Self-supervised Weighted Information Bottleneck for Multi-view Clustering

Zhengzheng Lou, **Chaoyang Zhang**, **Hang Xue**, **Yangdong Ye**, **Qinglei Zhou** and **Shizhe Hu** *
School of Computer and Artificial Intelligence, Zhengzhou University, China
ieshizhehu@gmail.com

Abstract

Multi-view clustering (MVC) is a long-standing topic in machine learning and data mining community, focusing on investigating and utilizing the relationships among views for final consistent data cluster structure discovery. Generally, weighted MVC is one of the popular methods working by learning and applying the view weight/importance on each view for fully exploring the complementary information across views. However, most existing weighted MVCs only consider the quality of each view, ignoring the vital role of pseudo label self-supervision information in weight learning. In this work, we propose a novel self-supervised weighted information bottleneck (SWIB) method for solving the multi-view clustering problem. It combines the weighted information from different views based on information bottleneck theory, and the view weight learning mechanism is newly designed by simultaneously taking into accounting both the quality of view-contained information and the self-supervised information on the data partition of each view. Experimental results on multi-view text, multi-feature image, multi-angle video, and multi-modal text-image dataset as well as large-scale datasets show the superiority of the SWIB method. To our knowledge, this is the first work incorporating the self-supervised learning into weighted multi-view clustering.

1 Introduction

Multi-view clustering (MVC) [Bickel and Scheffer, 2004] is a hot topic in machine learning and computer vision community, and it focuses on investigating and utilizing the close relationships among different views for learning the final consistent data cluster structure. And it has been successfully applied into many computer vision and pattern recognition fields, such as medical image analysis and image retrieval. Of all the MVC methods, short for MVCs, weighted MVC is one of the effective methods working by learning and applying the view weight or importance on multiple views for fully exploring the complementary information across views.

Recently, many weighted MVCs have been proposed and a few typical ones are listed in the followings. In the early period, a k-means based multi-view clustering method [Cai *et al.*, 2013] is designed with automatic weight learning, and also the model robustness is ensured. Additionally, Wang et al. [Wang *et al.*, 2020] proposed a multi-view graph-based method for data clustering problem, which produces an effective unified fused graph matrix with rank constrained parameter-free weight learning strategy, and further this learned fused matrix benefits the graph matrix learning of each view. More recently, Xia et al. [Xia *et al.*, 2023] introduces an effective tensorized bipartite graph learning method for MVC problem, worked by utilizing the between-view similarity and within-view similarity. It also incorporates the anchor point learning mechanism for efficient model learning. However, most existing weighted MVCs only consider the quality of each view, ignoring the vital role of pseudo label self-supervision information in weight learning.

In this work, we propose a novel self-supervised weighted information bottleneck (SWIB) method for solving the multi-view clustering problem, as shown in Figure 1. It combines the weighted information from different views based on information bottleneck theory, and the view weight learning mechanism is newly designed by simultaneously taking into accounting both the quality of view-contained information and the self-supervised information on the data partition of each view. Specifically, the view-contained part focuses on the quality of the relevant feature information of each view. The self-supervised part fully employs the pseudo-labels in each iteration to guide the learning of view weights. Both of the two parts are combined together with a proper balance between them to learn an effective view weights reflecting the importance of each view. The challenges of incorporating the self-supervised information into weight learning lie in two aspects. First, data samples from different views are always heterogeneous, thus directly using the self-supervised information for computing the importance of each view is very difficult. Second, lots of existing methods still need many parameters to regularize the weight distribution during the weight learning process, which makes the parameter tuning difficult. Thus, incorporating the self-supervised information into weight learning without bringing or only bringing one additional parameter is still challenging. Experimental re-

