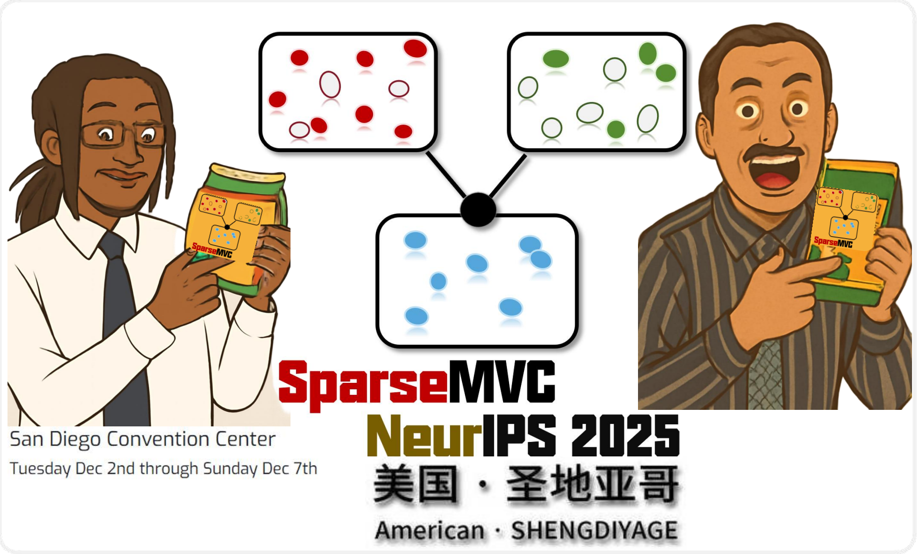
NEURAL INFORMATION PROCESSING SYSTEMS

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

Ruimeng Liu · Xin Zou · Chang Tang* ·
Xiao Zheng · Xingchen Hu · Kun Sun · Xinwang Liu

刘睿萌



San Diego Convention Center
Tuesday Dec 2nd through Sunday Dec 7th

SparseMVC

SCAN ME

Web link: https://neurips.cc/virtual/2025/poster/117045

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

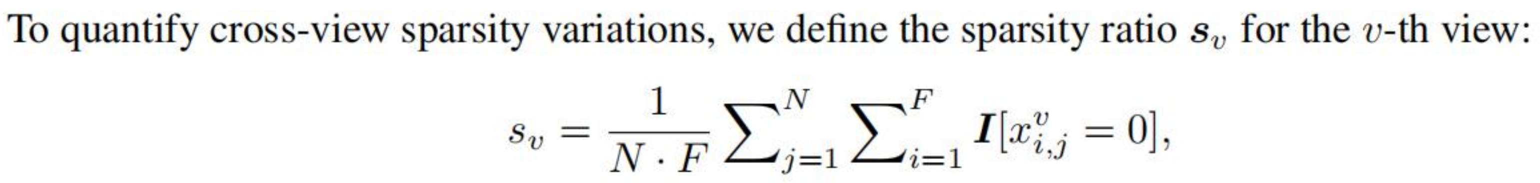
We explicitly identify, analyze, and define the problem of cross-view sparsity variations in multi-view data,

Identify problem, solve problem.

and to propose a dedicated framework SparseMVC that offers a targeted and principled solution.

1.Multi-view clustering： **Multi-view** refers to **data composed of multiple views from distinct sources.** **Clustering**, which emphasizes **unsupervised learning**, essentially treats clustering algorithms as probes to assess the **representational performance** of the features extracted by the network.

2.View sparsity ratio in multi-view datasets：



To quantify cross-view sparsity variations, we define the sparsity ratio s_v for the v-th view:

$$s_v = \frac{1}{N \cdot F} \sum\nolimits_{j=1}^{N} \sum\nolimits_{i=1}^{F} \boldsymbol{I}[x_{i,j}^{v} = 0],$$

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

Identify problem: cross-view sparsity variations

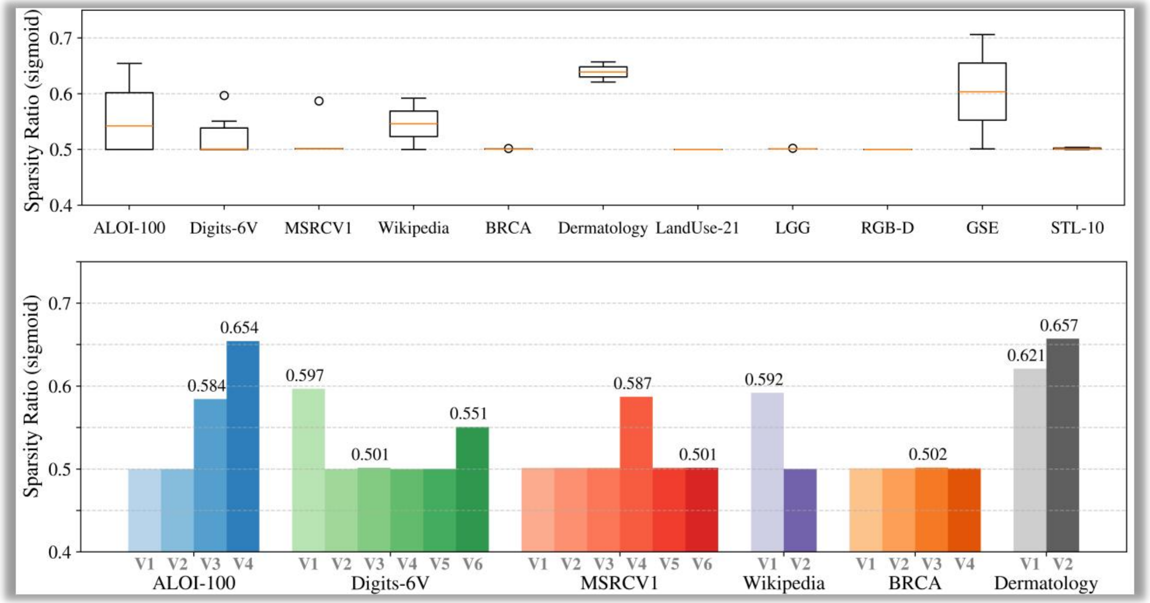
1. Multi-view data consist of **multiple distinct data sources.**

2. **Sparsity is a prevalent characteristic** in multi-view data.

