

Sparse and low-rank regularized deep subspace clustering

Wenjie Zhu^{*}, Bo Peng

Key Laboratory of Electromagnetic Wave Information Technology and Metrology of Zhejiang Province, College of Information Engineering, China Jiliang University, Hangzhou 310018, China

ARTICLE INFO

Article history:

Received 17 March 2020

Received in revised form 27 June 2020

Accepted 28 June 2020

Available online 3 July 2020

Keywords:

Subspace clustering

Self-expressive matrix

Low-rank

Deep neural network

ABSTRACT

Subspace clustering aims at discovering the intrinsic structure of data in unsupervised fashion. As ever in most of approaches, an affinity matrix is constructed by learning from original data or the corresponding hand-crafted feature with some constraints on the self-expressive matrix (SEM), which is then followed by spectral clustering algorithm. Based on successful applications of deep technologies, it has become popular to simultaneously accomplish deep feature and self-representation learning for subspace clustering. However, deep feature and SEM in previous deep methods are lack of precise constraints, which is sub-optimal to conform with the linear subspace model. To address this, we propose an approach, namely sparse and low-rank regularized deep subspace clustering (SLR-DSC). In the proposed SLR-DSC, an end-to-end framework is proposed by introducing sparse and low-rank constraints on deep feature and SEM respectively. The sparse deep feature and low-rank regularized SEM implemented via fully-connected layers are encouraged to facilitate a more informative affinity matrix. In order to solve the nuclear norm minimization problem, a sub-gradient computation strategy is utilized to cater to the chain rule. Experiments on the data sets demonstrate that our method significantly outperforms the competitive unsupervised subspace clustering approaches.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Collections of high-dimensional data in the real-world context can often be well represented by multiple low-dimensional subspaces across different classes. Due to extreme conditions of data samples, e.g. PIE (pose, illumination, and expression) for face images, the identity-specific segmentation can be accomplished by searching for the intrinsic structure of data with complex patterns. Subspace clustering exactly refers to the problem of partitioning data samples that lie in a union of low-dimensional subspaces. As for pattern analysis, it is rarely practical to collect a large amount of labeled data. Therefore, subspace clustering in unsupervised fashion has draw attentions from an increasing number of researchers in recent decades.

Subspace clustering has been successfully applied in face clustering [1,2], motion segmentation [3,4], feature selection [5], hyperspectral image processing [6,7] and so on. Most previous state-of-the-art approaches to subspace clustering follow a two-stage framework: (i) computing an affinity matrix by representing the original data samples over themselves with a SEM, and then (ii) adopting the spectral clustering technique to obtain clustering results. Therefore, related works focus on improving the SEM which contributes to an affinity matrix. The remarkable works can be

classified into two categories: traditional subspace clustering [8–16] and deep subspace clustering [17–19] according to the way to learn feature, deep or not.

Traditional Subspace Clustering. An optimal SEM makes much difference in subspace clustering approaches based on the self-expressive representation [20–23]. As well-behaved prior knowledge for multiple pattern analysis applications, sparsity is normally regarded as one of important characteristics of SEM. The representative sparse subspace clustering [24] computes the sparse representation to construct an affinity matrix by solving the ℓ_0 or ℓ_1 minimization problem. Then, the subspace clustering problem is posed as a non-convex problem to achieve the low-rank representation [25]. Inspired by their works, there have been various subspace clustering algorithms emerging in the past few years [26–31]. By introducing the kernel trick and neural networks, nonlinear subspace clustering [32,33] has been investigated to mine the nonlinear relation. Considering the fact that the two-stage framework may lead to sub-optimal results, a joint optimization framework is proposed to learn the affinity matrix and segmentation simultaneously [34]. The affinity matrix in subspace clustering is obtained by the optimization involved all of the images which is not applicable to the large-scale data set. For this, some of the related works [35,36] propose scalable subspace clustering by decomposing the original problem into some sequences which can be operated online. The subspace clustering algorithms based on sparse and low-rank representation

^{*} Corresponding author.

E-mail address: zhwj@cjl.u.edu.cn (W. Zhu).

have applied successfully in image processing [37] and object tracking [38].

Deep Subspace Clustering. With the remarkable success in computer vision and pattern classification, deep learning technique has attracted great concern due to its promising feature description ability. Sparse constraint has been applied in machine learning tasks [17,18,39]. Sparsity prior is introduced into deep neural net work to accomplish the subspace clustering, in which the hidden representation is enforced to preserve the predefined sparse relation, but the input is hand-crafted feature rather than original data [17,18]. With the aim to achieve an end-to-end representation learning for subspace clustering, the encoders and decoders are applied in series to facilitate neural networks. Then, a fully-connected layer is integrated into deep neural network as a self-expression scheme to bridge the deep feature learning and SEM optimization in DSC [19]. The compared deep subspace clustering frameworks are shown in Fig. 1. Along the framework of DSC, multimodal deep subspace clustering [40] and the multi-view DSC has been investigated [41]. In these deep subspace clustering algorithms, the deep feature is not constrained and SEM is merely constrained using ℓ_1 norm or ℓ_2 norm so that the affinity matrix is not informative enough.

Based on the discussion above, the proposed approach named Sparse and Low-Rank regularized Deep Subspace Clustering (SLR-DSC), in which deep feature and SEM are regularized with sparsity and low-rankness respectively to make the affinity matrix more informative. The contributions of our work are highlighted in two aspects:

- (1) Sparse regularized deep feature assists in improving data feature with distinguishable tendency by transforming data samples from original space to nonlinear feature space.
- (2) Low-rankness is utilized to improve SEM. For catering to chain rule of neural network, a sub-gradient computation strategy is employed to solve the nuclear norm minimization problem. To our best of knowledge, this is the first work that applies low-rank constraint on the SEM into neural network for subspace clustering.

