WILEY | Hindawi

Research Article

Robust Semisupervised Nonnegative Local Coordinate Factorization for Data Representation

Wei Jiang [iD],[1] Qian Lv,[1] Chenggang Yan,[2] Kewei Tang,[1] and Jie Zhang[1]

[1]*School of Mathematics, Liaoning Normal University, Dalian 116029, China*
[2]*Institute of Information and Control, Hangzhou Dianzi University, Hangzhou 541004, China*

Correspondence should be addressed to Wei Jiang; swxxjw@aliyun.com

Obtaining an optimum data representation is a challenging issue that arises in many intellectual data processing techniques such as data mining, pattern recognition, and gene clustering. Many existing methods formulate this problem as a nonnegative matrix factorization (NMF) approximation problem. The standard NMF uses the least square loss function, which is not robust to outlier points and noises and fails to utilize prior label information to enhance the discriminability of representations. In this study, we develop a novel matrix factorization method called robust semisupervised nonnegative local coordinate factorization by integrating robust NMF, a robust local coordinate constraint, and local spline regression into a unified framework. We use the $l_{2,1}$ norm for the loss function of the NMF and a local coordinate constraint term to make our method insensitive to outlier points and noises. In addition, we exploit the local and global consistencies of sample labels to guarantee that data representation is compact and discriminative. An efficient multiplicative updating algorithm is deduced to solve the novel loss function, followed by a strict proof of the convergence. Several experiments conducted in this study on face and gene datasets clearly indicate that the proposed method is more effective and robust compared to the state-of-the-art methods.

1. Introduction

Owing to the rapid development of data collection and storage techniques, there has been an increase in the demand for effective data representation approaches [1] to cope with image and gene information, particularly in the fields of pattern recognition, machine learning, and gene clustering. For large databases, an efficient representation of data [2–4] can improve the performance of numerous intelligent learning systems such as those used for classification and clustering analysis. In many application fields, the input samples are represented in high-dimensional form, which is infeasible for direct calculation. The efficiency and effectiveness of learning models exponentially decrease with each increase in the dimensionality of input samples, which is generally referred to as the "curse of dimensionality." Accordingly, dimensionality reduction [5–7] is becoming increasingly important as it can overcome the curse of dimensionality, enhance the learning speed, and even offer critical insights into the essence of the issue. In general, dimensionality reduction methods can be divided into two categories: feature extraction [5, 8, 9] and selection [10–14]. Feature selection involves selecting discriminative and highly related features from an input feature set, whereas feature extraction combines original features to form new features of data variables.

In recent years, there has been an increasing interest in feature extraction. Many feature extraction methods are designed to obtain a low-dimensional feature of high-dimensional data. These methods include singular value decomposition (SVD), principal component analysis (PCA) [5], nonnegative matrix factorization (NMF) [15, 16], and concept factorization (CF) [17]. Despite the different motivations of these models, they can all be interpreted as matrix decomposition, which often finds two or more low-dimensional matrices to approximate the original matrix. Factorization leads to a reduced representation of high-dimensional data and belongs to the category of methods employed for dimension reduction.

Unlike PCA [5] and SVD, NMF [15, 16] factorizes a sample matrix as a product of two matrices constrained by nonnegative elements. One matrix comprises new basis vectors that reveal the semantic structure, and the other matrix can be regarded as the set of coefficients composed of linear combinations of all sample points based on the new bases. Owing to their ability to extract the most discriminative features and their feasibility in computation, many extension versions [4, 18, 19] of NMF have been developed from various perspectives to enhance the original NMF. Sparseness-constrained NMF [20] has been introduced by adding l_1 norm minimization on the learned factor matrices to enhance sparsity for data representation. Fisher's criterion [21] has been incorporated into NMF formulation and is used to achieve discriminant representation. The semi- and convex-NMF formulations [22] relax the nonnegativity constraint of NMF by allowing the basis and coefficient matrices to have mixed signs, thereby extending the applicability of the method. Liu et al. [23] proposed a constrained NMF in which the label information is incorporated into the standard NMF for data representation. Cai et al. [24] extended NMF and proposed a graph-regularized NMF (GNMF) scheme, which imposes intrinsic geometry latent in a high-dimensional dataset onto the traditional NMF using an affinity graph. Chen et al. [9] presented a nonnegative local coordinate factorization (NLCF) method that imposes locality constraint onto the original NMF to explore faithful intrinsic geometry.

Traditional NMF and its variants usually adopt the square Euclidean distance to measure the approximation error. Although it has a solid theoretical foundation in mathematics and has shown encouraging performance in most cases, the square Euclidean distance is not always optimal for decomposition of a data matrix. The squared error has proved to be the best for both Gaussian and Poisson noise [25]. However, in real-world applications, data that violate the assumptions are usually involved. The squared loss is sensitive to outlier points and noises when the reconstruction error is measured. Even a single outlier point may sometimes easily dominate the objective function. In recent years, some variants have been presented to enhance the robustness of the classical NMF. A robust type of NMF that factorizes the sample matrix as the summation of two nonnegative matrices and one sparse error matrix was presented by Zhang et al. [26]. Zhang et al. [27] presented a robust NMF (RNMF) using the $l_{2,1}$ norm objective function, which can deal with outlier points and noises. Zhang et al. [28] presented a robust nonnegative graph-embedding framework (RNGE) that can simultaneously cope with noisy labels, noisy data, and uneven distribution.

Supervised learning algorithms [29–32] generally can achieve better performance than unsupervised learning techniques when label information is available in many applications. The motivation of semisupervised learning methods [33–38] is to employ numerous unlabeled samples as well as relatively few labeled samples to construct a better high-dimensional data analysis model. A surge of research interest in graph-based semisupervised learning techniques [37–39] [40] has recently occurred. Gaussian fields and harmonic functions (GFHF) [33] is an efficient and effective semisupervised learning methods in which the predicted label matrix is reckoned on the graph with respect to manifold smoothness and label fitness. Xiang et al. [37] presented a method called local spline regression (LSR) in which an iterative algorithm is built on local neighborhoods through spline regression. Han et al. [38] presented a model of video semantic recognition using semisupervised feature selection via spline regression (S2FS2R). These methods not only consider label information but also employ the local and global structure consistency assumption.

Despite NMF's appealing advantages, it suffers from the following problems in real-world applications: (1) data may often be contaminated by noise and outliers due to illumination (e.g., specular reflections), image noises (e.g., scanned image data), occlusion (e.g., sunglasses and scarf in front of a face), among others. Although NMF can deal with noise in the test data to some extent, it will suffer from severe performance degradation when the training samples have noise. (2) In an NMF method, a data point may be represented by the base vectors, which are far from the data point, resulting in poor clustering performance. The standard NMF does not preserve the locality during its decomposition process, whereas local line coding can preserve such properties. (3) One of the challenges for classification tasks in the real world is the lack of labeled training data. Therefore, data labeled by an expert is often used as an alternative. Unfortunately, designating labels requires considerable human effort and is thus time-consuming and difficult to manage. In addition, an accurate label may require expert knowledge. However, unlabeled samples are relatively easy to obtain.

To address all the aforementioned issues, we present an efficient and effective matrix factorization framework called robust semisupervised nonnegative local coordinate factorization (RSNLCF) in which both data reconstruction functions and a local coordinate constraint regularization term are formulated in a $l_{2,1}$ norm manner to make our model robust to outlier points and noises. By integrating Green's functions and a set of primitive polynomials into the local spline, the local and global label consistency of data can be characterized based on their distribution. The main work of our study and its contributions are summarized as follows:

(i) The proposed RSNLCF model is robust to outlier points and noises as a result of employing the $l_{2,1}$ norm formulations of NMF and a local coordinate constraint regularization term. In addition, to guarantee that the data representation is discriminative, local spline regression over labels is exploited.

(ii) Unlike traditional dimension reduction approaches that treat feature extraction and selection separately, the proposed RSNLCF algorithm integrates the two aspects into a single optimization framework.

(iii) We present an efficient algorithm to solve the presented RSNLCF model and provide the proof of rigorous convergence and correctness analysis of our model.

The remainder of this paper is organized as follows. Related studies are introduced in Section 2. We introduce our RSNLCF method and the optimization scheme in Section 3 and offer a convergence proof in Section 4. We describe and analyze the results of our experiments in Section 5. We conclude and discuss future work in Section 6.

2. Related Work

In this section, we summarize the notations and definitions of norm used in this study and briefly review NMF.

2.1. Notations and Definitions. Matrices and vectors are denoted by boldface capital and lowercase letters, respectively. $\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}$ denotes the l_p norm of the vector $\mathbf{x} \in \mathbb{R}^n$. $\mathbf{x}^i$ and $\mathbf{x}_j$ denote the ith row and the jth column of matrix $\mathbf{X} = (x_{ij})$, respectively. x_{ij} is the element in the ith row and jth column of $\mathbf{X}$, $\mathrm{Tr}[\mathbf{X}]$ denotes the trace of $\mathbf{X}$ if $\mathbf{X}$ is a square matrix, and $\mathbf{X}^T$ denotes the transposed matrix of $\mathbf{X}$. The Frobenius norm of the matrix $\mathbf{X} \in \mathbb{R}^{M \times N}$ is defined as

$$\|\mathbf{X}\|_F = \sqrt{\sum_{i=1}^M \sum_{j=1}^N x_{ij}^2}. \tag{1}$$

The $l_{2,1}$ norm of a matrix is defined as

$$\|\mathbf{X}\|_{2,1} = \sum_{i=1}^M \|\mathbf{x}^i\|_2 = \sum_{i=1}^M \sqrt{\sum_{j=1}^N x_{ij}^2} = \mathrm{Tr}\left[\mathbf{X}^T \mathbf{D}\mathbf{X}\right], \tag{2}$$

where $\mathbf{D}$ is a diagonal matrix with $\mathbf{D}_{ii} = 1/2\|\mathbf{x}^i\|_2$. However, $\|\mathbf{x}^i\|_2$ could approach zero. For this case, we define $D_{ii} = 1/2 \|\mathbf{x}^i\|_2 + \varepsilon$, where ε is a very small constant.

Assume that the matrix samples are represented as $\mathbf{X} = \{\mathbf{x}_i|_{i=1}^L, \mathbf{x}_j|_{j=L+1}^N\}$, where $\mathbf{x}_i|_{i=1}^L, \mathbf{x}_j|_{j=L+1}^N$ denotes labeled and unlabeled data, respectively. The labels of $\mathbf{x}_i|_{i=1}^L$ are denoted as $l_i \in \{1, 2, \dots, L_c\}$ with L_c being the total number of categories. Let $\mathbf{F} \in \mathbb{R}^{L \times L_c}$ be a label indicator binary matrix with the jth entry $f_{ij} = 1$ if and only if $\mathbf{x}_i$ is labeled with the jth class; $f_{ij} = 0$ otherwise. We also introduce a predicted label matrix $\mathbf{Y} \in \mathbb{R}^{N \times L_c}$, where each row is the predicted label vector of the data $\mathbf{x}_i$.

2.2. NMF. Given a nonnegative matrix $\mathbf{X} \in \mathbb{R}_+^{M \times N}$, each column of $\mathbf{X}$ is a sample point. The main idea of NMF is to find two nonnegative matrices $\mathbf{U} = [u_{ik}] \in \mathbb{R}_+^{M \times K}$ and $\mathbf{V} = [v_{jk}] \in \mathbb{R}_+^{K+N}$ that minimize the Euclidean distance between $\mathbf{X}$ and $\mathbf{UV}$. The corresponding optimization problem is as follows:

$$\min_{\mathbf{U},\mathbf{V}} \quad \|\mathbf{X} - \mathbf{UV}\|_F^2 \tag{3}$$
$$\text{s.t.} \quad \mathbf{U} \geq 0, \mathbf{V} \geq 0,$$

where $\|\cdot\|_F$ is the Frobenius norm. To solve the objective function, Lee and Seung [15] proposed an iterative multiplicative updating algorithm as follows:

$$u_{jk}^{(t+1)} \leftarrow u_{jk}^{(t)} \frac{\left(\mathbf{XV}^T\right)_{jk}}{\left(\mathbf{UVV}^T\right)_{jk}},$$
$$\tag{4}$$
$$v_{ki}^{(t+1)} \leftarrow v_{ki}^{(t)} \frac{\left(\mathbf{U}^T\mathbf{X}\right)_{ki}}{\left(\mathbf{U}^T\mathbf{UV}\right)_{ki}}.$$

By NMF, each column of $\mathbf{U}$ and $\mathbf{u}_i$ can be viewed as the basis, while the matrix $\mathbf{V}$ can be treated as the set of the coefficients. Each sample point $\mathbf{x}_i$ is approximated by a linear combination of the K bases, weighted by components of $\mathbf{V}$.