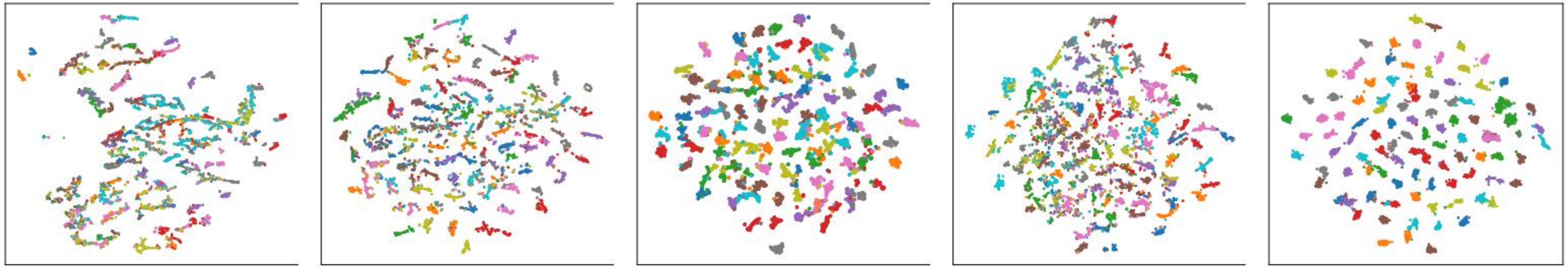
For a single dataset, does **sparsity variation exist among the internal views?**

Exist and exist widely

Sparsity ratios across views in multi-view datasets.



- **Top box plot**: Distribution, median (orange line), interquartile range (box), outliers (outside whiskers).
- **Bottom bar plot**: Sparsity ratios for each view in the datasets.
- **The sparsity ratios in Figure are transformed using the sigmoid function,** which shifts the baseline from 0 (the bottom of the image) to 0.5 (the middle of the image) for better visualization, as many views have very small sparsity values (less than 0.01) that would be hard to see with the original scale.



(a) GCFAgg (CVPR'23) (b) CPSPAN(CVPR'23) (c) SCMVC (TMM'24) (d) MVCAN (CVPR'24) (e) SparseMVC (Ours)

T-SNE visualization for comparison

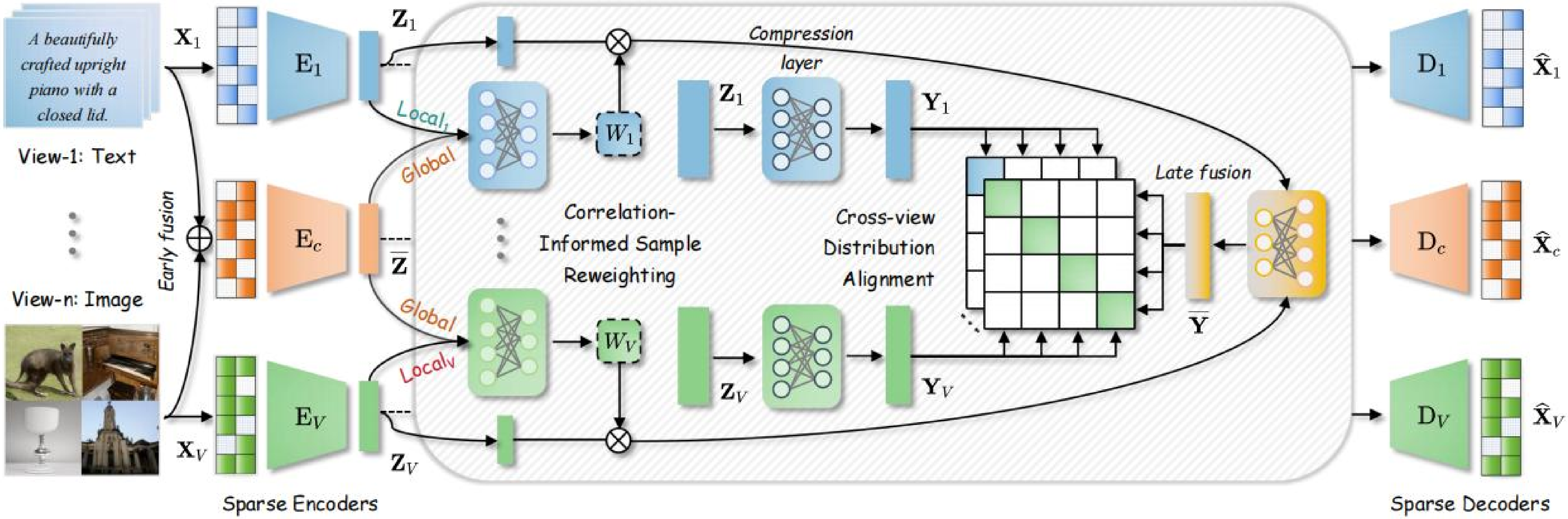
Datasets	Sparsity of Different Views	Accuracy [Ours]	Accuracy [2nd]	NMI [Ours]	NMI [2nd]
ALOI-100	[0.0001, 0.0001, 0.3415, 0.6383]	82.21 ↑4.49	77.72	92.65 ↑1.78	90.87
MSRCV1	[0.0049, 0.0048, 0.0048, 0.3478, 0.0051, 0.0048]	97.14 ↑5.71	91.43	94.22 ↑6.32	87.90
LGG	[0.0040, 0.0038, 0.0078, 0.0037]	83.15 ↑0.38	82.77	54.13 ↑0.49	54.13
Synthetic3d	[0.0017, 0.0017, 0.0017]	98.33 ↑0.16	98.17	98.33 ↑0.16	98.17

- Above table demonstrates that SparseMVC maintains robust performance across datasets with **significant** sparsity differences (e.g., a 6000-fold disparity between the maximum and minimum view sparsity ratios in **ALOI-100**) as well as those with **minimal** sparsity variations (e.g., no view sparsity differences in **Synthetic3D**).