2. Related work

Given a union of multiple subspaces spanned by collections of data samples, subspace clustering aims at partitioning them into some category-specific groups in unsupervised fashion. The self-expression framework intends to compute SEM and the corresponding affinity matrix which is then utilized for spectral clustering. Therefore, the optimization problem of self-expression based subspace clustering is:

$$\begin{aligned} \min_{\mathbf{C}} \quad & \frac{1}{2} \|\mathbf{X} - \mathbf{XC}\|_F^2 + f(\mathbf{C}), \\ \text{s.t.} \quad & \text{diag}(\mathbf{C}) = 0, \end{aligned} \quad (1)$$

where $\mathbf{X} \in \mathbb{R}^{d \times n}$ and $\mathbf{C} \in \mathbb{R}^{n \times n}$ denote an original sample set including d -dimensional n samples and SEM respectively. With different assumptions of subspace, $f(\cdot)$ is expressed using one or mixture of norm functions [42], e.g., ℓ_0 , ℓ_1 , ℓ_2 , nuclear norm [43] and so on. Based on the SEM scheme, there have been a variety of subspace clustering works [44–47], e.g. nonlinear subspace clustering [48,49]. $\mathbf{X}$ can be the original samples or hand-crafted feature. However, it is not necessarily guaranteed that the hand-crafted feature exploit robust descriptions of data samples.

Most contributions of deep subspace clustering rely on powerful nonlinear feature descriptions of neural network [50]. The simplified structure of competitive deep subspace clustering framework is presented in Fig. 1. As an extension of traditional sparse subspace clustering, the works [17,18] introduce a sparse

$\mathbf{C}$ into neural network framework to achieve deep feature with sparsity preservation, solving the optimization problem¹ as follows:

$$\begin{aligned} \min_{\mathbf{Y}} \quad & \frac{1}{2} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \frac{\lambda}{2} \|\mathbf{Y} - \mathbf{YC}\|_F^2, \\ \text{s.t.} \quad & \mathbf{Y} = g(\mathbf{X}), \hat{\mathbf{X}} = h(\mathbf{Y}), \end{aligned} \quad (2)$$

where $\mathbf{Y}$ denotes the deep feature derived from the encoder layers, $g(\cdot)$ and $h(\cdot)$ denote the mapping function implemented by encoder and decoder layers respectively. $\hat{\mathbf{X}}$ denotes the reconstructed data set. The first term attempts to minimize the reconstruction error and the second term encourages the deep feature to be updated towards sparsity preservation. Despite sparsity preservation of PARTY (deep subspace clustering with sparsity prior) in [17,18], the pre-computed SEM in the original space $\mathbf{C}$ may be sub-optimal for deep feature learning. Thus, DSC designs a fully-connected layer for SEM which is followed by a series of decoder layers. More precisely, DSC jointly learns deep feature and SEM. The loss function of DSC is defined as follows:

$$\begin{aligned} \min_{\mathbf{Y}, \mathbf{C}} \quad & \frac{1}{2} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \frac{\lambda}{2} \|\mathbf{Y} - \mathbf{YC}\|_F^2 + \|\mathbf{C}\|_{(\cdot)}, \\ \text{s.t.} \quad & \text{diag}(\mathbf{C}) = 0, \mathbf{Y} = g(\mathbf{X}), \hat{\mathbf{X}} = h(\mathbf{Y}), \end{aligned} \quad (3)$$

where $\|\mathbf{C}\|_{(\cdot)}$ indicates ℓ_1 or ℓ_2 norm in DSC.

From the compared works, it can be observed that PARTY involves a pre-computed SEM in original space and learns deep feature that preserves sparse relation. DSC jointly learns deep feature and SEM in feature space which is more robust than PARTY. However, both of them do not attach much weight on the sparsity of deep feature and low-rankness of SEM.

3. Proposed method:SLR-DSC

Inspired by the auto-encoder neural network for deep subspace clustering, we extend the DSC framework for sparse deep feature and low-rank SEM to improve informative affinity matrix. In order to cater to chain rule of neural network, a sub-gradient computation for minimizing nuclear norm is adopted in this work. Accordingly, we first present the objective of proposed SLR-DSC approach, and then give the optimization procedure by expressing the sub-gradient computation for solving nuclear norm minimization problem.

3.1. Objective

According to discussion above, we employ the same architecture presented in Fig. 1. The sparse deep feature and low-rank SEM are highlighted in proposed SLR-DSC, and our goal is to minimize the following objective:

$$\begin{aligned} \min \quad & \mathcal{J} = \mathcal{J}_E + \mathcal{J}_F + \mathcal{J}_L, \\ \text{s.t.} \quad & \text{diag}(\mathbf{C}) = 0, \mathbf{Y} = g(\mathbf{X}), \hat{\mathbf{X}} = h(\mathbf{Y}), \end{aligned} \quad (4)$$

where $\mathcal{J}_E = \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \|\mathbf{Y} - \mathbf{YC}\|_F^2$ measures the reconstruction errors of encoder-decoder scheme and self-representation learning, $\mathcal{J}_F = \|\mathbf{Y}\|_1$ encourages deep feature to be sparse for distinguishable information exploitation. The low-rank SEM can be achieved by introducing the third loss $\mathcal{J}_L = \|\mathbf{C}\|_* + \|\mathbf{C}\|_F^2$, where $\|\mathbf{C}\|_*$ denotes nuclear norm that can be replaced with the sum of absolute eigenvalues of $\mathbf{C}$. By attaching weights on each term in Eq. (4), we have:

$$\begin{aligned} \min_{\Theta_e, \Theta_d, \mathbf{C}} \quad & \left\{ \frac{1}{2} \|\mathbf{X} - \hat{\mathbf{X}}\|_F^2 + \frac{\lambda}{2} \|\mathbf{Y} - \mathbf{YC}\|_F^2 \right\} \\ & + \alpha \|\mathbf{Y}\|_1 + \beta \|\mathbf{C}\|_* + \gamma \|\mathbf{C}\|_F^2, \\ \text{s.t.} \quad & \text{diag}(\mathbf{C}) = 0, \mathbf{Y} = g(\Theta_e; \mathbf{X}), \hat{\mathbf{X}} = h(\Theta_d; \mathbf{Y}), \end{aligned} \quad (5)$$

¹ The weights and bias parameters associated with $g(\cdot)$ are target variables in [17,18], which is actually equivalent to achieving deep feature $\mathbf{Y}$.

