Semi-Supervised Subspace Clustering via Tensor Low-Rank Representation

Yuheng Jia, Guanxing Lu, Hui Liu, Junhui Hou, *Senior Member, IEEE*

Abstract—In this letter, we propose a novel semi-supervised subspace clustering method, which is able to simultaneously augment the initial supervisory information and construct a discriminative affinity matrix. By representing the limited amount of supervisory information as a pairwise constraint matrix, we observe that the ideal affinity matrix for clustering shares the same low-rank structure as the ideal pairwise constraint matrix. Thus, we stack the two matrices into a 3-D tensor, where a global low-rank constraint is imposed to promote the affinity matrix construction and augment the initial pairwise constraints synchronously. Besides, we use the local geometry structure of input samples to complement the global low-rank prior to achieve better affinity matrix learning. The proposed model is formulated as a Laplacian graph regularized convex low-rank tensor representation problem, which is further solved with an alternative iterative algorithm. In addition, we propose to refine the affinity matrix with the augmented pairwise constraints. Comprehensive experimental results on eight commonly-used benchmark datasets demonstrate the superiority of our method over state-of-the-art methods. The code is publicly available at https://github.com/GuanxingLu/Subspace-Clustering.

Index Terms—tensor low-rank representation, semi-supervised learning, subspace clustering, pairwise constraints.

I. INTRODUCTION

High-dimensional data are ubiquitously in many areas like image processing, DNA microarray technology, etc. The high-dimensional data can often be well approximated by a set of linear subspaces, but the subspace membership of a certain sample is unknown [1], [2]. Subspace clustering aims to divide the data samples into different subspaces, which is an important tool to model the high-dimensional data. The state-of-the-art subspace clustering methods [3], [4] are based on self-expressiveness, which represent high-dimensional data by the linear combination of itself, and enforce a subspace-preserving prior on the self-representation matrix. The representation coefficients capture the global geometric relationships of samples and can act as an affinity matrix, and the subspace segmentation can be obtained by applying spectral

Fig. 1: Illustration of the proposed method, which adaptively learns the affinity and enhances the pairwise constraints simultaneously by using their identical global low-rank structure.

clustering on the generated affinity matrix. The most well-known subspace clustering methods include sparse subspace clustering [3] and low-rank representation [4].

In many real-world applications, some supervisory information is available, e.g., the label information of a dataset, and the relationships between two samples. Generally, those supervisory information can be represented by two kinds of pairwise constraints, i.e., must-link constraints and cannot-link constraints indicating whether two samples belong to the same category or not. As the supervisory information is widespread and provides a discriminative description of the samples, many semi-supervised subspace clustering methods [5]–[13] were proposed to incorporate them. Based on the type of the supervisory information, we roughly divide these methods into three classes. The first kind of methods include the must-link constraints. For example, [5], [6] incorporated the must-links as hard constraints, which restricted the samples with a must-link to have exactly the same representation. The second kind of methods integrate the cannot-link constraints. For instance, [7] required the affinity relationship of two samples with a cannot-link to be 0. Liu *et al.* [8] first enhanced the initial cannot-links by a graph Laplacian term, and then inhibited the affinity of two samples with a cannot-link. The third kind of methods can incorporate both the must-links and cannot-links, which generally assume two samples with a must-link should have a higher affinity, while those with a cannot-link should have a lower affinity [9]–[13]. However, the above-mentioned semi-supervised subspace clustering methods exploit the supervisory information from a local perspective, but overlook the global structure of the pairwise constraints, which is also important to semi-supervised affinity matrix learning. In other words, the previous methods under-use the supervisory information to some extent.

To this end, we propose a new semi-supervised subspace clustering method shown in Fig. 1, which explores the supervisory information from a global manner. Specifically, in the ideal case, the pairwise constraint matrix is low-rank,

as if all the pairwise relationships of samples are available, we could encode the pairwise constraint matrix as a binary low-rank matrix. Meanwhile, the ideal affinity matrix is also low-rank, as a sample should only be represented by the samples from the same class. More importantly, they share an identical low-rank structure. Based on such an observation, we stack them into a 3-dimensional tensor, and regulate a global low-rank constraint to the formed tensor. By seeking the tensor low-rank representation, we can refine the affinity matrix with the available pairwise constraints, and at the same time, augment the initial pairwise constraints with the learned affinity matrix. Besides, we encode the local geometry structure of the data samples to complement the global low-rank prior. The proposed model is formulated as a convex optimization model, which can be solved efficiently. Finally, we use the augmented pairwise constraint matrix to further refine the affinity matrix. Extensive experiments on 8 datasets w.r.t. 2 metrics demonstrate that our method outperforms the state-of-the-art semi-supervised subspace clustering methods to a large extent.

