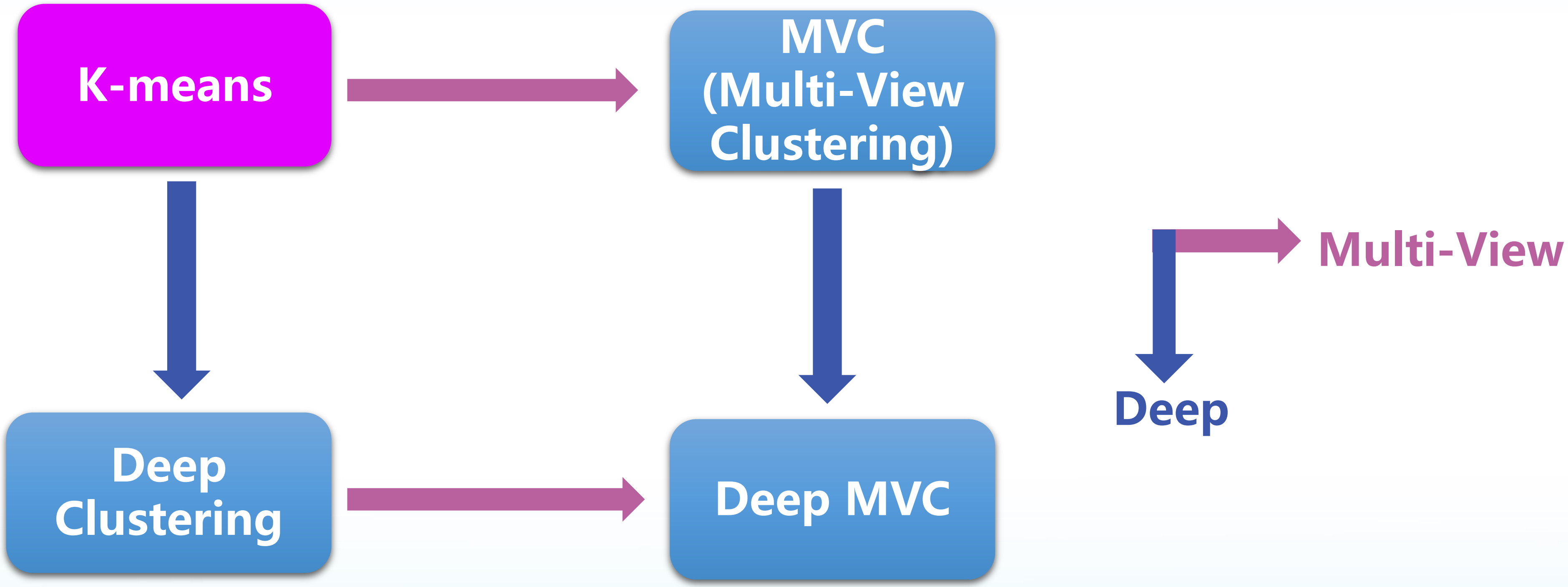
Self-Supervised Feature Learning for Deep Multi-View Clustering

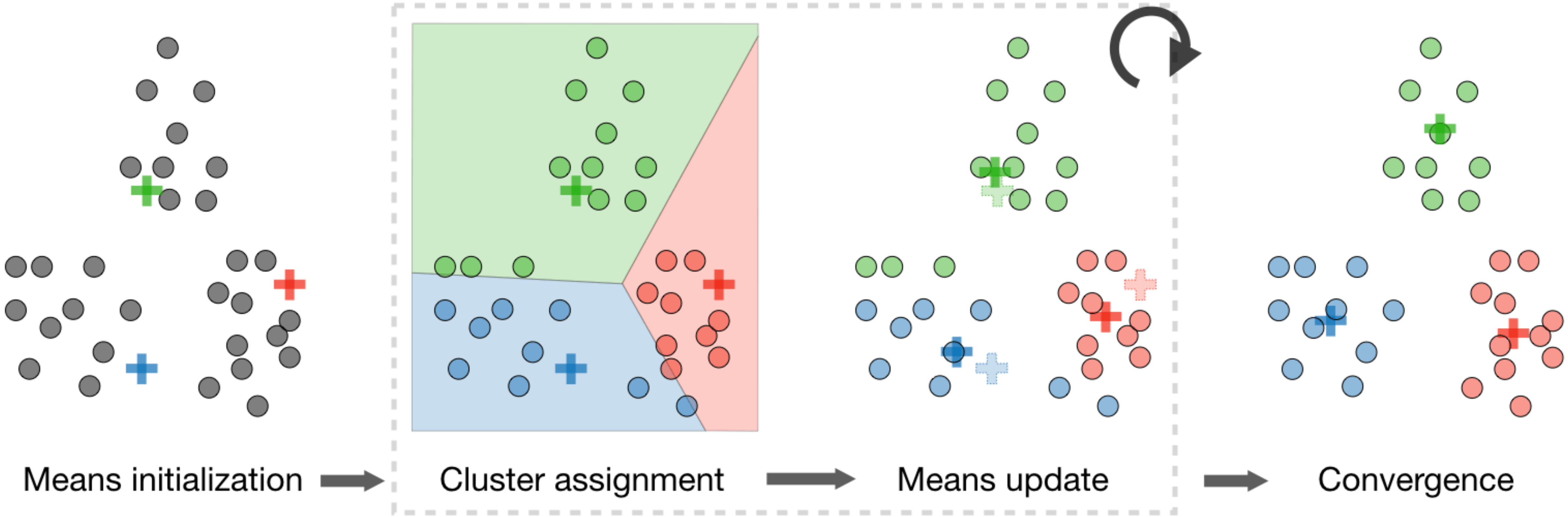
Yazhou Ren
yazhou.ren@uestc.edu.cn
School of Computer Science and Engineering, UESTC
https://yazhou-ren.github.io/

Traditional clustering method



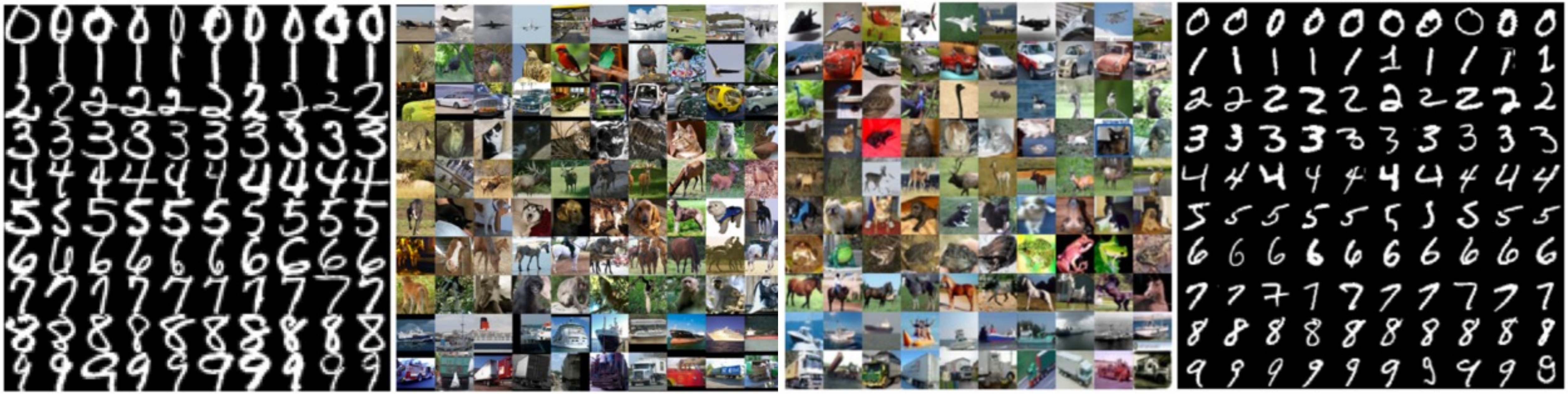
1967：K-means

$$\arg\min_{\mathbf{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}\in S_i} \left\|\mathbf{x}-\boldsymbol{\mu}_i\right\|^2$$

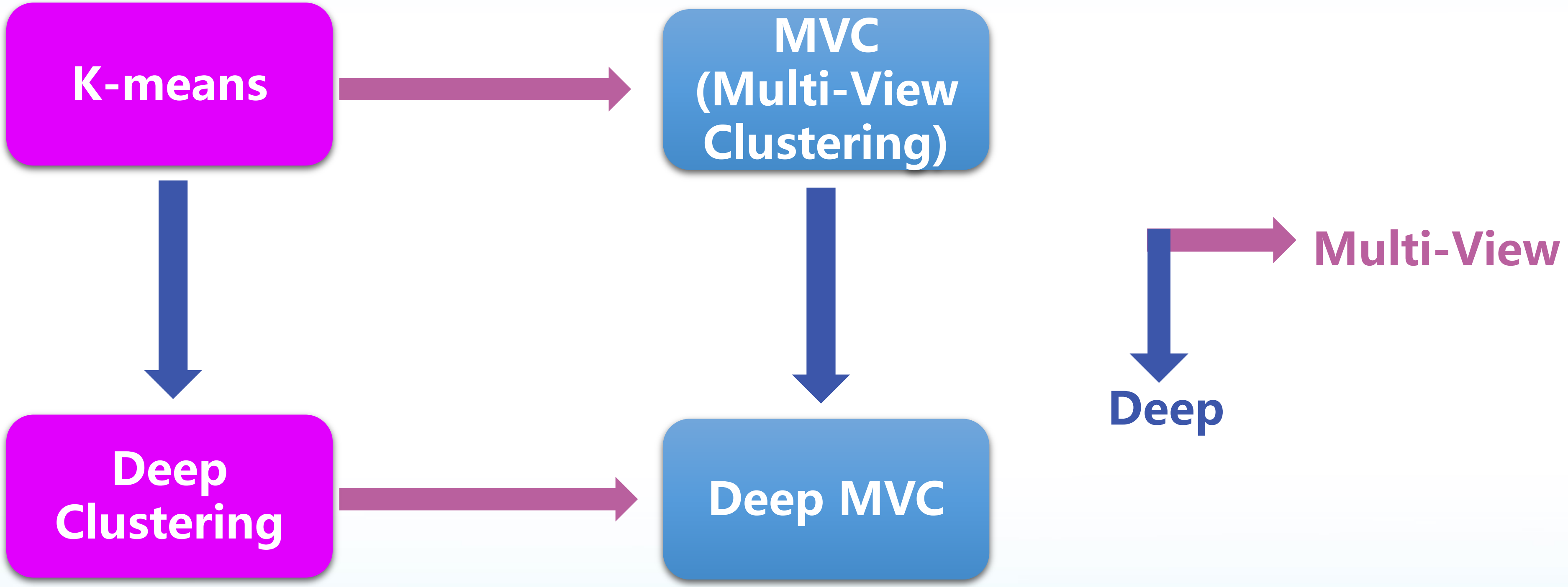


J. MacQueen. Some methods for classification and analysis of multivariate observations. *In Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