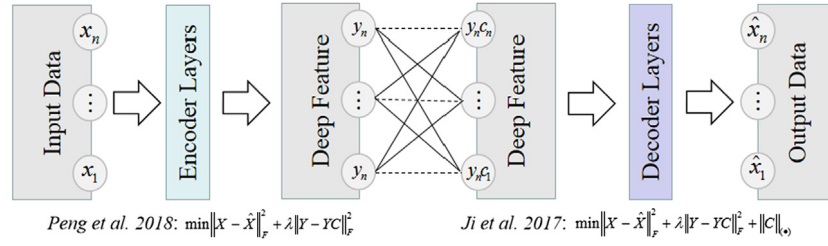


Fig. 1. Compared deep subspace clustering frameworks. The input data can be mapped through a mount of encoder layers to achieve deep feature, and then the reconstructed data can be obtained in the other side of neural network by decoding the self-expressive deep feature.

where λ , α , β and γ are four non-negative tradeoff parameters. Θ_e and Θ_d denote the parameters of encoder and decoder parameters of neural networks.

In the proposed objective, the mixture of nuclear norm and Frobenius norm is employed to jointly control the self-expressive matrix. It has been proved that there exists connections between them [51], however, the introduction of nuclear norm is not superfluous:

(1) Generally, the self-expressive matrix is enforced to be sparse or/and low-rank in the traditional subspace clustering task. In [19], Ji et al. introduce the ℓ_1 -norm and ℓ_2 -norm on the coefficient matrix $\mathbf{C}$ in DSCNet respectively. The experiments on three benchmark data sets show that the latter results in a lower clustering error. This can be attributed to the difficulty in optimization introduced by the ℓ_1 norm as it is non-differentiable at zero.

(2) In the traditional subspace clustering, one or mixture of norm functions, including ℓ_0 , ℓ_1 , ℓ_2 , nuclear norm. In the proposed objective, the matrix elastic net regularization [52] formulated with ℓ_2 norm and nuclear norm is utilized to improve the self-expressive matrix. The experiments of ablation study in the later section can illustrates the effectiveness of introduction of nuclear norm.

3.2. Optimization

Similar to [19], the coefficient matrix $\mathbf{C}$ is treated as a part of deep neural network, and back propagation is adopted to update the unknowns in Eq. (5). In order to address the problem in computing derivative of nuclear norm and meet the requirements of the chain rule, we introduce a sub-gradient computation strategy [53]. By decomposing $\mathbf{C}$ with singular value decomposition (SVD), $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices and $\mathbf{\Sigma}$ is the diagonal matrix with real numbers on the diagonal. Hence, the nuclear norm can be re-written as follows,

$$\begin{aligned} \|\mathbf{C}\|_* &= \text{trace}(\sqrt{\mathbf{C}^T \mathbf{C}}) \\ &= \text{trace}(\sqrt{(\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)^T (\mathbf{U}\mathbf{\Sigma}\mathbf{V}^T)}) \\ &= \text{trace}(\sqrt{\mathbf{V}\mathbf{\Sigma}^2 \mathbf{V}^T}) = \text{trace}(\sqrt{\mathbf{V}^T \mathbf{V} \mathbf{\Sigma}^2}) \\ &= \text{trace}(|\mathbf{\Sigma}|), \end{aligned} \quad (6)$$

where $|\mathbf{\Sigma}|$ denotes the element-wise absolute matrix of $\mathbf{\Sigma}$. Thus, the nuclear norm of $\mathbf{C}$ can be also defined as the ℓ_1 norm of singular values of the SVD of $\mathbf{C}$ [54].

Since $\mathbf{\Sigma}$ is diagonal, its sub-differential set can be computed:

$$\frac{\partial |\mathbf{\Sigma}|}{\partial \mathbf{C}} = |\mathbf{\Sigma}| \mathbf{\Sigma}^{-1} \frac{\partial \mathbf{\Sigma}}{\partial \mathbf{C}}, \quad (7)$$

where $\mathbf{\Sigma}^{-1}$ represents the Moore–Penrose pseudo-inverse matrix. Since $|\mathbf{\Sigma}|$ is not derivable at zero point, the sub-differential set depends on the cases on the different sides of zero. It is obvious that the elements of $|\mathbf{\Sigma}| \mathbf{\Sigma}^{-1}$ are 1 or -1 . Therefore, the results

of Eq. (7) can construct the sub-differential set [55]. Recalling $\mathbf{C} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, it can be checked that

$$\text{trace}(\partial \mathbf{\Sigma}) = \text{trace}(\mathbf{U}^T \partial \mathbf{C} \mathbf{V}) \quad (8)$$

With integration of Eqs. (6)–(8), we have

$$\begin{aligned} \frac{\partial \|\mathbf{C}\|_*}{\partial \mathbf{C}} &= \frac{\partial \text{trace}(|\mathbf{\Sigma}|)}{\partial \mathbf{C}} = \frac{\text{trace}(\partial |\mathbf{\Sigma}|)}{\partial \mathbf{C}} \\ &= \frac{\text{trace}(|\mathbf{\Sigma}| \mathbf{\Sigma}^{-1} \partial \mathbf{\Sigma})}{\partial \mathbf{C}} \\ &= \frac{\text{trace}(|\mathbf{\Sigma}| \mathbf{\Sigma}^{-1} \mathbf{U}^T \partial \mathbf{C} \mathbf{V})}{\partial \mathbf{C}} \\ &= (\mathbf{V} |\mathbf{\Sigma}| \mathbf{\Sigma}^{-1} \mathbf{U}^T)^T. \end{aligned} \quad (9)$$

Algorithm 1: SLR-DSC algorithm

Input: The collection of images $\mathbf{X}$, the parameters $\{\lambda, \alpha, \beta, \gamma\}$, the number of clusters k , the learning rate η , and the number of epochs M .

