Semi-supervised Multi-view Manifold Discriminant Intact Space Learning

Lu Han[1], Fei Wu[1] and Xiao-Yuan Jing[1,2]*
[1] College of Automation, Nanjing University of Posts and Telecommunications
Nanjing 210003, China
[e-mail: hanl@njupt.edu.cn, wufei_8888@126.com, jingxy_2000@126.com]
[2] State Key Laboratory of Software Engineering, School of Computer, Wuhan University
Wuhan 430072, China
Corresponding author: Xiao-Yuan Jing

Abstract

Semi-supervised multi-view latent space learning is gaining considerable popularity recently in many machine learning applications due to the high cost and difficulty to obtain the large amount of label information of data. Although some semi-supervised multi-view latent space learning methods have been presented, there is still much space for improvement: 1) How to learn latent discriminant intact feature representations by employing data of multiple views; 2) How to exploit the manifold structure of both labeled and unlabeled point in the learned latent intact space effectively. To address the above issues, we propose an approach called semi-supervised multi-view manifold discriminant intact space learning (SM^2DIS) for image classification in this paper. SM^2DIS aims to seek a manifold discriminant intact space for data of different views by making use of both the discriminant information of labeled data and the manifold structure of both labeled and unlabeled data. Experimental results on MNIST, COIL-20, Multi-PIE, and Caltech-101 databases demonstrate the effectiveness and robustness of our proposed approach.

1. Introduction

With the rapid development of data acquisition technology, there exists different forms of representations (features) with respect to a certain object [1-3]. For example, in the application of computer vision, multi-pose images can be obtained by shooting from different angles. Clothing images can be described by color, texture and shape. These multiple types of data are generally named multi-view data [4-6]. For classification problems, multi-view data usually contains more abundant information and representation ability, especially the complementary information, than that of single view. Thus, integrating information from different views can help to enhance the learning performance.

Recently, multi-view latent space learning, which aims to discover a latent feature representation to represent different views of one data, is becoming an important research direction in machine learning and image processing [7-10]. To exploit the intrinsic geometry of the data distribution in each view, graph regularized partial multi-view clustering (GPMVC) [11] is proposed and applied in the problem of partial-views (every view has missing information). Bi-level multi-view latent sparse learning (BLMV) [12] not only explores both shared and private latent factors of multiple views, but also reveals the nonlinearly inherent in data by taking the manifold structure of each view into consideration. Assuming that each single view of the data only captures partial information, and all views together can carry redundant information about the learned intact feature representation, multi-view intact space learning (MISL) [13] discovers a latent intact representation for each data of multiple views. When there exists labeled samples, these unsupervised multi-view latent space learning methods cannot employ the label information of labeled samples to guide classification.

Some researchers utilize the label information of multi-view data to extract effect discriminant information [14-21]. For many real-world applications such as image annotation, gene function prediction, and insider threat detection, hierarchical multi-latent space (HiMLS) [16] learns a hierarchical multi-latent space to jointly model the triple types of heterogeneity, i.e., task heterogeneity, view heterogeneity, and label heterogeneity of the data. Discriminative multi-view interactive image re-ranking (DMINTIR) [17] is well applied in multi-view interactive image re-ranking. By maximally preserving the discriminative information in the learned latent subspace, DMINTIR losslessly recovers the original feature space. Based on the assumption that there exists one single discriminative intact feature representation for different views of the data point, multiview intact and single-view classifier learning (MISC) [18] tries to recover the discriminative intact feature space for the data points and learns a classifier in the intact space with the help of class labels of the data. However, some favorable information in unlabeled samples cannot be effectively exploited.

Due to the high cost and the difficulty of obtaining the large amount number of label information of data, a lot of applications in machine learning can be regarded as semi-supervised learning problems [22-24]. Recently, multi-view latent space learning in semi-supervised setting has attracted a lot of attentions. Semi-supervised unified latent factor learning (SULF) [25] learns a predictive unified latent representation by leveraging both complementary information among multiple views and the supervision of partially labeled information. First, SULF simultaneously performs Nonnegative Matrix Factorization (NMF) with different view data matrices to exploit the information of multiple feature sets collaboratively. Second, a $l_{2,1}$ -norm regularized regression model is used to minimize the prediction error on partially labeled data with the latent representation. Multi-view concept learning (MCL) [26] is proposed to address the problem of learning a nonnegative conceptual

representation from multiple view data. In the learned common latent space across different views, MCL can reflect the semantic relationships between data items through graph embedding regularization on labeled items. Meanwhile, MCL allows each latent factor to be associated with a subset of views. As a result, conceptual patterns hidden in multi-view features can be captured flexibly. Partially shared latent factor (PSLF) [27] jointly exploits both consistency and complementary properties of multi-view data to obtain a compact and comprehensive partially shared latent representation, which is composed of the common latent factor shared by all views and view-specific latent factors. Adaptive multi-view semi-supervised nonnegative matrix factorization (AMVNMF) [28] uses label information as hard constraints to ensure data with same class label are clustered together, so that the discriminating power of new representations are enhanced.

1.1 Motivation and Contributions

Although some semi-supervised multi-view latent space learning methods have been presented, there is still much space for improvement. We elaborate it in the following two points:

(1) How to discover the latent intact feature representations for multiple views of data points: In practice, due to the noises in multiple views, the performances of most multi-view latent space learning methods are usually affected. MISL integrates the encoded complementary information in multiple insufficient views to discover a latent intact representation of the data. However, existing latent space based semi-supervised multi-view learning methods, such as SULF [25], MCL [26], PSLF [27], and AMVNMF [28], fails to discover latent intact (complete and not damaged) space due to the influence of noises in insufficient views.

(2) How to effectly use the local manifold structure from the unlabeled latent feature representations: the exploitation of the manifold structure of data will bring a better classification performance [29]. However, SULF, MCL, PSLF, and AMVNMF use the unlabeled information only in the process of matrix decomposition, and do not focus on exploiting the manifold structure of latent feature representations of the unlabeled data.

In this paper, we propose a semi-supervised multi-view manifold discriminant intact space learning (SM^2DIS) approach. The contributions of our work are summarized as follows:

(1) SM^2DIS learns the discriminant intact feature representation for each labeled training data point. To achieve this goal, SM^2DIS makes generated intra-class multi-view data points near to each other, and simultaneously makes generated inter-class multi-view data points as far as possible. The learned feature representations are intact as SM^2DIS uses all view generation matrices to restore the view feature space.

(2) SM^2DIS explores the manifold structure of data points in the learned intact space by employing the information of both labeled and unlabeled data points to construct nearest graph. As a result, the classification performance of our proposed SM^2DIS can be effectively enhanced.

(3) We propose an iterative solution for SM^2DIS. Experiments on several datasets show the effectiveness of the proposed SM^2DIS.

The rest of the paper is organized as follows: In Section 2, some related works including MISL, SULF, MCL, PSLF are introduced. The proposed SM^2DIS approach is introduced in Section 3. Section 4 reports the experimental results on four benchmark datasets. Finally, we give conclusion in Section 5.

