# A Graph-Based Algorithm for Supervised Image Classification

Ke Du[1], Jinlong Liu[2(✉)], Xingrui Zhang[2], Jianying Feng[2], Yudong Guan[2], and Stéphane Domas[1]

[1] FEMTO-ST Institute, UMR 6174 CNRS,
University of Bourgogne Franche-Comté, 90000 Belfort, France
[2] School of Electronics and Information Engineering,
Harbin Institute of Technology, Harbin 150000, China
`yq20@hit.edu.cn`

**Abstract.** Manifold learning is a main stream research track used for dimensionality reduction as a method to select features. Many variants have been proposed with good performance. A novel graph-based algorithm for supervised image classification is introduced in this paper. It makes the use of graph embedding to increase the recognition accuracy. The proposed algorithm is tested on four benchmark datasets of different types including scene, face and object. The experimental results show the validity of our solution by comparing it with several other tested algorithms.

**Keywords:** Graph-based · Supervised learning · Image classification

## 1 Introduction

In the last years, machine learning has been playing an important role in many domains, especially in image recognition and classification. It has shown the great power for effective learning. In supervised learning, a physical phenomenon is described by a mapping between predict or labeled data. In this domain, graph-based algorithms have drawn great attention [1–5]. A lot of efforts have been done by using graph-based learning methods to various topics, such as regression [6] and dimensionality reduction [7].

Techniques that address the latter problem were proposed to reduce the multi-dimensional data dimensionality. It aims to find relevant subsets for feature description. It yields a smaller set of representative features while preserving the optimal salient characteristics. Hence, not only the processing time can be decreased, but also a better generalization of the learning models can be achieved. The algorithms mentioned above rely on both the manifold structure and learning mechanism [8–10]. Therefore, in many cases, it is possible to achieve better performance than other conventional methods. However, all of these methods firstly define the characterized manifold structure and then perform a regression [5]. As a result, the constructed graphs have great effects on

the performance. Indeed, the graph spectral is fixed in the following regression steps.

Taking into consideration the above remarks, we introduce in this paper a graph-based algorithm for efficient supervised image classification. It applies the models of graph-based dimensionality reduction and sparse regression simultaneously. Besides, an iterative locally linear graph weight algorithm is applied to acquire graph weights and improve the recognition accuracy. Finally, we inspect the optimization problem of the proposed approach and we demonstrate the situations to solve it.

The rest of the paper is structured as follows. In Sect. 2, the graph embedding model is introduced. Section 3 details the proposed graph-based supervised classification algorithm. Section 4 presents the experiments carried out on benchmark datasets to verify the effectiveness of the proposed algorithm by comparing with other art-of-state algorithms. The analysis of the experimental results are also given. Finally, in Sect. 5, we draw conclusions and discuss the works for the future research.

## 2   Related Works

### 2.1   Notations and Preliminaries

In order to make the paper self-contained, the notations used in the paper are introduced. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_l, \mathbf{x}_{l+1}, \cdots, \mathbf{x}_{l+u}] \in \mathbb{R}^{d \times (l+u)}$ is defined as the sample data matrix, where $\mathbf{x}_i \big|_{i=1}^{l}$ and $\mathbf{x}_j \big|_{j=l+1}^{l+u}$ are the labeled and unlabeled samples, respectively. $l$ and $u$ are the total numbers of labeled and unlabeled samples, respectively, and $d$ is the sample dimension. Let $N$ be the total number of samples. The label of each sample $\mathbf{x}_i$ is denoted by $y_i \in 1, 2, ..., C$, where $C$ relates to the total number of classes. Let $\mathbf{S} \in \mathbb{R}^{(l+u) \times (l+u)}$ be the graph similarity matrix, where $\mathbf{S}_{ij}$ represents the similarity between $\mathbf{x}_i$ and $\mathbf{x}_j$ as given by the Cosine or the Gaussian Kernel ($\mathbf{S}$ is symmetric). To make it clear, Table 1 shows all the nations and descriptions in this paper.