II. PRELIMINARY

In this letter, we denote tensors by boldface Euler script letters, e.g., $\mathcal{A}$, matrices by boldface capital letters, e.g., $\mathbf{A}$, vectors by boldface lowercase letters, e.g., $\mathbf{a}$, and scalars by lowercase letters, e.g., a. $\|\cdot\|_{2,1}$, $\|\cdot\|_{\infty}$, $\|\cdot\|_F$ and $\|\cdot\|_*$ are the $\ell_{2,1}$ norm, the infinity norm, and the Frobenius norm, and the nuclear norm (i.e., the sum of the singular values) of a matrix. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{d \times n}$ is the data matrix, where $\mathbf{x}_i \in \mathbb{R}^{d \times 1}$ specifies the d-dimensional vectorial representation of the i-th sample, and n is the number of samples. Let $\boldsymbol{\Omega}_m = \{(i,j) \mid \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the same class}\}$ and $\boldsymbol{\Omega}_c = \{(i,j) \mid \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to different classes}\}$ stand for the available must-link set and cannot-link set. We can encode the pairwise constraints as a matrix $\mathbf{B} \in \mathbb{R}^{n \times n}$:

$$\mathbf{B}_{ij} = \begin{cases} 1, & \text{if } (i,j) \in \boldsymbol{\Omega}_m \\ -1, & \text{if } (i,j) \in \boldsymbol{\Omega}_c. \end{cases} \quad (1)$$

Subspace clustering aims to segment a set of samples into a group of subspaces. Particularly, self-expressive-based subspace clustering methods have attracted great attention, which learn a self-representation matrix to act as the affinity. For example, Liu et al. [4] proposed to learn a low-rank affinity matrix by optimizing

$$\min_{\mathbf{Z}, \mathbf{E}} \quad \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \quad \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \quad (2)$$

where $\mathbf{Z} \in \mathbb{R}^{n \times n}$ is the representation matrix, $\mathbf{E} \in \mathbb{R}^{d \times n}$ denotes the reconstruction error, and $\lambda > 0$ is a hyper-parameter.

Recently, many semi-supervised subspace clustering methods were proposed by incorporating the pairwise constraints [5]–[13]. How to include the supervisory information is the crux of semi-supervised subspace clustering. *Generally, the existing methods incorporate the pairwise constraints from a local perspective, i.e., expand (resp. reduce) the value of $\mathbf{Z}_{ij}$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ has a must-link (resp. cannot-link).*

Algorithm 1 Solve Eq. (4) by ADMM

Input: data $\mathbf{X}$, pairwise constraints $\boldsymbol{\Omega}_m, \boldsymbol{\Omega}_c$, hyper-parameters λ, β.

1: Initialize: $\mathcal{C}^{(0)} = \mathcal{Y}_2^{(0)} = 0$, $\mathbf{B}^{(0)} = \mathbf{Z}^{(0)} = \mathbf{D}^{(0)} = \mathbf{E}^{(0)} = \mathbf{Y}_1^{(0)} = \mathbf{Y}_3^{(0)} = 0$, $\rho = 1.1$, $\mu^{(0)} = 1\mathrm{e}{-3}$, $\mu_{\max} = 1\mathrm{e}10$.

2: **repeat**

3:    Update $\mathcal{C}$ by $\mathcal{C}^{(k+1)} = \mathcal{S}_{\frac{1}{\mu^{(k)}}}(\mathcal{M}^{(k)} + \mathcal{Y}_2^{(k)}/\mu^{(k)})$, where $\mathcal{S}$ is the tensor singular value thresholding operator [14];

4:    Update $\mathbf{Z}$ by $\mathbf{Z}^{(k+1)} = (\mathbf{I} + \mathbf{X}^\top \mathbf{X})^{-1}(\mathbf{X}^\top(\mathbf{X} - \mathbf{E}^{(k)}) + \mathcal{C}^{(k)}(:,:,2) + (\mathbf{X}^\top \mathbf{Y}_1^{(k)} - \mathcal{Y}_2^{(k)}(:,:,2))/\mu^{(k)})$;

5:    Update $\mathbf{B}$ by $\mathbf{B}^{(k+1)} = (\mu^{(k)}(\mathcal{C}^{(k)}(:,:,2) + \mathbf{D}^{(k)}) - (\mathcal{Y}_2^{(k)}(:,:,2) + \mathbf{Y}_3^{(k)}))/(\beta(\mathbf{L} + \mathbf{L}^\top) + 2\mu^{(k)}\mathbf{I})$;

6:    Update $\mathbf{D}$ by
$$\mathbf{D}_{ij}^{(k+1)} = \begin{cases} s, & \text{if } (i,j) \in \boldsymbol{\Omega}_m \\ -s, & \text{if } (i,j) \in \boldsymbol{\Omega}_c \\ \mathbf{B}_{ij}^{(k)} + \mathbf{Y}_{3ij}^{(k)}/\mu^{(k)}, & \text{otherwise;} \end{cases}$$

7:    Update $\mathbf{E}$ by
$$\mathbf{e}_j^{(k+1)} = \begin{cases} \dfrac{\left\|\mathbf{q}_j^{(k)}\right\|_2 - \lambda/\mu^{(k)}}{\left\|\mathbf{q}_j^{(k)}\right\|_2}\mathbf{q}_j^{(k)}, & \text{if } \left\|\mathbf{q}_j^{(k)}\right\|_2 \geq \lambda/\mu^{(k)} \\ 0, & \text{otherwise;} \end{cases}$$

