electronics

Article

Multi-View Projection Learning via Adaptive Graph Embedding for Dimensionality Reduction

Haohao Li [1], Mingliang Gao [2], Huibing Wang [3] and Gwanggil Jeon [1,4,*]

[1] Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou 310018, China; hhl820@zstu.edu.cn
[2] School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255000, China; mlgao@sdut.edu.cn
[3] College of Information and Science Technology, Dalian Maritime University, Dalian 116021, China; huibing.wang@dlmu.edu.cn
[4] Department of Embedded Systems Engineering, Incheon National University, Incheon 22012, Republic of Korea
* Correspondence: gjeon@inu.ac.kr

Abstract: In order to explore complex structures and relationships hidden in data, plenty of graph-based dimensionality reduction methods have been widely investigated and extended to the multi-view learning field. For multi-view dimensionality reduction, the key point is extracting the complementary and compatible multi-view information to analyze the complex underlying structure of the samples, which is still a challenging task. We propose a novel multi-view dimensionality reduction algorithm that integrates underlying structure learning and dimensionality reduction for each view into one framework. Because the prespecified graph derived from original noisy high-dimensional data is usually low-quality, the subspace constructed based on such a graph is also low-quality. To obtain the optimal graph for dimensionality reduction, we propose a framework that learns the affinity based on the low-dimensional representation of all views and performs the dimensionality reduction based on it jointly. Although original data is noisy, the local structure information of them is also valuable. Therefore, in the graph learning process, we also introduce the information of predefined graphs based on each view feature into the optimal graph. Moreover, assigning the weight to each view based on its importance is essential in multi-view learning, the proposed GoMPL automatically allocates an appropriate weight to each view in the graph learning process. The obtained optimal graph is then adopted to learn the projection matrix for each individual view by graph embedding. We provide an effective alternate update method for learning the optimal graph and optimal subspace jointly for each view. We conduct many experiments on various benchmark datasets to evaluate the effectiveness of the proposed method.

Keywords: multi-view learning; dimensionality reduction; graph learning; self-weighted learning

Citation: Li, H.; Gao, M.; Wang, H.; Jeon, G. Multi-View Projection Learning via Adaptive Graph Embedding for Dimensionality Reduction. *Electronics* **2023**, *12*, 2934. https://doi.org/10.3390/electronics12132934

Academic Editors: Mohit Mittal, Rocío Pérez de Prado and Valentina E. Balas

Received: 25 May 2023
Revised: 19 June 2023
Accepted: 25 June 2023
Published: 3 July 2023

1. Introduction

In real-world applications, samples can usually be collected by diverse data collection sources or various feature extraction methods, which are often represented with multiple views. Taking images for example, one color image can be represented by multiple types of descriptors [1–4], such as Gist [5], histogram of oriented gradient (HoG) [6], local binary patterns (LBP) [7], SIFT [8], etc. [9–11]. Since features from different views may characterize different specific information of one sample, complementarity among different views can be explored for improving the performances of single-view learning algorithms.

In many fields, most constructed features are usually high-dimensional, while the underlying structure can be described by a small number of parameters in most situations. Direct manipulations on these features are time-consuming and computationally expensive. Dimensionality reduction (DR), therefore, has become a basic preprocessing technique to deal with such data in most problems. In the past few decades, plenty of dimensionality reduction algorithms were proposed to seek the optimal subspace for original high-dimensional data based on different principles. One class of dimensionality reduction algorithms is based on manifold embedding, which seeks to construct the low-dimensional representation that maintains the graph affinity between the samples as much as possible. The graph embedding-based methods can be divided into linear methods and nonlinear methods. Linear methods aim to seek the appropriate projection matrix to project the high-dimensional data onto an optimal low-dimensional subspace. One of the most famous linear algorithms is principal component analysis (PCA) [12], which maximizes the global variance of data to obtain the low-dimensional subspace. Linear discriminant analysis (LDA) [13] is a supervised learning method that seeks to construct the projection matrix that can maximize the separation between the classes. In addition, there are also many representative and classical linear dimensionality reduction methods such as locality preserving projections (LPP) [14], neighborhood preserving embedding (NPE) [15], marginal Fisher analysis (MFA) [16], etc. [15,17,18].

Different from linear dimensionality reduction methods, nonlinear methods agree with one famous hypothesis that the observed high-dimensional data is actually mapped from a low-dimensional submanifold. There are many representative nonlinear dimensionality reduction algorithms such as locally linear embedding (LLE) [19], Isomap [20], Laplacian eigenmaps (LE) [21], etc. [22,23], that have been well studied. Moreover, these traditional DR methods construct a graph in advance and then find an optimal subspace that can preserve such a graph as much as possible. The graph construction and dimensionality reduction processes are separated, which leads these methods to seek the subspace utilizing a suboptimal graph. To address this issue, some algorithms such as graph-optimized locality preserving projections (GoLPP) [24], dimensionality reduction with adaptive graph (DRAG) [25] and joint graph optimization and projection learning (JGOPL) [26] are proposed to introduce the graph optimization into dimensionality reduction procedure. These methods aim to simultaneously seek an optimal graph and subspace in one objective function.

Although plenty of dimensionality reduction methods have good performance in dealing with high-dimensional data, most of them fail to extend to the multi-view setting directly. As they cannot effectively explore the inherent relation among different views features. In the past decade, multi-view learning has been well-developed in various fields [27–31]. Canonical correlation analysis (CCA) [32] and its multi-view version multi-view canonical correlation analysis (MCCA) [33] are famous algorithms that are widely adopted as a regularization term for multi-view learning. Distributed spectral embedding (DSE) [34] aims to construct one common low-dimensionality embedding based on the smooth principle. However, since the original multi-view data are invisible to the final learning process, it cannot well explore the complementary nature of different views. To overcome this problem, Xia et al. propose a nonlinear dimensionality reduction algorithm for multi-view data termed multi-view spectral embedding (MSE) [35], which effectively explores the complementary and compatible information from different views to construct one common low-dimensional embedding for all views. Kan et al. [36] extend LDA to a multi-view setting and propose multi-view discriminant analysis (MvDA) to project multi-view data to a common discriminative space. Ding et al. [37] propose a low-rank common subspace (LRCS) to seek one common linear subspace with low-rank constraint for each view based on a compatible principle that aims to reduce the semantic gap between different views. However, most multi-view dimensionality reduction algorithms construct the graph by original high-dimensional data, which is independent of subspace learning. Therefore, such DR results of these algorithms are sensitive to the graph construction. If the predefined graph is of low quality, the quality of the results of dimensionality reduction may also be low.

