JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

GCFAgg: Global and Cross-view Feature Aggregation for Multi-view Clustering

Weiqing Yan[1,5] Yuanyang Zhang[1] Chenlei Lv[2] Chang Tang[3*]
Guanghui Yue[4] Liang Liao[5] Weisi Lin[5]

1 2,4 3 5

ID: THU-AM-321

Outline

Background

Traditional learning-based method:

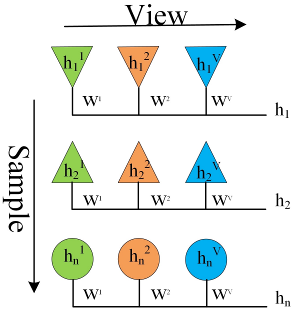
- Matrix factorization-based. (Wang *et al.*, 2022)
- Multi-view graph clustering. (Nie *et al.*, 2017, Hu *et al.*, 2020, Jing *et al.*, 2021, Lin *et al.*, 2021, Wang *et al.*, 2022)
- Multi-view subspace clustering. (Cao *et al.*, 2015, Kang *et al.*, 2020, Lv *et al.*, 2021, Sun *et al.*, 2021, Liu *et al.*, 2022)

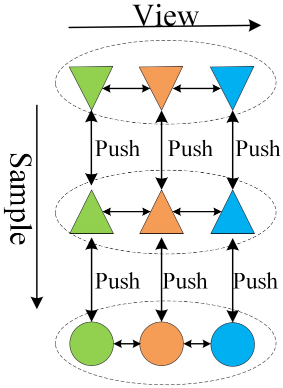
Original features or specified kernel features usually include noises and redundancy.

Deep representation learning-based method:

- Multi-level feature learning for contrastive multi-view clustering. (Xu *et al.*, 2022)
- Adversarial attention network for multi-modal clustering. (Zhou *et al.*, 2021)
- Deep incomplete multi-view clustering via contrastive prediction. (Lin *et al.*, 2021)
- Deep safe incomplete multi-view clustering (Tang *et al.*, 2022)

Shortcoming





Based on view-wise fusion models, Ignore a potential prior that is the presence of correlation between samples.

Distinguish the positive pair and negative pair from the sample-level. Be conflict with the clustering objective.

The Proposed framework

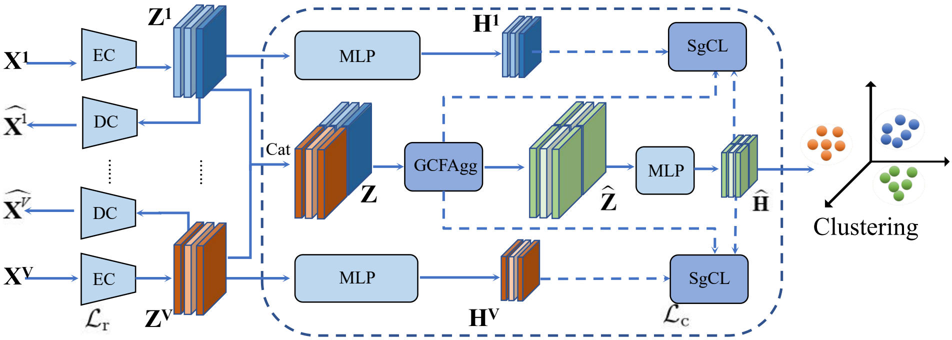


Figure: GCFAgg: global and cross-view feature aggregation module, SgCL:structure-guided multiview contrastive learning module.

GCFAgg

Cross-view Fusion

$$\begin{bmatrix} \mathbf{R}_{1:} \\ \mathbf{R}_{2:} \\ \vdots \\ \mathbf{R}_{n:} \end{bmatrix} = \begin{bmatrix} \mathbf{z}_1^1 & \mathbf{z}_1^2 & \cdots & \mathbf{z}_1^V \\ \mathbf{z}_2^1 & \mathbf{z}_2^2 & \cdots & \mathbf{z}_2^V \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{z}_n^1 & \mathbf{z}_n^2 & \cdots & \mathbf{z}_n^V \end{bmatrix} \begin{bmatrix} \mathbf{W}_{R1:} \\ \mathbf{W}_{R2:} \\ \vdots \\ \mathbf{W}_{RV:} \end{bmatrix} \tag{1}$$

$$\mathbf{R}_{j:} = \sum_{k=1}^{V} \mathbf{z}_j^k \mathbf{W}_{Rk:}; \mathbf{Q_1}_{j:} = \sum_{k=1}^{V} \mathbf{z}_j^k \mathbf{W}_{Q_1 k:}; \mathbf{Q_2}_{j:} = \sum_{k=1}^{V} \mathbf{z}_j^k \mathbf{W}_{Q_2 k:} \tag{2}$$

where $\mathbf{z}_j^k \in \mathbb{R}^{1\times d_v}$ is the feature representation of the j-th sample in the k-th view. W_{Q_1}, W_{Q_2}, W_R: achieve feature space transformation of cross-view.