8:    Update $\mathbf{Y}_1, \mathcal{Y}_2, \mathbf{Y}_3$, and μ by
$$\begin{cases} \mathbf{Y}_1^{(k+1)} = \mathbf{Y}_1^{(k)} + \mu^{(k)}(\mathbf{X} - \mathbf{XZ}^{(k+1)} - \mathbf{E}^{(k+1)}) \\ \mathcal{Y}_2^{(k+1)}(:,:,1) = \mathcal{Y}_2^{(k)}(:,:,1) + \mu^{(k)}(\mathbf{Z}^{(k+1)} - \mathcal{C}^{(k+1)}(:,:,1)) \\ \mathcal{Y}_2^{(k+1)}(:,:,2) = \mathcal{Y}_2^{(k)}(:,:,2) + \mu^{(k)}(\mathbf{B}^{(k+1)} - \mathcal{C}^{(k+1)}(:,:,2)) \\ \mathbf{Y}_3^{(k+1)} = \mathbf{Y}_3^{(k)} + \mu^{(k)}(\mathbf{B}^{(k+1)} - \mathbf{D}^{(k+1)}) \\ \mu^{(k+1)} = \min(\rho\mu^{(k)}; \mu_{\max}); \end{cases}$$

9: **until** convergence

III. PROPOSED METHOD

A. Model Formulation

As aforementioned, existing semi-supervised subspace clustering methods usually impose the pairwise constraints on the affinity matrix in a simple element-wise manner, which underuses the supervisory information to some extent. As studied in previous works [4], [8], the ideal affinity matrix $\mathbf{Z}$ is low-rank as a sample should be only reconstructed by the samples within the same class. Meanwhile, the ideal pairwise constraint matrix $\mathbf{B}$ is also low-rank, as it records the pairwise relationship among samples. Moreover, we observe that their low-rank structures should be identical. Accordingly, if we stack them to form a 3-D tensor $\mathcal{C} \in \mathbb{R}^{n \times n \times 2}$, i.e., $\mathcal{C}(:,:,1) = \mathbf{Z}$, and $\mathcal{C}(:,:,2) = \mathbf{B}$, the formed tensor $\mathcal{C}$ should be low-rank ideally. Therefore, we use a global tensor low-rank norm to exploit this prior and preliminarily formulate the problem as

$$\min_{\mathcal{C}, \mathbf{E}, \mathbf{B}, \mathbf{Z}} \quad \|\mathcal{C}\|_{\circledast} + \lambda\|\mathbf{E}\|_{2,1}$$
$$\text{s.t.} \quad \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \mathcal{C}(:,:,1) = \mathbf{Z}, \mathcal{C}(:,:,2) = \mathbf{B}, \quad (3)$$
$$\mathbf{B}_{ij} = s, (i,j) \in \boldsymbol{\Omega}_m, \mathbf{B}_{ij} = -s, (i,j) \in \boldsymbol{\Omega}_c.$$
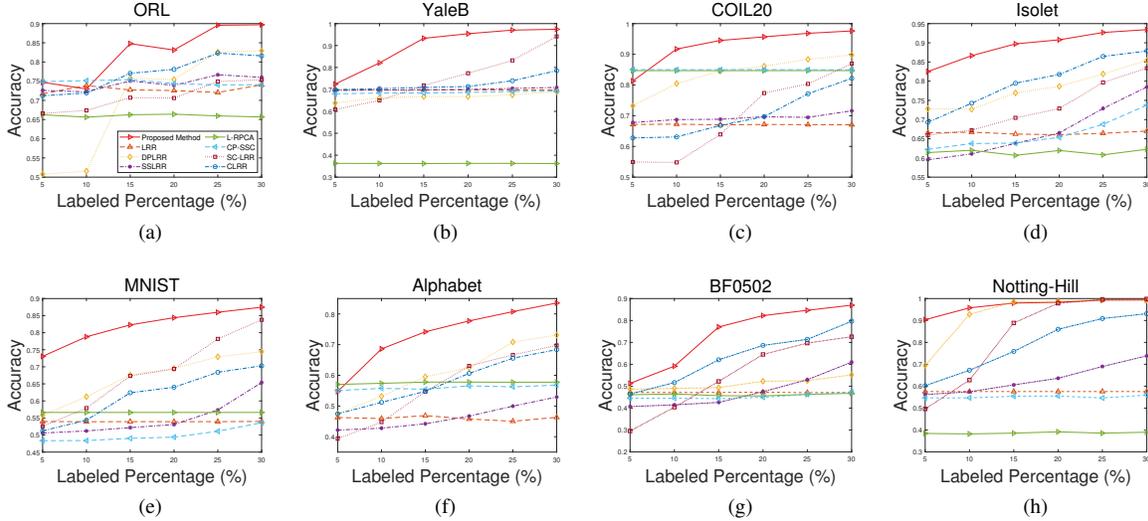
Fig. 2: Comparisons of the accuracy of all methods under various supervisory information. All subfigures share the same legend.
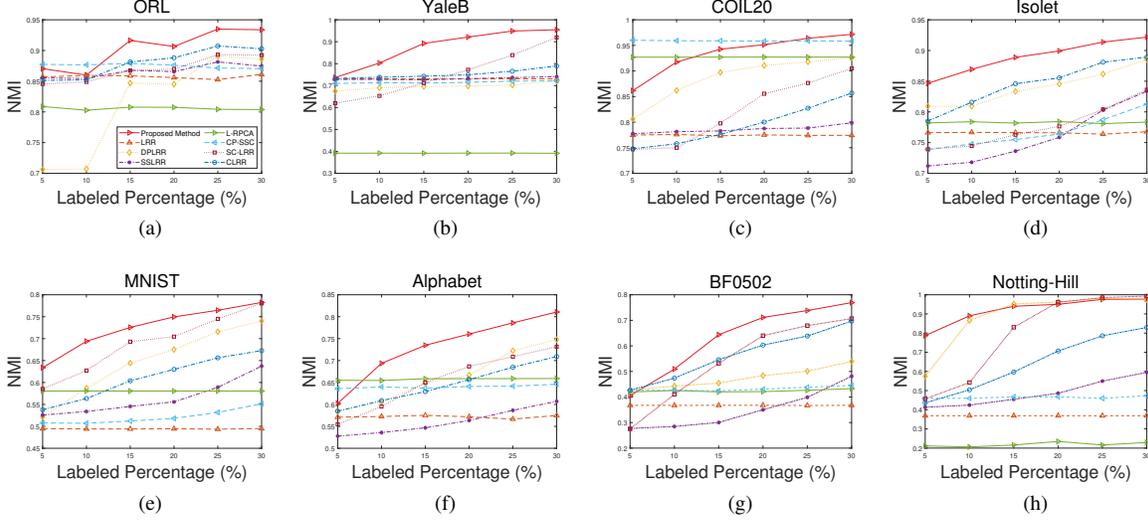
Fig. 3: Comparisons of the NMI of all methods under various supervisory information. All subfigures share the same legend.