Limitations of shallow models

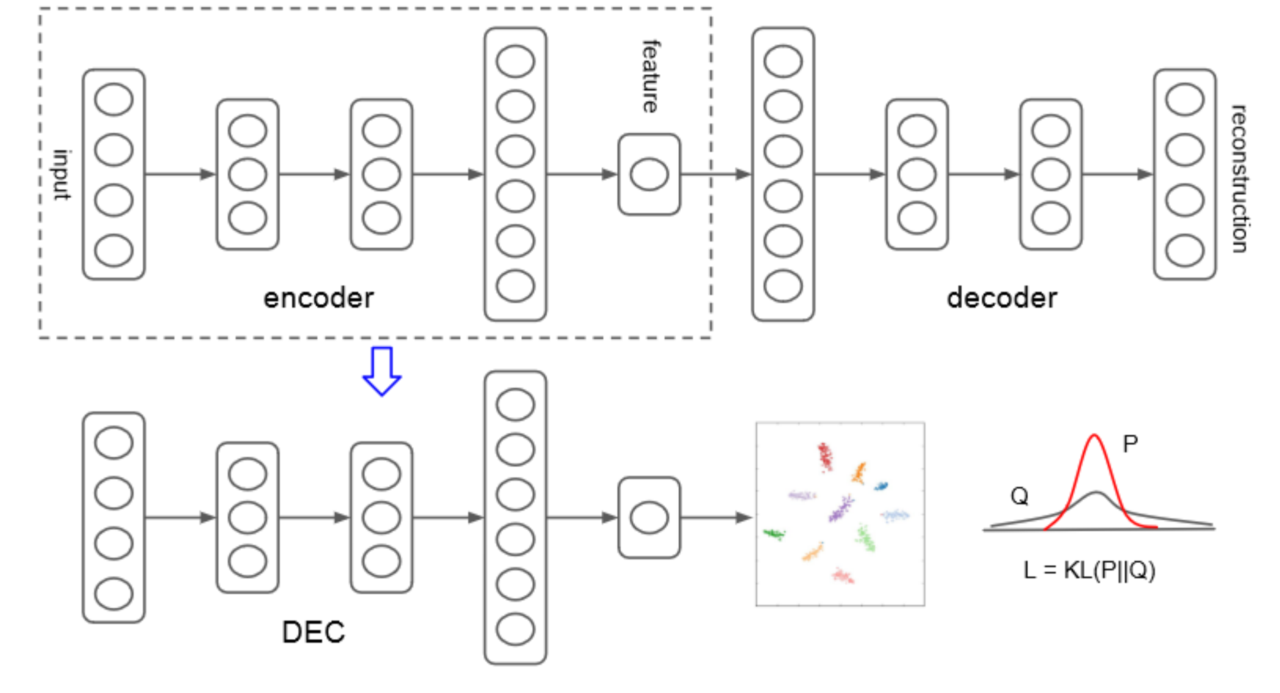


(a) USPS (b) STL-10 (c) CIFAR-10 (d) MNIST



For high-dimensional datasets

Deep Clustering→DEC (deep embedding clustering)



Soft assignment

$$q_{ij} = \frac{(1 + \|z_i - \mu_j\|^2/\alpha)^{-\frac{\alpha+1}{2}}}{\sum_{j'}(1 + \|z_i - \mu_{j'}\|^2/\alpha)^{-\frac{\alpha+1}{2}}}$$

Target distribution

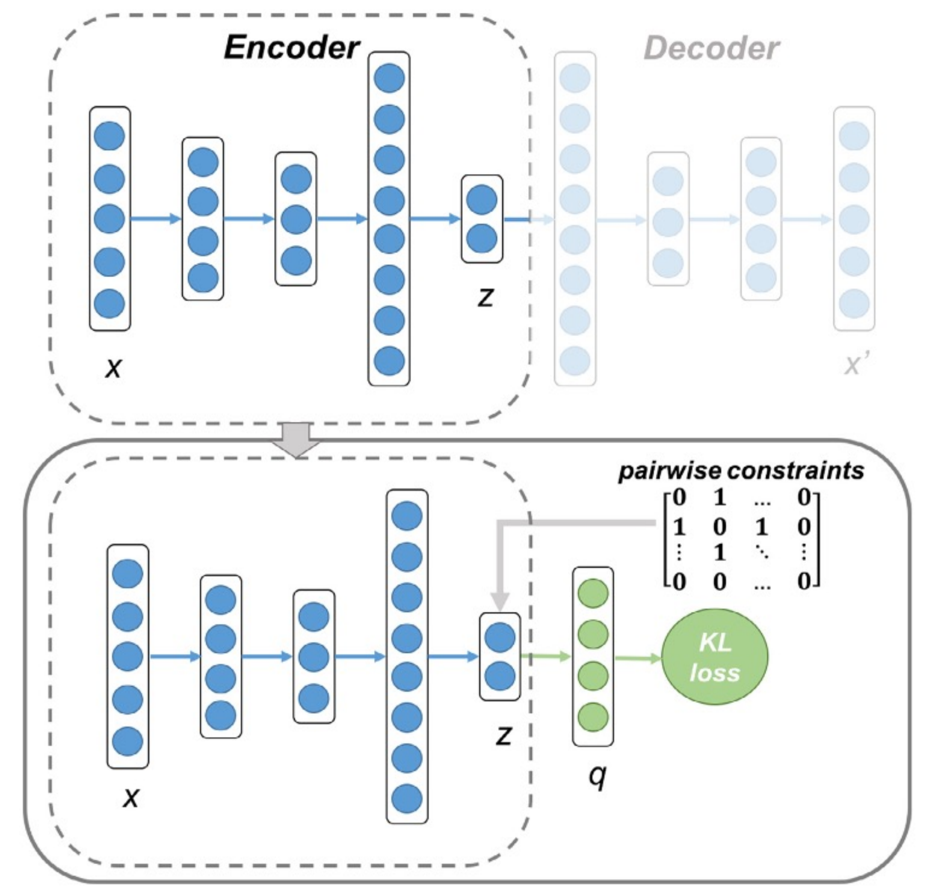
$$p_{ij} = \frac{q_{ij}^2/\sum_i q_{ij}}{\sum_j \left(q_{ij}^2/\sum_i q_{ij}\right)}$$

KL divergence minimization

$$L = \mathrm{KL}(P\|Q) = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}}$$

Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. In *ICML*, pages 478–487, 2016.

Deep Clustering →SDEC (Semisupervised DEC)

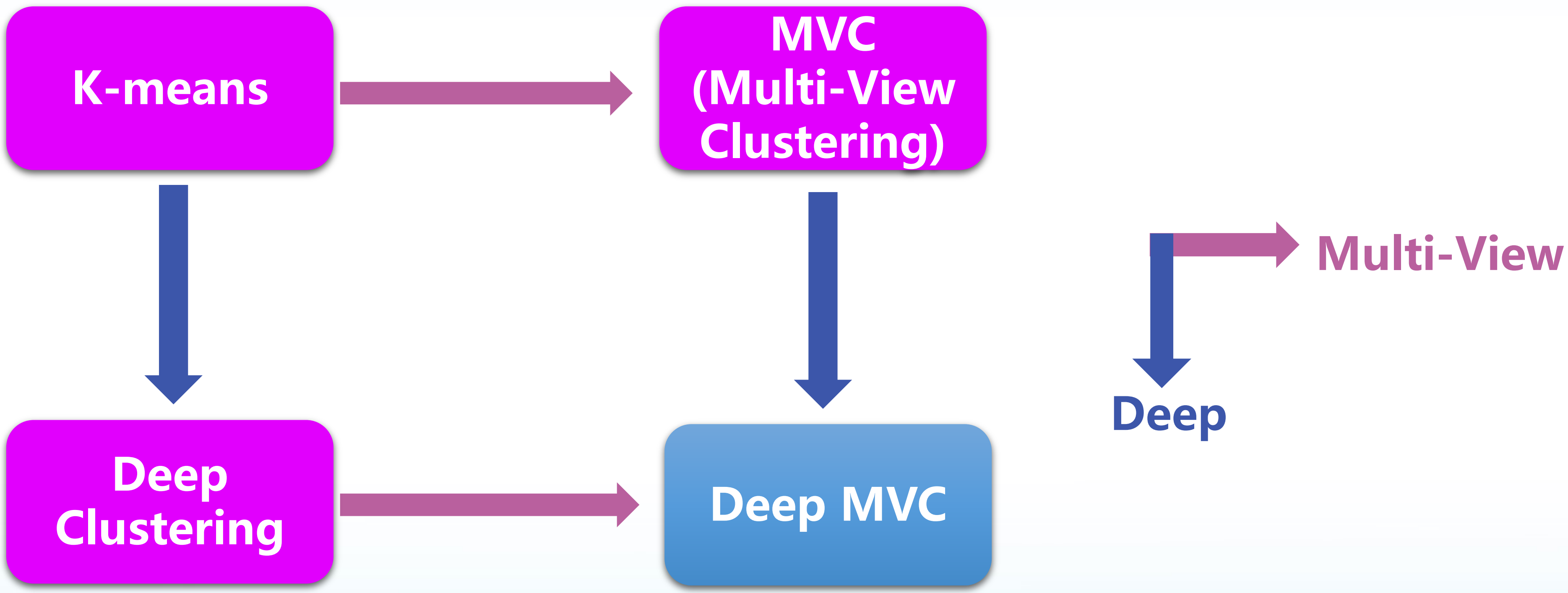


Clustering results measured by ACC(%).

