Deep Contrastive Multi-view Subspace Clustering

Lei Cheng[1], Yongyong Chen[1,2], and Zhongyun Hua[1,2(✉)]

[1] School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China
{cyy2020,huazhongyun}@hit.edu.cn
[2] Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China

Abstract. Multi-view subspace clustering has become a hot unsupervised learning task, since it could fuse complementary multi-view information from multiple data effectively. However, most existing methods either fail to incorporate the clustering process into the feature learning process, or cannot integrate multi-view relationships well into the data reconstruction process, which thus damages the final clustering performance. To overcome the above shortcomings, we propose the deep contrastive multi-view subspace clustering method (DCMSC), which is the first attempt to integrate the contrastive learning into deep multi-view subspace clustering. Specifically, DCMSC includes multiple autoencoders for self-expression learning to learn self-representation matrices for multiple views which would be fused into one unified self-representation matrix to effectively utilize the consistency and complementarity of multiple views. Meanwhile, to further exploit multi-view relations, DCMSC also introduces contrastive learning into multi-autoencoder network and Hilbert Schmidt Independence Criterion (HSIC) to better exploit complementarity. Extensive experiments on several real-world multi-view datasets demonstrate the effectiveness of our proposed method by comparing with state-of-the-art multi-view clustering methods.

Keywords: Multi-view Subspace clustering · Contrastive learning · Hilbert Schmidt Independence Criterion

1 Introduction

Multi-view clustering, finding a consensus segmentation of data across multiple views, has become a hot unsupervised learning topic. Unlike single-view clustering, it faces multiple different descriptions or sources of the same data. How to fully exploit the consistency and complementarity of different views is the most significant challenge for multi-view clustering. Currently, multi-view clustering has made great progress and has played an important role in many practical applications. Most traditional methods first learned one common representation and then performed some single-view clustering methods. However, they would

M. Tanveer et al. (Eds.): ICONIP 2022, CCIS 1791, pp. 692–704, 2023.
https://doi.org/10.1007/978-981-99-1639-9_58

ignore the high-dimensionality of data, and their performance is greatly reduced when the dimensions of each view are extremely unbalanced.

Subspace clustering refers to find the underlying subspace structures of the data under the popular assumption that high-dimensional data could be well described in several low-dimensional subspaces. Recently, self-representation-based subspace clustering models have achieved great success. It assumes that each data point can be represented as a linear combination of other data points. Given a single-view data matrix consisting of n column vectors where each column represents a sample, self-representation properties can be formalized as follows:

$$\min_{C} \mathcal{L}(X, C) + \mathcal{R}(C) \quad s.t. \quad X = XC, \tag{1}$$

where $X = [x_1, x_2, \cdots, x_n]$. $\mathcal{L}(X, C)$ represents the self-representation loss and $\mathcal{R}(C)$ is the regularization term.

Recently, multi-view subspace clustering methods have made great success by extending the single-view subspace clustering methods. In general, there are two main ways to exploit multi-view information. The first way is to learn a common representation first, and then self-representation is conducted on the learned common representation. For instance, the direct way is to concatenate all multi-view features to form a combined feature. The second approach is to first conduct self-representation on each view separately, and then fuse them. In recent years, to solve the problem of insufficient representation ability and possible non-linearity of original data, deep learning-based methods have been proposed. For example, a unified network architecture composed of multiple autoencoders are designed to integrate the process of feature learning and multi-view relationship exploration into data clustering in [1,3]. Among them, [3] propose a multi-view deep subspace clustering networks (MvDSCN) in which the multi-view self-representation relation is learned by the end-to-end manner. Although they have achieved good results, there are still the following limitations: (1) Data representation learning and clustering processes are independently handled, and multi-view relationship cannot play a role in the feature extraction process. (2) Multi-view data reconstruction process is performed independently within each view, which ignores comprehensive information from multiple data. (3) They are unable to effectively handle the imbalance of multi-view dimensions.

