0.535	0.656	0.714	0.519	0.481	0.705	0.933	0.719	0.762	0.910	**0.995**
	MSRC-V1+100%SI	0.508	0.599	0.704	0.476	0.328	0.619	0.914	0.672	0.761	0.861	**0.990**
NMI	MSRC-V1+50%SI	0.435	0.616	0.630	0.493	0.437	0.588	0.885	0.671	0.718	0.878	**0.989**
	MSRC-V1+100%SI	0.419	0.536	0.620	0.480	0.214	0.557	0.864	0.665	0.709	0.852	**0.979**
ACC	MSRC-V1+Diagonal	0.686	0.723	0.838	0.766	0.740	0.756	0.942	0.889	0.581	0.910	**0.986**
	MSRC-V1+Block	0.653	0.652	0.866	0.685	0.613	0.824	0.919	0.742	0.795	0.854	**0.981**
NMI	MSRC-V1+Diagonal	0.626	0.600	0.847	0.760	0.728	0.711	0.903	0.891	0.578	0.878	**0.968**
	MSRC-V1+Block	0.618	0.594	0.781	0.713	0.556	0.721	0.869	0.708	0.756	0.840	**0.960**

While WETMSC+l_2 is significantly improved compared with WETMSC-$l_{2,1}$, which also confirms that the assumption of Gaussian noise does exist in real dataset. Furthermore, by combining $l_{2,1}$ norm and weighted error entropy, WETMSC can achieve $12.1\%, 4.1\%, 0.9\%, 19.7\%, 9.2\%, 6.6\%$ improvement on ACC over the baseline algorithm. The results show that the weighted error entropy and $l_{2,1}$ constraints together could improve the clustering performance.

To demonstrate the robustness of the proposed WETMSC method, we add 5%, 10%, 15%, 20% of Gamma noise (GM), Gaussian noise (GS) (Fig. 1(a)) and Poisson noise (PS) into the UCI-3views dataset to construct 12 noises datasets. Besides, we also add 5%, 10% mixed noise (GM, GS, PS) on the MSCR-V1 dataset. In addition, we add 50% and 100% simulated illumination noise (SI) with different mean values, diagonal noise and block noise on MSCR-V1 dataset. Specifically, for SI noise (Fig. 1(c)), each disjoint $d^{(v)} \times 10$ region on each view of MSCR-V1 is added with Gaussian noise that obeys the i.i.d., their means and standard deviations increases gradually at a rate of 0.05 and 0.01 from left to right, respectively. For diagonal noise (Fig. 1(b)), each view of MSCR-V1 is destroyed by Gaussian noise with the mean of 0 and the standard deviation of 1 along the diagonal region. For block noise, each view of MSCR-V1 is destroyed by outliers generated from the uniform distribution from interval $[0, 15]$. Tables VI and VII show that the results of the proposed WETMSC method are still stable for different types of noise, indicating that the WETMSC exhibits superior robustness to i.i.d. and non-i.i.d. noise. However, the clustering performance of other comparison methods based on the i.i.d. assumption has been affected by different noise and cannot handle highly structured non-i.i.d. noise. The above gratifying results are due to the i.p.i.d. model introduced in the proposed method, which randomly processes complex noises.

D. Parameter Analysis and Convergence

The proposed WETMSC method introduces two balancing parameters: λ and β. To choose the optimal parameter values,