3. The Proposed RSNLCF Framework

In this section, we introduce our novel learning method for image clustering (RSNLCF), which is used to find an effective and robust representation of data.

3.1. Robust Sparse NMF. The square loss function based on the Frobenius norm is used to learn the data representations in NMF. However, it is very sensitive to outlier points and noises. Therefore, our robust representation model is represented as

$$\min_{\mathbf{U},\mathbf{V}} \quad \|\mathbf{X} - \mathbf{UV}\|_{2,1} + \lambda\|\mathbf{V}\|_{2,1}, \tag{5}$$

where $\lambda > 0$ is the regularization parameter. Because the $l_{2,1}$ norm reduces the components occupied by the large magnitude of error in the loss function, the corrupted samples never dominate the objective function. In this sense, the loss function $\|\mathbf{X} - \mathbf{UV}\|_{2,1}$ is insensitive to outlier points and noises. Meanwhile, the regularization term $\|\mathbf{V}\|_{2,1}$ ensures that $\mathbf{V}$ is sparse in rows. This means that some of $\mathbf{V}$'s rows approximate zero. Consequently, $\mathbf{V}$ can be considered the combination coefficient for the most discriminative features. Feature selection is then achieved by $\mathbf{V}$, where only the features related to the nonzero rows in $\mathbf{V}$ are chosen.

3.2. Robust Local Coordinate Constraint. Motivated by the concept of local coordinate coding [41], we present a robust local coordinate constraint as a regularization term for image clustering. First, we define coordinate coding.

Definition 1. Coordinate coding [41] can be written as concept pair (γ, C), where C is defined as a set of anchor points with d dimensions and γ is a map of $\mathbf{x} \in \mathbb{R}^d$ to $[\gamma_v(\mathbf{x})]_{v \in C} \gamma_v(\mathbf{x})v$. It induces the following physical approximation of $\mathbf{x}$ in $\mathbb{R}^d$: $\gamma(\mathbf{x}) = \sum_{v \in C} \gamma_v(\mathbf{x})v$.

For the local coordinate coding system, NMF can be considered as coordinate coding in which the columns of the matrix $\mathbf{U}$ can be viewed as a set of anchor points, and each column of the coefficient matrix $\mathbf{V}$ represents the corresponding coordinate coding for each data point. We might further hope that each sample point is represented as a linear combination of only a few proximate anchor points. A natural assumption here would be that if $\mathbf{x}_i$ is far away from the anchor points $\mathbf{u}_k$, then its coordinate coding v_{ki} with respect to $\mathbf{u}_k$ will tend to be zero and thus achieve sparsity and

locality simultaneously. The local coordinate constraint [41] can be defined as follows:

$$\min_{\mathbf{U},\mathbf{V}} \quad \sum_{i=1}^{N}\sum_{k=1}^{K} |v_{ki}|\|\mathbf{u}_k - \mathbf{x}_i\|_2^2 = \min_{\mathbf{U},\mathbf{V}} \quad \sum_{i=1}^{N} \left\| \left(\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U}\right)\Lambda_i^{1/2} \right\|_F^2, \tag{6}$$

where $\mathbf{x}_i$ denotes the ith column of $\mathbf{X}$, $\mathbf{u}_k$ is the kth column of $\mathbf{U}$, v_{ki} is the coordinate of $\mathbf{x}_i$ with respect to $\mathbf{u}_k$, and $\Lambda_i = \mathrm{diag}\,(\mathbf{v}_i) \in \mathbb{R}^{K \times K}$, $\mathrm{diag}\,(\mathbf{v}_i)$ indicates a conversion of the vector $\mathbf{v}_i$ into a diagonal matrix in which the kth diagonal element is v_{ki}.

The local coordinate constraint employs a square loss. When the dataset is corrupted by outlier points and noises, the local coordinate constraint may fail to achieve sparsity and locality simultaneously. In order to alleviate the side effect of noisy data, our robust local coordinate constraint can be formulated as

$$\min_{\mathbf{U},\mathbf{V}} \quad \sum_{i=1}^{N} \left\| \left(\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U}\right)\Lambda_i^{1/2} \right\|_{2,1}, \tag{7}$$

where the Frobenius norm-based square loss function has been substituted by the $l_{2,1}$ norm.

3.3. Local Spline Regression. In this subsection, we briefly introduce local spline regression [42].

Given N data points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\}$ sampled from the underlying submanifold M, we use set $\mathcal{N}(\mathbf{x}_i) = \{x_{i_j}\}_{j=1}^{k}$ to denote $\mathbf{x}_i$ and its $k-1$ nearest neighbor points, where $i_j \in \{1, 2, \ldots, N\}$, and $\mathbf{Y}_i = [y_{i1}, y_{i2}, \ldots, y_{ik}]^{\mathrm{T}}$ is the local predicted label matrix for the ith region. The task of local spline regression is to seek the predicting function $g_i : \mathbb{R}^M \to \mathbb{R}$ in order to map each data point $\mathbf{x}_{i_j} \in \mathbb{R}^M$ to the local predicted class label $y_{i_j} = g_i(x_{ij})$. The model of local spline regression can be expressed as

$$\min_{g_i} \quad \sum_{j=1}^{k} \left(y_{i_j} - g_i\left(x_{i_j}\right)\right)^2 + \gamma \mathcal{S}(g_i), \tag{8}$$

where $\mathcal{S}(g_i)$ is a regularization term and $\gamma > 0$ is a small positive regularization parameter to control the smoothness of the spline [42]. If $\mathcal{S}(g_i)$ is defined as a seminorm of a Sobolev space, g_i can be solved by the following objective function [43]:

$$g_i(\mathbf{x}) = \sum_{j=1}^{d} \beta_{i,j} p_j(\mathbf{x}) + \sum_{j=1}^{k} \alpha_{i,j} G_{i,j}(x), \tag{9}$$

where $d = C_{M+s-1}^{s}$, in which s is the order of the partial derivatives [43]. $\{p_j(x)\}_{j=1}^{d}$ and $G_{i,j}$ are a set of primitive polynomials and a Green's function, respectively. The coefficients α_i and β_i can be achieved by solving the following problem:

$$\begin{pmatrix} \mathbf{K}_i & \mathbf{P}_i^{\mathrm{T}} \\ \mathbf{P}_i & 0 \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \mathbf{Y}_i \\ 0 \end{pmatrix}, \tag{10}$$

where $\mathbf{K}_i$ is a symmetrical matrix with elements $K_{r,c} = G_{r,c}(x_{i_r} - x_{i_c})$, and $\mathbf{P}_i$ is a matrix with its elements $P_{i,j} = p_i(x_{i_j})$. The local spline regression model can then be expressed as [42]

$$\min_{\mathbf{Y}_i} \quad \mathbf{Y}_i^{\mathrm{T}} \mathbf{M}_i \mathbf{Y}_i, \tag{11}$$

where $\mathbf{M}_i$ is the upper left $k \times k$ submatrix of the inverse matrix of the coefficient matrix in (10). Because the local predicted label matrix $\mathbf{Y}_i$ is a part of the global predicted label matrix $\mathbf{Y}$, we can construct a selection matrix $\mathbf{S}_i \in \mathbb{R}^{k \times N}$ for each $\mathbf{Y}_i$ such that

$$\mathbf{Y}_i = \mathbf{S}_i \mathbf{Y}, \tag{12}$$

where the selection matrix $\mathbf{S}_i$ is defined as follows:

$$S_i(r, c) = \begin{cases} 1, & \text{if } r = i_c, \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

After the local predicted label matrices are established, we combine them by minimizing the following loss function:

$$\min_{\mathbf{Y}_i} \quad \sum_{i=1}^{N} \mathbf{Y}_i \mathbf{M}_i \mathbf{Y}_i = \sum_{i=1}^{N} \mathbf{Y}^{\mathrm{T}} \mathbf{S}_i^{\mathrm{T}} \mathbf{M}_i \mathbf{S}_i \mathbf{Y} = \mathbf{Y}^{\mathrm{T}} \mathbf{M} \mathbf{Y}, \tag{14}$$

where

$$\mathbf{M} = \sum_{i=1}^{N} \mathbf{S}_i^{\mathrm{T}} \mathbf{M}_i \mathbf{S}_i. \tag{15}$$

Based on the studies of [33, 34], the predicted label matrix $\mathbf{Y}$ of the labeled data points should be consistent with the ground truth labels matrix $\mathbf{F}$. With the consistence constraints, the objective function (14) can be written as follows:

$$\min_{\mathbf{Y}} \quad \mathrm{Tr}\left(\mathbf{Y}^{\mathrm{T}} \mathbf{M} \mathbf{Y}\right) + \eta \mathrm{Tr}\left(\left(\mathbf{Y} - \mathbf{F}\right)^{\mathrm{T}} \mathbf{E} (\mathbf{Y} - \mathbf{F})\right), \tag{16}$$

where $\mathbf{E}$ is a diagonal matrix whose diagonal elements are 1 for labeled data and 0 for unlabeled data, and the elements of $\mathbf{F}$ are defined as follows:

$$f_{ij} = \begin{cases} 1, & \text{if } x_i \text{ is labeled as class } j, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

When η is sufficiently large, the optimal solution $\mathbf{Y}$ to the problem (16) makes the second term approximately equal to zero. Thus, the objective function (16) guarantees local and global structural consistency over labels. All the elements of $\mathbf{Y}$ are restricted to be nonnegative.

3.4. Objective Function of RSNLCF. By combining the RNMF (5), robust local coordinate constraint (7), and semisupervised local spline regression (16) into a unified

framework, we can formulate the objective function as follows:

$$\min_{\mathbf{U},\mathbf{V}} \quad \|\mathbf{X} - \mathbf{UV}\|_{2,1} + \mu \sum_{i=1}^{N} \|(\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U})\Lambda_i^{1/2}\|_{2,1} + \lambda \|\mathbf{V}\|_{2,1}$$
$$+ \tau \mathrm{Tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{MY} + \eta(\mathbf{Y} - \mathbf{F})^{\mathrm{T}}\mathbf{E}(\mathbf{Y} - \mathbf{F})\right), \tag{18}$$

where τ and μ are two trade-off parameters. We call (18) our proposed RSNLCF.

4. Optimization

The objective function (18) involves the $l_{2,1}$ norm, which is nonsmooth and cannot have a closed form solution. Consequently, we propose to solve it as follows.

Denote $\mathbf{X} - \mathbf{UV} = [\mathbf{a}^1, \ldots, \mathbf{a}^M]^{\mathrm{T}}, (\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U})\Lambda_i^{1/2} = [\mathbf{b}^1, \ldots, \mathbf{b}^K]^{\mathrm{T}}$ and $\mathbf{V} = [\mathbf{v}^1, \ldots, \mathbf{v}^M]^{\mathrm{T}}$. When considering the nonnegative constraint on $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{Y}$, the objective function (18) could be reformulated as

$$\mathcal{O} = \mathrm{Tr}\left((\mathbf{X} - \mathbf{UV})^{\mathrm{T}}\mathbf{A}(\mathbf{X} - \mathbf{UV})\right)$$
$$+ \mu \mathrm{Tr}\left(\sum_{i=1}^{N}\left((\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U})\Lambda_i^{1/2}\right)\mathbf{B}\left((\mathbf{x}_i \mathbf{1}^{\mathrm{T}} - \mathbf{U})\Lambda_i^{1/2}\right)\right)^{\mathrm{T}}$$
$$+ \lambda \mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{CV}\right) + \tau \mathrm{Tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{MY} + \eta(\mathbf{Y} - \mathbf{F})^{\mathrm{T}}\mathbf{E}(\mathbf{Y} - \mathbf{F})\right),$$

$$\text{s.t.} \quad U \in \mathbb{R}^{M \times K} > 0, V \in \mathbb{R}^{K \times N} > 0, F \in \mathbb{R}^{N \times Lc} > 0, \tag{19}$$

where $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are three diagonal matrices with their diagonal elements given as $\mathbf{A}_{ii} = 1/2\|\mathbf{a}^i\|_2$, $\mathbf{B}_{ii} = 1/2\|\mathbf{b}^i\|_2$, and $\mathbf{C}_{ii} = 1/2\|\mathbf{v}^i\|_2$, respectively.

