Bilinear Embedding Label Propagation: Towards Scalable Prediction of Image Labels

Yuchen Liang, Zhao Zhang, *Member, IEEE*, Weiming Jiang, Mingbo Zhao, *Member, IEEE*, and Fanzhang Li

***Abstract*—Traditional label propagation (LP) is shown to be effective for transductive classification. To enable the standard LP to handle outside images, two inductive methods by label reconstruction or by direct embedding have been presented, of which the latter scheme is relatively more efficient, especially for testing. But almost all inductive LP models use 1D vectors of images as inputs, which may destroy the topology structure of image pixels and usually suffer from high complexity due to the high dimension of 1D vectors in reality. To preserve the topology among pixels and address the scalability issue for the embedding based scheme, we propose a simple yet efficient *Bilinear Embedding Label Propagation* (BELP) by including a bilinear regularization term in terms of tensor representation to correlate the image labels with their bilinear features. BELP performs label prediction over the 2D matrices rather than 1D vectors, since images are essentially matrices. Finally, labels of new images can be easily obtained by embedding them onto a spanned bilinear subspace solved from a joint framework. Simulations verified the efficiency of our approach.**

***Index Terms*—Bilinear embedding, image label prediction, inductive label propagation, semi-supervised learning.**

I. Introduction

SEMI-SUPERVISED LEARNING (SSL) for classification of images is comparatively more useful than the supervised learning in real application due to lack of adequate labeled images [1]–[8], [20]–[24]. In contrast, the unlabeled images are usually readily available in reality. Using small number of labeled images and large number of unlabeled ones to reduce labelling efforts and obtain enhanced performance are the major goals of SSL.

Label propagation (LP) [3]–[7], as a popular graph based SSL (G-SSL) method [7], [8]for image classification, has been attracting much attention in recent years due to its efficiency and specialty of preserving the geometry structures of labeled and unlabeled data [5], [17]. The main idea of LP is to spread label information of labeled images to unlabeled ones based on the relationships between them [3], [4]. Typical transductive LP models include *Gaussian Fields and Harmonic Function* (GFHF) [4], *Learning with Local and Global Consistency* (LLGC) [5], *Linear Neighborhood Propagation* (LNP) [7][8] and *Special Label Propagation* (SLP) [6], etc.

It is worth noting that the all aforementioned settings (i.e., GFHF, LLGC, LNP and SLP) are transductive models, since they gain the soft labels of images directly without delivering a mapping or a projection that can be used to handle outside images. To the best of our knowledge, there are two popular out-of-sample inductive extensions for LP. The first scheme is to reconstruct each new test image by using the soft labels of its neighbors [7], [8]. The other one is to compute a projection for embedding the test images to deliver their labels, of which *Embedded Label Propagation* (ELP) [11] is one representative approach. Note that the scheme by label reconstruction needs to search the neighbors of each new test image prior to label reconstruction, which would be computationally expensive for large-scale datasets, while the scheme by embedding handles new images by mapping, which is very efficient for testing. But it is noted that almost all existing inductive LP models use 1D vectors as inputs. While images are essentially matrices [13], [16], [25] so the process of converting 2D matrices into 1D vectors may destroy the topology structures of image pixels and usually suffer from high time complexity due to the high dimension of 1D vectors in practice during training phase.

To improve the efficiency of the training phase for the label embedding based inductive scheme, we consider performing label prediction in the 2D matrix space rather than in the 1D vector space. Specifically, an efficient *Bilinear Embedding Label Propagation* termed BELP is proposed by incorporating a bilinear regularization in the form of tensor representation [13], [16] to correlate the image labels with bilinear features. Since BELP is performed in the 2D matrix space, the topology information between image pixels can be effectively preserved during the training phase. So, the label prediction performance can be potentially enhanced. More importantly, the training process will be computationally efficient, since the involved matrices to be computed are of small size (i.e., image length or width) that is usually far smaller than the dimension of the vectorized representations of images [13], [16][25]. Besides, for evaluations on new images, their labels can also be efficiently obtained by embedding them onto a bilinear subspace spanned by two projections that are computed from a joint framework