To overcome the above shortcomings, we propose the Deep Contrastive Multi-view Subspace Clustering (DCMSC) method which mainly includes a base network to conduct self-representation learning and an additional module including Schmidt Independence Criterion (HSIC) regularizer and contrastive penalty. Specifically, the base network includes V independent autoencoders, each of which is used to extract the latent features of each original view, and the fully connected layer between the encoder and the decoder is used to obtain the self-representation matrix; HSIC part and contrastive learning part could effectively take advantage of multi-view relationships and mine the consistency and complementarity of multi-view data. Among them, the HSIC restriction module is used to punish the dependencies between the representations of each view and promote the diversity of subspace representations; the consistency of each sample point in each

view is achieved through the contrastive learning module. Finally, the combined result of the self-representation matrix obtained by all view-specific autoencoders is used to build the similarity matrix. The final clustering result is obtained by the spectral clustering algorithm. In summary, our contributions include:

1. For the first time, we integrate the general idea of contrastive learning into the multi-view subspace clustering problem and propose the deep contrastive multi-view subspace clustering method.
2. In DCMSC, the base network is mainly used to learn the view-specific self-representation matrix constrained by the additional network which could make good use of the multi-view relationship, so that the fusion using the learned V self-representation matrices has a more powerful representation ability, thereby achieving better clustering performance.
3. The contrastive learning regards different views in multiple views as data-enhanced versions and aims to explore the common semantics among multiple views while the Hilbert Schmidt Independence Criterion is used to discover the diversity of multi-view features. Extensive experiments on a wide range of datasets demonstrate that DCMSC achieves state-of-the-art clustering effectiveness.

2 Related Works

2.1 Subspace Clustering

Subspace clustering aims to reveal the inherent clustering structures of the data composed of multiple subspaces. Given a set of n samples $[x_1, x_2, \cdots, x_N] \in \mathbb{R}^{d \times N}$ in which d denotes the dimension of data, the basic model of self-representation subspace clustering methods can be described as follows:

$$\min_{C} \|C\|_p + \|X - XC\|_F^2, \tag{2}$$

where $\|\cdot\|_p$ is an arbitrary matrix norm. After optimizing the above formula, the obtained self-representation matrix C could describe the subspace clustering relationship between data points, and then is input into the spectral clustering algorithm to obtain the final clustering result. For example, Sparse Subspace Clustering (SSC) [21] aims to enhance sparsity of self-representation by imposing l_1-norm regularization on the self-representation matrix. To discover multi-subspace structures, Low-rank representation (LRR) [22] explored the multi-block diagonal properties of self-representation matrix. Essentially, self-representation based methods depend on assumption that each data point can be reconstructed by a linear combination of other points. However, the actual data may not meet this assumption. Many scholars have proposed using the kernel trick [4] to solve this problem, but the kernel technique is heuristic. With the powerful representation ability of neural networks, a large number of deep subspace clustering networks have been proposed to embed self-representation into deep autoencoder through fully connected layers, which has achieved the state-of-the-art performance. Deep adversarial subspace clustering leveraged the idea

of generative adversarial and added a GAN-like model into self-representation loss to evaluate clustering performance [23].

2.2 Multi-view Subspace Clustering

The multi-view clustering problem is faced with multiple representations of the same data. Compared with single-view data, multi-view data contains consensus information and complementary information from multiple views [5,6]. How to effectively fuse the information of each view is the key to the multi-view clustering task. For existing multi-view subspace clustering methods, there are currently three main categories. The first category is to perform self-representation learning on each view individually, and then fuse the results of individual self-representations. Divergent Multi-view Subspace Clustering (DiMSC) [5] proposed to exploit the complementarity from multi-view data by reducing redundancy. The second class firstly learns a common latent representation, and then performs self-representation learning on this latent representation. Latent Multi-view Subspace Clustering (LMSC) [6] explored complementary information from different views while building latent representation. The third category combines the above two ideas. Reciprocal Multi-layer Subspace Learning (RMSL) [2] simultaneously constructed the view-specific subspace representations and common representation to mutually restore the subspace structure of the data through the Backward Encoding Networks (BEN) and the Hierarchical Self-Representative Layers (HSRL). Multi-view Deep Subspace Clustering Network (MvDSCN) [3] proposed a network that can simultaneously learn view-specific self-representation and common self-representation, and leverages HSIC to capture nonlinear and higher-order inter-view relationships. However, due to their complex network structures and objective design, it is difficult to well optimize each objective function at the same time in the optimization process, and they ignore the role of each view describing the data in exploring the data representation and clustering structure.