Cross-sample Fusion

The structure relationship among samples is denoted as:

$$\mathbf{S} = \mathrm{softmax}\left(\frac{\mathbf{Q}_1\mathbf{Q}_2^T}{\sqrt{d}}\right). \tag{3}$$

The data representation enhanced by structure relationship is denoted as:

$$\hat{\mathbf{z}}_i = \sum_{j=1}^{n} \mathbf{S}_{ij}\mathbf{R}_{j:} \qquad \hat{\mathbf{Z}} = [\hat{\mathbf{z}}_1; \hat{\mathbf{z}}_2; \ldots; \hat{\mathbf{z}}_n] \tag{4}$$

where $\mathbf{R}_{j:}$ is the j-thcross-view fused feature representation.

SgCL:Structure-guided MV CL

Standard CL in the MVC task

The contrastive learning in MVC task is shown as follows:

$$\mathcal{L}_c = -\frac{1}{2N}\sum_{i=1}^{N}\sum_{v=1}^{V}\sum_{u\neq v}\log\frac{e^{C(\mathbf{H}_{i:}^{u},\mathbf{H}_{i:}^{v})/\tau}}{\sum_{j=1}^{N}\sum_{m=u,v}e^{C\left(\mathbf{H}_{i:}^{u},\mathbf{H}_{j:}^{m}\right)/\tau}-e^{1/\tau}} \tag{5}$$

Even if the *i*-th and *j*-th samples are from the same class, they are set as a negative pair.

SgCL

The proposed Structure-guided multiview Contrastive Learning is:

$$\mathcal{L}_c = -\frac{1}{2N}\sum_{i=1}^{N}\sum_{v=1}^{V}\log\frac{e^{C\left(\hat{\mathbf{H}}_{i:},\mathbf{H}_{i:}^{v}\right)/\tau}}{\sum_{j=1}^{N}e^{(1-\mathbf{S}_{ij})\,C\left(\hat{\mathbf{H}}_{i:},\mathbf{H}_{j:}^{v}\right)/\tau}-e^{1/\tau}} \tag{6}$$

When the structure relationship S_{ij} between the i-th and j-th sample is low (not from the same cluster), their corresponding representations are inconsistent.

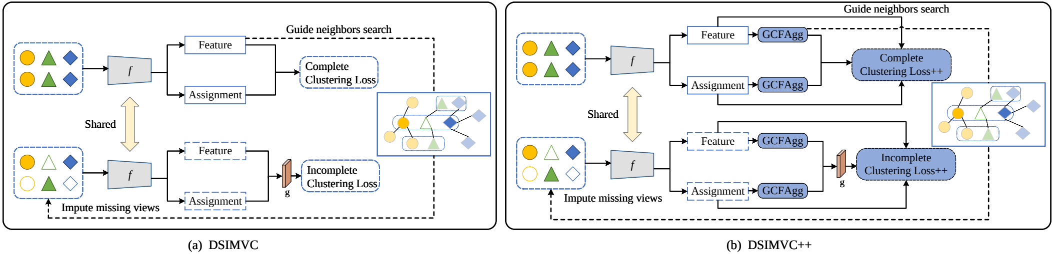


Figure: Improved incomplete MVC (DSIMVC++).

The feature alignment contrastive loss:

$$\mathcal{L}_{F++}\left(f\left(\mathcal{D}^c; w\right)\right) = \sum_{i=1}^{n_c} \sum_{q=1}^{m} \left[-\frac{2}{n_c} f_{\widehat{H}}\left(x_i; w_H\right)^T f_Z\left(x_i^q; w\right) + \frac{1}{n_c(n_c-1)} \sum_{j \neq i} \left(1 - S_{i,j}\right) \left(f_{\widehat{H}}\left(x_i; w_H\right)^T, f_Z\left(x_j^q; w\right) \right)^2 \right] \quad (7)$$

The clustering assignment alignment loss:

$$\mathcal{L}_{C++}\left(f\left(\mathcal{D}^c; w\right)\right) = -\frac{1}{K} \sum_{q=1}^{m} \sum_{j=1}^{K} \left[\log \frac{e^{\widehat{Q_j}^T Q_j^q}}{\sum_{s \neq j} e^{\widehat{Q_j}^T Q_s^q}} \right] \quad (8)$$

Experimental Setup

Datasets:

Table: Description of the multiview datasets.