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

solve problem： **cross-view sparsity variations**

Overview of SparseMVC



SparseMVC is a framework designed to address varying sparsity across views. It begins by utilizing sparse autoencoders with adaptive constraints, which dynamically adjust the coding strategy based on the probed $\boldsymbol{s}_v$, to generate latent features ($\boldsymbol{Z}$), making the reconstructed features ($\widehat{\boldsymbol{X}}$) approximate the original input features ($\boldsymbol{X}$). Subsequently, the correlation between the early-fused global features ($\bar{\boldsymbol{Z}}$) and view-specific features ($\{\boldsymbol{Z}_v\}_{v=1}^n$) guides the computation of sample-level weights ($\{\boldsymbol{W}_v\}_{v=1}^n$) via the attention mechanism within the correlation-informed sample reweighting module. Finally, the cross-view distribution alignment module enhances clustering performance by setting the late-fused global features $\overline{\boldsymbol{Y}}$ as the anchor latent representation, and then simultaneously aligning the multi-view feature distribution between $\overline{\boldsymbol{Y}}$ and each view-specific compressed feature ($\{\boldsymbol{Y}_v\}_{v=1}^n$).

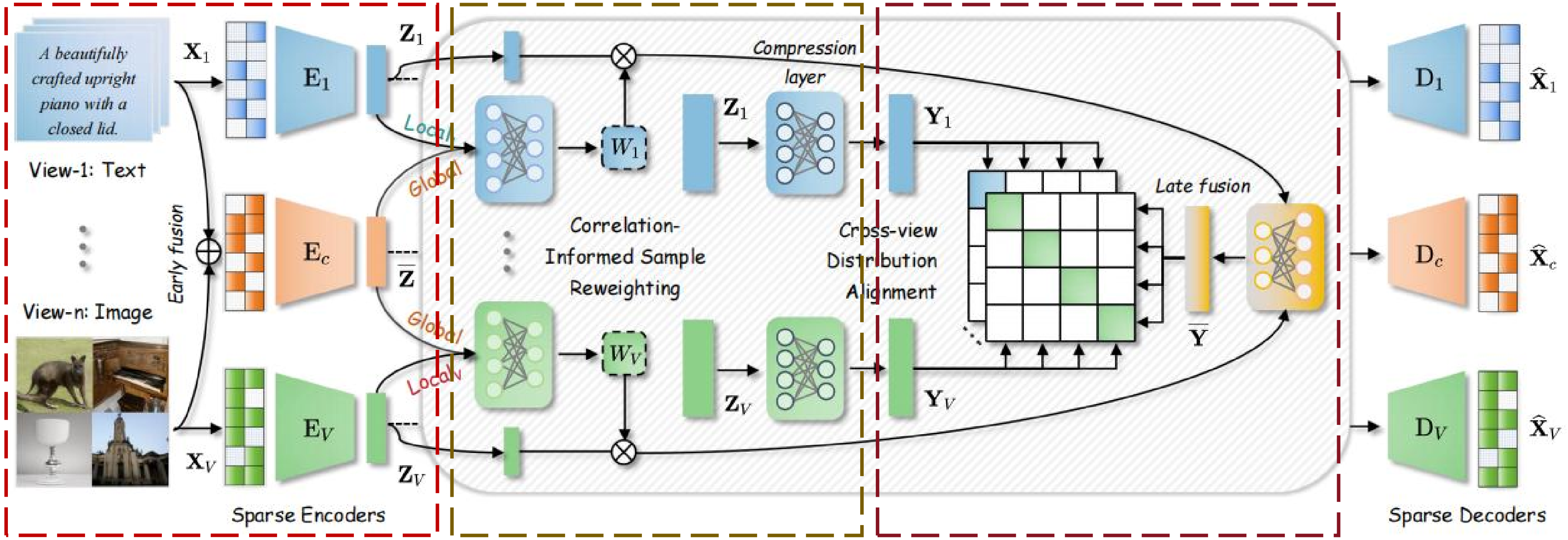
SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

Part 2. Correlation weights based on the early-fused global and view-specific features

How to enable autoencoders to adaptively encode features?

How to mitigate the side effects of encoding discrepancies?

How to align feature distributions across multiple views?