*Corresponding Author

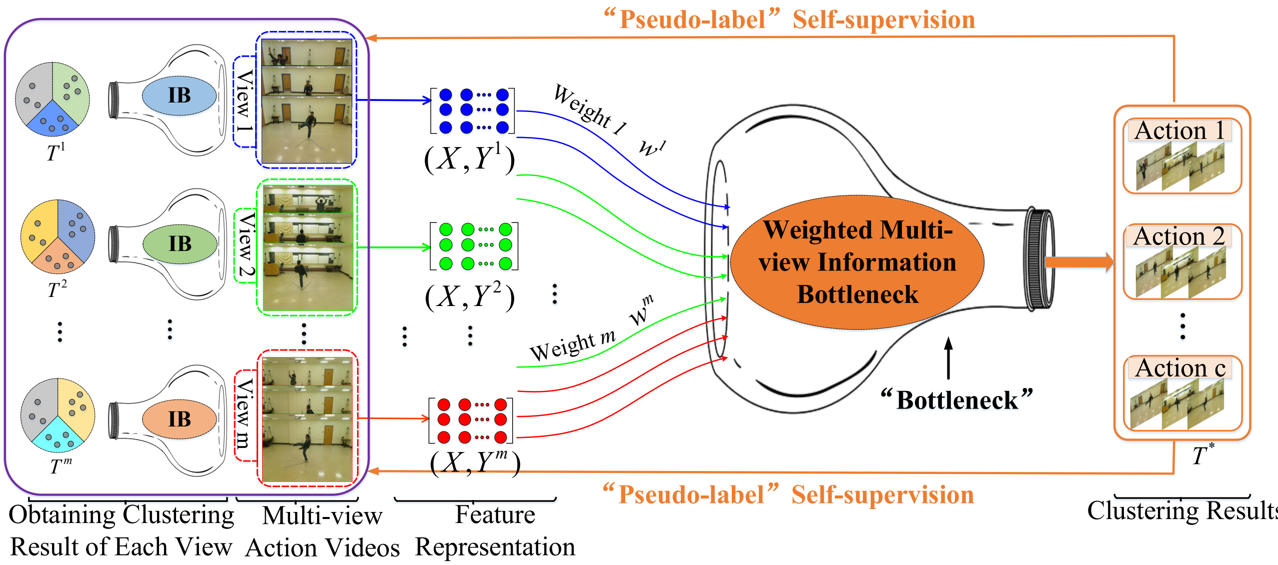

Figure 1: The overall framework of the proposed self-supervised weighted information bottleneck (SWIB) method. It combines the weighted information from different views based on information bottleneck theory, and the view weight learning mechanism is newly designed by simultaneously taking into accounting both the quality of view-contained information and the self-supervised information on the data partition of each view.

sults on various multi-view datasets, i.e., multi-view 20NGs text dataset, multi-feature COIL20 image dataset, multi-angle WVU video dataset, and multi-modal text-image PASCAL dataset as well as large-scale datasets (IAPR and ALOI) show the superiority of the SWIB method.

- To our knowledge, this is the first work incorporating the self-supervised learning into weighted multi-view clustering.
- A new view weight learning mechanism is designed by integrating both view-contained information and self-supervised information within and across views respectively.
- A novel SWIB method is proposed to comprehensively quantify the complementary information among views to enhance the MVC performance with only one parameter.
- Experimental results on various multi-view text, multi-feature image, multi-angle video, and multi-modal text-image dataset as well as large-scale datasets show the superiority of the SWIB method.

2 The Proposed SWIB Method

2.1 Prior Knowledge: Mutual Information and Information Bottleneck Theory

Information bottleneck (IB) [Tishby *et al.*, 1999] is an information-theoretic principle, and has been successfully applied to many real-world applications, such as unsupervised image segmentation [Bardera *et al.*, 2009] and classification [Lou *et al.*, 2013]. A more comprehensive survey on IB can be reached here [Hu *et al.*, 2024].

Recently, many multi-view learning methods via IB [Hu *et al.*, 2022b; Lou *et al.*, 2013; Hu *et al.*, 2022a; Zhang *et al.*, 2023] have been proposed and exhibited promising performance. Motivated by this, in this paper we tackle the multi-view clustering problem based on the IB theory.

IB regards the data clustering process as data compression of the input variable X into a compressed variable T, and meanwhile it preserves the relevant information about the feature variable Y as much as possible. The IB function is defined by

$$\mathcal{F}_{min}[p(t|x)] = I(T;X) - \beta I(T;Y). \tag{1}$$

where $I(T;X)$ denotes the mutual information between T and X, focusing on the data compactness quantification, $I(T;Y)$ denotes the mutual information between compact variable T and feature variable Y, focusing on the useful information preservation, $\beta \in (0, +\infty)$ is a balance parameter trading off both of them, and $p(t|x)$ denotes the data assignment probability.

To simplify the optimization, we give the max-version of the IB function

$$\mathcal{L}_{max}[p(t|x)] = I(T;Y) - \beta^{-1} I(T;X), \tag{2}$$

which is optimized by

$$p(t|x) = \frac{p(t)}{Z(x,\beta)} e^{-\beta D_{KL}[p(y|x)||p(y|t)]}, \tag{3}$$

where $p(t) = \sum_{x,y} p(x,y,t) = \sum_x p(x)p(t|x)$, $p(y|t) = \frac{1}{p(t)} \sum_x p(x,y)p(t|x)$, $Z(x,\beta)$ is a normalization function and D_{KL} is the *Kullback-Leibler* divergence [Cover and Thomas, 2006].

Note that we define the frequently-used concept of mutual information (MI) [Cover and Thomas, 2006] between two variables as

$$I(X,Y)=\sum_x\sum_y p(x,y)\log\frac{p(x,y)}{p(x)p(y)}. \quad (4)$$

where $p(x,y)$ denotes the joint probability distribution, $p(x)=\sum_y p(x,y)$ and $p(y)=\sum_x p(x,y)$.

2.2 Problem Formulation

Assume that we have data samples $\mathcal{X}=\{x_i\}_{i=1}^n$, denoted by a random variable X. These samples contain m different views, leading to m variables $\{Y^k\}_{k=1}^m$, where $Y^k\in R^{d^k}$ characterizes the samples from one view with d^k feature dimensionality. Based on that, we have m joint distributions $\{p(X,Y^k)_{k=1}^m\}$ reached by the classical Bag-of-Words model [Fei-Fei and Perona, 2005]. Our newly-designed SWIB method aims to explore and exploit the weighted complementary information among different views to obtain a good compact representations $p(t|x)$ of X to T. Note that we use c to indicate the number of clusters in each view.