Manuscript received July 11, 2015; revised September 13, 2015; accepted October 01, 2015. Date of publication October 08, 2015; date of current version October 15, 2016. This work was supported in part by the National Natural Science Foundation of China under Grants 61402310 and 61373093, the Major Program of Natural Science Foundation of Jiangsu Higher Education Institutions of China under Grant 15KJA520002, the Postdoctoral Science Foundation of Jiangsu Province of China under Grant1501091B, and by the Undergraduate Student Innovation Program of Soochow University under Grant2014xj069. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Arrate Munoz-Barrutia.

Y. Liang, Z. Zhang, W. Jiang, and F. Li are with the School of Computer Science and Technology, Soochow University, Suzhou 215006, China (e-mail: LeungYc183543@outlook.com; cszzhang@gmail.com; lfzh@suda.edu.cn).

M. Zhao is with Department of Electronic Engineering, City University of Hong Kong, Hong Kong (e-mail: mzhao4@cityu.edu.hk).

(*Corresponding author: Z. Zhang*.)

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2015.2488632

alternately. Hence, the inefficiency issue for training and the scalability issue for including outside images can be effectively solved, as can be observed from the results.

The letter is outlined as follows. First, Section II introduces the BELP framework. Section III shows the settings and results. Finally, the conclusion is drawn in Section IV.

II. The Proposed BELP Framework

A. Formulation

Given a set of training images $X = [x_1, x_2, \ldots, x_{N_T}] \in \mathbb{R}^{m\times n\times N_T}$ where $x_i \in \mathbb{R}^{m\times n}$ is an image, let $X_L = [x_1, x_2, \ldots, x_l] \in \mathbb{R}^{m\times n\times l}$ denote a labeled set and $X_U = [x_{l+1}, x_{l+2}, \ldots, x_{l+u}] \in \mathbb{R}^{m\times n\times u}$ be the unlabeled set, where l and u are the numbers of labeled and unlabeled images respectively, $N_T = l + u$ is the number of all training images. Let $F = [f_1, f_2, \ldots, f_{l+u}] \in \mathbb{R}^{(c+1)\times(l+u)}$ be the soft labels of all images and $Y = [y_1, y_2, \ldots, y_{l+u}] \in \mathbb{R}^{(c+1)\times(l+u)}$ denote the initial labels of the training images, similarly as SLP [6].

To enable the proposed criterion to handle images in matrix form directly, we propose to calculate two projections that can extract bilinear features from the images in the form of tensor representation $U^T x_i V$[13][16], and connect extracted features with image labels, which motivates us to define the following objective function for our BELP approach:

$$\begin{aligned}\underset{F,U,V}{Min}\ \widehat{J} = &\sum_{i=1}^{l+u}\left\| f_i - \sum_{j:x_j\in N(x_i)} W_{i,j} f_j \right\|_2^2 \\ &+ \sum_{i=1}^{l+u} \mu_i D_{ii} \left\| f_i - y_i \right\|_2^2, \\ &+ \alpha \sum_{i=1}^{l+u} \left\| U^T x_i V - f_i \right\|_2^2 + \beta \left(\|U\|_{2,1} + \gamma \|V\|_2^2 \right)\end{aligned} \tag{1}$$

where $N(x_i)$ is the neighbour set for image x_i, $U \in \mathbb{R}^{m\times(c+1)}$ and $V \in \mathbb{R}^{n\times 1}$ are two matrices that are used to extract bilinear features from the images, $\sum_{i=1}^{l+u} \|U^T x_i V - f_i\|_2^2$ is the cumulative reconstruction error over all training images measuring the difference between bilinear features $U^T x_i V$ and the estimated label f_i, $W_{i,j}$ are the coefficients for reconstructing x_i by its neighbors, D is a diagonal matrix with entries $D_{ii} = \sum_i W_{i,j}$, and ^{T} is the transpose of a matrix. Note that $U^T x_i V \in \mathbb{R}^{(c+1)\times 1}$ can approximate the soft label of each x_i by minimizing the reconstruction error. It is also worth noting that the $l_{2,1}$-norm regularization on U can ensure the computed U is sparse in rows so that discriminative features can be chosen in the latent subspace for predicting the soft labels of images and can also ensure the robustness to noise and outliers [19].