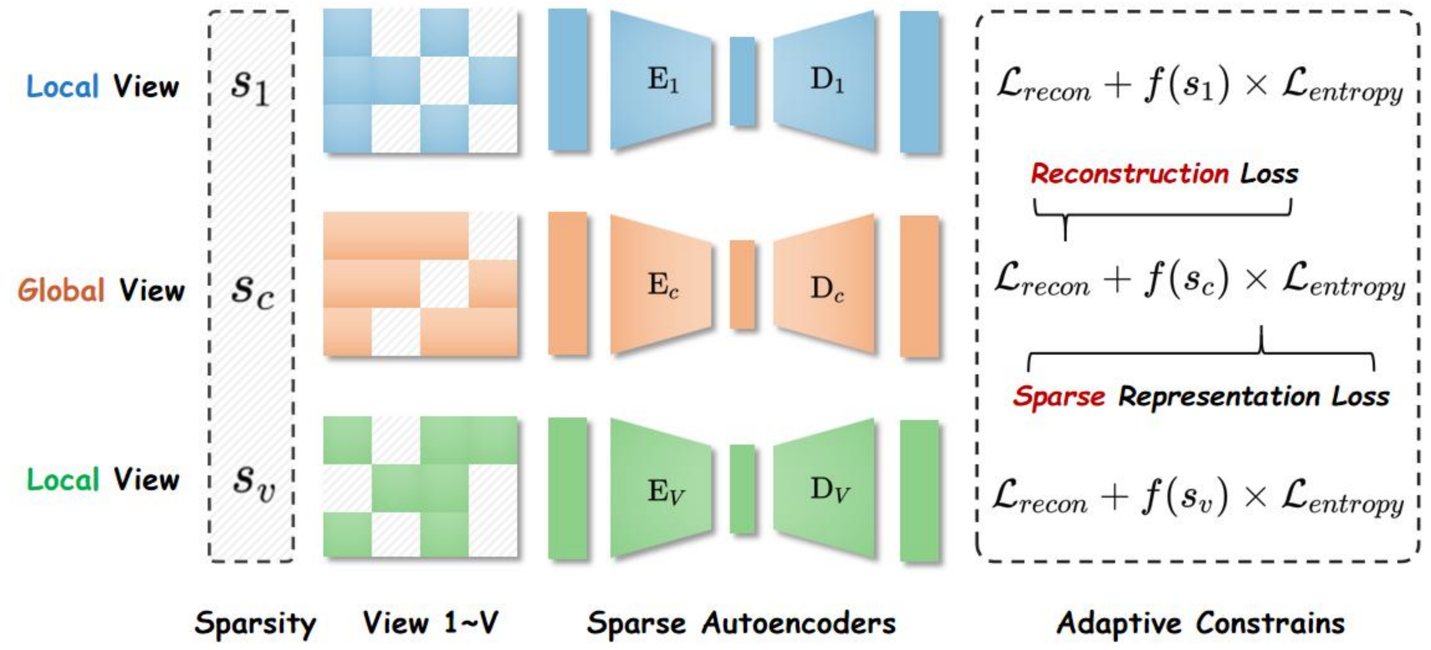


Part 1. Sparse autoencoders with constraints that vary based on view-level sparsity

Part 3. Sample pairs distribution alignment based on contrastive learning

- **SparseMVC** is specifically designed to handle **view-level sparsity variations**, a prevalent yet underexplored characteristic of multi-view data, through a complete data-driven and tightly integrated architecture.

Part 1.



$$\boxed{2}\quad \mathcal{L}_{\text{sparse}}^{v} = f(s_v) \cdot \mathcal{L}_{\text{entropy}}^{v} = f(s_v) \cdot \sum_{k=1}^{H} \left(\rho \log \frac{\rho}{\hat{h}_k^v} + (1-\rho) \log \frac{1-\rho}{1-\hat{h}_k^v}\right)$$

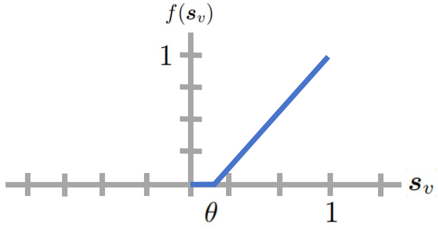
$$\boxed{1}\quad \mathcal{L}_{\text{recon}}^{v} = \frac{1}{N} \sum_{j=1}^{N} \left(\hat{\boldsymbol{x}}_j^v - \boldsymbol{x}_j^v\right)^2$$

$$\mathcal{L}_{\text{recon}}^{v} \oplus \mathcal{L}_{\text{sparse}}^{v}$$

$$\mathcal{L}_{\text{recon}}^{v} \oplus f(s_v) \cdot \mathcal{L}_{\text{entropy}}^{v}$$

We use the sparsity ratio of each view as prior knowledge:

$$f(s_v) = \begin{cases} 0, & \text{if } s_v \leq \theta, \\ \frac{s_v - \theta}{1-\theta}, & \text{if } s_v > \theta, \end{cases}$$



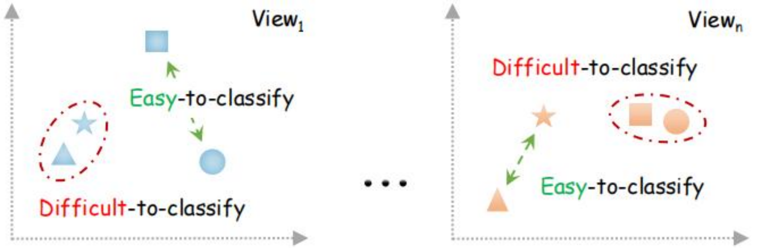
- **To address the sparsity variations across views within a dataset, we propose an adaptive encoding strategy that allows the encoders to automatically transition between standard and various degrees of sparse encoding**

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

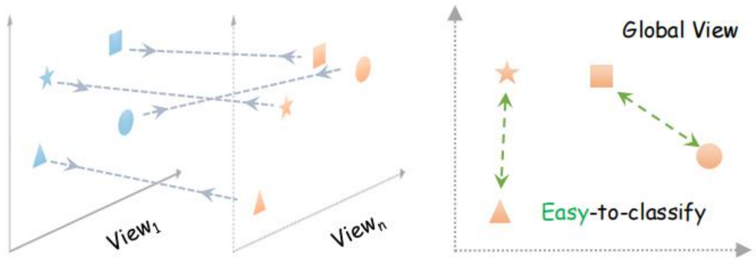
Part 2.

- To mitigate the enlarged representation gap caused by heterogeneous encoding, we incorporate a correlation-guided fusion strategy that leverages global-to-local feature relationships established in the early encoding stage to guide the weighting of local features during late fusion.

Part 3.