Datasets	Samples	Views	Clusters
Prokaryotic	551	3	4
Synthetic3d	600	3	3
MNIST-USPS	5000	2	10
CCV	6773	3	20
Hdigit	10000	2	10
Cifar10	50000	3	10
Cifar100	50000	3	100
YouTubeFace	101499	5	31
Caltech-5V	1400	5	7
NGs	500	3	5
Cora	2708	2	7
BDGP	2500	2	5
Fashion	10000	3	10

Compared methods:

- PLCMF (IEEE T CYBERNETICS, 2022)
- LMVSC (AAAI, 2020)
- SMVSC (ACM MM, 2021)
- FastMICE (IEEE TKDE, 2022)
- DEMVC (INS, 2021)
- CONAN (IEEE BigData, 2021)
- SiMVC (IEEE CVPR, 2021)
- CoMVC (IEEE CVPR, 2021)
- MFLVC (IEEE CVPR, 2022)

  Incomplete MVC methods:
- CDIMC (IJCAI, 2020)
- COMPLETER (IEEE CVPR, 2021)
- DIMVC (AAAI, 2022)
- DSIMVC (ICML, 2022)

Experimental results

Datasets	CCV			MNIST-USPS			Prokaryotic			Synthetic3d		
Metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
PLCMF	0.2294	0.1852	0.2557	0.6228	0.6594	0.6670	0.4446	0.0200	0.5681	0.8867	0.6555	0.8867
LMVSC	0.2014	0.1657	0.2396	0.5626	0.5039	0.6060	0.5753	0.1337	0.6294	0.9567	0.8307	0.9567
SMVSC	0.2182	0.1684	0.2439	0.7542	0.6883	0.7542	0.5590	0.1820	0.5717	0.9683	0.8665	0.9683
FastMICE	0.1997	0.1518	0.2341	0.9570	0.9332	0.9573	0.5629	0.2685	0.6500	0.9613	0.8490	0.9613
DEMVC	0.1942	0.2113	0.2169	0.8858	0.9100	0.8880	0.5245	0.3079	0.6969	0.8100	0.6136	0.8100
CONAN	0.1422	0.1016	0.1674	0.5722	0.5708	0.6178	0.4809	0.1589	0.5045	0.9650	0.8540	0.9650
SiMVC	0.1513	0.1252	0.2161	0.9810	0.9620	0.9810	0.5009	0.1945	0.6098	0.9366	0.7747	0.9366
CoMVC	0.2962	0.2865	0.2976	0.9870	0.9760	0.9890	0.4138	0.1883	0.6697	0.9530	0.8184	0.9520
MFLVC	0.3123	0.3162	0.3391	0.9954	0.9869	0.9898	0.4301	0.2216	0.5989	0.9650	0.8537	0.9650
Ours	**0.3543**	**0.3292**	**0.3812**	**0.9956**	**0.9871**	**0.9956**	**0.6225**	**0.3778**	**0.7314**	**0.9700**	**0.8713**	**0.9700**

Datasets	Hdigit			YouTubeFace			Cifar10			Cifar100		
Metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
PLCMF	0.9047	0.7965	0.9047	0.1473	0.1237	0.2875	0.8144	0.8265	0.8497	0.8260	0.9593	0.8698
LMVSC	0.9709	0.9293	0.9709	0.1479	0.1327	0.2816	0.9896	0.9721	0.9896	0.8482	0.9583	0.9582
SMVSC	0.8634	0.7683	0.8634	0.2587	0.2292	0.3321	0.9899	0.9730	0.9899	0.7429	0.9091	0.7529
FastMICE	0.9332	0.9258	0.9417	0.1825	0.1633	0.3028	0.9694	0.9622	0.9704	0.8257	0.9464	0.8298
DEMVC	0.3738	0.3255	0.4816	0.2487	0.0932	0.2662	0.4354	0.3664	0.4498	0.5048	0.8343	0.5177
CONAN	0.9562	0.9193	0.9562	0.1179	0.1178	0.1499	0.9255	0.8641	0.9255	0.6711	0.9441	0.9983
SiMVC	0.7854	0.6705	0.7854	0.0765	0.0481	0.2662	0.8359	0.7324	0.8359	0.5795	0.9225	0.5869
CoMVC	0.9032	0.8713	0.9032	0.1010	0.0851	0.2674	0.9275	0.8925	0.9275	0.6569	0.9345	0.6570
MFLVC	0.9442	0.8750	0.9440	0.2770	0.2952	0.3297	0.9918	0.9774	0.9918	0.8268	0.9560	0.8268
Ours	**0.9744**	**0.9305**	**0.9744**	**0.3262**	**0.3289**	**0.4007**	**0.9923**	**0.9781**	**0.9923**	**0.9597**	**0.9935**	**0.9605**