2. Related Work

2.1 Multi-view Intact Space Learning

Assume that each data of multiple views $\left\{z_i^v\right\}_{v=1}^M$ have an intact feature representation x_i, where z_i^v is the ith point in the vth view space, x_i is the ith data point in the latent intact space χ, multi-view intact space learning (MISL) [13] focuses on learning m view generation functions $W = \{W_1, W_2, ..., W_m\}$ and a latent intact space χ by solving the following problem:

$$\min_{\chi, W} \frac{1}{mn} \sum_{v=1}^m \sum_{i=1}^n \log\left(1 + \frac{\left\|z_i^v - W_v x_i\right\|_2^2}{c^2}\right) + C_1 \sum_{v=1}^m \left\|W_v\right\|_F^2 + C_2 \sum_{i=1}^n \left\|x_i\right\|_2^2 \tag{1}$$

where m denotes the number of views, n is the number of data points in the intact space χ, $W_v x_i$ is the generated multi-view data point of the ith data point with repect to the vth view, c is a constant scale parameter, C_1 and C_2 are non-negative parameters. The first term in formula (1) reflects the relationship between each view space and the generated multi-view data points, where the reconstruction error over the latent space is minimized by using Cauchy estimator [30]. The second and the third term in formula (1) are regularization terms.

2.2 Semi-supervised Unified Latent Factor Learning

Semi-supervised unified latent factor learning (SULF) [25] uses joint non-negative matrix factorization to learn a unified latent space shared by all view data. At the same time, a regularized regression model is adopted to minimize the prediction error of the partially labeled data using the latent representation. Assume X^k is the sample set of the kth view, Y is the label matrix with its element y_{ir} defined as: if the rth data point belongs to the ith class, the value of y_{ir} is set to 1; otherwise the value of y_{ir} is set to 0. The objective function of SULF is defined as

$$\min \sum_{k=1}^m \pi^k \left\|X^k - U^k V\right\|_F^2 + \lambda \left\|\Pi\right\|^2 + \beta \left\|W^T V_l - Y\right\|_F^2 + \gamma \left\|W\right\|_{2,1}$$
$$s.t. \ U, V, \Pi \geq 0, \sum_{k=1}^m \pi^k = 1 \tag{2}$$

where π^k, U^k are the weight vector and the base matrix with respect to the kth view respectively, $\Pi = \left(\pi^1, ..., \pi^k, ..., \pi^m\right)$ and λ control the smoothness of Π. SULF divides the common factor matrix V into two parts V_l and V_{ul}. V_l, which is consist of the first R column vectors in V, is the compact representation of R labeled data. V_{ul}, which is consist of the remained $n - R$ column vectors in V, is the compact representation of $n - R$ unlabeled data. For the labeled part V_l, SULF learns a linear classifier W to fit the label matrix Y.

2.3 Multi-view Concept Learning

Multi-view concept learning (MCL) [26] was proposed to capture conceptual factors from multi-view data by jointly learning basis matrix U^k for the kth view and the consensus encoding matrix V shared by all views. Besides, by constructing the within-class affinity graph G^a and the between-class penalty graph G^p for labeled data, MCL captures the semantic relationship between data items. The objective function of MCL can be formulated

as

$$\min_{\{U^k\}_{k=1}^m, V} \frac{1}{2}\sum_{k=1}^m \left\| X^k - U^k V \right\|_F^2 + \alpha \sum_{k=1}^m \left\| U^k \right\|_{1,\infty} + \frac{\beta}{2}\left\{ tr\left[V^l L^a \left(V^l\right)^T \right] - tr\left[V^l L^p \left(V^l\right)^T \right] \right\} + \gamma \left\| V \right\|_{1,1}$$

$$s.t. \ U_{iv}^k \geq 0, 1 \geq V_{vj} \geq 0, \forall i, j, v, k$$

(3)

where V^l denotes the labeled items in consensus encoding matrix V, L^a and L^p are the graph Laplacian matrices for the within-class affinity graph and the between-class affinity graph, respectively (see [26]), α and β are tunable parameters. MCL uses block coordinate descend method [31] to get the solution of problem (3).

2.4 Partially Shared Latent Factor Learning

Partially shared latent factor learning (PSLF) [27] jointly exploits the properties of consistency and complementarity of multi-view data. It aims to learn the partially shared latent factor matrix $V = \left[V_s^1; V_s^2; ...; V_s^m; V_C \right]$ through leveraging data matrices $\{X_k\}_{k=1}^m$ and the label matrix Y, where V_C is the common part shared by all views and $V_s^1, V_s^2, ..., V_s^m$ are m view-specific parts. Assume V_l, the first N_l columns of V, denotes the first N_l labeled instances, and V_u denotes the remained N_u unlabeled instances, PSLF solves the following problem:

$$\min \sum_{k=1}^m \pi^k \left\| X^k - U^k V^k \right\|_2^F + \lambda \left\| \Pi \right\|_2^2 + \beta \left(\left\| W^T V_l - Y \right\|_F^2 + \gamma \left\| W \right\|_{2,1} \right)$$

$$s.t. \ U^1, ..., U^m, V^1, ..., V^m, \Pi \geq 0, \sum_{k=1}^m \pi^k = 1$$

(4)

where W is the regression coefficient matrix, λ, β and γ are nonnegative parameters. As V^k can be divided into four components: $V_{s.l}^k, V_{s.u}^k, V_{c.l}, V_{c.u}$, the partially shared latent factor matrix V in formula (4) are solved by updating $V_{s.l}^k, V_{s.u}^k, V_{c.l}, V_{c.u}$ iteratively (see [27]). By obtaining the solutions of V_C and V_s^k, PSLF well reflects both the consistency information and complementary information of multi-view data simultaneously.

2.5 Adaptive Multi-view Semi-supervised Nonnegative Matrix Factorization

Adaptive multi-view semi-supervised nonnegative matrix factorization (AMVNMF) [28] is developed for accurate clustering multi-view data so that all data with the same label should be clustered regardless of their multiple views. Assume Z^k is the auxiliary matrix of the k th view, I_{n-N_l} is an $(n-N_l) \times (n-N_l)$ identity matrix, and C is the indicator matrix, we set $c_{ij} = 1$ if data point x_i belongs to the j th class, otherwise $c_{ij} = 0$. AMVNMF extended the traditional unsupervised NMF to a semi-supervised multi-view manner by solving the following problem:

$$\min \sum_{k=1}^m \left\| X^k - U^k \left(Z^k\right)^T A^T \right\|_F^2$$

$$s.t. \ U, Z \geq 0$$

(5)

where $A = \begin{pmatrix} C_{N_l \times C} & 0 \\ 0 & I_{n-N_l} \end{pmatrix}$ is the label constraint matrix A, which is constructed only based on the label information and consistent for all features (see [28]). To get the solution of formula (5), AMVNMF updates the variables iteratively and more details can be referenced to [28].

3. The Model of SM²DIS

In this section, our SM²DIS approach is introduced. First, we describe the semi-supervised multi-view manifold discriminant term, and then we give the objective function of SM²DIS. Finally, we introduce the classification strategy on SM²DIS.

3.1 Semi-supervised Multi-view Manifold Discriminant Term

Suppose each training sample x_k is characterized from M views, and $X = \left\{ x_k^j \mid x_k^j \in X^{(j)}, k = 1, 2, ..., n; j = 1, 2, ..., M \right\}$ denote the training sample sets. Each view set $X^{(j)} = \left[X_L^{(j)}, X_U^{(j)} \right]$ contains two parts: $X_L^{(j)}$ is composed of C class labeled samples and $X_U^{(j)}$ is composed of $n^{(C+1)}$ unlabeled samples. We define $l(\cdot)$ as the label information of samples, and the lables of unlabeled samples $\left\{ X_U^{(j)} \right\}_{j=1}^{M}$ are set to $C + 1$.