Data	k-means	KM-cst	AE+KM	AE+KM-cst	DEC	IDEC	SDEC
USPS	65.67	68.18	70.28	71.87	75.81	75.86	**76.39**
STL-10	28.31	29.09	34.00	35.15	37.40	36.99	**38.86**
CIFAR-10	23.75	23.91	23.89	24.36	26.26	25.02	**27.26**
MNIST	52.98	54.27	74.09	75.98	84.94	83.85	**86.11**
20NG	33.77	33.89	40.81	47.71	50.11	53.63	**78.12**

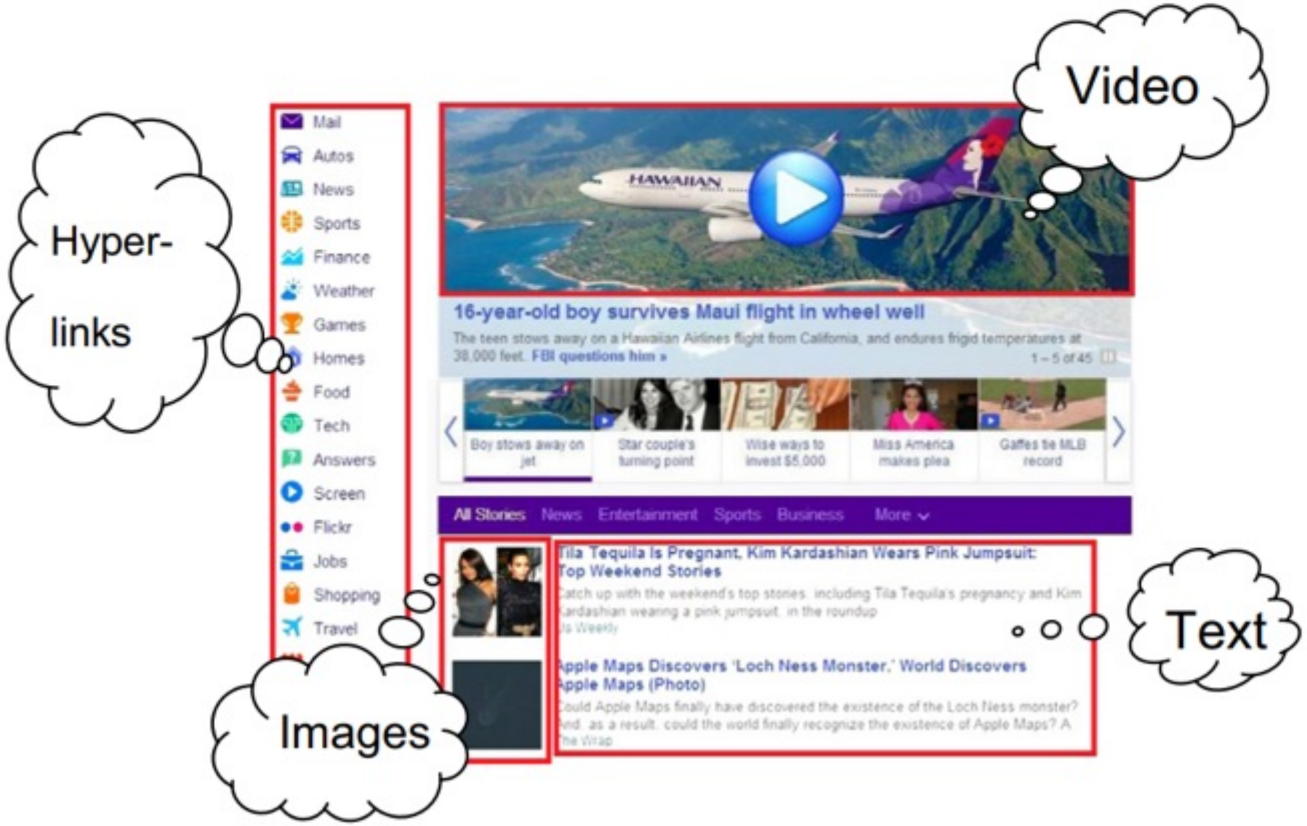
Yazhou Ren et al. Semi-supervised deep embedded clustering. *Neurocomputing*, 325:121–130, 2019.

A data example often has different observable views.



Multi-View Clustering

While in real-world, an object can be always describe by multiple views.
Conventional clustering methods only work on single-view data.



$$\min_{C^v,B} \sum_{v=1}^{m} ||X^v - C^v B||_F^2$$

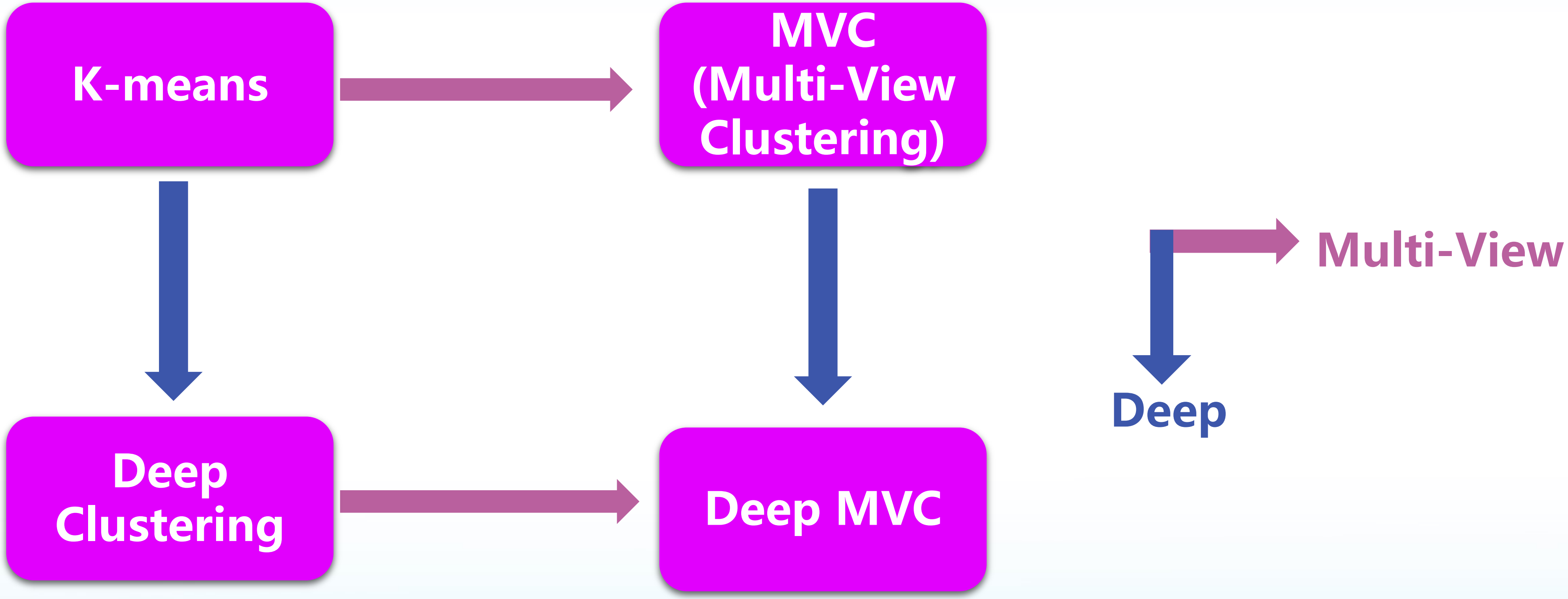
2011
Co-regularized multi-view spectral clustering
Kumar et al. in *NIPS*

2013
Multi-view k-means clustering on big data
Cai et al. in *IJCAI*

2015
Multi-view subspace clustering
Gao et al. in *ICCV*

2017
Self-weighted multi-view clustering with multiple graphs
Nie et al. in *IJCAI*

2021
Learning Smooth Representation for Multi-view Subspace Clustering,
Ren et al. in *ACM MM*



a. enough representation capability and are applicable for image clustering.
b. handle large-scale data clustering tasks.

Deep Multi-view Clustering

Multi-view clustering + Deep learning techniques

Representation learning

Self-supervised learning

Adversarial learning

Contrastive learning

……

Our Recent Work

Deep Embedded Multi-view Clustering with Collaborative Training (DEMVC, Information Sciences, 2021)

Deep embedded multi-view clustering with collaborative training

Jie Xu [a], Yazhou Ren [a,*], Guofeng Li [a], Lili Pan [b,c], Ce Zhu [b], Zenglin Xu [d,e]

[a] School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[b] School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China
[c] Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, Quzhou 324000, China
[d] Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China
[e] Center for Artificial Intelligence, Peng Cheng Lab, Shenzhen 518055, China

A R T I C L E I N F O

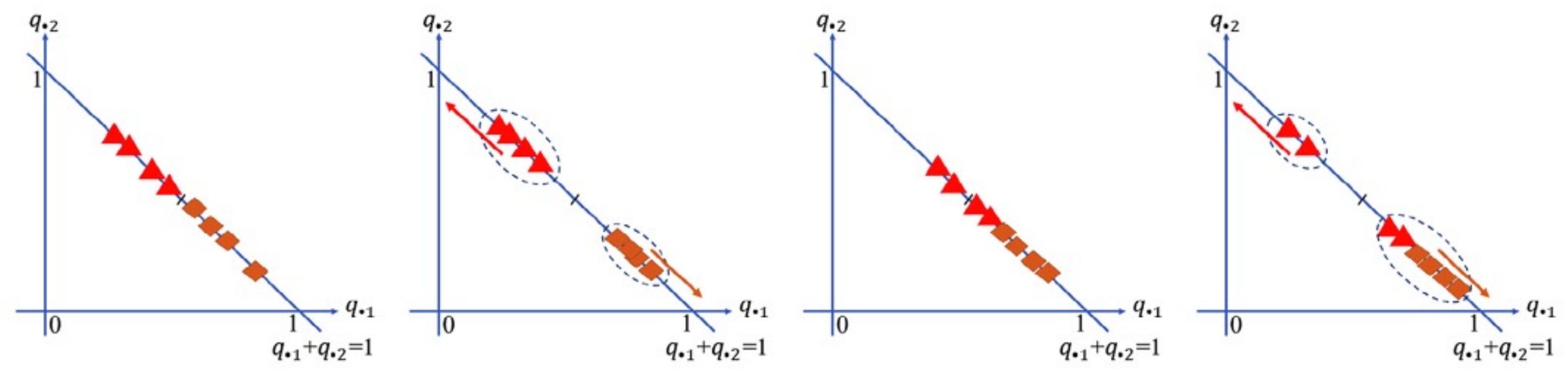
Article history:
Received 16 June 2020
Received in revised form 24 November 2020
Accepted 27 December 2020
Available online 26 January 2021

Keywords:
Deep embedded clustering
Multi-view clustering
Unsupervised learning
Collaborative training

A B S T R A C T

Multi-view clustering has attracted increasing attentions recently by utilizing information from multiple views. However, existing multi-view clustering methods are either with high computation and space complexities, or lack of representation capability. To address these issues, we propose deep embedded multi-view clustering with collaborative training (DEMVC) in this paper. Firstly, the embedded representations of multiple views are learned individually by deep autoencoders. Then, both consensus and complementary of multiple views are taken into account and a novel collaborative training scheme is proposed. Concretely, the feature representations and cluster assignments of all views are learned collaboratively. A new consistency strategy for cluster centers initialization is further developed to improve the multi-view clustering performance with collaborative training. Experimental results on several popular multi-view datasets show that DEMVC achieves significant improvements over state-of-the-art methods.