2.3 The Overall Objective Function

In this part, motivated from the popular self-supervised learning, we propose a novel self-supervised weighted information bottleneck method for addressing the multi-view clustering problem, where the overall objective function is defined as follows

$$\mathcal{F}_{SWIB}=\sum_{i=1}^m w^i[I(T^*;Y)-\beta^{-1}I(T^*;X)], \quad (5)$$

where w^i denotes the weight value of the i-th view, β denotes the balance parameter trading off the information compression and preservation. Generally, the compact variable is a highly compressed from the input data, leading to $|T^*|\ll|X|$. Thus, we usually set the parameter β as $+\infty$ in the practical usage [Hu *et al.*, 2021]. And the specific view weight learning mechanism is given in the following parts.

2.4 View Weight Learning Mechanism

In this part, we design a new view weight learning mechanism for the above SWIB method, where the weight learning mainly contains two parts, including view-contained and self-supervised parts. The whole weight learning is defined as

$$w^i=\lambda W^i_{VC}+(1-\lambda)W^i_{SS}, \quad (6)$$

where W^i_{VC} denotes the view-contained part, W^i_{SS} denotes the self-supervised part, and λ is the trade-off parameter balancing the two parts.

View-contained Part

In view-contained part, the view quality is evaluated by mutual information between the clustering partition variable T^* and relevant feature variable Y^i for weight learning, as shown in Figure 2. The detailed formulation is given as

$$I^i_{VC}=I(T^*;Y^i)=H(T^*)+H(Y^i)-H(T^*;Y^i),$$

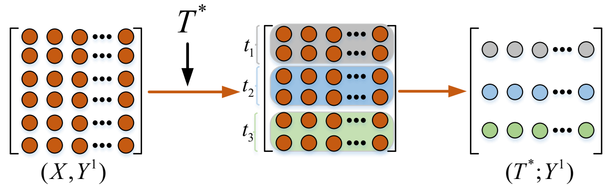

Figure 2: Process of computing the view-contained part for weight learning.

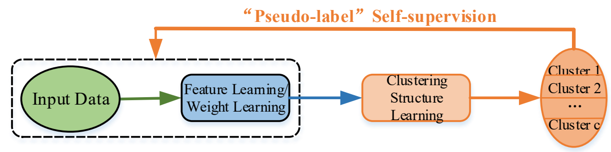

Figure 3: General idea of incorporating the self-supervision into multi-view learning paradigm.

where $H(T^*)$ and $H(T^i)$ are the entropy values of T^* and Y^i, and $H(T^*;Y^i)$ is the cross entropy of both variables.

Based on the above, we have the view-contained part for view weight learning as

$$W^i_{VC}=\frac{I^i_{VC}}{\sum_{i=1}^m I^i_{VC}} \quad (7)$$

Self-supervised Part

In self-supervised part, we incorporating the pseudo-labels as self-supervised signal for guiding the weight learning process, as shown in Figure 3. It works by using the metric of MI between the data partition variable T^* and the partition of each view T^i, shown as follows.

$$I^i_{SS}=\frac{I(T^*;T^i)}{[H(T^*)+H(T^i)]/2},$$

where $I(T^*;T^i)$ indicates the MI between the final data partition T^* and the data partition of the i-th view, $H(T^*)$ and $H(T^i)$ are their entropy values.

Based on the above, we have the self-supervised part for view weight learning as

$$W^i_{SS}=\frac{I^i_{SS}}{\sum_{i=1}^m I^i_{SS}} \quad (8)$$

2.5 Analysis on the Model

Observing from the proposed model, there are mainly two advantages compared with existing weighted multi-view clustering methods.

- Unlike existing models, the proposed method incorporates the self-supervised learning into view weight learning mechanism. To our knowledge, this is the first work doing so.
- Our weighted multi-view clustering method is one of the few models containing only one parameter in the objective function, which benefits for its usage in practical applications.

Algorithm 1 Algorithm for Optimizing the SWIB

Input: m joint probability distributions $\{p(X, Y^i)\}_{i=1}^m$, cluster number c, and the parameter λ.
Output: Final partition T^*.
Local Clustering:
  $\{T^1, T^2, \ldots, T^m\} \leftarrow$ Applying information bottleneck on $\mathcal{X}$ into c clusters on different views;
repeat
  $i \leftarrow 1$;
  while $i \leq m$ **do**
    for all $x \in \mathcal{X}$ **do**
      Draw x from its original cluster;
      Reassign x into clusters from the current i-th view, and calculate the merger cost;
      Merge x into a new cluster with minimal merger cost;
    end for
    $i \leftarrow i + 1$;
  end while
  Update the view weight values with the Eq. (6);
until Data partition unchanged or a fixed number of iterations finished

2.6 Optimization

In this part, we adopt the draw-and-merge sequential optimization method for solving the proposed method. It is a k-means like method by sequentially drawing each data sample from its original cluster and iteratively merging to a new cluster with minimal merger cost until the final clustering results unchanged or a fix number of iterations. The readers may refer to the related works [Lou *et al.*, 2013] for details, and the algorithm framework is given in Algorithm 1.

3 Experimental Validation

3.1 Datasets

We adopt multi-view text, multi-feature image, multi-angle video, and multi-modal text-image dataset as well as large-scale datasets to show the superiority of the SWIB method and the details are shown in Table 1.