The main purpose of our proposed SM²DIS approach is to learn a manifold discriminant intact space for all views of data. Let $h_k, k = 1, 2, ..., n$ denote the learned feature representations for the training samples $\left\{ x_k^j \right\}_{j=1}^{M}$ from M views, $H = \left[H_L, H_U \right] = \left\{ h_k \right\}_{k=1}^{n}$ denote the learned intact feature space, W_j denote the view generation function with respect to the j th view, $W_j h_k$ denote the generated multi-view data point of sample x_k^j. **Table 1** gives the explanations about the variables. On the one hand, all the learned intact feature representations for the labeled samples are discriminative. On the other hand, the intrinsic geometrical structure of both labeled and unlabeled intact feature representations can be employed simultaneously. Specifically, if two samples $x_k^j \in X$ and $x_t^r \in X$ have the same label or are unlabeled samples within the same neighborhood (at least one of them is unlabeled), the generated multi-view data points $W_j h_k$ and $W_r h_t$ should also be close to each other; if x_k^j and x_t^r are from different classes or are unlabed samples of different neighborhoods (at least one of them is unlabeled), $W_j h_k$ and $W_r h_t$ should keep as distant as possible to each other.

To explore the label information and the local geometric structure of the learned intact feature representations, according to graph embedding theory [32], we construct two nearest neighbor graphs P and $\tilde{P}$. The weight matrix S on graph P and the weight matrix $\overline{S}$ on graph $\tilde{P}$ are defined separately by:

$$S_{k,t}^{(j,r)} = \begin{cases} 1, \; if \; l\left(x_k^j \right) = l\left(x_t^r \right) and \; l\left(x_k^j \right) \neq C + 1 \\ x_k^{jT} \cdot x_t^r \big/ \left\| x_k^j \right\| \cdot \left\| x_t^r \right\|, \; if \; l\left(x_k^j \right) \; or \; l\left(x_t^r \right) = C + 1, and \; x_k^j \in N_K\left(x_t^r \right) \; or \; x_t^r \in N_K\left(x_k^j \right) \\ 0, \; else \end{cases} \quad (6)$$

$$\overline{S}_{k,t}^{(j,r)} = \begin{cases} 1, \; if \; l\left(x_k^j \right) \neq l\left(x_t^r \right), \; and \; l\left(x_k^j \right), \; l\left(x_t^r \right) \neq C + 1 \\ x_k^{jT} \cdot x_t^r \big/ \left\| x_k^j \right\| \cdot \left\| x_t^r \right\|, \; if \; l\left(x_k^j \right) \; or \; l\left(x_t^r \right) = C + 1, \; and \; x_k^j \notin N_K\left(x_t^r \right) \; or \; x_t^r \notin N_K\left(x_k^j \right) \\ 0, \; else \end{cases} \quad (7)$$

where $N_K\left(x_t^r \right)$ denotes the K nearest neighbors of sample x_t^r. Then we define the semi-supervised multi-view manifold discriminant term as

$$\sum_{j=1}^{M}\sum_{k=1}^{n}\sum_{r=1}^{M}\sum_{t=1}^{n}\left(\left\|W_jh_k - W_rh_t\right\|_2^2 S_{k,t}^{(j,r)} - \left\|W_jh_k - W_rh_t\right\|_2^2 \overline{S}_{k,t}^{(j,r)}\right) \tag{8}$$

Table 1. Explanation about the variables

Variable	Explanation
x_k	the k th training sample in the original space
X	feature sets containing M views
$X_L^{(j)}$	the labeled part of the j th view
$X_U^{(j)}$	the unlabeled part of the j th view
h_k	the feature representation learned from $\left\{x_k^j\right\}_{j=1}^{M}$
H	the learned manifold discriminant intact space
H_L	the labeled part in the learned manifold discriminant intact space
H_U	the unlabeled part in the learned manifold discriminant intact space
W_j	the learned view generation function for the j th view
W_jh_k	the generated multi-view data point which is generated by view generation matrix W_j over the feature representation h_k

3.2 The Objective Function of SM²DIS

Inspired by MISL [13], minimizing the reconstruction error over the intact space can well model the relationship between the view space and the learned intact space. Based on (8), we formulate the objective function of SM²DIS as follows:

$$\min_{V,W}\left(\begin{array}{l} \alpha\sum_{j=1}^{M}\sum_{k=1}^{n}\sum_{r=1}^{M}\sum_{t=1}^{n}\left(\left\|W_jh_k - W_rh_t\right\|_2^2 S_{k,t}^{(j,r)} - \left\|W_jh_k - W_rh_t\right\|_2^2 \overline{S}_{k,t}^{(j,r)}\right) \\ + \frac{1}{Mn}\sum_{j=1}^{M}\sum_{k=1}^{n}\log\left(1 + \frac{\left\|x_k^j - W_jh_k\right\|_2^2}{\gamma^2}\right) + c\left(\sum_{j=1}^{M}\left\|W_j\right\|_F^2 + \sum_{k=1}^{n}\left\|h_k\right\|_2^2\right) \end{array}\right) \tag{9}$$

where α and c are non-negative coefficient, γ is a constant scale parameter, $\sum_{j=1}^{M}\left\|W_j\right\|_F^2$ and

$\sum_{k=1}^{n}\left\|h_k\right\|_2^2$ are regularization terms to penalize the the view generation matrix and the intact space.

4. Optimization

The solution of the objective function (9) can be obtained in an iterative process. For each iteration, we use following strategies: 1) when updating view generation function W_j corresponding to the j th view, all data points $\{h_k\}_{k=1}^{n}$ in the learned intact space H are fixed; 2) when updating each data point in the learned intact space, view generation functions of all views are fixed.

4.1 Updating Rule for Generation Function W_j

Assume that $\{h_k\}_{k=1}^n$ are fixed, and when calculating W_j, all view generation functions $W_r (r \neq j)$ are fixed. Thus, the objective function in Eq. (9) can be reduced to:

$$J(W_j) = \min_{W_j} \left(\alpha(Z_1 + Z_2) + \frac{1}{n} \sum_{k=1}^n \log \left(1 + \frac{\left\| x_k^j - W_j h_k \right\|_2^2}{\gamma^2} \right) + c \left\| W_j \right\|_F^2 \right) \tag{10}$$

where

$$Z_1 = 2 \sum_{\substack{r=1, \\ r \neq j}}^M \sum_{k=1}^n \sum_{t=1}^n \left(\left\| W_r h_k - W_j h_t \right\|_2^2 S_{k,t}^{(r,j)} - \left\| W_r h_k - W_j h_t \right\|_2^2 \overline{S}_{k,t}^{(r,j)} \right) \tag{11}$$

$$Z_2 = \sum_{k=1}^n \sum_{t=1}^n \left(\left\| W_j h_k - W_j h_t \right\|_2^2 S_{k,t}^{(j,j)} - \left\| W_j h_k - W_j h_t \right\|_2^2 \overline{S}_{k,t}^{(j,j)} \right) \tag{12}$$

Let $\dfrac{\partial J(W_j)}{\partial W_j} = 0$, we get

$$W_j = \left(A_j + \sum_{k=1}^n x_k^j Q_k h_k^T \right) \left(B_j + \sum_{k=1}^n h_k Q_k h_k^T + nc \right)^{-1} \tag{13}$$

where Q_k, A_j, and B_j are separately defined in (14), (15) and (16)

$$Q_k = \frac{1}{\gamma^2 + \left\| x_k^j - W_j h_k \right\|_2^2} \tag{14}$$

$$A_j = 2\alpha n \sum_{\substack{r=1, \\ r \neq j}}^M \sum_{k=1}^n \sum_{t=1}^n \left(W_r h_k h_t^T S_{k,t}^{(r,j)} - W_r h_k h_t^T \overline{S}_{k,t}^{(r,j)} \right) \tag{15}$$

$$B_j = \alpha n \left(\begin{array}{c} 2 \displaystyle\sum_{\substack{r=1, \\ r \neq j}}^M \sum_{k=1}^n \sum_{t=1}^n \left(h_t h_t^T S_{k,t}^{(r,j)} - h_t h_t^T \overline{S}_{k,t}^{(r,j)} \right) \\ + \displaystyle\sum_{k=1}^n \sum_{t=1}^n \left((h_k - h_t)(h_k - h_t)^T S_{k,t}^{(j,j)} - (h_k - h_t)(h_k - h_t)^T \overline{S}_{k,t}^{(j,j)} \right) \end{array} \right) \tag{16}$$

The details of calculating each W_j can be referred to Algorithm 1.

