Hyper-Laplacian Regularized Multi-View Subspace Clustering With a New Weighted Tensor Nuclear Norm

QINGJIANG XIAO¹, SHIQIANG DU¹,², (Member, IEEE), JINMEI SONG¹,
YAO YU², AND YIXUAN HUANG²

¹Key Laboratory of China’s Ethnic Languages and Information Technology of Ministry of Education, Chinese National Information Technology Research Institute, Northwest Minzu University, Lanzhou, Gansu 730030, China
²College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, Gansu 730030, China

Corresponding author: Shiqiang Du (shiqiangdu@hotmail.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61866033, in part by the Natural Science Foundation of Gansu Province, China, under Grant 18JR3RA369, in part by the University Scientific Research Project of Gansu Province, China, under Grant 2017B-83, in part by the Introduction of Talent Research Project of Northwest Minzu University under Grant xbmuyjrc201904, in part by the Fundamental Research Funds for the Central Universities of Northwest Minzu University under Grant 31920200064 and Grant 1920200097, in part by Gansu Provincial First-Class Discipline Program of Northwest Minzu University under Grant 11080305, in part by the Leading Talent of National Ethnic Affairs Commission (NEAC), in part by the Young Talent of NEAC, and in part by the Innovative Research Team of NEAC (2018) 98.

ABSTRACT In this paper, we present a hyper-Laplacian regularized method WHLR-MSC with a new weighted tensor nuclear norm for multi-view subspace clustering. Specifically, we firstly stack the subspace representation matrices of the different views into a tensor, which neatly captures the higher-order correlations between the different views. Secondly, in order to make all the singular values have different contributions in tensor nuclear norm based on tensor-Singular Value Decomposition (t-SVD), we use weighted tensor nuclear norm to constrain the constructed tensor, which can obtain the class discrimination information of the sample distribution more accurately. Third, from a geometric point of view, the data are usually sampled from a low-dimensional manifold embedded in a high-dimensional ambient space, the WHLR-MSC model uses hyper-Laplacian graph regularization to capture the local geometric structure of the data. An effective algorithm for solving the optimization problem of WHLR-MSC model is proposed. Extensive experiments on five benchmark image datasets show the effectiveness of our proposed WHLR-MSC method.

INDEX TERMS Multi-view subspace clustering, hyper-Laplacian graph, low-rank tensor, weighted tensor nuclear norm.

I. INTRODUCTION

In the era of big data, the data are usually generated from different sources or collected from different views, these data are called multi-view data [1]. In practical applications, various image datasets include various multi-view representations. For example, a scene can be described by two views of image and text, and a picture can be described by three views of edge, fourier and texture. Due to the consistency and complementary information of multi-view data, in recent years, it has received extensive attention from researchers, and a series of multi-view clustering algorithms have been produced. At present, it can be roughly divided into five categories: collaborative training methods [2]–[4], matrix factorization methods [5]–[7], multi-kernel learning methods [8]–[10], subspace learning methods [11]–[15] and graph-based methods [16]–[20]. In this paper, we mainly study the subspace clustering methods of multi-view data.

In recent years, subspace clustering has become one of the research hotspots in the fields of signal processing [21], computer vision (such as face clustering [22], motion segmentation [23]), and pattern recognition [24]. In fact, the data in the high-dimensional space are often located in the low-dimensional subspace. The goal of subspace clustering is to divide the dataset in the high-dimensional space into different clusters, and each cluster corresponds to a low-dimensional subspace. Low-rank representation (LRR) is one of the most widely used techniques in subspace clustering.
we show our methods, including the weighted tensor nuclear norm, the proposed WHLR-MSC model, the related optimization details, the connection with related low-rank tensor methods and the corresponding convergence and complexity analysis. In Section IV, we give the experimental results and analysis. Finally, we conclude the whole paper in Section V.

II. RELATED WORK

This section briefly summarizes the hypergraph related knowledge and tensor nuclear norm.

A. HYPERGRAPH RELATED KNOWLEDGE

For hypergraph \( G = (V, E, S) \), \( E = \{e_j\} \) is the set of hyper edges. Each edge \( e_j \) connects more than two vertices and can be regarded as a subset of the vertex set \( V \). When each hyperedge contains only two vertices, the hypergraph degenerates into a normal graph. \( S \) represents the weight matrix of the hypergraph, and its element \( s(e_j) \) represents the weight of the hyper edge \( e_j \). Hypergraph can be represented by a \( |V| \times |E| \) incidence matrix \( H \). The elements of matrix \( H \) are expressed as \( h(v_i, e_j) \). If vertex \( v_i \) is in edge \( e_j \), then \( h(v_i, e_j) = 1 \), otherwise, \( h(v_i, e_j) = 0 \).

In addition, in order to better describe the hypergraph, the degree of the hypergraph vertex \( v_i \) and the degree of the hyper edge \( e_j \) are denoted as \( d(v_i) = \sum_{e_j \in E} s(e_j) h(v_i, e_j) \) and \( d(e_j) = \sum_{v_i \in V} h(v_i, e_j) \), respectively.

We use diagonal matrices \( D_V \) and \( D_E \) to represent the degree matrix of the vertex and the degree matrix of the hyperedge, and the diagonal elements are \( d(v_i) \) and \( d(e_j) \), respectively.

According to [28], the Laplacian matrix of the hypergraph is defined as

\[
L_H = D_V - HSD_E^{-1}H^T.
\] (1)

B. TENSOR NUCLEAR NORM

A tensor is a multidimensional array, which is also a generalization of the concept of a matrix. The definition of the tensor nuclear norm in the literature [29] is as follows

\[
\|A\|_* = \sum_{m=1}^{M} \zeta_m \|A(m)\|_* ,
\] (2)

where \( \zeta_m \)'s are constants, \( \zeta_m > 0 \) and \( \sum_{m=1}^{M} \zeta_m = 1 \). \( A \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_M} \). \( A(m) \) represents the mode-\( m \) matricization of the tensor \( A \), defined as \( \text{un fold}(A) = A(m) \in \mathbb{R}^{l_m \times (l_1 \times l_2 \times \cdots \times l_{m-1} \times l_{m+1} \times \cdots \times l_M)} \). \( \|A\|_* \) represents the nuclear norm of the tensor \( A \).