Multi-view Text. 20NGs dataset [1] has 500 documents from the popular 20 Newsgroups dataset. Every document is processed by three various methods, corresponding to each view of the dataset.

Multi-feature Image. COIL20 dataset [2] contains images of 20 different objects on a motorized turntable. These images are captured by a fixed camera with a turntable rotated in 360° at 5° change to make different object posture. We use 3 kinds of features, i.e., shape, color and texture representation, namely SIFT [Lowe, 2004], Color Attention [Khan *et al.*, 2009], and TPLBP [Wolf *et al.*, 2008] respectively to characterize each view.

Multi-angle Video. WVU dataset [3] contains human action videos of 10 types captured from different angles. The non-adjacent views 1, 3, 5 and 7 are used for experimental validation, where the setting is quite similar to real-world scenario, e.g., road video surveillance. We adopt space-time interest points technique [Laptev, 2005] for feature extraction with 1000 dimensionality.

Multi-modal Text-Image. PASCAL dataset [4] has 20 clusters with 1000 data samples, and they are captured from image and text aspect, where the images are extracted with SIFT representation and the documents are extracted with BoW model.

3.2 Compared Methods

We have the following compared methods classified into two types.

Classical Single/All-view clustering. k-Means (KM) and Information Bottleneck (IB) [Tishby *et al.*, 1999], and all-view version of them, namely, All-view KM and IB, short for AVKM and AVIB respectively, where multiple views are combined and the two methods are then applied.

State-of-the-art MVCs. We select 10 state-of-the-art MVCs for experimental comparison, in which RMKMC, MfIB, DEKM, MLAN, GMC and SMVSC are all based on learning weights.

1. MVIB [Gao *et al.*, 2007]: A typical multi-view IB method proposed in early years by imposing compatible constraints on individual clusterings.
2. RMKMC [Cai *et al.*, 2013]: A multi-view k-means clustering method with robustness guaranteed.
3. MfIB [Lou *et al.*, 2013]: A multi-feature IB method based on weight learning designed for unsupervised image clustering.
4. DEKM [Xu *et al.*, 2016]: A discriminatively embedded multi-view k-means method for addressing clustering problem.
5. MLAN [Nie *et al.*, 2018]: A weighted multi-view learning method without penalty parameters for alleviating outliers and noise.
6. GMC [Wang *et al.*, 2020]: A multi-view graph clustering method by learning a fused unified matrix regularized by a rank constraint.
7. SMVSC [Sun *et al.*, 2021]: A scalable MVC method with anchor learning for large-scale data clustering.
8. FPMVS-CAG [Wang *et al.*, 2022]: A fast MVC method by simultaneous graph learning and consensus anchor learning for large-scale setting.
9. OMSC [Chen *et al.*, 2022]: An efficient orthogonal MVC method with joint learning of representation and clustering.
10. TBGL [Xia *et al.*, 2023]: A tensorized bipartite graph learning for MVC by utilizing the between-view similarity and within-view similarity.

[1] http://lig-membres.imag.fr/grimal/data.html
[2] https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[3] https://community.wvu.edu/vkkulathumani/wvu-action.html
[4] https://aclanthology.org/W10-0721.pdf

Dataset	Type	# View	# Samples	# Clusters	# Dimensionality
20NGs	Text	3	500	5	2000
COIL20	Image	3	1440	20	1000
WVU	Video	4	650	10	1000
PASCAL	Text-Image	2	1000	20	(300, 500)
Large Dataset	Type	# View	# Samples	# Clusters	# Dimensionality
IAPR	Text-Image	2	7855	6	(1200, 500)
ALOI	Image	3	11025	100	(77, 64, 64)

Table 1: Details of various kinds of multi-view datasets.