In Eq. (1), the weight matrix W is constructed similarly as the LLE-style reconstruction weights:

$$\tilde{w}_{i,j} = \sum\nolimits_{r=1}^{l+u} \chi_{jr}^{(i)} \Big/ \left(\sum\nolimits_{u=1}^{l+u} \sum\nolimits_{t=1}^{l+u} \chi_{ut}^{(i)} \right), \tag{2}$$

where $\chi^{(i)} = (\aleph^{(i)})^{-1}$, $\aleph_{jr}^{(i)} = (x_i - x_j)^T (x_i - x_r)$, x_j, x_r are neighbours of x_i. After obtaining $\tilde{W} = [\tilde{w}_{i,j}] \in \mathbb{R}^{N_T\times N_T}$, we symmetrize and normalize $\tilde{W}$ by $\tilde{W} = (\tilde{W} + \tilde{W}^T)/2$ and $W = \tilde{D}^{-1/2}\tilde{W}\tilde{D}^{-1/2}$, where $\tilde{D}_{ii} = \sum_j \widetilde{W}_{i,j}$. Note that normalizing the weights can greatly reduce the effects brought by various density distributions [3].

TABLE I
Bilinear Embedding Label Propagation

Input: Training dataset $X = \left[x_1, x_2, \ldots, x_{N_T}\right] \in \mathbb{R}^{m\times n\times N_T}$;
Neighbor size K, parameters μ_i, α, β and γ.
Initialization: $V = \left[1, 0, 0, \ldots, 0\right] \in \mathbb{R}^{n\times 1}$.

While not converged do

1. Construct the normalized reconstruction weights W and define the initial label assignment Y;
2. Fix others and update the left-projection U by Eq.(4);
3. Fix others and update the right-projection V by Eq.(5);
4. Fix others and update the soft label matrix F by Eq.(7);
5. Fix others and update $\hat{X} = \left[\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{l+u}\right]$ by $\hat{x}_i = U^T x_i V$;
6. Check for convergence: if $\left\|F^{t+1} - F^t\right\|_F \le \varepsilon$, stop; else $t=t+1$; ***End while***
7. Output the soft label matrix F, projections U and V.

TABLE II
List of Used Datasets and Dataset Descriptions

Data Name	Size	#Points	#Classes (c)	Labeled (l)
Yale-B	32×32	2414	38	1c, 2c,…, 10c
CMU PIE	32×32	1428	68	1c, 2c,…, 5c
Butterflies	32×32	619	7	1c, 2c,…, 10c
Leaves	32×32	186	3	1c, 2c,…, 10c
Birds	32×32	600	6	1c, 2c,…, 10c
Faces	32×32	408	19	1c, 2c,…, 6c

Based on the above definitions, the first term in Eq. (1) can measure the manifold smoothness of the outputted labels, the second fitting term can ensure that the model can receive label information from the initial state [6], and the third term is to measure the cumulative reconstruction error. The trade-offs of the terms are controlled by μ_i, α, β and γ.