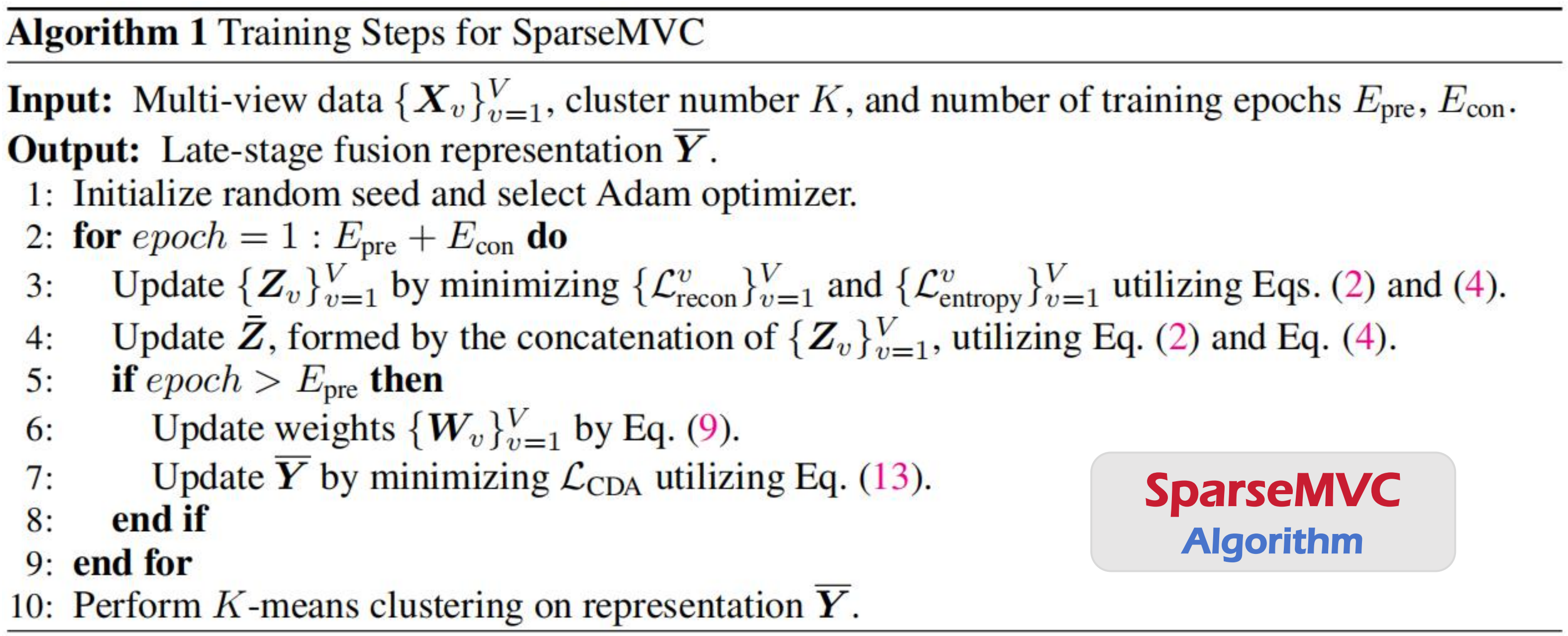


(a) Multi-view sample distribution in latent space



(b) Cross-view distribution alignment (c) Multi-view fusion

- **View distribution alignment based on contrast (with arbitrary view numbers).**



Algorithm 1 Training Steps for SparseMVC

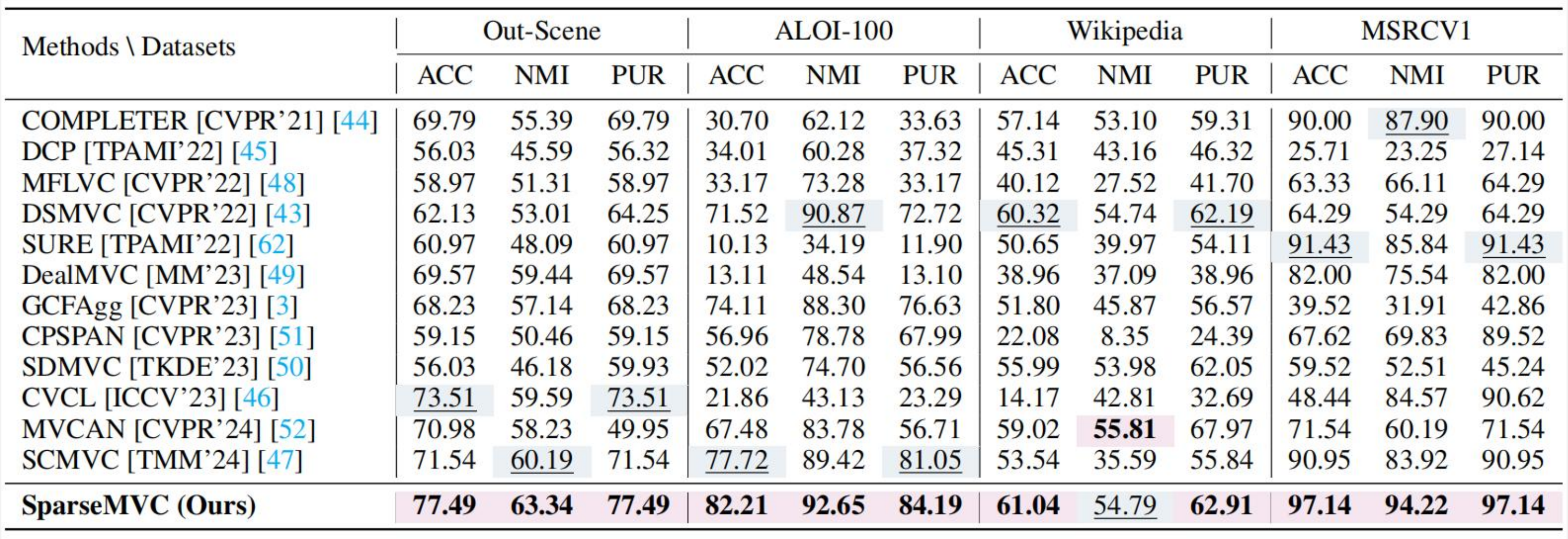
Input: Multi-view data $\{\boldsymbol{X}_v\}_{v=1}^V$, cluster number K, and number of training epochs $E_{\text{pre}}, E_{\text{con}}$.
Output: Late-stage fusion representation $\overline{\boldsymbol{Y}}$.

```
Initialize random seed and select Adam optimizer.
for $epoch = 1 : E_{\text{pre}} + E_{\text{con}}$ do
    Update $\{\boldsymbol{Z}_v\}_{v=1}^V$ by minimizing $\{\mathcal{L}_{\text{recon}}^v\}_{v=1}^V$ and $\{\mathcal{L}_{\text{entropy}}^v\}_{v=1}^V$ utilizing Eqs. (2) and (4).
    Update $\bar{\boldsymbol{Z}}$, formed by the concatenation of $\{\boldsymbol{Z}_v\}_{v=1}^V$, utilizing Eq. (2) and Eq. (4).
    if $epoch > E_{\text{pre}}$ then
        Update weights $\{\boldsymbol{W}_v\}_{v=1}^V$ by Eq. (9).
        Update $\overline{\boldsymbol{Y}}$ by minimizing $\mathcal{L}_{\text{CDA}}$ utilizing Eq. (13).
    end if
end for
Perform $K$-means clustering on representation $\overline{\boldsymbol{Y}}$.
```