Output: The clustering indexes of images.

- 1 Conduct the pre-trained processing using auto-encoder networks and achieve the initialized parameters of networks;
- 2 Initialize the number of epoch $m = 0$;
- 3 **while** $m < M$ **do**:
- 4 $m \leftarrow m + 1$;
- 5 Train the neural network using back propagation algorithm with both sparse and low-rank constraints simultaneously;
- 6 **end while**
- 7 Construct the affinity matrix $\mathbf{Z} = (|\mathbf{C}| + |\mathbf{C}|^T)/2$;
- 8 **spectral clustering**:
- 9 Compute the Laplacian matrix $\mathbf{L}$ via $\mathbf{L} = \sum_i \mathbf{Z}_i - \mathbf{Z}$;
- 10 Construct the new samples using the eigenvectors corresponding to the k smallest eigenvalues of $\mathbf{L}$;
- 11 Employ K-means algorithm on the new samples to obtain the clustering indexes.
- 12 **Return:** The clustering indexes of images.

The gradient in Eq. (9) is then passed down to the deep auto-encoder neural network and used in standard back propagation to compute the network parameter gradient. Due to the fully-connected layers, all of data samples are responsible for updating related parameters. Thus, the whole batch mechanism is utilized although it requires large memory. The proposed SLR-DSC algorithm is outlined in Algorithm 1. The pre-training strategy is utilized to initialize the networks, and the spectral clustering technique is introduced as the post-processing to obtain the clustering indexes of images.

3.3. Comparison with deep sparse auto-encoder

As illustrated in the previous subsections, the proposed SLR-DSC introduce the sparse constraint on the encoded deep feature.

Table 1
Neural network settings of data sets.

Data sets	COIL20			Extended Yale B			ORL		
Layers	ks	Channels	#par	ks	Channels	#par	ks	Channels	#par
Encoder-1	3×3	15	150	5×5	10	260	3×3	3	130
Encoder-2	–	–	–	3×3	20	1820	3×3	3	138
Encoder-3	–	–	–	3×3	30	5430	3×3	5	84
SE-layer	–	–	2 073 600	–	–	160 000	–	–	5 914 624
Decoder-1	3×3	15	136	3×3	30	5420	3×3	5	84
Decoder-2	–	–	–	3×3	20	1810	3×3	3	140
Decoder-3	–	–	–	5×5	10	251	3×3	3	126

It is true that SLR-DSC is somewhat similar to deep sparse AE on the ground that both algorithms leverage stacked encoder layers and decoder layers for deep feature extraction and reconstruction with the sparse constraint on the deep feature. However, the application of the two methods is totally different. Specifically, deep sparse AE is employed for the acquisition of deep feature during the process of image construction and SLR-DSC is designed for subspace clustering. In view of this, the sparse constraint in deep sparse AE is introduced for high-dimensional but sparse features with more powerful expression ability while the intension behind the sparse constraint in SLR-DSC is to make the deep features more discriminative and therefore enhance clustering performance.

3.4. Connection with PARTY and DSC

As discussed before, the proposed SLR-DSC follows the same architecture as PARTY and DSC (see Fig. 1). According to Eq. (5), the connections between ours and other two methods (PARTY and DSC) are summarized as follows:

- SLR-DSC, PARTY, and DSC follow encoder–decoder based neural networks in which SEM is designed as a fully-connected layer;
- Both proposed SLR-DSC and DSC can learn deep feature and SEM simultaneously. By contrast, the PARTY algorithm firstly make use of prior knowledge to construct a sparse SEM and then learn deep feature with the SEM;
- Compared with DSC which only achieves SEM by minimizing ℓ_1 norm or ℓ_2 norm, the proposed SLR-DSC ensures deep feature to be sparse and SEM to be low-rank, which makes feature descriptions more distinguishable and affinity matrix more informative.

Mathematically, we can observe from Eq. (5) that SLR-DSC degrades to DSC when setting $\alpha = \beta = 0$. Thus, our work is the generalized version of DSC by introducing sparse and low-rank regularization. Further setting $\gamma = 0$, it is converted to PARTY given a prior $\mathbf{C}$.

4. Experiments

To evaluate the performance of SLR-DSC for subspace clustering, we conduct the experiments on the face and object data sets. Meanwhile, the experimental results of some methods including traditional and deep subspace clustering algorithms are also provided as benchmarks. In addition, the impact of sparse constraint and low-rank constraint illustrated in Eq. (5) is analyzed respectively with the ablation study.

4.1. Experimental settings

4.1.1. Data sets

In order to evaluate the performance of the proposed SLR-DSC for subspace clustering, three benchmark data sets are employed

Table 2

Compared results of ACC and NMI (in %) of competitive methods on ORL data set.

Methods	LRSC	SSC	EnSC
ACC	67.75	67.00	69.25
NMI	80.82	82.06	82.20
Methods	AE+LRSC	AE+SSC	AE+EnSC
ACC	68.00	74.50	78.75
NMI	83.38	88.24	87.62
Methods	DSCNet1	DSCNet2	SLR-DSC
ACC	85.75	86.00	89.00
NMI	–	–	92.81

in the experiment, i.e. Extended Yale B face data set, COIL20 object data set, and ORL face data set.