Results on incomplete datasets

Dataset	Missing rates	0.1			0.3			0.5			0.7		
	Evaluation metrics	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
BDGP	CDIMC	0.8047	0.7008	0.8037	0.7467	0.6764	0.7527	0.6771	0.5451	0.6771	0.5611	0.3970	0.5776
	COMPLETER	0.4091	0.4180	0.4154	0.3963	0.3319	0.3115	0.3262	0.2747	0.4390	0.4359	0.4510	0.4090
	DIMVC	0.9640	0.8920	0.9120	0.9540	0.8660	0.8890	0.9470	0.8450	0.8730	0.9290	0.8020	0.8310
	DSIMVC	0.9827	0.9443	0.9827	0.9693	0.9034	0.9693	0.9529	0.8611	0.9529	0.9214	0.7937	0.9214
	DSIMVC++ (Our)	**0.9836**	**0.9455**	**0.9836**	**0.9698**	**0.9050**	**0.9698**	**0.9557**	**0.8685**	**0.9557**	**0.9332**	**0.8142**	**0.9332**
Synthetic3d	CDIMC	0.5965	0.3564	0.6000	0.5124	0.2239	0.5340	0.5330	0.2373	0.5663	0.4136	0.1387	0.4394
	COMPLETER	-	-	-	-		-	-	-	-	-		-
	DIMVC	**0.8183**	0.6701	0.8380	**0.8233**	0.5860	0.8241	**0.7968**	0.5355	0.7971	0.6689	0.3974	0.6774
	DSIMVC	0.7613	0.6744	0.8943	0.7378	0.6365	0.8773	0.7247	0.6090	0.8643	0.7043	0.5499	0.8242
	DSIMVC++ (Our)	0.7785	**0.6933**	**0.9005**	0.7530	**0.6693**	**0.9042**	0.7612	**0.6463**	**0.8952**	**0.7197**	**0.5900**	**0.8638**
Cora	CDIMC	0.2460	0.0111	0.3066	0.2222	0.0066	0.3024	0.2400	0.0052	0.3022	0.2518	0.0054	0.3025
	COMPLETER	0.2441	**0.4300**	0.3172	0.2542	**0.4130**	0.3242	0.2464	**0.4070**	0.3199	0.2540	0.1850	0.3055
	DIMVC	0.4384	0.2231	0.5079	0.3704	0.1470	0.4082	0.3561	0.1432	0.4275	0.2789	0.0718	0.3397
	DSIMVC	0.4402	0.3316	0.5445	0.4106	0.2924	0.5099	0.3764	0.2360	0.4742	0.3243	0.1628	0.4228
	DSIMVC++ (Our)	**0.4699**	0.3271	**0.5588**	**0.4484**	0.3035	**0.5544**	**0.4338**	0.2720	**0.5290**	**0.3554**	**0.1935**	**0.4620**
NGs	CDIMC	0.3072	0.0794	0.3216	0.2736	0.0478	0.2832	0.2532	0.0346	0.2620	0.2464	0.0270	0.2504
	COMPLETER	-	-	-	-	-	-	-	-	-	-	-	-
	DIMVC	0.3543	0.1493	0.3562	0.2120	0.0363	0.2138	0.2213	0.0588	0.2280	0.2598	0.0546	0.2645
	DSIMVC	0.5564	0.4599	0.6230	0.5178	0.3864	0.5854	**0.4672**	**0.2980**	0.5244	0.4090	0.2095	0.4746
	DSIMVC++ (Our)	**0.6358**	**0.5186**	**0.7090**	**0.6136**	**0.4428**	**0.6734**	0.4598	0.2801	**0.5310**	**0.4410**	**0.2298**	**0.5054**
Fashion	CDIMC	0.6500	0.6642	0.6696	0.5064	0.5121	0.5241	0.4484	0.4483	0.4553	0.3693	0.3668	0.3818
	COMPLETER	-	-	-	-	-	-	-	-	-	-	-	-
	DIMVC	0.7811	0.8578	0.8286	0.7132	0.7676	0.7614	0.7044	0.7447	0.7508	0.6128	0.6806	0.6693
	DSIMVC	0.8798	0.8623	0.8800	0.8680	0.8379	0.8687	0.8333	0.8025	0.8337	0.7825	0.7626	0.7825
	DSIMVC++ (Our)	**0.9360**	**0.8953**	**0.9360**	**0.9160**	**0.8657**	**0.9160**	**0.8969**	**0.8366**	**0.8969**	**0.8637**	**0.8015**	**0.8644**