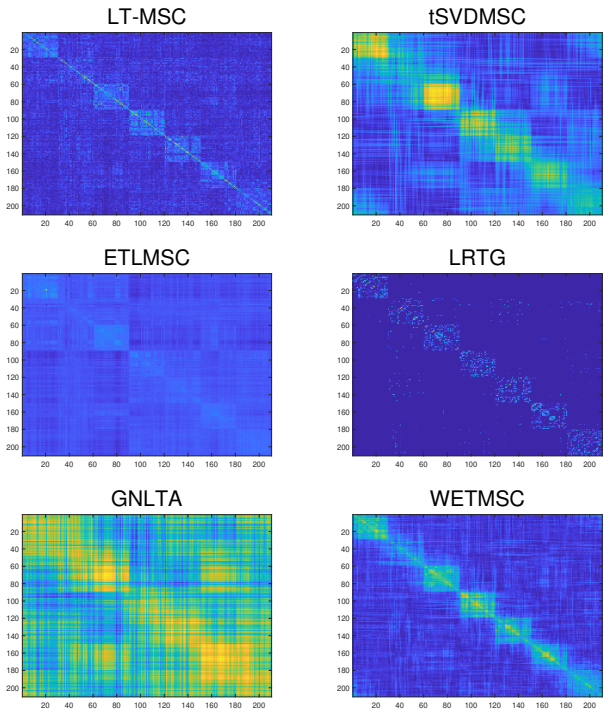

Fig. 4. Structural visualization comparison of the affinity matrices of LT-MSC, tSVDMSC, ETLMSC, LRTG, GNLTA and the proposed WETMSC on MSRC-V1 dataset.

we first fix one parameter and adjust the other parameter in the range $[0.01 : 0.01 : 0.1], [0.1 : 0.1 : 1]$, respectively. The corresponding results of different combinations of parameters λ and β are shown in Fig. 6. It can be seen that the proposed WETMSC achieves good performance of ACC and NMI metrics for all parameters in a wide range of parameter settings, which indicates that the WETMSC algorithm is insensitive to two free parameters.

We also show the relative error and ACC with iterations in Fig. 7, in which $Error1 = \frac{1}{M}\sum_{v=1}^{M}\|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)} - S^{(v)}\|_{\infty}$, $Error2 = \frac{1}{M}\sum_{v=1}^{M}\|Z^{(v)} - H^{(v)}\|_{\infty}$. Obviously, the iteration error monotonically decreases, and the corresponding ACC monotonically increases. After 20 iterations, the iteration error is close to 0, and the ACC is optimal. In addition, we also show the visualization results of clustering for two different iterations ($k = 5, 20$). As can be seen from Fig. 5, for a larger iteration number ($k = 20$), the structure of the clustering results is more obvious with fewer abnormal samples in each cluster than that for a smaller iteration number ($k = 5$). The above phenomena clearly show that the proposed algorithm is well convergent and usually converges within 20 iterations.

E. Running Time

Since the proposed WETMSC method combines the weighted error entropy and tensor low-rank constraint, it shows better performance than other MVC methods when the data contains noise and outliers. However, WETMSC also has limitations. In Fig. 8 (y axis being log scale), we present the running time comparison of the proposed method and the comparison MVSC methods. The results shown that WETMSC operates less efficiently. The reason is that the proposed WETMSC method needs to update the auxiliary variable $\mathbf{u}$ in the solution process. As such, its computational time is longer than that of most of other MVSC methods in comparison based on MSE or $l_{2,1}$ norm.

VII. Conclusions

This paper proposed a novel weighted error entropy regularized tensor learning method for MVSC (WETMSC). WETMSC not only preserves the high-order correlation among views by tensorizing the self-representation matrices from the view dimension, but also introduces weighted error entropy with i.p.i.d. model to encode non-i.i.d. and i.i.d. noise, robustly filtering out complex noise. Thus, WETMSC discovers the subspace structure and removes the complex noise disturbance simultaneously. To optimize the proposed objective function, we first transform the nonconvex weighted error entropy metric into a convex optimization problem resorting to the HQ technology, and finally adopt the alternating update strategy to solve the WETMSC model. The experimental results on six real-world datasets and their noisy datasets confirm the effectiveness of the proposed WETMSC method. The shortcoming of the proposed WETMSC method is high computational cost, which requires more time for multi-sample and high-dimensional datasets. The future research will be to investigate how to improve the computational efficiency of the proposed method for large-scale datasets.

References

[1] G. Jiang, J. Peng, H. Wang, Z. Mi, and X. Fu, "Tensorial multi-view clustering via low-rank constrained high-order graph learning," *IEEE Trans. Circuits Syst. Video Technol.*, 2022.

[2] A. Huang, Z. Wang, Y. Zheng, T. Zhao, and C.-W. Lin, "Embedding regularizer learning for multi-view semi-supervised classification," *IEEE Trans. Image Process.*, vol. 30, pp. 6997–7011, 2021.

[3] Y. Jia, H. Liu, J. Hou, S. Kwong, and Q. Zhang, "Multi-view spectral clustering tailored tensor low-rank representation," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 12, pp. 4784–4797, 2021.

[4] G. Lu, Y. Jia, and J. Hou, "Semi-supervised subspace clustering via tensor low-rank representation," *IEEE Trans. Circuits Syst. Video Technol.*, 2022.