The mechanism of deep embedded clustering to minimize KL divergence



(a) Before optimization (b) After optimization (c) Before optimization (d) After optimization

optimization using KL divergence

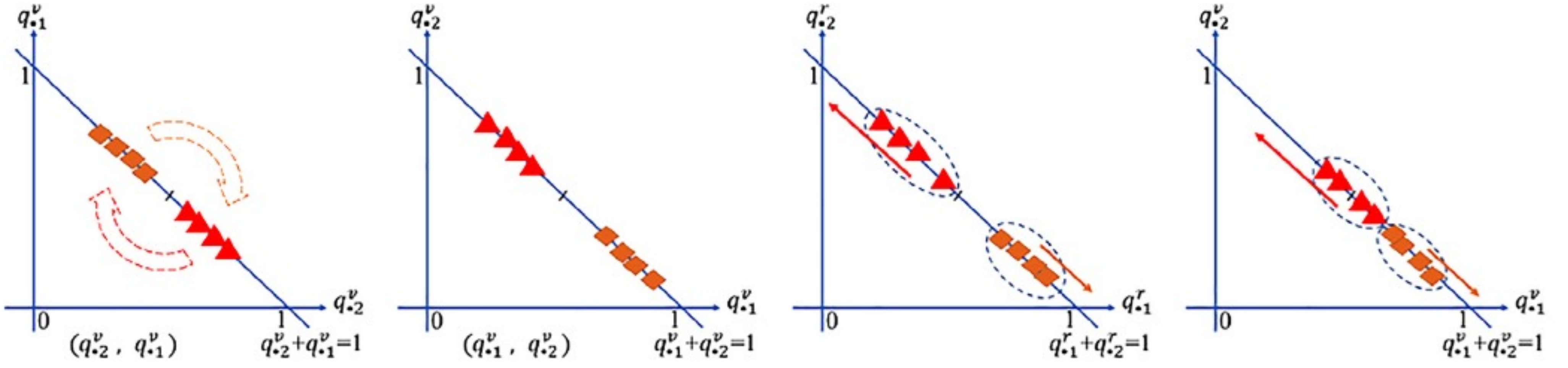
$$L_c^v = KL(P^v \| Q^v) = \sum_{i=1}^{N}\sum_{j=1}^{K} p_{ij}^v \log \frac{p_{ij}^v}{q_{ij}^v}$$

hard samples are prone to misclassification

$$L_r^v = \sum_{i=1}^{N} \left\| \boldsymbol{x}_i^v - g_{\Omega}^v\left(f_{\Theta}^v\left(\boldsymbol{x}_i^v\right)\right) \right\|_2^2$$

$$L = \sum_{v=1}^{V} L_r^v + \gamma \sum_{v=1}^{V} L_c^v$$

The mechanism by which DEMVC works



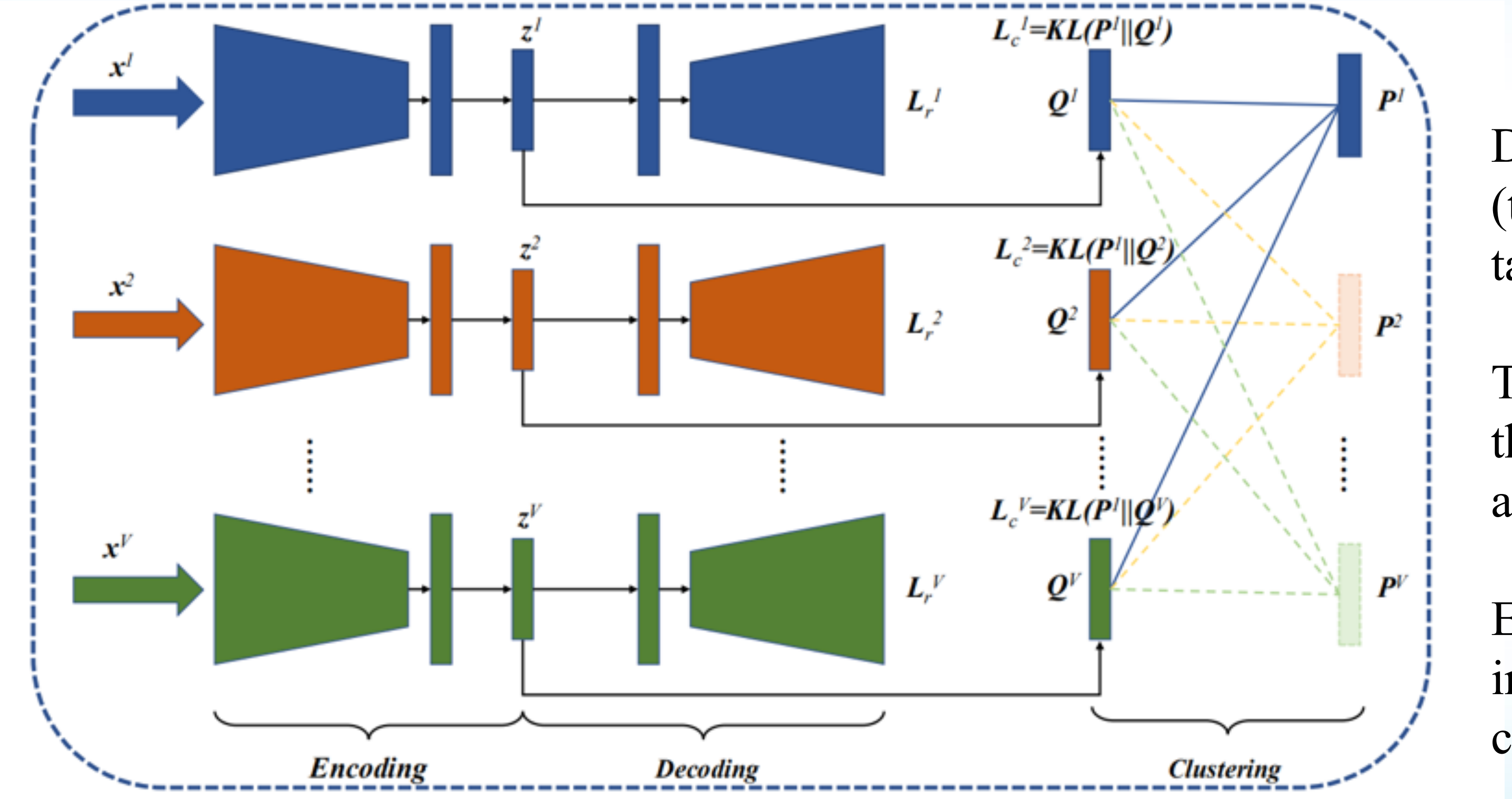
(a) Before coordinate alignment (b) After coordinate alignment (c) The referred view with correct samples (d) Another view with mispredicted samples

collaborative training

correct the mispredicted samples in other views

$$L = \sum_{v=1}^{V}\sum_{i=1}^{N}\|\boldsymbol{x}_i^v - g_\Omega^v(f_\Theta^v(\boldsymbol{x}_i^v))\|_2^2 + \gamma\sum_{v=1}^{V} KL(P^r\|Q^v)$$

The framework of DEMVC:



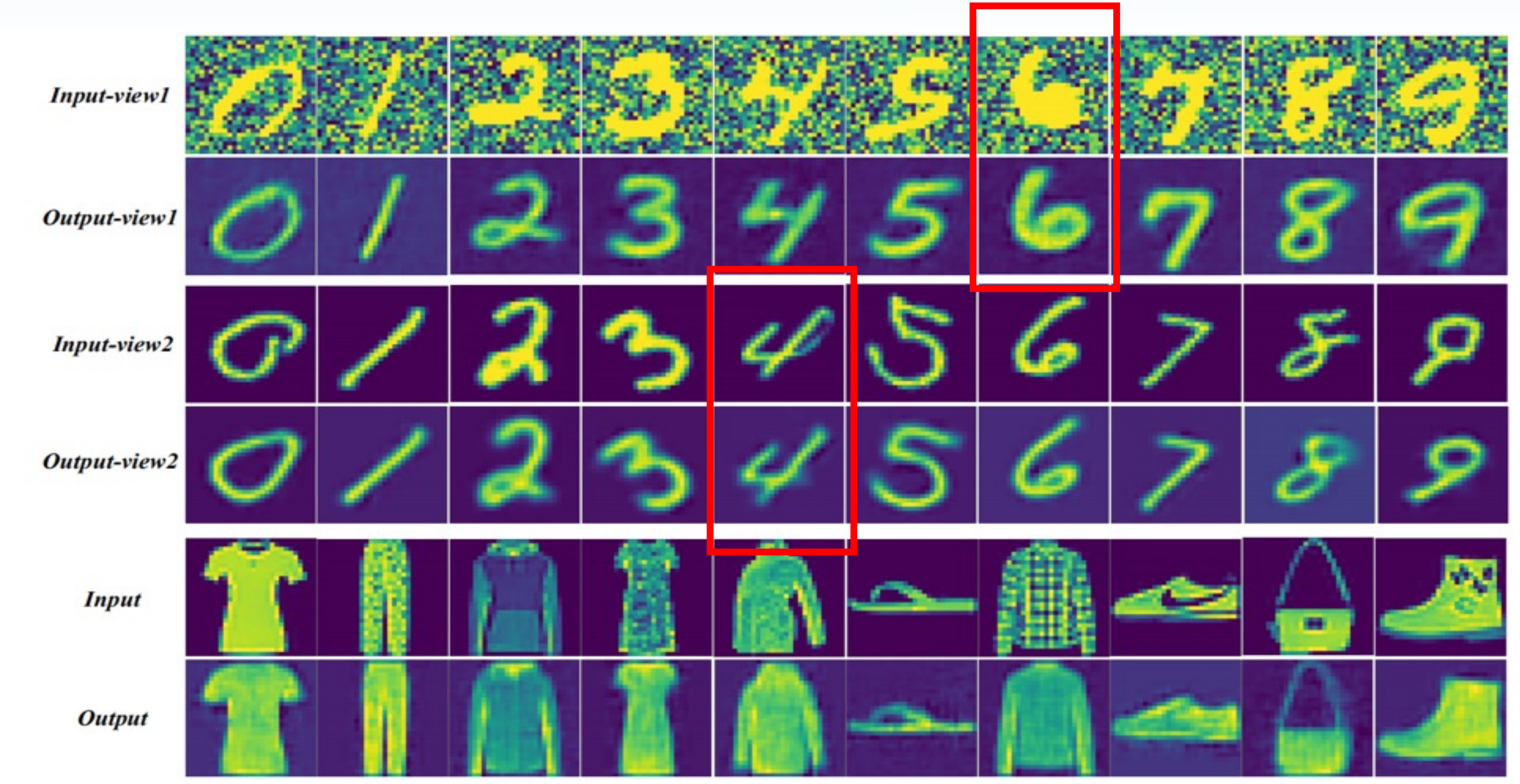
$$L^{V_s}_{k-means} = \sum_{i=1}^{N}\sum_{j=1}^{K}\left\| \mathbf{z}_i^{V_s} - \mathbf{c}_j^{V_s} \right\|^2$$

DEMVC applies k-means on one view (the referred view) to obtain an auxiliary target distribution.