To deal with these issues, in this paper, we incorporate graph optimization and low-dimensional subspace learning for multi-view data into one common framework to propose graph optimization multi-view projections learning (GoMPL) for dimensionality reduction. Since features from different views are exploited to describe the same sample, they usually admit the same underlying similarity structure. Based on this hypothesis, GoMPL aims to learn one intrinsic graph structure of samples and seek the subspace by preserving such a graph for each view simultaneously. In the whole learning procedure, the common graph is allowed to be adaptively adjusted based on low-dimensional representations of each view. Moreover, the information contained in original high-dimensional data is also important [25]. Therefore, we further regularize the target similarity graph as a centroid of the prespecified graph of each view, which introduce information on original data into the optimal graph. Specifically, the learned optimal graph is also employed to integrate the information from multi-view data, which avoids co-regularizing all the views to a common subspace [29]. Assigning an appropriate weight to each view based on some principles is essential in multi-view learning, so our proposed GoMPL provides a self-weighted scheme to automatically learn the weights in the graph learning process, which releases from predefining hyperparameters experientially. We provide an effective updating algorithm to solve the proposed GoMPL. Plenty of experiments on the various datasets evaluate the effectiveness of our proposed GoMPL. We summarize the contributions of our work as:

1. We propose a novel multi-view dimensionality algorithm called GoMPL which can seek one common underlying manifold structure for samples described by multi-view features and the appropriate subspace for each view simultaneously.
2. We adopt one optimal graph to learn the projections for all views. This graph can be further optimized based on both the low-dimensional representation of each view and the affinity of original high-dimensional data. Therefore, GoMPL can integrate the multi-view information by the common graph rather than co-regularize the low-dimensional representation of each view.
3. Since the information from the original affinity of each view is also important, GoMPL regularizes the target similarity graph as a centroid of the prespecified graph of each view. Moreover, different views may take different contributions to understand the underlying manifold structure; GoMPL adaptively allocates each view an appropriate weight without predefining hyper-parameters.

2. Methods

In this section, we first review one famous multi-view learning framework termed co-regularized multi-view spectral clustering (co-reg). Then we introduce a single-view DR method named graph optimization for dimensionality reduction with sparsity constraints (GODRSC) [38].

2.1. Co-Regularized Multi-View Spectral Clustering

Co-regularized multi-view spectral clustering (Co-reg) [27] is a famous multi-view clustering algorithm that extends traditional single-view spectral clustering into the multi-view setting. Co-reg incorporates the multi-view information to perform the spectral clustering based on the hypothesis that the clustering results of different views should be consistent. After giving the graph Laplacian based on each view feature, co-reg performs the spectral embedding for them based on this graph and the proposed co-regularization term, which is employed to regularize the embedding results of them based on clustering hypotheses. Given multi-view data set with N samples and m views, i.e., a set of matrices $\left\{X^{(v)} \in \mathbb{R}^{D^v \times N}\right\}_{v=1}^{m}$, the objective of co-reg can be described as follows:

$$\begin{aligned} &\max_{U^{(1)},U^{(2)},\dots,U^{(m)},U^{(*)}\in\mathbb{R}^{N\times K}} \sum_{v=1}^{m} Tr\left\{\left(U^{(v)}\right)^{T} L^{(v)} U^{(v)}\right\} - \sum_{v} \alpha D(U^{(v)},U^{(*)}) \\ &s.t. \quad \left(U^{(v)}\right)^{T} U^{(v)} = I, \quad \left(U^{(*)}\right)^{T} U^{(*)} = I \quad v = 1,2,\dots,m, \end{aligned} \tag{1}$$

where $L^{(v)}$ is the normalized graph Laplacian based on the similarity or kernel matrix of $X^{(v)}$, $\alpha > 0$ trades-off the embedding agreement and single-view spectral embedding term and $D(U,V) = -Tr\{UU^TVV^T\}$ is the co-regularization term which indicates disagreement between the clusterings of U and V. Therefore, by maximizing Equation (1), co-reg can make the embedding results towards the common consensus $U^{(*)}$ (centroid-based co-regularization).

2.2. Graph Optimization for Dimensionality Reduction with Sparsity Constraints

Graph optimization for dimensionality reduction with sparsity constraints (GODRSC) is a graph-optimized-based dimensionality reduction algorithm that has attracted wide attention recently. GODRSC integrates the dimensionality reduction process and sparse graph construction into one common objective, which can seek an optimal sparse reconstruction relation and subspace to preserve such relation simultaneously. For given $X = [x_1, x_2, ..., x_N] \in \mathbb{R}^{D\times N}, x_i \in \mathbb{R}^D$, GODRSC can obtain the changeable sparse reconstructive weights and the projection matrix simultaneously by optimizing the following objective function:

$$\begin{aligned} \min_{P,s_i} \quad & \frac{\sum_{i=1}^{N}\|P(x_i - Xs_i)\|_2^2}{\sum_{i=1}^{N}\|Px_i\|_2^2} + \sum \lambda_i\|s_i\|_1 \\ s.t. \quad & PP^T = I, \end{aligned} \tag{2}$$

where s_i is the reconstruction coefficient of ith sample, and P is the projection matrix. By optimizing Equation (2) in an alternating iteration scheme, GODRSC can learn the optimal sparse reconstruction relation and the low-dimensional subspace simultaneously.

3. Graph Optimization Multi-View Projections Learning

In this section, we first formulate the GoMPL in Section 3.1 in detail, which aims to learn one intrinsic graph structure and find the subspace preserving such a graph for each view simultaneously. In Section 3.2, we introduce the details of the optimization process of the proposed method, which employs the alternating iterative scheme to optimize the objective function of the proposed GoMPL. We also provide the convergence analysis of the proposed algorithm in Section 3.3.