B. Optimization

We first reformulate the criterion of BELP in matrix form as

$$\begin{aligned}\underset{F,U,V}{Min}\ \hat{J} = &tr\left(FLF^T\right) + tr\left((F-Y)QD(F-Y)^T\right) \\ &+ \alpha tr\left[\left(\hat{X}-F\right)\left(\hat{X}-F\right)^T\right] \\ &+ \beta\left(tr\left(U^T\Delta U\right) + \gamma\|V\|_2^2\right),\end{aligned} \tag{3}$$

where $L = (I - W - W^T + W^T W) = (I-W)^T(I-W)$ denotes the graph Laplacian, $\hat{X} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_{l+u}]$ with $\hat{x}_i = U^T x_i V \in \mathbb{R}^{(c+1)}$, Q is a diagonal matrix with entries being μ_i, $\Delta \in \mathbb{R}^{m\times m}$ is a diagonal matrix with entries being $1/2\|u^i\|_2$, $i = 1, \ldots, m$, and u^i is the i-th row vector of U. Parameter μ_l for each labeled image is set to 10^{10}, and parameter μ_u is set to 0 for unlabeled image.

We first compute U by fixing V. Because $\|A\|_2^2 = tr(AA^T)$, by setting the derivative of Eq. (3) w.r.t. U to zero, we can have

$$\partial\hat{J}/\partial U = 0 \Rightarrow U = \alpha(\alpha M_v + \beta\Delta)^{-1} M_F, \tag{4}$$

where $M_v = \sum_{i=1}^{l+u} x_i V V^T {x_i}^T$, $M_F = \sum_{i=1}^{l+u} x_i V {f_i}^T$, and the size of the matrix $M_v + \beta\Delta$ to be inversed is m by m.

After U is obtained at each iteration, we aim to compute V, Similarly by setting the derivative w.r.t. V to zero, we have

$$\partial\hat{J}/\partial V = 0 \Rightarrow V = \alpha(\alpha M_u + \beta\gamma I)^{-1}\hat{M}_F, \tag{5}$$

where $M_U = \sum_{i=1}^{l+u} {x_i}^T U U^T x_i$, $\hat{M}_F = \sum_{i=1}^{l+u} {x_i}^T U f_i$, and the size of the matrix $M_u + \beta\gamma I$ to be inversed is n by n.

TABLE III
COMPARISON OF MEAN AND HIGHEST ACCURACIES (%) ON IMAGE DATASETS (*THE BEST RECORDS ARE HIGHLIGHTED IN BOLD*)

Method / Data Name	SLP *Mean (%) Best (%)*	LNP *Mean (%) Best (%)*	GFHF *Mean (%) Best (%)*	LLGC1 *Mean (%) Best (%)*	LLGC2 *Mean (%) Best (%)*	ELP *Mean (%) Best (%)*	BELP *Mean (%) Best (%)*
Yale-B	42.29 59.90	41.91 59.56	36.14 56.58	35.51 56.68	33.29 53.23	52.07 **75.84**	**62.83** 75.64
CMU (lights change)	75.49 92.37	75.45 92.30	58.35 89.10	54.81 87.99	47.75 77.23	85.19 96.30	**99.99** **100.00**
Butterflies	27.66 34.98	25.19 32.90	27.11 34.95	28.17 35.02	29.16 35.57	26.61 33.39	**30.54** **38.77**
Caltech Leaves	**58.46** 67.42	48.72 59.10	57.30 67.42	57.55 67.38	56.86 66.13	53.94 62.62	58.04 **68.28**
Birds	29.92 36.11	24.35 30.69	28.91 35.60	29.27 36.41	29.49 36.23	26.36 30.54	**31.98** **39.40**
Caltech Faces	38.21 52.42	35.02 51.36	38.03 52.58	38.42 52.68	38.12 52.39	39.41 53.25	**49.03** **66.00**

When both U and V are computed, we aim at updating F. By removing terms that independent of F in Eq. (3), we can get

$$F = \arg\min_F tr\left(FLF^T\right) + tr\left((F-Y)QD(F-Y)^T\right) + \alpha tr\left[\left(\hat{X}-F\right)\left(\hat{X}-F\right)^T\right]. \quad (6)$$

Through taking the derivative $w.r.t$ variable F and setting it to zero, we can update F at each iteration as

$$F = \left(YQD + \alpha\hat{X}\right)(L + QD + \alpha I)^{-1}, \quad (7)$$

where the matrix $L + QD + \alpha I$ that will be inversed is of size $N_T \times N_T$. For complete presentation of the idea, we summarize the proposed BELP algorithm in Table I.