### 2.2   Graph Embedding

In graph embedding, each node of a constructed graph $\mathbf{G} = \{\mathbf{X}, \mathbf{S}\}$ relates to a data point $\mathbf{x}_i \in \mathbf{X}$ [11]. The graph embedding is aimed at finding an optimal matrix $\mathbf{Y}$ with a lower dimension that can make the best description of the similarity between the data well. The optimal $\mathbf{Y}$ is given by

$$\arg \min_{\mathbf{Y}} (\mathbf{Y}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{Y})$$
$$s.t. \ \mathbf{Y}^T \mathbf{X} \mathbf{D} \mathbf{Y}^T \mathbf{A} = \mathbf{I} \tag{1}$$

Where $\mathbf{L} = \mathbf{D} - \mathbf{S}$ gives the Laplacian matrix, $\mathbf{D}$ is a diagonal matrix and $\mathbf{I}$ is an identity matrix.

**Table 1.** Notations and descriptions.

| Notation | Description |
|---|---|
| $d$ | Dimensionality of original data |
| $N$ | Number of data samples |
| $l$ | Number of labeled samples |
| $u$ | Number of unlabeled samples |
| $C$ | Number of classes |
| $\mathbf{x}_i$ | The i-th original data sample |
| $y_i$ | The label of $\mathbf{x}_i$ |
| $\mathbf{S}$ | Graph similarity matrix |
| $\mathbf{W}$ | Linear transformation matrix |
| $\mathbf{D}$ | Diagonal matrix |
| $\mathbf{I}$ | Identity matrix |
| $\mathbf{L}$ | Laplacian matrix |
| $\mathbf{X}_l$ | Labeled train samples matrix |
| $\mathbf{X}_u$ | Unlabeled test samples matrix |
| $\mathbf{X}$ | Original data matrix |
| $\mathbf{Y}$ | Low dimensional matrix |

In fact, different algorithms for dimensionality reduction result in various intrinsic graphs $\mathbf{G} = \{\mathbf{X}, \mathbf{S}\}$. The most used algorithms to reduce the dimensionality include Principal Components Analysis (PCA), Linear Discriminant Analysis (LAD), Locally Linear Embedding (LLE) [12], Locality Preserving Projections (LPP) [2], ISOMAP [13], etc.

## 3    Proposed Algorithm

### 3.1    Similarity Matrix S

Firstly, a nearest neighbors method is used to determine $k$ neighbors ($k \leq N$) for each node. Asuming that $i$ and $j$ are two nodes linked by an edge, if $i$ is among the $k$ nearest neighbors of $j$, or if $j$ is among the $k$ nearest neighbors of $i$. It is obvious that this relation is symmetric.

Secondly, the similarity matrix $\mathbf{S}$ is computed. It is introduced in [14,15]. In order to acquire better performance for recognition and classification, the matrix $\mathbf{S}$ is computed in a high-dimensional data space. The regularizer $L_{1/2}$ is used as an unbiased estimator in this paper. It is used to improve the sparsity of matrix $\mathbf{S}$ for the minimization problem. Additionally, for graph embedding, the condition $\mathbf{S} \geq 0$ is added. The process of minimization can be presented as:

$$\min_{\mathbf{S} \geq 0} \sum_i \left\| \mathbf{x}_i - \sum_j \mathbf{S}_{i,j} \mathbf{x}_j \right\|^2 + \alpha \|\mathbf{S}\|_{\frac{1}{2}} + \beta \|\mathbf{S}\|^2$$

$$= \min_{\mathbf{S} \geq 0} \|\mathbf{X} - \mathbf{X}\mathbf{S}\|^2 + \alpha \|\mathbf{S}\|_{\frac{1}{2}} + \beta \|\mathbf{S}\|^2$$

$$\Rightarrow \min_{\mathbf{S} \geq 0} Tr\left(\tilde{\kappa} - 2\tilde{\kappa}\mathbf{S} + \mathbf{S}^T \tilde{\kappa} \mathbf{S}\right) + \alpha \|\mathbf{S}\|_{\frac{1}{2}} + \beta Tr\left(\mathbf{S}^T \mathbf{S}\right) \tag{2}$$

Where $\alpha$ and $\beta$ are the free parameters, $\tilde{\kappa}$ the kernel of $\mathbf{X}$ and $\|\mathbf{S}\|_{\frac{1}{2}} = \sum_i \sum_j \mathbf{S}_{i,j}^{1/2}$.