TABLE I: Detailed Comparisions of Accuracy and NMI under 30% Initial Labels.

Accuracy	ORL	YaleB	COIL20	Isolet	MNIST	Alphabet	BF0502	Notting-Hill	Average
LRR	0.7405	0.6974	0.6706	0.6699	0.5399	0.4631	0.4717	0.5756	0.6036
DPLRR	0.8292	0.6894	0.8978	0.8540	0.7442	0.7309	0.5516	0.9928	0.7862
SSLRR	0.7600	0.7089	0.7159	0.7848	0.6538	0.5294	0.6100	0.7383	0.6876
L-RPCA	0.6568	0.3619	0.8470	0.6225	0.5662	0.5776	0.4674	0.3899	0.5612
CP-SSC	0.7408	0.6922	0.8494	0.7375	0.5361	0.5679	0.4733	0.5592	0.6445
SC-LRR	0.7535	0.9416	0.8696	0.8339	0.8377	0.6974	0.7259	**0.9982**	0.8322
CLRR	0.8160	0.7853	0.8217	0.8787	0.7030	0.6837	0.7964	0.9308	0.8020
Proposed Method	**0.8965**	**0.9742**	**0.9761**	**0.9344**	**0.8747**	**0.8355**	**0.9934**	0.9934	**0.9193**
NMI	ORL	YaleB	COIL20	Isolet	MNIST	Alphabet	BF0502	Notting-Hill	**Average**
LRR	0.8611	0.7309	0.7742	0.7677	0.4949	0.5748	0.3675	0.3689	0.6175
DPLRR	0.8861	0.7205	0.9258	0.8853	0.7400	0.7477	0.5388	0.9748	0.8024
SSLRR	0.8746	0.7409	0.7986	0.8337	0.6373	0.6070	0.4810	0.5949	0.6960
L-RPCA	0.8038	0.3914	0.9271	0.7834	0.5805	0.6590	0.4329	0.2294	0.6009
CP-SSC	0.8705	0.7224	0.9583	0.8127	0.5516	0.6459	0.4453	0.4733	0.6850
SC-LRR	0.8924	0.9197	0.9048	0.8362	0.7803	0.7316	0.7068	**0.9931**	0.8456
CLRR	0.9028	0.7895	0.8568	0.8892	0.6727	0.7091	0.6970	0.8293	0.7933
Proposed Method	**0.9337**	**0.9548**	**0.9716**	**0.9218**	**0.7825**	**0.8107**	**0.7693**	0.9771	**0.8902**

In Eq. (3), we adopt the nuclear norm $\|\cdot\|_\circledast$ defined on tensor-SVD [14] to seek the low-rank representation, and other kinds of tensor low-rank norms are also applicable, e.g., [15]. We introduce a scalar s to constrain the maximum and minimum values of $\mathbf{B}$, promoting that $\mathbf{B}$ has a similar scale to $\mathbf{Z}$.

Empirically, s is set to the largest element of the learned affinity by LRR. By solving Eq. (3), the affinity matrix $\mathbf{Z}$ and pairwise constraint matrix $\mathbf{B}$ are jointly optimized according to the nuclear norm on $\mathcal{C}$, i.e., the supervisory information is transferred to $\mathbf{Z}$, and at the same time, the learned affinity

TABLE II: Ablation Study.