2.3 Contrastive Learning

[7,8] is a popular unsupervised learning paradigm in recent years, whose main idea is to make the similarity between positive pair as close as possible while negative pair as far as possible. This learning paradigm has achieved great success on computer vision [9]. For example, [11] proposed a one-stage online clustering method, which conducted contrastive learning both at instance-level and cluster-level. [12,13] introduced contrastive learning into multi-view clustering. For example, a contrastive multi-view encoding framework [13] has been designed to capture the latent scene semantics. Multi-level Feature Learning for Contrastive Multi-view Clustering (MFLVC) [10] proposes a flexible multi-view contrastive learning framework, which can simultaneously achieve the coherence goal of high-level features and semantic labels. However, to the best of our knowledge, there is no related work that exploits the idea of regarding each view as a data-augmented version in contrastive learning and applies the idea of contrastive learning into the multi-view subspace clustering task.

3 Proposed Method

Given dataset with V views $\{X^v \in \mathbb{R}^{d_v \times N}\}_{v=1}^{V}$, where $X^v = [x_1^v, x_2^v, \cdots, x_N^v]$ and d_v and N is the number of data points and features in the v^{th} view, respectively, our goal is to find a common self-representation matrix C that can express the relationship between data points among the multi-view data. In this section, we describe the Deep Contrastive Multi-view Subspace Clustering (DCMSC) in details.

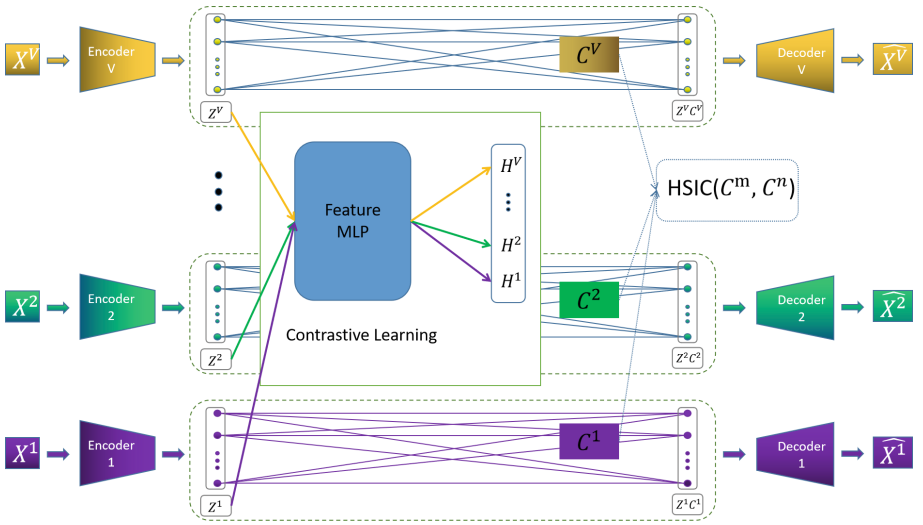

Fig. 1. Illustration of the proposed Deep Contrastive Multi-view Subspace Clustering (DCMSC) method. DCMSC builds V parallel autoencoders for latent feature extraction of view-specific data in which self-representation learning is conducted by a fully connected layer between encoder and decoder. Specifically, v^{th} original view X^v is encoded as Z^v and reconstructed to $\hat{X}^v$ through decoder. Self-representation matrix C^v is obtained by building a fully connected layer between the encoder and the decoder without activation function. Contrastive learning module is introduced into our network to exploit more common semantics information and HSIC constraint can effectively exploit the complementary information from multiple-view data, in which v^{th} high level semantic representation H^v is obtained for constructing contrastive loss. The final combination of all view-specific self-representation matrices further integrate the complementary and consistent information from multiple views.

3.1 The Proposed DCMSC

The network architecture of the proposed DCMSC method is shown in the Fig. 1, which consists of two modules, i.e., the base net which learns view-specific representation $\{C^v\}_{v=1}^{V}$ and the additional module including contrastive learning part and HSIC part which further exploits the multi-view relationship. In details, the

base net consists of V autoencoders, each of which conducts self-representation learning for each view-specific data. The encoder can be regarded as a function that simultaneously plays the role of dimensionality reduction and nonlinear conversion and the decoder is used to reconstruct the input features. Then self-representation is conducted by a fully connected layer without linear activation function and bias, which is built between the encoder and the decoder. Combining reconstruction loss in autoencoder into basic self-representation model i.e., Eq. (2), the loss of v^{th} autoencoder is summarized as follows:

$$\min_{C^v} \|C^v\|_F^2 + \|Z^v - Z^v C^v\|_F^2 + \|X^v - \hat{X}^v\|_F^2, \tag{3}$$

where Z^v is the output of encoder for v^{th} view X^v and $\hat{X}^v$ is the reconstruction of X^v.