3.1. The Proposed Method

For the given N samples with m views feature $\mathcal{X} = \{X^{(v)} \in \mathbb{R}^{D^{(v)}\times N}\}_{v=1}^{m}$, where $D^{(v)}$ is the dimensionalities of feature from vth view. GoMPL aims to fully extract the information from multi-view features to obtain the common underlying structure of samples. Although features from different views usually locate in different spaces, they are utilized to describe the same sample. Therefore, the affinity relation of these features is approximately equal. Based on this hypothesis, we exploit one common target graph to seek the optimal subspace for each view:

$$\begin{aligned} \min_{\mathcal{P}} & \sum_{v=1}^{m}\sum_{i,j=1}^{N}\left\|P^{(v)}x_i^{(v)} - P^{(v)}x_j^{(v)}\right\|^2 s_{ij} \\ s.t. \quad & P^{(v)}X^{(v)}\left(X^{(v)}\right)^T\left(P^{(v)}\right)^T = I \quad v = 1,2,...,m \end{aligned} \tag{3}$$

where $\mathcal{P} = \{P^{(1)}, P^{(2)}, ..., P^{(m)}\}$ and $P^{(v)} \in \mathbb{R}^{d\times D^{(v)}}$ is learned projection matrix to project vth view features into subspace, and s_{ij} describes the affinity between ith and jth samples. For the graph-embedding-based algorithm, one of the crucial objectives is designing a sufficiently smooth affinity on the data manifold. Therefore, it is essential to construct the common affinity graph $\boldsymbol{S}$ by integrating the multi-view information. To address this issue, a naive way is to take the average of the graph from each view feature as the common affinity graph. However, this graph $\boldsymbol{S}$ only considers predefined low-quality graphs, and the results of dimensionality reduction for each view may also be low quality. In order to perform dimensionality reduction based on the optimal graph, we combine the graph embedding for dimensionality reduction with graph optimization into unified frameworks to seek the optimal projections and the common graph jointly:

$$\begin{aligned}
&\min_{\mathcal{P},S} \sum_{v=1}^{m} \sum_{i,j=1}^{N} \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2 s_{ij} \\
&s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0, \\
&\qquad P^{(v)} X^{(v)} \left(X^{(v)} \right)^T \left(P^{(v)} \right)^T = I \quad v = 1,2,\ldots,m
\end{aligned} \tag{4}$$

Intuitively , this naive way completely depends on the transformed data to learn the common similarity graph, which may result in unstable dimensionality reduction performances. Thus, inspired by Dimensionality reduction with adaptive graph (DRAG) [25], we employ the information of original multi-view data to constrain the target graph. For the given predefined graph affinity $A^{(v)}$ based on X^v [14], the target common graph S should approximate each of them. With this constraint, we can obtain the final objective:

$$\begin{aligned}
&\min_{\mathcal{P},S} \sum_{v=1}^{m} \sum_{i,j=1}^{N} \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2 s_{ij} + \lambda \sum_{v=1}^{m} \left\| S - A^{(v)} \right\|_F \\
&s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0, \\
&\qquad P^{(v)} X^{(v)} \left(X^{(v)} \right)^T \left(P^{(v)} \right)^T = I \quad v = 1,2,\ldots,m
\end{aligned} \tag{5}$$

where $\lambda > 0$ is a hyperparameter that balances the trade-off. Although there are no weighting factors defined in the graph co-regularization term of Equation (5), this term can be reconstructed as a weighted average reconstruction error of different views by a self-weighted scheme in the procedure of optimization. These weights have a desired property that if the predefined graph $A^{(v)}$ approximates the target graph S the corresponding weight for view v is large, and vice versa. This indicates that these weights are meaningful for measuring the importance of each built similarity graph. Moreover, GoMPL can utilize original data information by graph co-regularization term. Therefore, the learned target graph S well reflects the common underlying similarity structure, which is crucial for dimensionality reduction.

3.2. Optimization

In order to optimize Equation (5), we employ an alternating iterative optimization scheme, which optimizes one variable with the others fixed. We first initialize $S = \frac{1}{m} \sum_{v=1}^{m} A^{(v)}$, and then the alternating iterative optimization process is as:

$\mathcal{P}$ problem : With S being fixed, we update $\mathcal{P}$ by solving:

$$\begin{aligned}
&\min_{\mathcal{P}} \sum_{v=1}^{m} \sum_{i,j=1}^{N} \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2 s_{ij} \\
&s.t. \quad P^{(v)} X^{(v)} \left(X^{(v)} \right)^T \left(P^{(v)} \right)^T = I \quad v = 1,2,\ldots,m
\end{aligned} \tag{6}$$

We can find that Equation (6) is independent for each individual $P^{(v)}$, so we can optimize it separately for each view. For vth view, the problem in Equation (6) can be reformulated by simple algebra as:

$$\begin{aligned} \left(P^{(v)}\right)^* &= \arg\min_{P^{(v)}} \sum_{v=1}^{m} \sum_{i,j=1}^{N} \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2 s_{ij} \\ &= \arg\min_{P^{(v)}} Tr\left(P^{(v)} X^{(v)} (\tilde{D} - \tilde{S})) \left(X^{(v)}\right)^T \left(P^{(v)}\right)^T \right) \\ & s.t. \quad P^{(v)} X^{(v)} \left(X^{(v)}\right)^T \left(P^{(v)}\right)^T = I \quad v = 1, 2, \ldots, m \end{aligned} \tag{7}$$

where $\tilde{D} = diag\{\tilde{d}_1, \tilde{d}_2, \ldots, \tilde{d}_N\}$ is a diag matrix with $\tilde{D}_i = \sum_{j=1}^{N} \tilde{s}_{ij}$ and $\tilde{s}_{ij}$ is the ijth element of matrix $\tilde{S} = S + S^T$. This problem is a classical generalized eigenproblem, and the optimal solution can be obtained as:

$$X^{(v)} (\tilde{D} - \tilde{S})) \left(X^{(v)}\right)^T \left(p^{(v)}\right)^T = \lambda X^{(v)} \left(X^{(v)}\right)^T \left(p^{(v)}\right)^T. \tag{8}$$