This auxiliary distribution is used to refine the deep autoencoders and clustering soft assignments for all views.

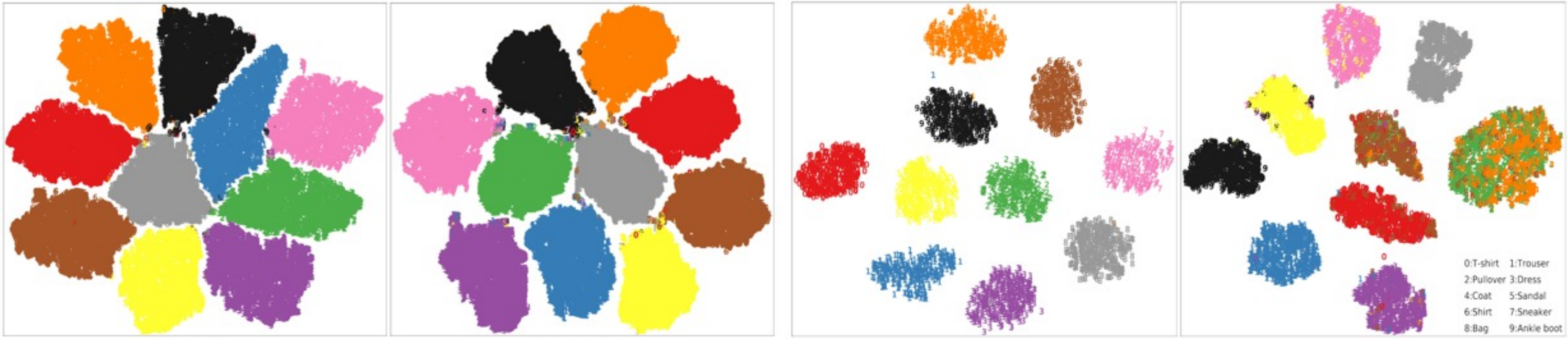
Each view will become the referred view in sequence to ensure that the multi-view clustering takes full advantage of all views.

Visualization of inputs and outputs:



This indicates DEMVC's good representation capability of sample features and reconstruction capability, which is the premise to improve clustering performance.

Visualization by *t*-sne:



(a) NoisyMnist (b) RotatingMnist (c) USPS (d) Fashion-10K

Linear complexity makes it easy to handle large-scale dataset (e.g. 70,000 examples)

Quantitative comparison:

Multi-view methods

Methods	Noisy-Rotated		MNIST-USPS		Caltech101-20	
	ACC	NMI	ACC	NMI	ACC	NMI
DCCA (ICML 2013)	97.00†	92.00†	97.42*	93.60*	-	46.48*
DCCAE (ICML 2015)	**97.50**†	**93.40**†	**98.00***	**94.70***	-	45.56*
DiMSC (CVPR 2015)	/	/	48.34*	36.02*	-	29.05*
LMSC (CVPR 2017)	/	/	78.60*	78.49*	-	**63.55***
BMVC (TPAMI 2018)	85.61	81.48	88.68	89.93	47.44	60.28
COMIC (ICML 2019)	/	/	47.76	64.16	62.32	60.56
DEMVC (ours)	**99.87**	**99.53**	**99.83**	**99.49**	**56.05**	**68.87**

Methods	MNIST-10K		Fashion-10K	
	ACC	NMI	ACC	NMI
DEC (ICML 2016)	83.41	79.22	56.70	61.29
IDEC (IJCAI 2017)	84.25	82.77	57.43	61.55
DCN (ICML 2017)	83.31‡	80.86‡	58.67‡	59.40‡
DEC-DA (ACML 2018)	97.93	95.81	53.55	59.91
k-SCN (ACCV 2018)	87.14‡	78.15‡	63.78‡	62.04‡
NCSC (ICML 2019)	94.09‡	86.12‡	72.14‡	68.60‡
DEMVC-2 views (ours)	**99.87**	**99.60**	**84.75**	**87.14**
DEMVC-3 views (ours)	**99.99**	**99.96**	**78.99**	**90.88**

Single-view methods

Our Recent Work

Self-supervised Discriminative Feature Learning for Multi-view Clustering (SDMVC, TKDE, 2022)

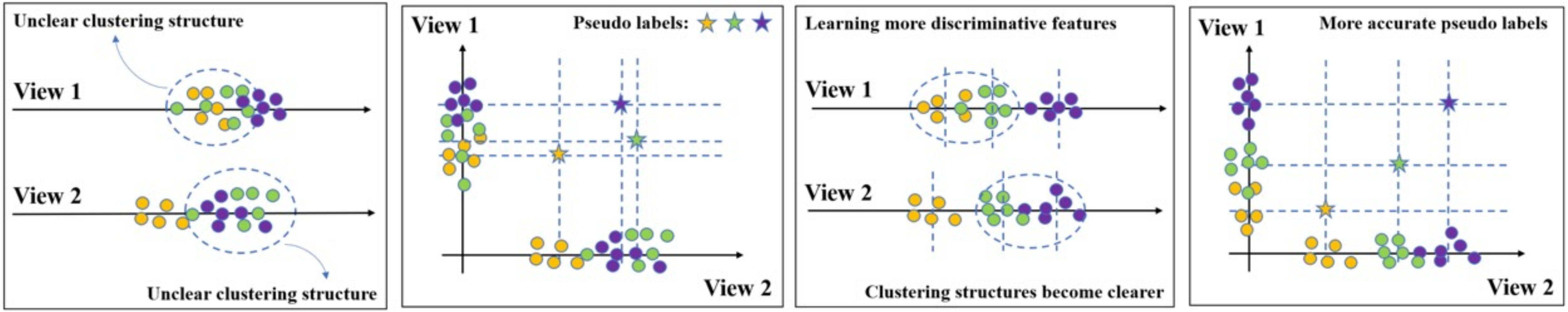
Self-Supervised Discriminative Feature Learning for Deep Multi-View Clustering

Jie Xu, Yazhou Ren, *Member, IEEE*, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, *Senior Member, IEEE*, Xiaorong Pu, Philip S. Yu, *Fellow, IEEE*, Lifang He, *Member, IEEE*

Abstract—Multi-view clustering is an important research topic due to its capability to utilize complementary information from multiple views. However, there are few methods to consider the negative impact caused by certain views with unclear clustering structures, resulting in poor multi-view clustering performance. To address this drawback, we propose self-supervised discriminative feature learning for deep multi-view clustering (SDMVC). Concretely, deep autoencoders are applied to learn embedded features for each view independently. To leverage the multi-view complementary information, we concatenate all views' embedded features to form the global features, which can overcome the negative impact of some views' unclear clustering structures. In a self-supervised manner, pseudo-labels are obtained to build a unified target distribution to perform multi-view discriminative feature learning. During this process, global discriminative information can be mined to supervise all views to learn more discriminative features, which in turn are used to update the target distribution. Besides, this unified target distribution can make SDMVC learn consistent cluster assignments, which accomplishes the clustering consistency of multiple views while preserving their features' diversity. Experiments on various types of multi-view datasets show that SDMVC outperforms 14 competitors including classic and state-of-the-art methods. The code is available at https://github.com/SubmissionsIn/SDMVC.

Index Terms—Multi-view clustering, Deep clustering, Unsupervised learning, Self-supervised learning.

How to improve multi-view clustering ?
How to extract complementary information from multiple views?

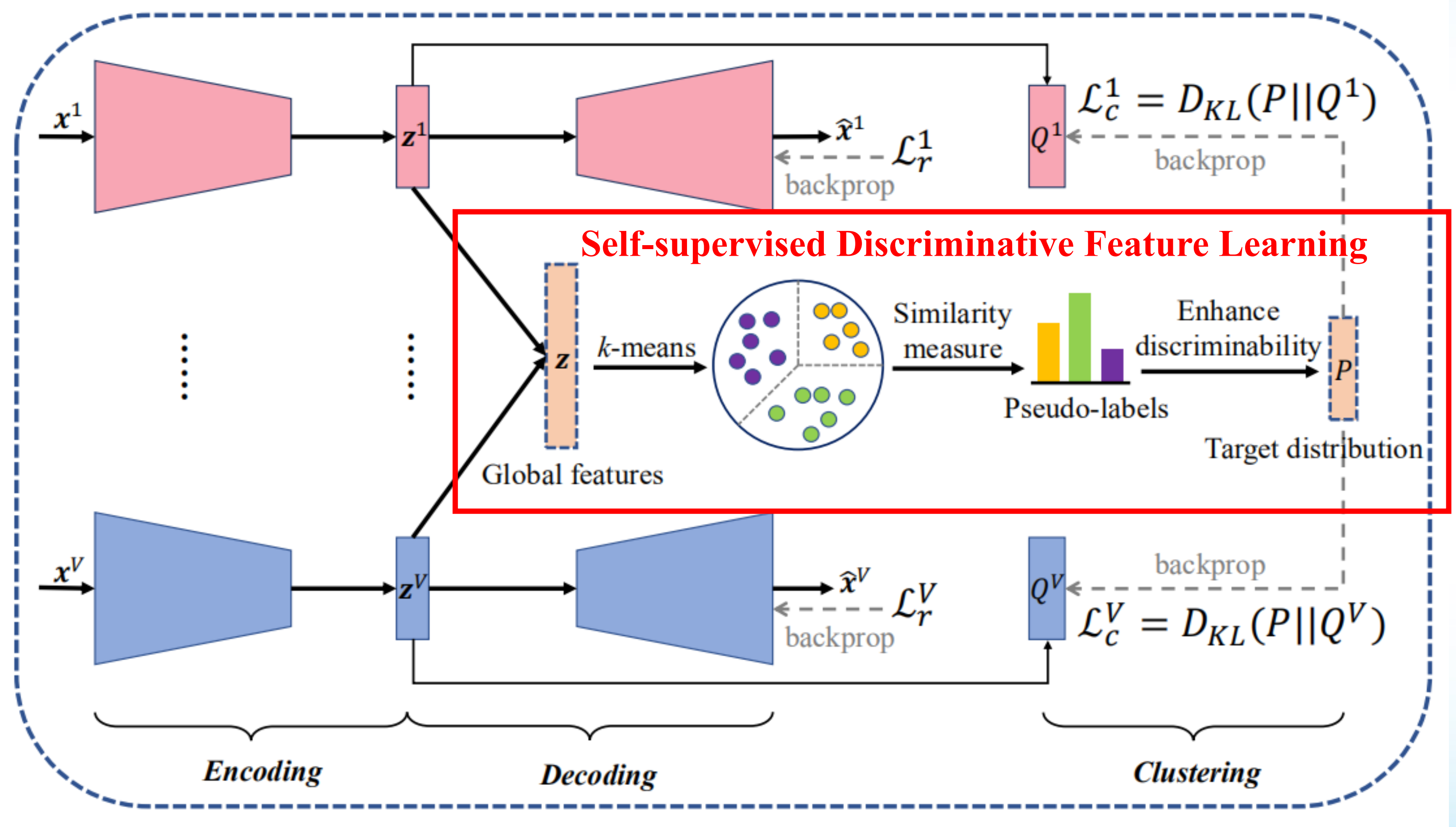


(a) (b) (c) (d)

The discriminability of different views' clustering structures is different.