The optimized solution of problem (2) can be obtained by solving \( M \) independent optimization problems. The number of singular values of \( A(m) \) is denoted by \( b \), the \( m \)-th (\( m = 1, 2, \cdots , M \)) optimization problem is

\[
\zeta_m \sum_{i=1}^{b} \sigma_i(A(m)),
\] (3)

where \( \sigma_i(A(m)) \) denotes the \( i \) largest singular value of \( A(m) \).
III. THE PROPOSED METHOD

In this section, we first give the weighted tensor nuclear norm and the objective function of our proposed WHLR-MSC method, then introduce an iterative alternating method for WHLR-MSC optimization, and finally discuss the relationship between WHLR-MSC and related low-rank tensor methods and analyze its convergence and complexity.

A. WEIGHTED TENSOR NUCLEAR NORM

In problem (3), the same parameter $\zeta_m$ is used to deal with each singular value, which is not very reasonable, because different singular values may have different importance, so they should be treated differently. For example, given any image, there is a big difference between its non-zero singular values, especially between the first few large singular values and the last few small singular values, the larger singular values are generally related to some significant parts of the image. Therefore, in order to retain the significant part, we should shrink the large singular values smaller, which is not considered in the existing tensor nuclear norm minimization. In order to solve this problem, we introduce the prior information of singular values, and propose a weighted tensor nuclear norm minimization method as shown below:

$$\|\mathbf{A}\|_{\omega,*} = \sum_{m=1}^{M} \xi_m \|\mathbf{A}(m)\|_{\omega,*} = \sum_{m=1}^{M} \xi_m \sum_{i=1}^{\min(1,\ldots,M-1)} \omega_i * \sigma_i(\mathbf{A}(m)).$$

where $\|\mathbf{A}\|_{\omega,*}$ is called the weighted nuclear norm of tensor $\mathbf{A} \in \mathbb{R}^{l_1 \times l_2 \times \cdots \times l_M}$, $\omega_i$ denotes the i element of vector $\omega$.

B. PROPOSED WHLR-MSC

The objective function of our proposed WHLR-MSC method is as follows

$$\sum_{A^{(v)}, B^{(v)}} \|\mathbf{A}\|_{\omega,*} + \lambda_1 \|\mathbf{B}\|_{2,1} + \lambda_2 \sum_{v=1}^{V} (\text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

s.t. $X^{(v)} = X^{(v)} A^{(v)} + B^{(v)}, v = 1, 2, \cdots, V,$

$$\mathbf{A} = \Phi(A^{(1)}, A^{(2)}, \cdots, A^{(V)}), \quad B = [B^{(1)}; B^{(2)}; \cdots; B^{(V)}],$$

where $\lambda_1$ and $\lambda_2$ are balance parameters. $\Phi(\cdot)$ stacks different subspace representation matrices into a 3-way tensor $\mathbf{A}$, the dimension of which is $N \times N \times V$. $\sum_{v=1}^{V} (\text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$ is the hyper-Laplacian regularized term. $\| \cdot \|_{2,1}$ represents the $\ell_{2,1}$-norm.

By using Eqs. (4) and (5), we get the following formula

$$\sum_{A^{(v)}, B^{(v)}} \sum_{m=1}^{M} \gamma_m \|\mathbf{A}(m)\|_{\omega,*} + \|\mathbf{B}\|_{2,1} + \gamma \sum_{v=1}^{V} (\text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

$$\text{s.t. } X^{(v)} = X^{(v)} A^{(v)} + B^{(v)}, v = 1, 2, \cdots, V,$$

$$Z = \Phi(A^{(1)}, A^{(2)}, \cdots, A^{(V)}), \quad B = [B^{(1)}; B^{(2)}; \cdots; B^{(V)}],$$

(6)

where $\gamma_m = \zeta_m/\lambda_1 > 0$ and $\gamma = \lambda_2/\lambda_1$.

An inexact augmented Lagrange multiplier (ALM) [30] is used to solve our objective function. In order to make the objective function separable, we introduce an auxiliary variable $G_m$ instead of $A(m)$, as shown below

$$\sum_{A^{(v)}, B^{(v)}} \sum_{m=1}^{M} \gamma_m \|G_m\|_{\omega,*} + \|\mathbf{B}\|_{2,1} + \gamma \sum_{v=1}^{V} (\text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

s.t. $P_m a = g_m, m = 1, 2, \cdots, M,$

$$X^{(v)} = X^{(v)} A^{(v)} + B^{(v)}, v = 1, 2, \cdots, V,$$

$$\mathbf{A} = \Phi(A^{(1)}, A^{(2)}, \cdots, A^{(V)}), \quad B = [B^{(1)}; B^{(2)}; \cdots; B^{(V)}],$$

(7)

where $g_m$ and $a$ are the vectorization of the matrix $G_m$ and the tensor $\mathbf{A}$, respectively. $P_m$, which is an alignment matrix corresponding to the mode- $m$ expansion, is a permutation matrix, whose function is to align the corresponding elements between $G_m$ and $A(m)$. We construct an augmented Lagrange function for Eq. (7) as follows

$$\mathcal{L}(A^{(1)}, \cdots, A^{(V)}; B^{(1)}, \cdots, B^{(V)}; G_1, \cdots, G_M) = \|\mathbf{B}\|_{2,1} + \sum_{m=1}^{M} (\gamma_m \|G_m\|_{\omega,*} + <\alpha_m, P_m a - g_m>)$$

$$+ \frac{\mu}{2} \|P_m a - g_m\|_F^2 + \sum_{v=1}^{V} (\gamma \text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

$$+ \sum_{v=1}^{V} (Y^{(v)T} X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}),$$

$$+ \frac{\mu}{2} \|X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}\|_F^2,$$

where $\mu$ is the penalty parameter, $\alpha_m$ and $Y_v$ are Lagrange multipliers.