4.2 Updating Rule for Each Point of Intact Space H

Assume that $\{W_j\}_{j=1}^M$ are fixed, and when calculating h_k, all other points in the intact space $h_t (t \neq k)$ are fixed. Thus, the objective function in Eq. (9) can be reduced to:

$$J(h_k) = \min_{h_k} \left(\alpha(Z_1' + Z_2') + \frac{1}{M} \sum_{j=1}^M \log \left(1 + \frac{\left\| x_k^j - W_j h_k \right\|_2^2}{\gamma^2} \right) + c \left\| h_k \right\|_F^2 \right) \tag{17}$$

where

$$Z_1' = 2 \sum_{j=1}^M \sum_{r=1}^M \sum_{\substack{t=1, \\ t \neq k}}^n \left(\left\| W_j h_t - W_r h_k \right\|_2^2 S_{t,k}^{(j,r)} - \left\| W_j h_t - W_r h_k \right\|_2^2 \overline{S}_{t,k}^{(j,r)} \right) \tag{18}$$

$$Z_2' = \sum_{j=1}^{M}\sum_{r=1}^{M}\left(\left\|W_j h_k - W_r h_k\right\|_2^2 S_{k,k}^{(j,r)} - \left\|W_j h_k - W_r h_k\right\|_2^2 \overline{S}_{k,k}^{(j,r)} \right) \tag{19}$$

Let $\dfrac{\partial J\left(h_k\right)}{\partial h_k} = 0$, we get

$$h_k = \left(D_k + \sum_{j=1}^{M} W_j^T Q_j W_j + Mc \right)^{-1}\left(E_k + \sum_{j=1}^{M} W_j^T Q_j x_k^j \right) \tag{20}$$

where Q_j, D_k and E_k are separately defined in (21), (22) and (23)

$$Q_j = \frac{1}{\gamma^2 + \left\|x_k^j - W_j h_k\right\|_2^2} \tag{21}$$

$$D_k = \alpha M \begin{pmatrix} 2\sum_{j=1}^{M}\sum_{r=1}^{M}\sum_{\substack{t=1,\\t\neq k}}^{n}\left(W_r^T W_r S_{t,k}^{(j,r)} - W_r^T W_r \overline{S}_{t,k}^{(j,r)} \right) \\[2mm] +\sum_{j=1}^{M}\sum_{r=1}^{M}\left(\left(W_j - W_r\right)^T\left(W_j - W_r\right) S_{k,k}^{(j,r)} - \left(W_j - W_r\right)^T\left(W_j - W_r\right)\overline{S}_{k,k}^{(j,r)} \right) \end{pmatrix} \tag{22}$$

$$E_k = 2\alpha M \sum_{j=1}^{M}\sum_{r=1}^{M}\sum_{\substack{t=1,\\t\neq k}}^{n}\left(W_r^T W_j h_t S_{t,k}^{(j,r)} - W_r^T W_j h_t \overline{S}_{t,k}^{(j,r)} \right) \tag{23}$$

The process of solving each h_k is similar to algorithm 1.

Algorithm 1. The procedure of solving W_j

Input: Given feature set X^j from the j th view, and parameters α, γ, c.

Output: W_j.

Initialize: all view generation functions $W_r(r \neq j)$, the intact space H, and W_j^0.

Compute covariance matrices S and $\overline{S}$ using Eq. (6) and (7).

Compute matrices A_j, and B_j using Eq. (15) and (16).

Compute Q_k using Eq. (14) when W_j is set to W_j^0.

While not converged do

1. Obtain the estimate of W_j using Eq. (13).

2. Update Q_k using the estimate of W_j.

End While

4.3 The Classification Schema of SM²DIS

For the given multi-view features $\left\{y^1, y^2, ..., y^M\right\}$ of a specific testing sample y, assume y^Δ is the feature representation of y in the learned manifold discriminant intact space, once the optimal view generation functions $W^\triangle = \left\{W_1^\triangle, W_2^\triangle, ..., W_M^\triangle\right\}$ is obtained, the optimal solution of y^Δ can be obtained by

$$\min_{y} \left(\frac{1}{M} \sum_{j=1}^{M} \log \left(1 + \frac{\left\| y^j - W_j^{\triangle} y \right\|_2^2}{\gamma^2} \right) + c \sum_{k=1}^{n} \|y\|_2^2 \right) \tag{24}$$

Then, we can use nearest neighbour classifier (NNC) [33] for classification. That is, the label of $y^{\triangle}$ is same to the feature representation point in H_L which has the minimal distance to $y^{\triangle}$.

Finally, we summarize the procedure of SM^{2}DIS in Algorithm 2.

Algorithm 2. Procedure of SM^{2}DIS

Input: Feature sets $\left\{ X^{(j)} \right\}_{j=1}^{M}$ and parameters α, γ, c.

Output: $\left\{ W_j \right\}_{j=1}^{M}$, $\left\{ h_k \right\}_{k=1}^{n}$.

Initilize: $\left\{ W_j \right\}_{j=1}^{M}$ and $\left\{ h_k \right\}_{k=1}^{n}$.

While not converged do

  Fix $\left\{ h_k \right\}_{k=1}^{n}$, update $\left\{ W_j \right\}_{j=1}^{M}$ using Eq. (13).

  Fix $\left\{ W_j \right\}_{j=1}^{M}$, update $\left\{ h_k \right\}_{k=1}^{n}$ using Eq. (20).

End While

Realize classification: For each testing sample, calculate the corresponding intact feature representation $y^{\triangle}$ by Eq. (24), and use NNC for classification.

4.4 Comparison with Related Works

In this section, we give a detailed comparison between our proposed SM^{2}DIS approach and compared methods:

Both MISL and our proposed SM^{2}DIS are based on the same assumption that each view only captures partial information about the learned latent intact feature representations. MISL [13] is an unsupervised learning method which integrates information from various views to find a latent intact space of data, while our SM^{2}DIS approach not only considers the discovery of the latent intact space, but also focuses on exploiting the discriminant information from labeled data and local geometrical structure of both labeled and unlabeled data in the learned latent intact space.

SULF [25], MCL [26], PSLF [27], and AMVNMF [28] are semi-supervised multi-view learning methods. As these methods don't take the view insufficiency problem stated in MISL into consideration, the learned feature representations are not intact. Our SM^{2}DIS learns the intact feature representations by using all view generation matrices to restore the view feature space. The detailed differences between our SM^{2}DIS and these methods are shown as follows:

(1) SULF models the label information as a factorization constraint on V_l, while our proposed SM^{2}DIS approach exploits the discriminant structure of the latent intact space. As a result, the learned feature representations $\left\{ h_k \right\}_{k=1}^{N}$ by our SM^{2}DIS approach are more discriminative. Besides, SM^{2}DIS utilizes the information of both labeled and unlabeled data for exploiting manifold structure of the learned latent intact space.