Complementary and consistent information in multiple views can be exploited by summing the self-representation matrices of each view at the end. The study in [3,5] has proposed to introduce HSIC which measures the nonlinear and high-order correlations into multi-view subspace clustering to exploit more complementary information. Here, we adopt the empirical definition of HSIC proposed in [3]:

$$\mathcal{L}_{\mathbf{hsic}} = \sum_{ij} HSIC(C^m, C^n), \tag{4}$$

where C^m and C^n denote m^{th} and n^{th} self-representation matrix respectively, $HSIC((C^m, C^n) = trace((C^m)^T C^m H (C^n)^T C^n H)$ and H is a $N \times N$ square matrix with element $1 - \frac{1}{N}$.

It is worth noting that data reconstruction only in specific view cannot well exploit multi-view relational information. To alleviate this problem, we propose to introduce contrastive learning into our framework. Specifically, our contrastive learning module consists of a fully connected layer shared by all views. As shown in Fig. 1, let H^m denotes the output of the contrastive learning module for the latent representation of the m^{th} view Z^m as m^{th} high level semantic representation. Each high-level feature h_i^m has $(VN-1)$ feature pairs, i.e., $\{h_i^m, h_j^n\}_{j=1,\cdots,N}^{n=1,\cdots,V}$, which consist of $(V-1)$ positive pairs $\{h_i^m, h_i^n\}_{n\neq m,\cdots,N}$ and $V(N-1)$ negative pairs left. Contrastive learning aims to maximize the similarities of positive pairs while minimize that of negative pairs. Specifically, the contrastive loss between H^m and H^n is defined as [10]:

$$\ell_{fc}^{(mn)} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{d(\mathbf{h}_i^m, \mathbf{h}_i^n)/\tau_F}}{\sum_{j=1}^{N}\sum_{v=m,n} e^{d(\mathbf{h}_i^m, \mathbf{h}_j^v)/\tau_F} - e^{1/\tau_F}}, \tag{5}$$

where $d(x, y)$ measures the similarity between sample x and sample y and τ_F denotes the temperature parameter. Inspired by NT-Xent [7], we apply cosine distance:

$$d(x, y) = \frac{\langle x, y \rangle}{\|x\|\|y\|}. \tag{6}$$

The final contrastive loss is designed as accumulated losses among all views:

$$\mathcal{L}_{\mathbf{con}} = \sum_{m=1}^{v} \sum_{n \neq m} \ell_{fc}^{(mn)}. \tag{7}$$

Mathematically, the loss function of DCMSC is formulated by combining the above tentative loss function in Eqs. (3), (4), (7) as follows:

$$\begin{aligned} \mathcal{L}_{\mathbf{final}} &= \mathcal{L}_{\mathbf{ae}} + \alpha_1 \mathcal{L}_{\mathbf{self}} + \alpha_2 \mathcal{L}_{\mathbf{reg}} + \alpha_3 \mathcal{L}_{\mathbf{hsic}} + \alpha_4 \mathcal{L}_{\mathbf{con}} \\ &= \sum_{v=1}^{V} \|X^v - \hat{X}^v\|_F^2 + \alpha_1 \sum_{v=1}^{V} \|Z^v - Z^v C^v\|_F^2 + \alpha_2 \sum_{v=1}^{V} \|C^v\|_F^2 \\ &\quad + \alpha_3 \sum_{ij}^{V} HSIC(Z^i, Z^j) + \alpha_4 \sum_{m=1}^{V} \sum_{n \neq m}^{V} \ell_{fc}^{(mn)}, \end{aligned} \tag{8}$$

where $\mathcal{L}_{\mathbf{ae}} = \sum_{v=1}^{V} \|X^v - \hat{X}^v\|_F^2$, $\mathcal{L}_{\mathbf{self}} = \sum_{v=1}^{V} \|Z^v - Z^v C^v\|_F^2$ and $\mathcal{L}_{\mathbf{reg}} = \sum_{v=1}^{V} \|C^v\|_F^2$. Parameters α_1, α_2, α_3, and α_4 are non-negative ones to balance different contributions of different terms.