C. OPTIMIZATION WHLR-MSC

Update $A^{(v)}$: Fixing $B^{(v)}$ and $G_m$, we solve the following subproblem to update $A^{(v)}$

$$A^{(v)*} = \arg \min_{A^{(v)}} \sum_{m=1}^{M} (<\alpha_m, P_m a - g_m> + \frac{\mu}{2} \|P_m a - g_m\|_F^2)$$

$$+ \frac{\mu}{2} \|G_m\|_{\omega,*} + \sum_{v=1}^{V} (\gamma \text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

$$+ \frac{\mu}{2} \|X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}\|_F^2,$$

$$= \arg \min_{A^{(v)}} \sum_{m=1}^{M} (<\Omega^v(\alpha_m), \Omega^v(P_m a - g_m)>)$$

$$+ \frac{\mu}{2} \|G_m\|_{\omega,*} + \sum_{v=1}^{V} (\gamma \text{tr}(\mathbf{A}^{(v)} L_h^{(v)} \mathbf{A}^{(v)T}))$$

$$+ \frac{\mu}{2} \|X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}\|_F^2.$$

(9)
where $\Omega^v(\cdot)$ represents the operation of selecting an element and reshaping it into a matrix corresponding to the $v$-th view. We get the close-form solution of $A^{(v)}$ as shown below

$$
A^{(v)} = \left( \sum_{m=1}^M T_m^{(v)} - \sum_{m=1}^M C_m^{(v)} + X^{(v)} T X^{(v)} - X^{(v)T} B^{(v)} + X^{(v)} Y_v B^{(v)} - X^{(v)T} X^{(v)} + 2\gamma L_h^{(v)} \right)^{-1},
$$

where $T_m^{(v)} = \Omega^v(g_m)$ and $C_m^{(v)} = \Omega^v(\alpha_m)$. $\Omega^v(\cdot)$ only selects the $N \times N$ elements corresponding to the $v$-th view, and reshapes them into $N \times N$ dimensional matrices $T_m^{(v)}$ and $C_m^{(v)}$ corresponding to $A^{(v)}$.

**Update $a$:** For the update of $a$, we first construct the tensor $A$ and then vectorize it.

**Update $B$:** Fixing $A^{(v)}$ and $G_m$, we solve the following subproblem to update $B$:

$$
B^* = \arg \min_B \|B\|_{2,1} + \sum_{v=1}^V \left( < Y_v^T, X^{(v)} - X^{(v)} A^{(v)} - B^{(v)} > - B^{(v)} > + \frac{\mu}{2} \|X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}\|_F^2 \right)
$$

$$
= \arg \min_B \frac{\mu}{2} \|B\|_{2,1} + \frac{1}{2} \|B - D\|_F^2.
$$

According to [25], the optimal solution of Eq. (11) is as follows

$$
B^*_{:,i,j} = \begin{cases} 
[D,:,i] - \frac{1}{\mu} \|D,:,i\|_2, & \|D,:,i\|_2 > \frac{1}{\mu} \\
0, & \text{otherwise}
\end{cases}
$$

where $D_{:,j}$ represents the $i$-th column of $D = [D^{(1)},\ldots,D^{(V)}]$, $D^{(j)} \in \mathbb{R}^{N \times 1}$ is the $j$-th view, and $j = 1,\ldots,V$.

**Update $Y_v$:**

$$
Y_v^* = Y_v + (X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}).
$$

**Update $G_m$:**

$$
G_m^* = \arg \min_{G_m} \gamma_m \|G_m\|_{\omega,\cdot} + \|\alpha_m, P_m a - g_m\|
$$

$$
= \arg \min_{G_m} \gamma_m \|G_m\|_{\omega,\cdot} + \frac{1}{2} \|\alpha_m, P_m a - g_m\|_F^2
$$

$$
= \arg \min_{G_m} \gamma_m \|G_m\|_{\omega,\cdot} + \frac{1}{2} \|G_m - F\|_F^2,
$$

where $\Omega_{\omega}(\cdot)$ shows the operation of reshaping a vector into a corresponding matrix according to the $m$-th way unfolding, $F = \Omega_{\omega}(P_m a + \eta_m g_m)$. Then, according to [31], the optimal solution of Eq. (14) is as follows

$$
G_m^* = \Gamma_{\omega,\cdot} \Gamma_{\omega,\cdot}^{-1} [F] = U F J_{\omega,\cdot} \Gamma_{\omega,\cdot}^{-1} (F) V_F^T,
$$

where $J_{\omega,\cdot} \Gamma_{\omega,\cdot}^{-1} (F) = \text{diag} (\beta_1,\beta_2,\ldots,\beta_l)$, $\beta_i = \text{sign}(\sigma_i (F)) \max (\sigma_i (F) - \frac{\sigma_{\max}}{\mu}) + \omega_i, 0)$.

**Update $g_m$:** We update $g_m$ by directly replacing the corresponding elements

$$
g_m^* \leftarrow g_m.
$$

**Update $\alpha_m$:**

$$
\alpha_m^* \leftarrow \alpha_m + P_m a - g_m.
$$

We summarize the entire solution process of Eq. (8) in Algorithm 1.