(2) MCL exploits the semantic relationship between labeled items in the learned latent space, while our SM^{2}DIS approach tries to reveal the local manifold structure of both labeled and

unlabeled points as well as the discriminant information in the latent intact space.

(3) PSLF extracts a compact and comprehensive partially shared latent representation for multi-view data. The label information used in sparse regression is to predict the cluster labels of labeled data, while our SM^{2}DIS approach is dedicated to extracting discriminant information from the latent intact space, which will be more favorable for classification. Besides, SM^{2}DIS reveals the relationship between the intact feature representations effectively.

(4) AMVNMF guarantees that all the samples sharing the same labels are grouped together by taking the label information as hard constraints, regardless they are coming from the same or different views. In addition to the use of label information, SM^{2}DIS considering the utilization of unlabeled samples to achieve a better classification performance.

5. Experimental Results and Analysis

5.1 Introduction of Databases

In this section, as in [34], four benchmark image databases are used to evaluate the effectiveness of our proposed SM^{2}DIS approach, including MNIST [35] handwritten digit database, COIL-20 [36] object database, Multi-PIE [37] database, and Caltech-101 [38] database.

MNIST database contains 1000 handwritten digit images from 10 (0~9) digits with 100 sample images per digit. The pixel size of each image is 28×28. COIL-20 object database consists of 1440 images from 20 objects with 72 samples per object. The pixel size of each image is 64×64. For MNIST and COIL-20 databases, Karhunen-Loeve (KL) transformation features [39], Gabor transformation features [40] and Local Binary Patterns (LBP) features [41] are extracted to express each image. Multi-PIE database consists of 75000 images from 337 different individuals. The size of each image is 64×64. For convience, 68 people from Multi-PIE dataset (24 samples per people) are selected for experiment. For Multi-PIE dataset, five feature sets, i.e., C05, C07, C09, C27 and C29, are used in the experiment. Caltech-101 dataset is a large-scale object dataset from 101 different categories such as airplanes, face, cellphone, etc.. Each image is roughly 300×200 pixels. For Caltech-101 dataset, we extract predefined distance matrices of PHOW color [42], geometric blur [43] and self-similarity features [44] to represent each image. For the above four databases, we use PCA transformation to reduce the dimension of features to 100.

5.2 Compared Methods and Experimental Settings

In this section, we compare the proposed SM^{2}DIS approach with several state-of-the-art methods on the above four databases, including three types of methods: semi-supervised multi-view feature learning methods: SULF [25], MCL [26], PSLF [27], AMVNMF [28]; unsupervised multi-view feature learning method MISL [13]; supervised multi-view feature learning method MISC [18].

To construct the semi-supervised scenario, the percentage of labeled samples in the training set is set to 40%, while the rest of training samples are chosen as unlabeled samples. Besides, to simulate the real enviroment, we add different values of signal-to-noise to MNIST, COIL-20, Multi-PIE, and Caltech-101 databases to evaluate the robustness of our approach. In the experiment, the value of signal noise ratio (SNR) is set to 5.

On the four databases, we randomly select 40, 36, 8, 30 samples per class separately for

training, and use the remained for testing (on Caltech-101 database, the maximum is 50 images per category for testing). For unsupervised method MISL, all training samples (including labeled and unlabeled) are used. For supervised method MISC, only the labeled part of training samples are chosen for learning. The tuning parameters in SM^{2}DIS, i.e., non-negative coefficient α and c, the constant scale parameter γ, the number of nearest neighbors K, the dimension d of the manifold discriminant intact space) and all the parameters of other compared methods, are evaluated by 5-fold cross validation to avoid over-fitting. During the process of experiment, we observe that the adjustment of constant scale parameter γ has little effect on the results, thus we set $\gamma = 60$ consistently on the four databases. For parameters α, c, K, and d, we set the values of them as follows: on the MNIST database, $\alpha=0.4$, $c=0.06$, $K = 37$, $d = 190$; on the COIL-20 database, $\alpha=0.5$, $c=0.05$, $K = 25$, $d = 210$; on the Multi-PIE database, $\alpha=0.3$, $c=0.08$, $K = 36$, $d = 180$; on the Caltech-101 database, $\alpha=0.7$, $c=0.04$, $K = 15$, $d = 160$. We report the average classification results on four datasets according to 20 times of random experiments.

5.3 Comparison with State-of-the-art Methods

5.3.1 Classification Accuracy Evaluation

In this subsection, we conduct experiment to evaluate the classification performance of our proposed SM^{2}DIS. **Table 2** shows the average classification results on MNIST, COIL-20, Multi-PIE and Caltech-101 databases, with best classification performance in bold. According to the Table, our SM^{2}DIS outperforms the compared methods consistently on the four databases.

Table 2. Average classification accuries on four databases (%)

Method	MNIST	COIL-20	Multi-PIE	Caltech-101
MISL[13]	71.54	75.37	74.11	57.00
SULF [25]	73.85	76.77	75.10	59.65
MCL [26]	75.02	79.32	77.44	61.28
PSLF [27]	74.14	78.57	75.58	60.45
AMVNMF[28]	72.76	76.28	74.35	58.97
MISC [18]	72.79	78.18	76.51	60.17
SM^{2}DIS	**77.37**	**81.48**	**79.02**	**62.68**

Compared with multi-view latent space learning methods MISL, SULF, MCL, PSLF, AMVNMF, and MISC, we analysis the reasons for the performance improvement as follows: MISL is an unsupervised multi-view learning method, where the label information is not used in the process of classification. MISC ignores the utilization of information of unlabeled multi-view data. SULF takes use of the label information of partial labeled samples and correlation information from multi-view data. PSLF jointly exploits the properties of consistency and complementarity of multi-view data. AMVNMF clusters the data with the same label regardless they are from the same view or not. However, the above three semi-supervised multi-view learning methods SULF, PSLF, and AMVNMF can not extract the useful discriminative information and local geometrical information from the learned latent space. MCL exploits the discriminant information to capture the semantic relationship between data items in the latent space. However, the exploition of manifold structure is not considered.

Table 3. Average classification accuries with different percentages of labeled training samples on four databases (%)

Database	Percentage	MISL [13]	SULF [25]	MCL [26]	PSLF [27]	AMVNMF [28]	MISC [18]	SM^{2}DIS
MNIST	10%	71.54	68.25	71.27	69.88	67.95	67.66	**72.85**
	20%	71.54	69.35	72.05	70.46	68.77	70.13	**73.72**
	30%	71.54	70.33	72.99	71.69	69.85	71.65	**75.00**
	40%	71.54	73.85	75.02	74.14	72.76	72.79	**77.37**
	50%	71.54	74.87	77.41	76.61	74.05	75.22	**79.71**
COIL-20	10%	75.37	71.67	74.78	73.45	70.95	70.22	**76.15**
	20%	75.37	73.21	75.97	74.52	72.78	73.39	**77.39**
	30%	75.37	74.87	77.28	76.65	74.09	75.62	**79.23**
	40%	75.37	76.77	79.32	78.57	76.28	78.18	**81.48**
	50%	75.37	78.68	81.83	79.97	77.72	79.27	**83.97**
Multi-PIE	10%	74.11	71.32	73.12	72.36	69.38	70.55	**74.67**
	20%	74.11	72.08	74.10	73.25	70.97	71.22	**75.83**
	30%	74.11	73.05	75.29	74.20	72.62	73.74	**77.05**
	40%	74.11	75.10	77.44	75.58	74.35	76.51	**79.02**
	50%	74.11	76.95	79.07	78.24	76.00	77.85	**81.50**
Caltech-101	10%	57.00	52.95	56.04	54.83	52.11	54.75	**57.35**
	20%	57.00	54.28	57.21	56.89	54.12	55.65	**58.25**
	30%	57.00	57.00	58.26	57.48	56.13	56.29	**59.36**
	40%	57.00	59.65	61.28	60.45	58.97	60.17	**62.68**
	50%	57.00	62.03	64.15	63.20	61.12	62.09	**65.37**

5.3.2 Classification Performance Evaluation with Different Percentages of Labeled Training Samples

To further analysis the classification performances on the four databases, we increase the percentage of labeled training samples from 10% to 50% with the step size of 10%, and list the results in **Table 3**. It can be observed that SM^{2}DIS performs better than compared methods consistently, which indicates that the intact feature representations learned by SM^{2}DIS are with better discriminative structure and manifold structure.