Model analysis

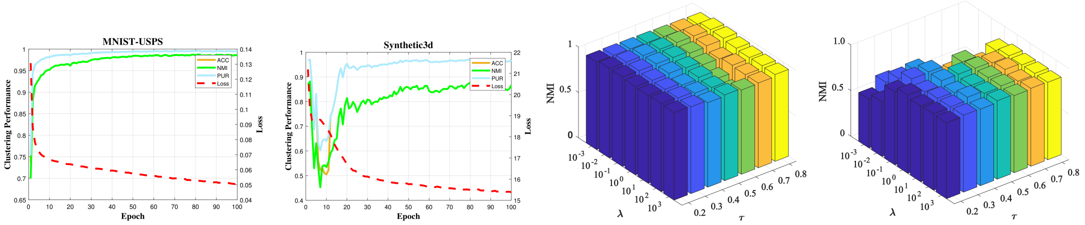
Table: Ablation study.

Datasets	Method	ACC	NMI	PUR
Prokaryotic	No-GCFAgg	0.4403	0.1906	0.5740
	No-SgCL	0.4804	0.2226	0.6534
	GCFAggMVC	**0.6225**	**0.3778**	**0.7314**
CCV	No-GCFAgg	0.2850	0.2740	0.3150
	No-SgCL	0.2020	0.1900	0.2560
	GCFAggMVC	**0.3543**	**0.3292**	**0.3812**
MNIST-USPS	No-GCFAgg	0.9753	0.9500	0.9753
	No-SgCL	0.6949	0.6656	0.7410
	GCFAggMVC	**0.9956**	**0.9871**	**0.9956**

Table: The ablation study for the SgCL.

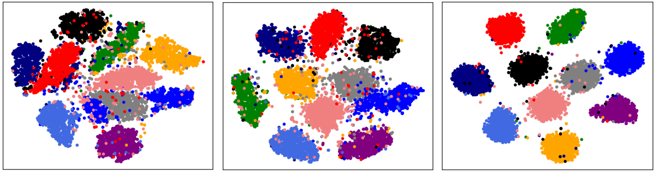
Datasets	Method	ACC	NMI	PUR
CCV	Standard CL	0.2711	0.2669	0.3046
	Standard CL with **S**	0.3046	0.3017	0.3363
	SgCL without **S**	0.2858	0.2833	0.3260
	SgCL	**0.3543**	**0.3292**	**0.3812**
MNIST-USPS	Standard CL	0.9562	0.9386	0.9562
	Standard CL with **S**	0.9768	0.9527	0.9768
	SgCL without **S**	0.9698	0.9327	0.9698
	SgCL	**0.9956**	**0.9871**	**0.9956**

Model analysis



(a) (b) (c) (d)

Figure: The convergence analysis and parameter analysis on MNIST-USPS, Synthetic3d.



(a) **Z** features (b) **H** features (c) $\hat{\mathbf{H}}$ features

Figure: The visualization results of different feature representations on different layers after convergence. Note that, **H** feature denotes the concatenation of all learnt $\mathbf{H}^v$.

Conclusion

- Propose a cross-sample and cross-view feature aggregation module (GCFAgg), which makes the representations of samples with highly structure relationship be more similar.
- Design a Structure-guided Contrastive Learning module, which addresses the conflict of negative pairs with the clustering objective.
- The proposed modules are suitable to the complete MVC task and incomplete MVC tasks.

ArXiv: `https://arxiv.org/pdf/2305.06799v1.pdf`
Code: `https://github.com/Galaxy922/GCFAggMVC`