C. Inclusion of Outside Images by BELP

We discuss the proposed BELP to involve outside images. After gaining U and V by transductive training, we can predict the soft label $f_{new} = U^T x_{new} V \in \mathbb{R}^{(c+1)}$ of each new $x_{new} \in \mathbb{R}^{m\times n}$ efficiently by embedding it onto the spanned bilinear subspace by U and V. Then, the hard label of each x_{new} can be similarly obtained by locating the position according to the largest entry in f_{new}. Note that this inductive process can also be considered as a dimension reduction process to reduce the dimension of each image from $m \times n$ to $(c+1) \times 1$.

III. SIMULATION RESULTS AND ANALYSIS

A. Experiment Setup

We mainly compare our BELP with SLP, LLGC, LNP, GFHF and ELP that are closely related to our present work. For fair comparison and avoiding estimating the kernel parameter, we use the LLE-reconstruction weights for each model. The size of neighbourhood K is set to 7 for all simulations. The ratio of the number of training images to the total number of images is set to 0.3 for each dataset. For inclusion of the outside images, ELP and BELP use their own projections, while other models apply the inductive scheme by label reconstruction.

For LLGC [5], the authors proposed an extra extension, so we test them separately as LLGC1 and LLGC2. For ELP, the two model parameters in the objective function of ELP are set to $\alpha = 10^{-6}$ and $\beta = 10^6$ as suggested by [11]. For our BELP, we choose the parameters α, β and γ using grid search from $\{10^{-6}, 10^{-4}, 10^{-2}, 10^0, 10^2, 10^4, 10^6\}$. All methods are implemented in Matlab R2014a, and we run all the simulations on a PC with Intel(R) Core(TM) i5-4590 CPU@3.30 GHz 3.30 GHz 8G.

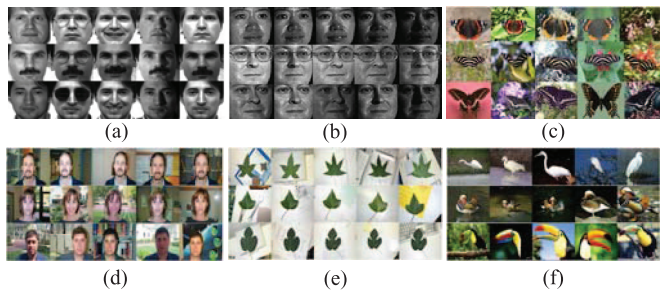

Fig. 1. Typical image examples of the image datasets. (a) Yale-B (b) CMU PIE (c) Butterflies (d) Caltech Faces (e) Caltech Leaves (f) Birds.

B. Data Preparation

In this study, we test Yale-B [13], Butterflies [15], CMU PIE faces (lights change) [12], Caltech Faces 1999, Caltech Leaves 1999 and Birds [18] image databases. The dataset descriptions are shown in Table II. Fig. 1 shows some image examples of the datasets. The Butterflies and Birds images are provided at http://www-cvr.ai.uiuc.edu/ponce_grp/data/. Faces and Leaves images are available from http://www.vision.caltech.edu/html-files/ archive.html. For CMU PIE, we consider various lighting conditions, thus pose & expression are fixed and there are 21 images in each subject. All the images are resized into 32×32 pixels, so each image corresponds to a 1024-dimensional data vector in VSM for SLP, LLGC, LNP, GFHF and ELP.