Percentage	Accuracy	ORL	YaleB	COIL20	Isolet	MNIST	Alphabet	BF0502	Notting-Hill	**Average**
	SSLRR	0.7223	0.6965	0.6874	0.6107	0.5121	0.4278	0.4150	0.5747	0.5808
	CLRR	0.7193	0.7032	0.6309	0.7424	0.5435	0.5120	0.5165	0.6728	0.6301
10	Eq. (3)	0.7298	0.7838	0.6744	0.8599	0.5224	0.5022	0.5786	0.8079	0.6824
	Eq. (4)	0.7298	0.7838	0.8708	0.8424	0.7659	0.6640	0.5779	0.9573	0.7740
	Eq. (5)	**0.7523**	**0.8696**	**0.9171**	**0.8665**	**0.7879**	**0.6862**	**0.5915**	**0.9576**	**0.8036**
	SSLRR	0.7390	0.6998	0.6966	0.6651	0.5308	0.4672	0.4750	0.6363	0.6137
	CLRR	0.7808	0.7130	0.6971	0.8176	0.6401	0.6064	0.6863	0.8598	0.7251
20	Eq. (3)	0.7860	0.9194	0.8101	0.9012	0.6661	0.6443	0.7554	0.9378	0.8025
	Eq. (4)	0.7860	0.9194	0.9364	0.9065	0.8366	0.7511	0.8077	0.9817	0.8657
	Eq. (5)	**0.8325**	**0.9548**	**0.9569**	**0.9078**	**0.8439**	**0.7772**	**0.8223**	**0.9831**	**0.8848**
	SSLRR	0.7600	0.7089	0.7159	0.7848	0.6538	0.5294	0.6100	0.7383	0.6876
	CLRR	0.8160	0.7853	0.8217	0.8787	0.7030	0.6837	0.7964	0.9308	0.8020
30	Eq. (3)	0.8893	0.9664	0.9096	0.9222	0.8370	0.7671	0.8083	0.9661	0.8832
	Eq. (4)	0.8893	0.9664	0.9710	0.9300	0.8745	0.8244	0.8631	0.9917	0.9138
	Eq. (5)	**0.8965**	**0.9742**	**0.9761**	**0.9344**	**0.8747**	**0.8355**	**0.8697**	**0.9934**	**0.9193**

matrix can also augment the initial pairwise constraints from a global perspective.

Besides, if two samples ($\mathbf{x}_i$ and $\mathbf{x}_j$) are close to each other in the feature space, we can expect they have a similar pairwise relationship, i.e., $\mathbf{B}(:,i)$ is close to $\mathbf{B}(:,j)$. To encode this prior, we first construct a kNN graph $\mathbf{W} \in \mathbb{R}^{n \times n}$ to capture the local geometric structure of samples, and use the local Laplacian regularization $\mathrm{Tr}(\mathbf{BLB}^\top)$ to replenish the global low-rank term, where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the Laplacian matrix with $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$ [16]. Therefore, our model is finally formulated as

$$\min_{\mathcal{C},\mathbf{E},\mathbf{B},\mathbf{Z}} \quad \|\mathcal{C}\|_{\circledast} + \lambda\|\mathbf{E}\|_{2,1} + \beta\,\mathrm{Tr}(\mathbf{BLB}^\top)$$
$$\text{s.t.} \quad \mathbf{X} = \mathbf{XZ} + \mathbf{E}, \mathcal{C}(:,:,1) = \mathbf{Z}, \mathcal{C}(:,:,2) = \mathbf{B}, \quad (4)$$
$$\mathbf{B}_{ij} = s, (i,j) \in \mathbf{\Omega}_m, \mathbf{B}_{ij} = -s, (i,j) \in \mathbf{\Omega}_c.$$

After solving Eq. (4), we first normalize each column of $\mathbf{Z}$ to $[0,1]$, and normalize $\mathbf{B}$ by $\mathbf{B} \leftarrow \mathbf{B}/s$. Then, we use the augmented pairwise constraint matrix $\mathbf{B}$ to repair $\mathbf{Z}$, i.e.,

$$\mathbf{Z}_{ij} \leftarrow \begin{cases} 1 - (1 - \mathbf{B}_{ij})(1 - \mathbf{Z}_{ij}), & \text{if } \mathbf{B}_{ij} \geq 0 \\ (1 + \mathbf{B}_{ij})\mathbf{Z}_{ij}, & \text{if } \mathbf{B}_{ij} < 0. \end{cases} \quad (5)$$

When $\mathbf{B}_{ij}$ is larger than 0, $\mathbf{x}_i$ and $\mathbf{x}_j$ are likely to belong to the same class, Eq. (5) will increase the corresponding element of $\mathbf{Z}$. Similarly, when $\mathbf{B}_{ij}$ is less than 0, $\mathbf{Z}_{ij}$ will be depressed. Therefore, Eq. (5) further enhances the affinity matrix by the augmented pairwise constraints. Finally, we apply spectral clustering [17] on $\mathbf{W} = (|\mathbf{Z}| + |\mathbf{Z}^\top|)/2$ to get the subspace segmentation.

B. Optimization Algorithm

As Eq. (4) contains multiple variables and constraints, we solve it by alternating direction method of multipliers (ADMM) [18]. Algorithm 1 summarizes the whole pseudo code. Due to the page limitation, the detailed derivation process can be found in the supplementary file.