### Algorithm 1 Algorithm of WHLR-MSC

1. **Input:** Multi-view data set: $X = \{X^{(1)},\ldots,X^{(V)}\}$ with $V$ views, parameters $\gamma_m, \gamma, \omega$ and cluster number $K$.
2. **Initialize:** $A^{(1)} = \cdots = A^{(V)} = 0$, $B^{(1)} = \cdots = B^{(V)} = 0$, $Y_1 = \cdots = Y_v = 0$, $\alpha_1 = \cdots = \alpha_M = 0$, $G_1 = \cdots = G_M = 0$, $\mu = 10^{-5}$, $\rho = 2$, $\epsilon = 10^{-7}$, $\mu_{\max} = 10^8$.
3. Construct matrices $H, L$ and $S$.
4. While not converge do
5. Construct hyper-Laplacian matrices $\{L_h^{(v)}\}_{v=1}^V$ from $\{A^{(v)}\}_{v=1}^V$ by using Eq. (1).
6. Update $A^{(v)}$ by Eq. (10).
7. Update $B^{(v)}$ by Eq. (11).
8. Update $Y_v$ by Eq. (13).
9. Update $a$.
10. For each of $M$ modes do
11. Update $G_m$ by Eq. (15).
12. Update $g_m$ by Eq. (16).
13. Update $\alpha_m$ by Eq. (17).
14. End
15. Update the parameter $\mu$ by $\mu = \min (\rho \mu, \mu_{\max})$.
16. Check the convergence conditions:
17. $\|X^{(v)} - X^{(v)} A^{(v)} - B^{(v)}\|_F < \epsilon$ and $\|P_m a - g_m\|_F < \epsilon$.
18. Obtain the similarity matrix by $W = \frac{1}{V} \sum_{v=1}^V |A^{(v)}| + |A^{(v)T}|$.
19. Use spectral clustering method on similarity matrix $W$.
20. **Output:** Clustering result $Q$.

### D. CONNECTION WITH RELATED LOW-RANK TENSOR METHODS

In this section, we discuss the connection between our proposed model and related low-rank tensor methods. For convenience, we rewrite the objective function (5) of WHLR-MSC by assuming that all singular values have the same contribution in the tensor nuclear norm, the Eq. (5) is changed as follows

$$
\sum_{A^{(v)},B^{(v)}} \|A\| + \lambda_1 \|B\|_{2,1} + \lambda_2 \sum_{v=1}^V \text{tr} (A^{(v)} L_h^{(v)} A^{(v)T})
$$

s.t. $X^{(v)} = X^{(v)} A^{(v)} + B^{(v)} + \epsilon, v = 1,2,\ldots,V$, $A = \phi(A^{(1)},A^{(2)},\ldots,A^{(V)})$, $B = [B^{(1)},B^{(2)},\ldots,B^{(V)}]$.
In this section, we evaluate the effectiveness of our proposed method. We conduct numerical experiments on five benchmark datasets. We use the WHLR-MSC method, which is the objective function of HLR-MSCLRT [13]. In the objective function (18) of HLR-MSCLRT, set $\lambda_2 = 0$, and then Eq. (18) is transformed as shown below:

$$\sum_{A^{(v)}, B^{(v)}} \|A^{(v)}\|_s + \lambda_1 \|B^{(v)}\|_{2,1}$$

s.t. $X^{(v)} = X^{(v)}A^{(v)} + B^{(v)}, v = 1, 2, \cdots, V$,

$A = \Phi(A^{(1)}, A^{(2)}, \cdots, A^{(V)}),$

$B = [B^{(1)}; B^{(2)}; \cdots; B^{(V)}].$ (19)

The Eq. (19) is the model of LT-MSC [12]. Therefore, the LT-MSC and HLR-MSCLRT methods can be regarded as two special cases of our proposed method.

### E. CONVERGENCE AND COMPLEXITY ANALYSIS

In this section, the convergence property and the complexity of our proposed method are analyzed. Since the objective function Eq. (5) is not smooth and has blocks $A^{(v)}, B, G_m$ and $\omega$, it is difficult to prove the theoretical convergence of the proposed WHLR-MSC method. However, WHLR-MSC converges to the local optimal point of the objective function can be proved by Eq. (6). The main reason is that fixing other variables, solving for each single variable, there is a closed-form solution, and each update of each single variable is convex. For the WHLR-MSC method we propose, in Section IV-F, Figure 2 shows the convergence curves of the proposed method on five datasets. We can easily find that the proposed method for solving WHLR-MSC converges quickly.

The main cost of ALM for solving Eq. (7) is to construct $[L_h^{(v)}]_{v=1}^V$, and update $A^{(v)}, B, G_m$ and $Y_v$. Define $N$ to represent the number of data samples, $V$ is the number of views. Regarding the construction of the hyper-Laplacian matrix, it costs $O(VN^2\log(N))$ for all the views. Update $A^{(v)}, G_m$ and $Y_v$ cost $O(VN^3)$, $O(VN^3)$ and $O(V)$ respectively for all views. As for the update of $B$, it costs $O(VN^2)$ in each iteration. Therefore, the total computational complexity for solving the WHLR-MSC model is $O((VN^2\log(N) + VN^3 + VN^2 + VN^3 + V)) + O(N^3)$. In which $r$ is the number of iterations and $O(N^3)$ is the cost of spectral clustering.

### IV. EXPERIMENTS AND ANALYSIS OF RESULTS

In this section, in order to evaluate the effectiveness of our proposed method, we conduct numerical experiments on five benchmark datasets, and use six popular evaluation metrics to measure the clustering results. Then we carry out parameter setting, performance evaluation, convergence analysis, computation cost and weighted values analysis for our proposed method. Finally, in order to visually display the clustering effect, we implement some visualizations of multi-view clustering methods. All experiments are conducted in Matlab 2019a on the computer with Intel(R) Core(TM) i5-4210U processors, 1.70 GHz CPU and 12 GB of memory.

#### A. DATASET DESCRIPTIONS

We select five widely used datasets, which involve different clustering tasks, including toy clustering, scene clustering and face clustering. They are COIL-20, MSRC-v1, Yale, Extended YaleB and ORL respectively. Their detailed descriptions can be found in [13]. We give a brief summary as shown in Table 1. Figure 1 provides some sample examples of images.

#### B. COMPARISON ALGORITHMS

We compare the proposed WHLR-MSC method with ten state-of-the-art methods, which are LRRbest [25], SPCbest [32], ConcatFea [8], ConcatPCA [8], SwMC [33], AMGL [16], MLAN [17], AWP [34], LT-MSC [12] and HLR-MSCLRT [13] respectively.