S problem: When $\mathcal{P}$ are fixed, we can formulate the subproblem of Equation (5) with respect to S as:

$$\begin{aligned} & \min_S \sum_{v=1}^{m} \sum_{i,j=1}^{N} \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2 s_{ij} + \lambda \sum_{v=1}^{m} \left\| S - A^{(v)} \right\|_F \\ & s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0. \end{aligned} \tag{9}$$

To solve the optimal graph S, we first simplify Equation (9). By setting $u_{ij}^{(v)} = \left\| P^{(v)} x_i^{(v)} - P^{(v)} x_j^{(v)} \right\|^2$ and $U^{(v)}$ as a matrix with ijth element being $u_{ij}^{(v)}$, the Equation (9) can be further reformulated as:

$$\begin{aligned} & \min_S Tr(US) + \lambda \sum_{v=1}^{m} \left\| S - A^{(v)} \right\|_F \\ & s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0, \end{aligned} \tag{10}$$

where $U = \sum_{v=1}^{m} U^{(v)}$. We exploit the Lagrange multiplier algorithm to solve the problem Equation (10). The Lagrange is formulated as:

$$\min_S Tr(US) + \lambda \sum_{v=1}^{m} \left\| S - A^{(v)} \right\|_F + \mathcal{C}(\Lambda, S) \tag{11}$$

where $\mathcal{C}(\Lambda, S)$ serves as a proxy for the constraints to S, and Λ is the Lagrange multipliers. We first take the derivative of the objective function in Equation (11) with respect to S and set it to 0, and then we can obtain an equation about S as:

$$U^T + \sum_{v=1}^{m} w^{(v)} \nabla_S \left(\left\| S - A^{(v)} \right\|_F^2 \right) + \nabla_S \mathcal{C}(\Lambda, S) = 0 \tag{12}$$

where the form of $w^{(v)}$ is:

$$w^{(v)} = \frac{1}{2\left\| S - A^{(v)} \right\|_F} \tag{13}$$

We can see from Equation (13) that $w^{(v)}$ is dependent on the graph S. In order to simplify this optimal problem, we set $w^{(v)}$ stationary. Then the solution of Equation (12) can be reformulated as:

$$\begin{aligned} &\min_{S} Tr(US) + \lambda \sum_{v=1}^{m} w^{(v)} \left\| S - A^{(v)} \right\|_F^2 \\ &s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0, \end{aligned} \tag{14}$$

which can be solved simpler. Then, the optimal S obtained from Equation (14) is further exploited to calculate the weights factor $w^{(v)}$ based on Equation (13). Therefore, we solve the original S-problem Equation (10) by alternately optimizing S and updating $w^{(v)}$ iteratively. If this process converges, we can know from Equation (11) that S will finally converge to the KKT condition of Equation (10). In practice, we usually set $w^{(v)} = \frac{1}{2\sqrt{\left\|S-A^{(v)}\right\|_F^2+\delta}}$ to avoid dividing by zero, where δ is a very small value.

Obviously, problem Equation (14) is independent between different columns of S. Thus, we can reformulate Equation (14) as the following problem for each column:

$$\begin{aligned} &\min_{s_i} \sum_{v=1}^{m} u_i^T s_i + \lambda \sum_{v=1}^{m} w^{(v)} \left\| s_i - a_i^{(v)} \right\|_2^2 \\ &s.t. \quad \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0. \end{aligned} \tag{15}$$

This problem Equation (15) is a quadratic programming (QP) problem, which can be effectively solved by the CVXOPT software package.

After $S, P^{(1)}, P^{(2)}, ..., P^{(m)}$ being solved, low-dimensional representations of features from vth view can be obtained as:

$$Y^{(v)} = P^{(v)} X^{(v)} \in \mathbb{R}^{d \times N} \tag{16}$$

The self-weighted scheme of our method is as follows: in S problem, we can find that when the iterative converges, the form of Equation (14) can be considered as the weighted sum of the distance between the learned graph S and predefined graph $A^{(v)}$ of each view. Moreover, according to Equations (13) and (14), when the distance between $A^{(v)}$ and S is large, the obtained weight factor of vth view $w^{(v)}$ will be small, which means that the information contained vth view feature is less important.

3.3. Convergence

We provide the convergence analysis of the GoMPL in this section. According to the alternating process summarized in Algorithm 1, the optimization procedure is divided as $\mathcal{P}$ problem in Equation (7) and S problem in Equation (10). For $\mathcal{P}$ problem, we can obtain the closed optimal projections of each view directly. Therefore, after updating $\mathcal{P}$ in each iteration, we can get the objective function value to decrease. So, we focus on analyzing the convergence of S-problem.

Algorithm 1: The optimization procedure of GoMPL.

Input:
1. The multi-view data of n samples: $X = \{X^{(v)} \in \mathbb{R}^{d_v \times n}\}_{v=1}^{m}$.
2. The regularization parameter λ.

Initialization::
1. Constructing the adjacency graph $A^{(v)}$ based on heat kernel function:
$\left(A^{(v)}\right)_{ij} = exp\left\{-\left\|x_i^{(v)} - x_j^{(v)}\right\|^2 / 2\sigma^2\right\}$.
2. Initializing centered graph $S = \frac{1}{m}\sum_{v=1}^{m} A^{(v)}$

Optimization of GoMPL:

Repeat:
1. **Update** $\{P^{(v)}\}_{v=1}^{m}$:
   for each v, fix $P^{(i)}, i \neq v$ and S, update $P^{(v)}$ based on Equation (7) by generalized eigenvalue decomposition.
2. **Update** S:
   Do:
      for each i, update the ith column of S by Equation (15)
      update the weight factors $w^{(v)}$ for each view by $w^{(v)} = \frac{1}{2\sqrt{\left\|S - A^{(v)}\right\|_F^2 + \delta}}$.
   Until convergence.
3. Calculate the low-dimensional representations by Equation (16)

Output:
The projection matrices and low-dimensional representation $P^{(v)}, Y^{(v)}, v = 1, 2, ..., m$ for all views.