5.3.3 Classification Performance Evaluation with Different Values of SNR

SNR is an important factor to measure the robustness of the learning method. **Fig. 1 (a) – (d)** show the average classification results on MNIST, COIL-20, Multi-PIE, and Caltech-101 databases with respect to different SNR values. We can observe that the classification accuracy of SM^{2}DIS drops more slowly than constract methods when the proportion of noise information in images becomes much more larger, i.e., the value of SNR is becoming much more smaller. Thus, our proposed SM^{2}DIS can get a better classification performance to deal with view-insufficiency problem.

5.3.4 Evaluation of the Influence of Parameters

This subsection focuses on evaluating the impact of important parameters in SM^{2}DIS (non-negative coefficient α and c, the number of nearest neighbors K, and the dimension d of the intact manifold discriminant space) on the classification performance. Taking the MNIST database as an example, **Fig. 2 (a) – (d)** illustrate the average classification accuracies with the change of parameter values. For non-negative coefficient parameter α, we observe

the performance variations of our proposed approach in the change interval of [0.1 1] with the step size of 1; for non-negative coefficient parameter c, we observe the performance variations of our proposed approach in the change interval of [0.01 0.1] with the step size of 0.01; for the number of nearest neighbors parameter K, we observe the performance variations of our approach in the change interval of [1 39] with the step size of 1; for the dimension of the intact manifold discriminant space parameter d, we observe the performance variations of our approach in the change interval of [10 300] with the step size of 10. From **Fig. 2 (a) – (d)**, SM^{2}DIS can reach a stable and best state when α is in the interval of [0.3 0.5], c is in the interval of [0.05 0.07], K is in the interval of [35 37], and d is in the interval of [160 220]. For convenience, we set $\alpha=0.4$, $c=0.06$, $K=37$, $d=190$ on MNIST database.

(a) MNIST.   (b) COIL-20.

(c) Multi-PIE.   (d) Caltech-101.
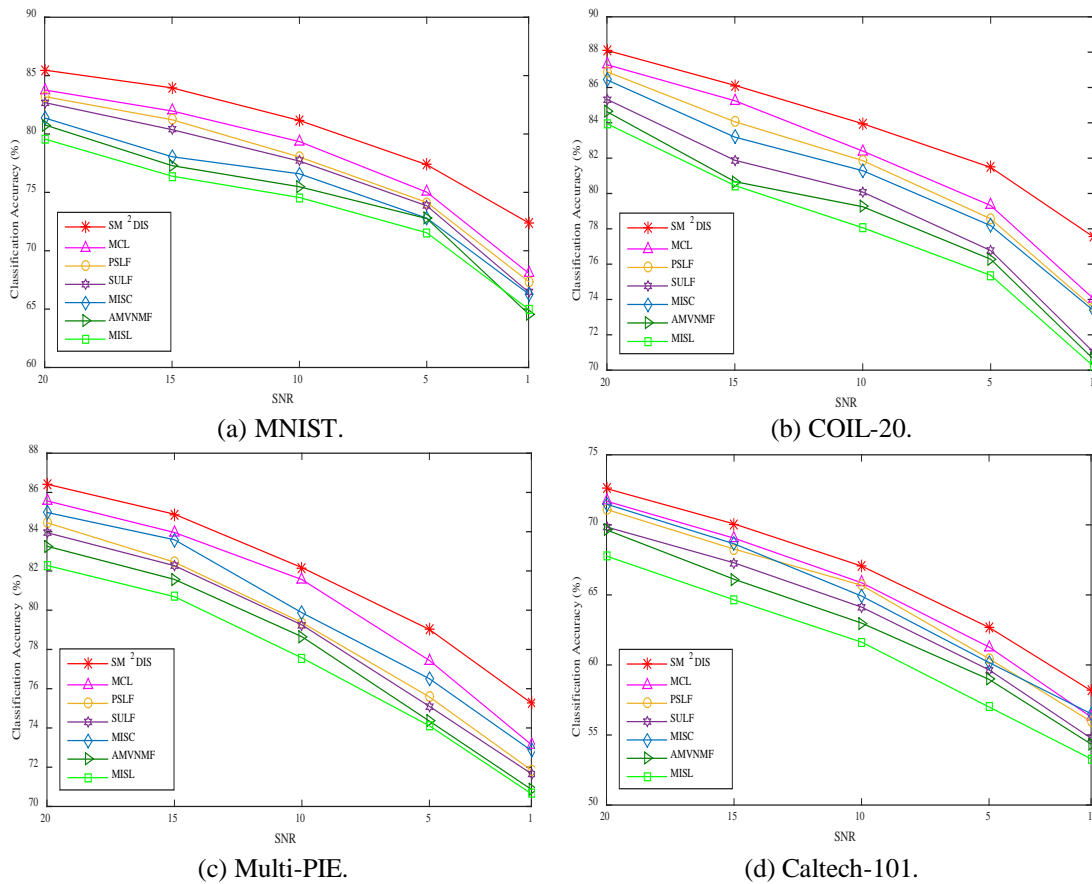
Fig. 1. Classification results on the four databases with different SNR values.