The clustering structures of different views can correct each other.

The framework of SDMVC:



Global feature

$$\mathbf{z}_i = [\mathbf{z}_i^1; \mathbf{z}_i^2; \ldots; \mathbf{z}_i^V] \in \mathbb{R}^{\sum_{v=1}^{V} d_v}$$

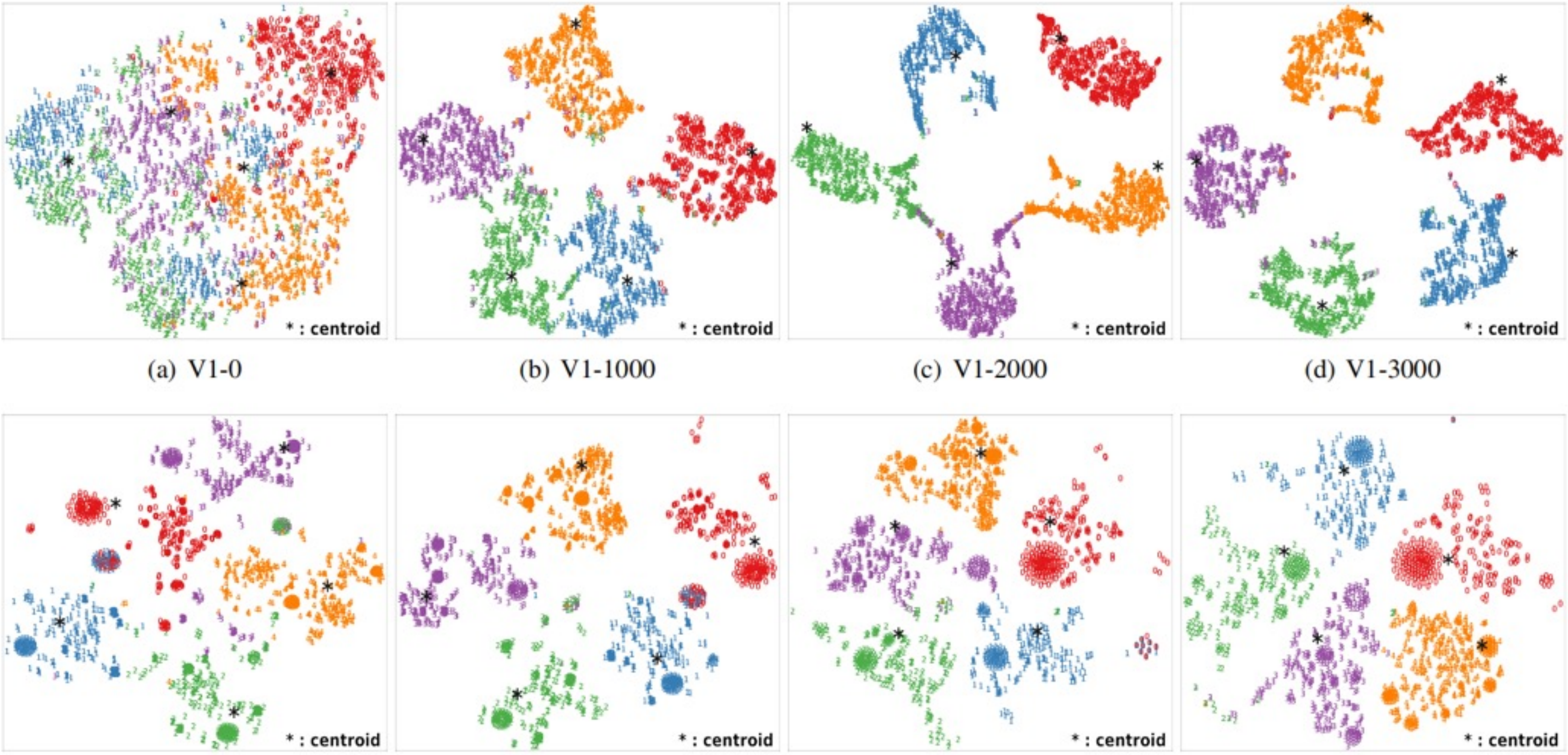
$$\min_{\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_K} \sum_{i=1}^{N} \sum_{j=1}^{K} \|\mathbf{z}_i - \mathbf{c}_j\|^2$$

Soft assignment

$$s_{ij} = \frac{(1 + \|\mathbf{z}_i - \mathbf{c}_j\|^2)^{-1}}{\sum_j (1 + \|\mathbf{z}_i - \mathbf{c}_j\|^2)^{-1}}$$

$$p_{ij} = \frac{(s_{ij})^2 / \sum_i s_{ij}}{\sum_j ((s_{ij})^2 / \sum_i s_{ij})}$$

Visualization the features in learning process (BDGP data set) :



(a) V1-0 (b) V1-1000 (c) V1-2000 (d) V1-3000

(e) V2-0 (f) V2-1000 (g) V2-2000 (h) V2-3000

The clustering structures of embedded features become clearer and clearer while their centroids are gradually separated.

Quantitative comparison:

	MNIST-USPS			Fashion-MV			BDGP			Caltech101-20		
	2 views, $K = 10$ 5,000 examples			3 views, $K = 10$ 10,000 examples			2 views, $K = 5$ 2,500 examples			6 views, $K = 20$ 2,386 examples		
Methods	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI	ACC	NMI	ARI
k-means (1967)	0.7678	0.7233	0.6353	0.7093	0.6561	0.5689	0.4324	0.5694	0.2604	0.4179	0.3351	0.2605
SC (2002)	0.6596	0.5811	0.4864	0.5354	0.5772	0.4261	0.5172	0.5891	0.3156	0.4620	0.4589	0.3933
DEC (2016)	0.7310	0.7146	0.6323	0.6707	0.7234	0.6291	0.9478	0.8662	0.8702	0.4268	0.6251	0.3767
IDEC (2017)	0.7658	0.7689	0.6801	0.6919	0.7501	0.6522	0.9596	0.8940	0.9025	0.4318	0.6253	0.3773
BMVC (2018)	0.8802	0.8945	0.8448	0.7858	0.7488	0.6835	0.3492	0.1202	0.0833	0.5553	0.6203	0.5038
MVC-LFA (2019)	0.7678	0.6749	0.6092	0.7910	0.7586	0.6887	0.5468	0.3345	0.2881	0.4221	0.5846	0.2994
COMIC (2019)	0.4818	0.7085	0.4303	0.5776	0.6423	0.4361	0.5776	0.6423	0.4361	0.6232	0.6838	0.6931
SAMVC (2020)	0.6965	0.7458	0.6090	0.6286	0.6878	0.5665	0.5386	0.4625	0.2099	0.5218	0.5961	0.4653
PVC (2020)	0.6500	0.6118	0.4964	–	–	–	0.4724	0.2972	0.2520	–	–	–
DEMVC (2021)	0.9976	0.9939	0.9948	0.7864	0.9061	0.7793	0.9548	0.8720	0.8901	0.5748	0.6781	0.5068
SDMVC (ours)	**0.9982**	**0.9947**	**0.9960**	**0.8626**	**0.9215**	**0.8405**	**0.9816**	**0.9447**	**0.9548**	**0.7158**	**0.7176**	**0.7265**

The clustering performance of SDMVC is better than other methods.

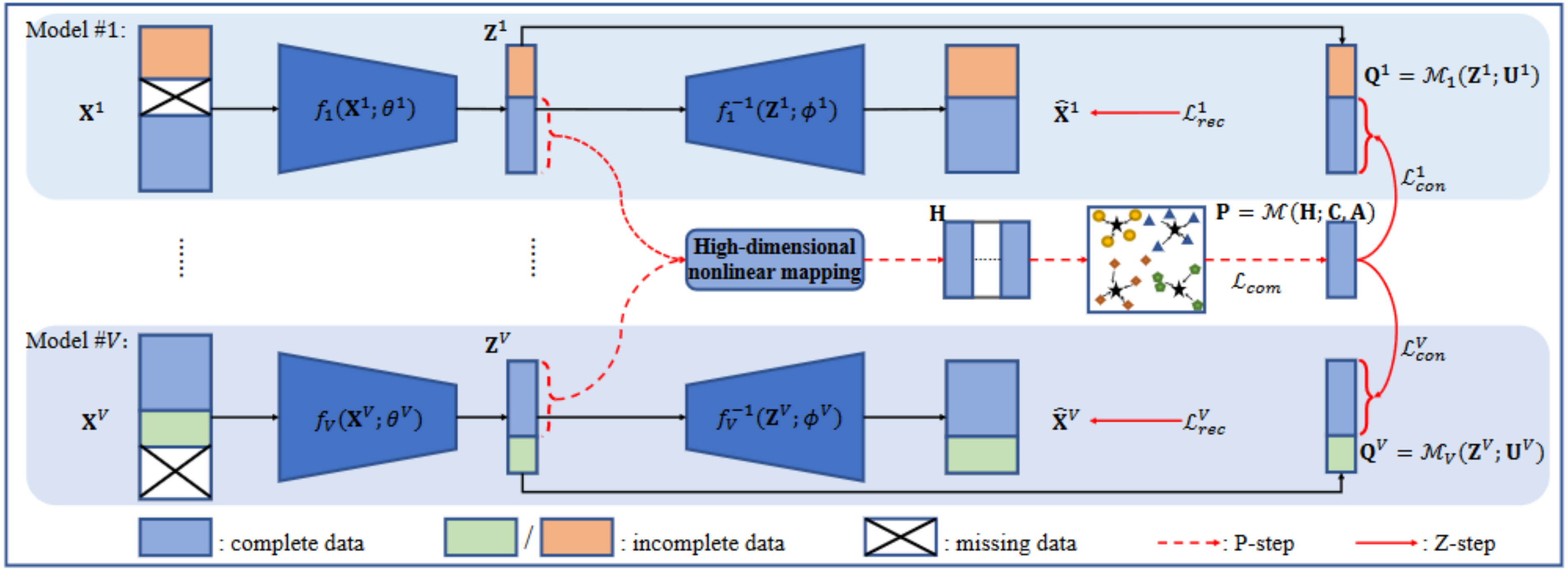
Hot Research Directions

- Incomplete MVC
- Deep hierarchical clustering
- Multi-view graph clustering
- ……

Incomplete MVC

How to handle multi-view data containing missing data in some views?