4.1. Update Rules. The objective function $\mathcal{O}$ of RSNLCF in (19) is not convex in $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{Y}$ together. Therefore, it is unrealistic to expect an algorithm to find the global minima. In this subsection, we describe our development of an iterative algorithm based on the Lagrangian multiplier method, which can achieve local minima. Following some algebraic steps, the objective function can be written as follows:

$$\mathcal{O} = \mathrm{Tr}\left(\mathbf{XX}^{\mathrm{T}}\mathbf{A} + \mathbf{UVV}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A} - 2\mathbf{XV}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A}\right.$$
$$\left. + \mu \sum_{i=1}^{N}\left(\mathbf{x}_i \mathbf{1}^{\mathrm{T}}\Lambda_i \mathbf{1}\mathbf{x}_i^{\mathrm{T}}\mathbf{B} - 2\mathbf{x}_i \mathbf{1}^{\mathrm{T}}\Lambda_i \mathbf{U}^{\mathrm{T}}\mathbf{B} + \mathbf{U}\Lambda_i \mathbf{U}^{\mathrm{T}}\mathbf{B}\right)\right)$$
$$+ \lambda \mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{CV}\right) + \tau \mathrm{Tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{MY} + \eta(\mathbf{Y} - \mathbf{F})^{\mathrm{T}}\mathbf{E}(\mathbf{Y} - \mathbf{F})\right). \tag{20}$$

To tackle the nonnegative constraint on $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{Y}$, the objective (20) can be rewritten as the Lagrangian multiplier.

$$\mathcal{L} = \mathrm{Tr}\left(\mathbf{XX}^{\mathrm{T}}\mathbf{A} + \mathbf{UVV}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A} - 2\mathbf{XV}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A}\right.$$
$$\left. + \mu \sum_{i=1}^{N}\left(\mathbf{x}_i \mathbf{1}^{\mathrm{T}}\Lambda_i \mathbf{1}\mathbf{x}_i^{\mathrm{T}}\mathbf{B} - 2\mathbf{x}_i \mathbf{1}^{\mathrm{T}}\Lambda_i \mathbf{U}^{\mathrm{T}}\mathbf{B} + \mathbf{U}\Lambda_i \mathbf{U}^{\mathrm{T}}\mathbf{B}\right)\right) \tag{21}$$
$$+ \lambda \mathrm{Tr}\left(\mathbf{V}^{\mathrm{T}}\mathbf{CV}\right) + \tau \mathrm{Tr}\left(\mathbf{Y}^{\mathrm{T}}\mathbf{MY} + \eta(\mathbf{Y} - \mathbf{F})^{\mathrm{T}}\mathbf{E}(\mathbf{Y} - \mathbf{F})\right)$$
$$- \mathrm{Tr}\left(\mathbf{\Psi}\mathbf{U}^{\mathrm{T}}\right) - \mathrm{Tr}\left(\mathbf{\Phi}\mathbf{V}^{\mathrm{T}}\right) - \mathrm{Tr}\left(\mathbf{\Theta}\mathbf{Y}^{\mathrm{T}}\right),$$

where $\mathbf{\Psi} = [\psi_{jk}]$, $\mathbf{\Phi} = [\phi_{ki}]$, and $\mathbf{\Theta} = [\theta_{is}]$ are the Lagrangian multipliers. Let the partial derivatives of the objective function (21) with respect to $\mathbf{U}$, $\mathbf{V}$, and $\mathbf{Y}$ be zero. Thus, we have

$$\mathbf{\Psi} = 2\mathbf{AUVV}^{\mathrm{T}} - 2\mathbf{AXV}^{\mathrm{T}} - 2\mu\mathbf{BXV}^{\mathrm{T}} + 2\mu\mathbf{BUH},$$
$$\mathbf{\Phi} = 2\mathbf{U}^{\mathrm{T}}\mathbf{AUV} - 2\mathbf{U}^{\mathrm{T}}\mathbf{AX} + \mu\left(\mathbf{G} - 2\mathbf{U}^{\mathrm{T}}\mathbf{BX} + \mathbf{D}\right) + 2\lambda\mathbf{CV},$$
$$\mathbf{\Theta} = 2\tau\mathbf{MY} + 2\tau\eta\mathbf{E}(\mathbf{Y} - \mathbf{F}), \tag{22}$$

where $\mathbf{H}$ is a diagonal matrix whose entries are row sums of $\mathbf{V}$. $\mathbf{G} = (\mathbf{g}, \ldots, \mathbf{g})^{\mathrm{T}}$ is a $K \times N$ matrix whose columns are $\mathbf{g} = \mathrm{diag}\left(\mathbf{X}^{\mathrm{T}}\mathbf{BX}\right) \in \mathbb{R}^N$. $\mathbf{D} = (\mathbf{d}, \ldots, \mathbf{d})$ is a $K \times N$ matrix, and $\mathbf{d} = \mathrm{diag}\left(\mathbf{U}^{\mathrm{T}}\mathbf{BU}\right) \in \mathbb{R}^K$.

Based on the Karush-Kuhn-Tucker conditions [44] $\psi_{jk} u_{jk} = 0$, $\phi_{ki} v_{ki} = 0$ and $\theta_{is} y_{is} = 0$, we obtain

$$\left(\mathbf{AUVV}^{\mathrm{T}}\right)_{jk} u_{jk} - \left(\mathbf{AXV}^{\mathrm{T}}\right)_{jk} u_{jk} - \mu\left(\mathbf{BXV}^{\mathrm{T}}\right)_{jk} u_{jk}$$
$$+ \mu\left(\mathbf{BUH}\right)_{jk} u_{jk} = 0,$$
$$2\left(\mathbf{U}^{\mathrm{T}}\mathbf{AUV}\right)_{ki} v_{ki} - 2\left(\mathbf{U}^{\mathrm{T}}\mathbf{AX}\right)_{ki} v_{ki} + 2\lambda\left(\mathbf{CV}\right)_{ki} v_{ki} \tag{23}$$
$$+ \mu\left(\mathbf{G} - 2\mathbf{U}^{\mathrm{T}}\mathbf{BX} + \mathbf{D}\right)_{ki} v_{ki} = 0,$$
$$\left(\mathbf{MY}\right)_{is} y_{is} + \eta\left(\mathbf{E}(\mathbf{Y} - \mathbf{F})\right)_{is} y_{is} = 0.$$

The corresponding equivalent formulas are as follows:

$$\left(\mathbf{AUVV}^{\mathrm{T}}\right)_{jk} u_{jk}^2 - \left(\mathbf{AXV}^{\mathrm{T}}\right)_{jk} u_{jk}^2 - u\left(\mathbf{BXV}^{\mathrm{T}}\right)_{jk} u_{jk}^2$$
$$+ \mu\left(\mathbf{BUH}\right)_{jk} u_{jk}^2 = 0, \tag{24}$$

$$2\left(\mathbf{U}^{\mathrm{T}}\mathbf{AUV}\right)_{ki} v_{ki}^2 - 2\left(\mathbf{U}^{\mathrm{T}}\mathbf{AX}\right)_{ki} v_{ki}^2 + 2\lambda\left(\mathbf{CV}\right)_{ki} v_{ki}^2$$
$$+ \mu\left(\mathbf{G} - 2\mathbf{U}^{\mathrm{T}}\mathbf{BX} + \mathbf{D}\right)_{ki} v_{ki}^2 = 0. \tag{25}$$

$$\left(\mathbf{MY}\right)_{is} y_{is}^2 + \eta\left(\mathbf{E}(\mathbf{Y} - \mathbf{F})\right)_{is} y_{is}^2 = 0. \tag{26}$$

Solving (24), (25), and (26), we obtain the following update rules, given by

$$u_{jk}^{(t=1)} \leftarrow u_{jk}^{(t)} \sqrt{\frac{\left(\mathbf{AXV}^{\mathrm{T}} + \mu\mathbf{BXV}^{\mathrm{T}}\right)_{jk}}{\left(\mathbf{AUVV}^{\mathrm{T}} + \mu\mathbf{BUV}\right)_{jk}}}, \tag{27}$$

$$v_{ki}^{(t=1)} \leftarrow v_{ki}^{(t)} \sqrt{\frac{2\left(\mathbf{U}^{\mathrm{T}}\mathbf{AX} + \mu\mathbf{U}^{\mathrm{T}}\mathbf{BX}\right)_{ki}}{\left(2\mathbf{U}^{\mathrm{T}}\mathbf{AUV} + \mu\mathbf{G} + \mu\mathbf{D} + 2\lambda\mathbf{CV}\right)_{ki}}}, \tag{28}$$

$$y_{is}^{(t+1)} \leftarrow y_{is}^{(t)} \sqrt{\frac{\eta\left(\mathbf{EF}\right)_{is}}{\left(\mathbf{MY} + \eta\mathbf{EY}\right)_{is}}}. \tag{29}$$

In this manner, we obtain the solver for the objective function (19).

4.2. Convergence Analysis. In this subsection, we demonstrate that the objective function (20) converges to a local optimum by using the update rules (27), (28), and (29) after finite iterations. We adopt the auxiliary function approach [16] to prove the convergence. Here, we first introduce the definition of an auxiliary function.

Definition 1. $Z(q, q')$ is an auxiliary function for $F(q)$ if the following properties are satisfied:

$$Z(q, q') \geq F(q), Z(q, q) = F(q). \tag{30}$$

Lemma 1. *If Z is an auxiliary function for F, then F is nonincreasing under the update:*

$$h^{(t+1)} = \underset{q}{\operatorname{argmin}} Z(q, q^{(t)}). \tag{31}$$

Proof 1.

$$F(q^{(t+1)}) \leq Z(q^{(t+1)}, q^{(t)}) \leq Z(q^{(t)}, q^{(t)}) = F(q^{(t)}). \tag{32}$$

Lemma 2. *For any nonnegative matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, $\mathbf{B} \in \mathbb{R}^{k \times k}$, $\mathbf{S} \in \mathbb{R}^{n \times k}$, $\mathbf{S}' \in \mathbb{R}^{n \times k}$, and $\mathbf{A}, \mathbf{B}$ are symmetric, and then the following inequality holds*

$$\sum_{i=l}^{n} \sum_{p=1}^{k} \frac{(\mathbf{A}\mathbf{S}'\mathbf{B})_{ip} \mathbf{S}_{ip}^2}{\mathbf{S}'_{ip}} \geq \operatorname{Tr}(\mathbf{S}^{\mathrm{T}}\mathbf{A}\mathbf{S}\mathbf{B}). \tag{33}$$

The convergence of the algorithms is demonstrated in the following:

For given $\mathbf{X}$, the optimizing objective function (20) w.r.t. $\mathbf{V}$ is equivalent to minimizing

$$\mathcal{O}(\mathbf{V}) = \operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{V}) - 2\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{X}\mathbf{V}^{\mathrm{T}}) + \mu\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{G}) \\ - 2\mu\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{X}\mathbf{V}^{\mathrm{T}}) + \mu\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{D}) + \lambda\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{C}\mathbf{V}). \tag{34}$$

Theorem 1. *The following function*

$$Z(\mathbf{V}, \mathbf{V}') = \sum_{ki} \frac{(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{V}')_{ki} \mathbf{V}_{ki}^2}{\mathbf{V}'_{ki}} - 2\sum_{ki} (\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{X})_{ki} \mathbf{V}'_{ki} \\ \cdot \left(1 + \log \frac{\mathbf{V}_{ki}}{\mathbf{V}_{ki}}\right) + \mu\sum_{ki} \mathbf{G}_{ki} \frac{\mathbf{V}_{ki}^2 + (\mathbf{V}')_{ki}^2}{2\mathbf{V}'_{ki}} \\ - 2\mu\sum_{ki} (\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{X})_{ki} \mathbf{V}'_{ki}\left(1 + \log \frac{\mathbf{V}_{ki}}{\mathbf{V}'_{ki}}\right) \\ + \mu\sum_{ki} \mathbf{D}_{ki} \frac{\mathbf{V}_{ki}^2 + (\mathbf{V}')_{ki}^2}{2\mathbf{V}'_{ki}} + \lambda\sum_{ki} \frac{(\mathbf{C}\mathbf{V}')_{ki} \mathbf{V}_{ki}^2}{\mathbf{V}'_{ki}} \tag{35}$$

is an auxiliary function for $\mathcal{O}(\mathbf{V})$.