Theorem 1. *In each iteration of updating* **S***, the target equation* (10) *will monotonically decrease.*

In order to prove Theorem 1 , we first introduce one lemma [39]:

Lemma 1. *For any* $f > 0$ *and* $g > 0$*, the following relationship holds:*

$$f - \frac{f^2}{2g} \leq g - \frac{g^2}{2g} \tag{17}$$

Proof of Theorem 1. Let S^k denote kth iteration of S problem. According to the algorithm, we have:

$$\begin{aligned} S^{k+1} =& arg \min_{S^k} Tr\left(US^k\right) + \lambda \sum_{v=1}^{m} w_{S^k}^{(v)} \left\|S^k - A^{(v)}\right\|_F^2 \\ s.t. \quad & \sum_{j=1}^{N} s_{ij} = 1, \quad s_{ij} \geq 0, \end{aligned} \tag{18}$$

By bringing $w_{S^k}^{(v)} = \frac{1}{2\left\|S^k - A^{(v)}\right\|_F}$ into Equation (18), we can obtain:

$$Tr\left(US^{k+1}\right) + \lambda \sum_{v=1}^{m} \frac{\left\|S^{k+1} - A^{(v)}\right\|_F^2}{2\left\|S^k - A^{(v)}\right\|_F} \leq Tr\left(US^k\right) + \lambda \sum_{v=1}^{m} \frac{\left\|S^k - A^{(v)}\right\|_F^2}{2\left\|S^k - A^{(v)}\right\|_F} \tag{19}$$

Moreover, according to Lemma 1, we have:

$$\sum_{v=1}^{m} \left\|S^{k+1} - A^{(v)}\right\|_F - \sum_{v=1}^{m} \frac{\left\|S^{k+1} - A^{(v)}\right\|_F^2}{2\left\|S^k - A^{(v)}\right\|_F} \leq \sum_{v=1}^{m} \left\|S^k - A^{(v)}\right\|_F - \sum_{v=1}^{m} \frac{\left\|S^k - A^{(v)}\right\|_F^2}{2\left\|S^k - A^{(v)}\right\|_F} \tag{20}$$

Since regularization parameter λ is positive, we can sum Equation (20)*λ and Equation (19), and obtain:

$$Tr\left(US^{k+1}\right) + \lambda \sum_{v=1}^{m} \left\| S^{k+1} - A^{(v)} \right\|_F \leq Tr\left(US^{k}\right) + \lambda \sum_{v=1}^{m} \left\| S^{k} - A^{(v)} \right\|_F \quad (21)$$

This inequality in Equation (21) illustrates that the objective function Equation (10) for solving S problem will monotonically decrease by iterative optimization. Therefore, the obtained S will finally converge to the KKT condition of Equation (10), and we can get a local optimal solution of the final objective function Equation (5). □

4. Experiments

4.1. Datasets and Comparing Methods

To evaluate the effectiveness of our GoMPL, we compare our algorithm with several representative dimensionality reduction methods using five datasets, including BBCSport, BBC, ORL, COIL-20, and Outdoor Scene. Among these five datasets, BBCSport and BBC are multi-view textual datasets. ORL, COIL-20 and Outdoor Scene are image datasets. The details of employed datasets are shown in Table 1. The methods employed for comparison are 1. Graph-optimized locality preserving projections (GoLPP) [24] with the best single view (GoLPP_BSV), 2. Graph optimization for dimensionality reduction with sparsity constraints (GODRSC) [38] with the best single view (GODRSC_BSV), 3. Multi-view canonical correlation analysis (MCCA) [33], 4. Multi-view spectral embedding (MSE) [35], 5. Multi-view dimensionality co-reduction (McDR) [40], 6. CMSRE [41], 7. Our method (GoMPL).

4.2. Text Classification

In this section, we conduct text classification experiments on two multi-view textual datasets: BBCSport and BBC. For single-view dimensionality reduction methods, the performances of the best single view are selected. For multi-view methods, after dimensionality reduction, we perform classification by the low-dimensional representation of each view and demonstrate the average performance of different views. The predefined graph of each view is built in the same way as LPP [14]. We find the tradeoff parameter $\lambda = 0.6$ by 5 folds cross-validation. All the features are normalized first.

For the BBCsport dataset, we randomly select 60% samples to train the models. All algorithms are performed to find the subspaces with different dimensionalities. Then we adopt 3NN classifier to perform the final classification. All the algorithms are conducted 20 times with different random training samples and Table 2 shows the classification accuracies.

For BBC dataset, we randomly select 411 samples as the training ones. All algorithms are performed to construct the optimal subspaces with different dimensionalities. All DR methods are trained 20 times with different random training samples and Table 3 shows the results.

From the results shown in Tables 2 and 3, we can find that the proposed GoMPL obtains the best performances for two multi-view document datasets. Meanwhile, for the single-view learning method, GODRSC can outperform some multi-view learning algorithms, which means that incorporating the graph-optimized into the dimensionality reduction framework can obtain better results.

4.3. Face Recognition

In this section, we conduct a face recognition experiment on the ORL face datasets to evaluate the proposed GoMPL. The tradeoff parameter $\lambda = 0.6$ is obtained by 5-fold cross-validation. For the ORL face image dataset, we extract 3 types of features: intensity, LBP and Gabor. Seven samples of each class are selected to train the models. All algorithms are performed to construct the optimal subspaces with different dimensionalities. We run these experiments 20 times with different random training samples and Table 4 shows the results.

It can be seen that the proposed GoMPL obtains the best performances in most situations. Therefore, it is a good dimensionality reduction algorithm to deal with multi-view features for face recognition tasks.

Table 1. The detailed information of all datasets.

Datasets	BBCSport	BBC	ORL	COIL-20	Outdoor Scene
Sizes	554	685	400	1400	2688
Classes	5	5	40	20	8
Views	2	4	3	3	4

Table 2. Classification accuracies (%) with different dimensions on BBCsport dataset.

Method	Dim = 10		Dim = 30		Dim = 50	
	Mean	Max	Mean	Max	Mean	Max
GoLPP_BSV	78.27	82.89	79.44	87.03	81.01	89.41
GODRSC_BSV	80.14	83.94	80.80	87.56	83.47	87.83
MCCA	79.25	84.96	81.55	88.69	81.78	86.26
MSE	79.68	83.57	80.09	87.76	84.08	86.34
McDR	80.45	85.36	81.45	88.41	83.29	90.85
CMSRE	79.96	87.41	80.53	88.95	82.76	90.34
Ours	**84.60**	**89.41**	**83.76**	**91.12**	**85.76**	**92.88**

Table 3. Classification accuracies (%) with different dimensions on BBC dataset.