There are 38 individuals and 64 face images for each person in the Extended Yale B data set. The images with size 168×192 are down-sampled to 42×42 . Some of the images are shown in Fig. 2(a). Another challenging face data set employed in this paper is ORL face data set. The example face images on the ORL face data set are shown in Fig. 2(b). There are totally 165 face images across 15 different persons on the ORL face data set. COIL20 object is the last data set employed in this experiment, in which 20 different objects contribute to 1440 object images. Some of object images are shown in Fig. 2(c). In the experiment, the images from ORL data set and COIL20 data set are normalized to 32×32 .

4.1.2. Training details

Following the DSC, the pre-trained auto-encoder neural networks are utilized to initialize the parameters Θ . The proposed method is implemented in Python with Tensorflow-1.14, and the learning rate $\eta = 10^{-3}$ for updating related variables. We did not tune the weights of loss terms and set parameters empirically, $\lambda = \gamma = 0.5$, $\alpha = \beta = 1.0$. The structure of neural network is summarized in Table 1, where kernel size (ks) and channels are the convolutional layer parameters, and the number of parameters (#par) indicates the number of neural network nodes as well. A big batch using the whole data is feed into the neural networks in the pre-training stage. According to [19], it is to be numerically more stable and converge faster than stochastic gradient descent using randomly sampled mini-batches which is verified in this work as well.

4.1.3. Competitive methods

In this experiment, some state-of-the-art subspace clustering algorithms in previous works are employed for comparison: Low Rank Subspace Clustering (LRSC) [25], Sparse Subspace Clustering (SSC) [1], Elastic Net Subspace Clustering (EnSC) [56], LRSC, SSC, and EnSC with the pre-trained convolutional auto-encoder features (AE+LRSC, AE+SSC, AE+EnSC), Deep Subspace Clustering with Sparsity Prior (PARTY) [17], Structured AutoEncoders for subspace clustering (StructAE) [18], and Deep Subspace Clustering Networks (DSCNet) [19] which includes ℓ_1 (DSCNet1) and ℓ_2 (DSCNet2) versions.



(a) Extended Yale B



(b) ORL



(c) COIL20

Fig. 2. Example images on the data sets in this experiment.**Table 3**

Compared results of ACC and NMI (in %) of competitive methods on Extended Yale B and COIL20 data sets.

Methods	ACC		NMI	
	Extended Yale B	COIL20	Extended Yale B	COIL20
LRSC	67.43	74.72	77.19	80.69
SSC	62.54	86.04	72.51	94.50
EnSC	61.14	82.43	65.52	91.50
AE+LRSC	63.57	73.82	74.25	80.22
AE+SSC	74.75	77.64	78.82	83.36
AE+EnSC	63.16	83.75	73.37	92.27
PARTY-DSIFT	88.55	85.76	90.85	91.19
PARTY-HOG	92.08	85.50	96.91	91.19
StructAE-SDSIFT	94.70	91.04	96.58	94.67
StructAE-LPQ	45.65	84.38	50.24	89.72
DSCNet1	96.67	93.05	–	–
DSCNet2	97.33	94.86	–	–
SLR-DSC	98.12	97.64	96.73	97.63

4.2. Experimental results

To evaluate performance of the proposed SLR-DSC, we conduct experiments on the aforementioned three benchmark data sets with SLR-DSC and the competitive methods. The clustering performance measured by ACC and NMI on the employed data sets are illustrated in Tables 2 and 3. For fair comparison, we report the experimental results of PARTY and StructAE on the Extended Yale B and COIL20 data sets in the corresponding works.

As shown in Tables 2 and 3, the methods based on PARTY or StructAE perform significantly better than those using auto-encoder (AE+), which indicates that the introduction of prior

knowledge makes deep feature extraction more effective. Besides, it is noticeable that deep subspace clustering models can bring relatively satisfying outcomes, which results from the introduction of SEM learning. Meanwhile, among all deep subspace clustering techniques, the clustering result of the proposed SLR-DSC reaches the highest level. The success of our approach is achieved by imposing the sparse constraint on deep feature and the low-rank constraint on the SEM.

In order to further verify the robustness of the proposed SLR-DSC with regard to an increasing number of clusters, we select $K = 2, 5, 10, 15$ classes (each class includes 72 images,) from the COIL20 data set in order. The clustering results with different numbers of clusters are shown in Fig. 3. Seen from Fig. 3, with regard to the number of clusters, the performance of DSCNet and the proposed SLR-DSC remains at the highest level which illustrates the superiority of deep subspace clustering technique. Besides, thanks to the additionally introduced sparse constraint on the deep feature and low-rank constraint on the self-expressive coefficient matrix, the proposed SLR-DSC always outperforms DSCNet. Meanwhile, comparing traditional subspace clustering technique (LRSC, SSC and EnSC) with the two-stage methods (AE+LRSC, AE+SSC, AE+EnSC), we observe that simply combining the pre-trained auto-encoder features do not necessarily improve the performance of LRSC, SSC and EnSC, which confirms the benefits of the joint optimization of all parameters in one network.

Since the proposed SLR-DSC pursues sparse deep features, we also show the sparsity of the deep features. As the result of feeding the original data samples into encoder layers, one sparse deep feature is depicted in Fig. 4. After counting zero elements of the displayed deep feature, its sparsity rates are $16/80 = 20\%$, $221/1080 = 20.46\%$, and $137/3840 = 3.57\%$ on the ORL, YaleB, and COIL20 data sets respectively.