5.3.5 Evaluation of Convergence

In this subsection, we conduct experiment to observe the convergency of our proposed SM^{2}DIS approach. Taking the MNIST database as an example, **Fig. 3.** shows the change of objective function value in formula (9) versus the increase of iteration times. It is observed that the value of the objective function decreases fast as the iteration number increases, which indicates the effectiveness of our approach.
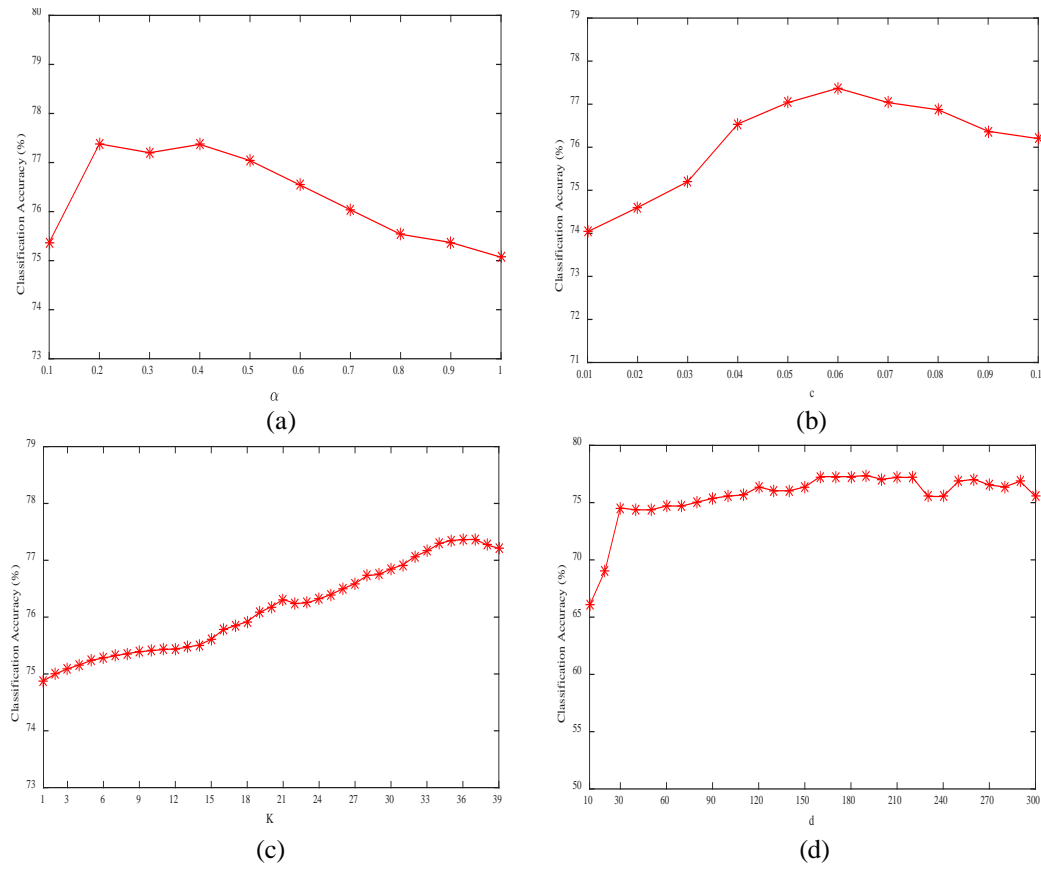
Fig. 2. (a) parameter α; (b) parameter c; (c) parameter K; (d) parameter d.
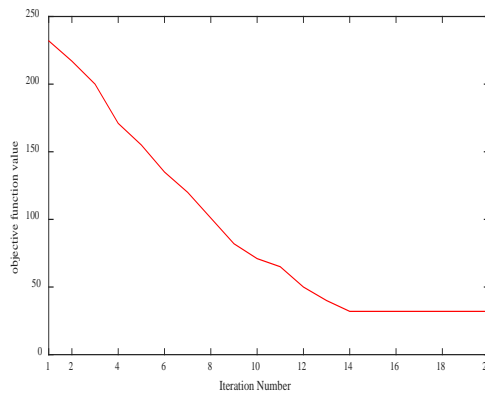
Fig. 3. Convergence curve on MNIST database.

5.3.6 Evaluation of Computation Cost

In this subsection, we evaluate the average training time of SM^{2}DIS and compared methods on MNIST, COIL-20, Multi-PIE, and Caltech-101 databases. **Table 4** lists the corresponding results. The hardware configuration in the experiment is 2.4 GHz CPU, 4GB memory. It can be observed from **Table 4** that the computation cost of our SM^{2}DIS is comparable with that of MCL method, and the computation cost in Caltech-101 database is larger than that in other databases as the size of Caltech-101 database is relatively larger.

Table 4. Average training time (s) between SM^{2}DIS and compared methods on four databases

Method	MNIST	COIL-20	Multi-PIE	Caltech-101
MISL[13]	18.27	26.45	31.86	81.95
SULF [25]	23.22	35.48	46.59	107.80
MCL [26]	38.06	49.77	54.13	116.40
PSLF [27]	32.86	41.15	48.91	112.05
AMVNMF[28]	39.87	40.28	42.43	102.26
MISC [18]	34.75	39.17	44.26	75.99
SM^{2}DIS	41.00	48.21	53.24	123.09

Conclusion

In this paper, we propose a new semi-supervised multi-view intact space learning approach named SM^{2}DIS for discovering manifold and discriminative intact feature representations for data of multiple views. SM^{2}DIS can effectively exploit the discriminant information from labeled intact feature representations by employing the complementary information of multi-view data. Besides, SM^{2}DIS exploits the manifold structure from both labeled and unlabeled intact feature representations. Experiments on MNIST, COIL-20, Multi-PIE, and Caltech-101 databases have shown the effectiveness and robustness of our approach.

For the future work, we will evaluate the effectiveness and robustness of our approach on more databases, and extend SM^{2}DIS to its kernel version to further enhance the classification performance.

References

[1] M. Yang, and S. L. Sun, "Multi-view uncorrelated linear discriminant analysis with applications to handwritten digit recognition," in *Proc. of the International Joint Conference on Neural Networks*, pp. 4175-4181, July 6-11, 2014. Article (CrossRef Link)

[2] S. L. Sun, X. J. Xie, and M. Yang, "Multi-view uncorrelated discriminant analysis," *IEEE Transactions on Cybernetics*, vol. 46, no. 12, pp. 3272-3284, 2016. Article (CrossRef Link)

[3] X. Y. Jing, Q. Liu, F. Wu, B. Xu, and Y. Zhu, "Web page classification based on uncorrelated semi-supervised intra-view and inter-view manifold discriminant feature extraction," in *Proc. of the International Conference on Artificial Intelligence*, pp. 2255-2261, July 25-31, 2015.

[4] M. Liu, Y. Luo, D. C. Tao, C. Xu, and Y. G. Wen, "Low-rank multi-view learning in matrix completion for multi-label image classification," in *Proc. of the AAAI Conference on Artificial Intelligence*, pp. 2778-2784, January 25-30, 2015.

[5] Z. Y. He, C. Chen, J. J. Bu, P. Li, and D. Cai, "Multi-view based multi-label propagation for image annotation," *Neurocomputing*, vol. 168, pp. 853-860, 2015. Article (CrossRef Link)

[6] C. Deng, Z. T. Lv, W. Liu, J. Z. Huang, D. C. Tao, and X. B. Gao, "Multi-View Matrix Decomposition: A New Scheme for Exploring Discriminative Information," in *Proc. of the International Conference on Artificial Intelligence*, pp. 3438-3444, July 25-31, 2015.

[7] W. Y. Chang, C. P. Wei, and Y. C. F. Wang, "Multi-view nonnegative matrix factorization for clothing image characterization," in *Proc. of the International Conference on Pattern Recognition*, pp. 1272-1277, Augest 24-28, 2014. Article (CrossRef Link)

[8] Sun S, "A survey of multi-view machine learning," *Neural Computing and Applications*, vol. 23, no. 7, pp. 2031-2038, 2013. Article (CrossRef Link)

[9] Guo Y, "Convex subspace representation learning from multi-view data," in *Proc. of the AAAI Conference on Artificial Intelligence*, pp. 387-393, July 14-18, 2013.

[10] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4333-4341, July 21-26, 2017. Article (CrossRef Link)

[11] R. Nishant, N. Sumit, C. Santanu, and D. Om, "Partial Multi-view clustering using graph regularized NMF," in *Proc. Of the International Conference on Pattern Recognition*, pp. 2192-2197, December 4-8, 2016. Article (CrossRef Link)

[12] Z. Xue, G. R. Li, S. H. Wang, W. G. Zhang, and Q. M. Huang, "Bi-Level multi-view latent space learning," *IEEE Transactions on Circuits and Systems for Video Technology*, In Press, pp. 1-14, 2016. Article (CrossRef Link)

[13] C. Xu, D. C. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Transsactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 12, pp. 2531-2544, 2015. Article (CrossRef Link)

[14] M. N. Kan, S. G. Shan, H. H. Zhang, S. H. Lao, and X. L. Chen, "Multi-view discriminant analysis," in *Proc. of the European Conference on Computer Vision*, pp. 808-821, October 7-13, 2012.