#### C. EVALUATION METRICS

We use six widely used evaluation metrics to quantitatively evaluate the performance of all clustering methods. These metrics are accuracy (ACC), normalized mutual information

---

**TABLE 1. Statistics of different datasets (# feature).**

| Datasets  | Instances | Views | Clusters | Objective | view1 | view2 | view3 | view4 | view5 |
|-----------|-----------|-------|----------|-----------|-------|-------|-------|-------|-------|
| MSRC-v1   | 210       | 5     | 7        | Scene     | 24    | 576   | 512   | 256   | 254   |
| ORL       | 400       | 3     | 40       | Face      | 4096  | 3304  | 6750  | -     | -     |
| Yale      | 165       | 3     | 15       | Face      | 4096  | 3304  | 6750  | -     | -     |
| YaleB     | 650       | 3     | 10       | Face      | 2500  | 3304  | 6750  | -     | -     |
| COIL-20   | 1440      | 3     | 20       | Toy       | 1024  | 3304  | 6750  | -     | -     |
(NMI), Adjusted Rand Index (ARI), F-score, Precision and Recall respectively. Their detailed definition can be found in [35]. The higher the value of each metric, the better the clustering performance.

D. PARAMETER SETTING

In the WHLR-MSC method we proposed, a total of $M+2$ parameters are involved. For simplicity, we set $\gamma_m, m = 1, \cdots, M$ with equal value, and adjust the parameters $\gamma$ and $\omega$ accordingly. In our experiments, $\gamma_m$ and $\gamma$ are selected in the range of [0.01, 0.05, 0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3] and [0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3, 1.5, 1.7], respectively, and $\omega_j \in (0, 100]$ to obtain the best results. Specifically, $\gamma_m$ and $\gamma$ are set to 0.9 and 0.1, respectively, the weight vector $\omega$ is set to $(1;10;10;1.5)$ on MSRC-v1 dataset. $\gamma_m$ and $\gamma$ are set to 0.8 and 0.1, respectively, the weight vector $\omega$ is set to $(1;10;50)$ on ORL dataset. $\gamma_m$ and $\gamma$ are set to 0.1 and 0.9, respectively, the weight vector $\omega$ is set to $(1;0.5;10)$ on Yale dataset. $\gamma_m$ and $\gamma$ are set to 0.1 and 0.9, respectively, the weight vector $\omega$ is set to $(1;10;100)$ on YaleB dataset. $\gamma_m$ and $\gamma$ are set to 0.01 and 0.9, respectively, the weight vector $\omega$ is set to $(0.5;1;5)$ on COIL-20 dataset. In our method, the hyperedge is constructed in terms of $k$-nearest neighbor, and as a rule of thumb, we set $k = 5$.

E. PERFORMANCE EVALUATION

In order to compare each clustering algorithm fairly, we run all clustering algorithms 10 times and report their mean values and standard deviations. At the same time, we carefully adjust the clustering algorithms with parameters to achieve the best performance. We give the experimental results shown in Tables 2-6. From these tables, we draw the following conclusions.

1) Compared with other methods, our proposed WHLR-MSC method achieves the best clustering performance on all the five datasets. For example, for the ORL dataset, on the six evaluation metrics, our method achieves a significant improvement of approximately 9.5%, 4.2%, 12.8%, 12.5%, 15.7% and 8.8%, respectively, compared with the second-best WHLR-MSCLRT method. On YaleB dataset, our method improves about 1.1%, 1.7%, 2.1%, 1.9%, 2.0% and 1.8% in terms of ACC, NMI, ARI, F-score, Precision and Recall, respectively, compared with the second-best WHLR-MSCLRT method. The main reason is that the weighted tensor nuclear norm fully considers the prior information of singular values, which is of great significance in practice.

2) Multi-view clustering methods are roughly divided into two categories: tensor-based methods and matrix-based methods. The tensor-based methods, including LT-MSC,
TABLE 4. Clustering results (%) on Yale.

| Method        | ACC    | NMI    | ARI    | F-score | Precision | Recall  |
|---------------|--------|--------|--------|---------|-----------|---------|
| LRR_{best}    | 71.5 ± 1.1 | 70.7 ± 1.1 | 49.2 ± 1.9 | 52.5 ± 1.7 | 49.0 ± 2.0 | 56.5 ± 1.3 |
| SPC_{best}    | 69.1 ± 0.6 | 73.0 ± 0.5 | 54.3 ± 0.8 | 57.1 ± 0.8 | 55.9 ± 0.8 | 58.4 ± 0.8 |
| ConcatFea     | 71.0 ± 1.7 | 73.3 ± 1.1 | 55.0 ± 1.1 | 57.9 ± 1.0 | 55.5 ± 1.1 | 60.5 ± 0.9 |
| ConcatPCA     | 63.2 ± 1.2 | 63.0 ± 0.8 | 41.5 ± 1.2 | 45.3 ± 1.1 | 43.2 ± 1.2 | 47.5 ± 1.1 |
| SwMC          | 63.4 ± 4.6 | 65.8 ± 4.3 | 40.2 ± 5.2 | 44.4 ± 4.7 | 38.4 ± 5.3 | 53.1 ± 4.3 |
| AMGL          | 61.0 ± 3.8 | 63.1 ± 2.0 | 36.4 ± 2.1 | 41.0 ± 1.8 | 34.6 ± 2.5 | 50.3 ± 2.2 |
| MLAN          | 57.6 ± 0.0 | 66.3 ± 0.0 | 44.6 ± 0.0 | 48.5 ± 0.0 | 41.6 ± 0.0 | 58.3 ± 0.0 |
| AWP           | 51.5 ± 0.0 | 57.5 ± 0.0 | 37.4 ± 0.0 | 41.8 ± 0.0 | 35.9 ± 0.0 | 50.2 ± 0.0 |
| LT-MSC        | 73.5 ± 0.3 | 75.9 ± 0.5 | 59.2 ± 0.9 | 61.8 ± 0.9 | 59.5 ± 1.1 | 64.2 ± 0.6 |
| HLR-MSCLRT    | 74.3 ± 1.4 | 78.3 ± 1.1 | 62.4 ± 2.7 | 64.8 ± 2.5 | 62.3 ± 3.1 | 67.4 ± 1.9 |
| WHLR-MSC      | 75.5 ± 0.6 | 79.5 ± 0.6 | 64.8 ± 1.2 | 67.0 ± 1.2 | 65.1 ± 1.4 | 69.1 ± 0.9 |