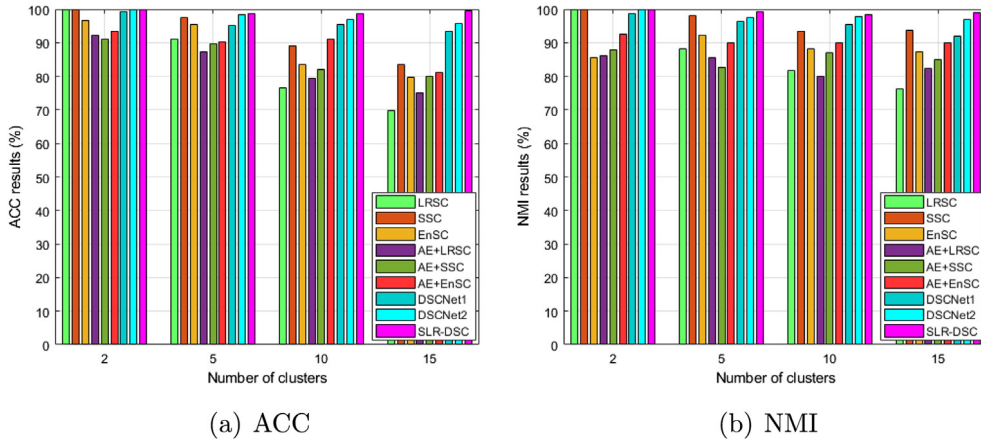


Fig. 3. ACC and NMI results of competitive algorithms versus different numbers of clusters on the COIL20 data set.

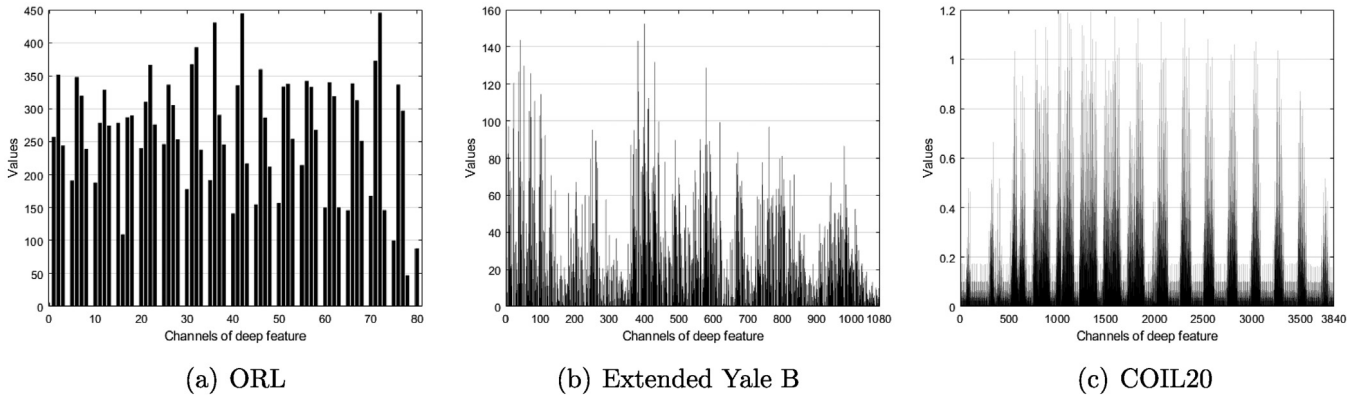


Fig. 4. Sparse deep feature derived from SLR-DSC on the data sets.

Table 4
ACC and NMI results on the data set for ablation study.

ACC	ORL	Extended Yale B	COIL20
DSCNet2	86.00	97.33	94.86
SLR-DSC($\alpha = 0$)	88.25	97.88	97.05
SLR-DSC($\beta = 0$)	87.50	97.62	96.60
SLR-DSC	89.00	98.12	97.64
NMI	ORL	Extended Yale B	COIL20
SLR-DSC($\alpha = 0$)	92.14	96.48	97.57
SLR-DSC($\beta = 0$)	92.37	96.12	97.14
SLR-DSC	92.81	96.73	97.63

Table 5
Training time (s) of one epoch on the data sets.

Methods	ORL	Extended Yale B	COIL20
DSCNet1	0.06	21.48	1.02
DSCNet2	0.06	20.76	0.98
SR-DSC	0.07	22.91	1.09
LR-DSC	0.09	23.37	1.23
SLR-DSC	0.11	26.21	1.31

4.3. Ablation study

As shown in Eq. (5), the sparse constraint on deep features and low-rank constraint on SEM are the highlights in our work. Thus, we carry out the ablation study to reveal the contributions of the two constraints. Table 4 reports the ACC and NMI with the degraded versions of SLR-DSC. By setting $\alpha = 0$, only low-rank constraint works and deep feature is dense. Setting $\beta = 0$, deep feature is expected to be sparse. According to the results in Table 4, both SLR-DSC($\alpha = 0$) and SLR-DSC($\beta = 0$) perform worse than SLR-DSC, which illustrates the effectiveness of proposed objective. Moreover, the clustering results of SLR-DSC($\alpha = 0$) are slightly higher than those of SLR-DSC($\beta = 0$), which indicates the contribution of low-rank regularization in this work.

4.4. Comparison of computational time

Due to the introduction of nuclear norm in the proposed SLR-DSC algorithm, the SVD procedure needs an external cost during the learning process. Theoretically, the computation complexity of SVD is $\mathcal{O}(n^3)$, where n is the number of images. To compare the computation time at the training phase, we report the training time in terms of second by counting one epoch using DSCNet1, DSCNet2, and proposed SLR-DSC, including two degraded versions SR-DSC and LR-DSC in Table 5. From Table 5, we can dramatically observe that DSCNet2 needs the least training time owing to the ℓ_2 norm. The SVD procedure actually makes SLR-DSC require relatively more computation complexity. Besides, the computational time varies largely from data sets. Two aspects, the number of images and the number of layers, lead to relatively much more computational time on the Extended Yale B data set. Thanks to the less number of layers on the COIL20 data set, we can use less computational time.