Thus, Eq. (2) could be rewritten as:

$$\min_{\mathbf{S} \geq 0} Tr\left[\left(\tilde{\kappa} - 2\tilde{\kappa}\mathbf{S} + \mathbf{S}^T \tilde{\kappa} \mathbf{S}\right) + \beta \mathbf{S}^T \mathbf{S}\right] + \alpha \|\mathbf{S}\|_{\frac{1}{2}} \tag{3}$$

Furthermore, Eq. (3) is equivalent to

$$\min_{\mathbf{S} \geq 0} Tr\left[\mathbf{S}^T \left(\beta \mathbf{I} + \tilde{\kappa}\right) \mathbf{S} - 2\tilde{\kappa}\mathbf{S} + \tilde{\kappa}\right] + \alpha \|\mathbf{S}\|_{\frac{1}{2}} \tag{4}$$

It should be noticed that minimizing Eq.(4) is subjected to $\mathbf{S} \geq 0$. Let $\varsigma \geq 0$ be the corresponding Lagrange multipliers. The Lagrange function $F(\mathbf{S})$ can be presented as:

$$F(\mathbf{S}) = Tr\left[\mathbf{S}^T \left(\beta \mathbf{I} + \tilde{\kappa}\right) \mathbf{S} - 2\tilde{\kappa}\mathbf{S} + \tilde{\kappa}\right] + \alpha \|\mathbf{S}\|_{\frac{1}{2}} + Tr\left(\zeta \mathbf{S}^T\right) \tag{5}$$

Then, partial derivative of both sides leads to

$$\frac{\partial F(\mathbf{S})}{\partial \mathbf{S}_{ij}} = \left(-2\tilde{\kappa} + 2\tilde{\kappa}\mathbf{S} + 2\beta \mathbf{S} + \frac{1}{2}\alpha \mathbf{S}^{-\frac{1}{2}} + \zeta\right)_{ij} \tag{6}$$

Where $\mathbf{S}^{-\frac{1}{2}}$ is equivalent to the inverse matrix of principal square-rooting matrix $\mathbf{S}^{\frac{1}{2}}$.

Then, the Karush-Kuhn-Tucker(KKT) condition $\zeta \mathbf{S} = 0$ for $\mathbf{S}$ is

$$\left(-2\mathbf{X} + 2\mathbf{X}\mathbf{S} + 2\beta \mathbf{S} + \frac{1}{2}\alpha \mathbf{S}^{-\frac{1}{2}} + \zeta\right)_{ij} \mathbf{S}_{ij} = 0 \tag{7}$$

Eq. (7) can be reformulated as:

$$(-\tilde{\kappa}_{ij} + (\tilde{\kappa}\mathbf{S} + \beta \mathbf{S} + \frac{1}{2}\alpha \mathbf{S}^{-\frac{1}{2}})_{ij})\mathbf{S}_{ij} = 0 \tag{8}$$

An iterative process to retrieve $\mathbf{S}$ is expressed by

$$\mathbf{S}_{ij} \leftarrow \frac{\mathbf{X}}{(\mathbf{X}\mathbf{S} + \beta \mathbf{S} + \frac{1}{4}\alpha \mathbf{S}^{-\frac{1}{2}})_{ij}} \mathbf{S}_{ij} \tag{9}$$

In fact, Eq. (9) only shows the computation for one iteration and it repeats many times until the result is convergence. Finally, we acquire the similarity matrix $\mathbf{S}$ for graph projection.

## 3.2   Graph Embedding Learning

The work described in [16] proposed a novel graph-based embedding framework for feature selection with unsupervised learning, named Joint Embedding Learning and Sparse Regression (JELSR). This unsupervised method aims at ranking the original features by performing non-linear embedding learning and sparse regression concurrently. JELSR inspired us to develop a method with graph embedding algorithm for supervised learning in the domain of image classification.