TABLE 5. Clustering results (%) on YaleB.

| Method        | ACC    | NMI    | ARI    | F-score | Precision | Recall  |
|---------------|--------|--------|--------|---------|-----------|---------|
| LRR_{best}    | 45.8 ± 0.6 | 40.9 ± 0.7 | 18.3 ± 0.4 | 27.8 ± 0.4 | 23.7 ± 0.3 | 33.6 ± 0.7 |
| SPC_{best}    | 42.9 ± 0.0 | 40.6 ± 0.1 | 19.8 ± 0.0 | 29.0 ± 0.0 | 25.0 ± 0.0 | 34.5 ± 0.0 |
| ConcatFea     | 48.1 ± 0.1 | 48.5 ± 0.1 | 28.2 ± 0.1 | 36.4 ± 0.1 | 31.6 ± 0.1 | 42.8 ± 0.1 |
| ConcatPCA     | 15.7 ± 0.4 | 7.3 ± 0.8 | 3.7 ± 0.1 | 14.6 ± 0.2 | 12.7 ± 0.2 | 17.1 ± 0.7 |
| SwMC          | 44.9 ± 0.2 | 44.1 ± 0.2 | 18.7 ± 0.2 | 29.0 ± 0.1 | 22.5 ± 0.1 | 40.7 ± 0.1 |
| AMGL          | 39.0 ± 2.1 | 36.8 ± 2.2 | 11.5 ± 1.4 | 23.2 ± 0.7 | 17.4 ± 1.2 | 35.4 ± 5.2 |
| MLAN          | 34.3 ± 0.0 | 31.5 ± 0.0 | 9.0 ± 0.0 | 21.0 ± 0.0 | 15.6 ± 0.0 | 32.1 ± 0.0 |
| AWP           | 40.6 ± 0.0 | 40.8 ± 0.0 | 21.7 ± 0.0 | 30.8 ± 0.0 | 26.1 ± 0.0 | 37.6 ± 0.0 |
| LT-MSC        | 46.4 ± 0.3 | 46.6 ± 0.2 | 26.8 ± 0.3 | 34.6 ± 0.2 | 32.2 ± 0.3 | 37.3 ± 0.2 |
| HLR-MSCLRT    | 64.7 ± 0.4 | 66.0 ± 0.6 | 48.5 ± 0.9 | 53.8 ± 0.8 | 51.5 ± 0.9 | 56.3 ± 0.7 |
| WHLR-MSC      | 65.8 ± 0.3 | 67.7 ± 0.5 | 50.6 ± 0.6 | 55.7 ± 0.5 | 53.5 ± 0.6 | 58.1 ± 0.5 |

TABLE 6. Clustering results (%) on COIL-20.

| Method        | ACC    | NMI    | ARI    | F-score | Precision | Recall  |
|---------------|--------|--------|--------|---------|-----------|---------|
| LRR_{best}    | 71.8 ± 2.1 | 80.5 ± 1.5 | 64.1 ± 3.3 | 66.0 ± 3.1 | 62.2 ± 4.0 | 70.3 ± 2.2 |
| SPC_{best}    | 87.7 ± 0.0 | 93.9 ± 0.0 | 84.8 ± 0.0 | 85.6 ± 0.0 | 80.4 ± 0.0 | 91.5 ± 0.0 |
| ConcatFea     | 85.1 ± 0.0 | 94.4 ± 0.0 | 83.2 ± 0.0 | 84.2 ± 0.0 | 75.6 ± 0.0 | 94.9 ± 0.0 |
| ConcatPCA     | 86.0 ± 0.0 | 92.0 ± 0.0 | 83.9 ± 0.1 | 84.7 ± 0.1 | 83.2 ± 0.1 | 86.4 ± 0.0 |
| SwMC          | 85.2 ± 1.0 | 94.5 ± 0.0 | 83.3 ± 0.0 | 84.3 ± 0.0 | 75.7 ± 0.0 | 94.9 ± 0.0 |
| AMGL          | 76.7 ± 6.1 | 87.2 ± 3.2 | 66.6 ± 8.7 | 68.6 ± 8.0 | 55.8 ± 9.8 | 90.5 ± 1.8 |
| MLAN          | 84.2 ± 0.0 | 92.5 ± 0.0 | 80.3 ± 0.0 | 81.3 ± 0.0 | 72.5 ± 0.0 | 92.7 ± 0.0 |
| AWP           | 71.2 ± 0.0 | 86.4 ± 0.0 | 72.0 ± 0.0 | 73.7 ± 0.0 | 60.0 ± 0.0 | 95.4 ± 0.0 |
| LT-MSC        | 71.8 ± 3.3 | 80.7 ± 1.9 | 65.4 ± 2.7 | 67.2 ± 2.5 | 63.6 ± 3.1 | 71.2 ± 2.4 |
| HLR-MSCLRT    | 87.1 ± 0.6 | 94.0 ± 0.4 | 86.0 ± 0.8 | 86.7 ± 0.8 | 84.8 ± 0.6 | 88.7 ± 1.0 |
| WHLR-MSC      | 88.3 ± 4.0 | 95.6 ± 1.8 | 87.2 ± 5.5 | 87.9 ± 5.2 | 81.5 ± 7.8 | 95.7 ± 1.7 |

HLR-MSCLRT and our proposed method, on all the datasets except COIL-20, have a significant improvement in clustering performance than other methods. The main reason is that tensor-based methods can explore high-order correlations between views. At the same time, it can mine the complementary information between different views more comprehensively.

3) We observe that ConcatFea has better clustering performance than LRR_{best} and SPC_{best} on all the datasets except COIL-20, which indicates that the performance of directly concatenating features is dependent on the data. On most datasets, clustering methods based on multi-view perform better than clustering methods based on single view, which shows that compared with the simple expansion of single view data, the full integration of multi-view data will improve its separability.