[5] D. Huang, C.-D. Wang, and J.-H. Lai, "Fast multi-view clustering via ensembles: Towards scalability, superiority, and simplicity," *IEEE Trans. Knowl. Data Eng.*, 2023.

[6] M.-S. Chen, L. Huang, C.-D. Wang, and D. Huang, "Multi-view clustering in latent embedding space," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 3513–3520.

[7] D. Huang, C.-D. Wang, J.-S. Wu, J.-H. Lai, and C.-K. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1212–1226, 2019.

[8] S. Huang, H. Zhang, J. Xue, and A. Pižurica, "Heterogeneous regularization-based tensor subspace clustering for hyperspectral band selection," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[9] M. Lan, M. Meng, J. Yu, and J. Wu, "Generalized multi-view collaborative subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 6, pp. 3561–3574, 2021.

[10] S. Huang, H. Zhang, and A. Pižurica, "Hybrid-hypergraph regularized multiview subspace clustering for hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–16, 2021.

[11] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, 2013.

[12] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, 2013.

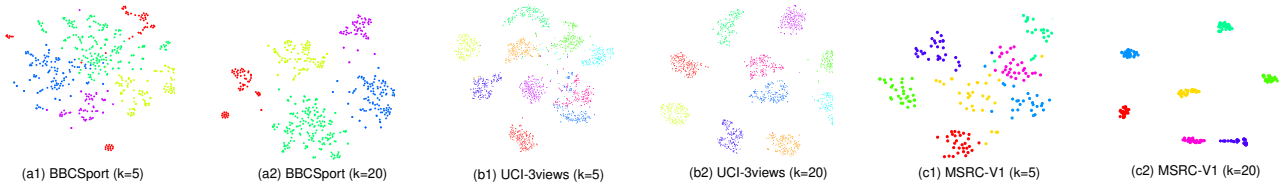

Fig. 5. The proposed WETMSC uses t-SNE on BBCSport, UCI-3views and MSRC-V1 datasets when the iterations k= 5 and 20. Different colors indicate different categories in each dataset.

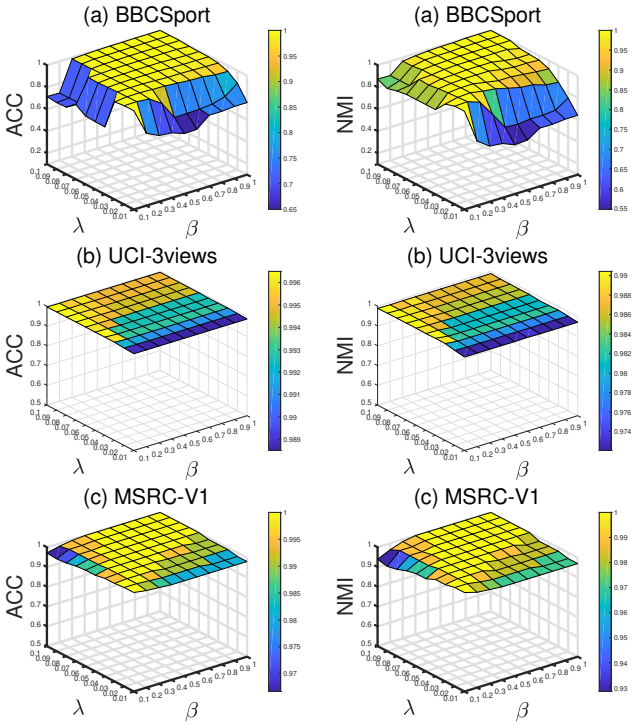

Fig. 6. Parameter tuning with respect to λ and β on (a) BBSport, (b) UCI-3views, (c) MSRC-V1 datasets.

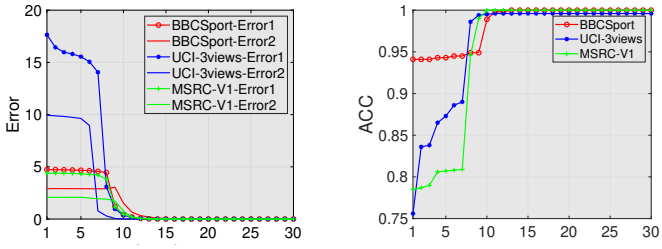

Fig. 7. Error versus iterations on BBSport, UCI-3views and MSRC-V1 datasets.