Proof 1. In one sense, $Z(\mathbf{V}, \mathbf{V}) = \mathcal{O}(\mathbf{V})$ is obvious. However, we need to prove that $Z(\mathbf{V}, \mathbf{V}') \geq \mathcal{O}(\mathbf{V})$. To accomplish this, we compare (34) and (35) to find out that $Z(\mathbf{V}, \mathbf{V}') \geq \mathcal{O}(\mathbf{V})$.

By applying Lemma 2, we obtain

$$\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{V}\mathbf{V}^{\mathrm{T}}) \leq \sum_{ki} \frac{(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{V}')_{ki} \mathbf{V}_{ki}^2}{\mathbf{V}'_{ki}}, \\ \lambda\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{C}\mathbf{V}) \leq \lambda\sum_{ki} \frac{(\mathbf{C}\mathbf{V}')_{ki} \mathbf{V}_{ki}^2}{\mathbf{V}'_{ki}}. \tag{36}$$

To obtain the upper bound for the third and fifth terms, we use the inequality $a^2 + b^2 \geq 2ab$, which holds for any a, $b \geq 0$, and these third and fifth terms in $\mathcal{O}(\mathbf{V})$ are bounded by

$$\mu\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{G}) \leq \mu\sum_{ki} \mathbf{G}_{ki} \frac{\mathbf{V}_{ki}^2 + (\mathbf{V}')_{ki}^2}{2\mathbf{V}'_{ki}}, \\ \mu\operatorname{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{D}) \leq \mu\sum_{ki} \mathbf{D}_{ki} \frac{\mathbf{V}_{ki}^2 + (\mathbf{V}')_{ki}^2}{2\mathbf{V}'_{ki}}. \tag{37}$$

To obtain lower bounds for the remaining terms, we adopt the inequality $z \geq 1 + \log z$, $\forall z$, and then

$$2\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{X}\mathbf{V}^{\mathrm{T}}) \geq 2\sum_{ki} (\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{X})_{ki} \mathbf{V}'_{ki}\left(1 + \log \frac{\mathbf{V}_{ki}}{\mathbf{V}_{ki}}\right), \\ 2\mu\operatorname{Tr}(\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{X}\mathbf{V}^{\mathrm{T}}) \geq 2\mu\sum_{ki} (\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{X})_{ki} \mathbf{V}'_{ki}\left(1 + \log \frac{\mathbf{V}_{ki}}{\mathbf{V}'_{ki}}\right). \tag{38}$$

Summing all inequalities, we can obtain $Z(\mathbf{V}, \mathbf{V}') \geq \mathcal{O}(\mathbf{V})$ which obviously satisfies $Z(\mathbf{V}, \mathbf{V}') \geq \mathcal{O}(\mathbf{V})$. Therefore, $Z(\mathbf{V}, \mathbf{V}')$ is an auxiliary function of $\mathcal{O}(\mathbf{V})$.

Theorem 2. *The updating rule (28) can be obtained by minimizing the auxiliary function $Z(\mathbf{V}, \mathbf{V}')$.*

Proof 1. To find the minimum of $Z(\mathbf{V}, \mathbf{V}')$, we set the derivative $\partial Z(\mathbf{V}, \mathbf{V}')/\partial \mathbf{V}_{ki} = 0$ and obtain

$$\frac{\partial Z(\mathbf{V}, \mathbf{V}')}{\partial \mathbf{V}_{ki}} = \frac{2(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{U}\mathbf{V}')_{ki} \mathbf{V}_{ki}}{\mathbf{V}'_{ki}} - \frac{2(\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{X})_{ki} \mathbf{V}'_{ki}}{\mathbf{V}_{ki}} \\ + \mu\frac{\mathbf{G}_{ki}\mathbf{V}_{ki}}{\mathbf{V}'_{ki}} - \mu\frac{2(\mathbf{U}^{\mathrm{T}}\mathbf{B}\mathbf{X})_{ki} \mathbf{V}'_{ki}}{\mathbf{V}_{ki}} \\ + \mu\frac{\mathbf{D}_{ki}\mathbf{V}_{ki}}{\mathbf{V}'_{ki}} + \lambda\frac{2(\mathbf{C}\mathbf{V}')_{ki} \mathbf{V}_{ki}}{\mathbf{V}'_{ki}}. \tag{39}$$

Thus, by simple algebraic formulation, we can obtain the iterative updating rule for $\mathbf{V}$ as (28).

Based on the properties of the auxiliary, we prove that the objective function (20) monotonically decreases under the updating v_{ki}.

The converge proofs showing that updating u_{jk} and y_{is} can be accomplished using (27) and (29) are similar to the aforementioned.

5. Experiments and Discussion

We systematically evaluated the performance of our presented RSNLCF method and compared it to the popular clustering methods.

5.1. Datasets. Three standard face datasets and the gene dataset were selected to evaluate different methods. The four datasets are described as follows:

(i) *Extended YaleB dataset*: the extended YaleB dataset contains 2414 frontal face images of 38 individuals. In this dataset, the size of each face image is 192×168 and each image was acquired from 64 illuminate conditions and nine individual poses. Each image was resized to 32×32 in our experiments.

(ii) *ORL face dataset*: the OR dataset contains 400 images of 40 individuals. All images were captured at different times and with different variations including lighting, face expressions (open and closed eyes, smiling, and not smiling), and specific facial details (glasses and no glasses). The original images had a size of 92×112. Each image was rescaled to 32×32.

(iii) *AR dataset*: the AR dataset contains over 4000 frontal face images of 126 individuals (70 men and 56 women) with different facial expressions, illumination conditions, and occlusions (sunglasses and scarf). All individuals participated in two photo sessions, and 26 images of each individual were captured. Each image was scaled to 32×32.

(iv) *Leukemia dataset*: the leukemia dataset contains data related to and samples of acute myelogenous leukemia (AML) and acute lymphoblastic leukemia (ALL). ALL can be further classified as T and B subtypes. This dataset consists of 5000 genes in 38 set of tumor data and contains 19 samples of B cell ALL B, eight samples of T cell ALL T, and 11 samples of AML.

5.2. Experimental Design. In this section, we describe our evaluation metrics, the compared methods, and our parameter selection.

5.2.1. Evaluation Metrics. In our experiments, two widely used metrics (i.e., accuracy (Acc) and normalized mutual information (NMI)) were adopted to evaluate the clustering results [45]. We evaluated the algorithms by comparing the cluster labels of each data point with its label provided by the dataset. The Acc metric is defined as follows:

$$\text{Acc} = \frac{\sum_{i=1}^{n} \delta(\text{map}(r_i), l_i)}{n}, \quad (40)$$

where n refers to the total number of samples, r_i denotes the cluster label of x_i, and l_i is the true class label. In addition, $\delta(x, y)$ is the delta function that is equal to 1 if $x = y$ and 0 otherwise, and $\text{map}(r_i)$ is the mapping function that maps the obtained label r_i to the equivalent label from the dataset. The best mapping function can be determined by using the Kuhn-Munkres algorithm [46]. The value of Acc is equal to 1 if and only if the clustering result and the true label are identical. The second measure is the NMI, which is adopted in order to evaluate the quality of clusters. Given a clustering result, the NMI is defined as follows:

$$\text{NMI} = \frac{\sum_{i=1}^{c} \sum_{j=1}^{c} n_{i,j} \log \left(n_{i,j}/n_i \widehat{n}_j \right)}{\sqrt{\left(\sum_{i=1}^{c} n_i \log \left(n_i/n \right) \right) \left(\sum_{j=1}^{c} \widehat{n}_j \log \left(\widehat{n}_j/n \right) \right)}}, \quad (41)$$

where n_i denotes the number of images contained in the ith cluster C_i based on clustering results, $\widehat{n}j$ is the number of images belonging to the C_j', and $n_{i,j}$ is the number of images that are in the intersection of C_i and C_j'.

5.2.2. Compared Methods. To verify the clustering performance of our RSNLCF, several popular methods were compared using the same dataset. The methods are listed as follows:

(i) RNMF using $l_{2,1}$ norm [27]

(ii) Semisupervised graph-regularized NMF (semi-GNMF) [24]

(iii) Constrained NMF (CNMF) [16]

(iv) Local centroid-structured NMF (LCSNMF) [47]

(v) Unsupervised robust seminonnegative graph embedding through the $l_{2,1}$ norm (URNGE) [28]

(vi) Nonnegative local coordinate factorization (NLCF) [9]

(vii) Our proposed RSNLCF

Sample images are shown in Figure 1.

5.2.3. Parameter Selection. Some parameters had to be tuned in the evaluated algorithms. To compare different algorithms fairly, we ran them using different parameters and chose the best average performance obtained for comparison. We set the number of clusters to be the same as the true number of categories on three image datasets and the leukemia dataset. Note that there was no parameter selection for RNMF and CNMF when the number of clusters was given. The regularization parameters were searched over the grid {0.001, 0.01, 0.1, 1, 10, 100, 1000} for semi-GNMF, URNGE, NLCF, and RSNLCF. The neighborhood size k to build the graph was chosen from $\{1, 2, \ldots, 10\}$, and the 0-1 weighting scheme was adopted for its simplicity in the graph-based methods of semi-GNMF and URNGE. We applied the approach
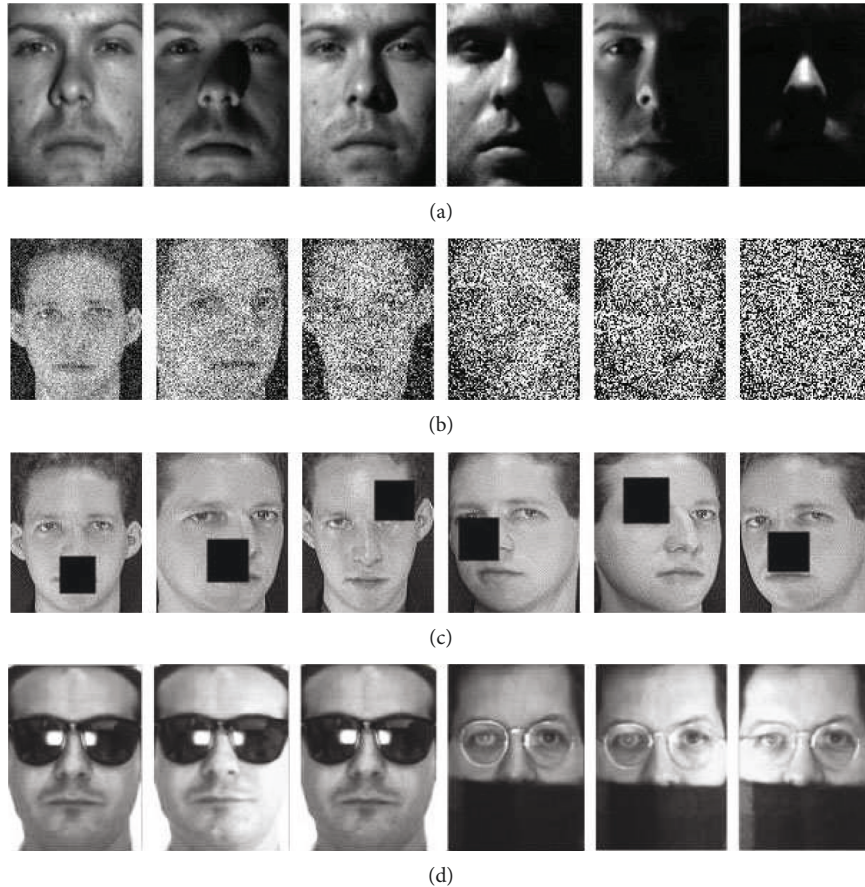
Figure 1: Sample images. (a) Extended YaleB dataset, (b) ORL dataset with random pixel corruption, (c) ORL dataset with random block occlusions, and (d) AR dataset with contiguous occlusions by sunglasses and scarves.

presented in literature [16] to adjust automatically the value of λ for LCSNMF.