Method	Dim = 10		Dim = 30		Dim = 50	
	Mean	Max	Mean	Max	Mean	Max
GoLPP_BSV	57.69	67.44	65.72	78.48	62.78	74.13
GODRSC_BSV	60.39	67.41	63.89	77.76	64.31	78.43
MCCA	63.79	73.59	72.62	79.58	70.65	80.63
MSE	64.47	69.95	70.23	72.47	73.75	81.38
McDR	65.03	72.88	79.34	90.16	80.20	92.33
CMSRE	65.81	76.51	76.08	88.01	80.68	90.76
Ours	**70.31**	**79.73**	**76.99**	**87.56**	**81.77**	**92.75**

Table 4. Recognition accuracies (%) with different dimensions on ORL dataset.

Method	Dim = 10		Dim = 30		Dim = 50	
	Mean	Max	Mean	Max	Mean	Max
GoLPP_BSV	60.36	67.16	63.58	75.69	63.77	76.12
GODRSC_BSV	62.31	66.58	64.13	75.84	65.27	77.78
MCCA	63.73	70.51	65.14	73.26	69.09	79.88
MSE	63.47	71.26	67.19	78.37	70.43	81.75
McDR	64.92	71.98	74.39	84.21	75.13	83.39
CMSRE	66.15	72.11	73.21	82.48	76.33	84.52
Ours	**70.28**	**80.53**	**75.68**	**86.35**	**84.61**	**90.23**

4.4. Image Retrieval

In order to evaluate GoMPL on image retrieval, we conduct some experiments on COIL-20 and outdoor scene datasets. For the COIL-20 dataset, we extract the same types of features as ORL. For the outdoor scene dataset, we extract the feature using GIST, LBP, HOG and color moment. The parameter λ is set to 0.4 and 0.3 in the experiment of COIL-20 and outdoor scene dataset respectively by 5-fold cross-validation.

For the experiment on the COIL-20 dataset, we randomly select 400 images as the query images and the rest 1040 images are relevant ones. All these methods were conducted to project all samples to a 60-dimensional subspace. Then we employ L_1 distance to measure the similarity between the low-dimensional representations. All the algorithms are conducted 20 times with different queries and we show the results in Table 5.

There are 2688 color images that come from 8 categories in the outdoor scene dataset. We randomly assign 10 images as the query images for each category. All these methods were conducted to project all samples to a 100-dimensional subspace. Then we employ L_1 distance to measure the similarity. All the algorithms are conducted 20 times with different queries, and we show the results in Figure 1.

From the results shown in Table 5 and Figure 1, it can be seen that GoMPL obtains the best performances in most situations. Although other multi-view learning methods consider the correlations between different views, they adopt a predefined graph that is constructed by original noisy data. Therefore, they can’t utilize the clear structure of the multi-view data for dimensionality reduction.

4.5. Parameter Tuning and Convergence

In our proposed approach, $\lambda > 0$ is an essential parameter to control the strength of utilizing original high-dimensional data information. Moreover, since GoMPL employs the iterative scheme to solve the optimization problem, we also need to test the convergence rate of our approach. Therefore, to show the impact of the regularization parameter λ and analyze the convergence, we conduct another experiment on the ORL dataset to demonstrate the performance of GoMPL. Figure 2a shows the accuracies of GoMPL with different values of λ. From Figure 2a, we can find that the value of λ exercises considerable influence on the performance of GoMPL. Moreover, GoMPL can obtain the best performance when $\lambda = 0.6$ on the COIL-20 dataset. From Figure 2b, we can see that the proposed GoMPL converges in less than 10 iterations.

Table 5. The performance of different algorithms on the COIL-20 dataset.

Method	Criteria			
	Precision	**Recall**	**MAP**	**F1-Score**
GoLPP_BSV	71.76	56.34	83.07	31.56
GODRSC_BSV	74.21	59.94	88.27	33.15
MCCA	73.89	58.46	87.03	32.47
MSE	76.53	59.58	87.62	33.5
McDR	78.26	60.73	88.74	34.2
CMSRE	77.43	61.22	86.19	34.19
Ours	**81.66**	**62.13**	**90.87**	**35.28**