Based on graph embedding and sparse regression optimization function, we can optimize it by making the following operation:

$$\ell(\mathbf{W}, \mathbf{Y}) = \underset{\mathbf{W}, \mathbf{Y} \ s.t. \mathbf{Y}^T \mathbf{Y} = \mathbf{I}}{\arg\min} (trace(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) + \mu(\left\| \mathbf{W^T X} - \mathbf{Y} \right\|_2^2 + \gamma \|\mathbf{W}\|_{2,1})) \quad (10)$$

Where $\gamma$ and $\mu$ are two regularization parameters. $\mathbf{W}$ represents the linear transform matrix, $m$ is the graph embedding dimensionality, and $\mathbf{Y}$ denotes the data matrix of embedding non-linear projection of $\mathbf{X}$. The $\ell_{2,1}$ norm of $\mathbf{W}$ is given by $\|\mathbf{W}\|_{2,1} = \sum_{i=1}^{d} \|\hat{\mathbf{w}}_i\|_2$. $\hat{\mathbf{w}}_i$ is the $i$-th row of $\mathbf{W}$.

Respecting to the matrix $\mathbf{W}$, we can get the derivative of $\ell(\mathbf{W}, \mathbf{Y})$ as follows,

$$\frac{\partial \ell(\mathbf{W}, \mathbf{Y})}{\partial \mathbf{W}} = 2\mathbf{X}\mathbf{X}^T\mathbf{W} - 2\mathbf{X}\mathbf{Y}^T + 2\gamma\mathbf{U}\mathbf{W} = 0 \quad (11)$$

Where $\mathbf{U} \in \mathbb{R}^{d \times d}$ is a diagonal matrix. The $i$-th diagonal element is $\mathbf{U}_{ii} = \frac{1}{2\|\hat{\mathbf{w}}_i\|_2}$.

Thus, we have the equation as follows:

$$\mathbf{W} = (\mathbf{X}\mathbf{X}^T + \gamma\mathbf{U})^{-1}\mathbf{X}\mathbf{Y}^T \quad (12)$$

Equation (10) can be reformulated as:

$$
\begin{aligned}
\ell(\mathbf{W}, \mathbf{Y}) &= \underset{\mathbf{W}, \mathbf{Y} \ s.t. \mathbf{Y}^T \mathbf{Y} = \mathbf{I}}{\arg\min} (trace(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) + \mu(\left\| \mathbf{W^T X} - \mathbf{Y} \right\|_2^2 + \gamma \|\mathbf{W}\|_{2,1}) \\
&= tr(\mathbf{Y}\mathbf{L}\mathbf{Y}^T) + \mu(tr(\mathbf{W}^T\mathbf{X}\mathbf{X}^T\mathbf{W}) - 2tr(\mathbf{W}^T\mathbf{X}\mathbf{Y}^T) \\
&\quad + tr(\mathbf{Y}\mathbf{Y}^T) + \gamma tr(\mathbf{W}^T\mathbf{U}\mathbf{W})) \\
&= tr(\mathbf{Y}\mathbf{L}\mathbf{Y}^T) + \mu(-tr(\mathbf{W}^T(\mathbf{X}\mathbf{X}^T + \gamma\mathbf{U})\mathbf{W}) + tr(\mathbf{Y}\mathbf{Y}^T)) \\
&= tr(\mathbf{Y}(\mathbf{L} + \mu\mathbf{I} - \mu\mathbf{X}^T\mathbf{A}^{-1}\mathbf{X})\mathbf{Y}^T) \quad (13)
\end{aligned}
$$

Where $\mathbf{A} = \mathbf{X}\mathbf{X}^T + \gamma\mathbf{U}$.

Taking the objective function and the constraint $\mathbf{Y}\mathbf{Y}^T = \mathbf{I}$ into account, the optimization problem turns to

$$\underset{\mathbf{Y}}{\arg\min}\, tr(\mathbf{Y}(\mathbf{L} + \mu\mathbf{I} - \mu\mathbf{X}^T\mathbf{A}^{-1}\mathbf{X})\mathbf{Y}^T) \ \ s.t. \ \mathbf{Y}\mathbf{Y}^T = \mathbf{I} \quad (14)$$

If $\mathbf{A}$ and $\mathbf{L}$ are fixed, The Eigen decomposition of matrix $(\mathbf{L} + \mu\mathbf{I} - \mu\mathbf{X}^T\mathbf{A}^{-1}\mathbf{X})$ can be used as the solution to the optimization problem in Eq. (14). We select $m$ eigenvectors corresponding to the $m$ smallest eigenvalues in order. These eigenvectors are suitable to build a graph-based embedding which is used for image classification.