The computational complexity of Algorithm 1 is dominated by steps 3-5. Specifically, the step 3 solves the t-SVD of an $n \times n \times 2$ tensor with the complexity of $\mathcal{O}(2n^2 \log 2 + 2n^3)$ [14]. The steps 4-5 involve matrix inverse and matrix multiplication operations with the complexity of $\mathcal{O}(n^3)$. Note that in step 4, the to be inversed matrix $(\mathbf{I} + \mathbf{X}^\top \mathbf{X})$ is unchanged, which only needs to be calculated once in advance. In summary, the
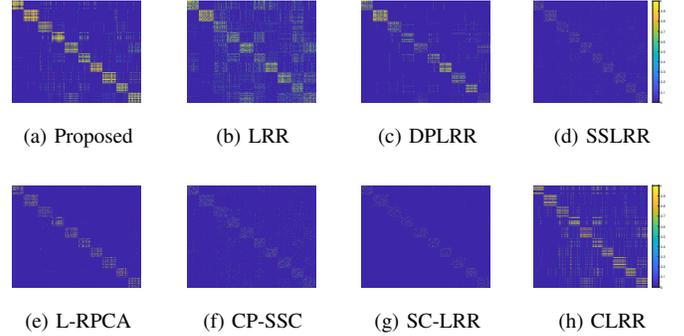
Fig. 4: Visual comparison of the affinity matrices learned by different methods on MNIST. The learned affinity matrices were normalized to [0,1]. Zoom in the figure for a better view.
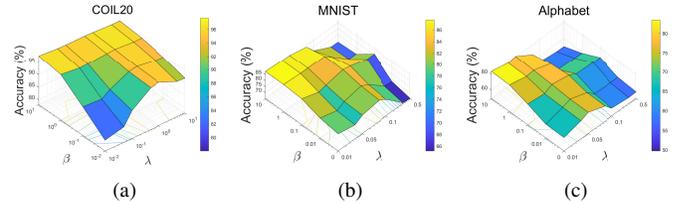
Fig. 5: Influence of the hyper-parameters on clustering performance.

overall computational complexity of Algorithm 1 is $\mathcal{O}(n^3)$ in one iteration.

IV. EXPERIMENTS

In this section, we evaluated the proposed model on 8 commonly-used benchmark datasets, including ORL, YaleB, COIL20, Isolet, MNIST, Alphabet, BF0502 and Notting-Hill[1]. Those datasets cover face images, object images, digit images, spoken letters, and videos. To generate the weakly-supervisory information, following [7], we inferred the pairwise constraints from the randomly selected labels.

We compared our method with LRR [4] and six state-of-the-art semi-supervised subspace clustering methods, including DPLRR [8], SSLRR [7], L-RPCA [19], CP-SSC [20], SC-LRR [9] and CLRR [5]. We performed standard spectral

[1]ORL, YaleB, COIL20, Isolet and MNIST can be found in http://www.cad.zju.edu.cn/home/dengcai/Data/data.html, BF0502 in https://www.robots.ox.ac.uk/~vgg/data/nface/index.html, Notting-Hill in https://sites.google.com/site/baoyuanwu2015/.
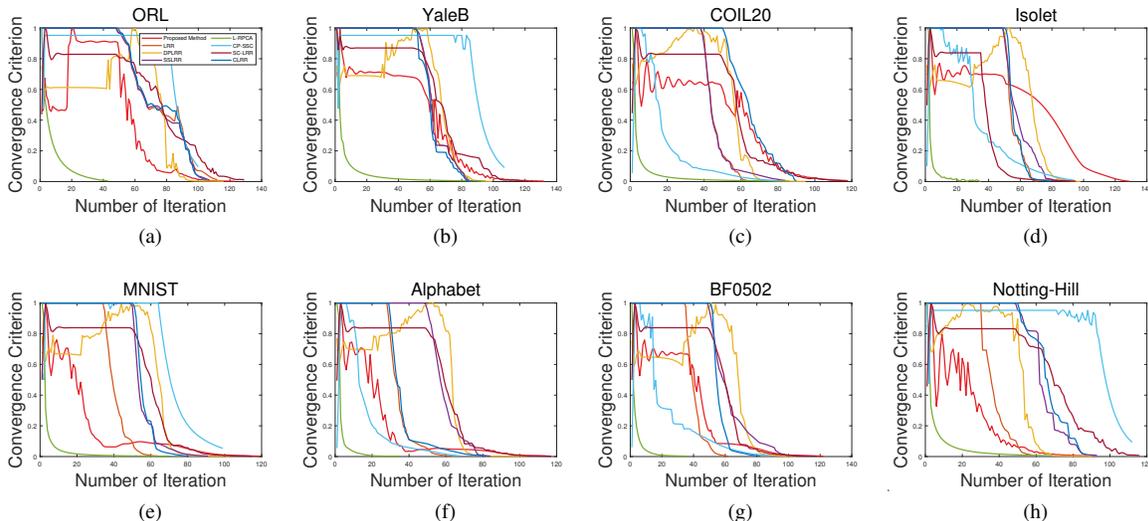
Fig. 6: Convergence behavior comparisons of different methods on eight datasets. The longitudinal axis is normalized.
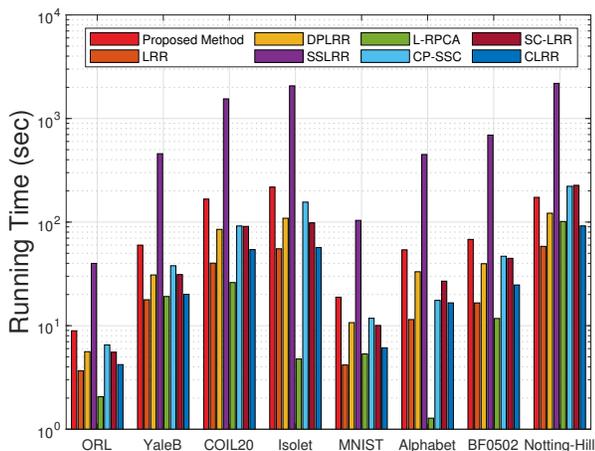
Fig. 7: Running time comparisons of different methods on eight datasets.