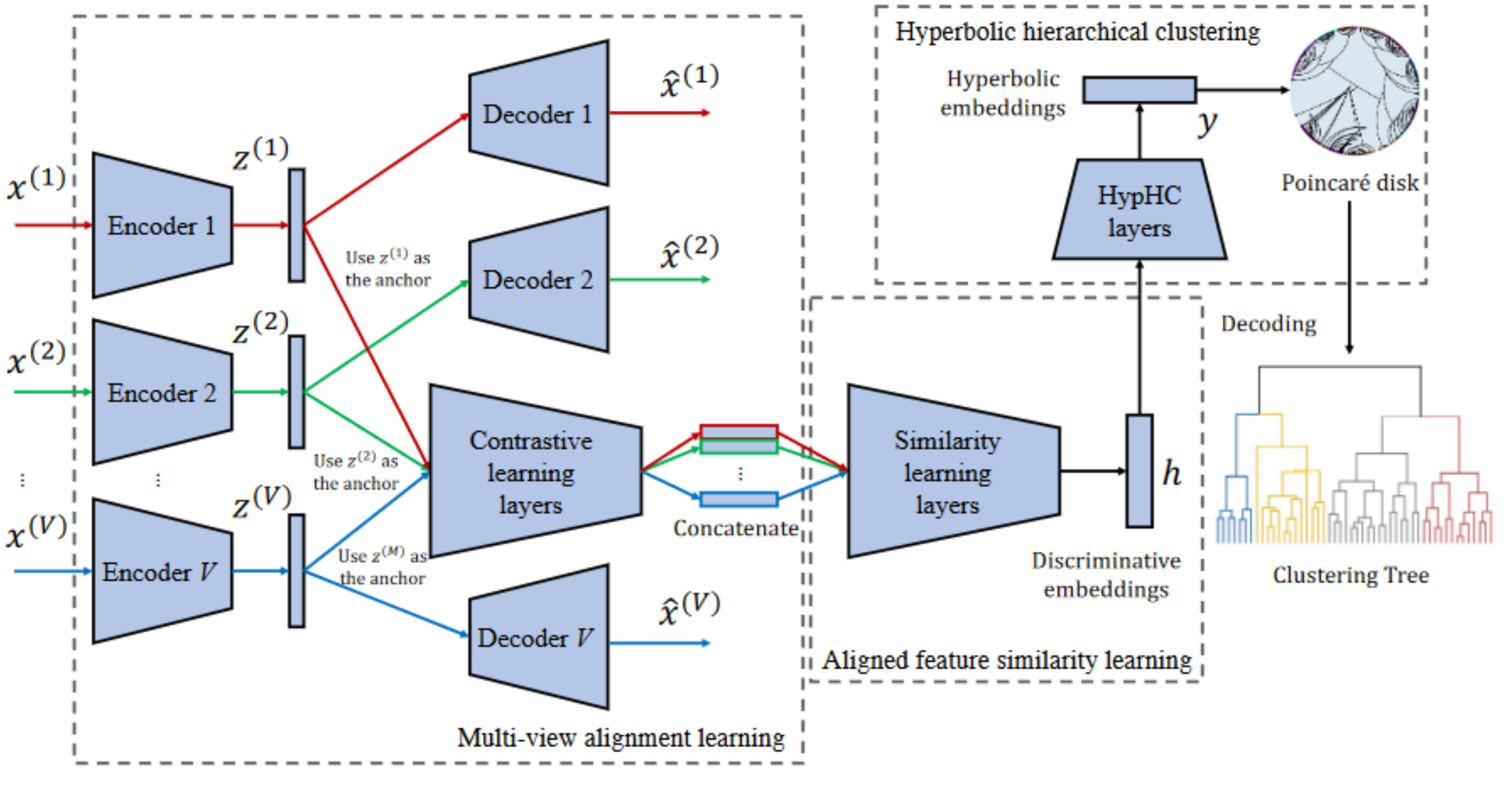
An imputation-free and fusion-free deep IMVC framework.



Jie Xu, Chao Li, **Yazhou Ren** et al. Deep incomplete multi-view clustering via mining cluster complementarity. In *AAAI*, pages 8761–8769, 2022.

Deep hierarchical clustering

How to better understand the hierarchical structure of multi-view data?



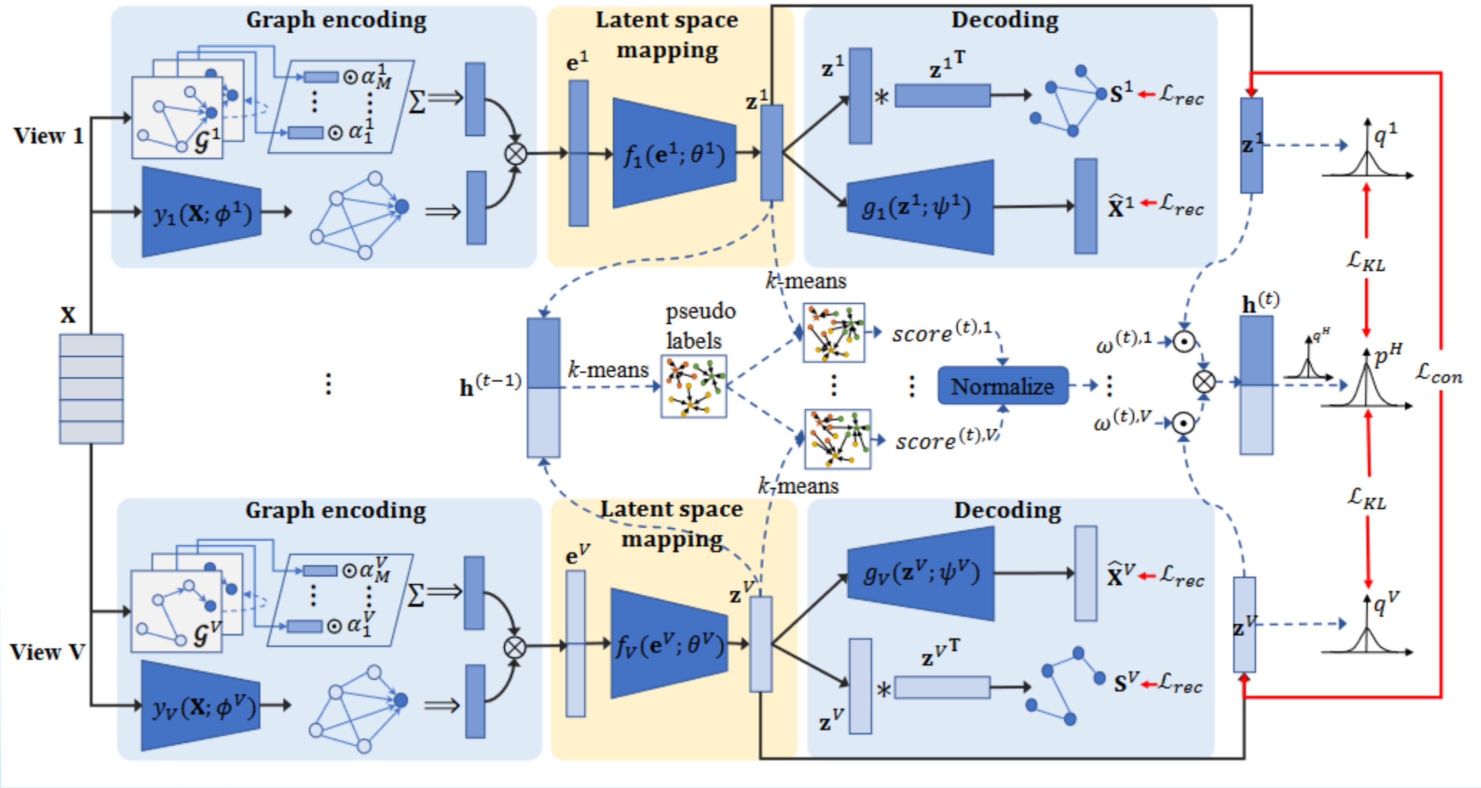
Consists of three parts: multi-view alignment learning, aligned feature similarity learning, and continuous hyperbolic hierarchical clustering.

Fangfei Lin, Bing Bai, Kun Bai, **Yazhou Ren** et al. Contrastive multi-view hyperbolic hierarchical clustering. In *IJCAI*, pages 3250–3256, 2022.

Multi-view graph clustering

How to solve multi-view attributed graph clustering?