# 4   Experiments

We have tested our method on four different datasets. They contains scenes (8 Sports Event Categories Dataset and Scene 15 Dataset), faces (ORL Face Dataset) and objects (COIL-20 Object Dataset). These images have been used in different groups to train and test. The details of the experiments and results are described in the following.

## 4.1   Dataset Configurations

The details of how the images in the four datasets are configured are listed as follows.

**8 Sports Event Categories Dataset** includes 8 sports event categories (provided by Li and Fei-Fei) [17]. We have used 130 images in every category, thus a total of 1040.

**Scene 15 Dataset** includes 4485 gray level images of 15 different scenes including indoor and outdoor scenes [18]. We use 130 images in every category, thus a total of 1950.

**ORL Face Dataset** consists of 10 different images of each 40 distinct subjects [19].

**COIL-20 Objects Dataset** contains 1440 images of 20 objects (provided by Columbia Object Image Library) [20]. We select 70 images out of 72 for each object as a subset.

We have tested different distributions between training and testing images. For the first three datasets, we have used 50% and 70% of images for training twice, leaving 50% and 30% for testing, respectively. For the last dataset, we have used 10% and 20% of images for training, remaining 90% and 80% for testing, respectively.

## 4.2   Graph Performance Comparison

In this experiment, the graph calculated from the similarity matrix **S** is firstly tested with by comparing with that of other classical similarity measure algorithms, such as KNN graph and $\ell_1$ graph. Table 2 displays the performance of graphs based on different similarity measure algorithms. In order to make the comparison, Laplacian Eigenmaps (LE) is chosen as the projection algorithm and the classification algorithm is 1NN classifier. From the results, it can be concluded that the kernelized sparse non-negative graph matrix **S** is able to produce a graph weight matrix much better than the KNN graph and $\ell_1$ graph methods.

## 4.3   Effect of Proposed Algorithm

The block-based Local Binary Patterns (LBP) is used as the image descriptor, where the number of blocks is set to $10 \times 10$. The LBP descriptor is the

**Table 2.** The best average recognition rates (%) on 10 random splits of different graph algorithms.

| Datasets | 8 Sports | | Scene 15 | | ORL Face | |
|---|---|---|---|---|---|---|
| Training images | 50% | 70% | 50% | 70% | 50% | 70% |
| KNN graph | 52.31 | 54.31 | 42.36 | 45.33 | 89.80 | 92.08 |
| $\ell_1$ graph | 53.81 | 57.31 | 46.72 | 49.23 | 89.95 | 93.67 |
| **Proposed algorithm** | **54.83** | **57.44** | **50.49** | **52.67** | **92.10** | **94.50** |

uniform one having 59 features. For ORL Face and COIL-20 Objects datasets, we use image raw brightnesses. The proposed algorithm is tested by comparing with the following five algorithms including LLE, Supervised Laplacian Eigenmaps (SLE) [21], Manifold Regularized Deep Learning Architecture (MRDL) [14], Semi-Supervised Discriminant Embedding (SDE)[22] and S-ISOMAP [23]. For MRDL method, we used two layers. Image classification is carried out in the obtained subspace using the Nearest Neighbor Classifier (NN). The experimental results are listed in Tables 3, 4, 5, and represented as graphs in Figs. 1 and 2.

**Table 3.** The best average recognition rates (%) of 8 Sports Event Categories Dataset on 10 random splits.

| 8 Sports scene | P = 50% | P = 70% |
|---|---|---|
| LLE | 44.92 | 49.10 |
| SLE | 51.40 | 50.90 |
| MRDL | 51.77 | 52.85 |
| S-ISOMAP | 51.88 | 54.68 |
| SDE | 51.98 | 55.96 |
| **Proposed algorithm** | **55.92** | **57.60** |