3.2 Optimization

The whole process of the proposed LDLRSC is summarized in Algorithm 1. We first pre-train the network without self-representation layer for more effective training in fine-tune stage and prevention of possible all-zero solution [3]. After the fine-tune stage, the final self-representation matrix C is calculated as $C = \sum_{v=1}^{V} C^v$. Generally, we can construct the affinity matrix simply by $(|C| + |C|^T)/2$ for spectral clustering. Here, we adopt the heuristic employed by SSC [21], which has been proved beneficial for clustering.

4 Experiment

4.1 Experimental Settings

Datasets. We conduct experiments on 6 benchmark multi-view datasets to evaluate our proposed DCMSC, including 4 classical datasets: Yale, ORL, Still DB and BBCSport and 2 bigger datasets: Caltech and BDGP. More details are listed in Table 1.

Evaluation Metrics. We adopt 4 widely used metrics to evaluate the clustering performance: accuracy (ACC), normalized mutual information (NMI), purity (PUR) and The F-measure. Note that higher values indicate better performance for the above 4 metrics. Parameters will be optimized to achieve the best clustering performance for all experiments. The average metric of 10 trials over each dataset is reported.

Algorithm 1. DCMSC

Input: Multi-view data $[X^1, X^2, \cdots, X^V]$;
Maximum iteration T_{max};
Trade-off parameters $\alpha_1, \alpha_2, \alpha_3, \alpha_4$;
The number of cluster K;
Output: Clustering result L;
1: Pre-train V autoencoders without self-representation layer;
2: Initialize the self-expression layer and contrastive learning net;
3: **while** $t \leq T_{max}$ **do**
4: Calculate the loss (8) and its gradient;
5: Do forward propagation;
6: **end while**
7: Calculate the final self-representation matrix $C = \sum_{v=1}^{V} C^v$;
8: Run algorithm employed by [21] to obtain affinity matrix A;
9: Run spectral clustering to get the clustering results L.

Table 1. The details of the datasets.

Datasets	#Samples	#Views	#Classes	Dimension of features
Yale	165	3	11	4096 /3304/6750
ORL	400	3	10	4096/3304/6750
Still DB	476	3	6	200/200/200
BBCSport	544	2	5	3183/3203
BDGP	2,500	2	5	1750/79
Caltech-3V	1,400	3	7	40/254/1984
Caltech-5V	1,400	5	7	40/254/1984/512/928

Comparison Methods. The comparison methods include some traditional state-of-the-art methods for both multi-view subspace clustering and deep multi-view clustering: BestSV [24], LRR [22], RMSC [31], DSCN [25], DCSC [26], DC [27], DMF [28], LMSC [6], MSCN [29], MvDSCN [3], RMSL [2], MVC-LFA [15], COMIC [16], IMVTSC [18], CDIMC-net [30], EAMC [17], SiMVC [20], CoMVC [20], MFLVC [10].

Implementation. We implement our DCMSC method on TensorFlow-2 in Python and evaluate its performance on several baseline methods. Adam optimizer is adopted for the gradient descent and the learning rate of the network is set to $1e^{-3}$. We choose ReLU as the activation function in the network except the self-expression layer.

4.2 Experimental Results

We compared DCMSC mainly with 8 subspace-based multi-view clustering algorithms on 4 datasets. To evaluate the superiority and robustness of our method, we also conduct experiments on 2 big datasets and compared its performance with 6 state-of-the-art multi-view clustering algorithms. The results are given in

Table 2 and Table 3. From Table 2, we can see that the proposed DCMSC significantly outperforms all methods on the first two datasets and performs comparable performance on the last two datasets. Obviously, DCMSC boosts the clustering performance by a large margin over other methods on Yale. The improvement of the proposed DCMSC over the second-best method FMR are 10.1%, 10.2%, and 18.5% with respect to NMI, ACC, and F-measure, respectively. From Table 3, there are following results: (1) our method obtain can also obtain competitive clustering performance on big data; (2) DCMSC greatly improves the clustering performance on Caltech-5V. In addition, we observe that although RMSL behaves on some benchmark datasets of multi-view subspace clustering, it does not obtain very competitive performances on BDGP and Caltech. In contrast, our method still maintains decent performance on other datasets.