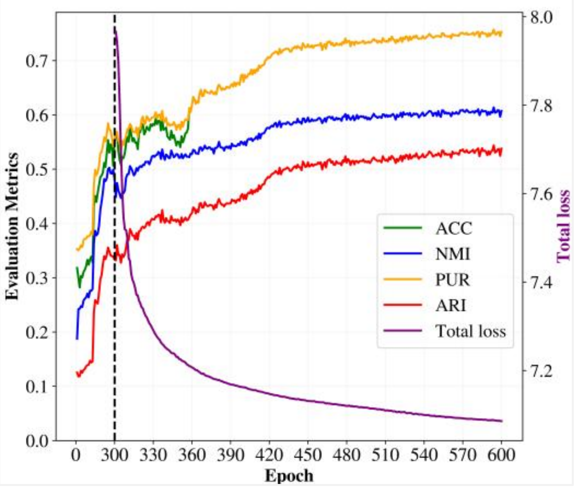
SparseMVC
Algorithm

$$\mathcal{L}_{\text{total}} = \sum_{v=1}^{V} \left(\mathcal{L}_{\text{recon}}^v + f(s_v) \cdot \mathcal{L}_{\text{entropy}}^v\right) + \lambda_{\text{CR}} \cdot \mathcal{L}_{\text{CDA}}$$

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering



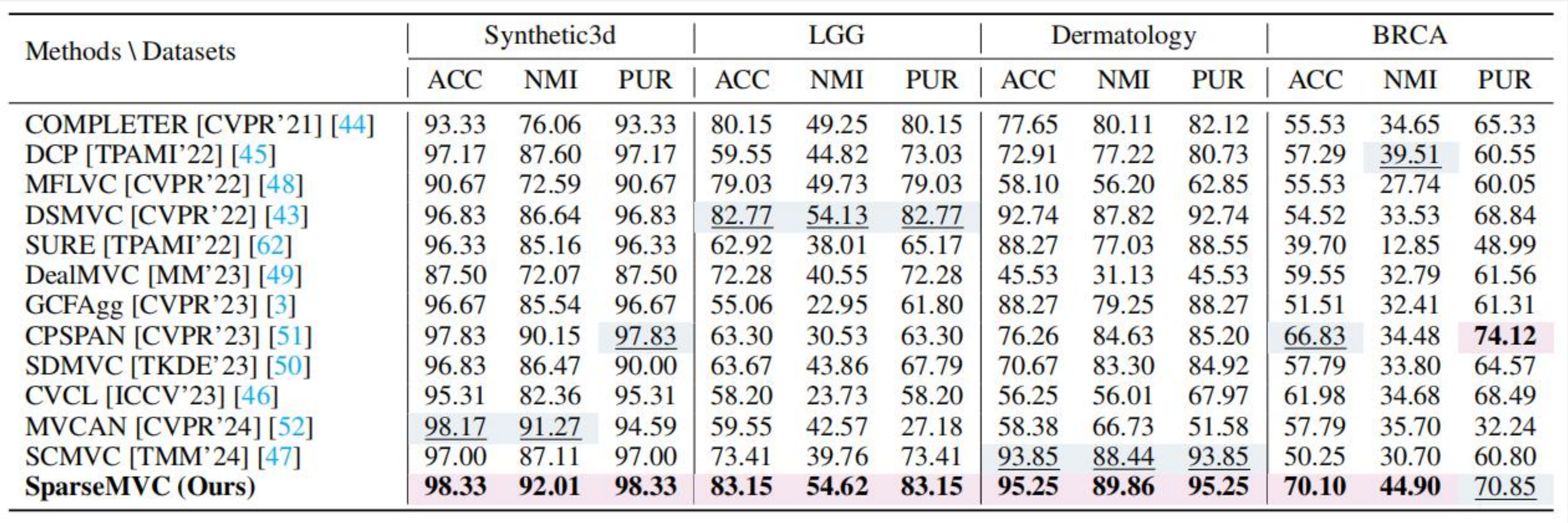
Methods \ Datasets	Out-Scene			ALOI-100			Wikipedia			MSRCV1		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER [CVPR'21] [44]	69.79	55.39	69.79	30.70	62.12	33.63	57.14	53.10	59.31	90.00	87.90	90.00
DCP [TPAMI'22] [45]	56.03	45.59	56.32	34.01	60.28	37.32	45.31	43.16	46.32	25.71	23.25	27.14
MFLVC [CVPR'22] [48]	58.97	51.31	58.97	33.17	73.28	33.17	40.12	27.52	41.70	63.33	66.11	64.29
DSMVC [CVPR'22] [43]	62.13	53.01	64.25	71.52	90.87	72.72	60.32	54.74	62.19	64.29	54.29	64.29
SURE [TPAMI'22] [62]	60.97	48.09	60.97	10.13	34.19	11.90	50.65	39.97	54.11	91.43	85.84	91.43
DealMVC [MM'23] [49]	69.57	59.44	69.57	13.11	48.54	13.10	38.96	37.09	38.96	82.00	75.54	82.00
GCFAgg [CVPR'23] [3]	68.23	57.14	68.23	74.11	88.30	76.63	51.80	45.87	56.57	39.52	31.91	42.86
CPSPAN [CVPR'23] [51]	59.15	50.46	59.15	56.96	78.78	67.99	22.08	8.35	24.39	67.62	69.83	89.52
SDMVC [TKDE'23] [50]	56.03	46.18	59.93	52.02	74.70	56.56	55.99	53.98	62.05	59.52	52.51	45.24
CVCL [ICCV'23] [46]	73.51	59.59	73.51	21.86	43.13	23.29	14.17	42.81	32.69	48.44	84.57	90.62
MVCAN [CVPR'24] [52]	70.98	58.23	49.95	67.48	83.78	56.71	59.02	**55.81**	67.97	71.54	60.19	71.54
SCMVC [TMM'24] [47]	71.54	60.19	71.54	77.72	89.42	81.05	53.54	35.59	55.84	90.95	83.92	90.95
SparseMVC (Ours)	**77.49**	**63.34**	**77.49**	**82.21**	**92.65**	**84.19**	**61.04**	54.79	**62.91**	**97.14**	**94.22**	**97.14**