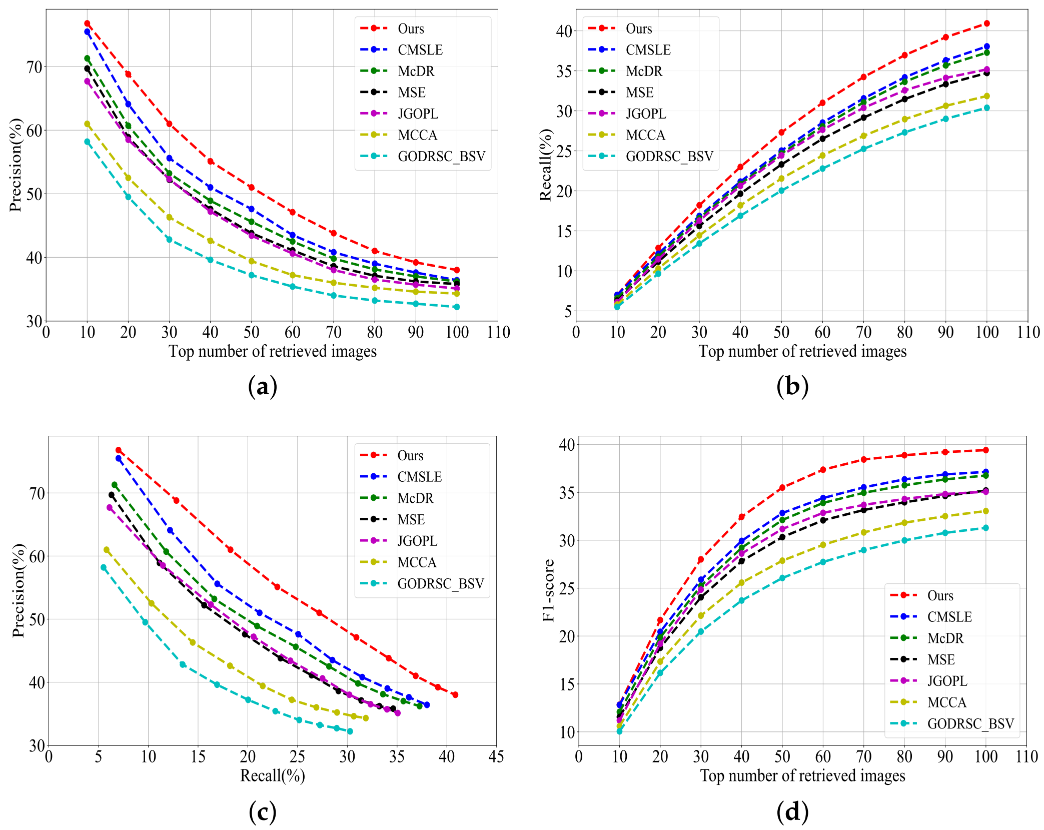

Figure 1. The performance of different algorithms on the outdoor scene dataset. (**a**) Precision. (**b**) Recall. (**c**) PR-Curve. (**d**) F1-Score.

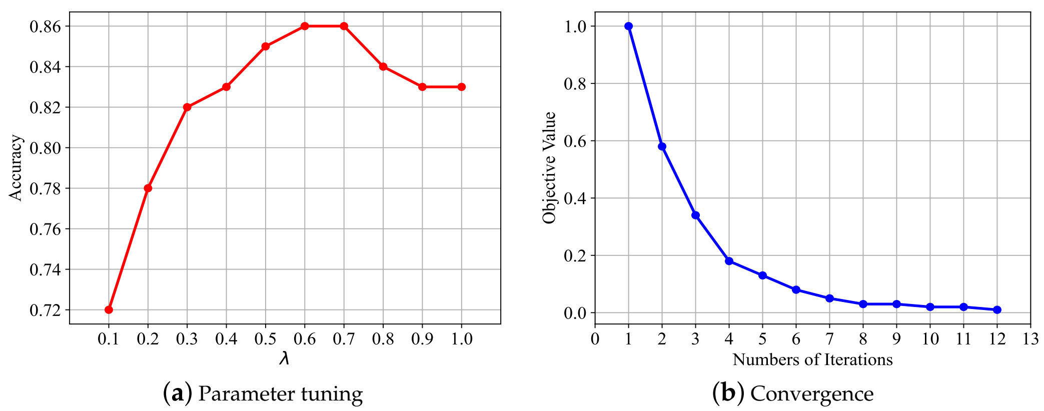

Figure 2. Results of face recognition with different λ (**a**); Objective function value with the number of iterations (**b**).

5. Conclusions

By introducing graph optimization, we propose a novel dimensionality reduction algorithm for multi-view data termed GoMPL, which integrates dimensionality reduction and graph optimization into unified frameworks to construct the projection matrices for different views and optimize the graph jointly. Therefore, the proposed GoMPL performs dimensionality reduction based on the optimal high-quality graph for each view. Moreover, the learned optimal graph by the proposed GoMPL can also integrate the information from multi-view data without co-regularizing the low-dimensional representations. Furthermore, to consider the information of original high-dimensional multi-view data, GoMPL regularizes the common target graph to approximate the predefined graphs based on them.

Plenty of experiments demonstrate that the proposed GoMPL can effectively explore the underlying intrinsic manifold structure of samples described by features from multiple views, and find more appropriate subspace for each view features than compared algorithms.

Our proposed GoMPL can explore the complementarity among multiple views for subspace learning. However, there are still some issues that require further clarification and possible future investigations. First, since the dimensionality reduction is constructed based on graph embedding it leads to computational cost on the matrix operations. Specifically, matrix inversion and eigendecomposition exploited in our method make our algorithm with high computational costs. However, in many practical applications, the scale of datasets is very large. Therefore, to deal with large-scale datasets, we will utilize DeepWalk [42] technique to accelerate the graph embedding speed for the proposed GoMPL. Second, GoMPL learns one common graph for all views to perform dimensionality reduction, which is not flexible enough. In future work, we will consider learning an optimal graph for each view to explore the underlying geometric structure of samples from multi-view data more flexibly.

Author Contributions: Methodology, G.J.; Software, H.L.; Validation, M.G. and H.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Science Foundation of Zhejiang Sci-Tech University (ZSTU) (Grant No. 11432932612227).

Data Availability Statement: Data openly available in a public repository.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Xu, C.; Tao, D.; Xu, C. Multi-View Intact Space Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 2531–2544. [CrossRef] [PubMed]
2. Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; Cao, X. Latent Multi-view Subspace Clustering. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4333–4341.
3. Wang, H.; Yao, M.; Jiang, G.; Mi, Z.; Fu, X. Graph-Collaborated Auto-Encoder Hashing for Multiview Binary Clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, 1–13. [CrossRef] [PubMed]
4. Wang, H.; Jiang, G.; Peng, J.; Deng, R.; Fu, X. Towards Adaptive Consensus Graph: Multi-view Clustering via Graph Collaboration. *IEEE Trans. Multimed.* **2022**, 1–13. [CrossRef]
5. Torralba, A.; Murphy, K.P.; Freeman, W.T.; Rubin, M.A. Context-based vision system for place and object recognition. In Proceedings of the Proceedings Ninth IEEE International Conference on Computer Vision, Nice, France, 13–16 October 2003; Volume 1, pp. 273–280.
6. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 1, pp. 886–893.
7. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 971–987. [CrossRef]
8. Lowe, D. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
9. Wang, J.; Yang, J.; Yu, K.; Lv, F.; Huang, T.; Gong, Y. Locality-constrained Linear Coding for image classification. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3360–3367.
10. Bi, X.; Pun, C.M.; Yuan, X.C. Multi-Level Dense Descriptor and Hierarchical Feature Matching for Copy-Move Forgery Detection. *Inf. Sci.* **2016**, *345*, 226–242. [CrossRef]
11. Wang, H.; Peng, J.; Chen, D.; Jiang, G.; Zhao, T.; Fu, X. Attribute-Guided Feature Learning Network for Vehicle Reidentification. *IEEE Multimed.* **2020**, *27*, 112–121. [CrossRef]
12. Wold, S. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
13. Balakrishnama, S.; Ganapathiraju, A. Linear discriminant analysis-a brief tutorial. *Inst. Signal Inf. Process.* **1998**, *18*, 1–8.
14. He, X.; Niyogi, P. Locality preserving projections. *Adv. Neural Inf. Process. Syst.* **2003**, *16*.
15. He, X.; Cai, D.; Yan, S.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision, Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.
16. Yan, S.; Xu, D.; Zhang, B.; Zhang, H.J. Graph embedding: A general framework for dimensionality reduction. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005; Volume 2, pp. 830–837.