4.3 Visualization

To intuitively show the superiority of DCMSSC, we visualized the affinity matrix A on BBCSport, ORL and Yale in Fig. 2, where A_{ij} denotes the similarity between sample x_i and sample x_j. Affinity A could be obtained from the final self-representation matrix C by algorithm employed by [21]. Noting the data points are sorted by classes on the above 3 datasets, the affinity matrix A should have a block-diagonal structure ideally. From Fig. 2, we can see that the affinity A learned by our proposed DCMSC well exhibits the block-diagonal property compared with MvDSCN.

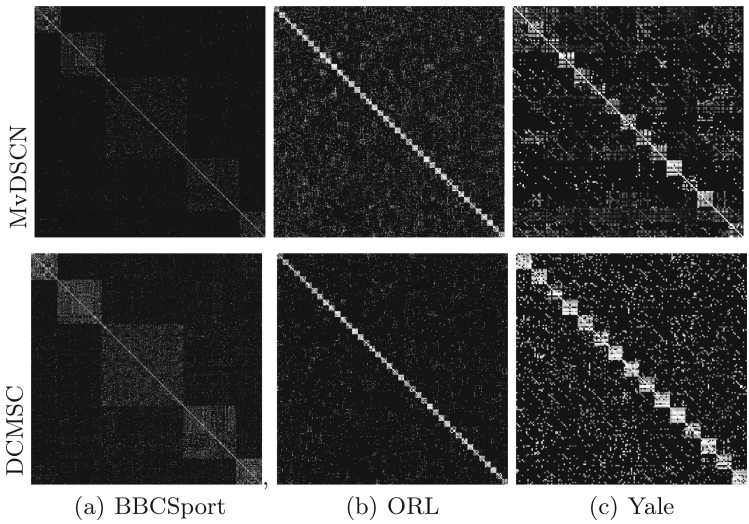

Fig. 2. Visualization of learned affinity matrix on BCCSport, ORL and Yale.

4.4 Ablation Studies

We conducted ablation studies on $\mathcal{L}_{\mathbf{con}}$ on Yale and Caltech-5V to illustrate the effectiveness of our contrastive learning module. Table 4 shows that our method

Table 2. Results of all methods on four small datasets. Bold indicates the best and underline indicates the second-best.

Datasets	Yale			ORL			Still DB			BBCSport		
Metrics	NMI	ACC	F-measure	NMI	ACC	F-measure	NMI	ACC	F-measure	NMI	ACC	F-measure
BestSV	0.654	0.616	0.475	0.903	0.777	0.711	0.104	0.297	0.221	0.715	0.836	0.768
LRR	0.709	0.697	0.547	0.895	0.773	0.731	0.109	0.306	0.240	0.690	0.832	0.774
RMSC	0.684	0.642	0.517	0.872	0.723	0.654	0.106	0.285	0.232	0.608	0.737	0.655
DSCN	0.738	0.727	0.542	0.883	0.801	0.711	0.216	0.323	0.293	0.652	0.821	0.683
DCSC	0.744	0.733	0.556	0.893	0.811	0.718	0.222	0.325	0.301	0.683	0.843	0.712
DC	0.756	0.766	0.579	0.865	0.788	0.701	0.199	0.315	0.280	0.556	0.724	0.492
LMSC	0.702	0.670	0.506	0.931	0.819	0.758	0.137	0.328	0.269	0.826	0.900	0.887
DMF	0.782	0.745	0.601	0.933	0.823	0.773	0.154	0.336	0.265	0.821	0.890	0.889
MSCN	0.769	0.772	0.582	0.928	0.833	0.787	0.168	0.312	0.261	0.813	0.888	0.854
MvDSCN	0.797	0.824	0.626	0.943	0.870	0.834	**0.245**	0.377	**0.320**	0.835	0.931	0.860
RMSL	0.831	0.879	0.828	0.950	0.881	0.842	0.135	0.336	0.293	**0.917**	**0.976**	**0.954**
DCMSC	**0.944**	**0.955**	**0.907**	**0.970**	**0.931**	**0.911**	0.156	**0.388**	0.284	0.864	0.953	0.907

Table 3. Results of all methods on BDGP, Caltech-3V and Caltech-5V. Bold indicates the best and underline indicates the second-best.