C. Recognition Results

We describe the classification results on Yale-B, CMU (lights change), Butterflies, Birds, Leaves and Faces image databases in Table III. The averaged and highest accuracy of each method are reported under different numbers (l) of labeled images, as shown in Table II. For each l on certain image set, we average the results over 15 random splits of training/test set.

From the results, we find that our BELP gains a promising and stable superiority performance, i.e., it obtains comparable or even higher accuracy than the other models in most cases, especially on Yale-B, CMU PIE and Faces image databases. Although BELP has a lower accuracy on Yale-B than ELP and delivers a lower accuracy on Caltech Leaves than SLP, the gap between existing methods and our BELP is very small.

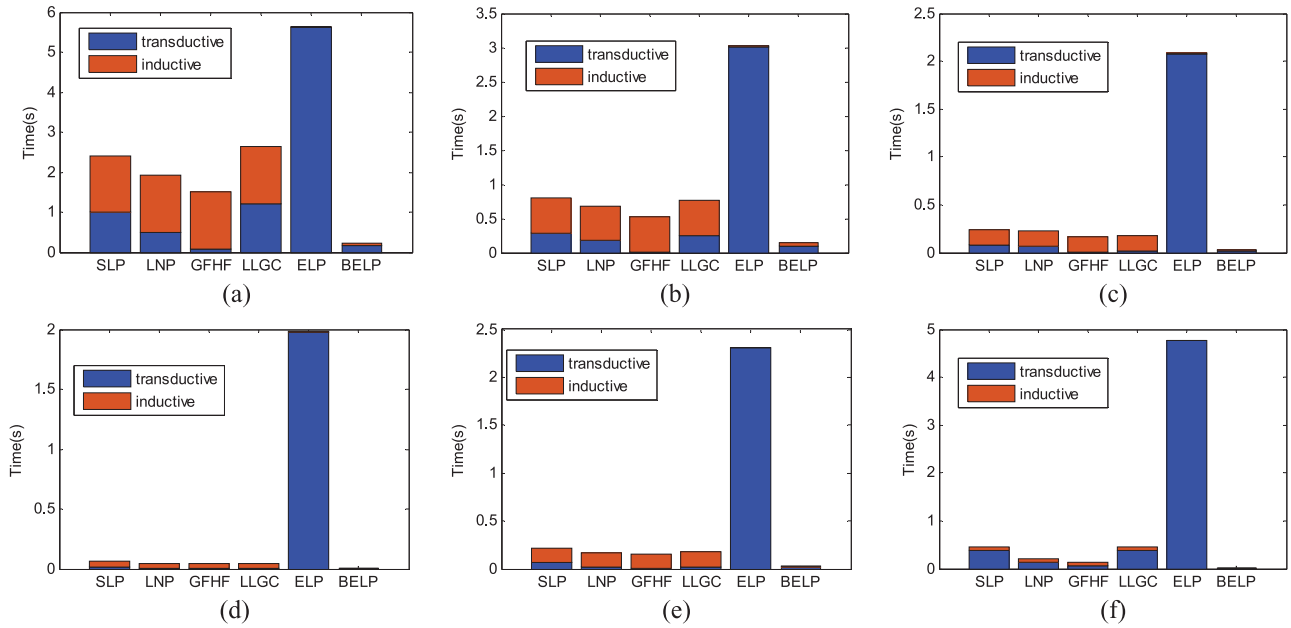

Fig. 2. The visualizations of needed time in training (transductive) and testing (inductive) phases. (a) Yale-B (b) CMU (lights change) (c) Butterflies (d) Caltech Leaves (e) Birds (f) Caltech Faces.