Method	20NGs		COIL20		WVU		PASCAL	
	Acc	NMI	Acc	NMI	Acc	NMI	Acc	NMI
KM	23.68±0.39	6.41±0.64	52.05±2.25	64.59±1.22	27.42±4.30	33.87±5.82	45.95±3.66	50.74±2.42
IB	85.48±10.57	76.07±7.68	68.43±3.89	78.44±2.61	52.23±2.50	50.75±0.84	50.19±6.08	51.65±2.54
AVKM	21.12±0.53	1.49±0.45	44.56±6.84	59.30±7.40	27.22±5.60	24.28±7.23	14.94±0.54	10.90±0.48
AVIB	91.80±9.88	92.72±6.71	77.52±5.99	91.42±2.55	53.48±1.63	51.37±0.36	24.70±0.14	30.23±0.15
MVIB (DASFAA'07)	94.22±1.37	83.21±3.18	62.19±10.50	73.54±6.81	60.08±3.31	62.73±2.97	23.00±1.58	25.26±1.42
RMKMC (IJCAI'13)	48.20±3.43	36.86±4.12	53.06±3.82	72.02±2.99	46.15±3.77	52.41±4.69	51.20±2.12	52.74±2.36
MfIB (IJCAI'13)	93.76±2.89	85.11±4.54	78.99±3.88	88.54±1.27	56.35±6.67	59.62±4.22	50.14±3.40	57.75±1.79
DEKM (CVPR'16)	31.60±0.00	13.48±0.00	40.69±0.00	59.52±0.00	54.77±0.00	58.79±0.00	15.20±0.00	13.14±0.00
MLAN (TIP'18)	96.40±0.11	89.18±0.17	86.91±0.40	94.49±0.00	37.85±5.85	39.39±5.14	51.73±0.07	51.58±0.06
GMC (TKDE'20)	98.20±0.00	93.92±0.00	73.33±0.00	92.11±0.00	46.62±0.00	55.15±0.00	27.00±0.00	34.96±0.00
SMVSC (ACM MM'21)	63.70±5.53	50.69±5.18	66.15±3.93	82.42±1.16	45.43±2.62	44.89±2.09	36.46±1.22	37.45±1.03
FPMVS-CAG (TIP'22)	73.80±0.00	59.23±0.00	69.17±0.00	85.41±0.00	49.08±0.00	49.39±0.00	39.00±0.00	38.65±0.00
OMSC (KDD'22)	73.80±0.00	59.23±0.00	72.01±0.00	83.69±0.00	56.15±0.00	59.22±0.00	33.30±0.00	34.71±0.00
TBGL (TPAMI'23)	89.11±0.00	83.45±0.00	79.17±0.00	89.24±0.00	58.40±0.00	59.66±0.00	38.90±0.00	56.59±0.00
SWIB	**99.80±0.00**	**99.30±0.00**	**89.06±4.55**	**94.68±2.28**	**62.12±6.72**	**64.58±4.35**	**55.09±3.50**	**61.03±2.02**

Table 2: Clustering performance with Acc and NMI on various kinds of datasets

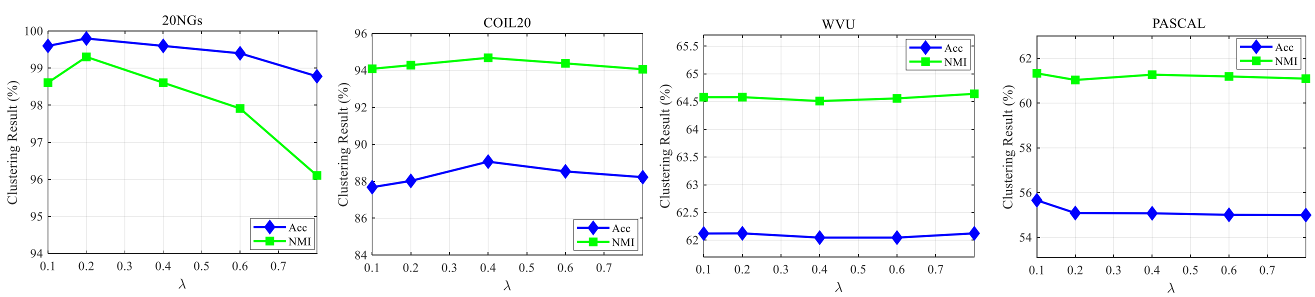

Figure 4: Parameter analysis of our method on different datasets.

3.3 Settings of Experiments

For single-view methods, the best clustering Accuracy (Acc) and Normalized Mutual Information (NMI) values on multiple views are shown. For state-of-the-art MVCs, we refer to the settings of parameter values in their papers and give the optimal performance with respect to the ideal parameter setting. For our SWIB method, we set β as $+\infty$ and search the parameter λ from the range $\{0.1, 0.2, 0.4, 0.6, 0.8\}$. The detailed parameter selection is given in the following subsection.

3.4 Analysis on Multi-view Clustering Results

The frequently-used clustering metrics of Acc and NMI are used for verifying the effectiveness of the proposed SWIB method. The lower values of both reveal poorer clustering performance. For all the methods, we take 10 runs for all the multi-view text, image, video and text-image datasets, and then reveal them in Table 2 with average deviation, where $\lambda = 0.2$ for 20NGs, WVU and PASCAL datasets, and $\lambda = 0.4$ for COIL20 dataset.

From the table, we have the following conclusions. Our SWIB method consistently outperform the clustering results reached by single/all-view KM and IB methods on all the datasets with a significant improvement, e.g., WVU and PASCAL dataset. This clearly reflects the advantages of the proposed method over single/all-view clustering methods. However, on some datasets, the view-concatenated all-view methods can not always beat the results obtained by their single-