5.3. Face Clustering under Illumination Variations.

The robustness of the approaches to illumination changes was tested widely with the extended YaleB dataset. Figure 1(a) shows some samples from this dataset. We used only the frontal face images of the first 18 individuals. Our experiments were performed with various numbers of clusters. For the fixed cluster number k, the images of k categories from the extended YaleB dataset were randomly selected and mixed for evaluation. For semisupervised methods semi-GNMG, CNMF, and URNGE, eight face images per individual were randomly chosen as labeled samples; the rest of the dataset was used as unlabeled samples. On the clustering set, the compared methods were used to achieve new data representations. For a fair comparison, we used k-means to cluster samples based on the new data representations. The results of k-means are related to initialization. We repeated the experiments 20 times with different initialization parameters. The clustering results were measured by the commonly used evaluation metrics, Acc and NMI. Table 1 shows the detailed clustering results on different clustering numbers. The final row shows the average clustering accuracy (NMI) over k. Compared with the second best method, our method (RSNLCF) achieves an 11.41% improvement in clustering

accuracy. For mutual information, it achieved a 10.63% improvement over the second best algorithm.

5.4. Face Clustering under Pixel Corruptions.

Two experiments were designed to test the robustness of RSNLCF against random pixel corruptions on the ORL face dataset. For the semisupervised algorithms of semi-GNMG, CNMF, URNGE, and RSNLCF, three images per individual were randomly chosen as labeled samples, and the remaining images were used as unlabeled samples. In the first experiment, each image was corrupted by replacing the pixel value with independent and identically distributed samples whose lower and upper bounds were the minimum and maximum pixel value of the image, respectively. The corrupted pixels of each image varied from 10 to 90% in increments of 10%. Figure 1(b) shows several examples. Because the corrupted pixels were randomly selected for each test sample, we repeated the experiments 20 times. Figure 2 displays the recognition accuracies over different levels of corruption. The recognition accuracies of the methods decreased rapidly as the level of corruption increased. From Figure 2, which depicts the recognition accuracies, we can observe that the proposed method consistently outperformed the others. When the samples had a high percentage of pixel corruption, the methods failed to obtain improved recognition performance because of inadequate discriminative information.

TABLE 1: Clustering performance on the extended YaleB dataset.

k	RNMF	Semi-GNMF	CNMF	LCSNMF	URNGE	NLCF	RSNLCF
Acc (%)							
2	78.43 ± 16.23	90.47 ± 15.25	76.61 ± 15.34	92.25 ± 17.34	91.02 ± 18.83	89.54 ± 13.26	96.25 ± 15.64
4	69.52 ± 15.01	84.85 ± 14.43	65.62 ± 12.36	88.75 ± 14.76	86.33 ± 16.76	83.03 ± 12.74	94.63 ± 14.53
6	53.45 ± 5.55	82.36 ± 9.63	63.74 ± 6.75	86.14 ± 4.94	84.23 ± 5.66	77.64 ± 13.24	92.37 ± 9.85
8	52.76 ± 3.46	83.71 ± 7.81	65.42 ± 6.41	85.79 ± 6.91	85.39 ± 5.78	75.83 ± 7.33	90.18 ± 6.63
10	53.24 ± 3.47	77.05 ± 4.17	63.68 ± 5.25	78.04 ± 5.79	74.58 ± 2.43	70.42 ± 4.87	88.31 ± 4.14
12	55.11 ± 4.53	72.84 ± 3.15	63.25 ± 4.34	75.65 ± 4.36	73.34 ± 2.76	71.72 ± 7.23	87.26 ± 3.56
14	52.05 ± 3.26	71.57 ± 2.24	61.58 ± 3.86	73.91 ± 3.43	72.16 ± 2.14	68.52 ± 2.13	87.11 ± 2.34
16	51.97 ± 3.17	67.85 ± 2.45	59.91 ± 3.42	69.35 ± 3.82	68.35 ± 1.82	65.46 ± 3.14	85.67 ± 3.16
18	51.53 ± 3.32	67.58 ± 3.01	57.33 ± 2.97	71.64 ± 3.58	69.09 ± 2.23	65.83 ± 3.71	85.32 ± 3.01
Avg.	57.56	77.59	64.13	80.17	78.28	74.22	89.69
NMI (%)							
2	82.91 ± 18.13	92.51 ± 18.52	78.81 ± 16.28	94.36 ± 19.84	93.35 ± 17.46	91.71 ± 17.15	98.46 ± 14.75
4	72.63 ± 17.92	87.24 ± 16.35	71.58 ± 13.41	92.71 ± 15.14	90.78 ± 15.46	85.53 ± 15.94	96.38 ± 13.52
6	63.42 ± 5.21	85.91 ± 6.47	68.37 ± 8.86	90.25 ± 9.48	86.13 ± 6.48	82.82 ± 12.03	94.35 ± 7.51
8	62.14 ± 2.96	86.86 ± 3.72	70.54 ± 7.97	89.37 ± 8.16	88.73 ± 6.52	80.34 ± 5.02	93.82 ± 5.73
10	64.06 ± 3.01	84.25 ± 3.26	69.72 ± 6.33	87.54 ± 5.42	86.82 ± 4.36	79.16 ± 2.92	92.73 ± 6.22
12	63.41 ± 3.05	80.82 ± 2.33	68.98 ± 4.64	85.28 ± 5.37	84.94 ± 4.53	80.47 ± 3.77	90.69 ± 4.91
14	58.99 ± 2.11	79.34 ± 3.44	67.67 ± 5.78	84.91 ± 4.98	83.12 ± 3.98	77.55 ± 1.76	90.43 ± 3.01
16	57.23 ± 2.17	77.85 ± 2.97	65.98 ± 4.58	82.11 ± 4.12	81.66 ± 3.54	76.46 ± 2.44	89.27 ± 4.62
18	55.59 ± 2.03	77.51 ± 2.25	64.67 ± 3.98	83.79 ± 4.43	82.36 ± 3.28	77.59 ± 1.99	89.21 ± 4.75
Avg.	64.49	86.02	69.59	87.81	86.43	81.29	97.06

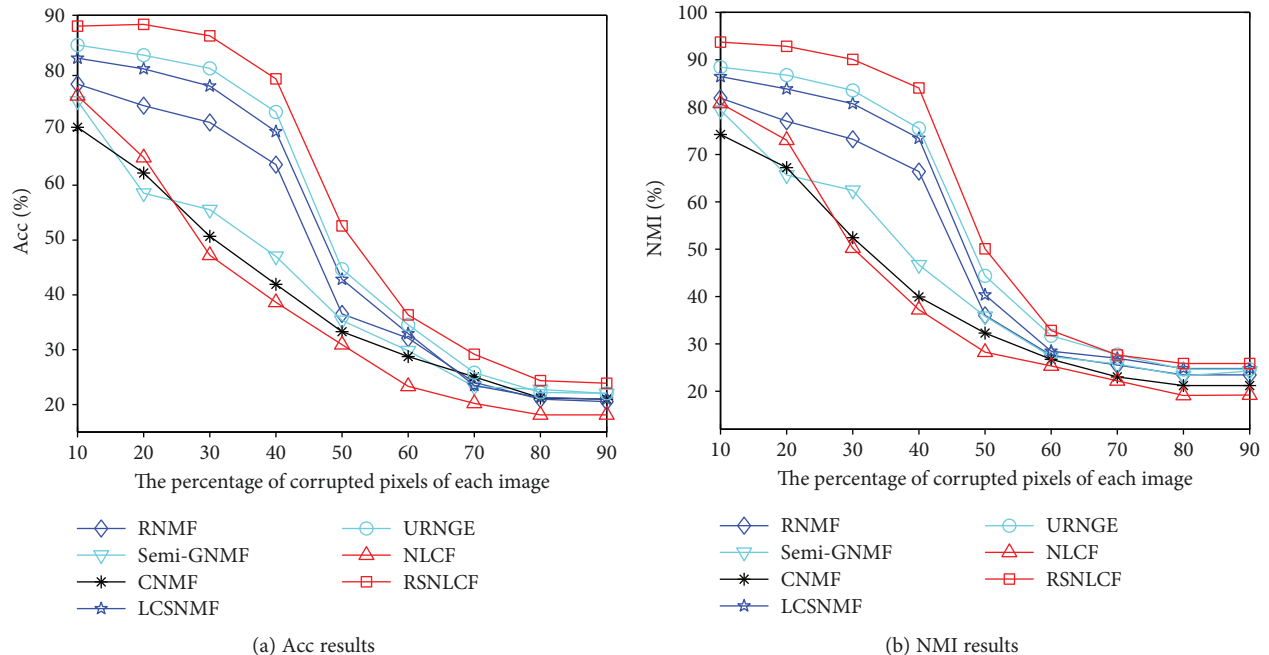(a) Acc results

(b) NMI results

FIGURE 2: Clustering Acc and NMI curves across percentages of corrupted pixels of each image for the compared methods on the ORL dataset.

In the second experiment, 40% of the pixels randomly selected from each sample were replaced by setting the pixel value as 255. The number of corrupted samples of each individual is gradually increased from 10 to 90%. We conducted the evaluations 20 times at different corruption percentages and computed the average recognition accuracies of Acc and NMI. Figure 3 illustrates clustering Acc and NMI curves of RSNLCF and the proposed method's six competitors

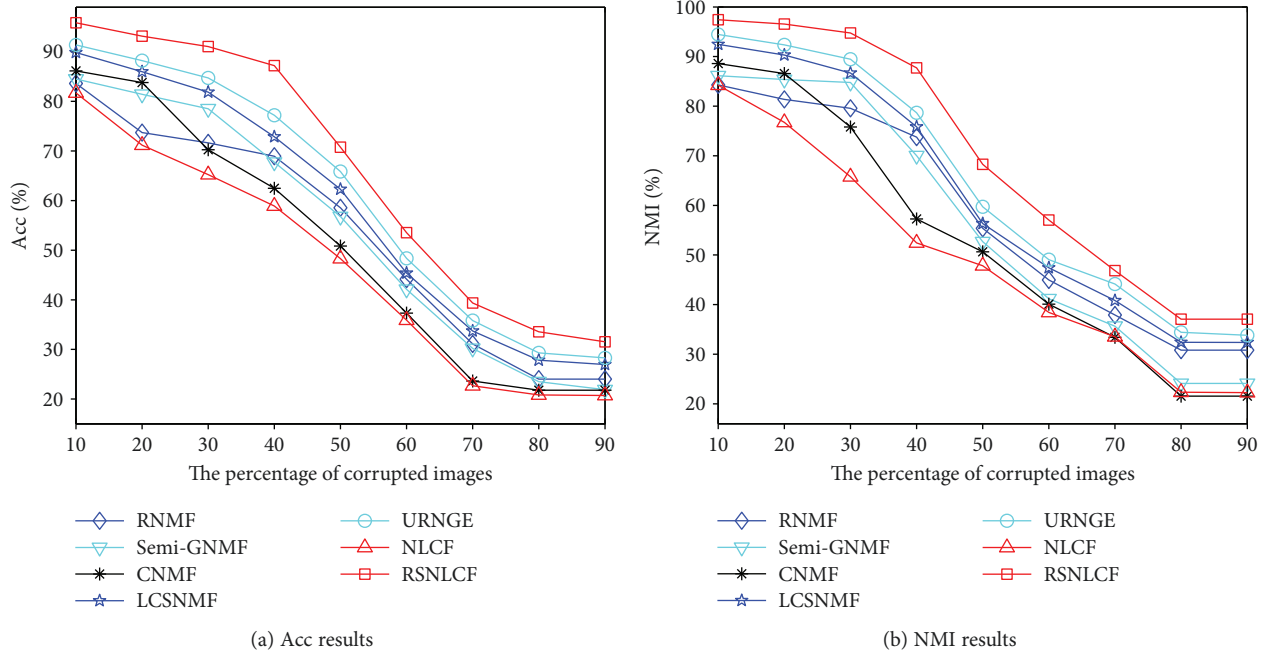(a) Acc results

(b) NMI results

FIGURE 3: Clustering Acc and NMI curves across percentages of corrupted images for the compared methods on the ORL dataset.

versus the percentage of corrupted images. From Figure 3, which depicts the comparison results on the ORL dataset, we can clearly see that the RSNLCF obtained the best recognition accuracy in all situations.