5. Conclusion

In this paper, we proposed a dictionary learning method based on deep neural network for subspace clustering. In this work, we proposed a sparse and low-rank regularized deep neural network for subspace clustering. Sparse deep feature and low-rank SEM were achieved by architecture of SLR-DSC. For purpose of realizing back propagation to update variables, the sub-gradient computation strategy for minimizing nuclear norm were employed. Despite approximate derivative of nuclear norm, the clustering performance illustrates that the proposed SLR-DSC outperforms the state-of-the-art traditional subspace clustering approaches and previous deep subspace clustering methods on the data sets. Despite the successful integration of low-rank constraint into deep neural networks, the deep subspace clustering framework suffers from loading all the images in one batch, which is memory-consuming and not update-friendly to neural networks. Therefore, the scalable deep subspace clustering will be the future work.

CRedit authorship contribution statement

Wenjie Zhu: Methodology, Supervision, Writing - review & editing. **Bo Peng:** Software, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work is in part by Natural Science Foundation of Zhejiang Province, China (NO. LQ20F030015).

References

- [1] E. Elhamifar, R. Vidal, Sparse subspace clustering: Algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (11) (2013) 2765–2781.
- [2] K. Somandepalli, S. Narayanan, Reinforcing self-expressive representation with constraint propagation for face clustering in movies, in: *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019, pp. 4065–4069.
- [3] S. Rao, R. Tron, R. Vidal, Y. Ma, Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (10) (2010) 1832–1845.
- [4] P. Ji, Y.R. Zhong, H.D. Li, M. Salzmann, Null space clustering with applications to motion segmentation and face clustering, in: *IEEE International Conference on Image Processing*, 2014.
- [5] H. Zhai, H.Y. Zhang, L.P. Zhang, P.X. Li, Laplacian-regularized low-rank subspace clustering for hyperspectral image band selection, *IEEE Trans. Geosci. Remote Sens.* 57 (3) (2019) 1723–1740.
- [6] B. Tu, X. Yang, N. Li, C. Zhou, D. He, Hyperspectral anomaly detection via density peak clustering, *Pattern Recognit. Lett.* 129 (2019) 144–149.
- [7] B. Tu, X. Zhang, X. Kang, J. Wang, J.A. Benediktsson, Spatial density peak clustering for hyperspectral image classification with noisy labels, *IEEE Trans. Geosci. Remote Sens.* 57 (7) (2019) 5085–5097.
- [8] Q. Li, Z. Sun, Z. Lin, R. He, T. Tan, Transformation invariant subspace clustering, *Pattern Recognit.* 59 (2016) 142–155.
- [9] J. Chen, H. Mao, Y. Sang, Z. Yi, Subspace clustering using a symmetric low-rank representation, *Knowl.-Based Syst.* 127 (2017) 46–57.
- [10] S. Du, Y. Ma, Y. Ma, Graph regularized compact low rank representation for subspace clustering, *Knowl.-Based Syst.* 118 (2017) 56–69.
- [11] M. Brbić, I. Kopriva, Multi-view low-rank sparse subspace clustering, *Pattern Recognit.* 73 (2018) 247–258.
- [12] H. Chen, W. Wang, X. Feng, Structured sparse subspace clustering with within-cluster grouping, *Pattern Recognit.* 83 (2018) 107–118.
- [13] J. Song, G. Yoon, K. Hahn, S.M. Yoon, Subspace clustering via structure-enforced dictionary learning, *Neurocomputing* 362 (2019) 1–10.
- [14] Y. Chen, Z. Yi, Locality-constrained least squares regression for subspace clustering, *Knowl.-Based Syst.* 163 (2019) 51–56.
- [15] M. Liao, X. Gu, Subspace clustering based on alignment and graph embedding, *Knowl.-Based Syst.* 188 (2020) 105029.
- [16] Z. Kang, X.J. Zhao, C. Peng, H.Y. Zhu, J.T. Zhou, X. Peng, W.Y. Chen, Z.L. Xu, Partition level multiview subspace clustering, *Neural Netw.* 122 (2020) 279–288.
- [17] X. Peng, S.J. Xiao, J.S. Feng, W.Y. Yau, Y. Zhang, Deep subspace clustering with sparsity prior, in: *International Joint Conference on Artificial Intelligence*, 2016.
- [18] X. Peng, J.S. Feng, J.S. Xiao, W. Yau, J.T. Zho, S.F. Yang, Structured autoencoders for subspace clustering, *IEEE Trans. Image Process.* 27 (10) (2018) 5076–5086.
- [19] P. Ji, T. Zhang, H.D. Li, M. Salzmann, I. Reid, Deep subspace clustering networks, in: *Advanced Neural Information Processing Systems*, 2017.
- [20] V.M. Patel, R. Vidal, Kernel sparse subspace clustering, in: *IEEE International Conference on Image Processing*, 2014.
- [21] M. Yin, Y. Guo, J.B. Gao, H.Z. S., X.S. L., Kernel sparse subspace clustering on symmetric positive definite manifolds, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.
- [22] J.L. Xu, J.W. Han, F.P. Nie, X.L. Li, Re-weighted discriminatively embedded k-means for multi-view clustering, *IEEE Trans. Image Process.* 26 (6) (2017) 3016–3027.
- [23] C.Q. Zhang, H.Z. Fu, Q.H. Hu, X.C. Cao, Y. Xie, D.C. Tao, Generalized latent multi-view subspace clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (1) (2019) 86–99.
- [24] E. Elhamifar, R. Vidal, Sparse subspace clustering, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.
- [25] R. Vidal, P. Favaro, Low rank subspace clustering, *Pattern Recognit. Lett.* 43 (1) (2014) 47–61.
- [26] V.M. Patel, H.V. Nguyen, R. Vidal, Latent space sparse subspace clustering, in: *IEEE International Conference on Computer Vision*, 2013.
- [27] X. Peng, H.J. Tang, L. Zhang, Z. Yi, S.J. Xiao, A unified framework for representation-based subspace clustering of out-of-sample and large-scale data, *IEEE Trans. Neural Netw. Learn. Syst.* 27 (12) (2013) 2499–2512.
- [28] C.G. Li, R. Vidal, A structured sparse plus structured low-rank framework for subspace clustering and completion, *IEEE Trans. Signal Process.* 64 (24) (2016) 6557–6570.
- [29] Y.Z. Yang, J.S. Feng, N. Jojic, J.C. Yang, T.S. Huang, L0-sparse subspace clustering, in: *European Conference on Computer Vision*, 2016.
- [30] X.F. Zhu, Z.S. C., Y.G. Li, Z.J. L., Y.L. F., Low-rank sparse subspace for spectral clustering, *IEEE Trans. Knowl. Data Eng.* 31 (8) (2018) 1532–1543.
- [31] C.G. Li, C. You, R. Vidal, On geometric analysis of affine sparse subspace clustering, *IEEE J. Sel. Top. Sign. Proces.* 12 (6) (2018) 1520–1533.
- [32] A. Babaeian, M. Babae, A. Bayestehtashk, M. Bandarabadi, Nonlinear subspace clustering using curvature constrained distances, *Pattern Recognit. Lett.* 68P1 (2015) 118–125.
- [33] W. Zhu, W. Zhu, J. Lu, J. Zhou, Nonlinear subspace clustering, in: *IEEE International Conference on Image Processing*, 2018.
- [34] C.G. Li, C. You, R. Vidal, Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework, *IEEE Trans. Image Process.* 26 (6) (2017) 2988–3001.
- [35] J. Shen, P. Li, H. Xu, Online low-rank subspace clustering by basis dictionary pursuit, in: *International Conference on Machine Learning*, 2016.
- [36] M. Abdolali, N. Gillis, M. Rahmati, Scalable and robust sparse subspace clustering using randomized clustering and multilayer graphs, *Signal Process.* 163 (2019) 166–180.
- [37] R.S. Kasra, K. Mahdi, T.-d. Raimon, T. Laura, G. Richard, The application of subspace clustering algorithms in drill-core hyperspectral domaining, in: *Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing*, 2019.
- [38] L. Ma, Z. Liu, Hybrid sparse subspace clustering for visual tracking, in: *International Conference on Pattern Recognition*, 2018.
- [39] W. Zhu, B. Peng, H. Wu, B. Wang, Query set centered sparse projection learning for set based image classification, *Appl. Intell.* (2020) <http://dx.doi.org/10.1007/s10489-020-01730-3>.
- [40] M. Abavisani, V.M. Patel, Deep multimodal subspace clustering networks, *IEEE J. Sel. Top. Signal Process.* 12 (6) (2018) 1601–1614.
- [41] P.F. Zhu, B.Y. Hui, C.Q. Zhang, D.W. Du, L.Y. Wen, Q.H. Hu, Multi-view deep subspace clustering networks, 2018, arXiv:1908.01978.
- [42] M. Abdolali, M. Rahmati, Robust subspace clustering for image data using clean dictionary estimation and group lasso based matrix completion, *J. Vis. Commun. Image Represent.* 61 (2019) 303–314.
- [43] M. Abavisani, M.P. Vishal, Multimodal sparse and low-rank subspace clustering, *Inf. Fusion* 39 (2018) 168–177.
- [44] Y. Wang, Y.Y. Tang, L. Li, Minimum error entropy based sparse representation for robust subspace clustering, *IEEE Trans. Signal Process.* 63 (15) (2015) 4010–4021.
- [45] C. Su, Z. Wu, M. Yin, K. Li, W. Sun, Subspace clustering via independent subspace analysis network, in: *IEEE International Conference on Image Processing*, 2017.
- [46] M. Rahmani, K.G. Atia, Innovation pursuit: A new approach to subspace clustering, *IEEE Trans. Signal Process.* 65 (23) (2017) 6276–6291.