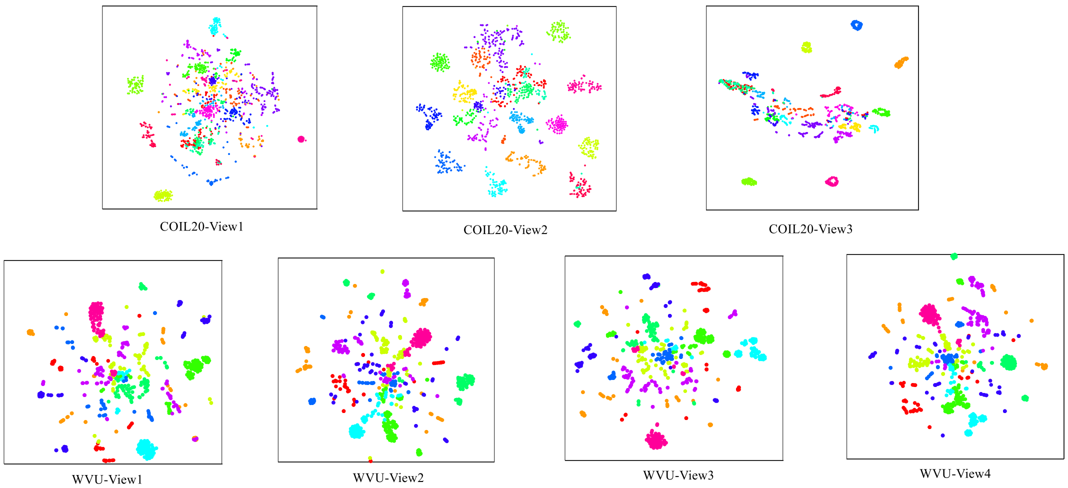

Figure 5: t-SNE analysis of our method on COIL20 and WVU datasets.

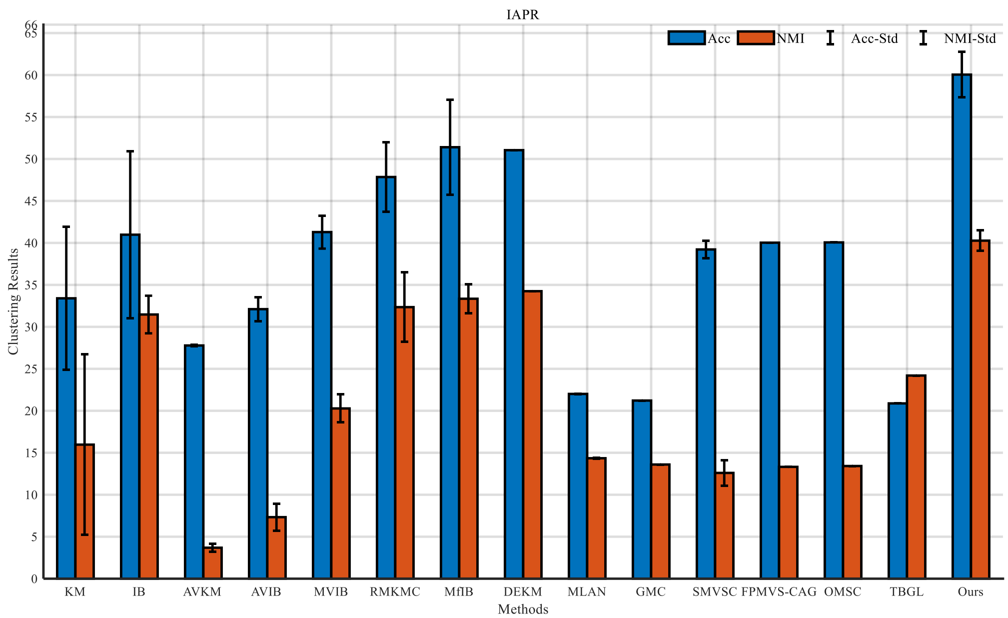

Figure 6: Clustering results on IAPR dataset. Note that -Std denotes the standard deviation.

view version methods. To some extent, this suggests that we have to resort to the paradigm of multi-view clustering. Actually, from the results, the proposed SWIB reaches much better results when compared with state-of-the-art MVCs methods, e.g., RMKMC, GMC, SMVSC, and OMSC method. This probably benefits from the new design of view weight learning mechanism with combination of view-contained information from individual views and self-supervised information from pseudo-labels of each iteration.

3.5 Parameter Selection

For our SWIB method, we set β as $+\infty$ and search the parameter λ from the range $\{0.1, 0.2, 0.4, 0.6, 0.8\}$. Figure 4 shows the clustering Acc and NMI results corresponding to each parameter setting. From this figure, we find that the clustering results maintain almost stable on the parameter search range, and suggests that it is not difficult to select a better parameter in usage and also exhibits possible practicability of SWIB method into real-world scenario. Note that we set $\lambda = 0.2$ for 20NGs, WVU and PASCAL datasets, and $\lambda = 0.4$ for COIL20 dataset.

3.6 Visualization Validation

For further illustrating the clustering results of the SWIB method, we select two datasets of COIL20 and WVU for clustering visualization with the popular method of t-SNE. The results are revealed in Figure 5, and we obtain that a majority of these views suggest a very compact cluster structure and a margin between different clusters, e.g., the view 2 and 3 of COIL20 dataset. This phenomenon further reflects that our