clustering in [17] on all the methods to generate the clustering result. We adopted the clustering accuracy and normalized mutual information (NMI) to measure their performance. For both metrics, the larger, the better. For fair comparisons, we carefully tuned the hyper-parameters of the compared methods through exhaustive grid search to achieve the best result. To comprehensively evaluate the different methods, for each dataset, we randomly selected various percentages of initial labels ($\{5\%, 10\%, 15\%, 20\%, 25\%, 30\%\}$) to infer the pairwise constraints. We used the same label information for all the compared methods in every case. To reduce the influence of the random selection, we repeated the experiments 10 times with the randomly selected labels each time, and reported the average performance.

A. Comparison of Clustering Accuracy

Figs. 2-3 compare the clustering accuracy and NMI of different methods under various numbers of pairwise con-

straints, and Table I shows the clustering performance of different methods with 30% initial labels of each datasets as the supervisory information. From those figures and table, we can draw the following conclusions.

1) With more pairwise constraints, all the semi-supervised subspace clustering methods generally perform better, which indicates the effectiveness of including supervisory information in subspace clustering.
2) The proposed method outperforms the other methods significantly. For example our method improves the accuracy value from 0.61 to 0.78 on MNIST and the NMI value from 0.72 to 0.89 on YaleB when compared with the best companions. According to Table I, the proposed method improves the average clustering accuracy of the best compared methods from 0.83 to 0.92. Moreover, the proposed method almost always achieves the best clustering performance with varied supervisory information.
3) The compared methods may be sensitive to different datasets (e.g., SC-LRR achieves the second-best performance on YaleB and MNIST, but performs relatively bad on ORL and COIL20), and sensitive to diverse clustering metrics (e.g., CP-SSC performs well in NMI but poorly in clustering accuracy). On the contrary, the proposed method is robust to distinct datasets and metrics.

Besides, we visualized the affinity matrices learned by different methods on MNIST in Fig. 4, where it can be seen that our method produces more dense and correct connections, leading to the most salient block diagonal affinity. This is owing to the used global tensor low-rank regularization, and further explains the good clustering results reported in Figs. 2-3 and Table I.

B. Hyper-Parameter Analysis

Fig. 5 illustrates how the two hyper-parameters λ and β affect the performance of our method on COIL20, MNIST and Alphabet. It can be seen that the proposed model is relatively

robust to hyper-parameters around the optimal. To be specific, we recommend to set $\lambda = 0.01$ and $\beta = 10$.

C. Convergence Speed

Fig. 6 shows the convergence behaviors of all the compared algorithms on all the datasets. Note that the convergence criteria of all the methods are the same, i.e., the residual error of a variable in two successive steps is less than $1e-3$. We can conclude that L-PRCA usually converges fastest among all the methods. While the proposed algorithm also converges fast compared with other methods like SSLRR and DPLRR. Moreover, the proposed algorithm gets converged within 130 iterations on all the eight datasets.

Fig. 7 compares the average running time of all eight methods on each dataset. Note that we implemented all the methods with MATLAB on a Windows desktop with a 2.90 GHz Intel(R) i5-10400F CPU and 16.0 GB memory. We can observe that the proposed method takes an average time of 95.97s each run on all the eight datasets. It is slightly higher than LRR, but significantly lower than SSLRR. We also need to point out that the proposed method performs much better than the compared methods in clustering performance.

D. Ablation Study

We investigated the effectiveness of the priors/processes involved in our model by comparing the clustering accuracy of Eqs. (3)-(5). The compared methods include two well-known element-wise semi-supervised subspace clustering methods SSLRR and CLRR. As Table II shows, the results of Eq. (3) outperform SSLRR and CLRR significantly on all the datasets, which demonstrates the advantage of the global tensor low-rank prior over the element-wisely fusion strategy. Besides, Eqs. (4) and (5) improve the performance of the proposed model successively, which indicates that both the graph regularization and the post-refinement processing contribute to our model.

V. CONCLUSION

We have proposed a novel semi-supervised subspace clustering model. We first stacked the affinity matrix and pairwise constraint matrix to form a tensor, and then imposed a tensor low-rank prior on it to learn the affinity matrix and augment the pairwise constraints simultaneously. In addition to the global tensor low-rank term, we added a Laplacian regularization term to model the underlying local geometric structure. Furthermore, the learned affinity matrix was refined by the augmented pairwise constraints. The proposed model was formulated as a convex problem, and solved by ADMM. The experimental results demonstrated that our model outperforms other semi-supervised subspace clustering methods significantly.