- [47] J. Wang, X. Wang, F. Tian, C. Liu, H. Yu, Constrained low-rank representation for robust subspace clustering, *IEEE Trans. Cybern.* 47 (12) (2017) 4534–4546.
- [48] W. Zhu, J. Lu, J. Zhou, Nonlinear subspace clustering, in: *IEEE International Conference on Image Processing*, 2017.
- [49] S. Hechmi, A. Gallas, E. Zagrouba, Multi-kernel sparse subspace clustering on the Riemannian manifold of symmetric positive definite matrices, *Pattern Recognit. Lett.* 125 (2019) 21–27.
- [50] W. Huang, M. Yin, J. Li, S. Xie, Deep clustering via weighted k-subspace network, *IEEE Signal Process. Lett.* 26 (11) (2019) 1628–1632.
- [51] X. Peng, C. Lu, Z. Yi, H. Tang, Connections between nuclear-norm and frobenius-norm-based representations, *IEEE Trans. Neural Netw. Learn. Syst.* 29 (1) (2016) 218–224.
- [52] H. Li, N. Chen, L. Li, Error analysis for matrix elastic-net regularization algorithms, *IEEE Trans. Neural Netw. Learn. Syst.* 23 (5) (2012) 737–748.
- [53] X.T. Zhen, M.Y. Yu, X.F. He, S. Li, Multi-target regression via robust low-rank learning, *IEEE Trans. Pattern Anal. Mach. Intell.* 40 (2) (2018) 497–504.
- [54] T. Bouwmans, S. Javed, H. Zhang, Z. Lin, R. Otazo, On the applications of robust PCA in image and video processing, *Proc. IEEE* 106 (8) (2018) 1427–1457.
- [55] G.A. Watson, Characterization of the subdifferential of some matrix norms, *Linear Algebra Appl.* 170 (1992) 33–45.
- [56] Y. Chong, C.G. Li, D.P. Robinson, R. Vidal, Oracle based active set algorithm for scalable elastic net subspace clustering, in: *IEEE Conference on Computer Vision and Pattern Recognition*, 2016.