5.5. Face Clustering under Contiguous Occlusions. We validated the robustness of RSNLCF against partial block occlusions (see Figure 1(c) for examples). Two experiments were conducted on the ORL face dataset. For the semisupervised algorithms of semi-GNMG, CNMF, URNGE, and RSNLCF, we randomly selected three samples from each category and used their category number as the label information. The first experiment was performed with a fixed contiguous block occlusion size of 40×40 pixels. We chose r of the face samples of each individual for occlusion, with r varying from 10 to 90%. The position of the block was randomly selected. The evaluations were performed 20 times for each r, and the means of Acc and NMI were recorded. Figure 4 shows the means of clustering Acc and NMI of the compared methods on different percentages of corrupted images. As shown in Figure 4, the performances of NMF, RNMF, semi-GNMF, CNMF, URNGE, and NLCF were lower than that of RSNLCF. With an increasing number of occluded samples, the clustering accuracy of RSNLCF dropped and thus matched expectations considerably.

In the second experiment, we simulated various levels of contiguous occlusions in each image by using an unrelated image of size $p \times p$ with $p \in \{5, 10, 20, \dots, 80\}$. The evaluations were conducted 20 times at each occlusion level, and the average Acc and NMI curves were recorded. Figure 5 plots clustering Acc and NMI results of the compared methods under different occlusion levels. Although the clustering accuracy of each method degraded with each increment

in occlusion level, RSNLCF consistently exceeded other methods. When the occlusion size increased to 50×50, the occluding part dominated the image and caused the clustering performance to diminish rapidly.

5.6. Face Clustering under Real Occlusions. We evaluated the robustness of RSNLCF against real malicious occlusions. The AR dataset adopted in this experiment contains 2600 frontal face images from 100 individuals (50 males and 50 females from two photo sessions). Figure 1(d) shows some face samples with real occlusions by sunglasses and scarf. Note that because RNMF, LCSNMF, and NLCF are unsupervised algorithms, we did not compare them here. In this experiment, we randomly selected r face images per individual as labeled samples, in which r was varied from four to 18, respectively, in increments of two. The remaining images were unlabeled samples. For each configuration, we conducted 20 test runs with each method. The mean and the standard deviation of clustering accuracy were recorded. Table 2 tabulates the detailed clustering results by Acc and NMI on the AR dataset and shows our algorithm achieved 8.55, 12.82, and 14.53% Acc improvement over URNGE, CNMF, and semi-GNMF, respectively.

For NMI, the recognition rate of RSNLCF was 7.06, 9.66, and 10.87% higher than URNGE, CNMF, and semi-GNMF, respectively.

5.7. Gene Data Clustering on the Leukemia Dataset. Finally, we assessed clustering performance on the leukemia dataset. The gene expression dataset was rather challenging in terms of clustering issues, because it contains numerous features but only a few samples. We filtered out genes with max/min < 15 and max − min < 500, leaving a total of 1999 genes.
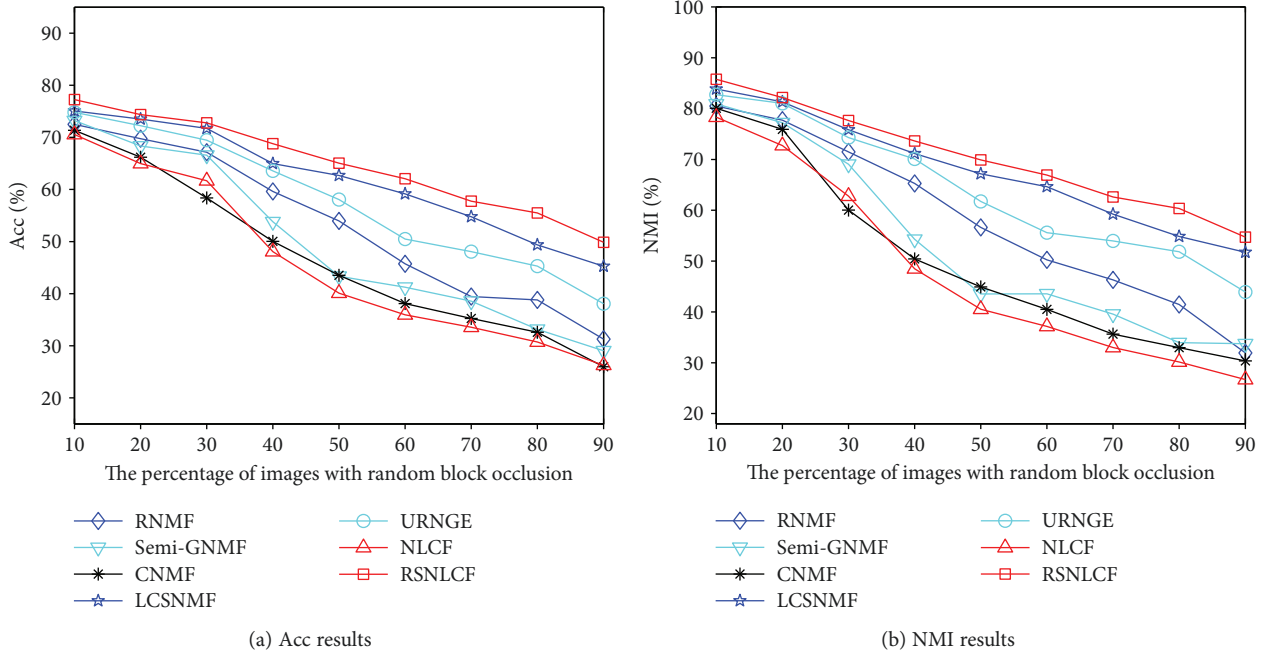
(a) Acc results

(b) NMI results

FIGURE 4: Clustering Acc and NMI curves of the compared methods on percentages of corrupted images with random block occlusions for the ORL dataset.

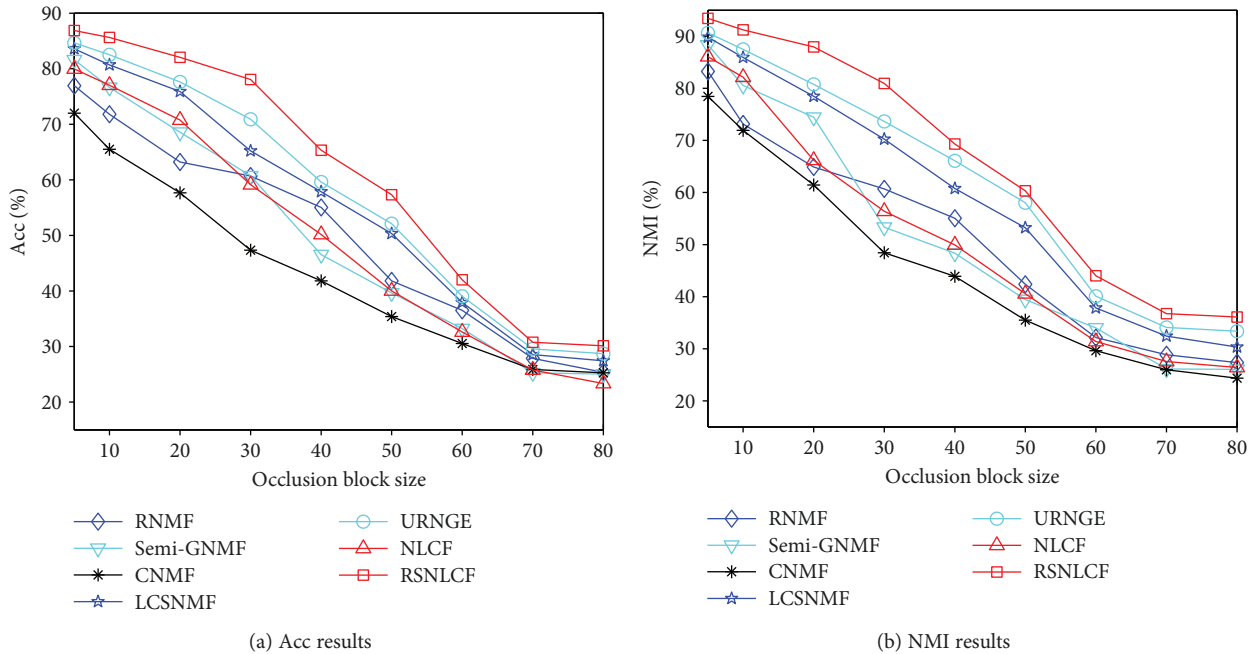(a) Acc results

(b) NMI results

FIGURE 5: Clustering Acc and NMI curves of the compared methods under different occlusion levels with each image in the ORL dataset.

Note that because RNMF, LCSNMF, and NLCF are unsupervised algorithms, we did not compare them here. For each category of data, $c = 2, 3, 4, 5, 6, 7, 8$ samples were randomly chosen and labeled, with the remaining samples being unlabeled. As the samples were randomly selected, for each c, we repeated each experiment 20 times and calculated the average clustering accuracy. Figure 6 plots clustering Acc and NMI results of the compared methods under different numbers of labeled samples. We can observe that our RSNLCF approach achieved the best clustering performance of all the compared approaches.