In the future, we will investigate how to incorporate our work with the existing semi-supervised learning neural networks [21]–[24]. For example, we can use the proposed pairwise constraint enhancement as a loss function to train the neural networks in an end-to-end manner. Moreover, we will improve our method by solving the noisy pairwise constraints problem.

REFERENCES

[1] R. Vidal, "Subspace clustering," *IEEE SPM*, vol. 28, no. 2, pp. 52–68, 2011.
[2] X. Peng, Y. Li, I. W. Tsang, H. Zhu, J. Lv, and J. T. Zhou, "Xai beyond classification: Interpretable neural clustering." *JMLR*, vol. 23, pp. 6–1, 2022.
[3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE TPAMI*, vol. 35, no. 11, pp. 2765–2781, 2013.
[4] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE TPAMI*, vol. 35, no. 1, pp. 171–184, 2013.
[5] J. Wang, X. Wang, F. Tian, C. H. Liu, and H. Yu, "Constrained low-rank representation for robust subspace clustering," *IEEE TCYB*, vol. 47, no. 12, pp. 4534–4546, 2017.
[6] C. Yang, M. Ye, S. Tang, T. Xiang, and Z. Liu, "Semi-supervised low-rank representation for image classification," *SIVP*, vol. 11, no. 1, pp. 73–80, 2017.
[7] L. Zhuang, Z. Zhou, S. Gao, J. Yin, Z. Lin, and Y. Ma, "Label information guided graph construction for semi-supervised learning," *IEEE TIP*, vol. 26, no. 9, pp. 4182–4192, 2017.
[8] H. Liu, Y. Jia, J. Hou, and Q. Zhang, "Learning low-rank graph with enhanced supervision," *IEEE TCSVT*, vol. 32, no. 4, pp. 2501–2506, 2022.
[9] K. Tang, R. Liu, Z. Su, and J. Zhang, "Structure-constrained low-rank representation," *IEEE TNNLS*, vol. 25, no. 12, pp. 2167–2179, 2014.
[10] C. Zhou, C. Zhang, X. Li, G. Shi, and X. Cao, "Video face clustering via constrained sparse representation," in *Proc. IEEE ICME*, 2014, pp. 1–6.
[11] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *Proc. IEEE CVPR*, 2015, pp. 277–286.
[12] Z. Zhang, Y. Xu, L. Shao, and J. Yang, "Discriminative block-diagonal representation learning for image recognition," *IEEE TNNLS*, vol. 29, no. 7, pp. 3111–3125, 2018.
[13] C.-G. Li, J. Zhang, and J. Guo, "Constrained sparse subspace clustering with side-information," in *Proc. IEEE ICPR*, 2018, pp. 2093–2099.
[14] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE TPAMI*, vol. 42, no. 4, pp. 925–938, 2020.
[15] S. Wang, Y. Chen, L. Zhang, Y. Cen, and V. Voronin, "Hyper-laplacian regularized nonconvex low-rank representation for multi-view subspace clustering," *IEEE TSIPN*, vol. 8, pp. 376–388, 2022.
[16] Y. Chen, X. Xiao, and Y. Zhou, "Multi-view subspace clustering via simultaneously learning the representation tensor and affinity matrix," *PR*, vol. 106, p. 107441, 2020.
[17] B. Peng, J. Lei, H. Fu, C. Zhang, T.-S. Chua, and X. Li, "Unsupervised video action clustering via motion-scene interaction constraint," *IEEE TCSVT*, vol. 30, no. 1, pp. 131–144, 2020.
[18] Y.-P. Zhao, L. Chen, and C. L. P. Chen, "Laplacian regularized nonnegative representation for clustering and dimensionality reduction," *IEEE TCSVT*, vol. 31, no. 1, pp. 1–14, 2021.
[19] D. Zeng, Z. Wu, C. Ding, Z. Ren, Q. Yang, and S. Xie, "Labeled-robust regression: Simultaneous data recovery and classification," *IEEE TCYB*, vol. 52, no. 6, pp. 5026–5039, 2022.
[20] K. Somandepalli and S. Narayanan, "Reinforcing self-expressive representation with constraint propagation for face clustering in movies," in *Proc. IEEE ICASSP*, 2019, pp. 4065–4069.
[21] W. Zhang, Q. M. J. Wu, and Y. Yang, "Semisupervised manifold regularization via a subnetwork-based representation learning model," *IEEE TCYB*, pp. 1–14, 2022.
[22] H. Zhao, J. Zheng, W. Deng, and Y. Song, "Semi-supervised broad learning system based on manifold regularization and broad network," *IEEE TCAS-I*, vol. 67, no. 3, pp. 983–994, 2020.
[23] Y. Li, P. Hu, Z. Liu, D. Peng, J. T. Zhou, and X. Peng, "Contrastive clustering," in *Proc. AAAI*, vol. 35, no. 10, 2021, pp. 8547–8555.
[24] Y. Li, M. Yang, D. Peng, T. Li, J. Huang, and X. Peng, "Twin contrastive learning for online clustering," *IJCV*, vol. 130, no. 9, pp. 2205–2221, 2022.