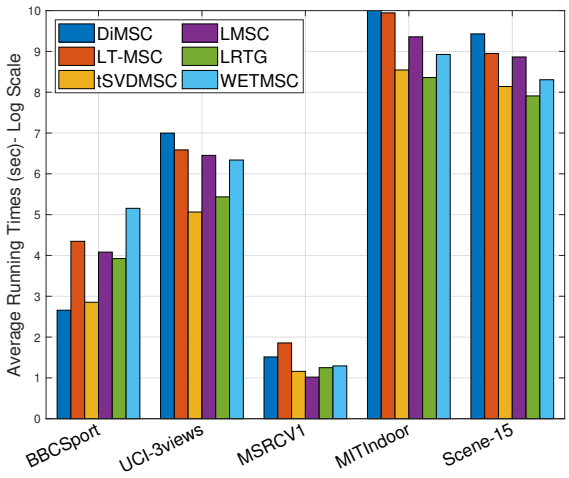

Fig. 8. The running time of compared methods and the proposed WETMSC on five datasets.

[13] C. Lu, H. Min, Z. Q. Zhao, L. Zhu, D. S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2012, pp. 347–360.

[14] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 04, 2020, pp. 4412–4419.

[15] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 1582–1590.

[16] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.

[17] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, 2018.

[18] Y. Chen, X. Xiao, Z. Hua, and Y. Zhou, "Generalized nonconvex low-rank tensor approximation for multi-view subspace clustering," *IEEE Trans. Image Process.*, vol. 30, pp. 4022–4035, 2021.

[19] Y. Tang, Y. Xie, C. Zhang, and W. Zhang, "Constrained tensor representation learning for multi-view semi-supervised subspace clustering," *IEEE Trans. on Multimedia*, 2021.

[20] Z. Li, C. Tang, X. Liu, X. Zheng, W. Zhang, and E. Zhu, "Consensus graph learning for multi-view clustering," *IEEE Trans. on Multimedia*, vol. 24, pp. 2461–2472, 2021.

[21] Y. Xie, J. Liu, Y. Qu, D. Tao, W. Zhang, L. Dai, and L. Ma, "Robust kernelized multiview self-representation for subspace clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 868–881, 2020.

[22] L. Xing, B. Chen, J. Wang, S. Du, and J. Cao, "Robust high-order manifold constrained low rank representation for subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 533–545, 2020.

[23] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Trans. on Multimedia*, vol. 22, no. 8, pp. 1985–1997, 2019.

[24] ——, "Multi-view clustering via simultaneously learning graph regularized low-rank tensor representation and affinity matrix," in *Proc. 2019 IEEE Int. Conf. Multimedia Expo*. IEEE, 2019, pp. 1348–1353.

[25] Y. Chen, S. Wang, F. Zheng, and Y. Cen, "Graph-regularized least squares regression for multi-view subspace clustering," *Knowl.-Based Syst.*, vol. 194, p. 105482, 2020.

[26] Y. Li, J. Zhou, J. Chen, J. Tian, L. Dong, and X. Li, "Robust matrix factorization via minimum weighted error entropy criterion," *IEEE Trans. Comput. Soc. Syst.*, 2021.

[27] Y. Zhang, Y. Wang, X. Chen, X. Jiang, and Y. Zhou, "Spectral–spatial feature extraction with dual graph autoencoder for hyperspectral image clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 12, pp. 8500–8511, 2022.