5.8. Parameter Sensitivity. In our proposed method, several parameters were tuned beforehand. We observed that RSNLCF is insensitive to τ in the range of $[10^{-3}, 10^3]$. Accordingly, we fixed η to be 10^6 and τ to be 10 for both the

TABLE 2: Clustering performances on the AR dataset.

r	Acc (%)				NMI (%)			
	Semi-GNMG	CNMF	URNGE	RSNLCF	Semi-GNMG	CNMF	URNGE	RSNLCF
4	58.29 ± 3.74	55.93 ± 3.72	63.28 ± 3.94	82.29 ± 2.54	68.65 ± 5.34	67.79 ± 3.51	72.37 ± 4.29	86.27 ± 2.28
6	64.53 ± 3.62	60.56 ± 4.54	69.06 ± 4.31	86.07 ± 4.19	72.49 ± 3.38	68.28 ± 4.43	76.85 ± 4.32	89.46 ± 4.93
8	67.37 ± 3.22	69.17 ± 3.93	74.19 ± 3.71	87.25 ± 2.68	78.13 ± 3.67	80.33 ± 3.25	81.24 ± 2.36	91.57 ± 2.28
10	74.87 ± 4.13	76.57 ± 2.76	79.73 ± 2.38	88.48 ± 3.85	83.48 ± 3.36	84.52 ± 2.37	87.63 ± 3.48	92.84 ± 3.58
12	79.38 ± 3.03	82.32 ± 2.83	88.01 ± 3.52	92.64 ± 2.23	85.21 ± 2.02	88.32 ± 3.76	90.27 ± 3.52	94.12 ± 2.93
14	85.95 ± 3.31	90.58 ± 2.53	91.49 ± 2.36	93.54 ± 2.17	88.27 ± 2.37	92.46 ± 2.43	92.34 ± 2.86	96.43 ± 3.54
16	86.39 ± 3.71	91.35 ± 4.54	92.92 ± 2.42	94.83 ± 2.79	89.38 ± 2.58	93.25 ± 3.38	93.94 ± 2.48	97.07 ± 2.72
18	88.83 ± 3.27	92.82 ± 4.55	94.73 ± 2.67	96.71 ± 2.35	91.03 ± 2.64	94.38 ± 3.46	95.52 ± 2.69	98.91 ± 2.76
Avg.	75.70	77.41	81.68	90.23	82.46	83.67	86.27	93.33

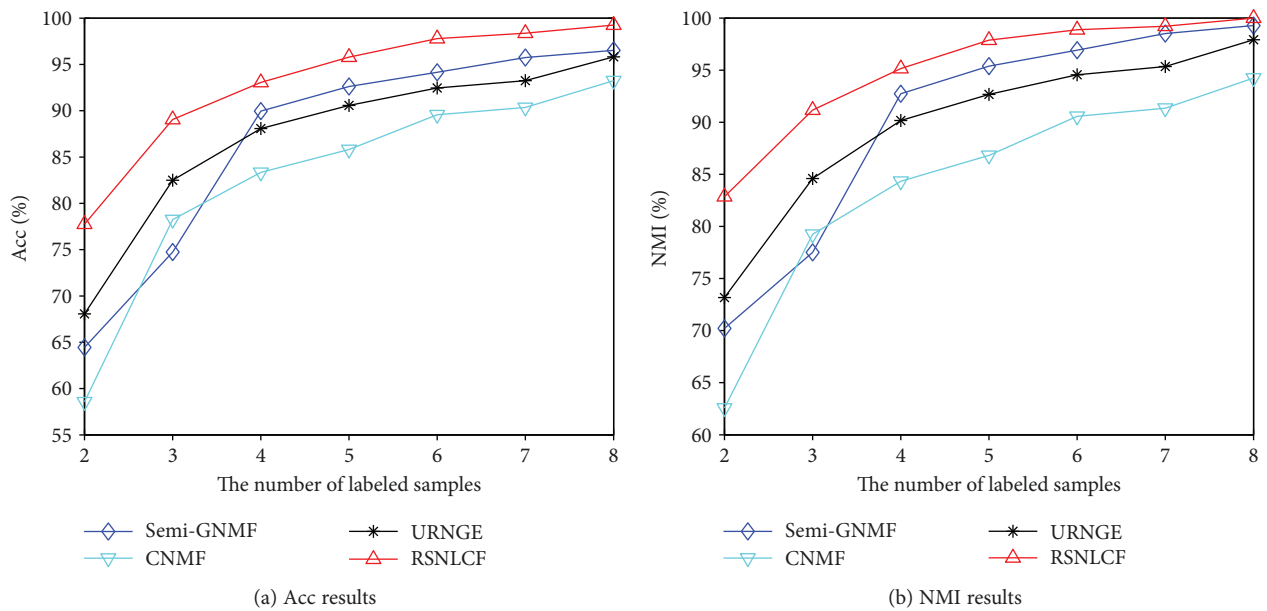

(a) Acc results   (b) NMI results

FIGURE 6: Clustering Acc and NMI curves of the compared methods under different numbers of labeled samples for the leukemia dataset.

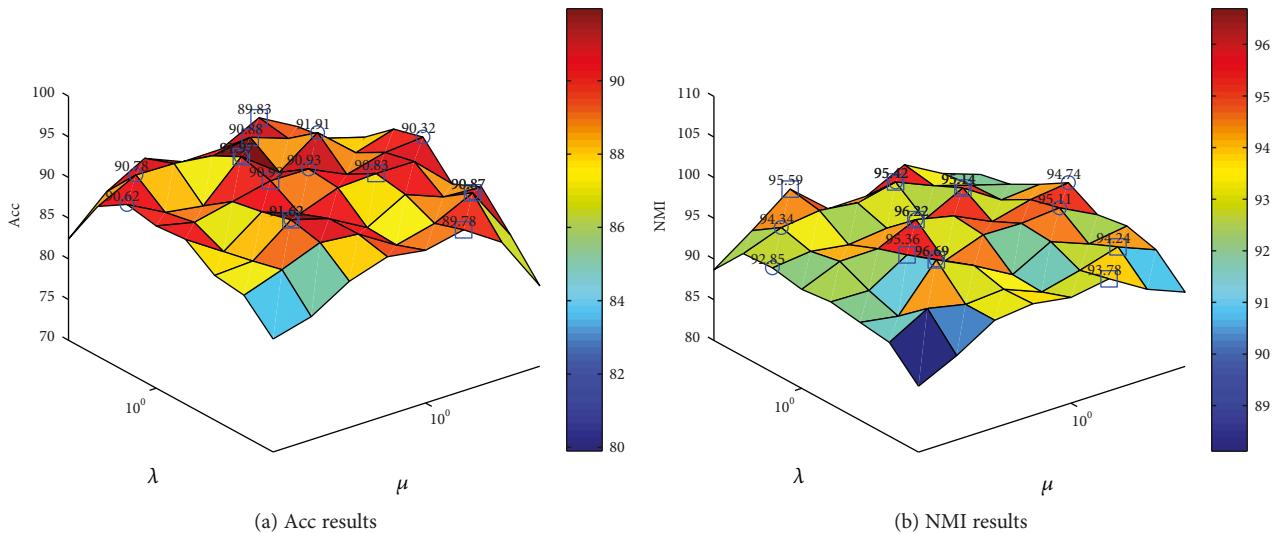(a) Acc results   (b) NMI results

FIGURE 7: Clustering accuracy of the proposed method with respect to the parameters μ and λ on the extended YaleB dataset.
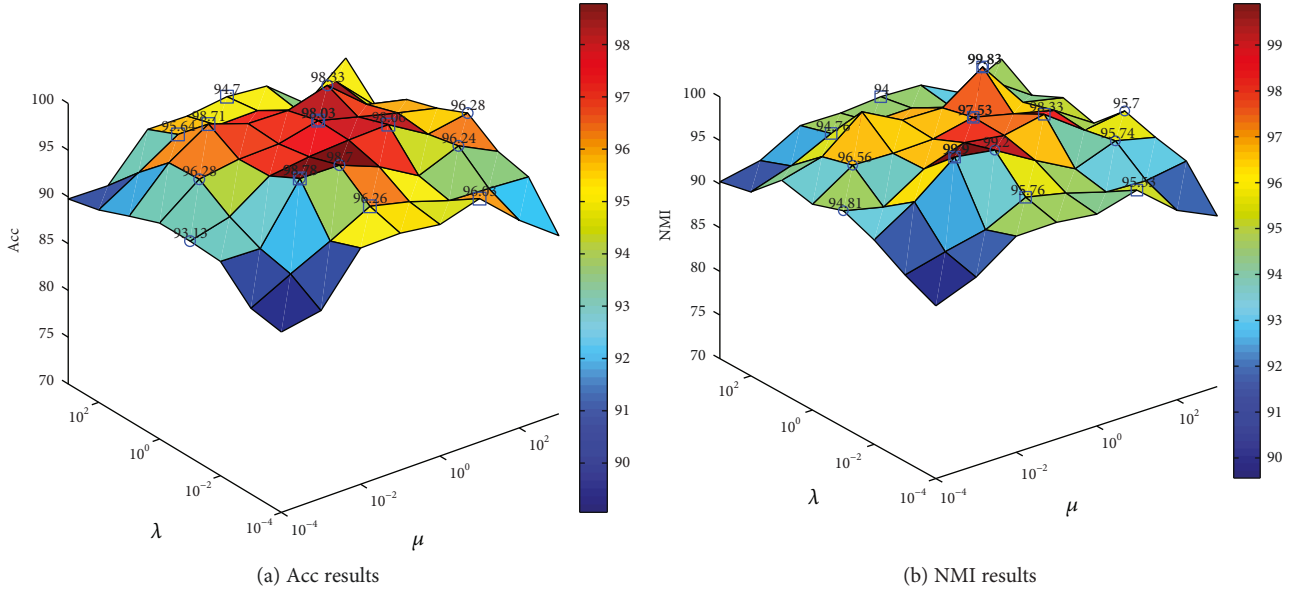
(a) Acc results

(b) NMI results

FIGURE 8: Clustering accuracy of the proposed method with respect to the parameters μ and λ on the leukemia dataset.

extended YaleB and leukemia datasets. To study the sensitivity of RSNLCF with respect to the remaining parameters (i.e., μ and λ), we varied these parameters. In the experiment, we plotted the Acc and NMI of RSNLCF with respect to μ and λ. Figures 7 and 8 show clearly the 3D results of RSNLCF. The horizontal axes are the parameters μ and λ, and the vertical axis represents the clustering accuracy of RSNLCF. In the 3D graphs, the square/circle marker indicates the best μ /λ for varying μ/λ. Next to each marker at the cross point is a digit number representing the value of Acc or NMI. We can notice from Figures 7 and 8 that the clustering performance varied with different combinations of μ and λ. However, it is unknown theoretically how to choose the best parameter. The regularization parameters should be associated with the characteristics of the dataset.

5.9. Convergence Analysis. In the previous section, we proved the convergence of our presented method. In our study, an experiment was performed to compare all algorithms' speed of convergence on the extended YaleB and leukemia datasets. The two parameters μ and λ were both fixed at 10. The time is measured using a computer with Intel Core™ I7 2600 and 16 GB memory. Figure 9 demonstrated the objective function value versus computational time for different algorithms. The horizontal and vertical axes here represent training times and the value of the objective function, respectively. We can observe from Figure 9 that the objective function value of all algorithms decreases steadily with the time increase, and RSNLCF requires less time than other graph-based methods, demonstrating that the proposed method was effective and efficient.

5.10. Overall Observations and Discussion. In our experiments, we considered several groups of experiments based on different databases, where the extended YaleB mainly involved illumination changes, the ORL database focused on pixel corruptions and block occlusions, the AR database included face images with different facial variations, sunglasses, and scarf occlusions, and the leukemia dataset contained a large number of features but only a few samples. From the aforementioned experimental results, we gained the following several attractive insights:

(i) In most cases, the performance of CNMF was usually lower than that of the graph-based approach, which demonstrates the superiority of intrinsic geometrical structure representation in discovering potential discriminative information.

(ii) Regardless of the datasets, our RSNLCF algorithm outperformed all six other methods. The reason lies in the fact that RSNLCF is designed for simultaneous application to local and global consistencies over labels simultaneously to uncover an underlying subspace structure. In addition, RSNLCF proved robust to outlier points and noises as a result of employing the $l_{2,1}$ norm formulations of NMF and the local coordinate constraint regularization term.

(iii) Future research on this topic will include how to use multicore processors [48, 49] to accelerate our proposed method and how to extend the idea of semisupervised learning to the existing clustering algorithms.