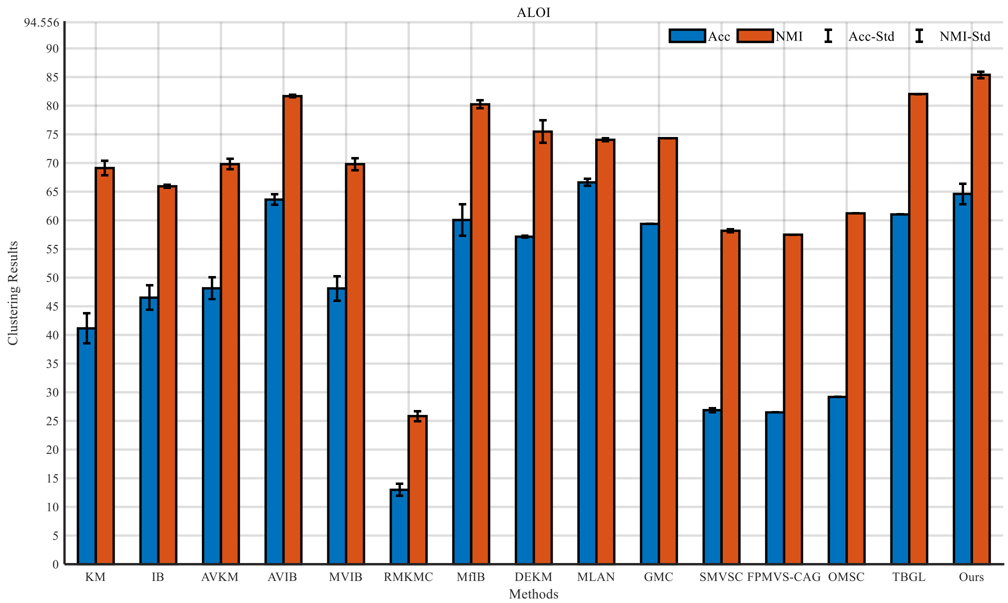

Figure 7: Clustering results on ALOI dataset. Note that -Std denotes the standard deviation.

SWIB method is capable of obtaining satisfactory clustering results with the new effective view weight learning mechanism.

3.7 Effectiveness on Large Datasets

To further verify the scalability of the proposed SWIB method, we conduct additional experiments on two large-scale multi-view dataset, i.e., multi-modal text-image IAPR dataset and mult-feature ALOI image dataset. IAPR dataset contains 7855 samples of 6 clusters and we extract SIFT image representation and BoW text representation for view description. Additionally, ALOI image dataset has 11025 image samples of 100 clusters, and we extract color similarity, RGB and Haralick feature representations as different views. We show the clustering results of them with compared methods and SWIB method on Figure 6 and Figure 7, where $\lambda = 0.2$ for both of the large multi-view datasets. It is seen that the proposed SWIB method consistently works better than the classical single/all-view clustering methods and most state-of-the-art MVC methods (except the second best Acc on ALOI dataset where MLAN obtains the best Acc) on both of the large datasets. These two figures also reveal the scalability of the proposed SWIB and the potential application into practical multi-view environments.

4 Conclusion

In this paper, a novel self-supervised weighted information bottleneck (SWIB) method is designed for addressing the challenging multi-view clustering problem. We combine the weighted information from different views based on information bottleneck theory, and then introduce a new view weight learning mechanism by simultaneously taking into accounting both the quality of view-contained information and the self-supervised information on the data partition of each view. Experimental results on multi-view text, image, video, and multi-modal text-image dataset as well as large-scale multi-view datasets have demonstrated its effectiveness. In the future, we will focus more on the deep version of the weighted MVC method [Huang *et al.*, 2023; Cui *et al.*, 2023], reliable multi-view learning [Xu *et al.*, 2024], incomplete multi-view clustering [Wen *et al.*, 2023; Wen *et al.*, 2021] and extend to more practical applications.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 62206254), the China Postdoctoral Science Foundation (No. 2023M743186), and the National Natural Science Foundation of China (No. 62176239).

Contribution Statement

All authors make contributions for the paper's methodology and experiments. Zhengzheng Lou and Chaoyang Zhang contributed equally to this work and should be considered co-first authors. Shizhe Hu is the corresponding author.

References

[Bardera *et al.*, 2009] Anton Bardera, Jaume Rigau, Imma Boada, Miquel Feixas, and Mateu Sbert. Image segmentation using information bottleneck method. *IEEE Trans. Image Process.*, 18(7):1601–1612, 2009.

[Bickel and Scheffer, 2004] Steffen Bickel and Tobias Scheffer. Multi-view clustering. In *Proceedings of the 4th IEEE International Conference on Data Mining (ICDM)*, pages 19–26, 2004.

[Cai *et al.*, 2013] Xiao Cai, Feiping Nie, and Heng Huang. Multi-view k-means clustering on big data. In *Proceed-*

ings of the 23rd International Joint Conference on Artificial Intelligence, pages 2598–2604, 2013.

[Chen *et al.*, 2022] Mansheng Chen, Chang-Dong Wang, Dong Huang, Jian-Huang Lai, and Philip S. Yu. Efficient orthogonal multi-view subspace clustering. In *The 28th ACM SIGKDD*, pages 127–135, 2022.

[Cover and Thomas, 2006] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[Cui *et al.*, 2023] Chenhang Cui, Yazhou Ren, Jingyu Pu, Xiaorong Pu, and Lifang He. Deep multi-view subspace clustering with anchor graph. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI*, pages 3577–3585, 2023.

[Fei-Fei and Perona, 2005] Li Fei-Fei and Pietro Perona. A bayesian hierarchical model for learning natural scene categories. In *IEEE CVPR*, pages 524–531, 2005.

[Gao *et al.*, 2007] Yan Gao, Shiwen Gu, Jianhua Li, and Zhining Liao. The multi-view information bottleneck clustering. In *12th International Conference on Database Systems for Advanced Applications, DASFAA*, volume 4443, pages 912–917, 2007.