[15] M. N. Kan, S. G. Shan, H. H. Zhang, S. H. Lao, and X. L. Chen, "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 188-194, 2016. Article (CrossRef Link)

[16] P. Yang, H. Davulcu, Y. D. Zhu, and J. R. He, "A generalized hierarchical multi-latent space model for heterogeneous learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3154-3168, 2016. Article (CrossRef Link)

[17] J. Li, C. Xu, W. K. Yang, C. Y. Sun, and D. C. Tao, "Discriminative multi-view interactive image re-ranking," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3113-3127, 2017. Article (CrossRef Link)

[18] Q. J. Wang, H. Y. Lv, J. Yue, and E. Mitchell, "Supervised multiview learning based on simultaneous learning of multiview intact and single-view classifier," *Neural Computing & Applications*, vol. 28, no. 8, pp. 2293-2301, 2017. Article (CrossRef Link)

[19] T. Diethe, D.R. Hardoon, and J. Shawe-Taylor, "Multi-view fisher discriminant analysis," in *Proc. of NIPS Workshop on Learning from Multiple Sources*, pp. 1-8, December 8-11, 2008.

[20] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: a discriminative latent space," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2160-2167, June 16-21, 2012. Article (CrossRef Link)

[21] Y. W. Guo, X. Q. Ding, and J. H. Xue, "MiLDA: a graph embedding approach to multi-view face recognition," *Neurocomputing*, vol. 151, no. 3, pp. 1255-1261, 2015. Article (CrossRef Link)

[22] X. Y. Jing, F. Wu, X. W. Dong, S. G. Shan, and S. C. Chen, "Semi-supervised multi-view correlation feature learning with application to webpage classification," in *Proc. of the AAAI Conference on Artificial Intelligence*, pp. 1374-1381, February 4-9, 2017.

[23] H. Tao, C. P. Hou, F. P. Nie, J. B. Zhu, and D. Y. Yi, "Scalable multi-view semi-supervised classification via adaptive regression," *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4283-4296, 2017. Article (CrossRef Link)

[24] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbors," in *Proc. of the AAAI Conference on Artificial Intelligence*, pp. 2408-2414, February 4-9, 2017.

[25] Y. Jiang, J. Liu, Z. Li, and H. Lu, "Semi-supervised unified latent factor learning with multi-view data," *Machine Vision and Applications*, vol. 25, no. 7, pp. 1635-1645, 2014. Article (CrossRef Link)

[26] Z. Y. Guan, L. J. Zhang, J. Y. Peng, and J. P. Fan, "Multi-view concept learning for data representation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27 no. 11, pp. 3016-3028, 2015. Article (CrossRef Link)

[27] J. Liu, Y. Jiang, Z. C. Li, Z. H. Zhou, and H. Q. Lu, "Partially shared latent factor learning with multiview data," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 6, pp. 1233-1246, 2015. Article (CrossRef Link)

[28] J. Wang, X. Wang, F. Tian, C. H. Liu, H. Yu, and Y. Liu, "Adaptive multi-view semi-supervised nonnegative matrix factorization," in *Proc. of the International Conference on Neural Information Processing*, pp. 435-444, October 16-21, 2016. Article (CrossRef Link)

[29] D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *Proc. of the International Conference on Artificial Intelligence*, pp. 714-719, January 6-12, 2007.

[30] I. Mizera, and C. H. Muller, "Breakdown points of Cauchy regresion-scale estimators," *Statistics & Probability Letters*, vol. 57, no. 1, pp. 79-82, 2002. Article (CrossRef Link)

[31] C. J. Lin, "Projected gradient methods for nonnegative matrix factorization," *Neural Computation*, vol. 19, no. 10, pp. 2756-2779, 2007. Article (CrossRef Link)

[32] R. P. Wang, and X. L. Chen, "Manifold discriminant analysis," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 429-436, June 20-25, 2009. Article (CrossRef Link)

[33] Y. M. Chen, and J. H. Chiang, "Face recognition using combined multiple feature extraction based on fourier-mellin approach for single example image per person," *Pattern Recognition Letters*, vol. 31, no. 13, pp. 1833-1841 , 2010. Article (CrossRef Link)

[34] L. Han, X. Y. Jing, and F. Wu, "Multi-view local discriminantion and canonical correlation analysis for image classification," *Neurocomputing*, vol. 275C, pp. 1087-1098 , 2018. Article (CrossRef Link)

[35] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324 , 1998. Article (CrossRef Link)

[36] H. Murase, and S. K. Nayar, "Visual learning and recognition of 3-D objects from appearance," *International Journal of Computer Vision*, vol. 14, no. 1, pp. 5-24, 1995. Article (CrossRef Link)

[37] D. Cai, X. He, J. Han, and H. J. Zhang, "Orthogonal laplacian faces for face recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3608-3614, 2006. Article (CrossRef Link)

[38] F. F. Li, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 594-611, 2006. Article (CrossRef Link)

[39] K. Fukunaga, and W. L. Koontz, "Application of the karhunen-loeve expansion to feature selection and ordering," *IEEE Transactions on Computers*, vol. 19, no. 4, pp. 311-318, 1970. Article (CrossRef Link)

[40] S.E. Grigorescu, N. Petkov, and P. Kruizinga, "Comparison of texture features based on Gabor filters," *IEEE Transactions on Image Processing*, vol. 11, no. 10, pp. 1160-1167, 2002. Article (CrossRef Link)

[41] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037-2041, 2006. Article (CrossRef Link)

[42] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *Proc. of the International Conference on Computer Vision*, pp. 1-8, October 14-21, 2007. Article (CrossRef Link)

[43] H. Zhang, A. Berg, M. Maire, and J. Malik, "SVM-KNN: discriminative nearest-neighbor classification for visual category recognition," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2126-2136, June 17-22, 2006. Article (CrossRef Link)

[44] E. Shechtman, and M. Irani, "Matching local self-similarities across image and videos," in *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-8, June 17-22, 2007. Article (CrossRef Link)

Lu Han is a Ph.D Candidate in the College of Automation, Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China. Her research interests include pattern recognition, artificial intelligence, and software engineering.

Fei Wu received the Ph.D. degree from Nanjing University of Posts and Telecommunications (NUPT), Nanjing, China, in 2016. He is currently a lecturer with the College of Automation, NUPT. He is an author of more than twenty scientific papers. His research interests include pattern recognition, artificial intelligence, and software engineering.
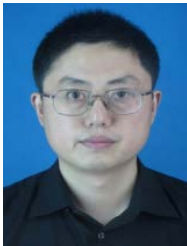
Xiao-Yuan Jing received the Doctoral degree of Pattern Recognition and Artificial Intelligence in the Nanjing University of Science and Technology, 1998. He was a Professor with the Department of Computer, Shenzhen Research Student School, Harbin Institute of Technology, 2005, and a Research Fellow with the Department of Computing, Hong Kong Polytechnic University. Now he is a Professor with the State Key Laboratory of Software Engineering, School of Computer, Wuhan University, and the College of Automation, Nanjing University of Posts and Telecommunications, China. His research interests include pattern recognition, machine learning, software engineering and artificial intelligence. He has published over 70 papers in the international conferences and journals like CVPR, IJCAI, AAAI, ICSE, FSE, ICME, IEEE TIP, TCSVT, TSMC-B, PR and so on.