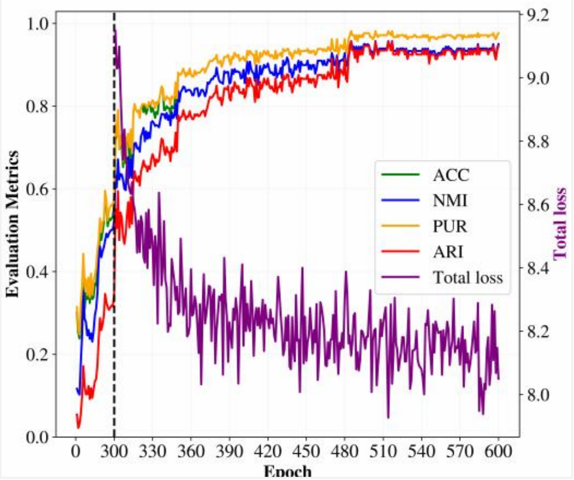
Out-Scene
big dataset

Comparative Results

Convergence Analysis



Methods \ Datasets	Synthetic3d			LGG			Dermatology			BRCA		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
COMPLETER [CVPR'21] [44]	93.33	76.06	93.33	80.15	49.25	80.15	77.65	80.11	82.12	55.53	34.65	65.33
DCP [TPAMI'22] [45]	97.17	87.60	97.17	59.55	44.82	73.03	72.91	77.22	80.73	57.29	39.51	60.55
MFLVC [CVPR'22] [48]	90.67	72.59	90.67	79.03	49.73	79.03	58.10	56.20	62.85	55.53	27.74	60.05
DSMVC [CVPR'22] [43]	96.83	86.64	96.83	82.77	54.13	82.77	92.74	87.82	92.74	54.52	33.53	68.84
SURE [TPAMI'22] [62]	96.33	85.16	96.33	62.92	38.01	65.17	88.27	77.03	88.55	39.70	12.85	48.99
DealMVC [MM'23] [49]	87.50	72.07	87.50	72.28	40.55	72.28	45.53	31.13	45.53	59.55	32.79	61.56
GCFAgg [CVPR'23] [3]	96.67	85.54	96.67	55.06	22.95	61.80	88.27	79.25	88.27	51.51	32.41	61.31
CPSPAN [CVPR'23] [51]	97.83	90.15	97.83	63.30	30.53	63.30	76.26	84.63	85.20	66.83	34.48	**74.12**
SDMVC [TKDE'23] [50]	96.83	86.47	90.00	63.67	43.86	67.79	70.67	83.30	84.92	57.79	33.80	64.57
CVCL [ICCV'23] [46]	95.31	82.36	95.31	58.20	23.73	58.20	56.25	56.01	67.97	61.98	34.68	68.49
MVCAN [CVPR'24] [52]	98.17	91.27	94.59	59.55	42.57	27.18	58.38	66.73	51.58	57.79	35.70	32.24
SCMVC [TMM'24] [47]	97.00	87.11	97.00	73.41	39.76	73.41	93.85	88.44	93.85	50.25	30.70	60.80
SparseMVC (Ours)	**98.33**	**92.01**	**98.33**	**83.15**	**54.62**	**83.15**	**95.25**	**89.86**	**95.25**	**70.10**	**44.90**	70.85



MSRCV1
small dataset

SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

NEURAL INFORMATION PROCESSING SYSTEMS

Ablation Study

1.Loss Function

Datasets	$\mathcal{L}_{\text{recon}}$	$\mathcal{L}_{\text{entropy}}$	$\mathcal{L}_{\text{CDA}}$	ACC	NMI	PUR
ALOI-100	✓			45.27	71.21	30.41
	✓	✓		66.35	81.65	70.62
	✓		✓	64.11	80.23	67.33
	✓	✓	✓	**82.21**	**92.65**	**84.19**
Dermatology	✓			70.11	74.41	59.69
	✓	✓		75.70	83.36	69.36
	✓		✓	70.95	71.06	83.80
	✓	✓	✓	**95.25**	**89.86**	**95.25**
MSRCV1	✓			58.57	48.63	32.76
	✓	✓		70.48	65.27	52.08
	✓		✓	92.38	87.62	92.38
	✓	✓	✓	**97.14**	**94.22**	**97.14**