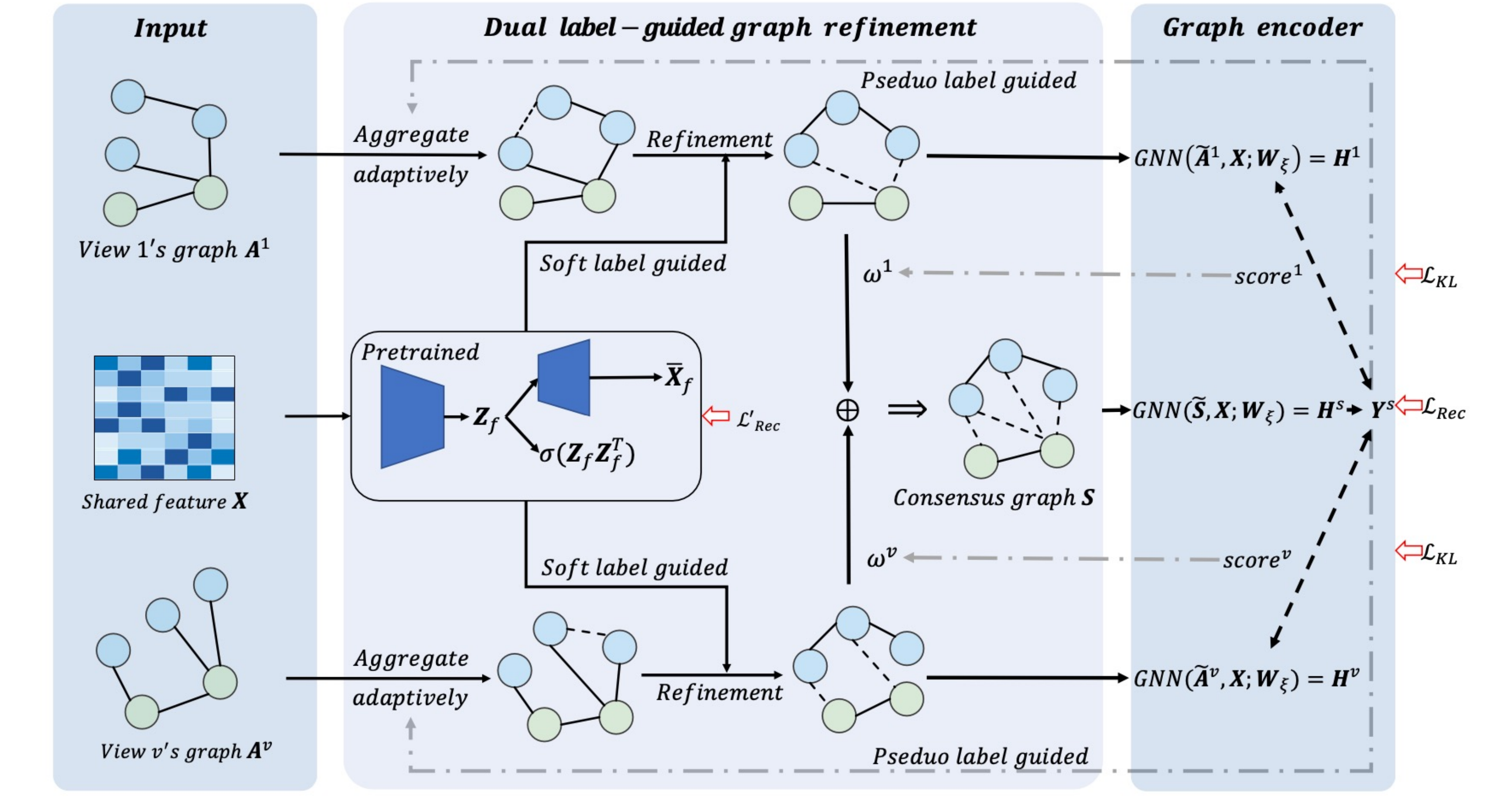
A shared-attribute multi-graph clustering with global self-attention.



Jianpeng Chen, Zhimeng Yang, Jingyu Pu, Xiaorong Pu, **Yazhou Ren** et al. Shared-attribute multi-graph clustering with global self-attention. In *ICONIP*, pages 1–12, 2022.

Multi-view graph clustering

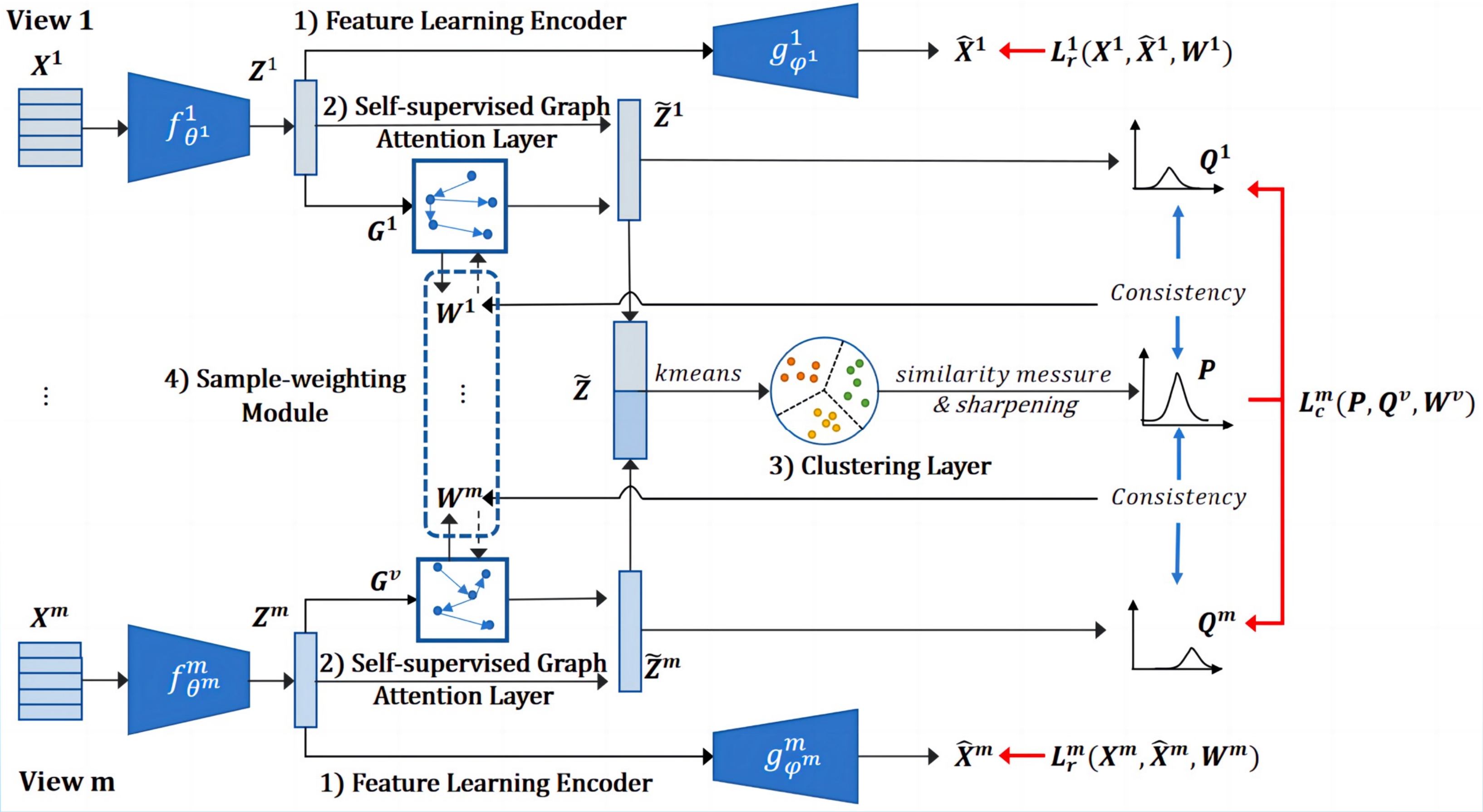
Existing MVGC methods are often sensitive to the given graphs, especially influenced by the low quality graphs, i.e., they tend to be limited by the homophily assumption.



Yawen Ling, Jianpeng Chen, **Yazhou Ren** et al. Dual Label-Guided Graph Refinement for Multi-View Graph Clustering. In AAAI, 2023.

Multi-view graph clustering

MVGC methods are limited due to the insufficient consideration in utilizing the self-supervised information and graph information.



Zongmo Huang, **Yazhou Ren** et al. Self-Supervised Graph Attention Networks for Deep Weighted Multi-View Clustering. In AAAI, 2023.

References

1. Yawen Ling, Jianpeng Chen, **Yazhou Ren***, Xiaorong Pu, Jie Xu, Xiaofeng Zhu, Lifang He. Dual Label-Guided Graph Refinement for Multi-View Graph Clustering. In *AAAI*, 2023.
2. Zongmo Huang, **Yazhou Ren***, Xiaorong Pu, Shudong Huang, Zenglin Xu, Lifang He. Self-Supervised Graph Attention Networks for Deep Weighted Multi-View Clustering. In *AAAI*, 2023.
3. Jie Xu, Huayi Tang, **Yazhou Ren***, Liang Peng, Xiaofeng Zhu, Lifang He. Multi-level Feature Learning for Contrastive Multi-view Clustering. In *CVPR*, pages 16051–16060, 2022.
4. Jie Xu, Chao Li, **Yazhou Ren***, Liang Peng, Yujie Mo, Xiaoshuang Shi, Xiaofeng Zhu. Deep Incomplete Multi-view Clustering via Mining Cluster Complementarity. In *AAAI*, pages 8761–8769, 2022.
5. Jie Xu, **Yazhou Ren***, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, S Yu Philip, Lifang He. Self-Supervised Discriminative Feature Learning for Deep Multi-View Clustering. *TKDE*, pages 1–12, 2022.
6. Jianpeng Chen, Zhimeng Yang, Jingyu Pu, Xiaorong Pu, **Yazhou Ren***, Li Gao, Lifang He. Shared-Attribute Multi-Graph Clustering with Global Self-Attention. In *ICONIP*, pages 1–12, 2022.
7. Jie Xu, **Yazhou Ren***, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, Lifang He. Multi-VAE: Learning Disentangled View-common and View-peculiar Visual Representations for Multi-view Clustering. In *ICCV* , pages 9234-9243, 2021.
8. Zongmo Huang, **Yazhou Ren*,** Xiaorong Pu, Lifang He. Non-Linear Fusion for Self-Paced Multi-View Clustering. In *ACM MM*, pages 3211-3219, 2021.
9. Zongmo Huang, **Yazhou Ren***, Xiaorong Pu, Lili Pan, Dezhong Yao, Guoxian Yu. Dual Self-Paced Multi-view Clustering. *Neural Networks*, 140: 184-192, 2021.
10. Jie Xu, **Yazhou Ren***, Guofeng Li, Lili Pan, Ce Zhu, Zenglin Xu. Deep Embedded Multi-view Clustering with Collaborative Training. *Information Sciences*, 573: 279-290, 2021.

Tutors

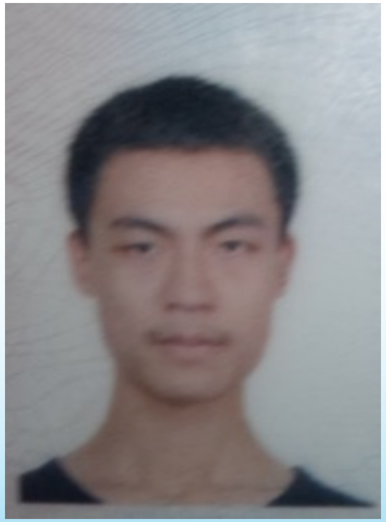
yazhou.ren@uestc.edu.cn

Yazhou Ren

Xiaorong Pu

Students

Jie Xu



Zongmo Huang

Jianpeng Chen

Yawen Ling

Zhimeng Yang

Thanks！

Download slides at: https://yazhou-ren.github.io/