4) In short, compared with the two methods of LT-MSC and HLR-MSCLRT, the WHLR-MSC method we proposed
TABLE 7. Average running time on all datasets (in seconds).

| Datasets | ConcatFea | ConcatPCA | SwMC | AMGL | MLAN | AWP | LT-MSC | HLR-MSCLRT | WHLR-MSC |
|----------|-----------|-----------|------|------|------|-----|--------|------------|----------|
| MSRC-v1 | 0.109     | 0.118     | 1.402| 0.306| 0.315| 0.122| 12.269  | 6.778      | 7.330    |
| ORL      | 0.744     | 1.142     | 3.208| 1.116| 1.592| 0.603| 107.952 | 35.908     | 56.387   |
| Yale     | 0.194     | 0.268     | 0.458| 0.198| 0.339| 0.155| 27.676  | 10.921     | 11.621   |
| YaleB    | 1.125     | 2.141     | 30.413| 2.185| 3.657| 0.956| 231.727 | 127.021    | 147.216  |
| COIL-20  | 4.290     | 8.797     | 231.805| 19.317| 22.493| 11.047| 1174.888 | 818.158    | 796.924  |

achieves the best performance on all the five datasets. In addition, these two methods are most related to our method, which shows that our method uses graph regularization term to capture the local structure of the data and uses weighted tensor nuclear norm to treat singular values differently, which is effective for improving clustering performance.

F. CONVERGENCE ANALYSIS
As shown in Figure 2, we analyze the convergence on the MSRC-v1, Yale, YaleB, ORL and COIL-20 datasets. Among them, we record the reconstruction error of each iteration. Reconstruction Error = ∥X^{(v)} - X^{(v)}A^{(v)} - B^{(v)}∥_∞. We can see that WHLR-MSC converges very quickly. After 20 iterations, the error is close to 0, which shows that our algorithm can stably converge after a certain number of iterations.

FIGURE 2. Empirical convergence versus iterations.

G. COMPUTATION COST
In Table 7 we report the average running time (in seconds) of the proposed WHLR-MSC method and the eight multi-view clustering methods on all datasets. We can easily see that the multi-view subspace clustering methods (including LT-MSC, HLR-MSCLRT and WHLR-MSC) cost longer than other comparison methods. Compared with LT-MSC and HLR-MSCLRT, the average running time of our proposed method is at an intermediate level. In short, although our proposed WHLR-MSC method is time-consuming, it significantly improves the clustering performance.

H. WEIGHTED VALUES ANALYSIS
We analyze the influence of weighted coefficient ω on clustering performance on MSRC-v1 and ORL datasets respectively, as shown in Figure 3 and Figure 4. In each figure, the x-axis and y-axis represent the weighted value and the evaluation metric value, respectively. We can easily see that the weighted parameter ω has a large influence for clustering effect, which shows the importance of fully considering the prior knowledge of singular values. On the MSRC-v1 and ORL datasets, our method overall has good performances (ACC and NMI) with ω = [1 100 10 0.5] and ω = [1 10 50]. This may be because each singular value of the matrix has different roles to the robustness of clustering algorithms in the presence of illumination and other noises.

FIGURE 3. Analysis of the weighted coefficient ω over the MSRC-v1 dataset (γ_m = 0.9, γ = 0.1).

FIGURE 4. Analysis of the weighted coefficient ω over the ORL dataset (γ_m = 0.8, γ = 0.1).

I. VISUALIZATION
The visualization of the clustering results of the nine multi-view clustering algorithms on the MSRC-v1 dataset is shown in Figure 5. We can clearly see that almost all multi-view clustering algorithms can divide the MSRC-v1 dataset into different clusters. However, compared with the other eight methods, the distribution of similar samples in the WHLR-MSC is more compact.
V. CONCLUSION

In this paper, we propose a hyper-Laplacian regularized multi-view subspace clustering with a new weighted tensor nuclear norm (WHLR-MSC) method. In our model, low-rank tensor is used to capture the global structure of the data and explore the high-order correlation between different views, hyper-Laplacian graph regularization is used to capture the local geometric structure of the data. In addition, the weighted tensor nuclear norm fully considers the prior information of the singular values, and can more accurately obtain the classification information of the sample distribution. An efficient algorithm is proposed to solve and optimize our model. Extensive experiments on five benchmark image datasets show that our method is superior to state-of-the-art methods.

REFERENCES

[1] Y. Yang and H. Wang, “Multi-view clustering: A survey,” Bid Data Mining Anal., vol. 1, no. 2, pp. 83–107, Jun. 2018.
[2] A. Kumar and H. Daume, “A co-training approach for multi-view spectral clustering,” in Proc. 28th Int. Conf. Mach. Learn. (ICML), 2011, pp. 393–400.
[3] A. Kumar, P. Rai, and H. Daume, “Co-regularized multi-view spectral clustering,” in Proc. Adv. Neural Inf. Process. Syst., vol. 24, 2011, pp. 1413–1421.
[4] H. Xu, X. Zhang, W. Xia, Q. Gao, and X. Gao, “Low-rank tensor constrained co-regularized multi-view spectral clustering,” Neural Netw., vol. 132, pp. 245–252, Dec. 2020.
[5] Z. Wang, X. Kong, H. Fu, M. Li, and Y. Zhang, “Feature extraction via multi-view non-negative matrix factorization with local graph regularization,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2015, pp. 3500–3504.
[6] 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, Oct. 2018.
[7] D. Cai, X. He, J. Han, and T. S. Huang, “Graph regularized nonnegative matrix factorization for data representation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 8, pp. 1548–1560, Aug. 2011.
[8] W. Weng, W. Zhou, J. Chen, H. Peng, and H. Cai, “Enhancing multi-view clustering through common subspace integration by considering both global similarities and local structures,” Neurocomputing, vol. 378, pp. 375–386, Feb. 2020.
[9] M. Gönen and A. A. Margolin, “Localized data fusion for kernel K-means clustering with application to cancer biology,” in Proc. Adv. Neural Inf. Process. Syst., vol. 27, 2014, pp. 1305–1313.
[10] L. Du, P. Zhou, L. Shi, H. Wang, M. Fan, W. Wang, and Y.-D. Shen, “Robust multiple kernel K-means using $\ell_{1,1}$-norm,” in Proc. 24th Int. Joint Conf. Artif. Intell., 2015, pp. 3476–3482.
[11] Y. Chen, X. Xiao, and Y. Zhou, “Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix,” Pattern Recognit., vol. 106, Oct. 2020, Art. no. 107441.
[12] 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. (ICCV), Dec. 2015, pp. 1582–1590.
[13] G.-F. Lu, Q.-R. Yu, Y. Wang, and G. Tang, “Hyper-Laplacian regularized multi-view subspace clustering with low-rank tensor constraint,” Neural Netw., vol. 125, pp. 214–223, May 2020.
[14] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, “Tensor-SVD based graph learning for multi-view subspace clustering,” in Proc. AAAI Conf. Artif. Intell., 2020, vol. 34, no. 4, pp. 3930–3937.
[15] 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. (CVPR), Jun. 2015, pp. 586–594.