17. Qiao, L.; Chen, S.; Tan, X. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2010**, *43*, 331–341. [CrossRef]
18. Xu, D.; Yan, S.; Tao, D.; Lin, S.; Zhang, H.J. Marginal Fisher Analysis and Its Variants for Human Gait Recognition and Content-Based Image Retrieval. *IEEE Trans. Image Process.* **2007**, *16*, 2811–2821. [CrossRef] [PubMed]
19. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [CrossRef]
20. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [CrossRef] [PubMed]
21. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [CrossRef]
22. Hinton, G.E.; Roweis, S. Stochastic Neighbor Embedding. In *Advances in Neural Information Processing Systems*; Becker, S., Thrun, S., Obermayer, K., Eds.; MIT Press: British, BC, Canada, 2002; Volume 15.
23. Zhang, Z.; Zha, H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM J. Sci. Comput.* **2004**, *26*, 313–338. [CrossRef]
24. Zhang, L.; Qiao, L.; Chen, S. Graph-optimized locality preserving projections. *Pattern Recognit.* **2010**, *43*, 1993–2002. [CrossRef]
25. Qiao, L.; Zhang, L.; Chen, S. Dimensionality reduction with adaptive graph. *Front. Comput. Sci.* **2013**, *7*, 745–753. [CrossRef]
26. Yi, Y.; Wang, J.; Zhou, W.; Fang, Y.; Kong, J.; Lu, Y. Joint graph optimization and projection learning for dimensionality reduction. *Pattern Recognit.* **2019**, *92*, 258–273. [CrossRef]
27. Kumar, A.; Rai, P.; Daume, H. Co-regularized Multi-view Spectral Clustering. In *Advances in Neural Information Processing Systems*; Shawe-Taylor, J., Zemel, R., Bartlett, P., Pereira, F., Weinberger, K., Eds.; Curran Associates, Inc.: Granada, Spain, 2011; Volume 24.
28. Guo, J.; Zhu, W. Partial Multi-View Outlier Detection Based on Collective Learning. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
29. Wang, H.; Wang, Y.; Zhang, Z.; Fu, X.; Zhuo, L.; Xu, M.; Wang, M. Kernelized Multiview Subspace Analysis By Self-Weighted Learning. *IEEE Trans. Multimed.* **2021**, *23*, 3828–3840. [CrossRef]
30. Wang, H.; Peng, J.; Zhao, Y.; Fu, X. Multi-Path Deep CNNs for Fine-Grained Car Recognition. *IEEE Trans. Veh. Technol.* **2020**, *69*, 10484–10493. [CrossRef]
31. Wang, H.; Peng, J.; Jiang, G.; Xu, F.; Fu, X. Discriminative feature and dictionary learning with part-aware model for vehicle re-identification. *Neurocomputing* **2021**, *438*, 55–62. [CrossRef]
32. Hardoon, D.R.; Szedmak, S.; Shawe-Taylor, J. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.* **2004**, *16*, 2639–2664. [CrossRef] [PubMed]
33. Rupnik, J.; Shawe-Taylor, J. Multi-view canonical correlation analysis. In Proceedings of the Conference on Data Mining and Data Warehouses (SiKDD 2010), Ljubljana, Slovenia, 12 October 2010; pp. 1–4.
34. Long, B.; Yu, P.S.; Zhang, Z. A general model for multiple view unsupervised learning. In Proceedings of the 2008 SIAM International Conference on Data Mining, SIAM, Atlanta, GA, USA, 24–26 April 2008; pp. 822–833.
35. Xia, T.; Tao, D.; Mei, T.; Zhang, Y. Multiview spectral embedding. *IEEE Trans. Syst. Man, Cybern. Part B (Cybernetics)* **2010**, *40*, 1438–1446.
36. Kan, M.; Shan, S.; Zhang, H.; Lao, S.; Chen, X. Multi-View Discriminant Analysis. In Proceedings of the European Conference on Computer Vision, Florence, Italy, 7–13 October 2012; pp. 188–194.
37. Ding, Z.; Fu, Y. Low-rank common subspace for multi-view learning. In Proceedings of the 2014 IEEE International Conference on Data Mining, Shenzhen, China, 14–17 December 2014; pp. 110–119.
38. Zhang, L.; Chen, S.; Qiao, L. Graph optimization for dimensionality reduction with sparsity constraints. *Pattern Recognit.* **2012**, *45*, 1205–1210. [CrossRef]
39. Nie, F.; Huang, H.; Cai, X.; Ding, C. Efficient and Robust Feature Selection via Joint l2,1-Norms Minimization. In Proceedings of the 23rd International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 6–9 December 2010; Volume 2, pp. 1813–1821.
40. Zhang, C.; Fu, H.; Hu, Q.; Zhu, P.; Cao, X. Flexible multi-view dimensionality co-reduction. *IEEE Trans. Image Process.* **2016**, *26*, 648–659. [CrossRef]
41. Wang, H.; Peng, J.; Fu, X. Co-regularized multi-view sparse reconstruction embedding for dimension reduction. *Neurocomputing* **2019**, *347*, 191–199. [CrossRef]
42. Perozzi, B.; Al-Rfou, R.; Skiena, S. DeepWalk: Online Learning of Social Representations. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; Association for Computing Machinery: New York, NY, USA; pp. 701–710.