Datasets	BDGP			Caltech-3V			Caltech-5V		
Evaluation metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
RMSL [2] (2019)	0.849	0.630	0.849	0.596	0.551	0.608	0.354	0.340	0.391
MVC-LFA [15] (2019)	0.564	0.395	0.612	0.551	0.423	0.578	0.741	0.601	0.747
COMIC [16] (2019)	0.578	0.642	0.639	0.447	0.491	0.575	0.532	0.549	0.604
CDIMC-net [19] (2020)	0.884	0.799	0.885	0.528	0.483	0.565	0.727	0.692	0.742
EAMC [17] (2020)	0.681	0.480	0.697	0.389	0.214	0.398	0.318	0.173	0.342
IMVTSC-MVI [18] (2021)	0.981	0.950	0.982	0.558	0.445	0.576	0.760	0.691	0.785
SiMVC [20] (2021)	0.704	0.545	0.723	0.569	0.495	0.591	0.719	0.677	0.729
CoMVC [20] (2021)	0.802	0.670	0.803	0.541	0.504	0.584	0.700	0.687	0.746
MFLVC [10] (2022)	**0.989**	**0.966**	**0.989**	0.631	0.566	0.639	0.804	0.703	0.804
DCMSC	0.985	0.957	0.985	**0.890**	**0.785**	**0.890**	**0.914**	**0.825**	**0.914**

achieves good results even without $\mathcal{L}_{\mathbf{con}}$, and the better effect could be obtained with the $\mathcal{L}_{\mathbf{con}}$, which shows contrastive learning works to improve the performance for multi-view subspace task due to its ability to exploit more comprehensive relationship in multi-view data.

Table 4. Ablation studies for contrastive learning structures on Yale and Caltech-5V.

Datasets	Yale			Caltech-5V		
Evaluation metrics	ACC	NMI	PUR	ACC	NMI	PUR
w/o $\mathcal{L}_{\mathbf{con}}$	0.912	0.826	0.912	0.874	0.782	0.874
w/ $\mathcal{L}_{\mathbf{con}}$	**0.955**	**0.944**	**0.955**	**0.914**	**0.825**	**0.914**

5 Conclusion

In this paper, we proposed a novel method named Deep Contrastive Multi-view Subspace Clustering (DCMSC) to exploit the multi-view relationship by combining multiple self-representation matrix and introducing contrastive learning into the networks for exploring more consistent information. DCMSC consists of the base network composed of V autoencoders by which V view-specific self-representation matrices are learned. In addition, HSIC regularizer and contrastive learning module are included in our base network to exploit more comprehensive information. Experiments on both benchmark datasets and two bigger datasets verify the superiority and robustness of our method compared with the state-of-the-arts methods.

Acknowledgements. This work was supported in part by the National Natural Science Foundation of China under Grants 62071142 and 62106063, by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011406, by the Shenzhen College Stability Support Plan under Grants GXWD20201230155427003-20200824210638001 and GXWD20201230155427003-20200824113231001, by the Guangdong Natural Science Foundation under Grant 2022A1515010819, and by Guangdong Provincial Key Laboratory of Novel Security Intelligence Technologies under Grant 2022B1212010005.

References

1. Rui, M., Zhiping, Z.: Deep multi-view subspace clustering network with exclusive constraint. In: 2021 40th Chinese Control Conference (CCC), pp. 7062–7067 (2021)
2. Li, R., Zhang, C., Fu, H., Peng, X., Zhou, T., Hu, Q.: Reciprocal multi-layer subspace learning for multi-view clustering. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 8172–8180 (2019)
3. Zhu, P., Hui, B., Zhang, C., Du, D., Wen, L., Hu, Q.: Multi-view deep subspace clustering networks. arXiv Preprint arXiv:1908.01978 (2019)
4. Patel, V., Van Nguyen, H., Vidal, R.: Latent space sparse subspace clustering. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 225–232 (2013)
5. Cao, X., Zhang, C., Fu, H., Liu, S., Zhang, H.: Diversity-induced multi-view subspace clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–594 (2015)
6. Zhang, C., Hu, Q., Fu, H., Zhu, P., Cao, X.: Latent multi-view subspace clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4279–4287 (2017)
7. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning, pp. 1597–1607 (2020)
8. Wang, T., Isola, P.: Understanding contrastive representation learning through alignment and uniformity on the hypersphere. In: International Conference on Machine Learning, pp. 9929–9939 (2020)