**Table 4.** The best average recognition rates (%) of Scene 15 Dataset on 10 random splits.

| Scene 15 dataset | P = 50% | P = 70% |
|---|---|---|
| LLE | 44.26 | 47.42 |
| SLE | 50.48 | 50.65 |
| MRDL | 46.59 | 47.91 |
| S-ISOMAP | 42.74 | 45.28 |
| SDE | 46.10 | 48.07 |
| **Proposed algorithm** | **51.83** | **58.59** |

**Table 5.** The best average recognition rates (%) of COIL-20 Object Dataset on 10 random splits.

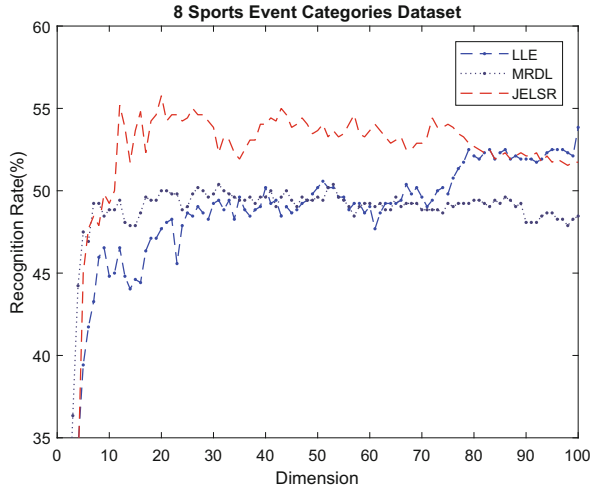| COIL-20 object | P = 10% | P = 20% |
|---|---|---|
| LLE | 91.81 | 94.71 |
| SLE | 82.03 | 88.56 |
| MRDL | 88.00 | 88.86 |
| **Proposed algorithm** | **93.80** | **96.88** |



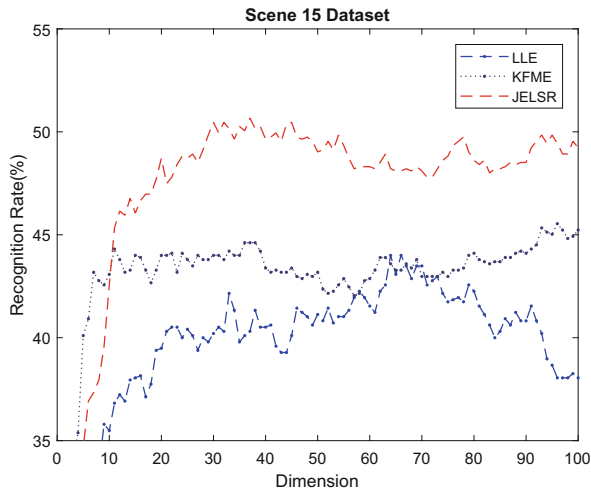**Fig. 1.** Recognition accuracy vs. feature dimension for 8 Sports Event Categories Dataset.



**Fig. 2.** Recognition accuracy vs. feature dimension for Scene 15 Dataset.

As presented by the results, we can draw the following conclusions. Generally, the proposed non-linear graph embedding method has enhanced performances compared with the other algorithms tested on different datasets in Tables 3, 4 and 5. Especially, compared with the MRDL algorithm, the best recognition rate of COIL-20 Object Dataset is increased by 15.80%. As the curves shown in Figs. 1 and 2, the recognition rates do not increase along with the dimension of features. Therefore, the proposed method can perform well without using large quantity of features. It can reduce the time and space complexity of training and classification.

## 5    Conclusions

By emplying a novel procedure, we proposed an image classification algorithm related to kernelized sparse non-negative graph matrix and graph-based sparse regression method. It is intended to reduce the feature dimensionality and improve the recognition accuracy in image classification. Experiments are carried out on benchmark datasets including scene, faces and object datasets to check the effectiveness of our algorithm. From the experimental results, it is obvious that the introduced algorithm outperforms the others tested. In the future, some optimization will be made to ensure the robustness of sparse regression. Some modifications are also needed to ameliorate the performance of our proposed graph-based supervised algorithm for image classification.