[16] F. Nie, J. Li, and X. Li, “Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification,” in Proc. IJCAI, 2016, pp. 1881–1887.

[17] F. Nie, G. Cai, and X. Li, “Multi-view clustering and semi-supervised classification with adaptive neighbours,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 2408–2414.

[18] 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, May 2019.

[19] H. Wang, Y. Yang, B. Liu, and H. Fujita, “A study of graph-based system for multi-view clustering,” Knowl.-Based Syst., vol. 163, pp. 1009–1019, Jan. 2019.

[20] K. Zhan, C. Zhang, J. Guan, and J. Wang, “Graph learning for multi-view clustering,” IEEE Trans. Cybern., vol. 48, no. 10, pp. 2887–2895, Oct. 2018.

[21] A. Adler, M. Elad, and Y. Hel-Or, “Probabilistic subspace clustering via sparse representations,” IEEE Signal Process. Lett., vol. 20, no. 1, pp. 63–66, Jan. 2013.

[22] X. Zhang, D. Phung, S. Venkatesh, D. S. Pham, and W. Liu, “Multi-view subspace clustering for face images,” in Proc. Int. Conf. Digit. Image Computing: Tech. Appl. (DICTA), Nov. 2015, pp. 1–7.

[23] S. Li, K. Li, and Y. Fu, “Temporal subspace clustering for human motion segmentation,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 4453–4461.

[24] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, “Latent multi-view subspace clustering,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 4279–4287.

[25] 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, Jan. 2013.

[26] 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.

[27] 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, Dec. 2019.

[28] D. Zhou, J. Huang, and B. Scholkopf, “Learning with hypergraphs: Clustering, classification, and embedding,” in Proc. Adv. Neural Inf. Process. Syst., vol. 19, 2006, pp. 1601–1608.

[29] J. Liu, P. Musialski, P. Wonka, and J. Ye, “Tensor completion for estimating missing values in visual data,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 1, pp. 208–220, Jan. 2013.

[30] Z. Lin, R. Liu, and Z. Su, “Linearized alternating direction method with adaptive penalty for low-rank representation,” 2011, arXiv:1109.0367. [Online]. Available: https://arxiv.org/abs/1109.0367

[31] K. Chen, H. Dong, and K.-S. Chan, “Reduced rank regression via adaptive nuclear norm penalization,” Biometrika, vol. 100, no. 4, pp. 901–920, 2012.

[32] A. Y. Ng, M. I. Jordan, and Y. Weiss, “On spectral clustering: Analysis and an algorithm,” in Proc. Adv. Neural Inf. Process. Syst., 2002, pp. 849–856.

[33] F. Nie, J. Li, and X. Li, “Self-weighted multiview clustering with multiple graphs,” in Proc. 26th Int. Joint Conf. Artif. Intell., Aug. 2017, pp. 2564–2570.

[34] F. Nie, L. Tian, and X. Li, “Multiview clustering via adaptively weighted proximities,” in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp. 2022–2030.

[35] K. Zhan, F. Nie, J. Wang, and Y. Yang, “Multiview consensus graph clustering,” IEEE Trans. Image Process., vol. 28, no. 3, pp. 1261–1270, Mar. 2019.

Q. Xiao et al.: Hyper-Laplacian Regularized Multi-View Subspace Clustering

QINGJIANG XIAO received the B.S. degree in computer science and technology from Henan Urban Construction College, Pingdingshang, China, in 2019. He is currently pursuing the M.S. degree with the College of China National Institute of Information Technology, Northwest Minzu University, Lanzhou, China. His current research interests include machine learning and software engineering.

SHIQIANG DU (Member, IEEE) received the Ph.D. degree in radio physics from the School of Information Science and Engineering, Lanzhou University, Lanzhou, China, in 2017. He was a Visiting Scholar with the Department of Mathematics and Statistics, Utah State University, Logan, UT, USA, in 2013. From 2019 to 2020, he was a Visiting Scholar with the Department of Engineering, University of Modena and Reggio Emilia, Modena, Italy. He is currently an Associate Professor with the College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou. His current research interests include machine learning, image processing, neural networks, and pattern recognition.

JINMEI SONG received the B.S. degree in computer science and technology from Hainan Normal University, Haikou, China, in 2020. She is currently pursuing the M.S. degree with the College of China National Institute of Information Technology, Northwest Minzu University, Lanzhou, China. Her current research interests include machine learning and software engineering.

YAO YU received the B.S. degree in information management and information system from Hunan University of Chinese Medicine, Changsha, China, in 2020. She is currently pursuing the M.S. degree with the College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, China. Her current research interests include machine learning and software engineering.

YIXUAN HUANG received the B.S. degree in computer science and technology from the College of Mathematics and Computer Science, Northwest Minzu University, Lanzhou, China, in 2019, where he is currently pursuing the M.S. degree. His current research interests include machine leaning and software engineering.

* * *