6. Conclusion

In this study, we proposed a novel matrix decomposition method (RSNLCF) to learn an efficient representation for data in a semisupervised learning scenario. An efficient iterative algorithm for RSNLCF was also presented. The convergence of the presented method was theoretically proved. Extensive experiments over diverse datasets demonstrated
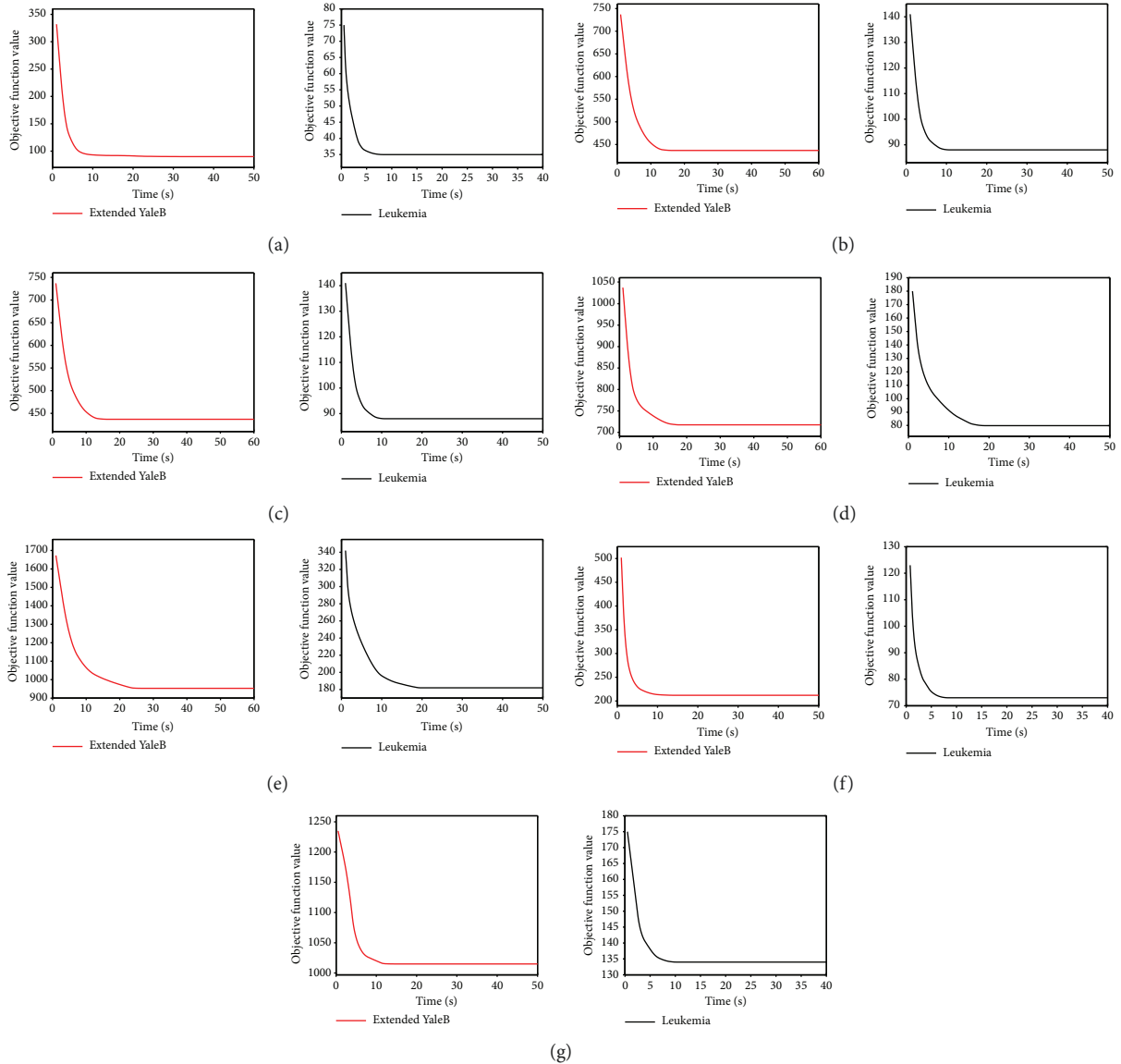
FIGURE 9: The curve of objective function value versus computational time on the extended YaleB and leukemia datasets. (a) RNMF, (b) semi-GNMF, (c) CNMF, (d) LCSNMF, (e) URNGE, (f) NLCF, and (g) RSNLCF.

that the presented method is quite effective and robust at learning an efficient data representation for clustering tasks. More importantly, experimental results revealed that our optimization algorithm quickly converges, indicating that our method can be utilized to solve practical problems.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.

[2] H. Qi, K. Li, Y. Shen, and W. Qu, "Object-based image retrieval with kernel on adjacency matrix and local combined features," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 8, no. 4, pp. 1–18, 2012.

[3] J. Wei, L. Min, and Z. Yongqing, "Neighborhood preserving convex nonnegative matrix factorization," *Mathematical Problems in Engineering*, vol. 2014, 8 pages, 2014.

[4] P. Li, J. Bu, C. Chen, Z. He, and D. Cai, "Relational multimanifold coclustering," *IEEE Transactions on Cybernetics*, vol. 43, no. 6, pp. 1871–1881, 2013.

[5] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, NY, USA, 1973.

[6] J. Zhao, L. Shi, and J. Zhu, "Two-stage regularized linear discriminant analysis for 2-D data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1669–1681, 2015.

[7] Y. Gao, X. Wang, Y. Cheng, and Z. J. Wang, "Dimensionality reduction for hyperspectral data based on class-aware tensor neighborhood graph and patch alignment," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 8, pp. 1582–1593, 2015.

[8] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.

[9] Y. Chen, J. Zhang, D. Cai, W. Liu, and X. He, "Nonnegative local coordinate factorization for image representation," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 969–979, 2013.

[10] C. Hou, F. Nie, X. Li, D. Yi, and Y. Wu, "Joint embedding learning and sparse regression: a framework for unsupervised feature selection," *IEEE Transactions on Cybernetics*, vol. 44, no. 6, pp. 793–804, 2014.

[11] Z. Ma, F. Nie, Y. Yang, J. R. R. Uijlings, and N. Sebe, "Web image annotation via subspace-sparsity collaborated feature selection," *IEEE Transactions on Multimedia*, vol. 14, no. 4, pp. 1021–1030, 2012.

[12] F. Nie, H. Huang, X. Cai, and C. H. Ding, "Efficient and robust feature selection via joint $\ell_{1,2}$-norms minimization," *Advances in Neural Information Processing Systems*, pp. 1813–1821, 2010.

[13] Z. Li, J. Liu, Y. Yang, X. Zhou, and H. Lu, "Clustering-guided sparse structural learning for unsupervised feature selection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 9, pp. 2138–2150, 2014.

[14] L. Du and Y. D. Shen, "Unsupervised feature selection with adaptive structure learning," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '15*, pp. 209–218, Sydney, NSW, Australia, 2015.

[15] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[16] H. Liu, Z. Wu, X. Li, D. Cai, and T. S. Huang, "Constrained nonnegative matrix factorization for image representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 7, pp. 1299–1311, 2012.

[17] W. Xu and Y. Gong, "Document clustering by concept factorization," in *Proceedings of the 27th annual international conference on Research and development in information retrieval - SIGIR '04*, pp. 202–209, Sheffield, UK, 2004.

[18] R. Zhi, M. Flierl, Q. Ruan, and W. B. Kleijn, "Graph-preserving sparse nonnegative matrix factorization with application to facial expression recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 41, no. 1, pp. 38–52, 2011.

[19] J. Ye and Z. Jin, "Dual-graph regularized concept factorization for clustering," *Neurocomputing*, vol. 138, pp. 120–130, 2014.

[20] P. O. Hoyer, "Non-negative sparse coding. Neural Networks for Signal Processing, 2002," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 557–565, Martigny, Switzerland, 2002.

[21] S. Zafeiriou, A. Tefas, I. Buciu, and I. Pitas, "Exploiting discriminant information in nonnegative matrix factorization with application to frontal face verification," *IEEE Transactions on Neural Networks*, vol. 17, no. 3, pp. 683–695, 2006.

[22] C. H. Q. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[23] H. Liu, Z. Yang, Z. Wu, and X. Li, "A-optimal non-negative projection for image representation," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1592–1599, Providence, RI, USA, 2012.

[24] D. Cai, X. He, J. Han, and T. S. Huang, "Graph regularized nonnegative matrix factorization for data representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1548–1560, 2011.

[25] A. Cichocki, R. Zdunek, and S. Amari, "Csiszár's divergences for non-negative matrix factorization: family of new algorithms," in *Independent Component Analysis and Blind Signal Separation. ICA 2006*, pp. 32–39, Springer, Berlin Heidelberg, 2006.

[26] L. Zhang, Z. Chen, M. Zheng, and X. He, "Robust non-negative matrix factorization," *Frontiers of Electrical and Electronic Engineering in China*, vol. 6, no. 2, pp. 192–200, 2011.

[27] H. Zhang, Z.-J. Zha, S. Yan, M. Wang, and T.-S. Chua, "Robust nonnegative matrix factorization using L21-norm," in *Proceedings of the 20th ACM international conference on Information and knowledge management*, pp. 673–682, Glasgow, Scotland, UK, 2011.

[28] H. Zhang, Z.-J. Zha, S. Yan, M. Wang, and T.-S. Chua, "Robust non-negative graph embedding: towards noisy data, unreliable graphs, and noisy labels," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2464–2471, Providence, RI, USA, 2012.

[29] C. Yan, H. Xie, D. Yang, J. Yin, Y. Zhang, and Q. Dai, "Supervised hash coding with deep neural network for environment perception of intelligent vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 284–295, 2018.

[30] C. Yan, H. Xie, S. Liu, J. Yin, Y. Zhang, and Q. Dai, "Effective Uyghur language text detection in complex background images for traffic prompt identification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 1, pp. 220–229, 2018.

[31] W. Zhang, B. Ma, K. Liu, and R. Huang, "Video-based pedestrian re-identification by adaptive spatio-temporal appearance model," *IEEE Transactions on Image Processing*, vol. 26, no. 4, pp. 2042–2054, 2017.

[32] W. Zhang, S. Hu, K. Liu, and J. Yao, "Motion-free exposure fusion based on inter-consistency and intra-consistency," *Information Sciences*, vol. 376, no. C, pp. 190–201, 2017.

[33] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using Gaussian fields and harmonic functions," in *Proceedings of the 20th International conference on Machine learning (ICML-03)*, pp. 912–919, Washington, DC, USA, 2003.

[34] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," *Advance in Neural Information Processing Systems*, vol. 16, no. 16, pp. 321–328, 2003.

[35] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[36] F. Wang and C. Zhang, "Label propagation through linear neighborhoods," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 1, pp. 55–67, 2008.

[37] S. Xiang, F. Nie, and C. Zhang, "Semi-supervised classification via local spline regression," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2039–2053, 2010.

[38] Y. Han, Y. Yang, Y. Yan, Z. Ma, N. Sebe, and X. Zhou, "Semisupervised feature selection via spline regression for video semantic recognition," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 2, pp. 252–264, 2015.

[39] W. Zhang, C. Qu, L. Ma, J. Guan, and R. Huang, "Learning structure of stereoscopic image for no-reference quality assessment with convolutional neural network," *Pattern Recognition*, vol. 59, pp. 176–187, 2016.

[40] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, "Deep neural networks for wireless localization in indoor and outdoor environments," *Neurocomputing*, vol. 194, pp. 279–287, 2016.

[41] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Advances in Neural Information Processing Systems*, pp. 2223–2231, MIT Press, 2009.

[42] S. Xiang, F. Nie, C. Zhang, and C. Zhang, "Nonlinear dimensionality reduction with local spline embedding," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 9, pp. 1285–1298, 2009.

[43] J. Duchon, "Splines minimizing rotation-invariant seminorms in Sobolev spaces," in *Constructive Theory of Functions of Several Variables*, pp. 85–100, Springer, Berlin, Heidelberg, 1977.

[44] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge university press, 2004.

[45] W. Xu, X. Liu, and Y. Gong, "Document clustering based on non-negative matrix factorization," in *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval - SIGIR '03*, pp. 267–273, Toronto, Canada, 2003.

[46] L. Lovsz and M. D. Plummer, *Matching Theory*, American Mathematical Society, 2009.

[47] H. Gao, F. Nie, and H. Huang, "Local centroids structured non-negative matrix factorization," in *Thirty-First AAAI Conference on Artificial Intelligence (AAAI 2017)*, pp. 1905–1911, San Francisco, CA, USA, 2017.

[48] C. Yan, Y. Zhang, J. Xu et al., "Efficient parallel framework for HEVC motion estimation on many-core processors," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 12, pp. 2077–2089, 2014.

[49] C. Yan, Y. Zhang, J. Xu et al., "A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 573–576, 2014.