9. Van Gansbeke, W., Vandenhende, S., Georgoulis, S., Proesmans, M., Van Gool, L.: SCAN: learning to classify images without labels. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12355, pp. 268–285. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58607-2_16
10. Xu, J., Tang, H., Ren, Y., Peng, L., Zhu, X., He, L.: Multi-level feature learning for contrastive multi-view clustering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 16051–16060 (2022)
11. Li, Y., Hu, P., Liu, Z., Peng, D., Zhou, J., Peng, X.: Contrastive clustering. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 8547–8555 (2021)
12. Hassani, K., Khasahmadi, A.: Contrastive multi-view representation learning on graphs. In: International Conference on Machine Learning, pp. 4116–4126 (2020)
13. Tian, Y., Krishnan, D., Isola, P.: Contrastive multiview coding. In: Vedaldi, A., Bischof, H., Brox, T., Frahm, J.-M. (eds.) ECCV 2020. LNCS, vol. 12356, pp. 776–794. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58621-8_45
14. Ji, P., Zhang, T., Li, H., Salzmann, M., Reid, I.: Deep subspace clustering networks. In: Advances in Neural Information Processing Systems, pp. 24–33 (2017)
15. Wang, S., et al.: Multi-view clustering via late fusion alignment maximization. In: IJCAI, pp. 3778–3784 (2019)
16. Peng, X., Huang, Z., Lv, J., Zhu, H., Zhou, J.: COMIC: multi-view clustering without parameter selection. In: International Conference on Machine Learning, pp. 5092–5101 (2019)
17. Zhou, R., Shen, Y.: End-to-end adversarial-attention network for multi-modal clustering. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14619–14628 (2020)
18. Wen, J., et al.: Unified tensor framework for incomplete multi-view clustering and missing-view inferring. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 10273–10281 (2021)
19. Wen, J., Zhang, Z., Xu, Y., Zhang, B., Fei, L., Xie, G.: CDIMC-net: cognitive deep incomplete multi-view clustering network. In: IJCAI, pp. 3230–3236 (2020)
20. Trosten, D., Løkse, S., Jenssen, R., Kampffmeyer, M.: Reconsidering representation alignment for multi-view clustering. In: CVPR, pp. 1255–1265 (2021)
21. Elhamifar, E., Vidal, R.: Sparse subspace clustering: algorithm, theory, and applications. IEEE Trans. Pattern Anal. Mach. Intell. 2765–2781 (2013)
22. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. 171–184 (2012)
23. Zhou, P., Hou, Y., Feng, J.: Deep adversarial subspace clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1596–1604 (2018)
24. Ng, A., Jordan, M., Weiss, Y.: On spectral clustering: analysis and an algorithm. In: Advances in Neural Information Processing Systems, vol. 14 (2001)
25. Peng, X., Xiao, S., Feng, J., Yau, W., Yi, Z.: Deep subspace clustering with sparsity prior. In: IJCAI, pp. 1925–1931 (2016)
26. Jiang, Y., Yang, Z., Xu, Q., Cao, X., Huang, Q.: When to learn what: deep cognitive subspace clustering. In: Proceedings of the 26th ACM International Conference on Multimedia, pp. 718–726 (2018)
27. Caron, M., Bojanowski, P., Joulin, A., Douze, M. Deep clustering for unsupervised learning of visual features. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 132–149 (2018)
28. Zhao, H., Ding, Z., Fu, Y.: Multi-view clustering via deep matrix factorization. In: Thirty-First AAAI Conference on Artificial Intelligence (2017)

29. Abavisani, M., Patel, V.: Deep multimodal subspace clustering networks. IEEE J. Sel. Top. Signal Process. **12**, 1601–1614 (2018)
30. Wen, J., Zhang, Z., Xu, Y., Zhang, B., Fei, L., Xie, G.: CDIMC-net: cognitive deep incomplete multi-view clustering network. In: Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020, pp. 3230–3236 (2020)
31. Xia, R., Pan, Y., Du, L., Yin, J.: Robust multi-view spectral clustering via low-rank and sparse decomposition. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 28 (2014)