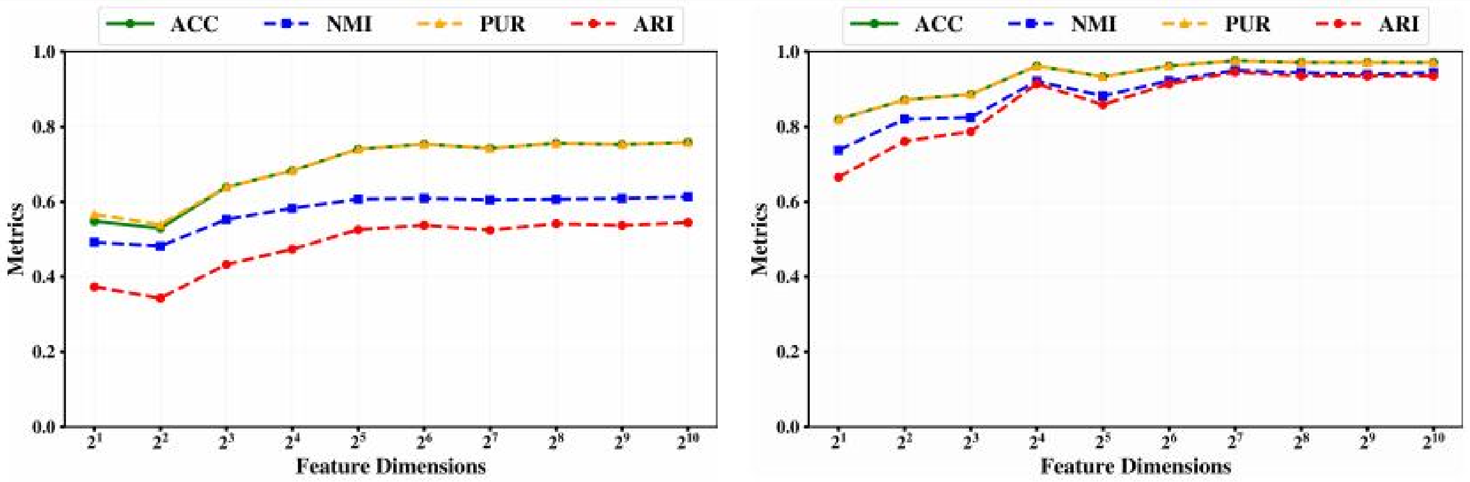
Column groups: Loss Function ($\mathcal{L}_{\text{recon}}$, $\mathcal{L}_{\text{entropy}}$, $\mathcal{L}_{\text{CDA}}$); Evaluation Metrics (ACC, NMI, PUR).

2.Network Components

Components \ Datasets	ALOI-100			Dermatology			MSRCV1		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
all sparse autoencoders w/o CSR	78.56	88.92	81.12	77.37	74.38	86.03	91.90	88.56	91.90
all sparse autoencoders	80.42	89.19	82.44	89.11	78.37	89.11	92.38	88.61	92.38
adaptive autoencoders w/o CSR	81.04	89.58	83.17	88.83	78.23	88.83	95.71	92.69	95.71
adaptive autoencoders (Ours)	**82.21**	**92.65**	**84.19**	**95.25**	**89.86**	**95.25**	**97.14**	**94.22**	**97.14**

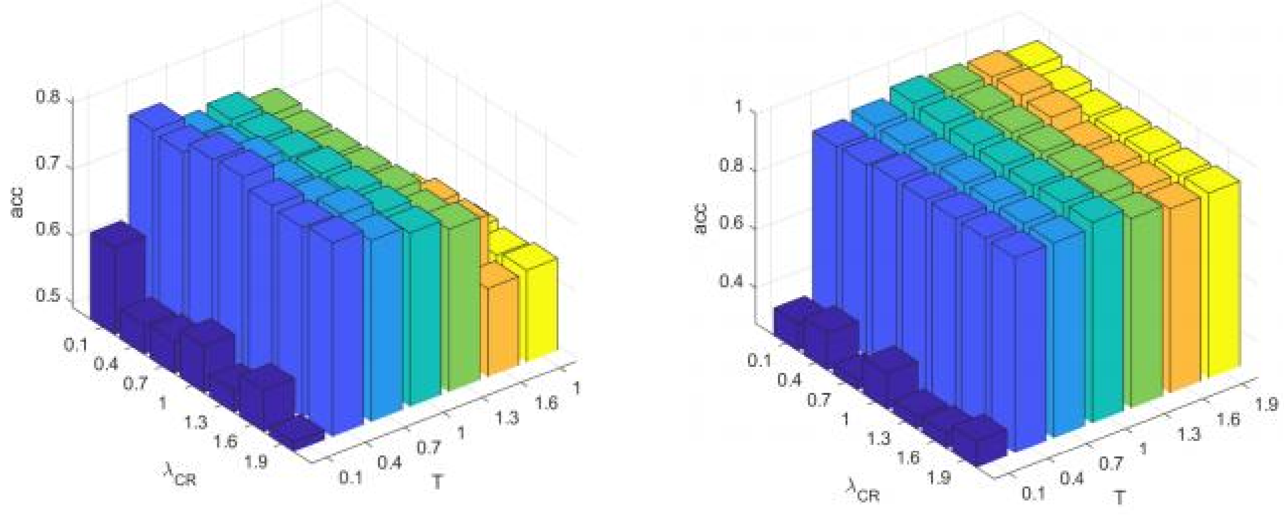
SparseMVC: Probing Cross-view Sparsity Variations for Multi-view Clustering

1.Feature (Zv) dimension /Latent space size
Sensitivity Analysis



(a) Out-Scene (b) MSRCV1

2.Loss Function Parameter
Sensitivity Analysis



(a) Out-Scene (b) MSRCV1

Limitations：

- SparseMVC does not incorporate a targeted design or specialized mechanisms to address view misalignment and random view missingness;
- the use of contrastive learning inherently introduces computational overhead, making it unlikely to rank among the fastest available approaches.

Conclusion：

- This paper highlights a frequently overlooked issue in deep multi-view learning: varying sparsity ratios across views;
- We systematically define, quantify, and analyze cross-view sparsity variation as a fundamental characteristic of multi-view data;
- Our framework advances the field by extending sparsity handling from the data-level to view-level and mitigating the adverse effects of encoding discrepancies through sample-level dynamic weighting;
- We hope this work inspires greater attention to the intrinsic characteristics of data and to the design of architectures driven by data.