TABLE IV
TOTAL TIME CONSUMPTION RESULTS (S) OF THE WHOLE STAGE (TRAINING AND TESTING)

Data Name \ Method	SLP	LNP	LLGC	GFHF	ELP	BELP
Yale-B	2.4164	1.9197	1.5064	2.6457	5.6405	**0.2341**
CMU (lights change)	0.8143	0.7080	0.5385	0.7802	3.1692	**0.1425**
Butterflies	0.2365	0.2243	0.1620	0.1833	2.0819	**0.0256**
Caltech Leaves	0.0606	0.0421	0.0416	0.0450	1.9808	**0.0067**
Birds	0.2227	0.1669	0.1563	0.1770	2.3057	**0.0263**
Caltech Faces	0.4552	0.2000	0.1314	0.4650	4.7695	**0.0194**

D. Comparison of Computational Time

We also explore the computational complexity of each method by describing the actual running time in Fig. 2 and Table IV. Fig. 2 illustrates the needed time for both transductive and inductive phases of each algorithm on several datasets. Table IV describes the actual run time of the whole stage, including both training and testing. From the results, we can find that: (1) The embedding based inductive methods (i.e., ELP and BELP) clearly outperform the reconstruction based inductive scheme (that is, GFHF, LLGC, LNP and SLP) for inclusion of outside images. For transductive classification, our presented BELP also has a very acceptable computational complexity compared with other existing models, especially by comparing with the most related inductive ELP method, which is mainly due to the proposed BELP is based on the 2D matrices of images rather than being performed in the high-dimensional vector space. It is also note that the needed time of the vector space based ELP is far slower than the transductive methods, e.g., GFHF, LLGC, LNP and SLP for transductive learning. In addition, since we use the same LLE-reconstruction weights for GFHF, LLGC, LNP and SLP, the runtime performances of them are comparable with each other in most cases.

From Fig. 2, we can see that our BELP is not the fastest model in either transductive or inclusion phase independently, but it should be noted that it has the shortest add-up time for the whole classification stage, as can be clearly observed from Table IV. Generally, the total time consumption of our BELP is about half of other LP algorithms. One reasonable explanation is that BELP works in the 2D matrix space and has a faster transductive convergence speed (Note that we will not show the convergence analysis due to page limitation). The superior efficiency of our BELP to all the other reconstruction based inductive models is because they need a separable neighbor search process for each new data. The runtime performance of our BELP is also superior to the VSM based inductive ELP, since ELP uses high-dimensional vectorized representations of the images for training, implying that 2D model will be faster than 1D model especially for images of large sizes.

IV. CONCLUDING REMARKS

We proposed a scalable bilinear embedding label propagation method which is mainly motivated by two facts: (1) The direct embedding based inductive scheme is more efficient than the label reconstruction based one for including outside images; (2) The existing embedding based scheme is performed in the 1D vector space. i.e., it has to convert 2D images into the 1D vectors when handling images, but such transformation may destroy the topology structures of image pixels and also it will be not efficient due to the high dimensionality of vectors. To preserve the topology and for efficiency, we present a scalable bilinear embedding label propagation that uses 2D matrices as inputs and learn an informative bilinear subspace spanned by two projections. The inclusion for the outside images by our method is also shown. Results of our investigated cases show our algorithm outperforms other related criteria in efficiency. More importantly, our model is shown to be the fastest one in terms of total computational time due to the power of handling 2D matrices directly by embedding. In future, we will explore a more natural 2D formulation for regression problem [25].

REFERENCES

[1] X. Zhu, Semi-supervised learning literature survey Univ. Wisconsin-Madison, Madison, WI, USA, Tech. Rep. 1530, 2005.

[2] Z. Zhang, M. B. Zhao, and T. W. S. Chow, "Graph based constrained semi-supervised learning framework via label propagation over adaptive neighborhood," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 9, pp. 2362–2376, Sep. 2015.

[3] X. Zhu and Z. Ghahramani, Learning from labeled and unlabeled data with label propagation Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. CMU-CALD-02-107, 2002.

[4] X. Zhu, Z. Ghahramani, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," in *Proc. Int. Conf. Machine Learning*, 2003.

[5] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. S. cholkopf, "Learning with local and global consistency," in *Advances in Neural Information Processing Systems*, 2004.