## References

1. Zhu, X., Ghahramani, Z., Lafferty, J.D.: Semi-supervised learning using gaussian fields and harmonic functions. In: 20th International Conference on Machine Learning, Washington DC, USA, pp. 912–919 (2003)
2. He, X., Niyogi, P.: Locality preserving projections. Adv. Neural Inf. Proc. Syst. **2**(5), 153–160 (2004)
3. Cheng, H., Liu, Z., Yang, J.: Sparsity induced similarity measure for label propagation. In: 12th IEEE International Conference on Computer Vision (ICCV), pp. 317–324. IEEE, Kyoto (2009)
4. Pei, X., Chen, C., Guan, Y.: Joint sparse representation and embedding propagation learning: a framework for graph-based semisupervised learning. IEEE Trans. Neural Netw. Learn. Syst. **28**(12), 2949–2960 (2017)
5. Shi, X., Guo, Z., Lai, Z., Yang, Y., Bao, Z., Zhang, D.: A framework of joint graph embedding and sparse regression for dimensionality reduction. IEEE Trans. Image Process. **24**(4), 1341–1355 (2015)
6. Ni, B., Yan, S., Kassim, A.: Learning a propagable graph for semisupervised learning: classification and regression. IEEE Trans. Knowl. Data Eng. **24**(1), 114–126 (2012)
7. Nie, F., Xu, D., Li, X., Xiang, S.: Semisupervised dimensionality reduction and classification through virtual label regression. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **41**(3), 675–685 (2011)

8. He, X., Cai, D., Han, J.: Semi-supervised discriminant analysis. In: 11th IEEE International Conference on Computer Vision (ICCV), pp. 1–7. IEEE, Rio de Janeiro (2007)
9. Yan, S., Xu, D., Yang, Q., Zhang, L., Tang, X., Zhang, H.J.: Discriminant analysis with tensor representation. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 526–532. IEEE, San Diego (2005)
10. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(1), 40–51 (2007)
11. Brand, M.: Continuous nonlinear dimensionality reduction by kernel eigenmaps. In: International Joint Conference on Artificial Intelligence (IJCAI), pp. 547–554. ACM, Acapulco (2010)
12. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)
13. Tenenbaum, J.B., De, S.V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290**(5500), 2319–2323 (2000)
14. Yuan, Y., Mou, L., Lu, X.: Scene recognition by manifold regularized deep learning architecture. IEEE Trans. Neural Netw. Learn. Syst. **26**(10), 2222–2233 (2015)
15. Kong, D., Ding, C.H.Q., Huang, H., Nie, F.: An iterative locally linear embedding algorithm. In: 29th International Conference on Machine Learning (ICML), Edinburgh, Scotland, UK (2010)
16. Hou, C., Nie, F., Li, X., Yi, D., Wu, Y.: Joint embedding learning and sparse regression: a framework for unsupervised feature selection. IEEE Trans. Cybern. **44**(6), 793–804 (2014)
17. Li, L.J., Li, F.F.: What, where and who? Classifying events by scene and object recognition. In: 11th IEEE International Conference on Computer Vision (ICCV), pp. 1–8. IEEE, Rio de Janeiro (2007)
18. Lazebnik, S., Schmid, C., Ponce, J.: Beyond bags of features: spatial pyramid matching for recognizing natural scene categories. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 2169–2178. IEEE, New York (2006)
19. Samaria, F.S., Harter, A.C.: Parameterisation of a stochastic model for human face identification. In: 2ed IEEE Workshop on Applications of Computer Vision, pp. 138–142. IEEE, Sarasota (2010)
20. Nene, S.A., Nayar, S.K., Murase, H.: Columbia object image library (coil-20). Technical report CUCS-005-96, Location (1996)
21. Raducanu, B., Dornaika, F.: A supervised non-linear dimensionality reduction approach for manifold learning. Pattern Recogn. **45**(6), 2432–2444 (2012)
22. Yu, G., Zhang, G., Domeniconi, C., Yu, Z., You, J.: Semi-supervised classification based on random subspace dimensionality reduction. Pattern Recogn. **45**(3), 1119–1135 (2012)
23. Geng, X., Zhan, D.C., Zhou, Z.H.: Supervised nonlinear dimensionality reduction for visualization and classification. IEEE Trans. Syst. Man Cybern. Part B (Cybern.) **35**(6), 1098–1107 (2005)