[Hu *et al.*, 2021] Shizhe Hu, Xiaoqiang Yan, and Yangdong Ye. Multi-task image clustering through correlation propagation. *IEEE Trans. Knowl. Data Eng.*, 33(3):1113–1127, 2021.

[Hu *et al.*, 2022a] Shizhe Hu, Zhengzheng Lou, and Yangdong Ye. View-wise versus cluster-wise weight: Which is better for multi-view clustering? *IEEE Trans. Image Process.*, 31:58–71, 2022.

[Hu *et al.*, 2022b] Shizhe Hu, Zenglin Shi, and Yangdong Ye. DMIB: dual-correlated multivariate information bottleneck for multiview clustering. *IEEE Trans. Cybern.*, 52(6):4260–4274, 2022.

[Hu *et al.*, 2024] Shizhe Hu, Zhengzheng Lou, Xiaoqiang Yan, and Yangdong Ye. A survey on information bottleneck. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, pages 1–20, 2024.

[Huang *et al.*, 2023] Zongmo Huang, Yazhou Ren, Xiaorong Pu, Shudong Huang, Zenglin Xu, and Lifang He. Self-supervised graph attention networks for deep weighted multi-view clustering. In *Thirty-Seventh AAAI Conference on Artificial Intelligence*, pages 7936–7943, 2023.

[Khan *et al.*, 2009] Fahad Shahbaz Khan, Joost van de Weijer, and María Vanrell. Top-down color attention for object recognition. In *Proc. ICCV*, pages 979–986, 2009.

[Laptev, 2005] Ivan Laptev. On space-time interest points. *Int. J. Comput. Vis.*, 64(2-3):107–123, 2005.

[Lou *et al.*, 2013] Zhengzheng Lou, Yangdong Ye, and Xiaoqiang Yan. The multi-feature information bottleneck with application to unsupervised image categorization. In *Proceedings of the 23rd International Joint Conference on Artificial Intelligence*, pages 1508–1515, 2013.

[Lowe, 2004] David G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004.

[Nie *et al.*, 2018] Feiping Nie, Guohao Cai, Jing Li, and Xuelong Li. Auto-weighted multi-view learning for image clustering and semi-supervised classification. *IEEE Trans. Image Process.*, 27(3):1501–1511, 2018.

[Sun *et al.*, 2021] Mengjing Sun, Pei Zhang, Siwei Wang, Sihang Zhou, Wenxuan Tu, Xinwang Liu, En Zhu, and Changjian Wang. Scalable multi-view subspace clustering with unified anchors. In *ACM Multimedia Conference*, pages 3528–3536, 2021.

[Tishby *et al.*, 1999] Naftali Tishby, Fernando C Pereira, and William Bialek. The information bottleneck method. In *Proc. Annual Allerton Conf. Communnication Control Computing*, pages 368–377, 1999.

[Wang *et al.*, 2020] Hao Wang, Yan Yang, and Bing Liu. GMC: graph-based multi-view clustering. *IEEE Trans. Knowl. Data Eng.*, 32(6):1116–1129, 2020.

[Wang *et al.*, 2022] Siwei Wang, Xinwang Liu, Xinzhong Zhu, Pei Zhang, Yi Zhang, Feng Gao, and En Zhu. Fast parameter-free multi-view subspace clustering with consensus anchor guidance. *IEEE Trans. Image Process.*, 31:556–568, 2022.

[Wen *et al.*, 2021] Jie Wen, Zheng Zhang, Zhao Zhang, Lunke Fei, and Meng Wang. Generalized incomplete multiview clustering with flexible locality structure diffusion. *IEEE Trans. Cybern.*, 51(1):101–114, 2021.

[Wen *et al.*, 2023] Jie Wen, Chengliang Liu, Gehui Xu, Zhihao Wu, Chao Huang, Lunke Fei, and Yong Xu. Highly confident local structure based consensus graph learning for incomplete multi-view clustering. In *CVPR*, pages 15712–15721, 2023.

[Wolf *et al.*, 2008] Lior Wolf, Tal Hassner, and Yaniv Taigman. Descriptor based methods in the wild. In *Real-Life Images workshop at ECCV*, 2008.

[Xia *et al.*, 2023] Wei Xia, Quanxue Gao, Qianqian Wang, Xinbo Gao, Chris Ding, and Dacheng Tao. Tensorized bipartite graph learning for multi-view clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):5187–5202, 2023.

[Xu *et al.*, 2016] Jinglin Xu, Junwei Han, and Feiping Nie. Discriminatively embedded k-means for multi-view clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pages 5356–5364, 2016.

[Xu *et al.*, 2024] Cai Xu, Jiajun Si, Ziyu Guan, Wei Zhao, Yue Wu, and Xiyue Gao. Reliable conflictive multi-view learning. In *AAAI*, pages 16129–16137, 2024.

[Zhang *et al.*, 2023] Chaoyang Zhang, Zhengzheng Lou, Qinglei Zhou, and Shizhe Hu. Multi-view clustering via triplex information maximization. *IEEE Trans. Image Process.*, 32:4299–4313, 2023.