[6] F. P. Nie, S. M. Xiang, Y. Liu, and C. S. Zhang, "A general graph-based semi-supervised learning with novel class discovery," *Neural Comput. Applicat.*, vol. 19, no. 4, pp. 549–555, 2010.

[7] F. Wang and C. S. Zhang, "Label propagation through linear neighborhoods," *IEEE Trans. Knowl. Data Eng.*, vol. 20, no. 1, pp. 55–67, 2008.

[8] F. Wang and C. S. Zhang, "Label propagation through linear neighborhoods," in *Proc. Int. Conf. Machine Learning*, Pittsburgh, PA, USA, 2006.

[9] F. P. Nie, S. M. Xiang, Y. Q. Jia, and C. S. Zhang, "Semi-supervised orthogonal discriminant analysis via label propagation," *Patt. Recognit.*, vol. 42, no. 11, pp. 2615–2627, 2009.

[10] G. Salton, A. Wong, and C. Yang, "A vector space model for automatic indexing," *Commun. ACM*, vol. 18, no. 11, pp. 613–620, 1975.

[11] X. Chang, F. P. Nie, Y. Yang, and H. Huang *et al.*, "Improved spectral clustering via embedded label propagation," *arXiv preprint arXiv:1411.6241*, 2014.

[12] X. F. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *NIPS*, 2005.

[13] Z. Zhang and T. W. Chow, "Tensor locally linear discriminative analysis," *IEEE Signal Process. Lett.*, vol. 18, no. 1, pp. 643–646, 2011.

[14] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, 2001.

[15] S. Lazebnik, C. Schmid, and J. Ponce, "Semi-local affine parts for object recognition," in *Brit. Machine Vision Conf.*, 2004.

[16] Z. Zhang and T. Chow, "Maximum margin multisurface support tensor machines with application to image classification and segmentation," *Exp. Syst. Applicat.*, vol. 39, no. 1, pp. 850–861, Jan. 2012.

[17] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[18] S. Lazebnik, C. Schmid, and J. Ponce, "A maximum entropy framework for part-based texture and object recognition," in *ICCV*, Beijing, China, 2005, vol. 1, pp. 832–838.

[19] C. P. Hou, F. P. Nie, X. L. Li, D. Y. Yi, and Y. Wu, "Joint embedding learning and sparse regression: A framework for unsupervised feature selection," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 793–804, 2014.

[20] Z. Zhang, L. Zhang, M. B. Zhao, W. M. Jiang, Y. C. Liang, and F. Z. Li, "Semi-supervised image classification by nonnegative sparse neighborhood propagation," in *Proc. ACM Int. Conf. Multimedia Retrieval*, Shanghai, China, Jun. 2015.

[21] Z. Zhang, T. W. S. Chow, and M. B. Zhao, "Trace ratio optimization based semi-supervised nonlinear dimensionality reduction for marginal manifold visualization," *IEEE Trans. Knowl. Data Eng.*, vol. 25, no. 5, pp. 1148–1161, May 2013.

[22] F. P. Nie, D. Xu, I. W. Tsang, and C. S. Zhang, "Flexible manifold embedding: A framework for semi-supervised and unsupervised dimension reduction," *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1921–1932, 2010.

[23] F. P. Nie, D. Xu, X. L. Li, and S. M. Xiang, "Semi-supervised dimensionality reduction and classification through virtual label regression," *IEEE Trans. Syst., Man, Cybern. B: Cybern.*, vol. 41, no. 3, pp. 675–685, 2011.

[24] F. P. Nie, H. Wang, H. Huang, and C. Ding, "Adaptive loss minimization for semi-supervised elastic embedding," in *Proc. 23rd Int. Joint Conf. Artificial Intelligence*, 2013.

[25] C. P. Hou, F. P. Nie, D. Y. Yi, and Y. Wu, "Efficient image classification via multiple rank regression," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 340–352, 2013.