[28] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view

spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, 2019.
[29] M. Najafi, L. He, and S. Y. Philip, "Error-robust multi-view clustering," in *Proc. IEEE Int. Conf. Big Data*. IEEE, 2017, pp. 736–745.
[30] H. Yong, D. Meng, W. Zuo, and L. Zhang, "Robust online matrix factorization for dynamic background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1726–1740, 2017.
[31] T. Hu, J. Fan, Q. Wu, and D.-X. Zhou, "Learning theory approach to minimum error entropy criterion," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 377–397, 2013.
[32] Y. Wang, Y. Y. Tang, and L. Li, "Robust face recognition via minimum error entropy-based atomic representation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5868–5878, 2015.
[33] D. Erdogmus and J. C. Principe, "An error-entropy minimization algorithm for supervised training of nonlinear adaptive systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1780–1786, 2002.
[34] L. Xing, B. Chen, J. Wang, S. Du, and J. Cao, "Robust high-order manifold constrained low rank representation for subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 2, pp. 533–545, 2021.
[35] C. Tang, X. Zhu, X. Liu, M. Li, P. Wang, C. Zhang, and L. Wang, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Trans. on Multimedia*, vol. 21, no. 7, pp. 1724–1736, 2018.
[36] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, 2020.
[37] M. Meng, M. Lan, J. Yu, and J. Wu, "Multiview consensus structure discovery," *IEEE Trans. Cybern.*, vol. 52, no. 5, pp. 3469–3482, 2022.
[38] Y. Chen, S. Wang, C. Peng, G. Lu, and Y. Zhou, "Partial tubal nuclear norm regularized multi-view learning," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 1341–1349.
[39] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Laplacian sparse coding, hypergraph laplacian sparse coding, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 92–104, 2012.
[40] M. Yin, J. Gao, and Z. Lin, "Laplacian regularized low-rank representation and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 3, pp. 504–517, 2016.
[41] S. Wang, Y. Chen, L. Zhang, Y. Cen, and V. Voronin, "Hyper-laplacian regularized nonconvex low-rank representation for multi-view subspace clustering," *IEEE Trans. Signal Inf. Process. over Networks*, vol. 8, pp. 376–388, 2022.
[42] J. Liu, C. Wang, J. Gao, and J. Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proc. SIAM Int. Conf. Data Min.* SIAM, 2013, pp. 252–260.
[43] H. Wang, Y. Yang, and T. Li, "Multi-view clustering via concept factorization with local manifold regularization," in *Proc. IEEE Int. Conf. Data Min.*, 2016, pp. 1245–1250.
[44] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, 2018.
[45] B. Yang, X. Zhang, Z. Lin, F. Nie, B. Chen, and F. Wang, "Efficient and robust multi-view clustering with anchor graph regularization," *IEEE Trans. Circuits Syst. Video Technol.*, 2022.
[46] M. Sun, P. Zhang, S. Wang, S. Zhou, W. Tu, X. Liu, E. Zhu, and C. Wang, "Scalable multi-view subspace clustering with unified anchors," in *Proc. ACM Int. Conf. Multimedia*, 2021, pp. 3528–3536.
[47] S. Wang, X. Liu, X. Zhu, P. Zhang, Y. Zhang, F. Gao, and E. Zhu, "Fast parameter-free multi-view subspace clustering with consensus anchor guidance," *IEEE Trans. Image Process.*, vol. 31, pp. 556–568, 2021.
[48] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI Conf. Artif. Intell.*, vol. 28, no. 1, 2014.
[49] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 586–594.
[50] S. Wang, Y. Chen, Y. Jin, Y. Cen, Y. Li, and L. Zhang, "Error-robust low-rank tensor approximation for multi-view clustering," *Knowl.-Based Syst.*, vol. 215, p. 106745, 2021.
[51] Y. Fan, J. Liang, R. He, B.-G. Hu, and S. Lyu, "Robust localized multi-view subspace clustering," *arXiv preprint arXiv:1705.07777*, 2017.
[52] R. He, B.-G. Hu, W.-S. Zheng, and X.-W. Kong, "Robust principal component analysis based on maximum correntropy criterion," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1485–1494, 2011.
[53] C. Yang, Z. Ren, Q. Sun, M. Wu, M. Yin, and Y. Sun, "Joint correntropy metric weighting and block diagonal regularizer for robust multiple kernel subspace clustering," *Information Sciences*, vol. 500, pp. 48–66, 2019.
[54] X. Zhang, B. Chen, H. Sun, Z. Liu, Z. Ren, and Y. Li, "Robust low-rank kernel subspace clustering based on the schatten p-norm and correntropy," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 12, pp. 2426–2437, 2019.
[55] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1116–1129, 2020.
[56] F. Nie, G. Cai, J. Li, and X. Li, "Auto-weighted multi-view learning for image clustering and semi-supervised classification," *IEEE Trans. Image Process.*, vol. 27, no. 3, pp. 1501–1511, 2018.
[57] Y. Chen, X. Xiao, C. Peng, G. Lu, and Y. Zhou, "Low-rank tensor graph learning for multi-view subspace clustering," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 1, pp. 92–104, 2021.